SCIENCE CHINA
Technological Sciences

•Article•

June 2024 Vol.67 No.6: 1893–1906
https://doi.org/10.1007/s11431-023-2616-3

Adaptive learning-based optimal tracking control system design and analysis of a disturbed nonlinear hypersonic vehicle model

AN Kai, WANG ZhenGuo & HUANG Wei*

Hypersonic Technology Laboratory, National University of Defense Technology, Changsha 410073, China

Received October 15, 2023; accepted February 1, 2024; published online May 28, 2024

We propose an adaptive learning-based optimal control scheme for height-velocity control models considering model uncertainties and external disturbances of hypersonic winged-cone vehicles. The longitudinal nonlinear model is first established and transformed into the control-oriented error equations, and the control scheme is organized by a steady-compensation combination. To overcome and eliminate the impact of model uncertainties and external disturbances, an adaptive radial basis function neural network (RBFNN) is designed by a q-gradient approach. Taking the height-velocity error system with estimated uncertainties into account, the adaptive learning-based optimal tracking control (ALOTC) scheme is proposed by combining the critic-only adaptive dynamic programming (ADP) framework and parameter optimization of system settling time. Furthermore, a novel weight update law is proposed to satisfy the online iteration requirements, and the algorithm convergence and closed-loop stability are discussed by the Lyapunov theory. Finally, four simulation cases are provided to prove the effectiveness, accuracy, and robustness of the proposed scheme for the hypersonic longitudinal control system.

optimal tracking control, adaptive dynamic programming, RBFNN, hypersonic vehicle

Citation: An K, Wang Z G, Huang W. Adaptive learning-based optimal tracking control system design and analysis of a disturbed nonlinear hypersonic vehicle model. Sci China Tech Sci, 2024, 67: 1893–1906, https://doi.org/10.1007/s11431-023-2616-3

1 Introduction

The great hypersonic flight systems, especially the height-velocity control system, play an enormous role in guiding the aircraft to achieve stable flight missions. Owing to the features of high speed, strong coupling, nonlinearity, and rapid response, the comprehensive performance design and analysis of the height-velocity tracking control system with disturbances are facing great challenges.

During the decades of development, the adaptive strategies have always been a significant focus on different control approaches for hypersonic flight control. Nair et al. [1] developed a mechanism for adjusting the proportional-integral-derivative (PID) controller parameters using the gradient method and Lyapunov stability theory. For the velocity-loop and altitude-loop subsystem, the proposed compound adaptive controller has essentially an error-negative feedback control principle with the adaptive disturbance parameter update laws [2]. Yin et al. [3] have proposed an adaptive scaling strategy for prescribed performance dynamic surface control, which the structural parameters are set according to initial flight state errors and constants. Huang et al. [4] introduced the fractional exponential term into dynamic inversion and backstepping control methods as an adaptive adjustment of controller convergence for velocity and altitude subsystems. In sliding mode control, the adaptivity of the control system is mainly reflected by the types of sliding mode surface, such as exponential sliding mode surface [5], terminal sliding mode surface [6], and super-twist sliding mode surface [7]. However, the application

*Corresponding author (email: gladrain2001@163.com)

scope of the above designed adaptive strategies is limited in the time-varying characteristic parameters of the hypersonic control system. Therefore, those methods for fast-variation parameter estimation are unable to effectively provide reliable approximation.

With the development of artificial intelligence technology and optimization theory, the intelligent control methods have received much attention for their advanced approximation ability and rapid response capability compared with traditional adaptive tricks. The neural network (NN) has the typical function properties with infinite approximation, which is regularly employed in the adaptive learning-based control methodologies. After Bellman first published the dynamic programming (DP) algorithm in 1966 [8], the branch of adaptive dynamic programming (ADP) was gradually expanded through the integration of optimal control techniques with reinforcement learning. Reinforcement learning mainly focuses on the interaction process between the agent and the current environment. After observing the present state from the environment, the agent makes a feasible decision by critic NN according to the total rewards calculated by actor NN. Therefore, the goal of reinforcement learning is to find a suitable strategy to maximize current and future expected returns. Based on the above actor-critic framework, the ADP approach is developed to obtain optimal policy and value by introducing optimal control theory.

To this end, many researchers have done lots of theoretical design and analysis about the ADP applied in hypersonic vehicle control systems [9,10]. Commonly, the major consideration of ADP-based controllers is to design the effective weight update law for actor and critic NN, which is deduced by the gradient descent method respectively [11–13]. Many existent main results are all obtained by combining the Lyapunov stability theory and backstepping control [14–17], the novel adaptive rules are generally deduced by innovative Lyapunov candidate functions. It seems that the more complex the update laws, the better the system control performance. However, the expected requirements are generally unsatisfied when the current update law is transferred to other control tasks. Therefore, how to design an adaptive learning control scheme to achieve desired control performance for different control objects and scenarios is the main focus of this article.

The model uncertainties caused by external disturbances and the introduced assumptions owing to the nonlinear properties of hypersonic vehicles bring enormous challenges to the overall flight control system. In order to eliminate the unstable chattering and potential divergence, the uncertainties can be approximated by the radial basis function neural network (RBFNN) surrogate model. Compared with the existent results [18,19], the RBFNN not only takes the computational efficiency into account, but also maintains the high approximation accuracy [20]. The RBFNN is commonly utilized to estimate the disturbances [21], uncertain state parameters [22], and unknown nonlinear functions [23]. However, the approximation performance of RBFNN is mainly determined by feasible adaptive weight rules and other flexible parameters. Therefore, how to design an effective RBFNN compensator to achieve advanced properties of rapid disturbance response and low approximation error is the second attention of this article.

Based on those discussions, we propose an adaptive learning-based optimal tracking control scheme with a novel RBFNN disturbance observer for the hypersonic height-velocity control system. The proposed scheme establishes a control-oriented model of longitudinal flight states to achieve integrated adaptive full-state optimal control. The control inputs are designed to include optimal control and compensation control. Based on the developed adaptive learning-based optimal tracking control (ALOTC) algorithm, the value iteration and parameter optimization are data-driven, and the optimal feedback control policy is obtained. The designed adaptive RBFNN disturbance observer is utilized to estimate the chattering disturbance and uncertainties. The main motivations and contributions can be unfolded as follows.

(1) As far as we all know, the ADP-based optimal control framework implements the iterative control just through the designed actor or critic update laws, which leads to the system response time can not achieving the optimal solution when the overall control sequence is being solved, such as refs. [9,15]. Compared with these results, the optimization and statistical analysis of response time are considered in the ALOTC scheme to improve the control performance of the tracking states.

(2) Based on the concepts of q-calculus [24], an adaptive q-parameter learning rule for the RBFNN disturbance observer is proposed by the stochastic descent algorithm. The convergence condition is provided for the proposed approach and extensive comparative analysis of this work is carried out with the existent results [25,26].

(3) In contrast with complex equations of critic NN in ref. [14], a novel single critic NN weight update law is proposed to obtain the optimal NN weights iteratively, which has concise expression but fast convergence based on the additional parameter error feedback term.

2 Problem formulation

As shown in Figure 1, the longitudinal dynamics in the earth and body coordinate system can be deduced in eq. (1), which can also be derived from 6-DOF full-state equations of the vehicle considering the lateral states are all zero.

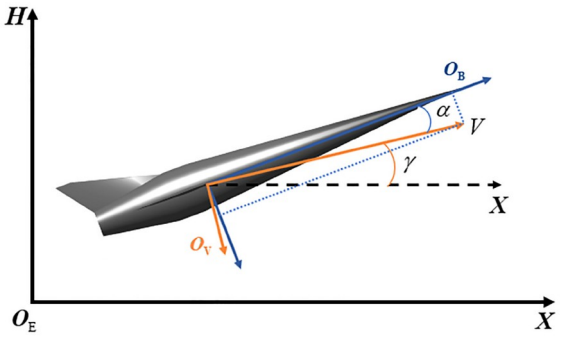

Figure 1 (Color online) The diagram of longitudinal motion.

$$\begin{cases} \dot{V} = \dfrac{T\cos\alpha - D}{m} - \dfrac{\mu\sin\gamma}{r^2}, \\ \dot{h} = V\sin\gamma, \\ \dot{\gamma} = \dfrac{L + T\sin\alpha}{mV} - \dfrac{\left(\mu - V^2 r\right)\cos\gamma}{Vr^2}, \\ \dot{\alpha} = \bar{q} - \dot{\gamma}, \\ \dot{\bar{q}} = \dfrac{M_{yy}}{I_{yy}}, \end{cases} \tag{1}$$

where V and h are the velocity and height of the hypersonic vehicle, respectively, γ, α, and $\bar{q}$ represent the angle information defined as flight path angle, angle of attack, and pitch angle rate. The physical parameters mass m and moment of inertia I_{yy} are provided for calculation. The term $g = \mu/r^2$ represents the acceleration of gravity, where μ and R_e in equation $r = h + R_e$ are the atmospheric parameters and earth radius. The aerodynamics, thrust, and moments are described as follows:

$$\begin{aligned} L &= C_L Q S_{\text{ref}}, \\ D &= C_D Q S_{\text{ref}}, \\ T &= C_T Q S_{\text{ref}}, \\ M_{yy} &= C_m Q S_{\text{ref}} c, \end{aligned} \tag{2}$$

where c and S_{ref} represent the mean aerodynamic chord and reference area. $Q = 0.5\rho V^2$ is the dynamic pressure, ρ is the atmospheric density. Based on the configuration in Figure 1, the nonlinear aerodynamic model is provided by ref. [27]. Therefore, the terms C_L, C_D, C_T, and C_m can be deduced as follows:

$$\begin{aligned} C_L &= C_L^{\alpha}\alpha + C_L^0, \\ C_D &= C_D^{\alpha^2}\alpha^2 + C_D^{\alpha}\alpha + C_D^0, \\ C_T &= C_T^{\phi}\phi + C_T^0, \\ C_m &= C_m^{\alpha} + C_m^{q} + C_m^{\delta_e}, \\ C_m^{\alpha} &= C_m^{\alpha^2}\alpha^2 + C_m^{\alpha}\alpha + C_m^0, \\ C_m^{q} &= \frac{cq}{2V}\left(C_m^{\alpha^2 q}\alpha^2 + C_m^{\alpha q}\alpha + C_m^{q}\right), \\ C_m^{\delta_e} &= c_e\left(\delta_e - \alpha\right), \end{aligned} \tag{3}$$

where ϕ and δ_e are the actual control variables throttle and elevator angle.

For the convenience of controller design, the dynamic equations are generally rewritten as an affine nonlinear model in eq. (4) according to ref. [28],

$$\dot{\mathbf{x}} = f(\mathbf{x}) + g(\mathbf{x})\mathbf{u} + d_j, \tag{4}$$

where $\mathbf{x} = [V,\ h,\ \gamma,\ \theta,\ \bar{q}]^{\mathrm{T}} \in \mathbb{R}^5$, $\mathbf{u} = [\phi,\ \delta_e]^{\mathrm{T}} \in \mathbb{R}^2$, $f(\mathbf{x})$ and $g(\mathbf{x})$ represent the nonlinear terms, $\|d_j\| \leqslant d_{j,\max}$, $j = V, h, \gamma, \theta, \bar{q}$ indicate the unknown model uncertainties and disturbances of the system, $d_{j,\max}$ is the constant upper bound, θ is the pitch angle.

Then, substitute eqs. (2) and (3) into eq. (1) and introduce the desired height signal h_d and velocity signal V_d, the control-oriented error dynamic equations can be deduced as shown in eqs. (5) and (6):

$$\begin{aligned} \dot{e}_V &= \dot{V} - \dot{V}_d \\ &= \frac{-D + C_T^0 Q S_{\text{ref}}\cos\alpha}{m} - \frac{\mu\sin\gamma}{r^2} + \frac{C_T^{\phi} Q S_{\text{ref}}\cos\alpha}{m}\phi - \dot{V}_d \\ &= f_V(\mathbf{x}) + g_V(\mathbf{x})\phi - \dot{V}_d + d_V, \end{aligned} \tag{5}$$

$$\begin{aligned} \dot{e}_h &= \dot{h} - \dot{h}_d = V\sin\gamma - \dot{h}_d + d_h, \\ \dot{e}_\gamma &= \dot{\gamma} - \dot{\gamma}_d^* \\ &= \frac{T\sin\alpha + \left(C_L^0 - C_L^{\alpha}\gamma\right) Q S_{\text{ref}}}{mV} \\ &\quad - \frac{\left(\mu - V^2 r\right)\cos\gamma}{Vr^2} + \frac{C_L^{\alpha} Q S_{\text{ref}}}{mV}\theta - \dot{\gamma}_d^* \\ &= f_\gamma(\mathbf{x}) + g_\gamma(\mathbf{x})\theta - \dot{\gamma}_d^* + d_\gamma, \\ \dot{e}_\theta &= \dot{\theta} - \dot{\theta}_d^* = \bar{q} - \dot{\theta}_d^* \\ &= f_\theta(\mathbf{x}) + g_\theta(\mathbf{x})\bar{q} - \dot{\theta}_d^* + d_\theta, \\ \dot{e}_{\bar{q}} &= \dot{\bar{q}} - \dot{\bar{q}}_d^* \\ &= \frac{\left(C_m^{\alpha} + C_m^{\bar{q}} - c_e\alpha\right) Q S_{\text{ref}} c}{I_{yy}} + \frac{c_e Q S_{\text{ref}} c}{I_{yy}}\delta_e - \dot{\bar{q}}_d^* \\ &= f_{\bar{q}}(\mathbf{x}) + g_{\bar{q}}(\mathbf{x})\delta_e - \dot{\bar{q}}_d^* + d_{\bar{q}}. \end{aligned} \tag{6}$$

In eq. (6), the desired flight path angle γ_d^* is derived in eq. (7) by the height error equation $\gamma_d^* = \arcsin\left(\left(\dot{e}_h + \dot{h}_d\right)/V\right)$ and the designed proportional-integral (PI) dynamic convergence mode $\dot{e}_h = -k_{h\text{p}} e_h - k_{h\text{I}}\int e_h$.

Similarly, both θ_d^* and $\bar{q}_d^*$ are obtained by the first-order differential equation (eq. (8)) to avoid the differential explosion problem.

$$\gamma_d^* = \arcsin\left(\frac{-k_{h\text{p}} e_h - k_{h\text{I}}\int e_h + \dot{h}_d}{V}\right), \tag{7}$$

$$\begin{aligned} \tau\dot{\theta}_d^* + \theta_d^* &= \theta, \quad \theta_d^*(0) = \theta(0), \\ \tau\dot{\bar{q}}_d^* + \bar{q}_d^* &= \bar{q}, \quad \bar{q}_d^*(0) = \bar{q}(0). \end{aligned} \tag{8}$$

Based on the aforementioned system transformation description, the final control-oriented error equations can be rewritten as the matrix form

$$\begin{aligned}&\dot{e}_V = f_V(\mathbf{x}) - \dot{V}_\mathrm{d} + d_V + g_V(\mathbf{x})\phi,\\&\begin{bmatrix}\dot{e}_\gamma\\ \dot{e}_\theta\\ \dot{e}_{\bar{q}}\end{bmatrix} = \begin{bmatrix}f_\gamma(\mathbf{x}) - \dot{\gamma}_\mathrm{d}^* + d_\gamma\\ f_\theta(\mathbf{x}) - \dot{\theta}_\mathrm{d}^* + d_\theta\\ f_{\bar{q}}(\mathbf{x}) - \dot{\bar{q}}_\mathrm{d}^* + d_{\bar{q}}\end{bmatrix} + \begin{bmatrix}g_\gamma(\mathbf{x}) & 0 & 0\\ 0 & g_\theta(\mathbf{x}) & 0\\ 0 & 0 & g_{\bar{q}}(\mathbf{x})\end{bmatrix}\begin{bmatrix}\theta\\ \bar{q}\\ \delta_\mathrm{e}\end{bmatrix},\end{aligned} \tag{9}$$

where $u = \phi$ for velocity subsystem and $u = [\theta, \bar{q}, \delta_\mathrm{e}]^\mathrm{T} \in \mathbb{R}^3$ for height subsystem.

Furthermore, considering the achievement of optimality of major dynamics and robustness of disturbance approximation, the integrated control policy is divided into two parts in this article for design as shown in eq. (10).

$$u = u_\mathrm{op} + u_\mathrm{comp}, \tag{10}$$

where u_op and u_comp represent the optimal controller and compensative controller, respectively.

To this end, the main purposes of this article conclude the two aspects.

(1) Design an adaptive learning-based optimal controller for the hypersonic height-velocity comprehensive control system (eq. (9)) to accurately track the desired signals h_d and V_d, and the angle states can also be stabilized by the virtual control.

(2) Construct an adaptive RBFNN-based compensation controller to implement the approximation of the unknown disturbances to enhance the robustness of the height-velocity control system.

3 Adaptive learning-based optimal control system design and analysis

3.1 Adaptive compensation controller design

In order to eliminate the impact of disturbances on the closed-loop system, an adaptive RBFNN-based disturbance observer is first introduced to approximate the unknown system uncertainties d_j on the right side of eq. (4). The conventional architecture of RBFNN is composed of an input layer, a hidden layer, and an output layer, in which the hidden layer is activated by a number of Gaussian kernel functions applied commonly. Let $\mathbf{x} \in \mathbb{R}^{m_0}$ be the input vector, then the overall mapping with the Gaussian kernel function of the RBFNN is provided as

$$\begin{aligned}&y = \sum_{i=1}^{m_1} w_i \phi_i(\| \mathbf{x} - \mathbf{c}_i \|) + b,\\&\phi_i(\| \mathbf{x} - \mathbf{c}_i \|) = \exp\left(-\frac{\| \mathbf{x} - \mathbf{c}_i \|^2}{b_w^2}\right),\end{aligned} \tag{11}$$

where m_1 is the number of neurons in the hidden layer, w_i are the weights of all the neurons connecting the hidden layer and output layer, ϕ_i are the Gaussian kernel function, b is the bias term of the output, $\mathbf{c}_i \in \mathbb{R}^{m_0}$ are the pre-defined center vector of the RBFNN, b_w is the width of the Gaussian kernel.

Commonly, the infinite approximation ability of RBFNN can be improved by the specific weight update rules, which can be deduced by the gradient descent algorithm. For the implementation of the gradient descent algorithm, the cost function is first defined to minimize the approximation error according to approximation value $\hat{d}_j$, which is shown in

$$\varepsilon_\mathrm{RBF} = \frac{1}{2}\left(d_j - \hat{d}_j\right)^2 = \frac{1}{2}e_d^2. \tag{12}$$

In the conventional gradient descent algorithm, the gradient of ε_RBF respect to w_i is evaluated by the chain rule as shown in

$$-\nabla_{w_i}\varepsilon_\mathrm{RBF} = -\frac{\partial \varepsilon_\mathrm{RBF}}{\partial e_d} \times \frac{\partial e_d}{\partial \hat{d}_j} \times \frac{\partial \hat{d}_j}{\partial w_i}, \tag{13}$$

then the weight update rule can be obtained as

$$w_i(k+1) = w_i(k) - \eta \nabla_{w_i}\varepsilon_\mathrm{RBF}, \tag{14}$$

where k is the current iteration step and η is the learning rate.

In order to achieve the desired convergence property, the regulation of derivation calculation is generally evaluated by different calculus rules. A q-factor gradient descent-based RBFNN has been proposed in ref. [25] which introduces the concept of q-calculus into the gradients of the cost function, the proposed convergence expression is given as shown in eq. (15), and the weights are updated by the following rule:

$$w_i(k+1) = w_i(k) + \eta\frac{(q(k)+1)}{2}\phi_i(\mathbf{x}, \mathbf{c}_i)e_d, \tag{15}$$

where q is the pre-designed convergence term. The robustness of q-RBFNN has been analyzed and guaranteed using the small gain theorem that the convergence condition of q must be satisfied as shown in

$$0 < q(k) < \frac{1}{\mu \left\| \phi(k) \right\|^2}, \tag{16}$$

where μ is the step size.

Based on the above proposed results, a novel convergence law of q is proposed and an adaptive RBFNN disturbance observer is trained as follows:

$$\begin{aligned}&w_i(k+1) = w_i(k) + \eta q(k)\phi_i(\mathbf{x}, \mathbf{c}_i)e_d,\\&q(k+1) = k_q q(k) + k_p\left(\exp\left(e_d^2\right) - 1\right),\end{aligned} \tag{17}$$

where $0 < k_q \leqslant 1$, $k_p \in \mathbb{R}^+$.

Then the disturbances in eqs. (5) and (6) can be approximated by

$$\hat{d}_j = \sum_{i=1}^{m_1} w_i \exp\left(-\frac{\left\| \mathbf{x} - \mathbf{c}_i \right\|^2}{b_w^2}\right) + b. \tag{18}$$

Based on the above analysis and results of the adaptive RBFNN disturbance observer, the compensative controller

can be designed as

$$\begin{aligned}
\phi_{\text{comp}} &= g_V^{-1}(\mathbf{x})(-f_V(\mathbf{x}_\text{d}) + \dot{V}_\text{d} - \hat{d}_V),\\
\theta_{\text{comp}} &= g_\gamma^{-1}(\mathbf{x})(-f_\gamma(\mathbf{x}_\text{d}) + \dot{\gamma}_\text{d}^* - \hat{d}_\gamma),\\
\overline{q}_{\text{comp}} &= g_\theta^{-1}(\mathbf{x})(-f_\theta(\mathbf{x}_\text{d}) + \dot{\theta}_\text{d}^* - \hat{d}_\theta),\\
\delta_{\text{e,comp}} &= g_{\overline{q}}^{-1}(\mathbf{x})(-f_{\overline{q}}(\mathbf{x}_\text{d}) + \dot{\overline{q}}_\text{d}^* - \hat{d}_{\overline{q}}),
\end{aligned} \tag{19}$$

where $\mathbf{x}_\text{d}$ is the desired state vector and $f_j(\mathbf{x}_\text{d})$ is the desired values of the major dynamic.

3.2 Adaptive optimal controller design

Substitute eqs. (10) and (19) into eq. (9), the major dynamic equations can be split from the error subsystems (eqs. (5) and (6)) as follows:

$$\begin{aligned}
\dot{e}_V &= f_V(\mathbf{x}) - f_V(\mathbf{x}_\text{d}) + g_V(\mathbf{x})\phi_\text{op}^* + \tilde{d}_V\\
&= \overline{f}_V(\mathbf{x}, \mathbf{x}_\text{d}) + g_V(\mathbf{x})\phi_\text{op} + \tilde{d}_V,
\end{aligned} \tag{20}$$

$$\begin{aligned}
\dot{e}_\gamma &= f_\gamma(\mathbf{x}) - f_\gamma(\mathbf{x}_\text{d}) + g_\gamma(\mathbf{x})\theta_\text{op} + \tilde{d}_\gamma\\
&= \overline{f}_\gamma(\mathbf{x}, \mathbf{x}_\text{d}) + g_\gamma(\mathbf{x})\theta_\text{op} + \tilde{d}_\gamma,\\
\dot{e}_\theta &= \overline{q}_\text{op} + \tilde{d}_\theta,\\
\dot{e}_{\overline{q}} &= f_{\overline{q}}(\mathbf{x}) - f_{\overline{q}}(\mathbf{x}_\text{d}) + g_{\overline{q}}(\mathbf{x})\delta_{\text{e,op}} + \tilde{d}_{\overline{q}}\\
&= \overline{f}_{\overline{q}}(\mathbf{x}, \mathbf{x}_\text{d}) + g_{\overline{q}}(\mathbf{x})\delta_{\text{e,op}} + \tilde{d}_{\overline{q}},
\end{aligned} \tag{21}$$

where the disturbance approximation error $\tilde{d}_i = d_i - \hat{d}_i$, and assume the condition $\|\tilde{d}_i\| \leqslant d_k$ is satisfied.

Assumption 1: For a set $\Omega \in \mathbb{R}$, the dynamic terms $f(\mathbf{x})$ and $g(\mathbf{x})$ are all Lipschitz continuous and all bounded by existent positive constants $b_{fV}, b_{f\gamma}, b_{f\overline{q}}, b_{gV}, b_{g\gamma}, b_{g\overline{q}}$ as follows:

$$\begin{aligned}
&\|\overline{f}_V(\mathbf{x}, \mathbf{x}_\text{d})\| \leqslant b_{fV}, \quad \|g_V(\mathbf{x})\| \leqslant b_{gV},\\
&\|\overline{f}_\gamma(\mathbf{x}, \mathbf{x}_\text{d})\| \leqslant b_{f\gamma}, \quad \|g_\gamma(\mathbf{x})\| \leqslant b_{g\gamma},\\
&\|\overline{f}_{\overline{q}}(\mathbf{x}, \mathbf{x}_\text{d})\| \leqslant b_{f\overline{q}}, \quad \|g_{\overline{q}}(\mathbf{x})\| \leqslant b_{g\overline{q}}.
\end{aligned} \tag{22}$$

Without loss of generality, in order to find the optimal control policy, considering the integral value function given by

$$J(e_i) = \int_t^\infty \exp(-\kappa(s-t))r_i(e_i, u_i)\text{d}s, \quad i = h, V, \tag{23}$$

where κ is the discount factor, $J(e_i) \in \mathbb{R}^+$, $r_i(e_i, u_i)$ are the reward function, which is generally designed as a quadratic function of the control inputs and state errors, as shown in eq. (24).

$$r_i = e_i^\text{T}\mathbf{Q}_i e_i + u_i^\text{T}\mathbf{R}_i u_i, \tag{24}$$

where $\mathbf{Q}_h \in \mathbb{R}^{3\times3}$, $\mathbf{R}_h \in \mathbb{R}^{3\times3}$ for $i = h$ and $\mathbf{Q}_V \in \mathbb{R}$, $\mathbf{R}_V \in \mathbb{R}$ for $i = V$.

Then the optimal cost function can be obtained as follows:

$$J^*(e_i) = \min_{u_{i,\text{op}} \in \Omega_u} \int_t^\infty \exp(-\kappa(s-t))r_i(e_i, u_{i,\text{op}})\text{d}s, \tag{25}$$

where $\Omega_u \in \mathbb{R}$ is the admissible control set.

According to the optimal control theory, the Hamilton-Jacobi-Bellman (HJB) equation can be deduced by taking the first-order time derivative of eq. (25):

$$H_i^*\left(e_i, u_{i,\text{op}}, J_i^*\right) = e_i^\text{T}\mathbf{Q}_i e_i + u_{i,\text{op}}^\text{T}\mathbf{R}_i u_{i,\text{op}} - \kappa J_i^* + \nabla J_i^* \cdot \dot{e}_i. \tag{26}$$

Combining eqs. (25) and (26), the optimality condition is defined as

$$\min_{u_{i,\text{op}} \in \Omega_u}\left\{H_i^*\left(e_i, u_{i,\text{op}}, J_i^*\right)\right\} = 0. \tag{27}$$

Then, the optimal control policy can be obtained from the time derivative of HJB equation (eq. (26)) as follows:

$$u_{i,\text{op}} = -\frac{1}{2}\mathbf{R}_i^{-1}g(\mathbf{x})^\text{T}\nabla J_i^*. \tag{28}$$

Remark 1: The terms $\mathbf{R}_i$ and $g(\mathbf{x})$ have been known from the aforementioned discussion, but the last term $\partial J_i^* / \partial \mathbf{x}$ must be calculated by solving the HJB equation (eq. (26)) because of the inherent nonlinearity. Therefore, an NN-approximation-based reinforcement learning algorithm is developed in the following sections to evaluate the cost function and optimal policy, which can avoid the large computation consumption of solving the HJB equation.

3.3 Reinforcement learning value function approximation: critic NN structure

Assume the optimal cost function is smooth, then the optimal value function can be approximated by an NN structure:

$$J^*(e_i) = \mathbf{W}_{i,\text{c}}^{*\text{T}}\sigma_\text{c}(e_i) + \varepsilon_{i,\text{c}}^*, \tag{29}$$

where $\mathbf{W}_{i,\text{c}}^* \in \mathbb{R}^{n\times1}$ is the ideal weight vector (n is the vector dimension), $\sigma_{i,\text{c}} \in \mathbb{R}^{n\times1}$ is the activation function and the second term $\varepsilon_{i,\text{c}}^* \in \mathbb{R}$, $|\varepsilon_{i,\text{c}}^*| \leqslant \varepsilon_{i,\text{bound}}$ is the bounded approximation error.

Then, the gradient of the approximated optimal value function J_i^* is defined as follows:

$$\nabla J^*(e_i) = \nabla \sigma_{i,\text{c}}^\text{T}\mathbf{W}_{i,\text{c}}^* + \nabla \varepsilon_{i,\text{c}}^*. \tag{30}$$

Assume the gradients of $\nabla \varepsilon_{i,\text{c}}^*$ are all bounded meeting $\|\nabla \varepsilon_{i,\text{c}}^*\| \leqslant b_{i,\varepsilon}$, then the estimation of the optimal value function is deployed as

$$\hat{J}(e_i) = \hat{\mathbf{W}}_{i,\text{c}}^\text{T}\sigma_{i,\text{c}}, \tag{31}$$

where $\hat{\mathbf{W}}_{i,\text{c}}$ is the estimation value of $\mathbf{W}_{i,\text{c}}^*$.

Substitute eq. (31) into eq. (28), and the approximation of optimal control policy is rewritten as

$$\hat{u}_{i,\text{op}} = -\frac{1}{2}\mathbf{R}_i^{-1}g(\mathbf{x})^\text{T}\left(\nabla \sigma_{i,\text{c}}^\text{T}\hat{\mathbf{W}}_{i,\text{c}}\right). \tag{32}$$

Then, the Bellman error function can be derived by the temporal-difference (TD) learning algorithm as follows:

$$\begin{aligned}\hat{J}\left(e_i(t-T)\right)&=\int_{t-T}^{\infty}\exp(-\kappa(s-(t-T)))r_i\left(e_i,\hat{u}_{i,\mathrm{op}}\right)\mathrm{d}s\\&=\int_{t-T}^{t}\exp(-\kappa(s-(t-T)))r_i\left(e_i,\hat{u}_{i,\mathrm{op}}\right)\mathrm{d}s\\&\quad+\int_{t}^{\infty}\exp(-\kappa(s-(t-T)))r_i\left(e_i,\hat{u}_{i,\mathrm{op}}\right)\mathrm{d}s\\&=\int_{t-T}^{t}\exp(-\kappa(s-(t-T)))r_i\left(e_i,\hat{u}_{i,\mathrm{op}}\right)\mathrm{d}s+\hat{J}\left(e_i(t)\right),\end{aligned}\tag{33}$$

$$\begin{aligned}\varepsilon_{i,\mathrm{c\text{-}B}}&=\hat{J}\left(e_i(t)\right)-\hat{J}\left(e_i(t-T)\right)+R_{i,\mathrm{step}}(t),\\&=\widehat{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\sigma_{\mathrm{c}}\left(e_i(t)\right)-\widehat{\mathbf{W}}_{\mathrm{c}}^{\mathrm{T}}\sigma_{\mathrm{c}}\left(e_i(t-T)\right)+R_{i,\mathrm{step}}(t),\\&=\widehat{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left(\sigma_{\mathrm{c}}\left(e_i(t)\right)-\sigma_{\mathrm{c}}\left(e_i(t-T)\right)\right)+R_{i,\mathrm{step}}(t),\\&=\widehat{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}+R_{i,\mathrm{step}}(t),\end{aligned}\tag{34}$$

where T is the time interval and $\Delta\sigma_{i,\mathrm{c}}=\sigma_{\mathrm{c}}\left(e_i(t)\right)-\sigma_{\mathrm{c}}\left(e_i(t-T)\right)$, $R_{i,\mathrm{step}}(t)$ is the one-time-step reward in eq. (14).

$$R_{i,\mathrm{step}}(t)=\int_{t-T}^{t}\exp(-\kappa(s-(t-T)))r_i\left(e_i,\hat{u}_{i,\mathrm{op}}\right)\mathrm{d}s.\tag{35}$$

Together with the performance index function $E_{i,\varepsilon}=0.5\varepsilon_{i,\mathrm{c\text{-}B}}^{2}$, the minimal $\varepsilon_{i,\mathrm{c\text{-}B}}$ can be obtained by the gradient descent method, and then the update law of $\dot{\widehat{\mathbf{W}}}_{i,\mathrm{c}}$ is designed as

$$\dot{\widehat{\mathbf{W}}}_{i,\mathrm{c}}=\left(\frac{\Gamma_{i,\mathrm{c}}}{\left(1+\Delta\sigma_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right)^{2}}+\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\right)\Delta\sigma_{i,\mathrm{c}}\varepsilon_{i,\mathrm{c\text{-}B}},\tag{36}$$

where the weight estimation error is determined by $\overline{\mathbf{W}}_{i,\mathrm{c}}=\mathbf{W}_{i,\mathrm{c}}^{*}-\widehat{\mathbf{W}}_{i,\mathrm{c}}$. The first term of the right-hand side represents the robust adaptive adjustment term, which can guarantee the convergence of $\overline{\mathbf{W}}_{i,\mathrm{c}}$, and $\mathbf{\Gamma}_{i,\mathrm{c}}\in\mathbb{R}^{n\times n}$ is a positive-definite matrix representing the gain coefficient. $\mathbf{k}_{i,W}\in\mathbb{R}^{n\times 1}$ can accelerate convergence with the appropriate value.

3.4 Stability analysis

The stability analysis of the proposed ALOTC scheme with RBFNN disturbance observer applied in the closed-loop system are discussed in this section. The detailed results are as follows.

Theorem 1: For the control-oriented dynamic model (eq. (9)) of the hypersonic vehicle, if the control input is designed as eq. (10) with the proposed ALOTC algorithm and disturbance observer, then the state errors and weight approximation errors can satisfy the ultimately uniformly bounded (UUB) convergence condition.

Proof: Define the Lyapunov candidate function as follows:

$$L_{\mathrm{p}}=J^{*}(e_i)+\frac{1}{2}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Gamma_{i,\mathrm{c}}^{-1}\overline{\mathbf{W}}_{i,\mathrm{c}},\tag{37}$$

$$\begin{aligned}\dot{L}_{\mathrm{p}}&=\dot{J}^{*}(e_i)+\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{\Gamma}_{i,\mathrm{c}}^{-1}\dot{\overline{\mathbf{W}}}_{i,\mathrm{c}}\\&=\nabla\sigma_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{W}_{i,\mathrm{c}}^{*}\left(f(e_i)+g(e_i)\hat{u}_{i,\mathrm{op}}^{*}+\tilde{d}_i\right)\\&\quad+\nabla\varepsilon_{i,\mathrm{c}}\left(f(e_i)+g(e_i)\hat{u}_{i,\mathrm{op}}^{*}+\tilde{d}_i\right)+\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{\Gamma}_{i,\mathrm{c}}^{-1}\dot{\overline{\mathbf{W}}}_{i,\mathrm{c}}.\end{aligned}\tag{38}$$

Then, the first term on the right-hand side can be replaced with the HJB equation (eq. (26)) as

$$\begin{aligned}&\nabla\sigma_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{W}_{i,\mathrm{c}}^{*}\left(f(e_i)+g(e_i)\hat{u}_{i,\mathrm{op}}+\tilde{d}_i\right)\\&=-e_i^{\mathrm{T}}\mathbf{Q}_i e_i-u_{i,\mathrm{op}}^{\mathrm{T}}\mathbf{R}_i u_{i,\mathrm{op}}+\kappa\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\sigma_{i,\mathrm{c}}+H_i^{*}\\&\quad+\nabla\sigma_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{W}_{i,\mathrm{c}}^{*}g(e_i)\tilde{u}_{i,\mathrm{op}}\\&\quad-\nabla\sigma_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{W}_{i,\mathrm{c}}^{*}g(e_i)u_{i,\mathrm{op}}\\&\quad+\nabla\sigma_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{W}_{i,\mathrm{c}}^{*}\tilde{d}_i,\end{aligned}\tag{39}$$

and the third term can be deduced by the expression of $\dot{\widehat{\mathbf{W}}}_{i,\mathrm{c}}$ in eq. (36) as

$$\begin{aligned}&\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{\Gamma}_{i,\mathrm{c}}^{-1}\dot{\overline{\mathbf{W}}}_{i,\mathrm{c}}\\&=\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{\Gamma}_{i,\mathrm{c}}^{-1}\left(\frac{\Gamma_{i,\mathrm{c}}}{\left(1+\Delta\sigma_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right)^{2}}+\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\right)\Delta\sigma_{i,\mathrm{c}}\varepsilon_{i,\mathrm{c\text{-}B}}\\&=\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left(\frac{\mathbf{I}_{n\times n}}{\left(1+\Delta\sigma_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right)^{2}}+\mathbf{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\right)\Delta\sigma_{i,\mathrm{c}}\varepsilon_{i,\mathrm{c\text{-}B}}.\end{aligned}\tag{40}$$

Substitute eq. (34) into eq. (40), then the equality can be derived as

(41)

where

$$\begin{aligned}\Delta\bar{\sigma}_{i,\mathrm{c}}&=\frac{\Delta\sigma_{i,\mathrm{c}}}{\left(1+\Delta\sigma_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right)^{2}},\\\chi_i&=\int_{t-T}^{t}\exp(-\kappa(s-(t-T)))\\&\quad\times\left(\kappa\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\sigma_{i,\mathrm{c}}+H_i^{*}+\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\nabla\sigma_{i,\mathrm{c}}\left(f(e_i)+g(e_i)u_{i,\mathrm{op}}\right)\right)\mathrm{d}s.\end{aligned}\tag{42}$$

If the time interval is enough small, the integral term can be approximated as

$$\begin{aligned}&\int_{t-T}^{t}\exp(-\kappa(s-(t-T)))\left(-u_{i,\mathrm{op}}^{\mathrm{T}}\mathbf{R}_i u_{i,\mathrm{op}}+\hat{u}_{i,\mathrm{op}}^{\mathrm{T}}\mathbf{R}_i\hat{u}_{i,\mathrm{op}}\right)\mathrm{d}s\\&\approx\frac{T}{4}\exp(-\kappa T)\widehat{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Psi_i\widehat{\mathbf{W}}_{i,\mathrm{c}}-\frac{T}{4}\exp(-\kappa T)\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\mathbf{W}_{i,\mathrm{c}}^{*}\\&=\frac{T}{4}\exp(-\kappa T)\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Psi_i\overline{\mathbf{W}}_{i,\mathrm{c}}-\frac{T}{2}\exp(-\kappa T)\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\overline{\mathbf{W}}_{i,\mathrm{c}}.\end{aligned}\tag{43}$$

Then, eq. (41) can be reformulated by

$$\begin{aligned}&\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{\Gamma}_{i,\mathrm{c}}^{-1}\dot{\overline{\mathbf{W}}}_{i,\mathrm{c}}\\&=\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left(\Delta\bar{\sigma}_{i,\mathrm{c}}+\mathbf{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right)\\&\quad\times\left(\chi_i+\frac{T}{4}\exp(-\kappa T)\left(\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Psi_i\overline{\mathbf{W}}_{i,\mathrm{c}}-2\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\overline{\mathbf{W}}_{i,\mathrm{c}}\right)\right),\end{aligned}\tag{44}$$

and

$$\begin{aligned}\dot{L}_{\mathrm{p}} =& -e_i^{\mathrm{T}}\mathbf{Q}_i e_i - u_{i,\mathrm{op}}^{\mathrm{T}}\mathbf{R}_i u_{i,\mathrm{op}} + \kappa \mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\sigma_{i,\mathrm{c}} + H_i^* \\ &+ \mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\nabla\sigma_{i,\mathrm{c}}g(e_i)\hat{u}_{i,\mathrm{op}} \\ &- \mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\nabla\sigma_{i,\mathrm{c}}g(e_i)u_{i,\mathrm{op}} + \mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\nabla\sigma_{i,\mathrm{c}}\tilde{d}_i \\ &+ \nabla\varepsilon_{i,\mathrm{c}}^*\left(f(e_i) + g(e_i)\hat{u}_{i,\mathrm{op}} + \tilde{d}_i\right) \\ &+ \overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2\left(\mathbf{W}_{i,\mathrm{c}}^* - \overline{\mathbf{W}}_{i,\mathrm{c}}\right) \\ &+ \overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2\boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left(\mathbf{W}_{i,\mathrm{c}}^* - \overline{\mathbf{W}}_{i,\mathrm{c}}\right) \\ &+ \overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left(\Delta\bar{\sigma}_{i,\mathrm{c}} + \boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right) \\ &\times\left(\chi_i + \frac{T}{4}\exp(-\kappa T)\left(\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Psi\overline{\mathbf{W}}_{i,\mathrm{c}} - 2\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\overline{\mathbf{W}}_{i,\mathrm{c}}\right)\right). \end{aligned} \tag{45}$$

According to Assumption 1 and the eigenvalues of matrix $\mathbf{Q}_i$, we can obtain eq. (46), where $\lambda_{\min}(\mathbf{Q}_i)>0$. Define the vector $\mathbf{z} = \left[e_i^{\mathrm{T}}, \overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\right]^{\mathrm{T}}$, then the expression can be transformed into the following form:

$$\begin{aligned}\dot{L}_{\mathrm{p}} \leqslant& -\lambda_{\min}(\mathbf{Q}_i)e_i^{\mathrm{T}}e_i - \frac{1}{4}\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\mathbf{W}_{i,\mathrm{c}}^* + \kappa\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\sigma_{i,\mathrm{c}} + H_i^* \\ &+ \frac{1}{2}\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\overline{\mathbf{W}}_{i,\mathrm{c}} + \chi_i\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Delta\bar{\sigma}_{i,\mathrm{c}} \\ &+ \left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{W}_{i,\mathrm{c}}^* + \mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\nabla\sigma_{i,\mathrm{c}}d_k \\ &+ \overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\begin{pmatrix}\left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2 + \chi_i\boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\Delta\sigma_{i,\mathrm{c}}^{\mathrm{T}} \\ +\left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2\boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\left(\mathbf{W}_{i,\mathrm{c}}^* - \overline{\mathbf{W}}_{i,\mathrm{c}}\right)^{\mathrm{T}}\end{pmatrix}\overline{\mathbf{W}}_{i,\mathrm{c}} \\ &+ \overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\left(\Delta\bar{\sigma}_{i,\mathrm{c}} + \boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right) \\ &\times\left(\frac{T}{4}\exp(-\kappa T)\left(\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}} - 2\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\right)\Psi_i\right)\overline{\mathbf{W}}_{i,\mathrm{c}} \\ &+ b_{i,\varepsilon}\left(b_{i,fm} + \frac{1}{2}\lambda_{\min}\left(\mathbf{R}_i^{-1}\right)b_{i,gm}^2\nabla\sigma_{i,\mathrm{c}}^{\mathrm{T}}\right), \end{aligned} \tag{46}$$

$$\dot{L}_{\mathrm{p}} \leqslant -\mathbf{z}^{\mathrm{T}}\begin{bmatrix}\lambda_{\min}(\mathbf{Q}_i)\mathbf{I}_{n\times n} & 0 \\ 0 & \left(-k_{i,\mathrm{c}} + d_3\right)\mathbf{I}_{n\times n}\end{bmatrix}\mathbf{z} + C, \tag{47}$$

where

$$\begin{aligned}\left\|D_1\right\| =& \left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2 + \chi_i\boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\Delta\sigma_{i,\mathrm{c}}^{\mathrm{T}} \\ &+ \left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2\boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\left(\mathbf{W}_{i,\mathrm{c}}^* - \overline{\mathbf{W}}_{i,\mathrm{c}}\right)^{\mathrm{T}} \leqslant k_{i,\mathrm{c}}, \\ \left\|D_2\right\| =& \left(\Delta\bar{\sigma}_{i,\mathrm{c}} + \boldsymbol{\Gamma}_{i,\mathrm{c}}^{-1}\mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right) \\ &\times\frac{T}{4}\exp(-\kappa T)\left(\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}} - 2\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\right)\Psi_i \leqslant d_3, \end{aligned} \tag{48}$$

$$\begin{aligned}C =& -\frac{1}{4}\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\mathbf{W}_{i,\mathrm{c}}^* + \kappa\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\sigma_{i,\mathrm{c}} + H_i^* + \frac{1}{2}\mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\Psi_i\overline{\mathbf{W}}_{i,\mathrm{c}} \\ &+ b_{i,\varepsilon}\left(b_{i,fm} + \frac{1}{2}\lambda_{\min}\left(\mathbf{R}_i^{-1}\right)b_{i,gm}^2\nabla\sigma_{i,\mathrm{c}}^{\mathrm{T}}\right) \\ &+ \chi_i\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\Delta\bar{\sigma}_{i,\mathrm{c}} + \left\|\Delta\sigma_{i,\mathrm{c}}\right\|^2\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\mathbf{W}_{i,\mathrm{c}}^* + \mathbf{W}_{i,\mathrm{c}}^{*\mathrm{T}}\nabla\sigma_{i,\mathrm{c}}d_k. \end{aligned} \tag{49}$$

Assume that $\|C\| \leqslant c_M$, it can be deduced that $\dot{L}_{\mathrm{p}}(z) \leqslant 0$ if $\|z\| \geqslant \sqrt{c_M / \lambda_{\min}(\mathbf{Q}_i)}$, so the state errors and weight estimation errors are all bounded. Furthermore, all the signals in the closed-loop system are bounded. The proof has been concluded.

3.5 Algorithm structure

In this section, the detailed algorithm structure discussed in this article is provided to illustrate the iteration process of the compound controller applied to the hypersonic vehicle system. In Algorithm 1, the total scheme is divided into three steps. Initializing the constant matrixes and other hyperparameters required by the calculation process is the first step. Then for every epoch j, the value iteration training is implemented to calculate the approximate value function and

Algorithm 1 ALOTC algorithm

Step 1. Initialization

1.1 Initialize positive-define constant matrixes $\mathbf{Q}_i$, $\mathbf{R}_i$, $\boldsymbol{\Gamma}_i$, $i = h, V$, and other hyperparameters including threshold settings δ_h and δ_V.

1.2 Initialize NN weight vector $\widehat{\mathbf{W}}_{i,\mathrm{c}}$, learning rate η_i, and discount factor κ.

1.3 Set the initial NN weight error estimate to Inf and the total iterative training epoch N_{epoch}.

Step 2. Iteration

Repeat every epoch*j*:

2.1 For an initial admissible control policy $\hat{u}_{i,\mathrm{op}}$ in eq. (32) and compensation control $\hat{u}_{i,\mathrm{comp}}$ in eq. (19), the error state $e_i(t-T)$ can be obtained.

2.2 Calculate the approximate value function $\hat{J}\left(e_i(t-T)\right)$ and its derivatives $\nabla\hat{J}\left(e_i(t-T)\right)$ by the previous error state $e_i(t-T)$.

2.3 Update the desired flight angle in eq. (6) from reference velocity and height.

2.4 Update the control policy $\hat{u}_{i,\mathrm{op}}(t)$ by using the dynamic equation information $g(\mathbf{x})$ as follows:

$$\hat{u}_{i,\mathrm{op}}(t) = -\frac{1}{2}\mathbf{R}_i^{-1}g(\mathbf{x})^{\mathrm{T}}\left(\nabla\sigma_{\mathrm{c}}^{\mathrm{T}}(e_i)\widehat{\mathbf{W}}_{i,\mathrm{c}}\right).$$

2.5 The next error state $e_i(t)$ and $\hat{J}\left(e_i(t)\right)$ can be obtained from the updated desired state and control policy.

2.6 If the convergence threshold δ_i is satisfied,

$$\left\|\widehat{\mathbf{W}}_{i,\mathrm{c}}(t) - \widehat{\mathbf{W}}_{i,\mathrm{c}}(t-T)\right\| \leqslant \delta_i,$$

then the iteration process is completed and the optimal weight vector $\mathbf{W}_{i,\mathrm{c}}^*$ and control policy $u_{i,\mathrm{op}}$ are obtained, else turn to step 2.7.

2.7 Update the critic NN structure parameter $\widehat{\mathbf{W}}_{i,\mathrm{c}}$ by

$$\dot{\widehat{\mathbf{W}}}_{i,\mathrm{c}} = \left(\frac{\boldsymbol{\Gamma}_{i,\mathrm{c}}}{\left(1 + \Delta\sigma_{i,\mathrm{c}}^{\mathrm{T}}\Delta\sigma_{i,\mathrm{c}}\right)^2} + \mathbf{k}_{i,W}\overline{\mathbf{W}}_{i,\mathrm{c}}^{\mathrm{T}}\right)\Delta\sigma_{i,\mathrm{c}}\varepsilon_{i,\mathrm{c\text{-}B}},$$

and then go back to step 2.2.

End

Step 3. Parameter optimization

3.1 Calculate the response time of the system under the activation of the compound controller.

3.2 Build optimization problem with the response time as the target function and control parameters (k_{hp}, k_{h1}) as input variables.

3.3 Applying genetic algorithm to solve this optimization problem.

control policy. Then, check whether the convergence threshold is satisfied, or continue to update the network parameters and return to step 2.2. Finally, analyze the signal response time of the system from the initial state to the stable state, and establish an optimization problem to obtain the optimal control coefficients $k_{h\mathrm{p}}$, $k_{h\mathrm{I}}$ in eq. (7) for rapid convergence of the system.

4 Simulation

In this section, the effectiveness and robustness of the ALOTC scheme applied in the hypersonic longitudinal control system are verified. The physical parameters are provided in Table 1 and the required hyperparameters are also provided in Table 2. In addition, the initial settings of $k_{h\mathrm{p}}$ and $k_{h\mathrm{I}}$ are 5 and 3, respectively. The critic NN structure parameters are considered in Table 3 and activation functions are provided in eq. (50):

$$\begin{aligned}\sigma_V &= \left[e_V, e_V^2, \tanh(e_V), \sin(e_V)\right]^{\mathrm{T}},\\ \sigma_h &= \left[e_\gamma, e_\theta, e_q, e_\gamma^2, e_\theta^2, e_q^2, \tanh(e_\gamma), \tanh(e_\theta), \tanh(e_q)\right]^{\mathrm{T}}.\end{aligned} \tag{50}$$

The desired height and velocity signals for simulation are defined in eq. (51),

$$\begin{aligned}h_{\mathrm{d}} &= \begin{cases}34000, & 0\ \mathrm{s} \leqslant t < 10\ \mathrm{s},\\ -20t^2+800t+28000, & 10\ \mathrm{s} \leqslant t < 20\ \mathrm{s},\\ 36000, & 20\ \mathrm{s} \leqslant t < 30\ \mathrm{s},\end{cases}\\ V_{\mathrm{d}} &= \begin{cases}4750, & 0\ \mathrm{s} \leqslant t < 10,\\ -2t^2+80t+4150, & 10\ \mathrm{s} \leqslant t < 20\ \mathrm{s},\\ 4950, & 20\ \mathrm{s} \leqslant t < 30\ \mathrm{s}.\end{cases}\end{aligned} \tag{51}$$

According to the initial flight states provided in Table 4, the superiority of the ALOTC scheme is validated by four simulation cases. These include the verification of the control system without/with uncertainties, comparison with other approaches, and offline implementation in different flight conditions. Additionally, the tolerant upper bounds of height and velocity are 10 m and 5 m/s respectively, and the tolerant band is defined as 0.95. The control performance discussed in this article includes error accumulation (EA), mean square error (MSE), settling time (ST), and steady-state error (SSE).

Table 1 Physical parameters of the hypersonic vehicle

Parameters	Value	Meaning
m	136809 kg	Mass
S_{ref}	334.73 m^2	Reference area
I_{yy}	9.49×10^6 kg m^2	Moment of inertia
c	24.384 m	Mean aerodynamic chord
R_{e}	6371368 m	Radius of earth
μ	3.93×10^{14} m^3/s^2	Gravitational constant

Table 2 The settings of some hyperparameters in the simulation

Parameters	Value	
	for $i = V$	for $i = h$
δ_h	1×10^{-10}	
δ_V	1×10^{-10}	
τ	0.001	
N_{epoch}	3	
$\mathbf{Q}_i$	1/100	$\mathbf{I}_{3\times3}$
$\mathbf{R}_i$	2	$50\mathbf{I}_{3\times3}$
Γ_i	$0.5\mathbf{I}_{4\times4}$	$0.5\mathbf{I}_{9\times9}$
$\mathbf{k}_{i,W}$	$0.002\mathbf{I}_{4\times4}$	$0.002\mathbf{I}_{9\times9}$

Table 3 Critic NN structure parameters setting in the simulation

Parameters	Value	
	for $i = V$	for $i = h$
Number of layers	1	1
Number of neurons n_i	4	9
η_i	0.001	0.001
κ	0.9	0.9

Table 4 The initial flight state of the hypersonic vehicle

Parameter	Value
V	4450.29 m/s
h	33.528 km
γ	0 rad
θ	0 rad
q	0 rad
α	0 rad

4.1 Verification of hypersonic vehicle system without uncertainties

Based on the above initial flight conditions and hyperparameters, the optimized $k_{h\mathrm{p}}$ and $k_{h\mathrm{I}}$ are 1.9934 and 7.3284, respectively. The simulation results of trajectory tracking performance are shown in Figures 2–7. It can be obtained from Figures 2 and 3, the closed-loop system has great tracking accuracy for both height and velocity commands under the ALOTC scheme, and the velocity tracking settling time is 0.53 s and height tracking is 0.96 s in 95% tolerance band. The flight path angle, pitch angular rate, and angle of attack curves are shown in Figure 4, which can converge to the desired state in a short period of time. Figure 5 shows that

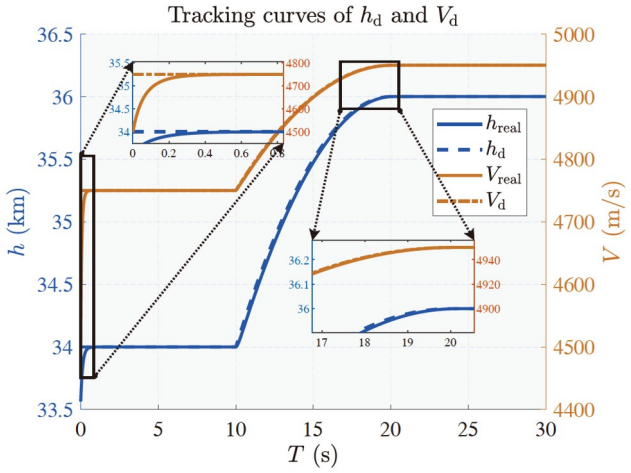

Figure 2 (Color online) Tracking trajectories of h_d and V_d.

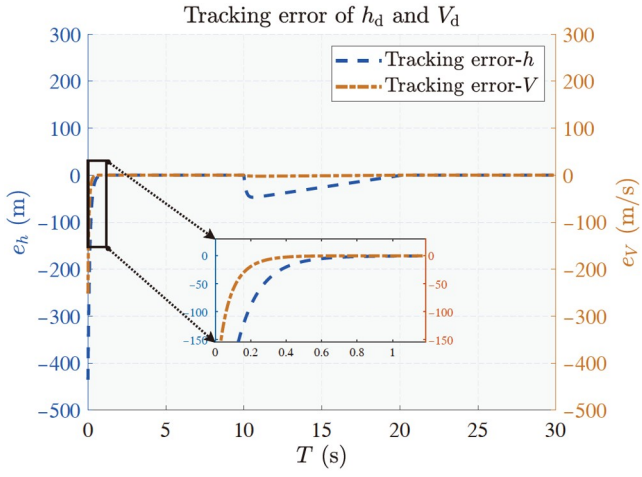

Figure 3 (Color online) Tracking errors of h_d and V_d.

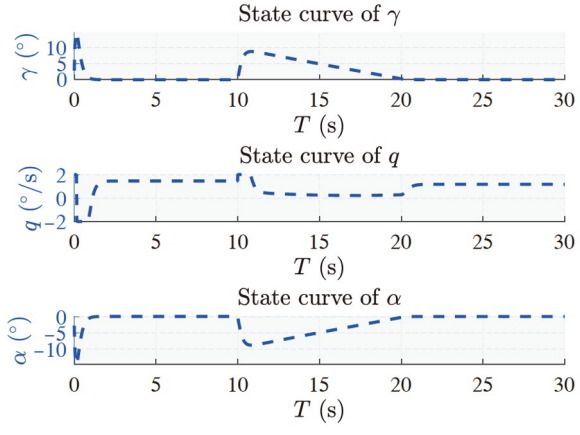

Figure 4 (Color online) Angle trajectories.

the actual control inputs are maintained to provide a feasible control performance.

The NN weight update curves can reach the stable values

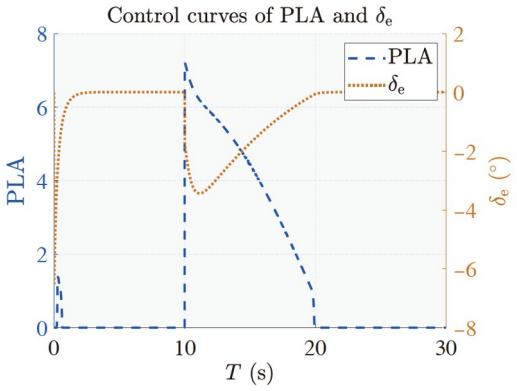

Figure 5 (Color online) Control inputs.

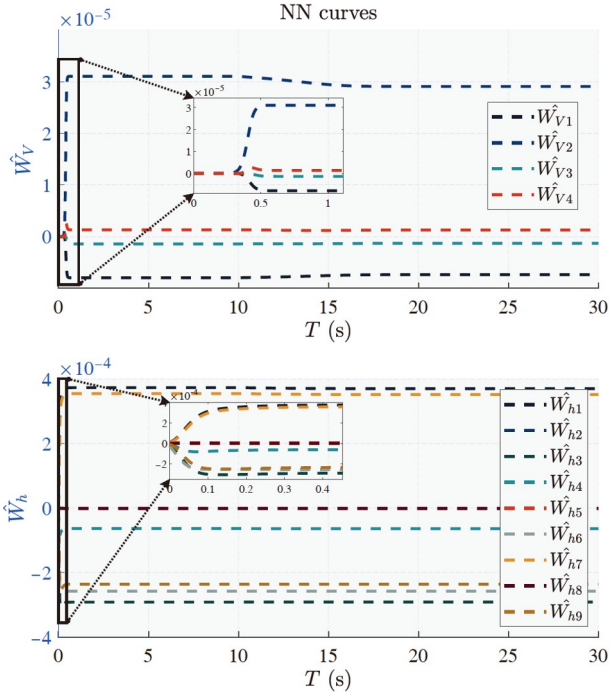

Figure 6 (Color online) NN convergence curves.

with a fast convergence speed in Figure 6. Additionally, the estimated value function in Figure 7 can also converge to the optimal value in a few iterations. In this simulation case, the obtained performance values are listed in Table 5. It is obvious that the rapid response ability can be guaranteed by short settling time and zero-approaching steady-state error.

4.2 Verification of hypersonic vehicle system with uncertainties

To further illustrate the robustness of the ALOTC scheme, different disturbances in eq. (53) are considered in the velocity and height subsystems respectively. In the actual flight task, the impact of uncertainties in angle states originates from the external disturbances and propagation of un-

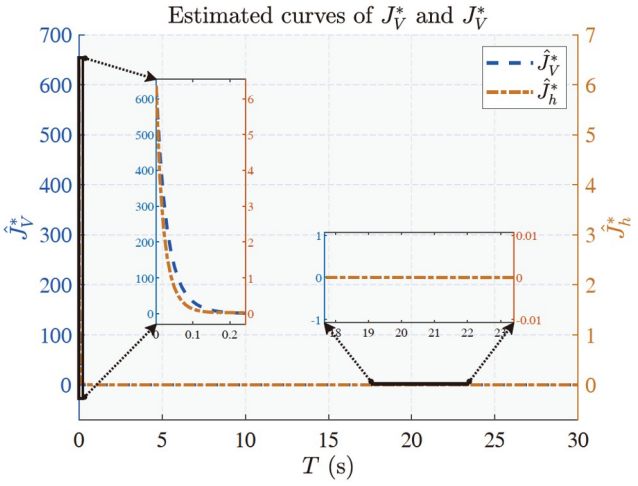

Figure 7 (Color online) Estimated value functions.

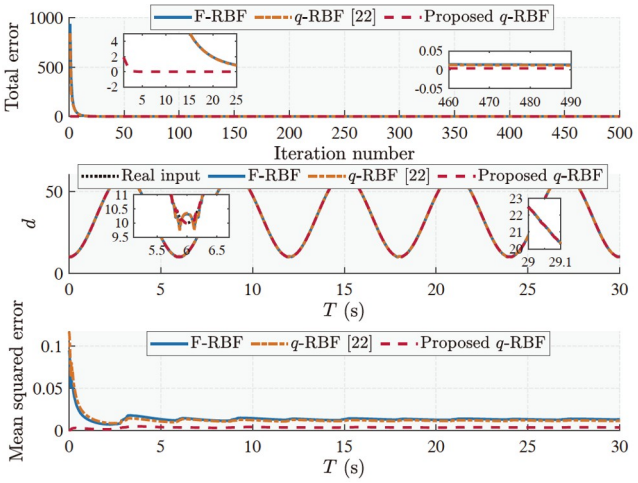

Figure 8 (Color online) Uncertainty estimation in case 1.

Table 5 Error information statistics of the proposed method applied in hypersonic vehicle

Tracking State	Error accumulation	Mean square error	Settling time	Steady-state error
h	6.74×10^3 m	4.55×10^7 m	0.62 s	0.5476 m
V	3.03×10^3 m/s	9.17×10^6 m/s	0.45 s	0.8583 m/s

certainties in velocity and height. In this article, the propagation of uncertainties is the major consideration, thus the uncertainties d_γ, d_θ, and d_q are considered as the first, second, and third-order derivatives of d_h, respectively.

The superiority of the proposed adaptive RBFNN disturbance observer is firstly discussed with two cases in eqs. (52) and (54), which the simulation results are shown in Figures 8 and 9.

$$d_V = 10 + 50\sin^2(\pi t / 6), \tag{52}$$

$$d_h = 20\sin^2(\pi t / 30), \tag{53}$$

$$d_i = a_1 r(t) + a_2 r(t-1) + a_3 r(t-2) + a_4[\cos(a_5 r(t)) + \exp(-|r(t)|)] + 0.1K(t), \tag{54}$$

where $K(t)$ is the Gaussian function with mean zero and variance one, a_1, a_2, a_3, a_4, a_5 are the constant parameters, $r(t)$ is the step function, and $i = h$, V.

From Figures 8 and 9, the effective disturbance rejection is obtained by the proposed q-RBFNN. Note that all three methods can achieve approximation convergence of disturbances to obtain satisfactory transient performance. Compared with the existent approaches, the q-RBFNN has a smaller cumulative error during the whole iteration period and superior performance in terms of steady-state error and mean squared error. Also, it has strong approximation ability for the random signals in Figure 9, resulting in better estimation precision. Introduce the discussed disturbances eqs.

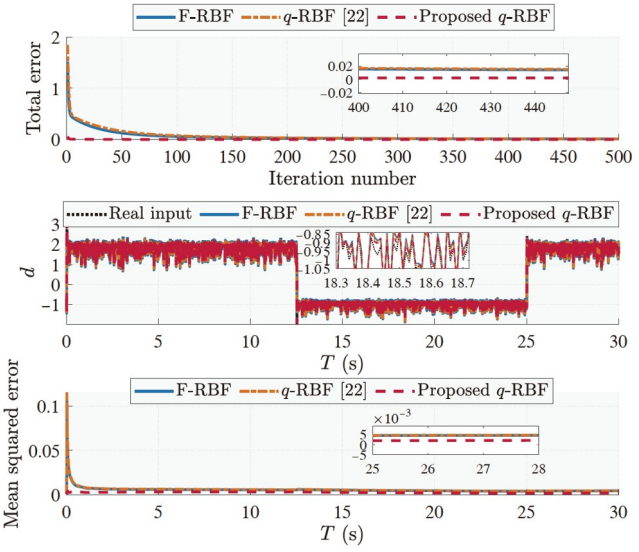

Figure 9 (Color online) Uncertainty estimation in case 2.

(52) and (53) into velocity and height subsystems, respectively. As illustrated in Figures 10–12, the great chatter occurs in the control system without the observer, leading to a significant steady-state error. At the same time, Figure 13 indicates that the control performance statistics for the disturbed system with an observer are all superior to the information without the observer. The control inputs are shown in Figure 14.

4.3 Comparison of different approaches with uncertainties

To demonstrate the advantages of the proposed ALOTC scheme, the existent adaptive backstepping approach in ref. [29] and the actor-critic method in ref. [9] are introduced as

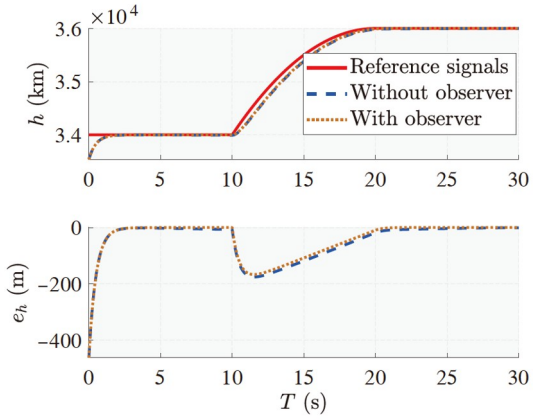

Figure 10 (Color online) Tracking trajectory and error of h_d.

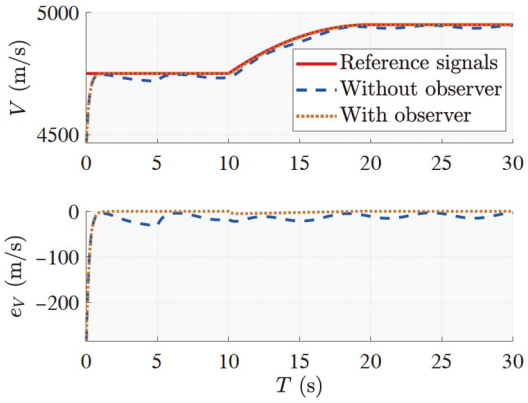

Figure 11 (Color online) Tracking trajectory and error of V_d.

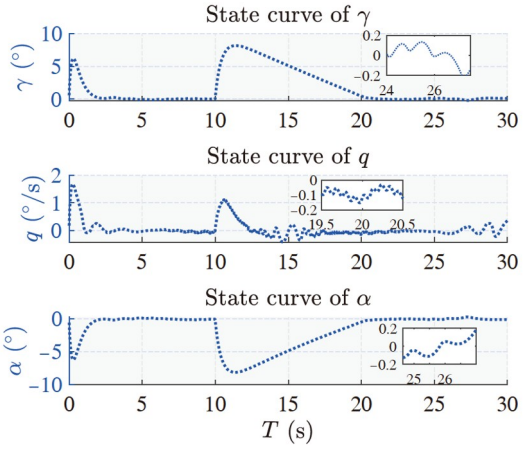

Figure 12 (Color online) Angle trajectories with the observer.

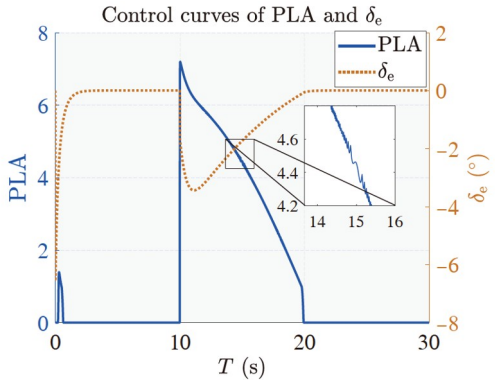

Figure 13 (Color online) Error information for the disturbed system without/with the observer.

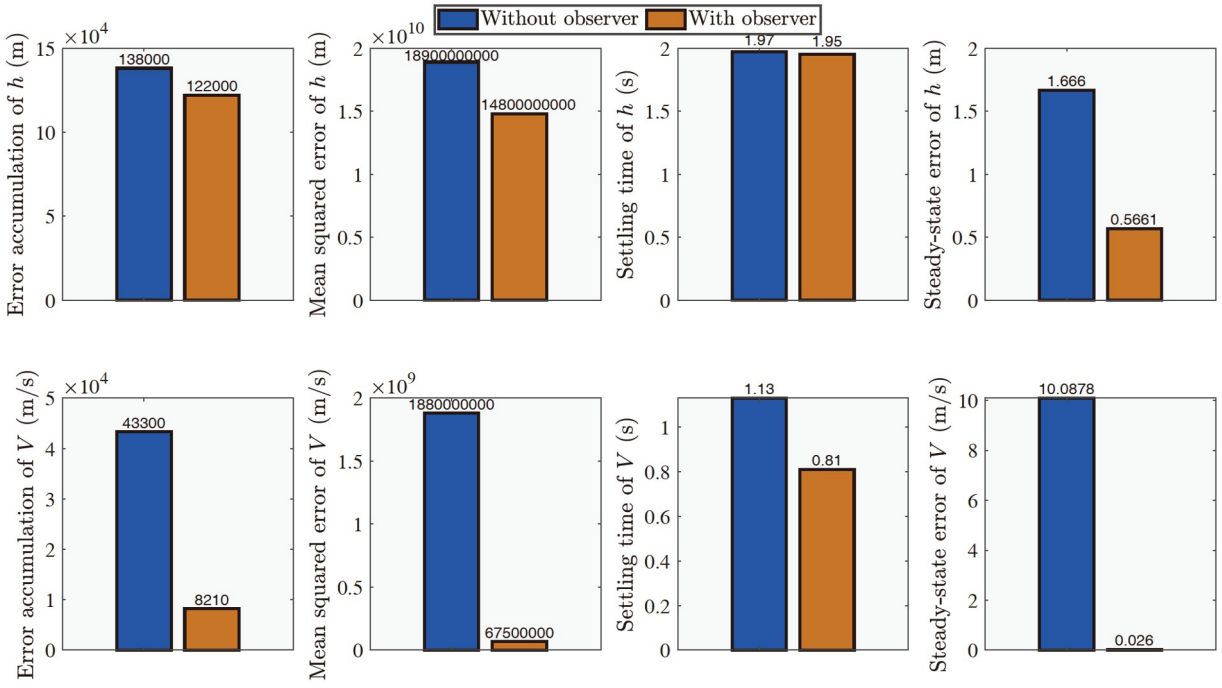

Figure 14 (Color online) Control inputs with the observer.

the comparative cases, and the results are depicted in Figures 15–17. It is obvious that the controllers can all achieve control purposes, but the adaptive backstepping and actor-critic methods cannot satisfy the requirements of convergence amplitude. In Figure 16, the actor-critic method has significant oscillations around the zero line. Figure 17 indicates good performance for the ALOTC scheme in MSE and ST for height and in MSE and SSE for velocity.

4.4 Offline implementation of the ALOTC control scheme

In this section, the offline implementation of the obtained ALOTC control scheme is considered in different initial flight states for hypersonic vehicles. The specific initial information is given as follows:

$$\mathbf{x} = [32.528\ \text{km},\ 4650\text{m/s},\ 0^\circ,\ 5^\circ,\ 0^\circ,\ 0^\circ]^{\mathrm{T}},$$

and the desired height and velocity signals for implementation simulation are extended in eq. (55):

$$h_{\mathrm{d}} = \begin{cases} 34000, & 0\ \text{s} \leqslant t < 10\ \text{s}, \\ -20t^2 + 800t + 28000, & 10\ \text{s} \leqslant t < 20\ \text{s}, \\ 36000, & 20\ \text{s} \leqslant t < 30\ \text{s}, \\ 35500, & 30\ \text{s} \leqslant t \leqslant 50\ \text{s}, \end{cases}$$

$$V_{\mathrm{d}} = \begin{cases} 4750, & 0\ \text{s} \leqslant t < 10\ \text{s}, \\ -2t^2 + 80t + 4150, & 10\ \text{s} \leqslant t < 20\ \text{s}, \\ 4950, & 20\ \text{s} \leqslant t < 30\ \text{s}, \\ 4750, & 30\ \text{s} \leqslant t \leqslant 50\ \text{s}. \end{cases} \tag{55}$$

The height and velocity tracking errors in Figure 18 clearly

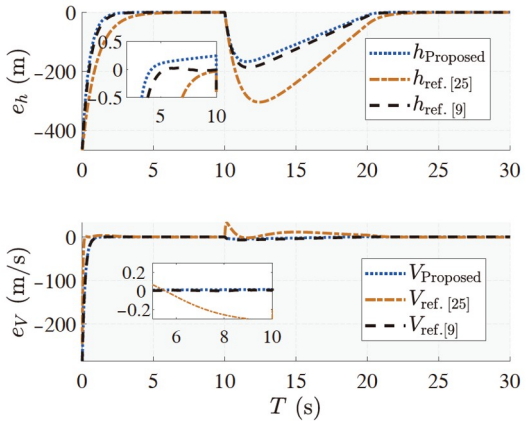

Figure 15 (Color online) Comparison of tracking errors of h_{d} and V_{d}.

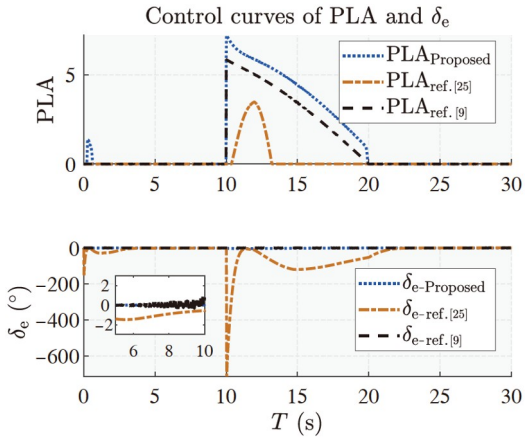

Figure 16 (Color online) Comparison of control inputs.

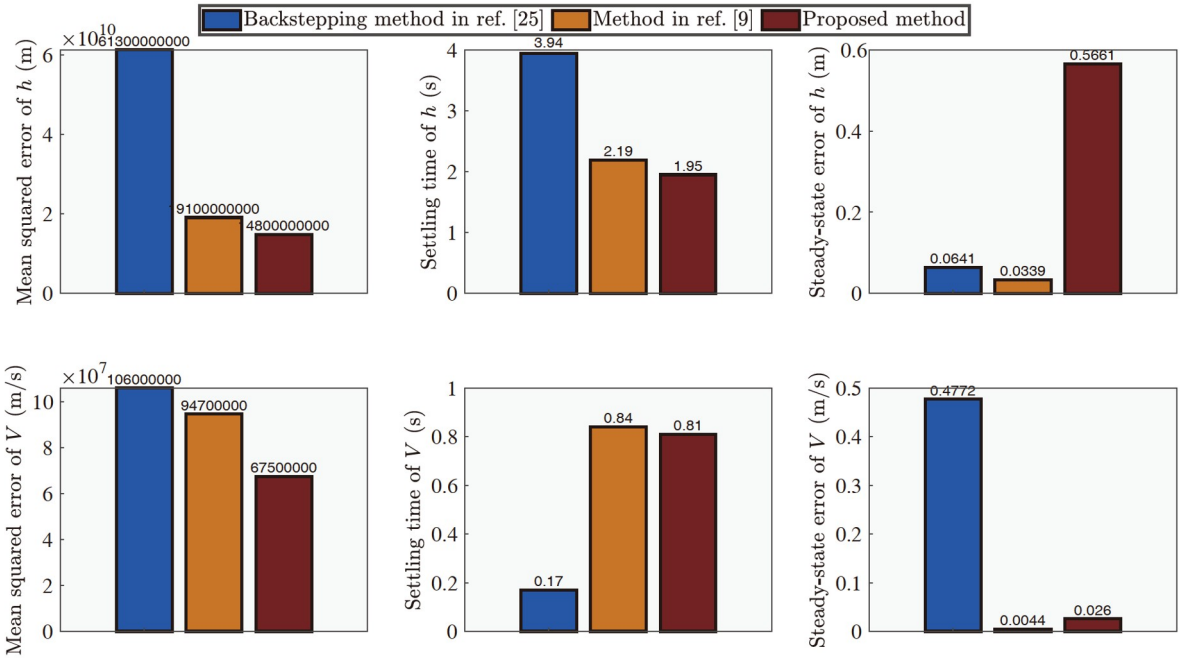

Figure 17 (Color online) Error information for the disturbed system with different methods.

illustrate that once the trained NN are applied to new initial states and reference commands, the desired tracking performance is still completed with possible significant accumulated errors. Figure 19 shows the effective control inputs. By statistical analysis, the ST is greater than 10 s for height and 0.26 s for velocity, resulting in the weak control performance. A fact can be obtained that the NN controllers trained with some specific trajectory samples can be applied offline to handle the similar-type trajectory control, which however cannot effectively address the maneuver trajectory control problem of gliding vehicles because the NN has not been exposed to similar samples during the training process, which is the partial work in our future research.

5 Conclusions

In this article, our purpose is to solve the height-velocity tracking control problem by proposing an adaptive learning approach for disturbed hypersonic vehicle systems. Based on the control-oriented dynamic equations, an integrated compound control model has been proposed to achieve the optimality and robustness of comprehensive control performance. Through the designed adaptive q-parameter update rule, the RBFNN was online trained to estimate the model uncertainties and external disturbances. Considering the system response time and approximation optimality, an ALOTC scheme was proposed to guarantee the convergence of weight parameters and the desired performance index. The simulation results indicated that the desired control performance and approximation accuracy are obtained for the hypersonic vehicle with the uncertainties and disturbances. In our future work, we will continue to focus on the application of intelligent algorithms in hardware-in-the-loop simulation.

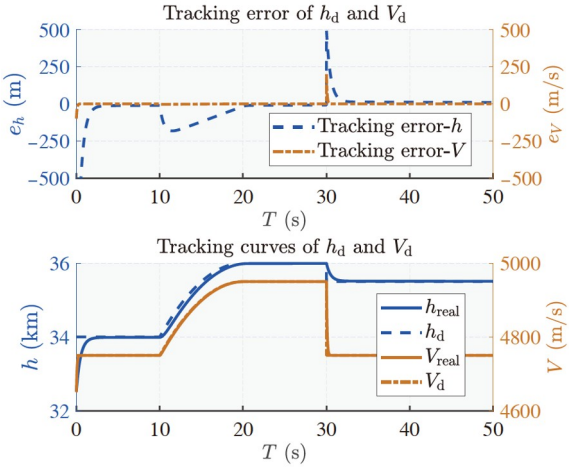

Figure 18 (Color online) Comparison of tracking errors of h_d and V_d.

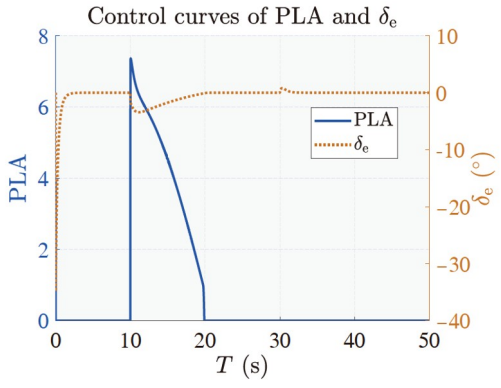

Figure 19 (Color online) Comparison of control inputs.

This work was supported by the Natural Science Foundation of Hunan Province (Grant No. 2021JJ10045), the National Natural Science Foundation of China (Grant No. 11972368), and the National Key R&D Program of China (Grant No. 2019YFA0405300).

1 Nair A P, Selvaganesan N, Lalithambika V R. Lyapunov based PD/PID in model reference adaptive control for satellite launch vehicle systems. Aerosp Sci Tech, 2016, 51: 70–77
2 An H, Wu Q, Wang G, et al. Adaptive compound control of air-breathing hypersonic vehicles. IEEE Trans Aerosp Electron Syst, 2020, 56: 4519–4532
3 Yin Z Y, Wang B, Xiong R T, et al. Attitude tracking control of hypersonic vehicle based on an improved prescribed performance dynamic surface control. Aeronaut J, 2024, 128: 875–895
4 Huang B, Li A, Xu B. Adaptive fault tolerant control for hypersonic vehicle with external disturbance. Int J Adv Robot Syst, 2017, 14: 172988141668713
5 Guo J, Wang G, Guo Z, et al. New adaptive sliding mode control for a generic hypersonic vehicle. Proc Inst Mech Eng Part G-J Aerosp Eng, 2018, 232: 1295–1303
6 Guo R, Ding Y, Yue X. Active adaptive continuous nonsingular terminal sliding mode controller for hypersonic vehicle. Aerosp Sci Tech, 2023, 137: 108279
7 An K, Guo Z, Huang W, et al. Leap trajectory tracking control based on sliding mode theory for hypersonic gliding vehicle. J Zhejiang Univ Sci A, 2022, 23: 188–207
8 Bellman R. Dynamic programming. Science, 1966, 153: 34–37
9 Wang N, Gao Y, Zhang X. Data-driven performance-prescribed reinforcement learning control of an unmanned surface vehicle. IEEE Trans Neural Netw Learn Syst, 2021, 32: 5456–5467
10 Xia R, Bu C, Yan X, et al. Finite-horizon optimal trajectory control of near space hypersonic vehicle with multi-constraints. Optim Control Appl Methods, 2024, 45: 302–320
11 Hu G, Guo J, Cieslak J, et al. Fault-tolerant control based on adaptive dynamic programming for reentry vehicles subjected to state-dependent actuator fault. Eng Appl Artif Intell, 2023, 123: 106450
12 Yang H, Hu Q, Dong H, et al. Optimized data-driven prescribed performance attitude control for actuator saturated spacecraft. IEEE ASME Trans Mechatron, 2023, 28: 1616–1626
13 Wang X, Li Y, Quan Z, et al. Optimal trajectory-tracking guidance for reusable launch vehicle based on adaptive dynamic programming. Eng Appl Artif Intell, 2023, 117: 105497
14 Lu J, Wei Q, Wang F Y. Parallel control for optimal tracking via adaptive dynamic programming. IEEE CAA J Autom Sin, 2020, 7: 1662–1674
15 Bao C, Wang P, Tang G. Data-driven based model-free adaptive optimal control method for hypersonic morphing vehicle. IEEE Trans Aerosp Electron Syst, 2022, 59: 3713–3725

16 Vrabie D, Lewis F. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. Neural Networks, 2009, 22: 237–246
17 Hu G, Guo J, Guo Z, et al. ADP-based intelligent tracking algorithm for reentry vehicles subjected to model and state uncertainties. IEEE Trans Ind Inf, 2023, 19: 6047–6055
18 He J, Qi R, Jiang B, et al. Adaptive output feedback fault-tolerant control design for hypersonic flight vehicles. J Franklin Inst, 2015, 352: 1811–1835
19 Sun J, Yi J, Pu Z. Augmented fixed-time observer-based continuous robust control for hypersonic vehicles with measurement noises. IET Control Theor Appl, 2019, 13: 422–433
20 Guo Y, Xu B. Finite-time deterministic learning command filtered control for hypersonic flight vehicle. IEEE Trans Aerosp Electron Syst, 2022, 58: 4214–4225
21 Zhao H W, Yang L. Global adaptive neural backstepping control of a flexible hypersonic vehicle with disturbance estimation. Aircr Eng Aerosp Tech, 2022, 94: 492–504
22 Wang F, Fan P, Fan Y, et al. Robust adaptive control of hypersonic vehicle considering inlet unstart. J Syst Eng Electron, 2022, 33: 188–196
23 Zhao H, Li R. Typical adaptive neural control for hypersonic vehicle based on higher-order filters. J Syst Eng Electron, 2020, 31: 1031–1040
24 Kac V, Cheung P. Quantum Calculus. New York: Springer, 2012
25 Khan S, Naseem I, Malik M A, et al. A fractional gradient descent-based RBF neural network. Circuits Syst Signal Process, 2018, 37: 5311–5332
26 Hussain S S, Usman M, Siddique T H M, et al. q-RBFNN: A quantum calculus-based RBF neural network, arXiv: 2106.01370
27 Xu H, Mirmirani M D, Ioannou P A. Adaptive sliding mode control design for a hypersonic flight vehicle. J Guid Control Dyn, 2004, 27: 829–838
28 Li Y, Qiang S, Zhuang X, et al. Robust and adaptive backstepping control for nonlinear systems using RBF neural networks. IEEE Trans Neural Netw, 2004, 15: 693–701
29 Liu X, Zhang Y, Wang S, et al. Backstepping attitude control for hypersonic gliding vehicle based on a robust dynamic inversion approach. Proc Inst Mech Eng Part I-J Syst Control Eng, 2014, 228: 543–552