Realizing asynchronous finite-time robust tracking control of switched flight vehicles by using nonfragile deep reinforcement learning

Haoyu Cheng[1]*, Ruijia Song[2], Haoran Li[1], Wencheng Wei[3], Biyu Zheng[1] and Yangwang Fang[1]

[1]Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China, [2]Xi'an Modern Control Technology Research Institute, Xi'an, China, [3]School of Astronautics, Northwestern Polytechnical University, Xi'an, China

In this study, a novel nonfragile deep reinforcement learning (DRL) method was proposed to realize the finite-time control of switched unmanned flight vehicles. Control accuracy, robustness, and intelligence were enhanced in the proposed control scheme by combining conventional robust control and DRL characteristics. In the proposed control strategy, the tracking controller consists of a dynamics-based controller and a learning-based controller. The conventional robust control approach for the nominal system was used for realizing a dynamics-based baseline tracking controller. The learning-based controller based on DRL was developed to compensate model uncertainties and enhance transient control accuracy. The multiple Lyapunov function approach and mode-dependent average dwell time approach were combined to analyze the finite-time stability of flight vehicles with asynchronous switching. The linear matrix inequalities technique was used to determine the solutions of dynamics-based controllers. Online optimization was formulated as a Markov decision process. The adaptive deep deterministic policy gradient algorithm was adopted to improve efficiency and convergence. In this algorithm, the actor−critic structure was used and adaptive hyperparameters were introduced. Unlike the conventional DRL algorithm, nonfragile control theory and adaptive reward function were used in the proposed algorithm to achieve excellent stability and training efficiency. We demonstrated the effectiveness of the presented algorithm through comparative simulations.

KEYWORDS

switched systems, asynchronous switching, deep reinforcement learning, nonfragile control, finite $H\infty$ control

1 Introduction

Aerospace technology has developed rapidly since the 20th century (Wang et al., 2021; Giacomin and Hemerly, 2022; Wang and Xu, 2022). To satisfy the requirements of scientific exploration, military attack, transportation, industrial assistance, and other domains (Bao et al., 2021), flight vehicle systems are becoming increasingly complex (Wu et al., 2021; Lee and Kim, 2022). As an effective tool for the analysis of complex nonlinear systems, switched systems exhibit considerable potential for use in fast time-variation (Hu et al., 2019), full envelope,

structural model mutation (Grigorie et al., 2022), re-modeling (Yue et al., 2019), among others (Chen et al., 2022; Yang et al., 2022).

Switched systems are a critical component of a series of discrete/continuous subsystems, and a switching signal controls the switching logic between these subsystems (Zhang et al., 2019). The switched system exhibits considerable potential for use in theoretical research and engineering applications (Sun and Lei, 2021), such as modeling (Huang et al., 2020), stability analysis (Yang et al., 2020; Zhang and Zhu, 2020), and control problems (Gong et al., 2020; Xiao et al., 2020). The stability analysis of the switched systems is typically used for controller design (Liu et al., 2020). The common Lyapunov function (CLF) method is widely used for stability analysis of arbitrary switching (Jiang et al., 2020). However, ensuring that a CLF is shared by all the subsystems remains challenging. This method is conservative to some degree, which leads to the research is required on the MLF and average dwell time (ADT) methods. Zhao et al. (2012) first studied the stability of the switched systems with ADT switching. In another study, the linear copositive function was extended to the MLF, and the multiple linear copositive Lyapunov function method was used to obtain a sufficient stability criterion for switched systems (Cheng et al., 2017). To obtain tight bounds on the dwell time, the mode-dependent average dwell time (MDADT) method was proposed to overcome the sharing problem of common parameters, and the worst cases were considered in the ADT method. The results were extended to a general case, and the properties of subsystems were considered. Generally, unstable modes may exist during the switching intervals. Therefore, a piecewise multi-Lyapunov function method was proposed in Zhao et al. (2017) for the stability analysis of unstable modes. To avoid dwelling for a long time in subsystems with poor performance and considering the MDADT methods, the slow switching is typically applied to stable modes, and fast switching is applied to unstable modes. Xu et al. (2019) proposed a time-dependent quadratic Lyapunov function method to solve the stability problem with all subsystems unstable. The bounded maximum ADT method is used to obtain the stability conditions of the linear switched system. However, these studies have only focused on infinite-time stability, whereas in finite time, the performance of the systems cannot be guaranteed. Unlike conventional Lyapunov stability, the FTS can achieve superior transient performance in finite time. Wei et al. (2020) proposed a novel MDADT switching signal. The dynamic decomposition technique was used to generate the switching signals, and sufficient conditions for FTS were detailed. For nonlinear switched systems with time delay, the Lyapunov-Razumikhin approach and Lyapunov-Krasovskii function method were used to investigate FTS problems (Wang et al., 2020). Furthermore, the tracking control is widely applied in flight vehicles (Liu et al., 2021). The finite-time tracking control problems in Wang et al. (2017) furthers research on finite-time robust tracking control of switched flight vehicles.

The tracking control problem for uncertain systems is investigated as follows (Liu et al., 2019; Chen et al., 2020; Lu et al., 2022): (1) constant parameter control, such as robust control, proportional integral derivative control, and optimal control, in which the worst case is considered for the bounded uncertainties and disturbances; (2) variable parameter control, such as adaptive and observer-based controls, in which the uncertainties and disturbances are compensated in real time; (3) learning-based control policy, such as reinforcement learning, which compensates uncertainties without prior knowledge

and learns a control law through trial and error. In constant parameter control, the model uncertainties and external disturbances are assumed to be bounded with known boundaries, which result in performance degradation and conservative control laws. The variable parameter control method can be used to mitigate the problem of time-varying uncertainties with unknown boundaries. However, the model uncertainties are assumed to be linearly parameterized with predefined structure and unknown time-varying parameters. The learning-based control method can be used for addressing system uncertainties with unknown boundaries and unknown structures (Yuan et al., 2017). However, this method cannot ensure stability, and computational complexities increase. A novel model-reference adaptive law and a switching logic were developed for uncertain switched systems. Ban et al. (2018) designed an H_∞ controller for polytopic uncertain switched systems. Introducing scalar parameters reduced the conservatism of the linear matrix inequality (LMI) conditions and simultaneously ensured robust H_∞ performance of the system. The problems of nonfragile control for nonlinear switched systems considering actuator failures and parametric uncertainties were studied in Sakthivel et al. (2018). The Lyapunov-Krasovskii function method and ADT approach were used to design a nonfragile reliable sampled-data controller. These studies have focused on control in the ideal environment. However, in practice, because of the limitation of network bandwidth, a network delay and packet loss always exist, which cause inevitable asynchronous switching. Thus, the control switching lags behind state switching. This phenomenon results in performance degradation and instability. Li and Deng (2018) investigated the pth moment exponential input-to-state stability (ISS) of the switched systems with asynchronous switching. The indefinite differentiable Lyapunov function was combined with ADT to establish the ISS conditions of the switched systems with Lévy noise. The conclusion of these results (Zhang and Zhu, 2019) were generalized in Li and Deng (2018), and the ISS problems, stochastic-ISS, and integral-ISS for asynchronously switched systems with asynchronous switching were investigated. Fast ADT switching was introduced to mitigate the increase in the Lyapunov-Krasovskii function when active subsystems matches the controller. However, in most existing results on controller design for flight vehicles, although stability and robustness can be attained, achieving optimal control performance in real-time challenging.

With improvement in the calculating ability of computing devices, machine learning has been widely applied in many fields, including the control field (Cheng and Zhang, 2018; Guo et al., 2019; Gheisarnejad and Khooban, 2021). Xu et al. (2019) proposed a model-driven DDPG algorithm for robotic multi-peg-in-hole assembly to avoid the analysis of the contact model. A feedback strategy and a fuzzy reward function were proposed to improve data efficiency and learning efficiency. In Tailor and Izzo (2019), optimal trajectory for a quadcopter model in two dimensions was investigated. A near-optimal policy was proposed to construct trajectories that satisfy Pontryagin's principle of optimality through supervised learning. With improved aircraft performance, the guidance and control system require rapidity, stability, and robustness. Therefore, deep learning and the exploration of reinforcement learning are an effective solution to this problem, which cannot be solved using conventional control. Cheng et al. (2019) and Gaudet et al. (2020) studied the fuel-optimal landing problems based on DRL. The optional control algorithms were designed considering the uncertainties of environment and

system parameters by using deep neural networks and policy gradient methods to ensure the real-time performance and optimality of the landing mission. The design of the reward function is a critical factor for controller/filter design with DRL. In this method, the final performance of the training networks was determined but not treated satisfactorily. This study is motivated to solve this problem.

However, the methods proposed in Tailor and Izzo (2019) and Gaudet et al. (2020) could not ensure the robustness and stability of the given system. Considering the advantages and limitations of the model-based and model-free methods, we proposed a novel nonfragile DRL for achieving asynchronously finite-time robust tracking control of switched flight vehicles. In this method, the best compromise was realized between system stability, robustness, and rapidity. The intelligent controller based on nonfragile H_∞ control and DRL was proposed to compensate model uncertainties and realize superior control performance. The FTS and finite-time robustness were realized by nonfragile H_∞ control, and the transient performance was optimized by using the adaptive deep deterministic policy gradient (ADDPG) algorithm. Because of the significance of reward function design in the training process, adaptive hyperparameters were introduced to construct a generalized reward function to improve the performance and achieve robustness. Therefore, the contributions of the paper can be summarized as follows:

(1) A novel control structure consisting of dynamics-based and learning-based controllers was proposed for the finite-time tracking control of switched flight vehicles. The robust control is focused on the worst case of uncertainties. However, transient performance is not ensured. The learning-based method, such as DRL, can address uncertainties with unknown boundaries and structures. However, stability is not guaranteed. Compared with the conventional method, in such a design structure, the advantages of both conventional robust control method and pure DRL are combined. The DRL is used to enhance control performance without exploiting their structures or boundaries, and the robustness is guaranteed by using model-based robust control. Thus, an optimal compromise between robustness and dynamic performance was achieved.

(2) The stability and robustness of closed-loop system were guaranteed by using non-fragile control theory. The restricted DRL algorithm was proposed, in which the boundaries of scheduling intervals were predefined. The scheduling of parameters can be viewed as the perturbation of parameters within a given interval. Compared with pure DRL, the proposed method improved training efficiency and ensured stability of the closed-loop system.

(3) The adaptive reward functions were proposed to realize rapid training convergence. The reward functions were crucial for the DRL algorithm. The conventional method of reward functions typically depends on the designing experience of the researchers, which degrade training efficiency and result in trial and error. Therefore, in the proposed method, adaptive factors for reward functions were used to improve training efficiency.

The rest of the paper is organized as follows. In Section 2, the structure of intelligent switched controllers is presented. In Section 3, the finite-time robust tracking control algorithm using DRL and H_∞

control was proposed. A numerical example is provided in Section 4. Finally, Section 5 presents the summary and directions for future studies.

2 Problem statement

The HiMAT vehicle was studied, which is an unmanned flight vehicle. Its nonlinear model can be described in Eq. (1).

$$
\begin{cases}
m_{\mathrm{f}}\dot{v} = T\cos\alpha - D - m_{\mathrm{f}}g\sin(\theta - \alpha) \\
\dot{\alpha} = \dfrac{-T\sin\alpha}{m_{\mathrm{f}}v} - \dfrac{L}{m_{\mathrm{f}}v} + q + \dfrac{g\cos(\theta - \alpha)}{v} \\
\dot{\varphi} = q \\
I_{\mathrm{y}}\dot{q} = M_{\mathrm{yy}} \\
\dot{h} = v\sin(\theta - \alpha) \\
\theta = \varphi - \alpha
\end{cases}
\tag{1}
$$

where m_{f} and v denote the mass and velocity of the flight vehicle, respectively. Here, α, θ, φ, and q are the attack angle, flight path angle, pitch angle, and pitch rate, respectively. Furthermore, M_{yy} and I_{y} are the pitch moment and the moment of inertia about the pitch axis, respectively. Furthermore, g denotes the gravitational constant. The notations of T, D, and L represent the thrust, drag force, and lift force, which can be expressed as follows:

$$
\begin{cases}
T = QSC_{\mathrm{T}} \\
D = QSC_{\mathrm{D}} \\
L = QSC_{\mathrm{L}}
\end{cases}
\tag{2}
$$

where $C_{\mathrm{L}}(\alpha) = C_{\mathrm{L}0}^{0} + C_{\mathrm{L}1}^{\alpha}\alpha$, $C_{\mathrm{D}}(\alpha) = C_{\mathrm{D}1}^{0} + C_{\mathrm{D}2}^{\alpha}\alpha + C_{\mathrm{D}3}^{\alpha^2}\alpha^2$, $C_{\mathrm{T}}(\delta_{\mathrm{c}}) = C_{\mathrm{T}}^{0} + C_{\mathrm{T}}^{\delta_{\mathrm{c}}}\delta_{\mathrm{c}}$, $Q = 0.5\rho V^2$, in which ρ and δ_{c} are the air density and throttle setting.

Based on Jacobian linearization, the nonlinear model of HiMAT vehicle can be converted into the linear model to bridge the connection between complex nonlinear and linear models. Therefore, the longitudinal short-period model of the HiMAT vehicle can be modeled as switched systems as follows:

$$
\begin{cases}
x(k+1) = A_i x(k) + B_i u(k) + D_i \omega(k) \\
y(k) = C_i x(k)
\end{cases}
\tag{3}
$$

where $\boldsymbol{x}(k) = [\alpha \quad q]^T \in \mathbf{R}^x$ is the state vector, $\omega(k) \in R^\omega$ represents the external disturbance that belongs to $L_2[0,\infty)$, $\boldsymbol{u}(k) = [\delta_{\mathrm{e}} \quad \delta_{\mathrm{v}} \quad \delta_{\mathrm{c}}]^T \in \mathbf{R}^u$ with δ_{e}, δ_{v}, and δ_{c} representing the elevator, elevon, and canard deflection, and $\boldsymbol{y}(k) \in \mathbf{R}^y$ denoting the control and output signals. Here, $\sigma(k) = i \to \Omega = \{1, 2, \cdots, n\}$ is the switching function, which is a piecewise continuous constant function. Furthermore, $n > 1$ is the number of subsystems. The characteristic of subsystems is assumed to depend on the switching signal, which are known previously. Here, $\boldsymbol{A}_i, \boldsymbol{B}_i, \boldsymbol{C}_i$, and $\boldsymbol{D}_i$ are system matrices with appropriate dimensions.

In the network environment, because of the limit source of network bandwidth, the packet dropouts should be considered. The packet dropouts are considered in the channel of sensors–controllers to satisfy the Bernoulli distribution (Cheng et al., 2018). Therefore, the measured output is described as follows:

$$\begin{cases} \tilde{y}(k) = \theta(k) y(k) \\ \mathrm{Pr}ob\{\theta(k) = 1\} = E\{\theta(k)\} = \rho \\ \mathrm{Pr}ob\{\theta(k) = 0\} = 1 - E\{\theta(k)\} = 1 - \rho \end{cases} \quad (4)$$

where $\tilde{y}(k)$ is the measured output, $\theta(k)$ represents a stochastic variable satisfying the Bernoulli distribution and takes value of $\{0,1\}$, and $\rho \in [0,1]$ is the probability of packet dropouts.

The control structure of switched flight vehicles to ensure stability and improve transient performance is displayed in Figure 1.

The controller diagram reveals that the controller is composed of two parts:

$$u(k) = u_{\mathrm{n}}(k) + u_{\mathrm{c}}(k) \quad (5)$$

where $u_{\mathrm{n}}(k)$ is the dynamics-based controller, and u_{c} is the learning-based controller, which are developed based on finite-time $H\circ$ control and DRL. The FTS and prescribed attenuation index are ensured by $u_{\mathrm{n}}(k)$, whose parameters can be obtained by the LMI technique. The transient performance is improved by $u_{\mathrm{c}}(k)$, whose parameters are scheduled by the ADDPG algorithm.

The tracking error of the output is defined as $e(k) = r_{\mathrm{c}}(k) - y(k)$, and the objective of tracking control is as follows:

$$\lim_{k \to \infty} e(k) = 0 \quad (6)$$

where $r_c(k)$ denotes the command signal.

We set the integral of tracking error as follows:

$$g(k) = \sum_{l=0}^{k-1} e(l) = \sum_{l=0}^{k-1} (r_{\mathrm{c}}(l) - y(l)) \quad (7)$$

The feedback controller is proposed as follows:

$$u_n(k) = K_{n,i}\tilde{x}(k) = K_{n,1i}x(k) + K_{n,2i}g(k) \quad (8)$$

where $\tilde{x}(k) = \begin{bmatrix} x(k)^T & g(k)^T \end{bmatrix}^T$, $K_{\mathrm{n},1i}$ and $K_{\mathrm{n},2i}$ are the gain matrices to be determined.

Nominal controller parameters $K_{\mathrm{n},1i}$ and $K_{\mathrm{n},2i}$ can be designed by the $H\circ$ control, the variation internal of learning-based controller $u_{\mathrm{c}}(k)$ in subsystem i can be perceived as the additional bounded uncertainties of the dynamics-based controller. Thus, the parameters vary in the interval $\left[K_{n,i} - \Delta \underline{K}_{c,i}, K_{n,i} + \Delta \bar{K}_{c,i} \right]$ and the stability of learning-based controller can be analyzed by using nonfragile control theory. Here, $\varnothing K_{c,i}$ is defined as the additional compensation to obtain the actual gain matrices as follows:

$$K_i = K_{n,i} + \Delta K_{c,i} \quad (9)$$

where $\varnothing \underline{K}_{c,i}$ and $\varnothing \bar{K}_{c,i}$ denote the lower and upper bounds of $\varnothing K_{c,i}$; set $\Delta K_{c,i} = M_i F_i N_i$, M_i and N_i are known parameters with appropriate dimensions, and F_i are uncertain matrices satisfying the following equation:

$$F_i^T F_i \,''\, I \quad (10)$$

Remark 1: The model of flight vehicle can be given based on switched systems. The variation of states in the envelope can be viewed as the switching between subsystems. The tracking controller is composed of two parts, namely dynamics-based controller $u_{\mathrm{n}}(k)$, which is developed based on finite-time $H\circ$ control to ensure stability and prescribed attenuation index; the learning-based controller $u_{\mathrm{c}}(k)$, which is based on
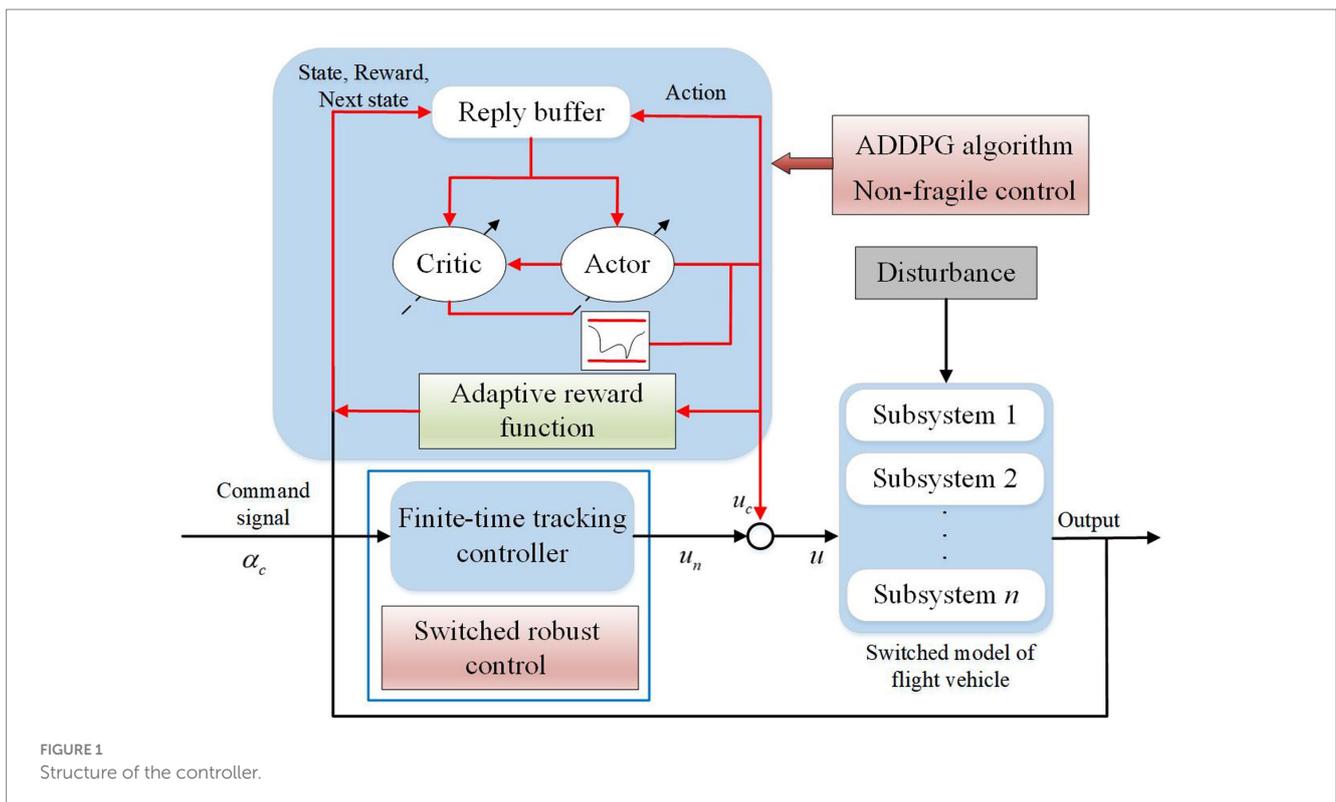

FIGURE 1
Structure of the controller.

ADDPG algorithm to achieve superior performance in real time. The output of $u_c(k)$ varies in the neighbor interval of $u_n(k)$ with given bounds. Therefore, the nonfragile control can be used to ensure the stability of $u_c(k)$. As mentioned, ensuring stability, robustness, and optimal performance simultaneously remains difficult. To improve training efficiency, adaptive factors for reward functions were applied in DDPG algorithm. With inspiration from the achievements in the DDPG algorithm and robust control, the advantages of model-based method ($H\circ$ control) and model-free method (DRL) were considered the problem.

Remark 2: The compensation of learning-based controller is considered as an additional gain value on the controller parameters with known bounds, which can be predefined and can presented by M_i and N_i. The optimal control policy can be realized in the scheduling interval by using the ADDPG algorithm.

The switching of controller always lags the switching of system mode because of packet dropouts. The ith subsystem is assumed to be activated at k_i, and the controller of ith subsystem is activated at $k_i + \Delta_i$, where $\varnothing_i$ denotes the length of unmatched periods. The condition in which unmatched and matched periods exist simultaneously is called asynchronous switching. The Lyapunov-like function decreases in matched periods and increases in unmatched periods with bounded rates, where a_i are introduced to represent the decreasing rate in matched periods, and b_i represent the increasing rate in unmatched periods. The increasing coefficients of the Lyapunov-like function at switching instants are set to be $\varnothing_i$.

For proof, the following assumptions are introduced.

Assumption 1 (Cheng et al., 2017): For given positive constant N_f, the time-varying exogenous disturbance $\omega(k)$ satisfies the following equation:

$$\sum_{k=0}^{N_f} \omega^T(k)\omega(k) \leq \bar{\omega} \qquad (11)$$

where $\bar{\omega}$ is the upper bound of external disturbance.

Assumption 2 (Cheng et al., 2017): The maximum number of consecutive data missing is set to be N_1, and the maximum probability of data missing is set to be $\bar{\rho}$.

According to the aforementioned statement, the closed-loop switched systems can be described as follows:

$$\begin{cases} \tilde{x}(k+1) = \tilde{A}_{ii}\tilde{x}(k) + \tilde{\theta}(k)\tilde{A}_{1i}\tilde{x}(k) + \tilde{B}_i\tilde{\omega}(k) \\ e(k) = \tilde{C}_i\tilde{x}(k) + \tilde{\theta}(k)\tilde{C}_{1i}\tilde{x}(k) + \tilde{D}_i\tilde{\omega}(k) \end{cases}, \forall k \in [k_i + \Delta_i, k_{i+1}) \ (12)$$

$$\begin{cases} \tilde{x}(k+1) = \tilde{A}_{ij}\tilde{x}(k) + \tilde{\theta}(k)\tilde{A}_{1i}\tilde{x}(k) + \tilde{B}_i\tilde{\omega}(k) \\ e(k) = \tilde{C}_i\tilde{x}(k) + \tilde{\theta}(k)\tilde{C}_{1i}\tilde{x}(k) + \tilde{D}_i\tilde{\omega}(k) \end{cases}, \forall k \in [k_i, k_i + \Delta_i)$$

where $\tilde{\omega}(k) = \begin{bmatrix} \omega^T(k) & r_c^T(k) \end{bmatrix}^T$, $\tilde{A}_{ii} = \begin{bmatrix} A_i + B_iK_{1i} & B_iK_{2i} \\ -\rho C_i & I \end{bmatrix}$,

$\tilde{A}_{1i} = \begin{bmatrix} 0 & 0 \\ -C_i & 0 \end{bmatrix}$, $\tilde{A}_{ij} = \begin{bmatrix} A_i + B_iK_{1j} & B_iK_{2j} \\ -\rho C_i & I \end{bmatrix}$, $\tilde{B}_i = \begin{bmatrix} D_i & 0 \\ 0 & I \end{bmatrix}$

$\tilde{C}_i = \begin{bmatrix} -\rho C_i & 0 \end{bmatrix}$, $\tilde{C}_{1i} = \begin{bmatrix} -C_i & 0 \end{bmatrix}$, $\tilde{D}_i = \begin{bmatrix} 0 & I \end{bmatrix}$, $\tilde{\theta}(k) = \theta(k) - \rho$.

Furthermore, the definitions of finite-time stable, finite-time boundedness, and finite-time H_∞ performance for switched systems are expressed as follows:

Definition 1 (Wei et al., 2020): For given appropriate constant positive matrix R_s, positive constants $c_1 > 0$, $c_2 > 0$, and N_f with

$c_1 < c_2$, respectively. The switched systems in Eq. (12) with $u(k) \equiv 0$ and $\omega(k) \equiv 0$ are finite-time stable with respect to (c_1, c_2, N_f, R_s) if Eq. (13) holds.

$$x^T(k_0)R_s x(k_0) \leq c_1 \Rightarrow x^T(k)R_s x(k) \leq c_2, \forall k \in \{1, 2, \ldots, N_f\} \ (13)$$

Definition 2 (Wei et al., 2020): For given appropriate constant positive matrix R_s, constants $c_1 > 0$, $c_2 > 0$, $\bar{\omega}$, and N_f with $c_1 < c_2$, respectively. The switched system in Eq. (12) is finite-time bounded (FTB) with respect to $(c_1, c_2, \bar{\omega}, N_f, R_s)$ such that the following expression holds:

$$x^T(k_0)R_s x(k_0) \leq c_1 \Rightarrow x^T(k)R_s x(k) \leq c_2, \forall k \in \{1, 2, \ldots, N_f\} \ (14)$$

where the external disturbance satisfies Assumption 1.

Definition 3 (Wei et al., 2020): For a given appropriate constant positive matrix R_s, constants $c_1 > 0$, $c_2 > 0$ for $\bar{\omega}$ and N_f with $c_1 < c_2$. The system in Eq. (12) exhibits finite-time $H\circ$ performance γ_d if the system is FTB and satisfies the following expression:

$$\sum_{s=0}^{N_f} e^T(s)e(s) \leq \gamma_d^2 \sum_{s=0}^{N_f} \tilde{\omega}^T(s)\tilde{\omega}(s) \qquad (15)$$

Thus, the main purposes of controller design is to ensure that the switched system is FTS with prescribed $H\circ$ performance γ_d with respect to $(c_1, c_2, \bar{\omega}, N_f, R_s)$, which is equivalent to design the robust controller, such that the following condition is satisfied:

1. The switched systems in Eq. (12) is FTB.
2. For given constant $\gamma_d > 0$, the system in Eq. (12) satisfies Eq. (15) under zero-initial situation for all external disturbance satisfies Eq. (11).

Based on the structure of control diagram, the design process is categorized into two steps:

Step 1: The scheduling interval of control parameters can be assumed to be the uncertain compensation of dynamics-based controller. Considering the controller uncertainties and asynchronous switching caused by packet dropouts, the finite-time $H\circ$ controllers are derived as dynamics-based controller according to nonfragile control theory and finite-time robust control theory in terms of LMI.

Step 2: The variations of controller parameters are assumed to be the action, and the dynamic model of flight vehicles is assumed to be the environment. The DRL algorithm was introduced to derive the learning-based controller to realize optimal control policy, in which the ADDPG algorithm was proposed as the model-free method in the actor–critic framework.

3 Main results

A dynamics-based controller was proposed to ensure stability and a prescribed performance index. The ADDPG algorithm was developed to realize performance and ensure controllers can adaptively schedule parameters.

3.1 Dynamics-based controller design

Definition 4 (Zhao et al., 2017): Given switching signal $\sigma(k)$ and any $0 \,''\, k_1 \,''\, k_2$, let $N_{\sigma i}(k_1, k_2)$ be the activated number of ith subsystem over the time interval (k_1, k_2). Here, $T_i(k_1, k_2)$ denotes the total running time of ith subsystem during the time interval (k_1, k_2), $i \in \mathbb{C}$. If positive numbers N_{0i} and τ_{ai}, exist such that

$$N_{\sigma i}(k_1, k_2) \leq N_{0i} + \frac{T_i(k_1, k_2)}{\tau_{ai}} \qquad (16)$$

then τ_{ai} is called the MDADT and N_{0i} is called the mode-dependent chatter bounds.

Lemma 1 (Cheng et al., 2017): For given symmetric matric Y, matrices F, $\tilde{M}$, and $\tilde{N}$, if a scalar $\varepsilon > 0$ exists such that

$$Y + \varepsilon^{-1} \tilde{M}^T \tilde{M} + \varepsilon \tilde{N}^T \tilde{N} < 0 \qquad (17)$$

then we can obtain the following:

$$Y + \tilde{M}^T F \tilde{N} + \tilde{N}^T F^T \tilde{M} < 0 \qquad (18)$$

where F satisfies $F^T F < I$.

Lemma 2 (Aristidou et al., 2014): For given matrix Q, which satisfies

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \qquad (19)$$

where $Q_{12} = Q_{21}^T$, and Q_{11} and Q_{22} are invertible matrices. Then we can conclude that the following three conditions are equivalent, which is called Schur Complement.

(1) $Q < 0$;
(2) $Q_{11} < 0, Q_{22} - Q_{12}^T Q_{11}^T Q_{12} < 0$;
(3) $Q_{22} < 0, Q_{11} - Q_{12} Q_{11}^T Q_{12}^T < 0$.

Theorem 1: Given system Eq. (12) and constant scalars $0 < a_i < 1$, $b_i > 0$, $\mu_i \geq 1$, $\gamma > 0$, if matrices $S_i > 0$, $S_j > 0$, $S_{ij} > 0$, and W_i, $\forall i, j \in \Omega, i \neq j$, then the following expression is obtained:

$$S_j \leq \mu_i S_i \qquad (20)$$

$$\begin{bmatrix} -S_i & 0 & \tilde{A}_{ii} S_i & \tilde{B}_i \\ * & -S_i & \tilde{\rho} \tilde{A}_{1i} S_i & 0 \\ * & * & -(1-a_i) S_i & 0 \\ * & * & * & -\gamma^2 W_i \end{bmatrix} < 0 \qquad (21)$$

$$\begin{bmatrix} -S_{ij} & 0 & \tilde{A}_{ij} S_j & \tilde{B}_i \\ * & -S_{ij} & \tilde{\rho} \tilde{A}_{1i} S_j & 0 \\ * & * & -(1+b_i)\left(S_{ij} - S_j - S_j^T\right) & 0 \\ * & * & * & -\gamma^2 W_i \end{bmatrix} < 0 \qquad (22)$$

then the switched system in Eq. (12) is FTB with respect to $(c_1, c_2, \bar{\omega}, N_f, R_s)$ if the MDADT satisfies the following equations:

$$\tau_{ai} \geq \tau_{ai}^* = \frac{N_f \ln \mu_i + N_f \Delta_i \ln \hbar_i}{\ln \dfrac{c_2}{\eta_1} - \ln \left[\dfrac{c_1}{\eta_2} + \gamma^2 \tilde{b}_{\max}^{N_f} \eta_3 \bar{\omega} \right] - N_f \ln \tilde{a}_i} \qquad (23)$$

$$\left(\frac{c_1}{\eta_2} + \gamma^2 \tilde{b}_{\max}^{N_f} \eta_3 \bar{\omega} \right) \tilde{a}_i^{N_f} \leq \frac{c_2}{\eta_1} \qquad (24)$$

where $\eta_1 = \max_{i \in \Omega}\left(\lambda_{\max}\left(\bar{S}_i\right), \lambda_{\max}\left(\bar{S}_{ij}\right) \right)$,

$\eta_2 = \min_{i \in \Omega}\left(\lambda_{\min}\left(\bar{S}_i\right), \lambda_{\min}\left(\bar{S}_{ij}\right) \right)$, $\eta_3 = \lambda_{\max}(W_i)$, $\tilde{\rho} = \sqrt{\rho(1-\rho)}$,

$\bar{S}_i = R_s^{1/2} S_i R_s^{1/2}$, $\bar{S}_j = R_s^{1/2} S_j R_s^{1/2}$, $\tilde{a}_i = 1 - a_i$, $\tilde{b}_i = 1 + b_i$, $\hbar_i = \tilde{b}_i / \tilde{a}_i$,

$\tilde{b}_{\max} = \max\left\{ \tilde{b}_i \right\}$.

Proof: For positive constant k, we define $k_0 = 0$ and $k_1, k_2, \ldots, k_i, k_{i+1}, \ldots k_n$ as the switching instants over the interval $[0, k]$, suppose the following Lyapunov functions exist:

$$V_i(k) = \tilde{x}^T(k) P_i \tilde{x}(k) \qquad (25)$$

Class κ_∞ functions exist as follows:

$$\kappa_{1i}\left|\tilde{x}(k)\right| \leq V_i(k) \leq \kappa_{2i}\left|\tilde{x}(k)\right| \qquad (26)$$

$$\Delta V_i(k) \leq \begin{cases} -a_i V_i(k), \forall k \in [k_0, k_1] \cup [k_i + \Delta_i, k_{i+1}] \\ b_i V_i(k), \forall k \in [k_i, k_i + \Delta_i] \end{cases} \qquad (27)$$

$$V_i(k) \leq \mu_i V_j(k) \qquad (28)$$

where $P_i > 0$ are Lyapunov matrices.

Define $\xi(k) = \begin{bmatrix} \tilde{x}^T(k) & \tilde{\omega}^T(k) \end{bmatrix}^T$, and combining with Eqs. (12) and (27), we can obtain the following expression:

$$\begin{aligned} &\Delta V_i(k) + a_i V_i(k) - \gamma^2 \tilde{\omega}^T(k) W_i \tilde{\omega}(k) \\ &= V_i(k+1) - V_i(k) + a_i V_i(k) - \gamma^2 \tilde{\omega}^T(k) W_i \tilde{\omega}(k) \\ &= E\left\{ \tilde{x}^T(k+1) P_i \tilde{x}(k+1) \right\} - (1-a_i) \tilde{x}^T(k) P_i \tilde{x}(k) \\ &\quad - \gamma^2 \tilde{\omega}^T(k) W_i \tilde{\omega}(k) \\ &= \xi^T(k) \left\{ \begin{bmatrix} \tilde{A}_{ii}^T \\ \tilde{B}_i^T \end{bmatrix} P_i \begin{bmatrix} \tilde{A}_{ii} & \tilde{B}_i \end{bmatrix} + \tilde{\rho}^2 \begin{bmatrix} \tilde{A}_{1i}^T \\ 0 \end{bmatrix} P_i \begin{bmatrix} \tilde{A}_{1i} & 0 \end{bmatrix} \right. \\ &\quad \left. + \begin{bmatrix} -(1-a_i) P_i & 0 \\ 0 & -\gamma^2 W_i \end{bmatrix} \right\} \xi(k) \\ &= \xi^T(k) \Pi_{ii} \xi(k) \end{aligned} \qquad (29)$$

$$\Delta V_i(k) - b_i V_i(k) - \gamma^2 \tilde{\omega}^T(k) W \tilde{\omega}(k)$$
$$= V_i(k+1) - V_i(k) - b_i V_i(k) - \gamma^2 \tilde{\omega}^T(k) W_i \tilde{\omega}(k)$$
$$= E\left\{\tilde{x}^T(k+1) P_{ij} \tilde{x}(k+1)\right\} - (1+b_i) x^T(k) P_{ij} x(k)$$
$$\quad - \gamma^2 \tilde{\omega}^T(k) W \tilde{\omega}(k)$$
$$= \xi^T(k) \left\{ \begin{bmatrix} \tilde{A}_{ij}^T \\ \tilde{B}_i^T \end{bmatrix} P_j \begin{bmatrix} \tilde{A}_{ij} & \tilde{B}_i \end{bmatrix} + \tilde{\rho}^2 \begin{bmatrix} \tilde{A}_{1i}^T \\ 0 \end{bmatrix} P_i \begin{bmatrix} \tilde{A}_{1i} & 0 \end{bmatrix} \right.$$
$$\left. + \begin{bmatrix} -(1+b_i) P_{ij} & 0 \\ 0 & -\gamma^2 W_i \end{bmatrix} \right\} \xi(k)$$
$$= \xi^T(k) \Pi_{ij} \xi(k) \tag{30}$$

where

$$\Pi_{ii} = \begin{bmatrix} -P_i & 0 & P_i \tilde{A}_{ii} & P_i \tilde{B}_i \\ * & -P_i & \tilde{\rho} P_i \tilde{A}_{1i} & 0 \\ * & * & -(1-a_i) P_i & 0 \\ * & * & * & -\gamma^2 W_i \end{bmatrix},$$

$$\Pi_{ij} = \begin{bmatrix} -P_{ij} & 0 & P_{ij} \tilde{A}_{ij} & P_{ij} \tilde{B}_i \\ * & -P_{ij} & \tilde{\rho} P_{ij} \tilde{A}_{1i} & 0 \\ * & * & -(1+b_i) P_{ij} & 0 \\ * & * & * & -\gamma^2 W_i \end{bmatrix}.$$

Setting $S_i = P_i^{-1}$ and performing a congruence transformation to Eqs. (29), (30) by matrices $diag\{S_i, S_i, S_i, I\}$ and $diag\{S_{ij}, S_{ij}, S_j, I\}$, we can obtain the following expression:

$$\begin{bmatrix} -S_i & 0 & \tilde{A}_{ii} S_i & \tilde{B}_i \\ * & -S_i & \tilde{\rho} \tilde{A}_{1i} S_i & 0 \\ * & * & -(1-a_i) S_i & 0 \\ * & * & * & -\gamma^2 W_i \end{bmatrix} < 0 \tag{31}$$

$$\begin{bmatrix} -S_{ij} & 0 & \tilde{A}_{ij} S_j & \tilde{B}_i \\ * & -S_{ij} & \tilde{\rho} \tilde{A}_{1i} S_j & 0 \\ * & * & -(1+b_i) S_j^T S_{ij}^{-1} S_j & 0 \\ * & * & * & -\gamma^2 W_i \end{bmatrix} < 0 \tag{32}$$

The inequality $(S_{ij} - S_j)^T S_{ij} (S_{ij} - S_j) \geq 0$ implies the following:

$$S_{ij} - S_j - S_j^T \geq -S_j^T S_{ij}^{-1} S_j \tag{33}$$

We can conclude that Eq. (31) is equivalent to Eq. (21) and Eq. (32) is equivalent to Eq. (22), such that the following expression holds true:

$$\Delta V_i(k) \leq \begin{cases} -a_i V_i(k) + \gamma^2 \tilde{\omega}^T(k) \tilde{\omega}(k), \forall k \in [k_0, k_1] \cup [k_i + \Delta_i, k_{i+1}] \\ b_i V_i(k) + \gamma^2 \tilde{\omega}^T(k) \tilde{\omega}(k), \forall k \in [k_i, k_i + \Delta_i] \end{cases} \tag{34}$$

Combining Eqs. (25), (26), (28), (34), we can obtain the following equations by iteration operation:

With the definitions of η_1 and η_2, we have the following expression:

$$V_{\sigma(k)}(k) \leq \tilde{a}_{\sigma(k_i+\Delta_i)}^{k-k_i-\Delta_i} V_{\sigma(k_i+\Delta_i)}(k_i+\Delta_i) + \gamma^2 \sum_{s=k_i+\Delta_i}^{k-1} \tilde{a}_{\sigma(k_i)}^{k-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s)$$

$$\leq \mu_{\sigma(k_i)} \tilde{a}_{\sigma(k_i+\Delta_i)}^{k-k_i-\Delta_i} \tilde{b}_{\sigma(k_i)}^{\Delta_i} V_{\sigma(k_i)}(k_i+\Delta_i)_- + \gamma^2 \sum_{s=k_i+\Delta_i}^{k-1} \tilde{a}_{\sigma(k_i)}^{k-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s) + \gamma^2 \sum_{s=k_i}^{k_i+\Delta_i-1} \tilde{a}_{\sigma(k_i)}^{k-k_i-\Delta_i} \tilde{b}_{\sigma(k_i)}^{k_i+\Delta_i-s-1} \tilde{\omega}^T(s) W_i \tilde{\omega}(s)$$

$$= \mu_{\sigma(k_i)} \tilde{a}_{\sigma(k_i+\Delta_i)}^{k-k_i} \hbar_{\sigma(k_i)}^{\Delta_i} V_{\sigma(k_i)}(k_i+\Delta_i)_- + \gamma^2 \sum_{s=k_i+\Delta_i}^{k-1} \tilde{a}_{\sigma(k_i)}^{k-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s) + \gamma^2 \sum_{s=k_i}^{k_i+\Delta_i-1} \tilde{a}_{\sigma(k_i)}^{k-s-1} \hbar_{\sigma(k_i)}^{k_i+\Delta_i-s-1} \tilde{\omega}^T(s) W_i \tilde{\omega}(s)$$

$$\leq \mu_{\sigma(k_i)} \tilde{a}_{\sigma(k_i)}^{k-k_i} \hbar_{\sigma(k_i)}^{\Delta_i} V_{\sigma(k_{i-1})}(k_i) + \gamma^2 \sum_{s=k_i}^{k-1} \tilde{b}_{\sigma(k_i)}^{k-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s)$$

$$\leq \mu_{\sigma(k_i)} \tilde{a}_{\sigma(k_i)}^{k-k_i} \hbar_{\sigma(k_i)}^{\Delta_i} \left[\mu_{\sigma(k_{i-1})} \tilde{a}_{\sigma(k_{i-1})}^{k_i-k_{i-1}} \hbar_{\sigma(k_{i-1})}^{\Delta_{i-1}} V_{\sigma(k_{i-1})}(k_{i-1}) + \gamma^2 \sum_{s=k_{i-1}+\Delta_{i-1}}^{k_i-1} \tilde{b}_{\sigma(k_{i-1})}^{k_i-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s) \right] + \gamma^2 \sum_{s=k_i}^{k-1} \tilde{b}_{\sigma(k_i)}^{k-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s) \tag{35}$$

$$\leq \cdots$$

$$\leq \tilde{a}_i^{T_i(k_0,k)} \tilde{a}_{i-1}^{T_{i-1}(k_0,k)} \cdots \tilde{a}_1^{T_1(k_0,k)} \hbar_i^{\Delta_i N_{\sigma,i}(k_0,k)} \hbar_{i-1}^{\Delta_{i-1} N_{\sigma,i-1}(k_0,k)} \cdots \hbar_1^{\Delta_1 N_{\sigma,1}(k_0,k)} \mu_i^{N_{\sigma,i}(k_0,k)} \mu_{i-1}^{N_{\sigma,i-1}(k_0,k)} \cdots \mu_1^{N_{\sigma,1}(k_0,k)} V_{\sigma(k_0)}(k_0)$$

$$+ \tilde{a}_i^{T_i(k_0,k)} \cdots \tilde{a}_1^{T_1(k_0,k)} \hbar_i^{\Delta_i N_{\sigma,i}(k_0,k)} \cdots \hbar_1^{\Delta_1 N_{\sigma,1}(k_0,k)} \mu_i^{N_{\sigma,i}(k_0,k)} \mu_{i-1}^{N_{\sigma,i-1}(k_0,k)} \cdots \mu_1^{N_{\sigma,1}(k_0,k)} \gamma^2 \sum_{s=k_0+\Delta_0}^{k_1-1} \tilde{b}_{\sigma(k_0)}^{k_1-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s)$$

$$+ \cdots + \gamma^2 \sum_{s=k_i+\Delta_i}^{k-1} \tilde{b}_i^{k-1-s} \tilde{\omega}^T(s) W_i \tilde{\omega}(s)$$

$$\leq \prod_{i=1}^n \mu_i^{N_{\sigma,i}(k_0,k)} \tilde{a}_i^{T_i(k_0,k)} \hbar_i^{\Delta_i N_{\sigma,i}(k_0,k)} V_{\sigma(k_0)}(k_0) + \gamma^2 \tilde{b}_{max}^{N_f} \sum_{s=k_0}^{k-1} \left(\prod_{i=1}^n \mu_i^{N_{\sigma,i}(s,k)} \tilde{a}_i^{T_i(s,k)} \hbar_i^{\Delta_i N_{\sigma,i}(s,k)} \right) \tilde{\omega}^T(s) W_i \tilde{u}(s)$$

$$\leq \prod_{i=1}^n \mu_i^{N_{\sigma,i}(k_0,k)} \tilde{a}_i^{T_i(k_0,k)} \hbar_i^{\Delta_i N_{\sigma,i}(k_0,k)} \left[V_{\sigma(k_0)}(k_0) + \gamma^2 \tilde{b}_{max}^{N_f} \eta_3 \bar{\omega} \right]$$

$$V_{\sigma(k_0)}(k_0) = \tilde{x}^T(k_0)P_i\tilde{x}(k_0)$$
$$= \tilde{x}^T(k_0)R_s^{1/2}\overline{S}_i^{-1}R_s^{1/2}\tilde{x}(k_0) \le \frac{1}{\eta_2}\tilde{x}^T(k_0)R_s\tilde{x}(k_0) \quad (36)$$

Moreover, using $\tilde{x}^T(k_0)R_s\tilde{x}(k_0) \le c_1$, we can obtain the following expression:

$$\tilde{x}^T(k)R_s\tilde{x}(k) \le \eta_1 V_{\sigma(k)}(k) \le \eta_1 \left[\prod_{i=1}^{n}\mu_i^{N_{\sigma,i}(k_0,k)}\tilde{a}_i^{T_i(k_0,k)}\hbar_i^{\Delta_i N_{\sigma,i}(k_0,k)}\right]$$

$$\left(\frac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\eta_3\overline{\omega}\right)$$

$$\le \exp\left\{\ln\eta_1 + \sum_{i=1}^{n}\begin{bmatrix}\dfrac{T_i(k_0,k)}{\tau_{ai}}\ln\mu_i + T_i(k_0,k)\ln\tilde{a}_i \\ +\Delta_i\dfrac{T_i(k_0,k)}{\tau_{ai}}\ln\hbar_i \\ +\ln\dfrac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\eta_3\overline{\omega}\end{bmatrix}\right\}$$

$$\le \exp\left\{\ln\eta_1 + \sum_{i=1}^{n}\begin{bmatrix}\dfrac{\ln\mu_i}{\tau_{ai}} + \ln\tilde{a}_i + \dfrac{\Delta_i\ln\hbar_i}{\tau_{ai}}\end{bmatrix}T_i(k_0,k) \\ +\ln\dfrac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\eta_3\overline{\omega}\right\}$$

Based on Definition 2, we have $\tilde{x}^T(k)R_s\tilde{x}(k) \le c_2$, which can be expressed as follows:

$$\exp\left\{\begin{array}{l}\ln\eta_1 + \sum_{i=1}^{n}\left[\dfrac{\ln\mu_i}{\tau_{ai}} + \ln\tilde{a}_i + \dfrac{\Delta_i\ln\hbar_i}{\tau_{ai}}\right] \\ T_i(k_0,k) + \ln\left[\dfrac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\lambda_3\overline{\omega}\right]\end{array}\right\} \le c_2$$
$$(37)$$
$$\Leftrightarrow \left(\frac{\ln\mu_i + \Delta_i\ln\hbar_i}{\tau_{ai}}\right)N_f \le \ln\frac{c_2}{\eta_1} - \ln\left[\frac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\eta_3\overline{\omega}\right] - N_f\ln\tilde{a}_i$$
$$\Leftrightarrow \tau_{ai} \ge \frac{N_f\ln\mu_i + N_f\Delta_i\ln\hbar_i}{\ln\dfrac{c_2}{\eta_1} - \ln\left[\dfrac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\eta_3\overline{\omega}\right] - N_f\ln\tilde{a}_i}$$

If Eqs. (23), (24) hold, then we can conclude that the following expression is true:

$$\ln\frac{c_2}{\eta_1} - \ln\left[\frac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\eta_3\overline{\omega}\right] - N_f\ln\tilde{a}_i > 0, \tau_{ai}$$
$$\ge \frac{N_f\ln\mu_i + N_f\Delta_i\ln\hbar_i}{\ln\dfrac{c_2}{\eta_1} - \ln\left[\dfrac{c_1}{\eta_2} + \gamma^2\tilde{b}_{\max}^{N_f}\eta_3\overline{\omega}\right] - N_f\ln\tilde{a}_i} \quad (38)$$

which is equivalent to $\tilde{x}^T(k)R_s\tilde{x}(k) \le c_2$. Thus, the switched system in Eq. (12) is FTB, which completes the proof.

The sufficient guarantees of FTS are given in Theorem 1, and the prescribed attenuation performance are discussed in Theorem 2.

Theorem 2: Given system Eq. (12) and constant scalars $0 < a_i < 1$, $b_i > 0$, $\mu_i \ge 1$, $\gamma > 0$, if matrices $\boldsymbol{S}_i > 0$, $\boldsymbol{S}_j > 0$, $\boldsymbol{S}_{ij} > 0$, and $\boldsymbol{W}_i$, $\forall i, j \in \Omega, i \ne j$, such that the following expression holds:

$$\boldsymbol{S}_j \le \mu_i\boldsymbol{S}_i \quad (39)$$

$$\begin{bmatrix} -S_i & 0 & 0 & 0 & \tilde{A}_{ii}S_i & \tilde{B}_i \\ * & -S_i & 0 & 0 & \tilde{\rho}\tilde{A}_{1i}S_i & 0 \\ * & * & -I & 0 & \tilde{C}_iS_i & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i}S_i & 0 \\ * & * & * & * & -(1-a_i)S_i & 0 \\ * & * & * & * & * & -\gamma^2I \end{bmatrix} < 0 \quad (40)$$

$$\begin{bmatrix} -S_{ij} & 0 & 0 & 0 & \tilde{A}_{ij}S_j & \tilde{B}_i \\ * & -S_{ij} & 0 & 0 & \tilde{\rho}\tilde{A}_{1i}S_j & 0 \\ * & * & -I & 0 & \tilde{C}_iS_j & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i}S_j & 0 \\ * & * & * & * & -(1+b_i)(S_{ij}-S_j-S_j^T) & 0 \\ * & * & * & * & * & -\gamma^2I \end{bmatrix} < 0 \quad (41)$$

then the system with MDADT satisfying the following expression is FTS with $H\circ$ performance γ_d with respect to $(0, c_2, \overline{\omega}, N_f, \boldsymbol{R}_s, \gamma_d)$.

$$\tau_{ai} \ge \tau_{ai}^* = \max\left\{\begin{array}{l}\dfrac{N_f\ln\mu_i + N_f\Delta_i\ln\hbar_i}{\ln\dfrac{c_2}{\eta_1} - \ln\left[\gamma^2\tilde{b}_{\max}^{N_f}\overline{\omega}\right] - N_f\ln\tilde{a}_i}, \\ \dfrac{\Delta_i\ln\hbar_i + \ln\mu_i}{-\ln\tilde{a}_i}\end{array}\right\} \quad (42)$$

$$\left(\gamma^2\tilde{b}_{\max}^{N_f}\overline{\omega}\right)\tilde{a}_i^{N_f} \le \frac{c_2}{\eta_1} \quad (43)$$

where $\eta_1 = \max_{i\in\Omega}\left(\lambda_{\max}(\overline{S}_i), \lambda_{\max}(\overline{S}_{ij})\right)$, $\tilde{\rho} = \sqrt{\rho(1-\rho)}$,

$\overline{S}_i = R_s^{1/2}S_iR_s^{1/2}$, $\overline{S}_j = R_s^{1/2}S_jR_s^{1/2}$, $\tilde{a}_i = 1 - a_i$, $\tilde{b}_i = 1 + b_i$, $\hbar_i = \tilde{b}_i/\tilde{a}_i$

$\gamma_d = \gamma\left(\dfrac{\tilde{a}_{\max}}{\tilde{a}_{\min}}\right)^{N_f/2}$, $\tilde{a}_{\max} = \max\{\tilde{a}_i\}$, $\tilde{a}_{\min} = \min\{\tilde{a}_i\}$, $\tilde{b}_{\max} = \max\{\tilde{b}_i\}$.

Proof: The Lyapunov functions are determined in Eq. (25). We can obtain the following equations under the zero-initial condition.

$$\Delta V_i(k) + a_iV_i(k) + E\left\{e^T(k)e(k)\right\} - \gamma^2\tilde{\omega}^T(k)\tilde{\omega}(k)$$
$$= E\left\{\tilde{x}(k+1)^TP_i\tilde{x}(k+1)\right\} - (1-a_i)\tilde{x}^T(k)P_i\tilde{x}(k)$$
$$+ E\left\{e^T(k)e(k)\right\} - \gamma^2\tilde{\omega}^T(k)\tilde{\omega}(k)$$
$$= \xi^T(k)\left\{\begin{bmatrix}\tilde{A}_{ii}^T \\ \tilde{B}_i^T\end{bmatrix}P_i\begin{bmatrix}\tilde{A}_{ii} & \tilde{B}_i\end{bmatrix} + \tilde{\rho}^2\begin{bmatrix}\tilde{A}_{1i}^T \\ 0\end{bmatrix}P_i\begin{bmatrix}\tilde{A}_{1i} & 0\end{bmatrix}\right.$$
$$+ \begin{bmatrix}\tilde{C}_i^T \\ \tilde{D}_i^T\end{bmatrix}\begin{bmatrix}\tilde{C}_i & \tilde{D}_i\end{bmatrix}$$
$$\left.+ \tilde{\rho}^2\begin{bmatrix}\tilde{C}_{1i}^T \\ 0\end{bmatrix}\begin{bmatrix}\tilde{C}_{1i} & 0\end{bmatrix} + \begin{bmatrix}-(1-a_i)P_i & 0 \\ 0 & -\gamma^2I\end{bmatrix}\right\}\xi(k)$$
$$= \xi^T(k)Z_{ii}\xi(k)$$
$$(44)$$

$$\Delta V_i(k) - b_i V_i(k) + E\left\{e^T(k)e(k)\right\} - \gamma^2 \tilde{\omega}^T(k)\tilde{\omega}(k)$$

$$= E\left\{\tilde{x}(k+1)^T P_{ij} \tilde{x}(k+1)\right\} - (1+b_i)\tilde{x}^T(k)P_{ij}\tilde{x}(k)$$

$$+ E\left\{e^T(k)e(k)\right\} - \gamma^2 \tilde{\omega}^T(k)\tilde{\omega}(k)$$

$$= \xi^T(k)\left\{ \begin{bmatrix} \tilde{A}_{ij}^T \\ \tilde{B}_i^T \end{bmatrix} P_{ij} \begin{bmatrix} \tilde{A}_{ij} & \tilde{B}_i \end{bmatrix} + \tilde{\rho}^2 \begin{bmatrix} \tilde{A}_{1i}^T \\ 0 \end{bmatrix} P_{ij} \begin{bmatrix} \tilde{A}_{1i} & 0 \end{bmatrix} \right.$$

$$+ \begin{bmatrix} \tilde{C}_i^T \\ \tilde{D}_i^T \end{bmatrix} \begin{bmatrix} \tilde{C}_i & \tilde{D}_i \end{bmatrix}$$

$$\left. + \tilde{\rho}^2 \begin{bmatrix} \tilde{C}_{1i}^T \\ 0 \end{bmatrix} \begin{bmatrix} \tilde{C}_{1i} & 0 \end{bmatrix} + \begin{bmatrix} -(1+b_i)P_{ij} & 0 \\ 0 & -\gamma^2 I \end{bmatrix} \right\} \xi(k)$$

$$= \xi^T(k)Z_{ij}\xi(k) \tag{45}$$

where

$$Z_{ii} = \begin{bmatrix} -P_i & 0 & 0 & 0 & P_i\tilde{A}_{ii} & P_i\tilde{B}_i \\ * & -P_i & 0 & 0 & \tilde{\rho}P_i\tilde{A}_{1i} & 0 \\ * & * & -I & 0 & \tilde{C}_i & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i} & 0 \\ * & * & * & * & -(1-a_i)P_i & 0 \\ * & * & * & * & * & -\gamma^2 I \end{bmatrix},$$

$$Z_{ij} = \begin{bmatrix} -P_{ij} & 0 & 0 & 0 & P_{ij}\tilde{A}_{ij} & P_{ij}\tilde{B}_i \\ * & -P_{ij} & 0 & 0 & \tilde{\rho}P_{ij}\tilde{A}_{1i} & 0 \\ * & * & -I & 0 & \tilde{C}_i & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i} & 0 \\ * & * & * & * & -(1+b_i)P_{ij} & 0 \\ * & * & * & * & * & -\gamma^2 I \end{bmatrix}$$

The system in Eq. (12) is stable with predefined performance such that

$$Z_{ii} < 0 \tag{46}$$

$$Z_{ij} < 0 \tag{47}$$

Setting $S_i = P_i^{-1}$ and performing congruence transformation to the aforementioned inequalities through $diag\{S_i, S_i, I, I, S_i, I\}$ and $diag\{S_{ij}, S_{ij}, I, I, S_j, I\}$, we can obtain the following expression:

$$\begin{bmatrix} -S_i & 0 & 0 & 0 & \tilde{A}_{ii}S_i & \tilde{B}_i \\ * & -S_i & 0 & 0 & \tilde{\rho}\tilde{A}_{1i}S_i & 0 \\ * & * & -I & 0 & \tilde{C}_iS_i & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i}S_i & 0 \\ * & * & * & * & -(1-a_i)S_i & 0 \\ * & * & * & * & * & -\gamma^2 I \end{bmatrix} < 0 \tag{48}$$

$$\begin{bmatrix} -S_i & 0 & 0 & 0 & \tilde{A}_{ii}S_i & \tilde{B}_i \\ * & -S_i & 0 & 0 & \tilde{\rho}\tilde{A}_{1i}S_i & 0 \\ * & * & -I & 0 & \tilde{C}_iS_i & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i}S_i & 0 \\ * & * & * & * & -(1-a_i)S_i & 0 \\ * & * & * & * & * & -\gamma^2 I \end{bmatrix} < 0 \tag{49}$$

Similar to the transformation in Eq. (33), we can obtain the following expression:

$$S_{ij} - S_j - S_j^T \geq -S_j^T S_{ij}^{-1} S_j \tag{50}$$

With Eqs.(40), (41), we have $Z_{ii} < 0$ and $Z_{ij} < 0$, which implies that the following expression:

$$\Delta V_i(k) \leq \begin{cases} -a_i V_i(k) - E\left\{e^T(k)e(k)\right\} + \gamma^2 \tilde{\omega}^T(k)\tilde{\omega}(k), \\ \forall k \in [k_i + \Delta_i, k_{i+1}) \\ b_i V_i(k) - E\left\{e^T(k)e(k)\right\} + \gamma^2 \tilde{\omega}^T(k)\tilde{\omega}(k), \\ \forall k \in [k_i, k_i + \Delta_i) \end{cases} \tag{51}$$

The following equation can be obtained by setting $\gamma^2 \tilde{\omega}^T(k)W_i\tilde{\omega}(k)$ as $\psi(s) = -E\left\{e^T(k)e(k)\right\} + \gamma^2 \tilde{\omega}^T(k)\tilde{\omega}(k)$. Moreover, the system in Eq. (12) is FTB with respect to $(0, c_2, \bar{\omega}, N_f, R_s)$ by setting $W_i = I$ and $c_1 = 0$.

$$V_i(k) \leq \prod_{i=1}^{n} \mu_i^{N_{\sigma,i}(k_0,k)} \tilde{a}_i^{T_i(k_0,k)} \hbar_i^{\Delta_i N_{\sigma,i}(k_0,k)} V_{\sigma(k_0)}(k_0)$$

$$+ \tilde{b}_{max}^{N_f} \sum_{s=k_0}^{k-1} \left(\prod_{i=1}^{n} \mu_i^{N_{\sigma,i}(s,k)} \tilde{a}_i^{T_i(s,k)} \hbar_i^{\Delta_i N_{\sigma,i}(s,k)} \right) \psi(s) \tag{52}$$

According to $V_{\sigma(k)}(k) \geq 0$ and zero-initial condition, we have the following expression:

$$\sum_{s=k_0}^{k-1} \left(\prod_{i=1}^{n} \mu_i^{N_{\sigma,i}(s,k)} \tilde{a}_i^{T_i(s,k)} \hbar_i^{\Delta_i N_{\sigma,i}(s,k)} \right) \psi(s) \geq 0$$

$$\Leftrightarrow \sum_{s=k_0}^{k-1} \left(\prod_{i=1}^{n} \mu_i^{N_{\sigma,i}(s,k)} \tilde{a}_i^{T_i(s,k)} \hbar_i^{\Delta_i N_{\sigma,i}(s,k)} \right) e^T(k)e(k) \tag{53}$$

$$\leq \gamma^2 \sum_{s=k_0}^{k-1} \left(\prod_{i=1}^{n} \mu_i^{N_{\sigma,i}(s,k)} \tilde{a}_i^{T_i(s,k)} \hbar_i^{\Delta_i N_{\sigma,i}(s,k)} \right) \tilde{\omega}^T(s)\tilde{\omega}(s)$$

Multiplying both sides of Eq. (53) by $\prod_{i=1}^{n} \left(\hbar_i^{\Delta_i} \mu_i\right)^{-N_{\sigma,i}(k_0,k)}$, we obtain the following equation:

$$\sum_{s=k_0}^{k-1} \left(\prod_{i=1}^{n} \mu_i^{-N_{\sigma,i}(k_0,s)} \tilde{a}_i^{T_i(s,k)} \hbar_i^{-\Delta_i N_{\sigma,i}(k_0,s)} \right) e^T(k)e(k)$$

$$\leq \gamma^2 \sum_{s=k_0}^{k-1} \left(\prod_{i=1}^{n} \mu_i^{-N_{\sigma,i}(k_0,s)} \tilde{a}_i^{T_i(s,k)} \hbar_i^{-\Delta_i N_{\sigma,i}(k_0,s)} \right) \tilde{\omega}^T(s)\tilde{\omega}(s) \tag{54}$$

Based on the definition of MDADT and Eq. (42), we have the following:

$$0 \leq N_{\sigma,i}(k_0,s) \leq \frac{T_i(k_0,s)}{\tau_{ai}^*} \leq \frac{-T_i(k_0,s)\ln\tilde{a}_i}{\Delta_i\left(\ln\tilde{b}_i - \ln\tilde{a}_i\right) + \ln\mu_i} \quad (55)$$

Combining with Eqs. (43), (45), we infer the following:

$$\sum_{s=k_0}^{k-1}\left(\prod_{i=1}^{n}\left(\mu_i\hbar_i^{\Delta_i}\right)^{\frac{T_i(k_0,s)\ln\tilde{a}_i}{\Delta_i\left(\ln\tilde{b}_i - \ln\tilde{a}_i\right)+\ln\mu_i}}\tilde{a}_i^{T_i(s,k)}\right)e^T(k)e(k)$$
$$\leq \gamma^2\sum_{s=k_0}^{k-1}\left(\prod_{i=1}^{n}\left(\mu_i\hbar_i^{\Delta_i}\right)^{\frac{T_i(k_0,s)\ln\tilde{a}_i}{\Delta_i\left(\ln\tilde{b}_i - \ln\tilde{a}_i\right)+\ln\mu_i}}\tilde{a}_i^{T_i(s,k)}\right)\tilde{\omega}^T(s)\tilde{\omega}(s) \quad (56)$$

Thus, we have the following equation:

$$\left(\mu_i\hbar_i^{\Delta_i}\right)^{\frac{T_i(k_0,s)\ln\tilde{a}_i}{\Delta_i\left(\ln\tilde{b}_i - \ln\tilde{a}_i\right)+\ln\mu_i}} = \tilde{a}_i^{T_i(k_0,s)} \quad (57)$$

Next, we have the following expression:

$$\sum_{s=k_0}^{k-1}\left(\prod_{i=1}^{n}\tilde{a}_i^{T_i(k_0,k)}\right)e^T(k)e(k) \leq \gamma^2\sum_{s=k_0}^{k-1}\left(\prod_{i=1}^{n}\tilde{a}_i^{T_i(k_0,k)}\right)\tilde{\omega}^T(s)\tilde{\omega}(s) \quad (58)$$

Setting $k-1 = N_f$, we can obtain the following:

$$\sum_{s=k_0}^{N_f}\left(\prod_{i=1}^{n}\tilde{a}_i^{T_i(k_0,k)}\right)e^T(k)e(k) \leq \gamma^2\sum_{s=k_0}^{k-1}\left(\prod_{i=1}^{n}\tilde{a}_i^{T_i(k_0,k)}\right)\tilde{\omega}^T(s)\tilde{\omega}(s)$$
$$\Leftrightarrow \tilde{a}_{\min}^{N_f}\sum_{s=k_0}^{N_f}e^T(k)e(k) \leq \gamma^2\tilde{a}_{\max}^{N_f}\sum_{s=k_0}^{k-1}\tilde{\omega}^T(s)\tilde{\omega}(s)$$
$$\Leftrightarrow \sum_{s=k_0}^{N_f}e^T(k)e(k) \leq \gamma^2\left(\frac{\tilde{a}_{\max}}{\tilde{a}_{\min}}\right)^{N_f}\sum_{s=k_0}^{k-1}\tilde{\omega}^T(s)\tilde{\omega}(s)$$

Therefore, the system Eq. (12) is FTB with given attenuation index $\gamma_d = \gamma\left(\frac{\tilde{a}_{\max}}{\tilde{a}_{\min}}\right)^{N_f/2}$, which completes the proof.

Based on Theorems 1 and 2, the parameters of finite-time tracking controller of switched systems is derived in Theorem 3.

Theorem 3: Given system Eq. (12) and constant scalars $0 < a_i < 1$, $b_i > 0$, $\mu_i \geq 1$, $\gamma > 0$, if positive matrices $\boldsymbol{S}_i$, $\boldsymbol{S}_j$ and $\boldsymbol{S}_{ij}$, $\forall i, j \in \Omega, i \neq j$, exist such that the following holds true:

$$\boldsymbol{S}_j \leq \mu_i\boldsymbol{S}_i \quad (59)$$

$$\begin{bmatrix} \Phi_{ii} & \Sigma_{1i} & \Sigma_{2i} \\ * & \Xi_{1i} & 0 \\ * & * & \Xi_{2i} \end{bmatrix} < 0 \quad (60)$$

$$\begin{bmatrix} \Phi_{ij} & \Sigma_{1j} & \Sigma_{2j} \\ * & \Xi_{1j} & 0 \\ * & * & \Xi_{2j} \end{bmatrix} < 0 \quad (61)$$

System Eq. (12) with MDADT satisfying Eqs. (42), (43) is finite-time stable with predefined attenuation index γ_d with respect to $\left(0, c_2, \bar{\omega}, N_f, \boldsymbol{R}_s, \gamma_d\right)$, and the parameters of robust controller can be expressed as follows:

$$\boldsymbol{K}_{n,1i} = \boldsymbol{U}_{1i}\boldsymbol{S}_{1i}^{-1} \quad (62)$$

$$\boldsymbol{K}_{n,2i} = \boldsymbol{U}_{2i}\boldsymbol{S}_{2i}^{-1} \quad (63)$$

where

$$S_i = \begin{bmatrix} S_{1i} & 0 \\ 0 & S_{2i} \end{bmatrix}, \qquad S_j = \begin{bmatrix} S_{1j} & 0 \\ 0 & S_{2j} \end{bmatrix}$$

$$\Phi_{ii} = \begin{bmatrix} -S_i & 0 & 0 & 0 & \varphi_{ii} & \tilde{B}_i \\ * & -S_i & 0 & 0 & \tilde{\rho}\tilde{A}_{1i}S_i & 0 \\ * & * & -I & 0 & \tilde{C}_iS_i & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i}S_i & 0 \\ * & * & * & * & -(1-a_i)S_i & 0 \\ * & * & * & * & * & -\gamma^2 I \end{bmatrix},$$

$$\Phi_{ij} = \begin{bmatrix} -S_{ij} & 0 & 0 & 0 & \varphi_{ij} & \tilde{B}_i \\ * & -S_{ij} & 0 & 0 & \tilde{\rho}\tilde{A}_{1i}S_j & 0 \\ * & * & -I & 0 & \tilde{C}_iS_j & \tilde{D}_i \\ * & * & * & -I & \tilde{\rho}\tilde{C}_{1i}S_j & 0 \\ * & * & * & * & -(1+b_i)\left(S_{ij}-S_j-S_j^T\right) & 0 \\ * & * & * & * & * & -\gamma^2 I \end{bmatrix},$$

$\Sigma_{1i} = \begin{bmatrix} \tilde{M}_{1i}^T & \tilde{N}_{1i}^T \end{bmatrix}$, $\quad \Sigma_{2i} = \begin{bmatrix} \tilde{M}_{2i}^T & \tilde{N}_{2i}^T \end{bmatrix}$, $\quad \Sigma_{1j} = \begin{bmatrix} \tilde{M}_{1j}^T & \tilde{N}_{1j}^T \end{bmatrix}$,

$\Sigma_{2j} = \begin{bmatrix} \tilde{M}_{2j}^T & \tilde{N}_{2j}^T \end{bmatrix}$, $\quad \Xi_{1i} = diag\{\varepsilon_{1i}, \varepsilon_{1i}\}$, $\quad \Xi_{2i} = diag\{\varepsilon_{2i}, \varepsilon_{2i}\}$,

$\Xi_{1j} = diag\{\varepsilon_{1j}, \varepsilon_{1j}\}$, $\quad \Xi_{2j} = diag\{\varepsilon_{2j}, \varepsilon_{2j}\}$, $\quad \varphi_{ii} = \begin{bmatrix} A_iS_{1i}+B_iU_{1i} & B_iU_{2i} \\ -\rho C_iS_{1i} & S_{2i} \end{bmatrix}$,

$\varphi_{ij} = \begin{bmatrix} A_iS_{2j}+B_iU_{1j} & B_iU_{2j} \\ -\rho C_iS_{1j} & S_{2j} \end{bmatrix}$, $\quad \tilde{M}_{1i} = \begin{bmatrix} M_{1i}^T\tilde{B}_i^T & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$,

$\tilde{N}_{1i} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & N_{1i}S_{1i} & 0 & 0 & 0 \end{bmatrix}$, $\quad \tilde{M}_{2i} = \begin{bmatrix} M_{2i}^T\tilde{B}_i^T & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$,

$\tilde{N}_{2i} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & N_{2i}S_{2i} & 0 & 0 \end{bmatrix}$, $\quad \tilde{M}_{1j} = \begin{bmatrix} M_{1j}^T\tilde{B}_i^T & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$,

$\tilde{N}_{1j} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & N_{1j}S_{1j} & 0 & 0 \end{bmatrix}$, $\quad \tilde{M}_{2j} = \begin{bmatrix} M_{2j}^T\tilde{B}_i^T & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$,

$\tilde{N}_{2j} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & N_{2j}S_{2j} & 0 & 0 \end{bmatrix}$, $\quad \varepsilon_{1i} > 0$, $\quad \varepsilon_{2i} > 0$, $\quad \varepsilon_{2j} > 0$.

Proof: According to Schur Complement (Aristidou et al., 2014) and Lemma 1, we can calculate the following equation:

$$\Phi_{ii} + \tilde{M}_{1i}^T F_{1i} \tilde{N}_{1i} + \tilde{N}_{1i}^T F_{1i}^T \tilde{M}_{1i} + \tilde{M}_{2i}^T F_{2i} \tilde{N}_{2i} + \tilde{N}_{2i}^T F_{2i}^T \tilde{M}_{2i} < \mathbf{0} \quad (64)$$

$$\Phi_{ij} + \tilde{M}_{1j}^T F_{1j} \tilde{N}_{1j} + \tilde{N}_{1j}^T F_{1j}^T \tilde{M}_{1j} + \tilde{M}_{2j}^T F_{2j} \tilde{N}_{2j} + \tilde{N}_{2j}^T F_{2j}^T \tilde{M}_{2j} < \mathbf{0} \quad (65)$$

Let $\boldsymbol{U}_{1i} = \boldsymbol{K}_{\mathrm{n},1i}\boldsymbol{S}_{1i}$, $\boldsymbol{U}_{2i} = \boldsymbol{K}_{\mathrm{n},2i}\boldsymbol{S}_{2i}$, $\boldsymbol{U}_{1j} = \boldsymbol{K}_{\mathrm{n},1j}\boldsymbol{S}_{1j}$, $\boldsymbol{U}_{2j} = \boldsymbol{K}_{\mathrm{n},2j}\boldsymbol{S}_{2j}$, Eq. (60) is equivalent to Eq. (40), and Eq. (61) is equivalent to Eq. (41). Therefore, the parameters of controller can be given according to Eqs. (59)–(61) by solving linear matrix inequalities Eqs. (62), (63).

3.2 Online scheduling based on the ADDPG algorithm

Based on the finite-time H_∞ control, the sufficient conditions to ensure the FTS and prescribed performance are presented. The process of online scheduling can be formulated as the Markov decision process (MDP). Because the control process is a series of continuous decision process, the ADDPG algorithm was proposed based on the actor–critic framework to realize superior control performance of switched flight vehicles.

The DRL is composed of an agent and the interacting environment. At each time, the agent obtains a state s_k, selects an action a_k, and can receive reward r_k and s_{k+1} by interacting with the environment, in which r_k is used to evaluate the performance of state-action pair at the time instant. In this study, the switched tracking controller can be viewed as the agent, whose purpose is maximizing the sum of the expected discounted reward function over a series of future steps:

$$R(k) = r_k + \gamma_d r_{k+1} + \gamma_d^2 r_{k+2} + \cdots + \gamma_d^{K_f - k} r_{K_f} = r_k + \gamma_d R_{k+1} \quad (66)$$

where $\gamma_{\mathrm{d}} \in [0,1]$ denotes the discount factor. Here, K_{f} denotes the terminal step of reinforcement learning. The value of reward depends on the action undertaken and the current state. The action and state are defined as follows:

$$a_k = \begin{bmatrix} \Delta K_{c,1i}(k) & \Delta K_{c,2i}(k) \end{bmatrix} \quad (67)$$

$$s_k = \begin{bmatrix} \alpha(k) & q(k) & r_c(k) & u_n(k) \end{bmatrix} \quad (68)$$

The ADDPG algorithm is provided based on the DDPG algorithm, in which the advantages of both deep Q learning and actor–critic framework are used to realize the optimal action, which is updated in continuous action spaces based on policy gradient theory. The ADDPG algorithm is realized in the following two sections: the action-value in each step is approximated by the critic network $Q\left(s_k, a_k | \varsigma^Q\right)$ with weights ς^Q, the current control policy is obtained by the actor network $\varpi\left(s_k | \varsigma^\varpi\right)$ with weights ς^ϖ. The weights of the critic network are updated by minimizing the loss function, which can be described as follows:

$$L\left(\varsigma^Q\right) = E_{(s,a)}\left[\left(Q\left(s_k, a_k | \varsigma^Q\right) - \bar{y}_k\right)^2\right] \quad (69)$$

where

$$\bar{y}_k = r_k\left(s_k, a_k\right) + \gamma_{\mathrm{d}} Q\left(s_{k+1}, \varpi\left(s_k | \varsigma^\varpi\right) \Big| \varsigma^Q\right).$$

The weights of actor network are updated according to the policy gradient in the following equations:

$$\varsigma^\varpi\left(k+1\right) = \varsigma^\varpi\left(k\right) + L_{\mathrm{an}} \nabla_{\varsigma^\varpi} J \quad (70)$$

$$\nabla_{\varsigma^\varpi} J = \mathrm{E}_\pi\left[\nabla_{\varsigma^\varpi} Q^\pi\left(s_k, \pi\left(s_k | \varsigma^\varpi\right) \Big| \varsigma^Q\right)\Big|_{s=s_k, a=\pi\left(s_k | \varsigma^\varpi\right)}\right]$$
$$= \mathrm{E}_\pi\left[\nabla_{\varsigma^\varpi} Q^\pi\left(s_k, \pi\left(s_k\right) \Big| \varsigma^Q\right) \nabla_{\varsigma^\varpi} \pi\left(s_k | \varsigma^Q\right)\right] \quad (71)$$

where L_{an} is the learning rate of $\varpi\left(s_k | \varsigma^\varpi\right)$.

To overcome the divergence of Q learning, two separated networks were adopted: the actor target network $\varpi'\left(s_k | \varsigma^{\varpi'}\right)$ and the critic target network $Q'\left(s_k, a_k | \varsigma^{Q'}\right)$, the mentioned two networks can update their weights as follows:

$$\varsigma^{\varpi'}\left(k+1\right) = L_{\mathrm{atn}}\varsigma^\varpi\left(k\right) + \left(1 - L_{\mathrm{atn}}\right)\varsigma^{\varpi'}\left(k\right) \quad (72)$$

$$\varsigma^{Q'}\left(k+1\right) = L_{\mathrm{ctn}}\varsigma^Q\left(k+1\right) + \left(1 - L_{\mathrm{ctn}}\right)\varsigma^{Q'}\left(k+1\right) \quad (73)$$

where L_{atn} and L_{ctn} are the learning rates.

Moreover, an exploration noise N_{a} is added to the actor to realize exploration and actual control policy, which is generated by actor and can be rewritten as follows:

$$a_k = \pi\left(s_k | \varsigma^\varpi\right) + N_{\mathrm{a}} \quad (74)$$

Unlike the conventional DDPG algorithm, the adaptive parameters were introduced to achieve superior convergence and robustness, respectively. By introducing robustness as a continuous parameter, the reward function enables the convenient exploration to realize adaptive training. The control policy is used to reduce the tracking error with lower control input and unsaturated actuator, therefore, the reward function depends on the tracking error, amplitude of control signal, and the saturation of actuator, which can be expressed as follows:

$$r_k = g_1 r_{\mathrm{e}1}(k) + g_2 r_{\mathrm{e}2}(k) + g_3 r_{\mathrm{e}3}(k) \quad (75)$$

$$r_{\mathrm{e}1} = \frac{|\upsilon_1 - 2|}{\upsilon_1}\left(\left(\frac{\left(e(k)/l_1\right)}{|\upsilon_1 - 2|} + 1\right)^{\frac{\upsilon_1}{2}} - 1\right) \quad (76)$$

$$r_{\text{e}2} = \frac{|v_2 - 2|}{v_2}\left|\left(\left(\frac{\left(\boldsymbol{u}(k)/l_2\right)}{|v_2 - 2|}+1\right)^{\frac{v_2}{2}}-1\right)\right| \qquad (77)$$

$$r_{\text{e}3} = \begin{cases} \delta_{\text{p}}, & |\boldsymbol{u}(k)| > \overline{u}(k) \\ 0, & |\boldsymbol{u}(k)| \leq \overline{u}(k) \end{cases} \qquad (78)$$

where $r_{\text{e}1}(k)$ represents the reward of tracking error, $r_{\text{e}2}(k)$ denotes the reward of control input, and $r_{\text{e}3}(k)$ is the reward of saturation, respectively. Here, g_1, g_2, and g_3 denote the weights of $r_{\text{e}1}(k)$, $r_{\text{e}2}(k)$, and $r_{\text{e}3}(k)$ in the reward function. Furthermore, v_1, v_2 are the adaptive shape parameters, which determine the robustness of the reward function. $l_1 > 0$ and $l_2 > 0$ are the parameters that controls the size of the quadratic bowl near the origin, respectively. Here, δ_{p} is predefined constant and $\overline{u}(k)$ denotes the upper bound of the actuator. Next, the final reward function $r_{\text{e}1}(k)$ and $r_{\text{e}2}(k)$ with adaptive parameters can be rewritten as follows:

$$r_{\text{e}1}(k) = \begin{cases} \frac{1}{2}\left(e(k)/l_1\right)^2 & \text{if } v_1 = 2 \\ \log\left(\frac{1}{2}\left(e(k)/l_1\right)^2+1\right) & \text{if } v_1 = 0 \\ 1-\exp\left(-\frac{1}{2}\left(e(k)/l_1\right)^2\right) & \text{if } v_1 = -\infty \quad (79) \\ \frac{|v_1-2|}{v_1}\left(\left(\frac{\left(e(k)/l_1\right)}{|v_1-2|}+1\right)^{\frac{v_1}{2}}-1\right) & \text{otherwise} \end{cases}$$

$$r_{\text{e}2}(k) = \begin{cases} \frac{1}{2}\left(u(k)/l_2\right)^2 & \text{if } v_2 = 2 \\ \log\left(\frac{1}{2}\left(u(k)/l_2\right)^2+1\right) & \text{if } v_2 = 0 \\ 1-\exp\left(-\frac{1}{2}\left(u(k)/l_2\right)^2\right) & \text{if } v_2 = -\infty \quad (80) \\ \frac{|v_2-2|}{v_2}\left(\left(\frac{\left(u(k)/l_2\right)}{|v_2-2|}+1\right)^{\frac{v_2}{2}}-1\right) & \text{otherwise} \end{cases}$$

The adaptive updating law of hyper parameters are defined as follows to improve transient performance and robustness of the algorithm:

$$\begin{cases} v_1^{(s)} = \left(v_{1\max} - v_{1\min}\right)\text{sigmoid}\left(v_{1\text{p}}^{(s)}\right) + v_{1\min} \\ v_2^{(s)} = \left(v_{2\max} - v_{2\min}\right)\text{sigmoid}\left(v_{2\text{p}}^{(s)}\right) + v_{2\min} \end{cases} \qquad (81)$$

$$\begin{cases} l_1^{(s)} = \text{softplus}\left(l_{1\text{p}}^{(s)}\right) + l_{1\min} \\ l_2^{(s)} = \text{softplus}\left(l_{2\text{p}}^{(s)}\right) + l_{2\min} \end{cases} \qquad (82)$$

where $v_{1\max}$ and $v_{1\min}$ denote the maximum and minimum values of $v_{1\max}$. Similarly, we can obtain the definitions of $v_{2\max}$, $v_{2\min}$, $l_{1\min}$, and $l_{2\min}$. The length of each segment is determined by training episodes.

Based on the statement, the pseudocode for the ADDPG algorithm proposed in this paper is presented in Algorithm 1.

Algorithm 1. Parameter optimization based on ADDPG
1. Set the variation range of controller parameters.
2. Design the switched tracking controllers for flight vehicles based on Theorem 3.
3. Randomly initialize the weights of networks $Q\left(s_k, a_k\|\varsigma^Q\right)$ and $\varpi\left(s_k\|\varsigma^\varpi\right)$ with ς^ϖ and ς^Q.
4. Initialize the weights of $\varpi'\left(s_k\|\varsigma^{\varpi'}\right)$ and $Q'\left(s_k, a_k\|\varsigma^{Q'}\right)$ with weights $\varsigma^{\varpi'} \leftarrow \varsigma^\varpi, \varsigma^{Q'} \leftarrow \varsigma^Q$.
5. Initialize the replay buffer, $episode = 0$
6. **for** $episode = 1$ to M do
7. Randomly initialize exploration noise N_{a}.
8. Randomly initialize the state vector of the agent with s_1, then the initial observation can be obtained.
9. **for** $t = 1$ to K do
10. Apply action $a_k = \pi\left(s_k\|\varsigma^\varpi\right) + N_{\text{a}}$ to the environment based on the state s_k and uncertain noise.
11. Receive the adaptive reward r_k and the state of next time instant s_{k+1}.
12. Store the variable transition pair in the replay buffer, which consists of s_k, a_k, r_k, and s_{k+1}.
13. Randomly sample a mini-batch of N transition pairs from the replay buffer R.
14. Set $y_k = r_k\left(s_k, a_k\right) + \gamma_p Q\left(s_{k+1}, \varpi\left(s_k\|\varsigma^\varpi\right)\|\varsigma^Q\right)$
15. Update the weights of network $Q\left(s_k, a_k\|\varsigma^Q\right)$ as follows: $L\left(\varsigma^Q\right) = E_{(s,a)}\left[\left(Q\left(s_k, a_k\|\varsigma^Q\right) - y_k\right)^2\right]$
16. Update the weights of network $\varpi\left(s_k\|\varsigma^\varpi\right)$ as follows: $\nabla_{\varsigma^\varpi} J = E_\pi\left[\nabla_{\varsigma^\varpi} Q^\pi\left(s_k, \pi(s_k)\|\varsigma^Q\right)\nabla_{\varsigma^\varpi} \pi\left(s_k\|\varsigma^Q\right)\right]$
17. Update the weights of target networks: $\varsigma^{\varpi'} \leftarrow L_{atn}\varsigma^\varpi + (1-L_{atn})\varsigma^{\varpi'}, \varsigma^{Q'} \leftarrow L_{ctn}\varsigma^Q + (1-L_{ctn})\varsigma^{Q'}$
18. **end for**
19. **end for**

Remark 3: Although the conventional DDPG algorithm can realize parameter optimization (Xu et al., 2019; Gaudet et al., 2020; Gheisarnejad and Khooban, 2021), guaranteeing data efficiency and system stability because it attempts to explore the optimal control policy for all possible action in the action space is difficult. Moreover, the proposed adaptive hyper parameters can increase robustness and achieve generalized case because the reward function determines training performance.

4 Numerical examples

In this study, the HiMAT vehicle is given to validate the proposed method. The three-view drawing and trim condition for operation points can be obtained from the study performed by Wang et al. (2015). The flight condition and the model of longitudinal motion dynamics are given as Wang et al. (2015).

Based on the trim condition within the flight envelope, the longitudinal motion dynamics can be described by switched systems. We set the sampling time $T_s = 0.02$ and obtain the system matrices A_i and B_i, which can be described as follows:

$$A_1 = \begin{bmatrix} 0.9804 & 0.0188 \\ 0.1768 & 0.9720 \end{bmatrix}, B_1 = \begin{bmatrix} -0.0049 & -0.0034 & 0.0007 \\ -0.1579 & -0.0979 & 0.0993 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.9728 & 0.0188 \\ 0.3773 & 0.9622 \end{bmatrix}, B_2 = \begin{bmatrix} -0.0075 & -0.0050 & 0.0014 \\ -0.2941 & -0.1765 & 0.1831 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 0.9688 & 0.0187 \\ 0.4968 & 0.9560 \end{bmatrix}, B_4 = \begin{bmatrix} -0.0096 & -0.0065 & 0.0021 \\ -0.4334 & -0.2895 & 0.2547 \end{bmatrix}$$

$$A_8 = \begin{bmatrix} 0.9766 & 0.0190 \\ 0.3312 & 0.9668 \end{bmatrix}, B_8 = \begin{bmatrix} -0.0077 & -0.0054 & 0.0018 \\ -0.3759 & -0.2798 & 0.2113 \end{bmatrix}$$

$$A_9 = \begin{bmatrix} 0.9725 & 0.0189 \\ 0.3344 & 0.9594 \end{bmatrix}, B_9 = \begin{bmatrix} -0.0099 & -0.0068 & 0.0026 \\ -0.5374 & -0.3793 & 0.2890 \end{bmatrix}$$

$$A_{12} = \begin{bmatrix} 0.9649 & 0.0188 \\ 0.2242 & 0.9509 \end{bmatrix}, B_{12} = \begin{bmatrix} -0.0136 & -0.0094 & 0.0042 \\ -0.9015 & -0.6166 & 0.4367 \end{bmatrix}$$

$$A_{18} = \begin{bmatrix} 0.9657 & 0.0191 \\ -0.9772 & 0.9523 \end{bmatrix}, B_{18} = \begin{bmatrix} -0.0061 & -0.0033 & 0.0023 \\ -0.4595 & -0.2426 & 0.2576 \end{bmatrix}$$

$$A_{19} = \begin{bmatrix} 0.9635 & 0.0192 \\ -1.2369 & 0.9507 \end{bmatrix}, B_{19} = \begin{bmatrix} -0.0066 & -0.0032 & 0.0019 \\ -0.5334 & -0.2569 & 0.2163 \end{bmatrix}$$

The switching of subsystems in the flight envelope is supposed to be 19-18-12-9-8-4-2-1, which is described in Figure 2.

The harmonics wind gust is considered in the paper, which is described in Eq. (83).

$$\begin{cases} p(k+1) = \begin{bmatrix} 0.9922 & 0.1247 \\ -0.1247 & 0.9922 \end{bmatrix} p(k) \\ d(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} p(k) \end{cases} \quad (83)$$

where $p(k)$ represents the state of external disturbance with initial value of $[0.01;0]$. Furthermore, a command filter was provided to improve the performance of the intelligent tracking controller, which can be generated as follows:

$$\begin{cases} J(k+1) = \begin{bmatrix} J_1(k+1) \\ J_2(k+1) \end{bmatrix} \\ = \begin{bmatrix} J_2(k) \\ 2\zeta_n \omega_n \left\{ S_v \left[\dfrac{\omega_n^2}{2\zeta_n \omega_n}(S_a - J_1(k)) \right] - J_2(k) \right\} \end{bmatrix} \\ z(k) = \begin{bmatrix} z_1(k) \\ z_2(k) \end{bmatrix} = \begin{bmatrix} J_1(k) \\ J_2(k) \end{bmatrix} \end{cases} \quad (84)$$

where $J(k)$ denotes the state vector; $z(k)$ represents the output of the filter; ζ_n and ω_n are the damping ratio and band width; S_a and S_v denote the transfer functions of the amplitude limiting and the rate limiting filters.

The parameters of the switched systems are given as $c_1 = 0, c_2 = 1.5$, $N_f = 25$, $\bar{\omega} = 5$, and $R = I$. Compared with the conventional ADT method, tighter bounds on FTS analysis can be obtained. The ADT method can be considered to be a special case of the MDADT method, and we can obtain that $\tau_{ai}^* \le \tau_a^*$, which is illustrated in Table 1. Therefore, the proposed method can realize limited conservative results than the ADT method. We set the probability of data missing as $\rho = 0.95$, the maximum number of consecutive data missing N_1 is set to be 5. Moreover, the matrices U_{1i}, U_{2i}, S_{1i}, and S_{2i} can be solved by Eqs. (62), (63) in Theorem 3. The dynamics-based controller was constructed, and its parameter matrices and structure are given as follows:

$$K_1 = \begin{bmatrix} 138.4164 & 2.4407 & -6.7702 \\ 167.8987 & 0.9985 & -7.5232 \\ 383.8513 & -4.9224 & -17.4799 \end{bmatrix}, K_2 = \begin{bmatrix} 101.5134 & 1.6037 & -3.8630 \\ 119.6668 & 0.8277 & -4.3093 \\ 276.3462 & -1.8812 & -10.0604 \end{bmatrix},$$

$$K_4 = \begin{bmatrix} 96.5387 & 1.2329 & -2.9000 \\ 93.0939 & 0.8237 & -2.1852 \\ 268.1340 & -0.7194 & -7.2606 \end{bmatrix}, K_8 = \begin{bmatrix} 175.2877 & 1.6404 & -6.3880 \\ 61.4511 & 1.1225 & -2.3991 \\ 391.6398 & -0.1709 & -14.2824 \end{bmatrix},$$

$$K_9 = \begin{bmatrix} 143.3570 & 1.4171 & -6.7477 \\ 71.8536 & 0.9124 & -3.4568 \\ 359.6986 & 0.5128 & -16.8429 \end{bmatrix}, K_{12} = \begin{bmatrix} 135.7652 & 1.2744 & -7.2143 \\ 84.5744 & 0.9719 & -5.9774 \\ 399.1680 & 1.8255 & -23.1364 \end{bmatrix}, \quad (85)$$

$$K_{18} = \begin{bmatrix} 221.0958 & 4.4685 & -11.2652 \\ 162.9903 & 3.1023 & -8.3565 \\ 555.0112 & 7.9902 & -29.2663 \end{bmatrix}, K_{19} = \begin{bmatrix} 251.0396 & 5.1211 & -13.9648 \\ 113.9665 & 2.3356 & -5.9639 \\ 762.3699 & 11.0262 & -40.3752 \end{bmatrix}$$
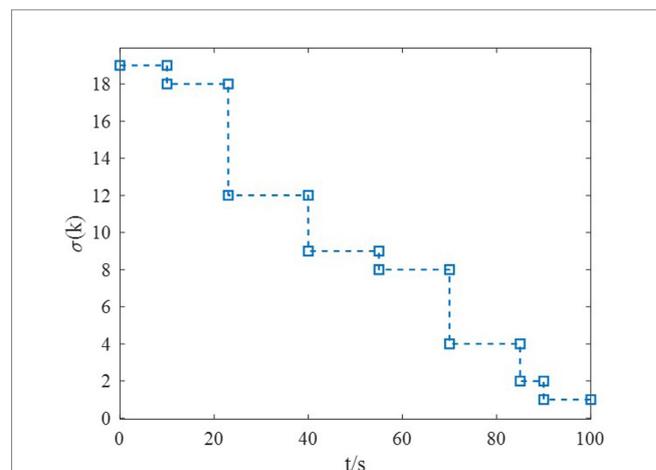

FIGURE 2
Switching logic of HiMAT in the flight envelope.

TABLE 1 Dwell time of various switching logics.

Switching logic	Parameter	Result
MDADT	$a_1 = 0.22, a_2 = 0.24, a_4 = 0.23, a_8 = 0.19, a_9 = 0.31, a_{12} = 0.26, a_{18} = 0.27, a_{19} = 0.28, b_1 = 0.03,$ $b_2 = 0.04, b_4 = 0.02, b_8 = 0.05, b_9 = 0.03, b_{12} = 0.04, b_{18} = 0.03, b_{19} = 0.03, \Delta_1 = 2, \Delta_2 = 1, \Delta_4 = 3, \Delta_8 = 3$ $, \Delta_9 = 2, \Delta_{12} = 1, \Delta_{18} = 3, \Delta_{19} = 1, \mu_1 = 1.15, \mu_2 = 1.21, \mu_4 = 1.30, \mu_8 = 1.11, \mu_9 = 1.12, \mu_{12} = 1.22,$ $\mu_{18} = 1.13, \mu_{19} = 1.25.$	$\tau_{a1}^* = 7.7101, \tau_{a2}^* = 4.3390$ $\tau_{a4}^* = 10.7211, \tau_{a8}^* = 16.8154$ $\tau_{a9}^* = 4.2968, \tau_{a12}^* = 3.7734$ $\tau_{a18}^* = 7.3806, \tau_{a19}^* = 3.4131$
ADT	$a = 0.19, b = 0.05, \Delta = 3, \mu = 1.11$	$\tau_a^* = 16.8154$
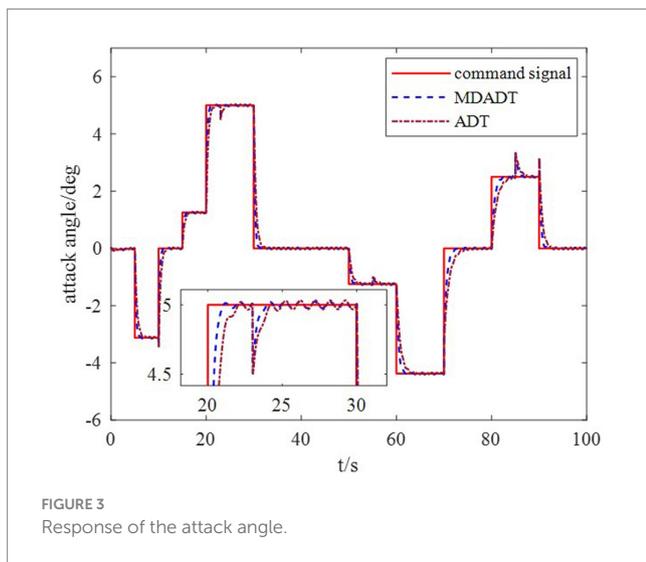
FIGURE 3
Response of the attack angle.
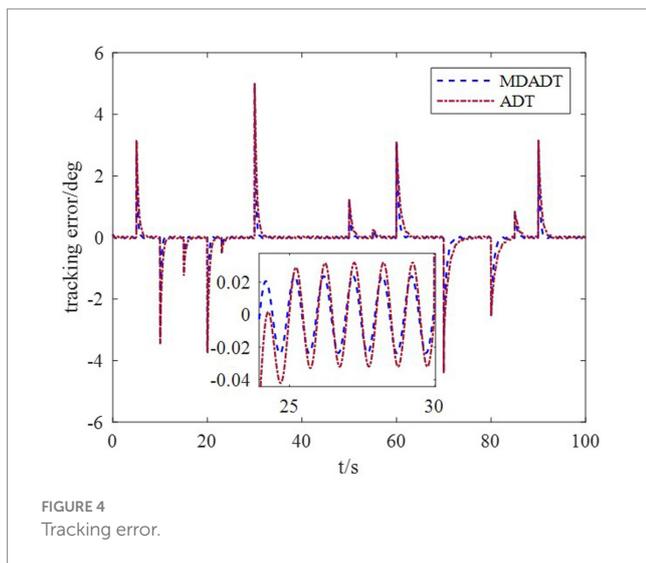
FIGURE 4
Tracking error.

Moreover, to overcome the problem of operation points with static instability, an angular rate compensator was introduced as follows:

$$T_f(s) = \frac{k_q(s + 1/t_q)}{s} \qquad (86)$$

where $T_f(s)$ denotes the transfer function of angular rate compensator, t_q and k_q are the parameters of compensator.

Next, we presented two examples to validate the proposed method.

TABLE 2 Parameters setting of the ADDPG.

Parameter	Value
Discount factor	0.9
Learning rate of the critic network	0.0005
Learning rate of the actor network	0.0005
Mini-batch size	32
Replay buffer size	1,000
Length of each segment	100
Activation function	ReLu
$[\nu_{1min}, \nu_{1max}]$	$[-3, 3]$
$[\nu_{2min}, \nu_{2max}]$	$[-10, 2]$
l_{1min}	10^{-6}
l_{2min}	10^{-6}

Example 1: The tighter bounds on the dwell time can be obtained by the proposed method according to the data in Table 1. Moreover, because the characteristic of each subsystem is considered, the transient performance can be achieved by using the MDADT method. The switching of subsystems is displayed in Figure 2. Notably, the parameters of flight vehicles switch at the switching instants. First, to compare the difference between the two switching logic mechanisms, the simulation results under ADT switching logic and MDADT switching logic are displayed in Figures 3, 4, in which the labels are defined as ADT and MDADT, respectively. Figures 3, 4 reveal that the curves of the attack angle highlight the tracking performance in the flight envelope of switched controllers under ADT switching logic and MDADT switching logic. Thus, the tracking error can converge within the given time interval, and the transient performance of MDADT method is superior. Moreover, in Figures 3, 4, we provide the detailed enlargement of simulation curves near the switching time and steady process. Switched controllers with MDADT logic can achieve better transient performance than the those of controllers with ADT logic. Furthermore, the MDADT method corresponds to smoother response. The switched controllers with MDADT logic can obtain excellent transient performance with tighter bounds on the dwell time, which is less conservative than the ADT logic.

Example 2. In this section, the feasibility of the ADDPG algorithm for flight aircraft is validated. The weights of actor network and critic network are updated such that the learning-based controller adaptively compensates the model uncertainties and external disturbance in the environment. The action of supplementary control is added to the dynamics-based controller, which constitutes the real-time finite-time adaptive tracking control for the flight vehicles. The design parameters of the ADDPG algorithm are defined in Table 2.
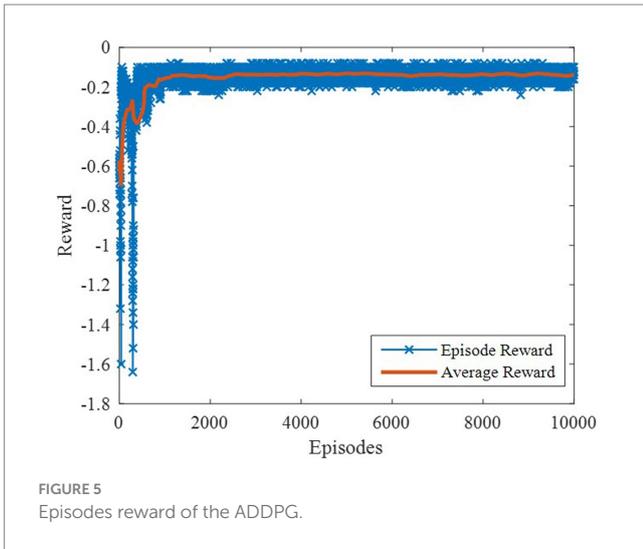
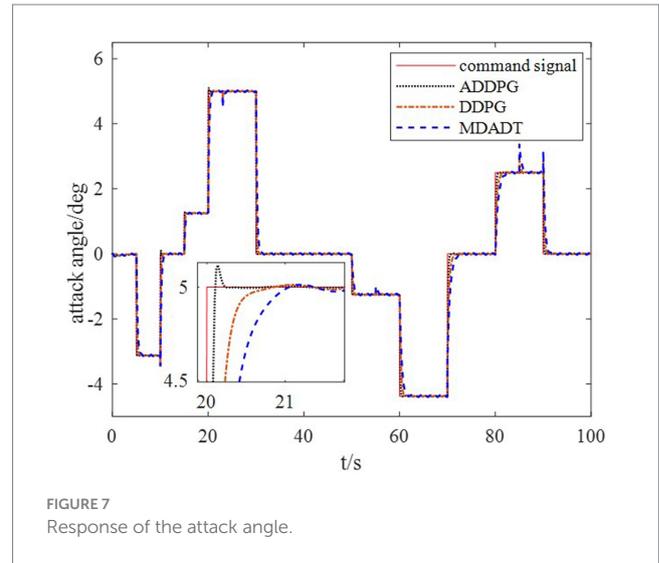FIGURE 5
Episodes reward of the ADDPG.
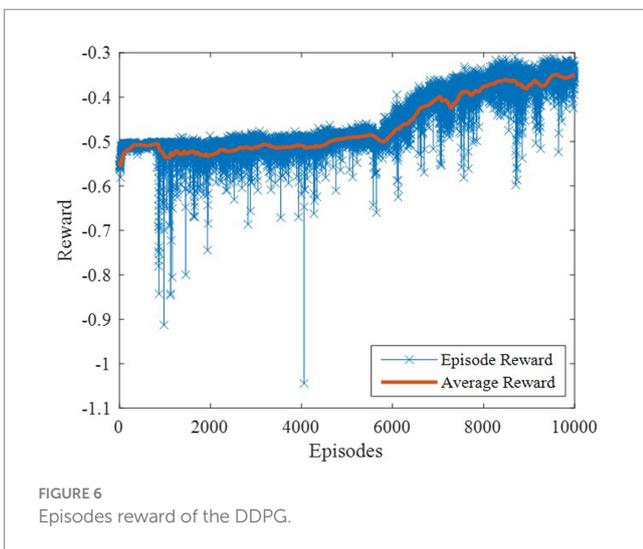

FIGURE 7
Response of the attack angle.

FIGURE 6
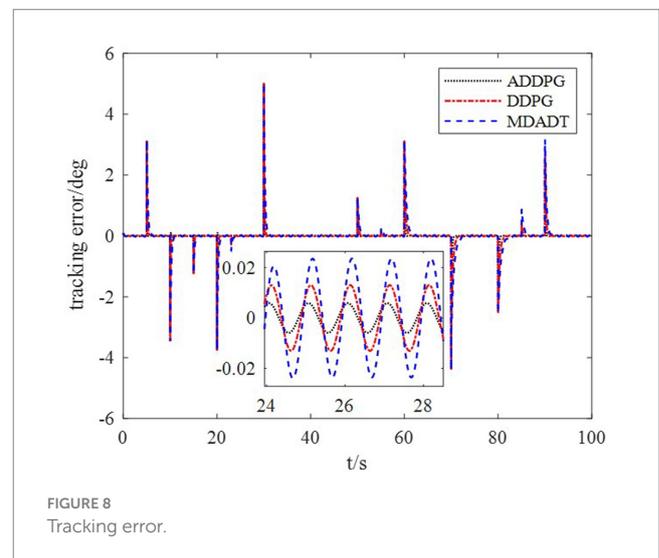Episodes reward of the DDPG.

FIGURE 8
Tracking error.

The input is divided into two paths for critic networks, corresponding to the observation and action. The number of neurons in the input layer of the observation path is the dimension of the observed states, which is represented by *obs*. The number of neurons in the input layer of the action path corresponding to the controller parameters. The critic networks are updated based on the adaptive moment estimation (Adam) algorithm. The regularization factor is set to be 2×10^{-4}.

We define the input of actor network is the observed states and the output is the compensated controller parameters. The activation function of fully connected layers is set to be ReLu and the activation function of output layer is tanh. The weights of actor network are updated based on the Adam algorithm. The variance of noise is set to be 0.1 and the variance decay rate is 1×10^{-5}. Because the stability and robustness of the closed-loop system are guaranteed by the switched control theory and robust control theory, we consider wind gust in the training environment, the perturbations of aerodynamic parameters and wind gust are introduced in the testing environment. Then the algorithms can be implemented on a desktop with Intel Core i7-10700K @3.80 GHz RAM 16.00 GB and operation system of Windows 10.

The DDPG algorithm was simulated to verify the advantages of the proposed method in terms of control performance and convergence for algorithms. The robust controller proposed by the MDADT method was designed as the dynamics-based controller. Both the ADDPG and DDPG algorithms are given in the simulation as the learning-based controller to compensate the unexpected uncertainties in the flight environment. The simulation results are displayed in Figures 5–9, in which the MDADT method, MDADT with DDPG method, and MDADT with ADDPG method are labeled as MDADT, DDPG, and ADDPG, respectively. As displayed in Figures 5, 6, the ADDPG algorithm outperformed the episodes reward convergence of DDPG algorithm, which required fewer episodes to converge in the neighbor of the origin. Therefore, the ADDPG algorithm outperformed the conventional DDPG algorithm in terms of the control performance and steady error. The responses of attack angle are displayed in Figure 7. Both DDPG and ADDPG algorithms could achieve convergence and efficient performance. However, the transient convergence of the ADDPG algorithm was superior to that of the DDPG algorithm. The tracking errors are displayed in Figure 8. The controller compensated with the DDPG and ADDPG algorithms can exhibit improved
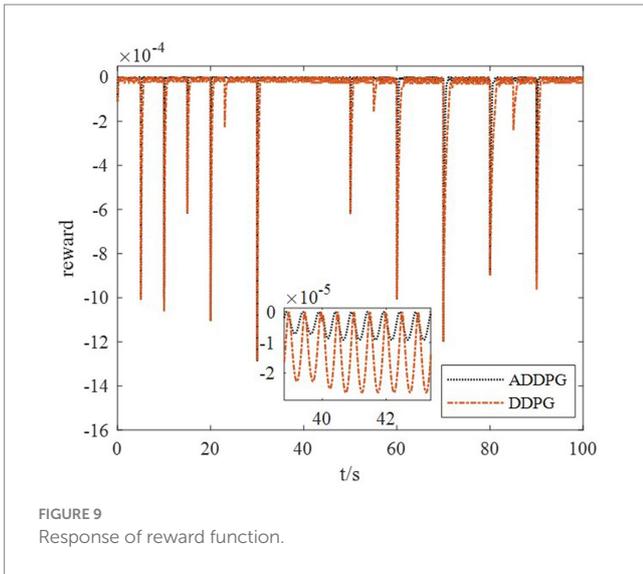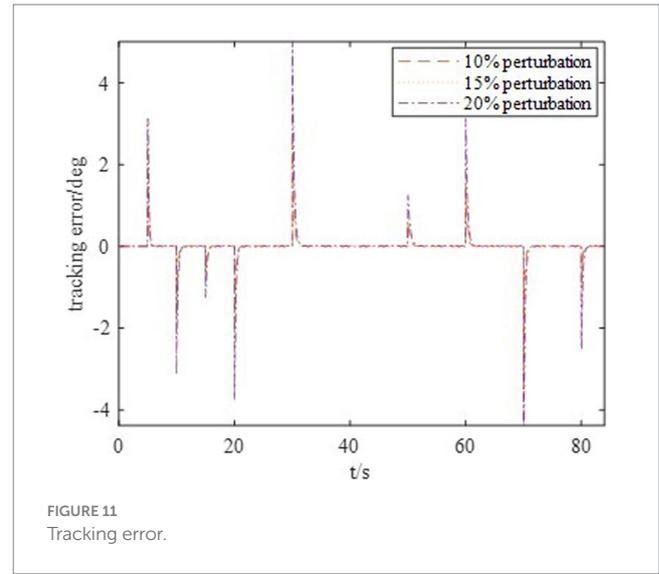
FIGURE 9
Response of reward function.

TABLE 3 Average tracking errors.

Methods	Value (deg)
ADT	0.2053
MDADT	0.1196
DDPG with MDADT	0.0711
ADDPG with MDADT	0.0326

FIGURE 11
Tracking error.

TABLE 4 Average tracking errors in the presence of aerodynamic perturbations.

Methods	Value (deg)
10% perturbation	0.0442
15% perturbation	0.0576
20% perturbation	0.0898
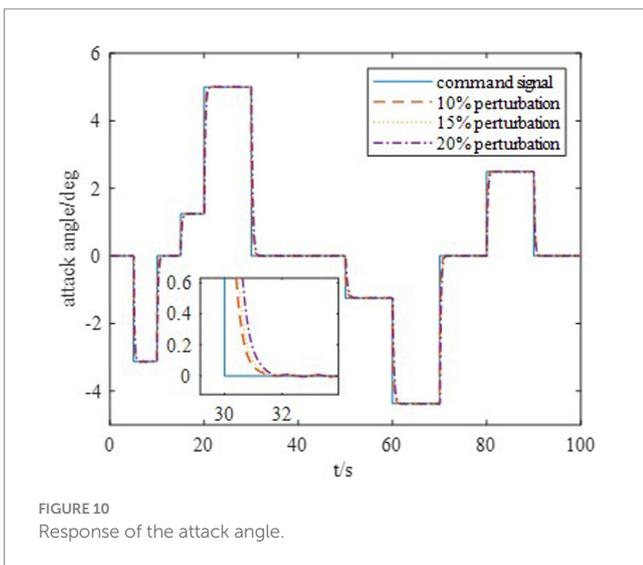
FIGURE 10
Response of the attack angle.

performance of steady-state response. However, the steady-state error of the ADDPG algorithm was less than that of the DDPG algorithm. The reward function of an episode is displayed in Figure 9. The ADDPG algorithm can achieve superior final performance.

The average tracking errors of methods are presented in Table 3. The online scheduling through DDPG and ADDPG can efficiently reduce the average tracking error; the adaptive reward function can improve the tracking performance. The proposed method can overcome the undesirable response caused by asynchronous switching and uncertainties in the flight environment.

Moreover, to show the effectiveness to deal with system uncertainties and disturbance, we give the simulation results of HiMAT vehicle with disturbances and uncertainties of aerodynamic parameters, which can also illustrate the potential application prospects for practical environment. The results are described in Figures 10, 11, in which we consider the cases where the aerodynamic parameter perturbations are 10, 15, and 20%. The responses of attack angle are given in Figure 10 and the tracking errors are given in Figure 11. The average tracking errors in the presence of aerodynamic perturbations are also provided in Table 4. We can see that the stability and tracking performance can be guaranteed with uncertainties and disturbances by using the proposed method, which illustrates that the proposed method can ensure the control accuracy, stability, and robustness simultaneously.

Remark 4: We draw inspiration from the traditional method of dealing with the sim-to-real transfer issue. Firstly, the nonlinear model is converted to a linear model by employing Jacobian linearization. Then we can design the nominal controller on the reference points. In most engineering applications, the stability margin is introduced and analyzed to ensure the robustness. Similarly, in this paper, we developed finite-time robust control theory to ensure the stability and attenuation performance. The uncertainties and disturbances in practical environment can be overcome. However, we noticed that it is difficult to realize optimal compromise between robustness and transient performance. The ADDPG algorithm is given to improve the control accuracy. Moreover, the non-fragile control theory is introduced, which ensures the stability and prescribed attenuation performance on the scheduling intervals.

Remark 5: The problem of finite-time tracking control for switched flight vehicles was investigated. According to the numerical examples, the advantages of the suggested control method to address the flight vehicle considering disturbances and uncertainties over the existing control methods are demonstrated, which can be described as follows: (1) Unlike the conventional model-based control methods, the proposed method was developed by using DRL, which can improve control performance and overcome the undesirable response caused by uncertainties. (2) In the proposed method, the advantages of model-based and model-free method are combined. The dynamics-based controller was developed to ensure stability and robustness, and the learning-based controller was proposed to compensate the uncertainties in the flight environment. (3) The established adaptive generalized reward function can improve convergence and robustness.

5 Conclusion

The finite-time control of switched flight vehicles with asynchronous switching was realized using a novel nonfragile DRL method. The flight vehicles were modeled as the switched system, and the asynchronous switching caused by packet dropouts was considered. The MDADT and MLF methods were used to ensure FTS and weighted prescribed attenuation index. LMIs were used to determine the solutions of the finite-time tracking controller. To compensate the external disturbance and improve tracking performance, the ADDPG algorithm based on the actor–critic framework was provided to optimize the parameters of tracking controllers. To improve optimization efficiency and decrease computational complexity, parameter optimization was assumed to be limited in the given range. The compensation of control policy in a given range is considered as the uncertainties of the controller parameters, and the FTS is ensured by nonfragile control theory. Compared with the conventional DDPG algorithm, the adaptive hyper parameters of reward function were introduced to achieve superior control performance and realize a general case. The FTS, robustness, and transient performance were ensured simultaneously by the proposed method. In the future, the following four points should be studied: (1) The event-triggered control structure should be considered to reduce the load and improve the robustness of information transformation. (2) The parallel optimization methods should be presented to improve training efficiency. (3) The fitting ability and generalization ability of neural networks should be studied to improve the robustness in the complex environment. (4) The semi physical simulations and flight tests of mini drones should be developed to further demonstrate the engineering feasibility of the proposed method.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

HC: Writing – original draft, Writing – review & editing. RS: Writing – original draft, Writing – review & editing. HL: Writing – review & editing, Writing – original draft. WW: Writing – review & editing, Writing – original draft. BZ: Writing – review & editing, Writing – original draft. YF: Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aristidou, P., Fabozzi, D., and Cutsem, V. T. (2014). Dynamic simulation of large-scale power systems using a parallel Schur-complement-based decomposition method. *IEEE Trans. Para. and Dis. Sys.* 25, 2561–2570. doi: 10.1109/TPDS.2013.252

Ban, J., Kwon, W., Won, S., and Kim, S. (2018). Robust H$_\infty$ finite-time control for discrete-time polytopic uncertain switched linear systems. *Nonlinear Anal-Hybri* 29, 348–362. doi: 10.1016/j.nahs.2018.03.005

Bao, C. Y., Wang, P., and Tang, G. J. (2021). Integrated method of guidance, control and morphing for hypersonic morphing vehicle in glide phase. *Chin. J. Aeronaut.* 34, 535–553. doi: 10.1016/j.cja.2020.11.009

Chen, L. H., Fu, S. S., Zhao, Y. X., Liu, M., and Qiu, J. B. (2020). State and fault observer design for switched systems via an adaptive fuzzy approach. *IEEE Trans. Fuzzy Syst.* 28, 2107–2118. doi: 10.1109/TFUZZ.2019.2930485

Chen, S. Z., Ning, C. Y., Liu, Q., and Liu, Q. P. (2022). Improved multiple Lyapunov functions of input–output-to-state stability for nonlinear switched systems. *Inf. Sci.* 608, 47–62. doi: 10.1016/j.ins.2022.06.025

Cheng, H. Y., Dong, C. Y., Jiang, W. L., Wang, Q., and Hou, Y. Z. (2017). Non-fragile switched H∞ control for morphing aircraft with asynchronous switching. *Chin. J. Aeronaut.* 30, 1127–1139. doi: 10.1016/j.cja.2017.01.008

Cheng, H. Y., Fu, W. X., Dong, C. Y., Wang, Q., and Hou, Y. Z. (2018). Asynchronously finite-time H∞ control for morphing aircraft. *Trans. Inst. Meas. Control.* 40, 4330–4344. doi: 10.1177/0142331217746737

Cheng, L., Wang, Z. B., and Jiang, F. H. (2019). Real-time control for fuel-optimal moon landing based on an interactive deep reinforcement learning algorithm. *Astrodynamics* 3, 375–386. doi: 10.1007/s42064-018-0052-2

Cheng, Y., and Zhang, W. D. (2018). Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels. *Neurocomputing* 272, 63–73. doi: 10.1016/j.neucom.2017.06.066

Gaudet, B., Linares, R., and Furfaro, R. (2020). Deep reinforcement learning for six degree-of-freedom planetary landing. *Adv. Space Res.* 65, 1723–1741. doi: 10.1016/j.asr.2019.12.030

Gheisarnejad, M., and Khooban, H. M. (2021). An intelligent non-integer PID controller-based deep reinforcement learning: implementation and experimental results. *IEEE Trans. Ind. Electron.* 68, 3609–3618. doi: 10.1109/TIE.2020.2979561

Giacomin, P. A. S., and Hemerly, E. M. (2022). A distributed, real-time and easy-to-extend strategy for missions of autonomous aircraft squadrons. *Inf. Sci.* 608, 222–250. doi: 10.1016/j.ins.2022.06.043

Gong, L. G., Wang, Q., Hu, C. H., and Liu, C. (2020). Switching control of morphing aircraft based on Q-learning. *Chin. J. Aeronaut.* 33, 672–687. doi: 10.1016/j.cja.2019.10.005

Grigorie, L. T., Khan, S., Botez, M. R., Mamou, M., and Mébarki, Y. (2022). Design and experimental testing of a control system for a morphing wing model actuated with miniature BLDC motors. *Chin. J. Aeronaut.* 33, 1272–1287. doi: 10.1016/j.cja.2019.08.007

Guo, Q., Zhang, Y., Celler, G. B., and Su, W. S. (2019). Neural adaptive backstepping control of a robotic manipulator with prescribed performance constraint. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 3572–3583. doi: 10.1109/TNNLS.2018.2854699

Hu, Q. L., Xiao, L., and Wang, C. L. (2019). Adaptive fault-tolerant attitude tracking control for spacecraft with time-varying inertia uncertainties. *Chin. J. Aeronaut.* 32, 674–687. doi: 10.1016/j.cja.2018.12.015

Huang, L. T., Li, Y. M., and Tong, S. C. (2020). Fuzzy adaptive output feedback control for MIMO switched nontriangular structure nonlinear systems with unknown control directions. *IEEE Trans. Syst. Man. Cybern. Syst.* 50, 550–564. doi: 10.1109/TSMC.2017.2778099

Jiang, W. L., Wu, K. S., Wang, Z. L., and Wang, Y. N. (2020). Gain-scheduled control for morphing aircraft via switching polytopic linear parameter-varying systems. *Aerosp. Sci. Technol.* 107:106242. doi: 10.1016/j.ast.2020.106242

Lee, S., and Kim, D. (2022). Deep learning based recommender system using cross convolutional filters. *Inf. Sci.* 592, 112–122. doi: 10.1016/j.ins.2022.01.033

Li, M. L., and Deng, F. Q. (2018). Moment exponential input-to-state stability of non-linear switched stochastic systems with Lévy noise. *IET Contr. Theory Appl.* 12, 1208–1215. doi: 10.1049/iet-cta.2017.1229

Liu, Y., Dong, C. Y., Zhang, W. Q., and Wang, Q. (2021). Phase plane design based fast altitude tracking control for hypersonic flight vehicle with angle of attack constraint. *Chin. J. Aeronaut.* 34, 490–503. doi: 10.1016/j.cja.2020.04.026

Liu, T. J., Du, X., Sun, X. M., Richter, H., and Zhu, F. (2019). Robust tracking control of aero-engine rotor speed based on switched LPV model. *Aerosp. Sci. Technol.* 91, 382–390. doi: 10.1016/j.ast.2019.05.031

Liu, L. J., Zhao, X. D., Sun, X. M., and Zong, G. D. (2020). Stability and l2-gain analysis of discrete-time switched systems with mode-dependent average dwell time. *IEEE Trans. Syst. Man. Cybern. Syst.* 50, 2305–2314. doi: 10.1109/TSMC.2018.2794738

Lu, Y., Jia, Z., Liu, X., and Lu, K. F. (2022). Output feedback fault-tolerant control for hypersonic flight vehicles with non-affine actuator faults. *Acta Astronaut.* 193, 324–337. doi: 10.1016/j.actaastro.2022.01.023

Sakthivel, R., Wang, C., Santra, S., and Kaviarasan, B. (2018). Non-fragile reliable sampled-data controller for nonlinear switched time-varying systems. *Nonlinear Anal-Hybri* 27, 62–76. doi: 10.1016/j.nahs.2017.08.005

Sun, Y. M., and Lei, Z. (2021). Fixed-time adaptive fuzzy control for uncertain strict feedback switched systems. *Inf. Sci.* 546, 742–752. doi: 10.1016/j.ins.2020.08.059

Tailor, D., and Izzo, D. (2019). Learning the optimal state-feedback via supervised imitation learning. *Astrodynamics* 3, 361–374. doi: 10.1007/s42064-019-0054-0

Wang, J. H., Ha, L., Dong, X. W., Li, Q. D., and Ren, Z. (2021). Distributed sliding mode control for time-varying formation tracking of multi-UAV system with a dynamic leader. *Aerosp. Sci. Technol.* 111:106549. doi: 10.1016/j.ast.2021.106549

Wang, Z. C., Sun, J., Chen, J., and Bai, Y. Q. (2020). Finite-time stability of switched nonlinear time-delay systems. *Int. J. Robust Nonlinear Control.* 30, 2906–2919. doi: 10.1002/rnc.4928

Wang, Z. L., Wang, Q., Dong, C. Y., and Gong, L. G. (2015). Closed-loop fault detection for full-envelope flight vehicle with measurement delays. *Chin. J. Aeronaut.* 28, 832–844. doi: 10.1016/j.cja.2015.04.009

Wang, H., and Xu, R. (2022). Heuristic decomposition planning for fast spacecraft reorientation under multiaxis constraints. *Acta Astronaut.* 198, 286–294. doi: 10.1016/j.actaastro.2022.06.012

Wang, F., Zhang, X. Y., Chen, B., Chong, L., Li, X. H., and Zhang, J. (2017). Adaptive finite-time tracking control of switched nonlinear systems. *Inf. Sci.* 421, 126–135. doi: 10.1016/j.ins.2017.08.095

Wei, J. M., Zhang, X. X., Zhi, H. M., and Zhu, X. L. (2020). New finite-time stability conditions of linear discrete switched singular systems with finite-time unstable subsystems. *J. Frankl. Inst.* 357, 279–293. doi: 10.1016/j.jfranklin.2019.03.045

Wu, C. H., Yan, J. G., Lin, H., Wu, X. W., and Xiao, B. (2021). Fixed-time disturbance observer-based chattering-free sliding mode attitude tracking control of aircraft with sensor noises. *Aerosp. Sci. Technol.* 111:106565. doi: 10.1016/j.ast.2021.106565

Xiao, X. Q., Park, H. J., Zhou, L., and Lu, G. P. (2020). Event-triggered control of discrete-time switched linear systems with network transmission delays. *Automatica* 111:108585. doi: 10.1016/j.automatica.2019.108585

Xu, J., Hou, Z. M., Wang, W., Xu, B. H., Zhang, K. G., and Chen, K. (2019). Feedback deep deterministic policy gradient with fuzzy reward for robotic multiple peg-in-hole assembly tasks. *IEEE Trans. Ind. Inform.* 15, 1658–1667. doi: 10.1109/TII.2018.2868859

Xu, X. Z., Mao, X., Li, Y., and Zhang, H. B. (2019). New result on robust stability of switched systems with all subsystems unstable. *IET Contr. Theory Appl.* 13, 2138–2145. doi: 10.1049/iet-cta.2019.0018

Yang, D., Li, X. D., and Song, S. J. (2020). Design of state-dependent switching laws for stability of switched stochastic neural networks with time-delays. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 1808–1819. doi: 10.1109/TNNLS.2019.2927161

Yang, D., Zong, G. D., Liu, Y. J., and Ahn, C. K. (2022). Adaptive neural network output tracking control of uncertain switched nonlinear systems: an improved multiple Lyapunov function method. *Inf. Sci.* 606, 380–396. doi: 10.1016/j.ins.2022.05.071

Yuan, S., Schutter, D. B., and Baldi, S. (2017). Adaptive asymptotic tracking control of uncertain time-driven switched linear systems. *IEEE Trans. Autom. Control* 62, 5802–5807. doi: 10.1109/TAC.2016.2639479

Yue, T., Xu, Z. J., Wang, L. X., and Wang, T. (2019). Sliding mode control design for oblique wing aircraft in wing skewing process. *Chin. J. Aeronaut.* 32, 263–271. doi: 10.1016/j.cja.2018.11.002

Zhang, L. X., Nie, L., Cai, B., Yuan, S., and Wang, D. Z. (2019). Switched linear parameter-varying modeling and tracking control for flexible hypersonic vehicle. *Aerosp. Sci. Technol.* 95:105445. doi: 10.1016/j.ast.2019.105445

Zhang, M., and Zhu, Q. X. (2019). Input-to-state stability for non-linear switched stochastic delayed systems with asynchronous switching. *IET Contr. Theory Appl.* 13, 351–359. doi: 10.1049/iet-cta.2018.5956

Zhang, M., and Zhu, Q. X. (2020). Stability analysis for switched stochastic delayed systems under asynchronous switching: a relaxed switching signal. *Int. J. Robust Nonlinear Control* 30, 8278–8298. doi: 10.1002/rnc.5240

Zhao, X. D., Shi, P., Yin, Y. F., and Nguang, S. K. (2017). New results on stability of slowly switched systems: a multiple discontinuous Lyapunov function approach. *IEEE Trans. Autom. Control* 62, 3502–3509. doi: 10.1109/TAC.2016.2614911

Zhao, X. D., Zhang, L. X., Shi, P., and Liu, P. (2012). Stability of switched positive linear systems with average dwell time switching. *Automatica* 48, 1132–1137. doi: 10.1016/j.automatica.2012.03.008