DOI: 10.13009/EUCASS2023-902
Aerospace Europe Conference 2023 – 10TH EUCASS – 9TH CEAS

Robust Adaptive Switching Control of Hypersonic Reentry Vehicle based on Q-learning

Xin Zhang†, Jiyuan Sun*, Shushi Wang* and Shuang Li**
** Nanjing University of Aeronautics and Astronautics*
29 Yudao Street, Qinhuai District, Nanjing, 210016, China
zhangxin2019@nuaa.edu.cn – sunjiyuan@nuaa.edu.cn
† Xin Zhang

Abstract

This paper proposes a robust adaptive switching control strategy for the hypersonic reentry vehicle based on the Q-learning algorithm. First, the kinematic and dynamic models of the hypersonic reentry vehicle (HRV) are established, apart from the system uncertainties and external disturbances. Then, the traditional PID controller is presented for this basic model. However, during the reentry phase, the strong coupling effects, high angle of attack, large disturbances, sharp changes of atmosphere environment, and system uncertainties inevitably depreciate the performance of the PID controller. To further improve HRV's control performance, the Q-leaning algorithm and zero-sum game theory are devised in the presence of unknown system matrices and disturbances. Therefore, the designed control scheme in this paper consists of the basic PID controller and the model-free Q-learning controller, while an adaptive switching strategy is generated by the attitude tracking errors and other performance evaluation factors based on the analytic hierarchy process method. As a result, the controller switching occurs mainly based on the better controller's judgement between PID and Q-learning controller by considering the responding control performance factors. Normally, when the attitude tracking error converges to a small value, the current controller in use will switch from PID to Q-learning. Finally, this proposed switching control system's stability analysis is conducted by the average dwell time method, while its feasibility, robustness and control performance are demonstrated by comparative simulations.

1. Introduction

The hypersonic reentry vehicle (HRV) has the characteristics of fast response, high flight speed, large envelop and global accessibility, which is of great significance in both military and civil fields [1]. Since HRV is a complex, strong coupling, nonlinear system with a harsh and varying reentry flight environment [2], it's quite a challenge for the attitude control design. Especially during the re-entry phase, it suffers strong coupling effects, severe external disturbance, nonlinearities, sharp changes of atmospheric environment, unknown uncertainties, redundant and heterogeneous actuators, and so on [3]. Therefore, even though HRV's attitude control research has attracted considerable attention during the past few decades, there are still many key technologies and new issues remain to be conquered, such as HRV attitude control system's fast stability, anti-interference capability, anti-saturation capability and so on.

At present, the control methods of HRV in the literature can be roughly divided into the linear control methods, nonlinear control methods, and intelligent control methods. As known, the conventional linear control methods have achieved good control performance in the conventional flight control, and widely used for the hypersonic vehicle as well, such as PID, pole placement, robust control, and so on. However, during the actual flight of HRV, some state information is quite difficult to accurately measure or estimate. The conventional controller's performance is inevitably depreciated for the fast time-varying, strong coupling and aerodynamic uncertainty of HRV system, since most of the linear control methods rely on the model accuracy. So, for the recent past research, the disturbance observer (DO) [4] is often introduced as a more sophisticated way of dealing with this kind of model uncertainties. It is not only widely researched, but also received abundant achievements, such as sliding mode DO, extended state observer [5], tracking differentiator based DO [6], high-gain DO. Meanwhile, quite a few researchers focus on the improvement of the conventional PID algorithm by introducing such as fractional calculus [7], fuzzy logic [8], neural networks [9], predictive control [10], and supervised learning [11], which on one hand retains the advantages and basic framework of PID controller, on the other hand improves its control capability under highly nonlinear and uncertain circumstances.

DOI: 10.13009/EUCASS2023-902

But, along with the higher and higher demands for HRV's attitude control precision, the weakness of linear control methods gradually emerges. For example, the above modified linear controllers can only ensure the state stability of the nonlinearly controller object near some equilibrium points, but not the global stability of the whole flight process [12], and has deficient ability of dealing with the system failures. To solve this problem and improve the system's online adaptive ability and robust performance, several nonlinear control algorithms are addressed, such as sliding-mode control [13], feedback linearization [14] and backstepping control [15]. Taking the backstepping as an example, it is normally devised by considering the system's nonlinearity and uncertainty, while the stableness of the designed control system is employed by the integration of the Lyapunov functions and intermediate virtual control variables. Because of its fast convergence speed, backstepping is quite suitable for the online control of dynamic systems. Therefore, it has been widely used in aerospace. Ref. [16] verifies the advantages of backstepping method in dealing with HRV control system's uncertainty and maintaining its stability, by the comparison with several different nonlinear control methods. However, even though the nonlinear control methods can effectively mitigate the strong nonlinear and coupling problems, most of them still require HRV's model accuracy. In addition, the nonlinear control methods normally require more complex design processes and calculation computations than the linear ones, which may bring new challenges for the HRV control system design.

During the past decade, the intelligent control technology including fuzzy control, expert control based on knowledge, neural-network control and integrated intelligent control, is developing gradually [17], since some of them can learn from large volumes of data, model-free, and the corresponding designed controllers can be trained to achieve higher robustness, precision and adaptable ability. As a remarkable strategy of intelligent methods, Q-learning scheme is model-free and learns an action-dependent value function to determine the optimal control action. The Q-learning based control schemes relies on the definition of an appropriate Q-function related to the action adopted. The optimal control input can be determined once the Q-function is obtained even when the model is unknown, which constitutes the main merit of Q-learning. Although there are numerous studies on Q-learning or intelligent control algorithms, the corresponding engineering application are relatively few, and most of the existing research on intelligent control are combined with traditional control approaches. The Q-learning technique is developed in Ref. [18] for model-free optimal tracking control of general non-affine nonlinear discrete-time systems with a critic-only implementation structure. In the spacecraft control field, Ref. [19] proposes a Q-learning attitude controller for the morphing aircraft, which shows the better control performance and robustness by the simulation. However, this designed control strategy is only designed for altitude tracking, and the controller' switching occurs only once, which is not suitable for HRV and its large angle attitude manoeuvre phase.

Therefore, motivated by the discussion above, to enhance the closed-loop attitude control performance for HRV, a robust adaptive switching control strategy based on Q-learning algorithm is provided. To be more specific, this paper is arranged as follows. Section II constructs the mathematical model of HRV, while the linearized control model are established for the re-entry phase. Section III starts with the prescription of the adaptive switching control scheme, which is designed to be able to switch from PID controller and Q-learning controller back and forth. Then, a related zero-sum differential game problem between the control input and disturbances is formulated, and solved to devise the Q-learning controller with unknown system matrices. Since the designed control scheme has two different kinds of controllers, PID and Q-learning, an adaptive switching strategy is then established by considering the attitude tracking error, and other control performance factors based on the analytic hierarchy process method. As a result, the switching from PID controller to Q-learning one or backward way occurs throughout the whole reentry phase. Then, the stability analysis of the designed switching control system is theoretically derived based on the mean dwell time approach. Section IV verifies the feasibility, robustness and better control performance of the proposed switching control scheme by numerical comparative simulations. Finally, the conclusion is summarized in Section V.

2. HRV model description

2.1 6-DOF dynamic model

The 6-DOF dynamic equations of the rigid-body HRV are made up of 3-DOF translational equations and 3-DOF rotational equations. The 3-DOF translational equations in the launch inertial coordinate system are given as follows:

$$m\begin{bmatrix} \dot{V} \\ V\dot{\theta}\cos\psi_v \\ -V\dot{\psi}_v \end{bmatrix} = \boldsymbol{C}_{\boldsymbol{n}}^{\boldsymbol{hv}}\begin{bmatrix} 0 \\ -mg \\ 0 \end{bmatrix} + \boldsymbol{C}_{\boldsymbol{v}}^{\boldsymbol{hv}}\boldsymbol{R} + \boldsymbol{C}_{\boldsymbol{b}}^{\boldsymbol{hv}}\boldsymbol{P} \tag{1}$$

where m is the mass of HRV; g is the acceleration of gravity; $V,\dot{V}$ are the flight velocity and acceleration of HRV, respectively; θ,ψ_v,γ_v are the flight-path angle, flight deflection angle and velocity bank angle, respectively; $\boldsymbol{R},\boldsymbol{P}$ is the

DOI: 10.13009/EUCASS2023-902

force vector generated by the aerodynamic surfaces, and propulsion engines, respectively; $\boldsymbol{C}_{\boldsymbol{n}}^{\boldsymbol{hv}}$ is the transformation matrix from the launch inertial to the ballistic coordinatesystem; $\boldsymbol{C}_{\boldsymbol{v}}^{\boldsymbol{hv}}$ is the transformation matrix from the velocity coordinate system to the ballistic coordinate system; $\boldsymbol{C}_{\boldsymbol{b}}^{\boldsymbol{hv}}$ is the transformation matrix from the body to the ballistic coordinate system. The expansions of $\boldsymbol{C}_{\boldsymbol{n}}^{\boldsymbol{hv}}, \boldsymbol{C}_{\boldsymbol{v}}^{\boldsymbol{hv}}, \boldsymbol{C}_{\boldsymbol{b}}^{\boldsymbol{hv}}$ are described as

$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{n}}^{\boldsymbol{hv}} &= \begin{bmatrix} \cos\theta\cos\psi_v & \sin\theta\cos\psi_v & -\sin\psi_v \\ -\sin\theta & \cos\theta & 0 \\ \cos\theta\sin\psi_v & \sin\theta\sin\psi_v & \cos\psi_v \end{bmatrix} \\
\boldsymbol{C}_{\boldsymbol{v}}^{\boldsymbol{hv}} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_v & -\sin\gamma_v \\ 0 & \sin\gamma_v & \cos\gamma_v \end{bmatrix} \\
\boldsymbol{C}_{\boldsymbol{b}}^{\boldsymbol{v}} &= \begin{bmatrix} \cos\alpha\cos\beta & -\sin\alpha & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta & \cos\alpha & \sin\alpha\sin\beta \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \\
\boldsymbol{C}_{\boldsymbol{b}}^{\boldsymbol{hv}} &= \boldsymbol{C}_{\boldsymbol{b}}^{\boldsymbol{v}} \cdot \boldsymbol{C}_{\boldsymbol{v}}^{\boldsymbol{hv}}
\end{aligned} \tag{2}$$

The 3-DOF rotational equations used for HRV's attitude control are built in the body coordinate system as

$$\begin{aligned}
\begin{bmatrix} \dot{\gamma} \\ \dot{\psi} \\ \dot{\varphi} \end{bmatrix} &= \begin{bmatrix} 1 & \sin\gamma\tan\psi & \cos\gamma\tan\psi \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma/\cos\psi & \cos\gamma/\cos\psi \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \\
\begin{bmatrix} \dot{\omega}_x \\ \dot{\omega}_y \\ \dot{\omega}_z \end{bmatrix} &= -\boldsymbol{I}_0^{-1}\boldsymbol{\Omega}\boldsymbol{I}_0^{-1} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} + \boldsymbol{I}_0^{-1}\left(\boldsymbol{M}_P + \boldsymbol{M}_R\right)
\end{aligned} \tag{3}$$

where γ, ψ, φ are the roll, yaw and pitch angle, respectively; $\omega_x, \omega_y, \omega_z$ are the roll, yaw and pitch angular velocity of HRV's body coordinate rotation from launch inertial coordinate, respectively; $\boldsymbol{M}_P, \boldsymbol{M}_R$ are the moment vectors generated by the propulsion engine and aerodynamic surfaces, respectively; $\boldsymbol{I}_0$ is the moment matrix of inertia. The expansions of $\boldsymbol{I}_0, \boldsymbol{\Omega}$ are

$$\boldsymbol{I}_0 = \begin{bmatrix} I_x & -I_{xy} & -I_{xz} \\ -I_{xy} & I_y & -I_{yz} \\ -I_{xz} & -I_{yz} & I_z \end{bmatrix}, \boldsymbol{\Omega} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \tag{4}$$

2.2 Linearized control model

The linearized control model is established based on the small perturbation assumption, while ignoring the earth's rotation, the product of inertia, the Coriolis, relative forces and moments. As the control variables are chose as α, ψ, γ for the pitch, yaw and roll channels respectively, while ignoring the system uncertainties, external disturbances, the coupling parameters and small values' high order terms, HRV's linearized control model can be derived as Eq. (5), where the dynamic pressure is $q = \rho V^2/2$, wherein ρ is the air density; S is the reference area, l is the reference length; I_x, I_y, I_z are moments of inertia; $C_y^{\alpha}, C_z^{\beta}, C_y^{\delta_z}$ are the aerodynamic force coefficients caused by α, β, δ_z, respectively; $m_x^{\delta_x}, m_y^{\delta_y}, m_z^{\delta_z}$ are the aerodynamic control moment coefficients of roll, yaw and pitch channels, respectively; m_{dx}, m_{dy}, m_{dz} are the aerodynamic damping moment coefficients of roll, yaw and pitch channels, respectively; c is the distance from HRV's pressure center to mass center.

DOI: 10.13009/EUCASS2023-902

$$\begin{cases} \dot{\omega}_z = -\dfrac{qSl^2 m_{dz}}{I_z V}\omega_z - \dfrac{qScC_y^{\alpha}}{I_z}\alpha + \dfrac{qSlm_z^{\delta_z}}{I_z}\delta_z + \dfrac{(\boldsymbol{M}_P)_z}{I_z} \\ \dot{\alpha} = \left(\dfrac{qScC_y^{\alpha}}{mV^2}+1\right)\omega_z + \dfrac{C_y^{\alpha} qS}{mV}\alpha + \dfrac{C_y^{\delta_z} qS}{mV}\delta_z + \dfrac{(\boldsymbol{P})_y}{mV} \end{cases}$$

$$\begin{cases} \dot{\omega}_y = -\dfrac{qSl^2 m_{dy}}{I_y V}\omega_y - \dfrac{qScC_z^{\beta}}{I_y}\beta + \dfrac{qSlm_y^{\delta_y}}{I_y}\delta_y + \dfrac{(\boldsymbol{M}_P)_y}{I_y} \\ \dot{\psi} = -\omega_y \end{cases} \quad \textbf{(5)}$$

$$\begin{cases} \dot{\omega}_x = -\dfrac{qSl^2 m_{dx}}{I_x V}\omega_x + \dfrac{qSlm_x^{\delta_x}}{I_x}\delta_x + \dfrac{(\boldsymbol{M}_P)_x}{I_x} \\ \dot{\gamma} = \omega_x \end{cases}$$

Let $\boldsymbol{x} = \left[\gamma, \psi, \alpha, \omega_x, \omega_y, \omega_z\right]^T$ stands by the state vector, $\boldsymbol{u} = \left[\boldsymbol{M}_P, \delta_x, \delta_y, \delta_z\right]^T$ denotes the control input, $\boldsymbol{d}$ represents the disturbance input due to linearization, external disturbances, etc. On the basis of Eq. (5), the control model of HRV can be given as follows:

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{G}\boldsymbol{d} \quad \textbf{(6)}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{G}$ denote the system matrices with appropriate dimensions, and are related to the derivatives of the forces and moments with respect to the state vector and control input.

3. Switching controller design

3.1 Switching control scheme

In this section, a switching control scheme is proposed, which is combined with two kinds of controllers, the gain scheduling PID controller, and Q-learning controller. The basic PID controller is designed to ensure the convergence of the attitude tracking errors, because of its simpleness, high robustness and strong engineering practicability. When the attitude tracking errors enter into a neighborhood of the origin with the PID controller, and the local linearization can be achieved for the attitude dynamics model, thus the Q-learning controller will be switched on to replace the PID controller. The reason for this replacement or switching is that, the Q-learning method is model-free and learns an action-dependent value function to determine the optimal control action, the controller designed based on this algorithm can further improve the control performance, compared with the PID controller. However, since HRV's flight environment of the re-entry phase is hostile, and large angle attitude manoeuvres may occur, which may lead to the tracking error suddenly get worse, and the model's local linearization cannot be achieved. In this situation, in order to maintain the stability of the control system, the PID controller will be switched on. Normally, the controller switching generates mainly based on the value of the attitude tracking error. In this paper, to further improve the control performance, an adaptive switching strategy is formulated by the attitude tracking error, its corresponding accumulated error and other performance evaluation factors based on the analytic hierarchy process method. As a result, the designed switching controller structure of this paper is described in Fig.1.

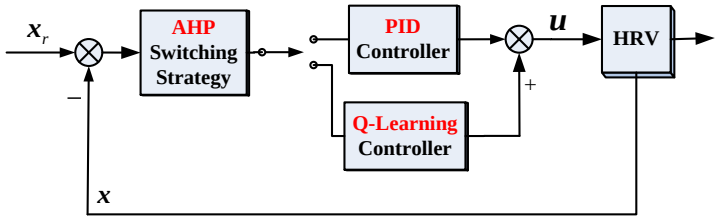

Figure 1: The control structure diagram

In this control scheme, the first part is to design the gain scheduling PID controller. Since the PID control method is quite simple and highly developed, its design procedure needs no further elaboration. To improve the flexibility and robustness of the designed PID controller, its control parameters is employed by varying the gains nonlinearly depending upon the reference trajectory, error, rate of error and/or system states. The tricky parts are the Q-learning controller design in the presence of unknown system, the adaptive switching strategy based on AHP, and how to ensure the switching system's stability, which will be elaborated in the following section 3.2, 3.3 and 3.4.

DOI: 10.13009/EUCASS2023-902

3.2 Q-learning controller

In this section, the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{G}$ of HRV's system(6) are unknown for the Q-learning controller design, which is treated as a zero-sum differential game problem with the following infinite horizon performance index:

$$J\left(\boldsymbol{x}(t_s),\boldsymbol{u},\boldsymbol{d}\right)=\int_{t_s}^{\infty}\left(\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}+\boldsymbol{u}^T\boldsymbol{W}\boldsymbol{u}-\chi^2\boldsymbol{d}^T\boldsymbol{d}\right)d\tau\triangleq\int_{t_s}^{\infty}r\left(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}\right)d\tau \tag{7}$$

where $\boldsymbol{M}\geq\boldsymbol{0}$ and $\boldsymbol{W}>\boldsymbol{0}$ are design matrices with appropriate dimensions, $\chi\geq\chi^*\geq 0$ denotes a user defined constant, and χ^* is the lower bound of χ such that the system (6) can be stabilized. Then, the value function for $\boldsymbol{u}$ and $\boldsymbol{d}$ is introduced as

$$S\left(\boldsymbol{x}(t),\boldsymbol{u},\boldsymbol{d}\right)=\int_{t}^{\infty}\left(\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}+\boldsymbol{u}^T\boldsymbol{W}\boldsymbol{u}-\chi^2\boldsymbol{d}^T\boldsymbol{d}\right)d\tau \tag{8}$$

Then, $\boldsymbol{u}$ and $\boldsymbol{d}$ are conceived as two players of the zero-sum differential game which is defined as

$$S^*\left(\boldsymbol{x}(t_s)\right)=\min_{\boldsymbol{u}}\max_{\boldsymbol{d}}\left[J\left(\boldsymbol{x}(t_s),\boldsymbol{u},\boldsymbol{d}\right)\right]=\min_{\boldsymbol{u}}\max_{\boldsymbol{d}}\left[\int_{t}^{\infty}\left(\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}+\boldsymbol{u}^T\boldsymbol{W}\boldsymbol{u}-\chi^2\boldsymbol{d}^T\boldsymbol{d}\right)d\tau\right] \tag{9}$$

where $\boldsymbol{u}$ acts as the minimizing player, $\boldsymbol{d}$ acts as the maximizing player, so that the saddle point $\left(\boldsymbol{u}^*,\boldsymbol{d}^*\right)$ can be found, and the following inequalities are satisfied for any $\boldsymbol{u}$ or $\boldsymbol{d}$:

$$J\left(\boldsymbol{x}(t_s),\boldsymbol{u}^*,\boldsymbol{d}\right)\leq J\left(\boldsymbol{x}(t_s),\boldsymbol{u}^*,\boldsymbol{d}^*\right)\leq J\left(\boldsymbol{x}(t_s),\boldsymbol{u},\boldsymbol{d}^*\right) \tag{10}$$

Let $\boldsymbol{u}=-\boldsymbol{K}\boldsymbol{x}$ and $\boldsymbol{d}=\boldsymbol{L}\boldsymbol{x}$ denote the state feedback control policy and disturbance policy, respectively, where $\boldsymbol{K}$ and $\boldsymbol{L}$ are matrices with appropriate dimensions. The value function $S\left(\boldsymbol{x}(t)\right)$ can be written as $S\left(\boldsymbol{x}(t)\right)=\boldsymbol{x}^T(t)\boldsymbol{P}\boldsymbol{x}(t)$ with $\boldsymbol{P}$ being a symmetric positive definite matrix. Particularly, the matrix $\boldsymbol{P}^*$ corresponding to the saddle point satisfies the following generalized algebraic Riccati equation:

$$\boldsymbol{A}^T\boldsymbol{P}^*+\boldsymbol{P}^*\boldsymbol{A}+\boldsymbol{M}-\boldsymbol{P}^*\boldsymbol{B}\boldsymbol{W}^{-1}\boldsymbol{B}^T\boldsymbol{P}^*+\chi^{-2}\boldsymbol{P}^*\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{P}^*=\boldsymbol{0} \tag{11}$$

where the saddle point of the zero-sum game is given by

$$\boldsymbol{u}^*=-\boldsymbol{K}^*\boldsymbol{x}=-\boldsymbol{W}^{-1}\boldsymbol{B}^T\boldsymbol{P}^*\boldsymbol{x} \tag{12}$$

$$\boldsymbol{d}^*=\boldsymbol{L}^*\boldsymbol{x}=\chi^{-2}\boldsymbol{G}^T\boldsymbol{P}^*\boldsymbol{x} \tag{13}$$

Moreover, the related value function can be represented by $S^*\left(\boldsymbol{x}(t_s)\right)=\boldsymbol{x}^T(t_s)\boldsymbol{P}^*\boldsymbol{x}(t_s)$ and the Hamiltonian function associated with Eqs.(7) and (8) becomes

$$H\left(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d},\frac{\partial S^*}{\partial\boldsymbol{x}}\right)=\frac{\partial S^*}{\partial\boldsymbol{x}^T}\left(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{u}+\boldsymbol{G}\boldsymbol{d}\right)+\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}+\boldsymbol{u}^T\boldsymbol{W}\boldsymbol{u}-\chi^2\boldsymbol{d}^T\boldsymbol{d} \tag{14}$$

The following Q-function is then introduced by combining the value function with the Hamiltonian:

$$\begin{aligned}Q^*\left(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}\right)&=S^*\left(\boldsymbol{x}\right)+H\left(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d},\frac{\partial S^*}{\partial\boldsymbol{x}}\right)\\&=\boldsymbol{x}^T\boldsymbol{P}^*\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{P}^*\left(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{u}+\boldsymbol{G}\boldsymbol{d}\right)+\left(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{u}+\boldsymbol{G}\boldsymbol{d}\right)^T\boldsymbol{P}^*\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}+\boldsymbol{u}^T\boldsymbol{W}\boldsymbol{u}-\chi^2\boldsymbol{d}^T\boldsymbol{d}\\&=\boldsymbol{U}^T\begin{bmatrix}\boldsymbol{Q}_{xx}&\boldsymbol{P}^*\boldsymbol{B}&\boldsymbol{P}^*\boldsymbol{G}\\\boldsymbol{B}^T\boldsymbol{P}^*&\boldsymbol{W}&\boldsymbol{0}\\\boldsymbol{G}^T\boldsymbol{P}^*&\boldsymbol{0}&-\chi^2\boldsymbol{I}\end{bmatrix}\boldsymbol{U}\triangleq\boldsymbol{U}^T\bar{\boldsymbol{Q}}\boldsymbol{U}\end{aligned} \tag{15}$$

where $\boldsymbol{U}=\left[\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}\right]^T$, $\boldsymbol{Q}_{xx}=\boldsymbol{P}^*+\boldsymbol{M}+\boldsymbol{P}^*\boldsymbol{A}+\boldsymbol{A}^T\boldsymbol{P}^*$.

DOI: 10.13009/EUCASS2023-902

Furthermore, a combination of Eqs.(11) - (15) entails that $Q^*\left(\boldsymbol{x},\boldsymbol{u}^*,\boldsymbol{d}^*\right)=S^*\left(\boldsymbol{x}\right)$, and the Bellman equation is hence satisfied for both $S^*\left(\boldsymbol{x}\right)$ and $Q^*\left(\boldsymbol{x},\boldsymbol{u}^*,\boldsymbol{d}^*\right)$, which is written as

$$Q^*\left(\boldsymbol{x}(t),\boldsymbol{u}^*(t),\boldsymbol{d}^*(t)\right)=Q^*\left(\boldsymbol{x}(t-\varepsilon),\boldsymbol{u}^*(t-\varepsilon),\boldsymbol{d}^*(t-\varepsilon)\right)-\int_{t-\varepsilon}^{t}r\left(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}\right)d\tau \tag{16}$$

where ε denotes the user defined time interval. Since the system matrices are unknown, the approximated expressions of the Q-function, control policy and disturbance policy are then adopted to generate the model-free controller. It follows that $Q^*\left(\boldsymbol{x},\boldsymbol{u}^*,\boldsymbol{d}^*\right)$ can be formulated as

$$Q^*\left(\boldsymbol{x},\boldsymbol{u}^*,\boldsymbol{d}^*\right)=vech\left(\bar{\boldsymbol{Q}}\right)^T\left(\boldsymbol{U}\otimes\boldsymbol{U}\right) \tag{17}$$

where $vech\left(\bar{\boldsymbol{Q}}\right)$ represents the half-vectorization of $\bar{\boldsymbol{Q}}$ with the off-diagonal elements multiplied by 2, and $\left(\boldsymbol{U}\otimes\boldsymbol{U}\right)$ denotes the Kronecker product which can be expressed as $\left\{U_iU_j\right\}$, $i=1,2,\cdots,9$, $j=1,2,\cdots,9$. Let $\boldsymbol{W}_c=vech\left(\bar{\boldsymbol{Q}}\right)$, whose estimated weight is denoted by $\hat{\boldsymbol{W}}_c$, and then the estimation of $Q^*\left(\boldsymbol{x},\boldsymbol{u}^*,\boldsymbol{d}^*\right)$ can be described by

$$Q^*\left(\boldsymbol{x},\boldsymbol{u}^*,\boldsymbol{d}^*\right)=\hat{\boldsymbol{W}}_c^T\left(\boldsymbol{U}\otimes\boldsymbol{U}\right) \tag{18}$$

Similarly, the approximations of the control policy and disturbance policy can be written as

$$\boldsymbol{u}=\hat{\boldsymbol{W}}_{au}^T\boldsymbol{x} \tag{19}$$

$$\boldsymbol{d}=\hat{\boldsymbol{W}}_{ad}^T\boldsymbol{x} \tag{20}$$

where $\hat{\boldsymbol{W}}_{au}$ and $\hat{\boldsymbol{W}}_{ad}$ are the weight estimations. Then, error related terms for derivation of the update laws of the weight estimations $\hat{\boldsymbol{W}}_c$, $\hat{\boldsymbol{W}}_{au}$ and $\hat{\boldsymbol{W}}_{ad}$ are constructed. With Eq.(16), the attitude error function $\boldsymbol{e}_c=[\gamma_r-\gamma,\psi_r-\psi,\alpha_r-\alpha]^T$ and related object function E_c for training $\hat{\boldsymbol{W}}_c$ are defined as

$$\begin{aligned}\boldsymbol{e}_c&=\hat{Q}\left(\boldsymbol{x}(t),\boldsymbol{u}(t),\boldsymbol{d}(t)\right)+\int_{t-\varepsilon}^{t}r\left(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}\right)d\tau-\hat{Q}\left(\boldsymbol{x}(t-\varepsilon),\boldsymbol{u}(t-\varepsilon),\boldsymbol{d}(t-\varepsilon)\right)\\&=\hat{\boldsymbol{W}}_c^T\left(\boldsymbol{U}(t)\otimes\boldsymbol{U}(t)\right)+\int_{t-\varepsilon}^{t}r\left(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}\right)d\tau-\hat{\boldsymbol{W}}_c^T\left(\boldsymbol{U}(t-\varepsilon)\otimes\boldsymbol{U}(t-\varepsilon)\right)\\E_c&=\frac{1}{2}\boldsymbol{e}_c^T\boldsymbol{e}_c\end{aligned} \tag{21}$$

The error functions $\boldsymbol{e}_{au}$, $\boldsymbol{e}_{ad}$ and related object functions E_{au}, E_{ad} for training $\hat{\boldsymbol{W}}_{au}$ and $\hat{\boldsymbol{W}}_{ad}$ are defined as

$$\boldsymbol{e}_{au}=\hat{\boldsymbol{W}}_{au}^T\boldsymbol{x}+\boldsymbol{W}^{-1}\hat{\boldsymbol{B}}^T\boldsymbol{P}^*\boldsymbol{x}\quad E_{au}=\frac{1}{2}\boldsymbol{e}_{au}^T\boldsymbol{e}_{au}\qquad\boldsymbol{e}_{ad}=\hat{\boldsymbol{W}}_{ad}^T\boldsymbol{x}-\chi^{-2}\hat{\boldsymbol{G}}^T\boldsymbol{P}^*\boldsymbol{x}\quad E_{ad}=\frac{1}{2}\boldsymbol{e}_{ad}^T\boldsymbol{e}_{ad} \tag{22}$$

where $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{G}}$ can be extracted from $\hat{\boldsymbol{W}}_c$. The weight estimations $\hat{\boldsymbol{W}}_c$, $\hat{\boldsymbol{W}}_{au}$ and $\hat{\boldsymbol{W}}_{ad}$ are then updated by a gradient based algorithm to minimize E_c, E_{au} and E_{ad}, respectively, which can be formulated as

$$\begin{aligned}&\dot{\hat{\boldsymbol{W}}}_c=-\mu_c\frac{\partial E_c}{\partial\hat{\boldsymbol{W}}_c}=-\mu_c\frac{\partial E_c}{\partial\boldsymbol{e}_c}\frac{\partial\boldsymbol{e}_c}{\partial\hat{\boldsymbol{W}}_c}=-\mu_c\boldsymbol{e}_c\Delta\boldsymbol{U}\qquad\dot{\hat{\boldsymbol{W}}}_{au}=-\mu_{au}\frac{\partial E_{au}}{\partial\hat{\boldsymbol{W}}_{au}}=-\mu_{au}\frac{\partial E_{au}}{\partial\boldsymbol{e}_{au}}\frac{\partial\boldsymbol{e}_{au}}{\partial\hat{\boldsymbol{W}}_{au}}=-\mu_{au}\boldsymbol{e}_{au}\boldsymbol{x}\\&\dot{\hat{\boldsymbol{W}}}_{ad}=-\mu_{ad}\frac{\partial E_{ad}}{\partial\hat{\boldsymbol{W}}_{ad}}=-\mu_{ad}\boldsymbol{x}\boldsymbol{e}_{ad}^T\qquad\qquad\Delta\boldsymbol{U}=\boldsymbol{U}(t)\otimes\boldsymbol{U}(t)-\boldsymbol{U}(t-\varepsilon)\otimes\boldsymbol{U}(t-\varepsilon)\end{aligned} \tag{23}$$

where μ_c, μ_{au} and μ_{ad} are positive constants which determine the learning rates. The update law for $\hat{\boldsymbol{W}}_c$ is further normalized to obtain

$$\dot{\hat{\boldsymbol{W}}}_c=-\mu_c\boldsymbol{e}_c\frac{\Delta\boldsymbol{U}}{\left(1+\Delta\boldsymbol{U}^T\Delta\boldsymbol{U}\right)^2} \tag{24}$$

DOI: 10.13009/EUCASS2023-902

Define $\tilde{\boldsymbol{W}}_c = \boldsymbol{W}_c - \hat{\boldsymbol{W}}_c$, $\tilde{\boldsymbol{W}}_{au} = \boldsymbol{W}_{au} - \hat{\boldsymbol{W}}_{au}$, and $\tilde{\boldsymbol{W}}_{ad} = \boldsymbol{W}_{ad} - \hat{\boldsymbol{W}}_{ad}$ as the weight estimation errors, based on the standard Lyapunov extension theorem and Ref [19], the following theorem can be given:

Theorem 1. Consider the closed-loop system formed of the plant (6), the approximation of the Q-function (18), the control policy (19) and disturbance policy (20) with the update laws given by Eq. (23) and (24). Suppose that the attitude error $\boldsymbol{e}_c$ are bounded, there exist the design parameters $\boldsymbol{M}, \boldsymbol{W}, \varepsilon, \chi, \mu_c, \mu_{au}, \mu_{ad}$ such that all signals of the closed-loop system with the state composed of $\boldsymbol{x}, \tilde{\boldsymbol{W}}_c, \tilde{\boldsymbol{W}}_{au}, \tilde{\boldsymbol{W}}_{ad}$ stay bounded.

The specific proof of Theorem 1 is provided in Ref. [19]. According to this theorem, by choosing the design parameters appropriately, $\boldsymbol{e}_c$ will stay bound, and the better control performance will hold.

3.3 Adaptive switching strategy design

In this section, an adaptive switching strategy is proposed. For the most research, the controllers' switching is generated by the attitude tracking error. However, in order to improve the adaptivity, safety and feasibility of controllers' switching, an analytic hierarchy process method is employed, which apart from the attitude error, its accumulated error, the control actuators' physical constraints and output smoothness, and other performance evaluation factors are all considered. Therefore, the switching strategy can be constructed as follows:

$$\boldsymbol{\kappa} = \boldsymbol{w}_1 \left\| \boldsymbol{e}_c \right\|^2 + \boldsymbol{w}_2 \left\| \boldsymbol{e}_{\max} \right\|^2 + \boldsymbol{w}_{i3} \left\| \boldsymbol{e}_I \right\|^2 + \boldsymbol{w}_4 \left\| \boldsymbol{u}_{\max} \right\|^2 + \boldsymbol{w}_{i5} \left\| \boldsymbol{u}_I \right\|^2 + \boldsymbol{w}_{i6} \left\| \dot{\boldsymbol{u}}_{\max} \right\|^2 + \cdots \tag{25}$$

where $\boldsymbol{e}_{\max}$ and $\boldsymbol{e}_I$ represent the maximum and integral values of $\boldsymbol{e}_c$; $\boldsymbol{u}_{\max}$, $\boldsymbol{u}_I$, $\dot{\boldsymbol{u}}_{\max}$ represent the maximum, integral and maximum differential values of $\boldsymbol{u}$. Then, let $s(\geq 1)$ denote the number of evaluation factors considered in this paper, $\boldsymbol{w}_j \ (j = 1, \cdots, s)$ denotes the weight values of the corresponding factors. It's known that, the switching strategy is almost impossible to make every evaluation factors optimal, its solution is to achieve the balance by constructing weight values for those evaluation factors. The weight values can be determined by the estimation matrix method, namely comparing the importance with each other. Therefore, the weight values in this section varies with the importance change of each evaluation factors. It also can be obtained by trials or numerical simulations.

In this paper, the controller only switches between PID and Q-learning, and generally the weight value of controller error is set bigger than the other weight values. In addition, $\boldsymbol{\kappa}$ can be set as a 3*1 vector, if each channel's controller switch is independent from the other two channels. Otherwise, $\boldsymbol{\kappa}$ can be set as a scalar.

3.4 Stability analysis

In this section, the stability of the whole switching control system is verified by the Lyapunov function based on the average dwell time method. Let τ_D denote the dwell time, which means that for the whole switching system, if every subsystem is stable and remain in use longer than τ_D after every switching, it can be concluded that the whole switching system is stable. Before demonstrating the stability analysis of the whole switching system, there are two lemmas are given. Also, let N denote the subsystem's number of the whole switching system, and for this paper, $N = 2$. $V_i \left(i = 1, \cdots, N \right)$ represents for the Lyapunov function of the corresponding subsystem.

Lemma 1. If $a_i, b_i, c_i, d_i, \eta_i \in K_\infty$ and a smooth function $V_i : \boldsymbol{R}^n \to \boldsymbol{R}^+$ exist, and for any bounded input $\boldsymbol{u}$, satisfy that

$$\begin{cases} a_i \left\| \boldsymbol{x} \right\|^2 \leq V_i(\boldsymbol{x}) \leq b_i \left\| \boldsymbol{x} \right\|^2 \\ \dot{V}_i(\boldsymbol{x}) \leq -c_i \left\| \boldsymbol{x} \right\|^2 + \eta_i \end{cases} \tag{26}$$

then each subsystem is stable during the time period of $[0, t)$.

Lemma 2. Suppose a series of continuously differentiable function $V_i : \boldsymbol{R}^n \to \boldsymbol{R}^+$, $a_i, b_i, d_i, \eta_i \in K_\infty$, and constant value $\mu > 1$ exist, for $\forall p, q \in \{1, \cdots, N\}$ and any bounded input $\boldsymbol{u}$ satisfy that

$$a_i \left\| \boldsymbol{x} \right\|^2 \leq V_p(\boldsymbol{x}) \leq b_i \left\| \boldsymbol{x} \right\|^2 \qquad \dot{V}_p(\boldsymbol{x}) \leq -\lambda_0 V_p(\boldsymbol{x}) + \eta_i \qquad V_p(\boldsymbol{x}) \leq \mu V_q(\boldsymbol{x}) \tag{27}$$

also, the dwell time satisfies $\tau_D > \ln \mu / \lambda_0$, then the switch system is stable during the time period of $[0, t)$.

By the above sections' analysis, the Lyapunov functions of PID and Q-leaning controller are given

DOI: 10.13009/EUCASS2023-902

$$V_1 = \frac{1}{2}\boldsymbol{e}_c^T\boldsymbol{e}_c \qquad V_2 = S^*(\boldsymbol{x}) + \frac{1}{2}\left\|\tilde{\boldsymbol{W}}_c\right\|^2 + \frac{1}{2}\left\|\tilde{\boldsymbol{W}}_{au}\right\|^2 + \frac{1}{2}tr\left(\tilde{\boldsymbol{W}}_{aw}^T\tilde{\boldsymbol{W}}_{aw}\right) \tag{28}$$

Moreover, for HRV's PID control subsystem or Q-learning control subsystem, Lemma 1 is satisfied. To ensure the whole switching system stable, the following theorem is given:

Theorem 2. Consider the closed-loop switching system formed of the plant(6), its control error are bounded, and its every subsystem both satisfy Lemma 1 and 2. So if the average dwell time for each controller's switching satisfies

$$\tau_D > \frac{\ln\mu}{\lambda_0} \tag{29}$$

then, the control error of the whole switching system (6) will stay bounded converge during the time period $[0,t)$.

Proof. For system (6), there exists a series of continuously differentiable function $V_\sigma : \boldsymbol{R}^n \to \boldsymbol{R}^+, \sigma \in \{1,\cdots,N\}$, and $d_\sigma \in K_\infty$ satisfy

$$a_\sigma\|\boldsymbol{x}\|^2 \le V_\sigma(\boldsymbol{x}) \le b_\sigma\|\boldsymbol{x}\|^2 \qquad \dot{V}_\sigma(\boldsymbol{x}) \le -\lambda_\sigma V_\sigma(\boldsymbol{x}) + \eta_\sigma \tag{30}$$

where $a_\sigma, b_\sigma, \lambda_\sigma > 0$ and all of them are constant values.

Denote that

$$\mu = \sup_{i,j\in\{1,\cdots,N\}}\left\{a_i/b_j\right\} \ge 1, \quad \lambda_0 = \inf_{\sigma\in\{1,\cdots,N\}}\left\{\lambda_\sigma\right\} \tag{31}$$

According to Lemma 2, it can be concluded that the control error for the switching system is bounded converge during the time period of $[0,t)$. In the other way, if during the time period of $[t_0, t_0+\tau_D)$, the system is switched and stayed to the i^{th} subsystem, then the upper bound of $V_i(\boldsymbol{x})$'s exponents is described as

$$V_i\left(\boldsymbol{x}(t_0+\tau_D)\right) \le e^{-(\lambda_\sigma)\tau_D}\left(V_i\left(\boldsymbol{x}(t_0)\right) - \frac{\eta_\sigma}{\lambda_\sigma}\right) + \frac{\eta_\sigma}{\lambda_\sigma} \tag{32}$$

Based on the above equation, when $V_i\left(\boldsymbol{x}(t_0)\right) > \eta_\sigma/\lambda_\sigma$ is satisfied, the control errors of the closed-loop system are bound. Therefore, when the dwell time is defined as

$$\tau_D \ge \sup_{i,j}\left(-\frac{b_i}{\lambda_i}\ln\frac{a_i}{b_j}\right) = \frac{\ln\mu}{\lambda_0}, \left(i, j \in \{1,2,\cdots,N\}\right) \tag{33}$$

for any controller's switching, the corresponding Lyapunov function for each subsystem is decreasing and every dwell time period is long enough to make sure the energy of each Lyapunov function is quite small, so that the stability of the whole switching system holds.

According to Theorem 2, it can be seen HRV's switching control system is bounded stabilized.

4. Numerical simulation

This section mainly presents the simulation results of the design control system during the initial re-entry phase. The simulation start time is 300s, whereas the mass of HRV is 16900kg. The initial flight condition is set

$$\begin{cases} x = 123030m \\ y = 106500m \\ z = 3660m \end{cases} \quad \begin{cases} V_x = -98m/s \\ V_y = -540m/s \\ V_z = 20m/s \end{cases} \quad \begin{cases} \varphi = 209^\circ \\ \psi = 0^\circ \\ \gamma = 180^\circ \end{cases} \tag{34}$$

Furthermore, there are two kinds of control actuators, attitude control thrusters and aerodynamic control surfaces. The thrusters mainly work for the high air space and their control output are discrete, while the aerodynamic surfaces need the dense atmosphere and high flight speed to maintain its control ability and their control output is continuous. Therefore, the control allocation scheme for these two kinds of actuators is needed but not elaborated in this paper. In addition, the Q-learning controller is trained offline and switched on mainly when the control errors are smaller than 0.5 degree.

DOI: 10.13009/EUCASS2023-902

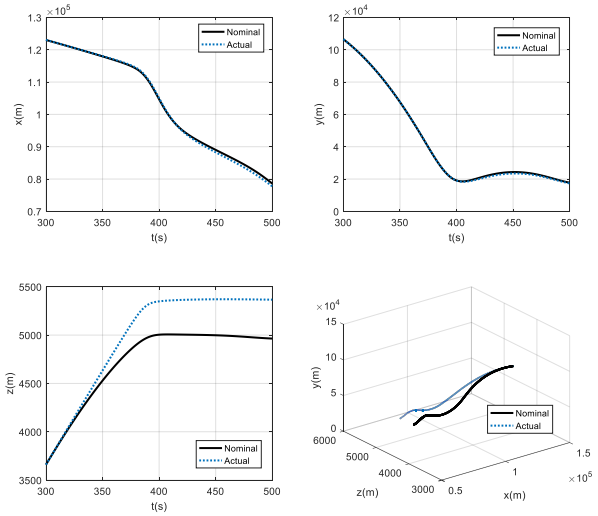

Figure 2: The flight trajectories of the control scheme

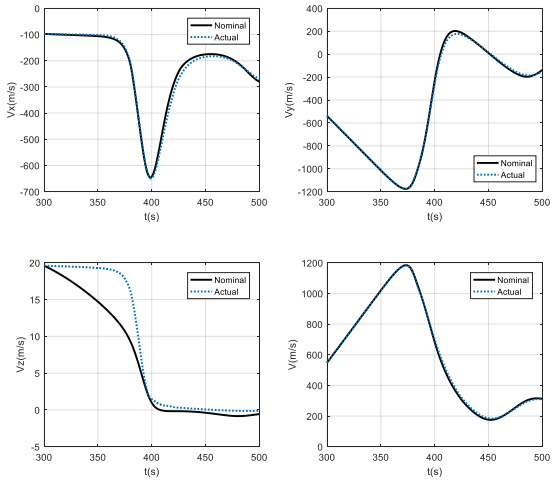

Figure 3: The flight velocities of the control scheme

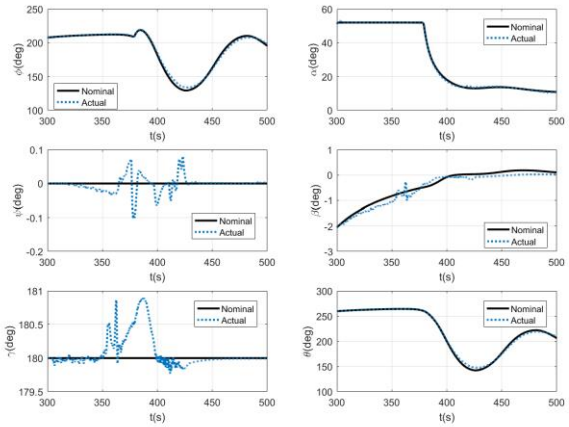

Figure 4: The attitude angles of the control scheme

The simulation results of the designed switching control scheme under the condition of no disturbance and uncertainty are given. Fig. 2, Fig.3 and Fig.4 show the simulation results of flight trajectories, velocities and attitude angles, respectively. Obviously according to these three figures, the actual results are quite close to the nominal data, and the control error maximums of α, ψ, γ are $1.36^{\circ}, 0.11^{\circ}, 0.90^{\circ}$, which also shows the validity and effectiveness of the scheme.

DOI: 10.13009/EUCASS2023-902

Then, to present advantages of the designed control scheme, the simulation comparison of different control schemes with or without the disturbance and uncertainty is given. In addition, the uncertainty parameter is set as $\Delta \boldsymbol{I}=10\%\boldsymbol{I}_0$, and the external disturbance is set as

$$\Delta \boldsymbol{d}=\begin{bmatrix}\left(1+\sin\left(\pi t/10\right)+\sin\left(\pi t/20\right)\right)\times 10^4\\ \left(1+\sin\left(\pi t/10\right)+\sin\left(\pi t/20\right)\right)\times 10^4\\ \left(1+\sin\left(\pi t/10\right)+\sin\left(\pi t/20\right)\right)\times 10^5\end{bmatrix} \tag{35}$$

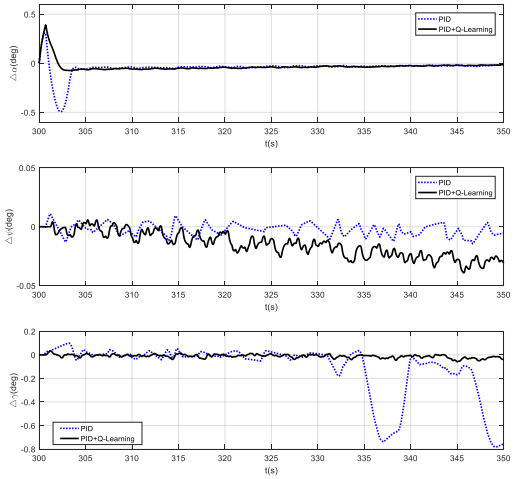

Figure 5: The control error comparison of different control schemes without disturbance or uncertainty

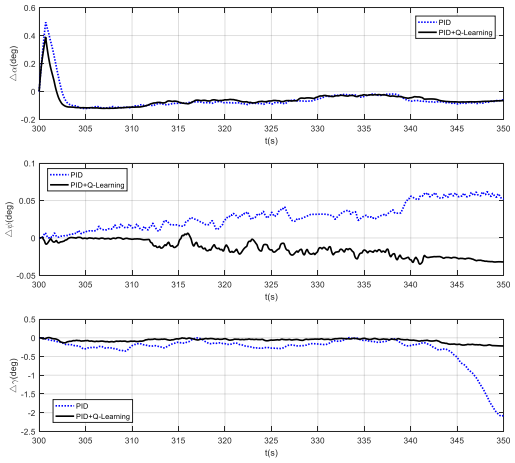

Figure 6: The control error comparison of different control schemes with disturbance and uncertainty

Fig. 5 shows the control error comparisons between PID controller only and the designed switching control scheme under the case without disturbance or uncertainty. And Fig.6 shows the comparisons with the disturbance and uncertainty. In these two figures, the designed switching control scheme is labeled as 'PID+Q-Learning' and compared with the PID controller labeled as 'PID'. According to the control error results in Fig.5, 'PID+Q-learning' is better than 'PID' in the pitch and roll channels, but its advantage is not quite obvious in the yaw channel. That's because HRV's yaw channel is quite coupled with the roll channel, so the controllers' performance in these two channels are coupled with each other, too. According to the above theoretical analysis sections, the designed switching control scheme should have superiority under the worse flight environment, which can be shown in Fig. 6. Especially in roll and yaw channels, the control error of 'PID' becomes quite bigger than 'PID+Q-Learning', or the one of 'PID' in Fig.5. This simulation results show that the designed switching control scheme has better control performance with or without disturbance or system uncertainties, but also shows higher robustness than the PID controller.

DOI: 10.13009/EUCASS2023-902

5. Conclusion

A robust adaptive switching control strategy is proposed for the hypersonic vehicle during its reentry phase based on Q-learning. The designed control scheme in this paper mainly consists of three parts, the PID controller employed for the initial control phase and large angle attitude maneuver phase, the Q-learning controller employed for the attitude tracking phase, and the adaptive switching strategy designed for the above two controllers' switching. Since the PID control method is simple and highly developed, the controller's design process is abbreviated. When the attitude control error is converged into a small value, the Q-learning controller is switched on, to further improve the control performance. Because the Q-learning controller is devised in the presence of unknown system matrices and disturbances, and treated as a zero-sum differential game problem with the control input and disturbance conceived as two opposite players. Moreover, the adaptive switching strategy established by the analytic hierarchy process method, can generate the controller's switching command based on the attitude tracking error and other performance evaluation factors. Then, this proposed switching control system's stability analysis is conducted by the average dwell time method, while its feasibility, robustness and control performance are demonstrated by comparative simulations. By comparing with the PID control systems with or without disturbances and uncertainties, the corresponding control performances are analyzed, which shows the advantages of the presented control scheme.

References

[1] P. Anjaly, S. Swaminathan. 2016. Integrated adaptive guidance and control for the entry phase of winged RLV. In: *2016 Indian Control Conf*. 324-329.

[2] Z. Dong, K. Liu, D. Li, J. Zhang. 2021. A dynamic control allocation approach for reentry compound attitude control design of aerospace vehicle. Journal of Astronautics. 42(6): 749-756.

[3] F. Wang, C. Hua, Q. Zong. 2016. Disturbance observer based finite time control design for reusable launch vehicle in re-entry phase. In: *2016 35th Chinese Control Conference (CCC) IEEE*. 10736-10741.

[4] W. Fang, G. Ying, Z. Zheng, H. Changchun, Z. Qun. 2017. Robust backstepping control of reusable launch vehicle in reentry phase based on disturbance observer. In *Chinese Control Conference. IEEE*. 4946-4951.

[5] L. Zhou, Z. Che, C. Yang. 2018. Disturbance observer-based integral sliding mode control for singularly perturbed systems with mismatched disturbances. *IEEE Access*. 6: 9854-9861.

[6] Z. Zhou, B. Zhang, D. Mao. 2018. Robust sliding mode control of PMSM based on a rapid nonlinear tracking differentiator and disturbance observer. *Sensors*. 18(4): 1031-1049.

[7] R. Javidan, H. Khuban. 2018. Optimal non-integer PID controller for a class of nonlinear systems: multi-objective modified black hole optimization algorithm. *Neural Comput. & Applic.* 30(1): 329-339.

[8] A. Moradvandi, M. Shahrokhi, S. Malek. 2018. Adaptive fuzzy decentralized control for a class of MIMO large-scale nonlinear state delay systems with unmodeled dynamics subject to unknown input saturation and infinite number of actuator failures. *Information Sciences*. 475: 121-141.

[9] Y. Huang, S. Li, J. Sun. 2018. Mars entry fault-tolerant control via neural network and structure adaptive model inversion. *Advances in Space Research*. 63: 557-571.

[10] S. Chao, S. Yang, B. Buckham. 2017. Trajectory Tracking Control of an Autonomous Underwater Vehicle Using Lyapunov-Based Model Predictive Control. *IEEE Transactions on Industrial Electronics*. 65(7): 5796-5805.

[11] Q. Hu, G. Niu, C. Wang. 2017. Spacecraft attitude fault-tolerant control based on iterative learning observer and control allocation. *Aerospace Science and Technology*. 75: 245-253.

[12] L. Chu, Q. Li, F. Gu, X. Du, Y. He, Y. Deng. 2022. Design, modeling, and control of morphing aircraft: a review," *Chinese Journal of Aeronautics*. 35(5): 220-246.

[13] Z. Guo, J. Guo, J. Zhou. 2018. Adaptive attitude tracking control for hypersonic reentry vehicles via sliding mode-based coupling effect-triggered approach. *Aerospace Science and Technology*, 78: 228–240.

[14] F. Ansarieshlaghi, P. Eberhard. 2019. Trajectory tracking control of a very flexible robot using a feedback linearization controller and a nonlinear observer. *ROMANSY 22–Robot Design, Dynamics and Control*. CISM International Centre for Mechanical Sciences (Courses and Lectures), Cham, Springer. 584: 26-33.

[15] X. Huang, Y. Yan. 2017. Saturated backstepping control of underactuated spacecraft hovering for formation flights. *IEEE Trans. Aerosp. Electron. Syst.* 53(4): 1988-2000.

[16] X. Zhang. 2019. Research on multi-mode adaptive control method for RLV. PhD Thesis. Harbin Institute of Technology, school of aeronautics.

[17] M. Ran, C. Wang, H. Liu, W. Wang, J. Lv. 2022. Research status and future development of morphing aircraft control technology. Acta Aeronautica et Astronautica Sinica. 43(10): 527449.

[18] B. Luo D. Liu, T. Huang, D. Wang. 2016. Model-free optimal tracking control via critic-only Q-learning. *IEEE Trans Neural Networks Learn Syst.* 27(10): 2134-2144.

[19] L. Gong, Q. Wang, C. Hu, C. Liu. 2020. Switching control of morphing aircraft based on Q-learning. *Chinese Journal of Aeronautics*. 33(2): 672-687.