The Aeronautical Journal (2024), 1–17
doi:10.1017/aer.2024.17

REGULAR PAPER

Reinforcement learning-based adaptive spiral-diving Manoeuver guidance method for reentry vehicles subject to unknown disturbances

T. Wu[1] and Z. Wang[1,2,3]

[1]Research Center for Unmanned System Strategy Development, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China, [2]Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China and [3]National Key Laboratory of Aerospace Flight Dynamics, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China
Corresponding author: Z. Wang; Email: nwpu@126.com

Received: 29 August 2023; **Revised:** 28 January 2024; **Accepted:** 8 February 2024

Keywords: Actor-critic design; spiral-diving manoeuver; adaptive guidance; reentry vehicles; anti-disturbance

Abstract

This paper proposed a reinforcement learning-based adaptive guidance method for a class of spiral-diving manoeuver guidance problems of reentry vehicles subject to unknown disturbances. First, the desired proportional navigation guidance law is designed for the vehicle based on the initial conditions, terminal constraints and the curve involute principle. Then, the first-order multivariable nonlinear guidance command tracking model considering unknown disturbances is established. And the controller design problem caused by the coupling of control variables is overcome by introducing the coordinate transformation technique. Moreover, the actor-critic networks and corresponding adaptive weight update laws are designed to cope with unknown disturbances. With the help of Lyapunov direct method, the stability of the system is proved. Subsequently, the range values of the guidance parameters are analysed. Finally, the validity as well as superiority of the proposed method are verified by numerical simulations.

Nomenclature

RBF	radial basis function
RL	reinforcement learning
VST	virtual sliding target

1.0 Introduction

Reentry vehicles are a class of unpowered aircraft capable of flying in near space, with special large lift-to-drag ratio aerodynamic profile, high manoeuverability and strong penetration ability. Different from traditional inertial vehicles, reentry vehicles rely on aerodynamic control and can achieve large lateral manoeuvering flight, which can significantly improve survivability [1, 2]. However, where there is a spear, there is a shield. Reconnaissance surveillance and interception technologies for reentry vehicles have also made great progress throughout the world [3, 4]. The interference and interception measures faced by the vehicles in the dive flight phase are more diverse than those in the midcourse and reentry phases, and the probability of being successfully intercepted will be greatly increased [5]. Therefore, it is a very interesting research topic to provide a manoeuver for reentry vehicles in the dive phase with high penetration ability and strong anti-disturbance ability [6, 7]. When the reentry vehicle does not

[†] Author's notes

have any prior information about the interceptor weapon, it is essential to choose a manoeuver mode that is difficult to predict and has high penetration probability. The typical ones are snake manoeuver, roller manoeuver and spiral manoeuver [8]. Among them, the spiral manoeuver has become a promising research direction in recent years because of the advantages of large manoeuver range, time-varying manoeuver frequency and unpredictable trajectory [9].

The main guidance techniques used for the dive phase are optimal guidance law, predictor-corrector guidance law, sliding mode variable structure guidance law and proportional navigation guidance law [10, 11]. An optimal guidance law design method based on block pulse function is proposed by Dou et al. [12]. It is able to optimise the landing angle, miss distance and control energy consumption simultaneously. Wang et al. [13] proposes an energy-based predictor-corrector guidance algorithm to design longitudinal and lateral guidance laws, respectively. Li and Qian [14] design a three-dimensional guidance law considering target manoeuver, impact angle constraint and input saturation by using integral sliding mode control and adaptive control. Dhananjay and Ghose [15] develop a proportional navigation guidance law incorporating a time-to-go estimation algorithm to strike stationary targets. Among many guidance methods, the proportional navigation guidance has been widely studied in the field of guidance because of its advantages of simplicity, high efficiency and small miss distance. Nevertheless, the proportional navigation guidance is not competent for flight missions that need to perform specific manoeuvering modes. For this reason, some studies have introduced the concept of virtual sliding target (VST) to extend the application scope of the proportional navigation guidance. Intuitively, it is the mapping of the real flight trajectory of the vehicle to the trajectory of the virtual sliding target. And the end point of the virtual sliding target overlaps with the real target point. As long as the vehicle flies in the direction pointing to the virtual sliding target, it will eventually hit the real target. In the Mozaffari et al. [16] and Raju and Ghose [17], specific manoeuvers of the vehicle are achieved by controlling the speed and direction of the virtual sliding target. In this case, the parameters of the virtual sliding target are selected empirically. By introducing a virtual sliding target, Hu et al. [18] designs a two-stage guidance method combining non-singular terminal sliding mode guidance law and proportional navigation guidance law, which can satisfy both impact angle and time constraints. For the stationary targets, an adaptive proportional guidance navigation law based on the virtual sliding target is proposed by He and Yan [19] to guide the vehicle to complete the spiral dive manoeuver. And the scope of application of this method is extended to low-speed moving targets in He et al. [20]. Despite the very large contributions made by the mentioned works, none of them consider the effect of unknown disturbances on the guidance effectiveness.

Fortunately, many nonlinear control methods are used to deal with the adverse effects of unknown disturbances on the controlled system. Some of these are inherently robust to unknown disturbances, while others combine various types of disturbance observers [21, 22]. Sliding mode control is frequently used to design controllers for vehicles because of its insensitivity to matched disturbances [23]. Shen et al. [24] proposes a continuous adaptive super-twisting sliding mode tracking control method, which combines a conventional super-twisting sliding mode controller with an adaptive gain technique to overcome bounded disturbances. By exploiting the constraint handling capability and enhanced anti-disturbance capability of model predictive control, Chai et al. [25] presents a robust model predictive attitude control algorithm. The nonlinear feedback law is designed, and the system constraints are tightened to ensure that robust constraints are satisfied for all allowed uncertainties. A resilient attitude control method for spacecraft is proposed by Cao and Xiao [26], which utilises a nonlinear disturbance observer to compensate for unknown disturbances. Xiang et al. [27] proposes an adaptive backstepping attitude control method for hypersonic vehicles with which a nonlinear disturbance observer is also used to estimate unknown disturbances. In Refs (28–30), a single group or two groups of neural networks are utilised to fit the unmodeled dynamics and external disturbances of hypersonic vehicles. In addition, reinforcement learning (RL) is also introduced into the design of the controller. Ouyang et al. [31] introduces actor-critic design into the tracking control problem of elastic joint robots to fit the system uncertainties. Shi et al. [32] proposes a robust adaptive safety control framework for hypersonic vehicles based

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

on reinforcement learning, in which the actor-critic networks are used to approximate the optimal controller. Wang et al. [33] develops a reinforcement learning-based adaptive tracking control method for a class of semi-Markovian non-Lipschitz uncertain systems, in which actor-critic networks are used to handle unmatched disturbances. The actor-critic networks in reinforcement learning not only inherit the good nonlinear processing ability of neural networks, but also introduce the error-related cost function. Therefore, they have better performance than neural networks in theory.

Combined with the previous discussion, this paper further studies the problem of spiral-diving manoeuver guidance for reentry vehicles considering unknown disturbances based on the results achieved in He and Yan [19] and He et al. [20]. The main contributions of this paper are as follows:

- Compared with He and Yan [19] and He et al. [20], this paper considers the adverse effects of unknown disturbances on the spiral-diving manoeuver of reentry vehicles. Specifically, the guidance command tracking control problem model considering unknown disturbances is established. This model is abstracted as a first-order coupled multivariable nonlinear system, which is more relevant to engineering practice.
- The coordinate transformation technique is employed to overcome the controller design challenge caused by the coupling of control variables. Combined with the recursive design technique, the first-order time derivative of the control variables is finally obtained, and the control variables can be obtained by integrating it.
- By designing the actor-critic networks and the corresponding adaptive weight update law, the unknown disturbances are compensated with high accuracy. Furthermore, by using the Lyapunov method, it is proved that the tracking errors are uniformly ultimately bounded. As a result, the assumption that the actual guidance commands are equivalent to the desired guidance commands made in the convergence analysis of the guidance parameters in He and Yan [19] and He et al. [20] is verified in this paper.

2.0 Problem formulation and preminimaries

2.1 Problem statement

The centroid dynamic model of unpowered reentry vehicle subjected to unknown disturbances is shown as follows:

$$\begin{cases} \dot{x} = V\cos\theta\cos\psi_v \\ \dot{y} = V\sin\theta \\ \dot{z} = -V\cos\theta\sin\psi_v \\ \dot{V} = -\dfrac{X}{m} - g\sin\theta \\ \dot{\theta} = \dfrac{Y\cos\gamma_v}{mV} - \dfrac{g}{V}\cos\theta + d_1(t) \\ \dot{\psi}_v = -\dfrac{Y\sin\gamma_v}{mV\cos\theta} + d_2(t) \end{cases} \tag{1}$$

where x, y, z are the positions of the vehicle in the inertial frame. V is the velocity. θ represents the angle between the velocity vector and the horizontal plane, i.e., the path angle. When the velocity vector is above the horizontal plane, the θ is positive. ψ_v represents the angle between the projection of the velocity vector in the horizontal plane and the x axis, i.e., the deflection angle, measured counterclockwise in the horizontal plane. α and γ_v are the angle-of-attack and back angle. m and g are mass and gravitational acceleration. $d_1(t)$ and $d_2(t)$ are the unknown disturbances to the vehicle. X and Y are the drag and lift forces, and the expression of Y is

$$Y = \left(C_y^0 + C_y^\alpha \alpha\right) qS_{ref} \tag{2}$$

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

where C_y^0 and C_y^α are the lift coefficients. S_{ref} is the reference area. q is the dynamic pressure, and its expression is

$$q = \frac{1}{2}\rho V^2 \tag{3}$$

where ρ is the atmospheric density.

Define $\xi = [\theta, \psi_v]^T$ as the state variables. Then, when the reentry vehicle performs a spiral-diving manoeuver, the desired guidance commands can be expressed as $\xi_d = [\theta_d, \psi_{vd}]^T$. Up to now, the problem of spiral-diving guidance for reentry vehicle subject to unknown disturbances can be essentially translated into the problem of tracking the desired guidance command. And this problem can be organized as follows:

$$\begin{cases} \dot{\xi} = f_1(\xi) + g_1(\xi, u) + d \\ \chi_o = \xi \end{cases} \tag{4}$$

where $u = \left[\alpha, \gamma_v\right]^T$ are the control variables. χ_o is the output vector. $d = [d_1, d_2]^T$. f_1 and g_1 are smooth nonlinear functions that can be expressed as

$$f_1(\xi) = \begin{bmatrix} -\frac{g}{V}\cos\theta \\ 0 \end{bmatrix} \tag{5}$$

$$g_1(\xi, u) = \begin{bmatrix} \frac{\left(C_y^0 + C_y^\alpha \alpha\right) qS_{ref}\cos\gamma_v}{mV} \\ -\frac{\left(C_y^0 + C_y^\alpha \alpha\right) qS_{ref}\sin\gamma_v}{mV\cos\theta} \end{bmatrix} \tag{6}$$

The tracking error can be expressed as

$$e_1 = \xi - \xi_d \tag{7}$$

Assumption 1. *The system represented by Equation 4 is controllable, which satisfies*

$$\left|\frac{\partial g_1(\xi, u)}{\partial u}\right| \neq 0 \tag{8}$$

Lemma 1. [34] For a Lyapunov function $L(t)$, if its initial value $L(0)$ is bounded and its time derivative satisfies

$$\dot{L}(t) \leq -\kappa L(t) + \delta \tag{9}$$

where $\kappa > 0$ and $\delta > 0$ are constants, then $L(t)$ is bounded.

Lemma 2. [35] For vectors $A \in \mathbb{R}^n$ and $B \in \mathbb{R}^n$, there always holds

$$2A^T B \leq \|A\|^2 + \|B\|^2 \tag{10}$$

$$\|AB\| \leq \|A\| \cdot \|B\| \tag{11}$$

Remark 1. As shown in Equation 1, the impact of unknown disturbances are considered in this paper when studying the spiral-diving guidance problem. This work is missing in He and Yan [19] and He et al. [20]. For this reason, this paper is complemented by the design of the desired guidance command tracking system as shown in Equation 4. Further, the objective of this paper is to design a reinforcement learning based adaptive controller for system 4 such that $\|e_1\| \leq \tilde{e}_1$ as $t \to \infty$, where $\tilde{e}_1$ is a sufficiently small positive constant.

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

2.2 Spiral trajectory parameters solving

The logarithmic spiral trajectory can be defined as

$$r_s = r(\vartheta) = r_0 e^{\vartheta \cot \Lambda} \tag{12}$$

where r_0 is the initial polar diameter. ϑ is the polar angle. Λ is the angle between the component of the vehicle velocity vector in yaw plane and the polar diameter. It can be found from Equation 12 that once the values of r_0 and Λ are acquired, the shape of the spiral trajectory can be determined uniquely.

Figure 3 in He et al. [20] shows the geometric representation of the spiral trajectory in the yaw plane. Where M_0 and M_s are the initial position and the current desired position of the vehicle. T is the target position. px_pz_p denotes the polar frame, the pole p coincides with the rotation centre of the spiral trajectory, the polar axis pz_p points to the M_0, and the polar axis px_p is perpendicular to pz_p. r_{s0}, r_s and r_{s1} are the polar diameters at M_0, M_s and T, and the corresponding polar angles and deflection angles are ϑ_0, ϑ, ϑ_1 and ψ_{vs0}, ψ_{vs}, ψ_{vs1}, respectively. η is the rotation angle of the polar frame with respect to the inertial frame. If let (x_0, z_0) and (x_1, z_1) represent the coordinates at M_0 and T, respectively, then the initial condition set H_0 and terminal constraint set H_1 of the vehicle can be defined as

$$H_0 = \{(x_0, z_0), \vartheta_0, \psi_{vs0}\}, \quad H_1 = \{(x_1, z_1), \psi_{vs1}\}$$

It is worth noting that the definition of deflection angle direction in this paper is contrary to that in He et al. [20]. Therefore, in order to facilitate understanding, it is necessary to deduce new expressions related to the determination of spiral trajectory parameters. Referring to Fig. 3 in He et al. [20], the main geometrical relation can be expressed as follows:

$$\frac{\pi}{2} + \psi_{vs} = \vartheta + \eta + \Lambda \tag{13}$$

Substituting the ψ_{vs0} and ψ_{vs1} into Equation 13, we get

$$\eta = \frac{\pi}{2} + \psi_{vs0} - \Lambda \tag{14}$$

$$\vartheta_1 = \psi_{vs1} - \psi_{vs0} \tag{15}$$

The coordinates of the p can be solved by

$$\begin{bmatrix} x_p \\ z_p \end{bmatrix} = \frac{1}{K_1 - K_0} \begin{bmatrix} K_1 x_0 - K_0 x_1 + K_0 K_1 (z_1 - z_0) \\ K_1 z_1 - K_0 z_0 - (x_1 - x_0) \end{bmatrix} \tag{16}$$

where K_0 and K_1 are slopes of the rays pM_0 and pT, respectively, and their expressions are:

$$K_i = \tan\left(\frac{\pi}{2} + \psi_{vsi} - \Lambda\right), \quad i = 0, 1 \tag{17}$$

Refer to He et al. [20] for the calculation of x_p and z_p when K_0 and K_1 do not exist. Next, the lengths of the polar diameters r_{s0} and r_{s1} can be calculated:

$$\|r_{s0}\| = \left| \frac{z_0 - z_p}{\sin(\psi_{vs0} - \Lambda)} \right| \tag{18}$$

$$\|r_{s1}\| = \left| \frac{z_1 - z_p}{\sin(\psi_{vs1} - \Lambda)} \right| \tag{19}$$

Dividing Equation 19 by Equation 18 and combining Equations 12 and 17 yields

$$In \left| \frac{\cos(\psi_{vs0} - \Lambda - \mu)}{\cos(\psi_{vs1} - \Lambda - \mu)} \right| = \vartheta_1 \cot \Lambda \tag{20}$$

where $\mu = \arctan\left[(x_0 - x_1) / (z_0 - z_1)\right]$.

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

Once the values in the sets H_0 and H_1 are given, Λ can be acquired by solving Equation 20. By substituting Λ into Equations 16, 18, 14 and 15, the polar coordinates (x_p, z_p), initial polar diameter r_0, the rotation angle η of the polar frame with respect to the inertial frame and the terminal polar angle ϑ_1 can be calculated, respectively.

Remark 2. This subsection presents the procedure for calculating the spiral trajectory parameters in the yaw plane. Without loss of generality, the spiral trajectory in three-dimensional space can be obtained by stretching the yaw plane spiral trajectory along the vertical direction.

2.3 Neural networks in reinforcement learning

Neural networks are an important part of reinforcement learning and are powerful in coping with nonlinearities. With the help of neural networks, a nonlinear function f can be expressed as

$$f = W^T \Phi\left(\bar{Z}\right) + \varepsilon\left(\bar{Z}\right) \tag{21}$$

where $W \in \mathbb{R}^l$ is the weight vector, l is the number of nodes in the hidden layer. $\bar{Z} = [\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_m]^T \in \mathbb{R}^m$ are the inputs of neural networks with dimension m. $\Phi\left(\bar{Z}\right) = [\varphi_{b1}, \varphi_{b2}, \ldots \varphi_{bl}]^T \in \mathbb{R}^l$ are the basis functions. $\varepsilon\left(\bar{Z}\right)$ is the function reconstruction error. The optimal approximation can be obtained by properly selecting the number of network nodes.

The radial basis functions (RBF) neural networks are selected as the basic network frame in this paper, and its basis functions can be described as

$$\varphi_j(Z) = \exp\left(-\frac{\|Z - \zeta_j\|^2}{\varsigma_j^2}\right) \tag{22}$$

where $\zeta_j = \left[\zeta_{j1}, \zeta_{j2}, \ldots, \zeta_{jm}\right]^T$ is the centre vector of the j-th node in the hidden layer. ς_j is the width value.

Lemma 3. [31] The basis function $\Phi\left(\bar{Z}\right)$ of neural networks is bounded, which satisfies $\left\|\Phi\left(\bar{Z}\right)\right\| \leq \Phi_M$ and $\left\|\dot{\Phi}\left(\bar{Z}\right)\right\| \leq \Phi_{dM}$, where Φ_M and Φ_{dM} are positive constants.

Lemma 4. [31] If the ideal weight W^* is obtained, then there exists $\left|\varepsilon\left(\bar{Z}\right)\right| \leq \varepsilon_m$ and $\left|\dot{\varepsilon}\left(\bar{Z}\right)\right| \leq \varepsilon_{dm}$, where ε_m and ε_{dm} are positive constants.

3.0 Main results

This section fully presents the reinforcement learning spiral-diving manoeuver guidance method proposed in this paper. The concept of virtual sliding target is employed to design the desired proportional navigation guidance law. With the help of coordinate transformation technique, the desired guidance command tracking controller design challenge arising from the coupling of control variables is overcome. And the actor-critic networks and the corresponding adaptive weight update law are designed to approximate the unknown disturbances. After proving that the tracking errors are uniformly ultimately bounded by using the Lyapunov method, the range of guidance parameters is derived. The diagram of proposed reinforcement learning spiral-diving manoeuver guidance framework for reentry vehicle is shown in Fig. 1.

3.1 Desired proportional navigation guidance law

Figure 4 in He et al. [20] displays the motion of the vehicle and the virtual sliding target in the yaw plane. Where M is the projection of the current position of the vehicle in the yaw plane, and M_s is the closest point of M to the spiral trajectory.

Assumption 2. *The polar angle at M is the same as the polar angle at M_s. Moreover, in order to keep the shape of the spiral trajectory invariant, the pole p is assumed to have the same dynamic properties as the target.*

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

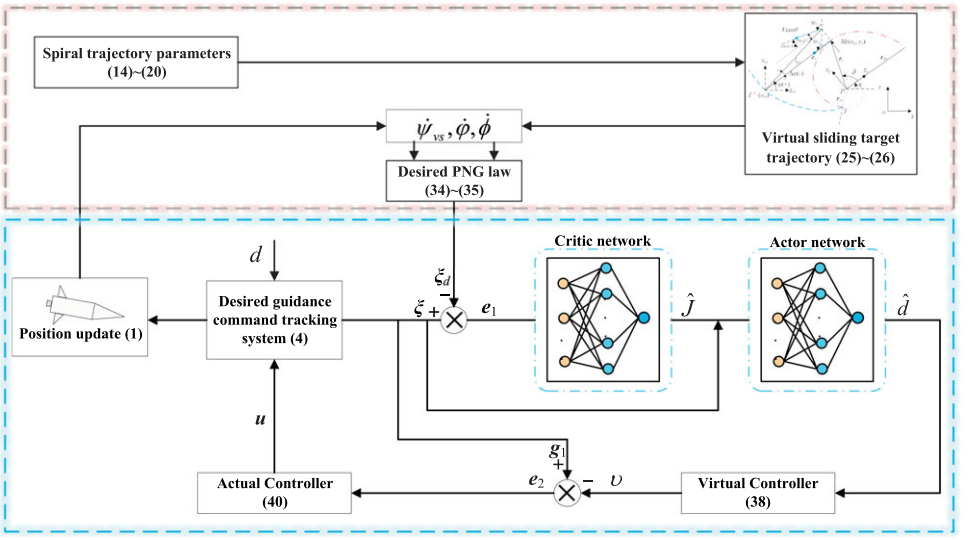

Figure 1. *Reinforcement learning adaptive spiral-diving manoeuver guidance framework.*

Under Assumption 2, the polar angle ϑ corresponding to the M can be obtained by solving the following equation:

$$\left(x - x_p\right)\sin\left(\vartheta + \eta + \Lambda\right) + \left(z - z_p\right)\cos\left(\vartheta + \eta + \Lambda\right) = ae^{\vartheta\cot\Lambda}\cos\Lambda \tag{23}$$

The time derivative of Equation 23 can be arranged to obtain the time derivative of ϑ:

$$\dot{\vartheta} = \frac{V\cos\theta\sin\left(\vartheta + \eta + \Lambda - \psi_v\right) - V_t\sin\left(\vartheta + \eta + \Lambda - \psi_{vt}\right)}{\left[\|r_s\|\cos^2\Lambda/\sin\Lambda - \left(x - x_p\right)\cos\left(\vartheta + \eta + \Lambda\right) + \left(z - z_p\right)\sin\left(\vartheta + \eta + \Lambda\right)\right]} \tag{24}$$

where V_t and ψ_{vt} are the size and direction angle of the target velocity, respectively. Note that ψ_{vt} is meaningless when the target is stationary, i.e., $V_t = 0$.

The trajectory of the virtual sliding target T' is designed based on the curve involute principle and is denoted as

$$r_{vt} = \begin{bmatrix} x_p \\ z_p \end{bmatrix} + r_s \begin{bmatrix} \sin\left(\vartheta + \eta\right) \\ \cos\left(\vartheta + \eta\right) \end{bmatrix} + l_{go} \begin{bmatrix} \sin\left(\vartheta + \eta + \Lambda\right) \\ \cos\left(\vartheta + \eta + \Lambda\right) \end{bmatrix} \tag{25}$$

where $r_{vt} = [x_{vt}, z_{vt}]^T$ represents the coordinate vector of the T'. l_{go} is the remaining length of the spiral trajectory and its value can be obtained by integrating Equation 12. The time derivative of Equation 25 yields

$$V_{vt} = \begin{bmatrix} \dot{x}_{vt} \\ \dot{z}_{vt} \end{bmatrix} = V_t \begin{bmatrix} \cos\psi_{vt} \\ -\sin\psi_{vt} \end{bmatrix} + l_{go}\dot{\vartheta} \begin{bmatrix} \cos\left(\vartheta + \eta + \Lambda\right) \\ -\sin\left(\vartheta + \eta + \Lambda\right) \end{bmatrix} \tag{26}$$

The virtual line-of-sight deflection angle from the current position of T' pointing to M is defined as

$$\varphi = \arctan\frac{x - x_{vt}}{z - z_{vt}} \tag{27}$$

Taking the time derivative of Equation 27, and combining with Equation 1 and Equation 26, the following equation can be obtained:

$$\dot{\varphi} = \frac{V\cos\theta}{s}\sin\left(\delta\psi_v\right) + \dot{\vartheta}\cos^2\left(\Delta\varphi\right) - \frac{V_t}{s}\cos\left(\psi_{vt} - \varphi\right) \tag{28}$$

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

where

$$\delta\psi_v = \frac{\pi}{2} + \varphi - \psi_v \tag{29}$$

$$\Delta\varphi = -\frac{\pi}{2} + \psi_{vs} - \varphi \tag{30}$$

$s = \sqrt{(x - x_{vt})^2 + (z - z_{vt})^2}$ is the remaining flight distance.

Similarly, the virtual line-of-sight path angle from the current position of T' pointing to M is defined as

$$\phi = \arctan\left(\frac{y}{s}\right) \tag{31}$$

Combining Equations 1 and 26, the time derivative of Equation 31 can be derived:

$$\dot{\phi} = \frac{V}{r}\left(\cos\phi\sin\theta + \sin\phi\cos\theta\cos\left(\delta\psi_v\right)\right) + \frac{l_{go}}{r}\dot{\vartheta}\sin\phi\sin\left(\Delta\varphi\right) + \frac{V_t}{r}\sin\phi\sin\left(\varphi - \psi_{vt}\right) \tag{32}$$

where $r = \sqrt{s^2 + y^2}$ is the distance of the vehicle from the virtual target.

Furthermore, taking the time deriving of Equation 13 yields

$$\dot{\psi}_{vs} = \dot{\vartheta} \tag{33}$$

Based on Equations 28, 32 and 33, the desired proportional navigation guiding law of the vehicle with respect to the virtual sliding target as shown in Equations 34 and 35 can be designed:

$$\dot{\psi}_{vd} = -\lambda_1\dot{\varphi} + \left(1 + \lambda_1\right)\dot{\psi}_{vs} \tag{34}$$

$$\dot{\theta}_d = -\lambda_2\dot{\phi} \tag{35}$$

where λ_1 and λ_2 are user-defined guidance parameters, and their value ranges will be determined later.

3.2 Reinforcement learning adaptive controller

For the first-order multivariate tightly coupled system 4 considering the effects of unknown disturbances, treat $g_1\left(\xi, u\right)$ as the virtual control variable and define

$$e_2 = g_1\left(\xi, u\right) - \upsilon\left(\xi, u\right) \tag{36}$$

where υ is the virtual control law.

Taking the time derivative of Equation 7 and combining it with Equation 36 yields

$$\dot{e}_1 = f_1 + e_2 + \upsilon + d - \dot{\xi}_d \tag{37}$$

so the virtual controller can be designed as

$$\upsilon = -k_1 e_1 - f_1 - \hat{d} + \dot{\xi}_d \tag{38}$$

where $k_1 > 0$ is a user-defined control gain. $\hat{d}$ is the estimation of d.

Taking the time derivative of Equation 36 leads to

$$\dot{e}_2 = \frac{\partial g_1}{\partial \xi}\dot{\xi} + \frac{\partial g_1}{\partial u}\dot{u} - \dot{\upsilon} \tag{39}$$

so the time derivative of the controller u can be designed as

$$\dot{u} = \left(\frac{\partial g_1}{\partial u}\right)^{-1}\left(-k_2 e_2 - e_1 - \frac{\partial g_1}{\partial \xi}\left(f_1 + g_1 + \hat{d}\right) + \dot{\upsilon}\right) \tag{40}$$

where $k_2 > 0$ is a user-defined control gain. By integrating Equation 40, u can be obtained.

From Equations 38 and 40, it is clear that how to obtain the estimate of d is the premise of designing the controller u. Ingeniously, the actor-critic networks in reinforcement learning provide a superior alternative to deal with the problem.

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

In the framework of the actor network, d can theoretically be expressed by

$$d = W_a^{*T}\Phi_a(\xi) + \varepsilon_a \tag{41}$$

where $\Phi_a \in \mathbb{R}^{l_a}$ is the basis function of dimension l_a, which satisfies $\|\Phi_a\| \le \Phi_{aM}$. $\varepsilon_a \in \mathbb{R}^2$ is the actor reconstruction error and satisfies $\|\varepsilon_a\| \le \varepsilon_{am}$. $W_a^* \in \mathbb{R}^{l_a \times 2}$ is the real actor network weight. The reality is that only the estimation of d can be obtained:

$$\hat{d} = \hat{W}_a^T\Phi_a(\xi) \tag{42}$$

where $\hat{W}_a \in \mathbb{R}^{l_a \times 2}$ is the estimated weight of the actor network.

In the framework of the critic network, the integral penalty function can be designed as

$$J(t) = \int_\tau^\infty \mathcal{L}(t)\,dt \tag{43}$$

where $\mathcal{L}(t) = e_1^T Q e_1$, $Q \in \mathbb{R}^{2\times 2}$ is a positive definite matrix. J can theoretically be expressed by

$$J = W_c^{*T}\Phi_c(e_1) + \varepsilon_c \tag{44}$$

where $\Phi_c \in \mathbb{R}^{l_c}$ is the basis function of dimension l_c, which satisfies $\|\dot{\Phi}_c\| \le \Phi_{cdM}$. ε_c is the critic reconstruction error and satisfies $|\dot{\varepsilon}_c| \le \varepsilon_{cdm}$. $W_c^* \in \mathbb{R}^{l_c}$ is the real critic network weight. The reality is that only the estimation of J can be obtained:

$$\hat{J} = \hat{W}_c^T\Phi_c(e_1) \tag{45}$$

where $\hat{W}_c \in \mathbb{R}^{l_c}$ is the estimated weight of the critic network.

Define the weight error of critic network as $\tilde{W}_c = \hat{W}_c - W_c^*$. In addition, define the critic error as

$$e_c = \mathcal{L} + \dot{\hat{J}} = \mathcal{L} + \hat{W}_c^T\dot{\Phi}_c \tag{46}$$

and the critic error function can be designed as

$$E_c = \frac{1}{2}e_c^2 \tag{47}$$

According to the gradient descent criterion, the adaptive update law of $\hat{W}_c$ can be deduced as follows:

$$\dot{\hat{W}}_c = -\lambda_c\left(\mathcal{L} + \hat{W}_c^T\dot{\Phi}_c\right)\dot{\Phi}_c - \lambda_c\hbar_c\hat{W}_c \tag{48}$$

where $\lambda_c > 0$ and $\hbar_c > 0$ are the user-defined learning rates of the critic network.

Define the weight error of actor network as $\tilde{W}_a = \hat{W}_a - W_a^*$. And define the approximation error H_a as

$$H_a = \hat{W}_a^T\Phi_a - W_a^{*T}\Phi_a = \tilde{W}_a^T\Phi_a \tag{49}$$

Then, the actor error can be defined as

$$e_a = H_a + \Omega_a\hat{J} \tag{50}$$

where $\Omega_a \in \mathbb{R}^{2\times 1}$ is the user-defined gain matrix satisfying $\|\Omega_a\| \le \Omega_{aM}$, and Ω_{aM} is a positive constant.

Furthermore, the actor error function can be designed as

$$E_a = \frac{1}{2}e_a^T e_a \tag{51}$$

According to the gradient descent criterion, the adaptive update law of $\hat{W}_a$ can be deduced as follows:

$$\begin{aligned}\dot{\hat{W}}_a &= -\lambda_a\Phi_a e_a^T - \lambda_a\hbar_a\hat{W}_a \\ &= -\lambda_a\Phi_a\left(\Phi_a^T\hat{W}_a + \hat{J}\Omega_a^T\right) - \lambda_a\hbar_a\hat{W}_a\end{aligned} \tag{52}$$

where $a > 0$ and $\hbar_a > 0$ are the user-defined learning rates of the actor network.

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

3.3 Stability and convergence analysis

Theorem 1. *Consider the Assumptions 1–2 and Lemmas 1–4, if the control law 40 is designed for system 4, the actor-critic networks 42 and 45 and the corresponding adaptive weight update laws 48 and 52 are designed to cope with d, and the Lyapunov candidate function as shown in Equation 53 is constructed, then the tracking error e_1 is uniformly ultimately bounded stable. As well, the weight errors $\hat{W}_a$ and $\hat{W}_c$ of the actor-critic networks are uniformly ultimately bounded.*

Proof . Construct the Lyapunov candidate function as follows:

$$L = L_1 + L_2 + L_3 \tag{53}$$

where

$$L_1 = \frac{1}{2} e_1^T e_1 + \frac{1}{2} e_2^T e_2 \tag{54}$$

$$L_2 = \frac{1}{2} Tr\left(\hat{W}_a^T \lambda_a^{-1} \hat{W}_a\right) \tag{55}$$

$$L_3 = \frac{1}{2} Tr\left(\hat{W}_c^T \lambda_c^{-1} \hat{W}_c\right) \tag{56}$$

By taking time derivative of Equation 54 and combining Equations 37–42, it can be deduced that

$$\begin{aligned}
\dot{L}_1 &= e_1^T \dot{e}_1 + e_2^T \dot{e}_2 \\
&= -k_1 e_1^T e_1 + e_1^T e_2 - e_1^T \hat{W}_a^T \Phi_a + e_1^T \varepsilon_a - k_2 e_2^T e_2 - e_2^T e_1 - e_2^T \frac{\partial g}{\partial \xi} \hat{W}_a^T \Phi_a + e_2^T \frac{\partial g}{\partial \xi} \varepsilon_a \\
&\le -k_1 e_1^T e + \frac{1}{2} e_1^T e_1 + \frac{1}{2} \Phi_a^T \hat{W}_a \hat{W}_a^T \Phi_a + \frac{1}{2} e_1^T e_1 + \frac{1}{2} \varepsilon_a^T \varepsilon_a \\
&\quad - k_2 e_2^T e_2 + \frac{1}{2} e_2^T \frac{\partial g}{\partial \xi} \left(\frac{\partial g}{\partial \xi}\right)^T e_2 + \frac{1}{2} \Phi_a^T \hat{W}_a \hat{W}_a^T \Phi_a + \frac{1}{2} e_2^T \frac{\partial g}{\partial \xi} \left(\frac{\partial g}{\partial \xi}\right)^T e_2 + \frac{1}{2} \varepsilon_a^T \varepsilon_a \\
&\le -\left(k_1 - 1\right) e_1^T e_1 - \left(k_2 - Tr\left(\frac{\partial g}{\partial \xi} \left(\frac{\partial g}{\partial \xi}\right)^T\right)\right) e_2^T e_2 + \Phi_{aM}^2 Tr\left(\hat{W}_a^T \hat{W}_a\right) + \varepsilon_{am}^2
\end{aligned} \tag{57}$$

By taking time derivative of Equation 55 and combining Equation 52, it can be deduced that

$$\begin{aligned}
\dot{L}_2 &= Tr\left(\hat{W}_a^T \lambda_a^{-1} \dot{\hat{W}}_a\right) \\
&= Tr\left(\tilde{W}_a^T \lambda_a^{-1} \left[-\lambda_a \Phi_a \left(\Phi_a^T \hat{W}_a + \hat{J} \Omega_a^T\right) - \lambda_a \hbar_a \hat{W}_a\right]\right) \\
&\le -Tr\left(\tilde{W}_a^T \Phi_a \Phi_a^T \tilde{W}_a\right) + \frac{1}{2} Tr\left(\tilde{W}_a^T \Phi_a \Phi_a^T \tilde{W}_a\right) + \frac{1}{2} Tr\left(W_a^T \Phi_a \Phi_a^T W_a\right) \\
&\quad + \frac{1}{2} Tr\left(\tilde{W}_a^T \Phi_a \Phi_a^T \tilde{W}_a\right) + \frac{1}{2} Tr\left(\tilde{W}_c^T \Phi_c \Omega_a^T \Omega_a \Phi_c^T \tilde{W}_c\right) \\
&\quad + \frac{1}{2} Tr\left(\tilde{W}_a^T \Phi_a \Phi_a^T \tilde{W}_a\right) + \frac{1}{2} Tr\left(W_c^T \Phi_c \Omega_a^T \Omega_a \Phi_c^T W_c\right) \\
&\quad - \hbar_a Tr\left(\tilde{W}_a^T \tilde{W}_a\right) + \frac{1}{2} \hbar_a Tr\left(\tilde{W}_a^T \tilde{W}_a\right) + \frac{1}{2} \hbar_a Tr\left(W_a^T W_a\right) \\
&\le -\frac{1}{2} \left(\hbar_a - \Phi_{aM}^2\right) Tr\left(\tilde{W}_a^T \tilde{W}_a\right) + \frac{1}{2} \left(\hbar_a + \Phi_{aM}^2\right) Tr\left(W_a^T W_a\right) \\
&\quad + \frac{1}{2} \Phi_{cM}^2 \Omega_{aM}^2 Tr\left(\tilde{W}_c^T \tilde{W}_c\right) + \frac{1}{2} \Phi_{cM}^2 \Omega_{aM}^2 Tr\left(W_c^T W_c\right)
\end{aligned} \tag{58}$$

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

Similarly, by taking time derivative of Equation 3.3 and combining Equation 48, it can be deduced that

$$\begin{aligned}
\dot{L}_3 &= Tr\left(\tilde{W}_c^T\lambda_c^{-1}\tilde{W}_c\right)\\
&= Tr\left(\tilde{W}_c^T\lambda_c^{-1}\left[-\lambda_c\left(\mathcal{L}+\hat{W}_c^T\dot{\Phi}_c\right)\dot{\Phi}_c-\lambda_c\hbar_c\hat{W}_c\right]\right)\\
&\le -Tr\left(\tilde{W}_c^T\dot{\Phi}_c\dot{\Phi}_c^T\tilde{W}_c\right)+Tr\left(\tilde{W}_c^T\dot{\Phi}_c\dot{\Phi}_c^T\tilde{W}_c\right)+Tr\left(W_c^T\dot{\Phi}_c\dot{\Phi}_c^TW_c\right)+\frac{1}{2}Tr\left(\tilde{W}_c^T\dot{\Phi}_c\dot{\Phi}_c^T\tilde{W}_c\right)\\
&\quad+\frac{1}{2}\varepsilon_{cdm}^2-\hbar_cTr\left(\tilde{W}_c^T\tilde{W}_c\right)+\frac{1}{2}\hbar_cTr\left(\tilde{W}_c^T\tilde{W}_c\right)+\frac{1}{2}\hbar_cTr\left(W_c^TW_c\right)\\
&\le -\frac{1}{2}\left(\hbar_c-\Phi_{cdM}^2\right)Tr\left(\tilde{W}_c^T\tilde{W}_c\right)+\frac{1}{2}\left(\hbar_c+2\Phi_{cdM}^2\right)Tr\left(W_c^TW_c\right)+\frac{1}{2}\varepsilon_{cdm}^2
\end{aligned}\tag{59}$$

At last, by taking time derivative of Equation 53 and substituting Equations 57–59, we get

$$\begin{aligned}
\dot{L} &= \dot{L}_1+\dot{L}_2+\dot{L}_3\\
&\le -\left(k_1-1\right)e_1^Te_1-\left(k_2-Tr\left(\frac{\partial g}{\partial\xi}\left(\frac{\partial g}{\partial\xi}\right)^T\right)\right)e_2^Te_2-\frac{1}{2}\left(\hbar_a-3\Phi_{aM}^2\right)Tr\left(\tilde{W}_a^T\tilde{W}_a\right)\\
&\quad-\frac{1}{2}\left(\hbar_c-\Phi_{cdM}^2-\Phi_{cM}^2\Omega_{aM}^2\right)Tr\left(\tilde{W}_c^T\tilde{W}_c\right)+\varepsilon_{am}^2+\frac{1}{2}\varepsilon_{cdm}^2\\
&\quad+\frac{1}{2}\left(\hbar_a+\Phi_{aM}^2\right)Tr\left(W_a^TW_a\right)+\frac{1}{2}\left(\hbar_c+2\Phi_{cdM}^2+\Phi_{cM}^2\Omega_{aM}^2\right)Tr\left(W_c^TW_c\right)
\end{aligned}\tag{60}$$

Equation 60 satisfies $\dot{L}\le-\kappa L\left(t\right)+\delta$ under the condition that $\left(k_1-1\right)>0$, $\left(k_2-Tr\left(\frac{\partial g}{\partial\xi}\left(\frac{\partial g}{\partial\xi}\right)^T\right)\right)>0$, $\left(\hbar_a-3\Phi_{aM}^2\right)>0$ and $\left(\hbar_c-\Phi_{cdM}^2-\Phi_{cM}^2\Omega_{aM}^2\right)>0$, where

$$\kappa=\min\left\{\begin{array}{l}2\left(k_1-1\right),2\left(k_2-Tr\left(\frac{\partial g}{\partial\xi}\left(\frac{\partial g}{\partial\xi}\right)^T\right)\right),\\ \left(\hbar_a-3\Phi_{aM}^2\right),\left(\hbar_c-\Phi_{cdM}^2-\Phi_{cM}^2\Omega_{aM}^2\right)\end{array}\right\}$$

$$\delta=\frac{1}{2}\left(\eta_a+\Phi_{aM}^2\right)Tr\left(W_a^TW_a\right)+\frac{1}{2}\Phi_{cM}^2\Omega_a^2Tr\left(W_c^TW_c\right)+\varepsilon_{am}^2+\frac{1}{2}\varepsilon_{cdm}^2$$

Therefore, e_1, e_2, $\tilde{W}_a$ and $\tilde{W}_c$ are uniformly ultimately bounded. □

Remark 3. e_1 is bounded indicating that $\theta\to\theta_d$, $\psi_v\to\psi_{vd}$ as $t\to\infty$, that is, the objective of this paper highlighted in Remark 1 is satisfied. $\tilde{W}_a$, $\tilde{W}_c$ are bounded indicating that $\hat{W}_a\to W_a^*$, $\hat{W}_c\to W_c^*$ as $t\to\infty$, that is, $\hat{d}\to d$ as $t\to\infty$. In conclusion, the designed actor-critic networks and the corresponding adaptive weight update laws can cope with unknown disturbances well.

Theorem 2. *For the spiral trajectory in the yaw plane, consider Theorem 1 and the geometric relationship shown in Fig. 4 of He et al. [20]. In addition, let the initial angle between the velocity vector of the vehicle and the virtual line-of-sight be such that* $|\delta\psi_v\left(0\right)|<\frac{\pi}{2}$. *If* $|\theta|<\frac{\pi}{2}$, *the guidance parameter* $\lambda_1<-1$ *renders* $s\to0$ *as* $t\to\infty$, *regardless of the value of V. The guidance parameter* $\lambda_1<-2$ *not only renders the flight trajectory converges to the spiral trajectory, but also renders the velocity vector of the vehicle converges to the virtual line-of-sight, meaning that* $\varphi-\psi_v\to-\frac{\pi}{2}$ *and* $\varphi-\psi_{vs}\to-\frac{\pi}{2}$.

Proof. It follows from Theorem 1 that

$$\psi_v=\psi_{vd}+e_{12}\tag{61}$$

where $|e_{12}|$ is an arbitrarily small constant. And the time derivative of Equation 61 gives

$$\dot{\psi}_v=\dot{\psi}_{vd}\tag{62}$$

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

Taking the time derivative of Equation 29 and substituting Equations 28, 34 and 62, the following equation can be derived:

$$\begin{aligned}\delta\dot{\psi}_v &= \dot{\varphi} - \dot{\psi}_v = (1+\lambda_1)\left(\dot{\varphi} - \dot{\psi}_{vs}\right) \\ &= (1+\lambda_1)\left(\frac{V\cos\theta}{s}\sin\left(\delta\psi_v\right) - \frac{V_t}{s}\cos\left(\psi_{vt} - \varphi\right) - \dot{\vartheta}\sin^2\left(\Delta\varphi\right)\right)\end{aligned} \tag{63}$$

Neglecting the second-order small quantity $\dot{\vartheta}\sin^2(\Delta\varphi)$ and the action term $\frac{V_t}{s}\cos(\psi_{vt} - \varphi)$ of the low-speed moving target in Equation 63, it can be rewritten as

$$\delta\dot{\psi}_v = (1+\lambda_1)\frac{V\cos\theta}{s}\sin\left(\delta\psi_v\right) \tag{64}$$

A similar treatment to time derivative of s produces

$$\dot{s} = -V\cos\theta\cos\left(\delta\psi_v\right) \tag{65}$$

Dividing Equation 65 by Equation 64, the Equation 66 can be obtained:

$$\frac{ds}{d\left(\delta\psi_v\right)} = -\frac{s}{1+\lambda_1}\frac{\cos\left(\delta\psi_v\right)}{\sin\left(\delta\psi_v\right)} \tag{66}$$

And the Equation 67 can be obtained by integrating the Equation 66:

$$s = \ell\left|\sin\left(\delta\psi_v\right)\right|^{-\frac{1}{1+\lambda_1}} \tag{67}$$

where $\ell > 0$ is the bounded integration constant.

If $\delta\psi_v(t)$ satisfies $0 < \delta\psi_v(0) < \frac{\pi}{2}$ at $t = 0$, then by substituting Equation 67 into Equation 64, we can get

$$\delta\dot{\psi}_v = (1+\lambda_1)\frac{V\cos\theta}{\ell}\left(\sin\left(\delta\psi_v\right)\right)^{\frac{\lambda_1+2}{\lambda_1+1}} \tag{68}$$

Note that $V\cos\theta > 0$ always holds no matter in which flight state. In the case $0 < \delta\psi_v(0) < \frac{\pi}{2}$, when the guidance parameter $\lambda_1 < -1$, there is $\delta\dot{\psi}_v < 0$, which indicates that $\delta\psi_v \to 0$ as $t \to \infty$. From Equation 67, it can be found that the remaining flight distance between the vehicle and the virtual sliding target $s \to 0$ as $\delta\psi_v \to 0$. Furthermore, from Equation 29, it can be found that $\varphi - \psi_v \to -\frac{\pi}{2}$ as $\delta\psi_v \to 0$. From Equation 68, when $\lambda_1 < -2$, there is $\delta\dot{\psi}_v \to 0$ as $\delta\psi_v \to 0$, so $\dot{\varphi} - \dot{\psi}_v \to 0$, and combining Equations 34 and 62, it can be observed that $\dot{\psi}_{vs} \to \dot{\psi}_v$. Therefore, the flight trajectory converges to the spiral trajectory. This means that $\Delta\varphi \to 0$, i.e. $\varphi - \psi_{vs} \to -\frac{\pi}{2}$. $\delta\psi_v \to 0$ and $\Delta\varphi \to 0$ indicate that velocity vector of the vehicle converges to the virtual line-of-sight. The same conclusion can be obtained when $-\frac{\pi}{2} < \delta\psi_v(0) < 0$. □

Theorem 3. *Considering Theorem 1 and Theorem 2, the guidance parameter* $\lambda_2 > 1$ *renders the distance from the vehicle to the virtual sliding target* $r \to 0$ *as* $t \to \infty$. *The guidance parameter* $\lambda_2 > 2$ *also renders* $\phi + \theta \to 0$.

Proof. It follows from Theorem 1 that

$$\theta = \theta_d + e_{11} \tag{69}$$

where $|e_{11}|$ is an arbitrarily small constant. And the time derivative of Equation 69 gives

$$\dot{\theta} = \dot{\theta}_d \tag{70}$$

Define $\sigma = \phi + \theta$, by taking its time derivative and combining Equations 35 and 70, we get

$$\dot{\sigma} = \dot{\phi} + \dot{\theta} = (1-\lambda_2)\dot{\phi} \tag{71}$$

Taking the time derivative of r, and then neglecting the low-speed moving target action term and noting that $\delta\psi_v \to 0$, $\Delta\varphi \to 0$ as $t \to \infty$, yields

$$\dot{r} = -V\cos\sigma \tag{72}$$

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

Table 1. *The parameters of two cases*

Symbols	Case 1	Case 2
λ_1, λ_2	$-3, 3$	-2.1 2.02
k_1, k_2	4, 4	4, 4
$\lambda_c, \hbar_c, \lambda_a, \hbar_a$	0.5, 2e-9, 0.0105, 0.0095	0.1, 1e-8, 0.01, 5
d	$\begin{bmatrix} -0.01\sin(0.023t) \\ -0.01\sin(0.023t) \end{bmatrix}$	$\begin{bmatrix} 0.02e^{-0.05t} \\ -0.02e^{-0.05t} \end{bmatrix}$

Similarly, there is

$$\dot{\phi} = \frac{V}{r}\sin\sigma \tag{73}$$

Dividing Equation 72 by Equation 71 yields

$$\frac{dr}{d\sigma} = \frac{r}{\lambda_2 - 1}\frac{\cos\sigma}{\sin\sigma} \tag{74}$$

And the Equation 75 can be obtained by integrating the Equation 74:

$$r = \ell' |\sin\sigma|^{\frac{1}{\lambda_2 - 1}} \tag{75}$$

where $\ell' > 0$ is the bounded integration constant. Combining Equations 71, 73 and 75, the Equation 76 can be organized:

$$\dot{\sigma} = (1 - \lambda_2)\frac{V}{\ell'}(\sin\sigma)^{\frac{\lambda_2 - 2}{\lambda_2 - 1}} \tag{76}$$

when $0 < \sigma\ (0) < \pi$, if the guidance parameter $\lambda_2 > 1$, then $\dot{\sigma} < 0$. Therefore $\sigma \to 0$ as $t \to \infty$. From Equation 75, it can be found that spatial distance from the vehicle to the virtual sliding target $r \to 0$ as $\sigma \to 0$. If $\lambda_2 > 2$, there is $\dot{\sigma} \to 0$. Because $\dot{\sigma} = (1 - \lambda_2)\dot{\phi}$, so $\dot{\phi} \to 0$. Thus, ϕ approaches a constant and θ approaches the negative of the same constant. That is, in the pitch plane, $\phi + \theta \to 0$. The same conclusion can be obtained when $-\pi < \sigma\ (0) < 0$. □

Remark 4. As can be seen from Equations 61, 62 and Equations 69, 70, the assumptions that $\dot{\psi}_v = \dot{\psi}_{vd}$ and $\dot{\theta} = \dot{\theta}_d$ made in the He and Yan [19] and He et al. [20] are verified.

4.0 Simulations

In this section, some simulations are presented to demonstrate the validity and superiority of the proposed reinforcement learning based adaptive spiral-diving guidance method. Specifically, the validity of the proposed method is verified by striking a stationary target and a low-speed moving target. For convenience, the former is denoted as Case 1, and the latter is denoted as Case 2. Otherwise, the superiority of the proposed method is demonstrated by comparing it with methods that without RL and RBF neural networks for unknown disturbances.

The parameters of the vehicle are: vehicle mass $m = 200$ kg, the reference area $S_{ref} = 1.8$ m. The initial position $(x_0, y_0, z_0)^T = (0, 32, 60)^T$ km, and the initial velocity $V_0 = 1200$ m/s. The initial path angle $\theta_0 = -2°$, and the initial deflection angle $\psi_{v0} = 28°$. The terminal deflection angle $\psi_{vf} = 363°$. Moreover, the gravitational acceleration $g = 9.81$ m/s ^{2}.The position of the stationary target is $(0, 0, 0)^T$ km, which is also the starting point of the low-speed moving target. Note that the low-speed target moves only in the horizontal plane with velocity $V_t = 9$ m/s and directional angle $\psi_{vt} = -90°$. Other parameters of the two cases are shown in Table 1.

The simulation results for Case 1 and Case 2 are shown in Figs. 2 and 3, respectively. The 3-D spiral trajectories of the vehicle in two cases are shown in Figs. 2(a) and 3(a). And corresponding trajectories

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

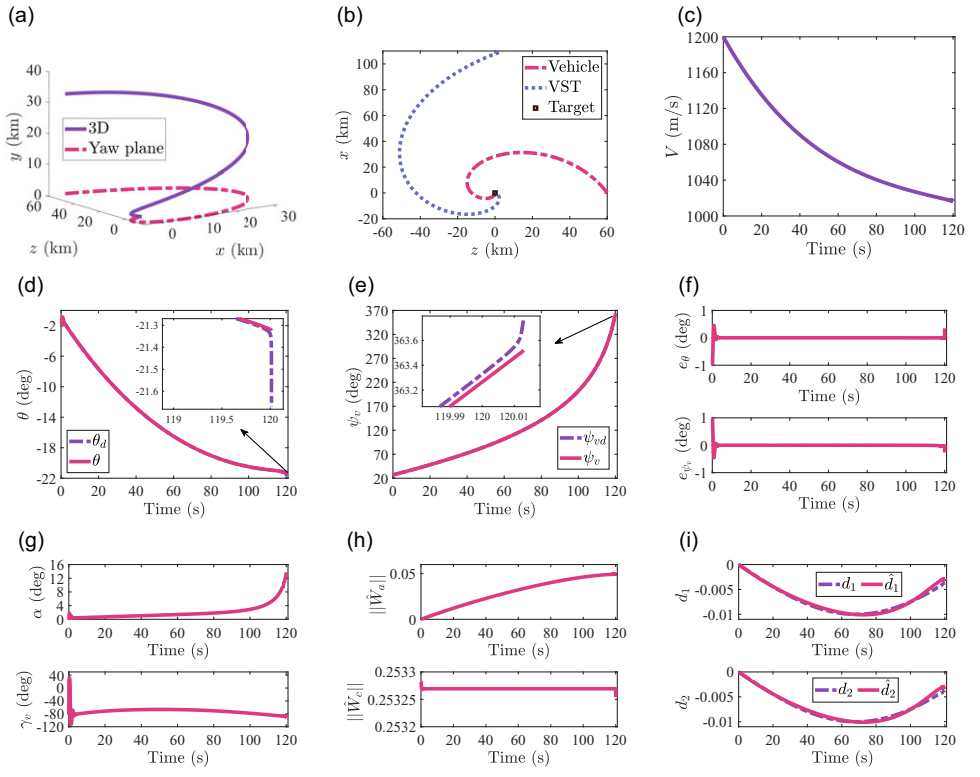

Figure 2. *Simulation profiles for Case 1. (a) 3-D trajectory of the vehicle. (b) Trajectories of the vehicle, the target and the virtual sliding target in yaw plane. (c) Vehicle velocity. (d) Path angle tracking curve. (e) Deflection angle tracking curve. (f) Tracking errors. (g) Control inputs. (h) Adaptive weights. (i) Real and estimated values of disturbances.*

of the vehicle, the target and the virtual sliding target in yaw plane are shown in Figs. 2(b) and 3(b). They indicate that the vehicle is able to hit the target in both cases, and the respective miss distances are 0.449 m and 0.6092 m. Figures 2(c) and 3(c) show the vehicle velocity response profiles in two cases. Figures 2(d) and 3(d) show the real path angle versus desired path angle for two cases. And Figs. 2(e) and 3(e) show the real deflection angle versus desired deflection angle for two cases. At the moment of hitting the target, the desired path angle and the deflection angle have a small jump. The reason for this is that the vehicle needs to slow down the descent rate in the y-direction to adjust the motion in the x, z-directions to reduce the miss distance. The tracking errors of the path angle and the deflection angle in two cases are shown in Figs. 2(f) and 3(f). So the uniform ultimate boundedness of the tracking error is proved. The control input profiles under reinforcement learning based adaptive law for two cases are depicted in Figs. 2(g) and 3(g). Figures 2(h) and 3(h) show the adaptive adjustment profiles of the weights in the two cases, and their effects are verified in Figs. 2(i) and 3(i). In other words, the unknown disturbances are well compensated. In brief, the above simulation results fully demonstrate the validity of the proposed method in this paper.

Without loss of generality, a comparative simulation of the proposed method with Without RL method and RBF method is included based on Case 2. The striking effects of the three methods are shown in Fig. 4. And as shown in Table 2, the miss distances under the three methods are 0.6092 m, 0.8718 m and 0.7602 m, respectively. The strike accuracy of the proposed method has improved by 30.12% and 19.86% compared to without RL method and RBF method.

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

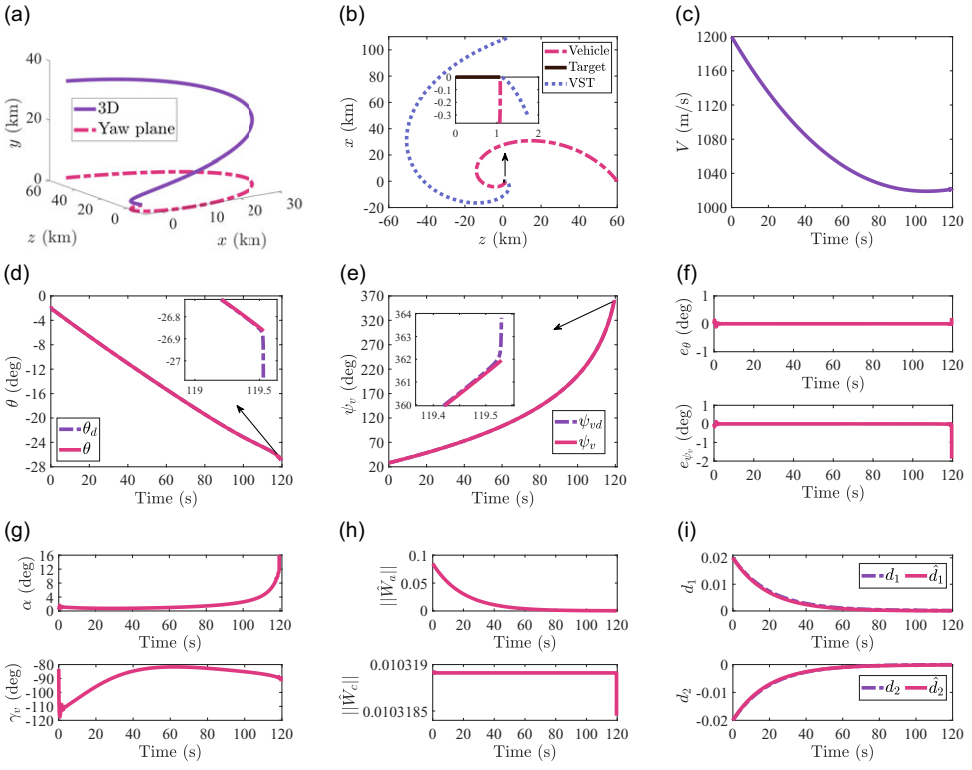

Figure 3. *Simulation profiles for Case 2. (a) 3-D trajectory of the vehicle. (b) Trajectories of the vehicle, the target and the virtual sliding target in yaw plane. (c) Vehicle velocity. (d) Path angle tracking curve. (e) Deflection angle tracking curve. (f) Tracking errors. (g) Control inputs. (h) Adaptive weights. (i) Real and estimated values of disturbances.*

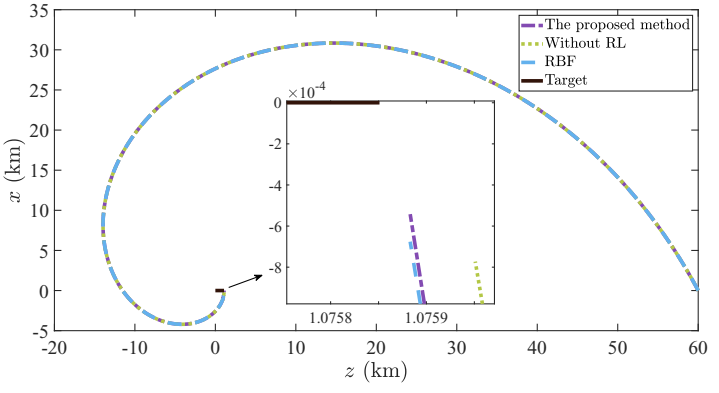

Figure 4. *The striking effects under three methods.*

5.0 Conclusion

In this paper, the reinforcement learning based adaptive method has been implemented for a class of spiral-diving manoeuver guidance problems of reentry vehicles subject to unknown disturbances. By designing the actor-critic networks and the corresponding adaptive weight update laws, the unknown disturbances are well compensated. In addition, by introducing the coordinate transformation technique,

https://doi.org/10.1017/aer.2024.17 Published online by Cambridge University Press

Table 2. *The miss distances under the three methods*

Three methods	Miss distances
The proposed method	0.6092
Without RL	0.8718
RBF	0.7602

the controller design problem caused by the coupling of control variables is overcome. As a result, a novel reinforcement learning based adaptive guidance framework has been constructed such that desired guidance commands can be tracked stably. Some numerical simulations have been provided to demonstrate the validity and superiority of the proposed method. Based on the work done in this paper, we will study the cooperative spiral-diving guidance of reentry vehicle formation.

Acknowledgements. This work was supported by the Foundation of National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics and The Foundation of Shanghai Astronautics Science and Technology Innovation, China.

References

[1] Bairstow S.H. Reentry guidance with extended range capability for low l/d spacecraft, PhD thesis, Massachusetts Institute of Technology, 2006.
[2] Zhang Mingang Y.D. A sinking trajectory planning and design method for reentry vehicle under multiple constraints. *Missiles Space Veh.*, 2022, **4**, (25–28). doi: 10.7654/j.issn.1004-7182.20220406
[3] Rahmani-Nejad A. An anti-hypersonic missiles hybrid optical and enhanced Railgun system. In *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies V*, Vol. 11869, SPIE, Washington State, USA, 2021, pp. 107–116.
[4] Liu S., Yan B., Liu R., Dai P., Yan J. and Xin G. Cooperative guidance law for intercepting a hypersonic target with impact angle constraint. *Aeronaut. J.*, 2022, **126**, (1300), pp 1026–1044. doi: 10.1017/aer.2021.117
[5] Yu L.L.X. Weaving maneuver trajectory design for hypersonic glide vehicles, *Acta Aeronaut. Astronaut. Sin.*, 2011, **32**, (2174–2181). doi: 10.11-1129/V.20110921.0830.001
[6] Dwivedi P., Bhale P., Bhattacharyya A. and Padhi R. Generalized state estimation and model predictive guidance for spiraling and ballistic targets. *J. Guid. Control Dyn.*, 2014, **37**, (1), pp 243–264. doi: 10.2514/1.60075
[7] He L., Yan X. and Tang S. Spiral-diving trajectory optimization for hypersonic vehicles by second-order cone programming. *Aerosp. Sci. Technol.*, 2019, **95**, p. 105427. doi: 10.1016/j.ast.2019.105427
[8] Li G., Zhang H. and Tang G. Maneuver characteristics analysis for hypersonic glide vehicles. *Aerosp. Sci. Technol.*, 2015, **43**, pp 321–328. doi: 10.1016/j.ast.2015.03.016
[9] Rusnak I and Peled-Eitan L. Guidance law against spiraling target. *J. Guid. Control Dyn.*, 2016, **39**, (7), 1694–1696. doi: 10.2514/1.G001646
[10] Yibo D., Xiaokui Y., Guangshan C. and Jiashun S. Review of control and guidance technology on hypersonic vehicle. *Chin. J. Aeronaut.*, 2022, **35**, (7), pp 1–18. doi: 10.1016/j.cja.2021.10.037
[11] Zhang W. and Wang B. A new guidance law for impact angle constraints with time-varying navigation gain. *Aeronaut. J.*, 2022, **126**, (1304), pp 1752–1770. doi: 10.1017/aer.2022.16
[12] Dou L. and Dou J. The design of optimal guidance law with multi-constraints using block pulse functions. *Aerosp. Sci. Technol.*, 2012, **23**, (1), pp 201–205. doi: 10.1016/j.ast.2011.02.009
[13] Qing R.M.W. Reentry guidance for hypersonic vehicle based on predictor-corrector method. *J. Beijing Univ. Aeronaut. Astronaut.*, 2013, **39** (1563–1567). doi: 10.13700/j.bh.1001-5965.2013.12.013
[14] Li T. and Qian H. Design of three-dimensional guidance law with impact angle constraints and input saturation. *IEEE Access*, 2020, **8**, pp 211474–211481. doi: 10.1109/ACCESS.2020.3038830
[15] Dhananjay N. and Ghose D. Accurate time-to-go estimation for proportional navigation guidance. *J. Guid. Control Dyn.*, 2014, **37**, (4), pp 1378–1383. doi: 10.2514/1.G000082
[16] Mozaffari M., Safarinejadian B. and Binazadeh T. Optimal guidance law based on virtual sliding target. *J. Aerosp. Eng.*, 2017, **30**, (3), p 04016097. doi: 10.1061/(ASCE)AS.1943-5525.0000692
[17] Raju P. and Ghose D. Empirical virtual sliding target guidance law design: an aerodynamic approach. *IEEE Trans. Aerosp. Electron. Syst.*, 2003, **39**, (4), pp 1179–1190. doi: 10.1109/TAES.2003.1261120
[18] Hu Q., Han T. and Xin M. New impact time and angle guidance strategy via virtual target approach. *J. Guid. Control Dyn.*, 2018, **41**, (8), pp 1755–1765. doi: 10.2514/1.G003436
[19] He L. and Yan X. Adaptive terminal guidance law for spiral-diving maneuver based on virtual sliding targets. *J. Guid. Control Dyn.*, 2018, **41**, (7), pp 1591–1601. doi: 10.2514/1.G003424
[20] He L., Yan X. and Tang S. Guidance law design for spiral-diving maneuver penetration. *Acta Aeronaut. Astronaut. Sin.*, 2019, **40**, (193–207). doi: 10.7527/S1000-6893.2019.22457

[21] Ning X., Liu J., Wang Z. and Luo C. Output constrained neural adaptive control for a class of kkvs with non-affine inputs and unmodeled dynamics. *Aeronaut. J.*, 2024, **128**, (1319), pp 134–151. doi: 10.1017/aer.2023.44
[22] Yin Z., Wang B., Xiong R., Xiang Z., Liu L., Fan H. and Xue C. Attitude tracking control of hypersonic vehicle based on an improved prescribed performance dynamic surface control. *Aeronaut. J.* 2023, pp 1–21. doi: 10.1017/aer.2023.79
[23] Mao Q., Dou L., Yang Z., Tian B. and Zong Q. Fuzzy disturbance observer-based adaptive sliding mode control for reusable launch vehicles with aeroservoelastic characteristic. *IEEE Trans. Ind. Inf.*, 2019, **16**, (2), pp 1214–1223. doi: 10.1109/TII.2019.2924731
[24] Shen G., Xia Y., Zhang J. and Cui B. Adaptive super-twisting sliding mode altitude trajectory tracking control for reentry vehicle. *ISA Trans.*, 2023, **132**, pp 329–337. doi: 10.1016/j.isatra.2022.06.023
[25] Chai R., Tsourdos A., Gao H., Chai S. and Xia Y. Attitude tracking control for reentry vehicles using centralised robust model predictive control. *Automatica*, 2022, **145**, p 110561. doi: 10.1016/j.automatica.2022.110561
[26] Cao L. and Xiao B. Exponential and resilient control for attitude tracking maneuvering of spacecraft with actuator uncertainties. *IEEE/ASME Trans. Mechatron.*, 2019, **24**, (6), pp 2531–2540. doi: 10.1109/TMECH.2019.2928703
[27] Xiang K., Yanli D., Peng Z. and Haibing L. Adaptive backstepping control for hypersonic vehicles with actuator amplitude and rate saturation. *Trans. Nanjing Univ. Aeronaut. Astronaut.*, 2019, **36**, (2), pp 242–252. doi: 10.16356/j.1005-1120.2019.02.007
[28] Cheng L., Wang Z. and Gong S. Adaptive control of hypersonic vehicles with unknown dynamics based on dual network architecture. *Acta Astronaut.*, 2022, **193**, pp 197–208. doi: 10.1016/j.actaastro.2021.12.043
[29] Cheng L., Wang Z., Jiang F. and Li J. Fast generation of optimal asteroid landing trajectories using deep neural networks. *IEEE Trans. Aerosp. Electron. Syst.*, 2019, **56**, (4), pp 2642–2655. doi: 10.1109/TAES.2019.2952700
[30] Zhang X., Chen K., Fu W. and Huang H. Neural network-based stochastic adaptive attitude control for generic hypersonic vehicles with full state constraints. *Neurocomputing*, 2019, **351**, pp 228–239. doi: 10.1016/j.neucom.2019.04.014
[31] Ouyang Y., Dong L., Wei Y. and Sun C. Neural network based tracking control for an elastic joint robot with input constraint via actor-critic design. *Neurocomputing*, 2020, **409**, pp 286–295. doi: 10.1016/j.neucom.2020.05.067
[32] Shi L., Wang X. and Cheng Y. Safe reinforcement learning-based robust approximate optimal control for hypersonic flight vehicles. *IEEE Trans. Veh. Technol.*, 2023. doi: 10.1109/TVT.2023.3264243
[33] Wang Z. and Liu J. Reinforcement learning based-adaptive tracking control for a class of semi-markov non-lipschitz uncertain system with unmatched disturbances. *Inf. Sci.* 2023, **626**, pp 407–427. doi: 10.1016/j.ins.2023.01.043
[34] He W. and Dong Y. Adaptive fuzzy neural network control for a constrained robot using impedance learning. *IEEE Trans. Neural Netwk. Learn. Syst.*, 2017, **29**, (4), pp 1174–1186. doi: 10.1109/TNNLS.2017.2665581
[35] Xing-Kai H. Young type inequalities for matrices. *J. East China Normal Univ. (Nat. Sci.)*, 2012, **2012**, (4), p 12. doi: 10.3969/j.issn.1000-5641.2012.04.002

Cite this article: Wu T. and Wang Z. Reinforcement learning-based adaptive spiral-diving Manoeuver guidance method for reentry vehicles subject to unknown disturbances. *The Aeronautical Journal*, https://doi.org/10.1017/aer.2024.17