




Research on Guidance Method of Hypersonic Vehicle Based on Reinforcement Learning and Dynamic Surface Control

Yin Diao¹^a, Baogang Lu²^b and Yingzi Guan³^c

¹Beijing Institute of Astronautical Systems Engineering, Beijing, China

²Science and Technology on Space Physics Laboratory, Beijing, China

³School of Astronautics, Harbin Institute of Technology, Harbin, China

Keywords: Hypersonic Vehicle, Reinforcement Learning, Dynamic Surface Control, Online Guidance.


Abstract: To meet the requirements of high-precision tracking of long-range hypersonic vehicle position and minimization of terminal velocity deviation, this paper completes the online generation of guidance commands based on dynamic surface control theory and reinforcement learning method. First, this paper transforms the control problem of three-dimensional under-actuated system into a problem of two-dimensional path-tracking and establishes the basic framework of the guidance system by using the control method of dynamic surface path-tracking, and the online optimal adjustment of guidance parameters is then accomplished through the online network of flight state deviation and reinforcement learning to achieve the minimization of the integrated deviation of the process position and terminal velocity. The simulation results show that the proposed guidance method can solve the high-precision position tracking problem of long-range hypersonic vehicles effectively, and it can reduce the terminal velocity deviation. The algorithm computation is small, which has good prospects for engineering applications.


1 INTRODUCTION


Hypersonic vehicles are fast, have long flight range, and can achieve flexible mission maneuvers, but modeling and perturbation deviations under complex environmental conditions can seriously affect the flight capability and mission execution effectiveness of hypersonic vehicles. Once the optimal trajectory satisfying the mission and constraints is planned offline or online, a high-precision trajectory tracking guidance system is the key to ensuring that the hypersonic vehicle is effective.

For a nonlinear system with large perturbations and strong uncertainties such as hypersonic vehicles, the backstepping method can divide the higher-order system into several lower-order subsystems based on modeling deviations, and achieve the asymptotic stability of the system through the rectification of the subsystems that meet the Lyapunov stability requirements (Xu et al, 2011). However, due to the need to derive the virtual control quantities, it is easy to lead to the problem of "complexity explosion". The

adaptive dynamic surface control (ADSC) method is based on the backstepping method and solves the "complexity explosion" problem by adding a first-order low-pass filter (Swaroop et al, 1997), so it has also been more widely used. The combination of extended observer and dynamic inversion technology improved the response speed and accuracy of attitude control for hypersonic vehicles (Liu et al, 2015). The Reference (Xu et al, 2014) designed a general hypersonic vehicle longitudinal controller based on adaptive dynamic surface method. In the literature (Wu et al, 2021), a finite-time control strategy is proposed by combining dynamic surface trajectory tracking control with sliding mode attitude control. The literature (Butt et al, 2010; Butt et al, 2013) combined dynamic surface control theory with neural networks for the design of tracking control systems, which better dealt with the effect of nonlinear terms in the system. After that, (Yu et al, 2014; Xu et al, 2016; Shin, 2017) used neural networks with different structures to compensate the nonlinear terms, so as to achieve accurate tracking of the states such as altitude

^a <https://orcid.org/0000-0002-1102-2745>

^b <https://orcid.org/0000-0001-8277-1922>

^c <https://orcid.org/0000-0001-7925-2400>

and speed. The literature (Hu et al, 2013) combined dynamic surface control with a dynamic inverse strategy to improve the robustness of the control system against the effects of model uncertainty. The literature (Aguiar et al, 2007) used a path tracking method to transform the 3D underdriven trajectory tracking problem into a 2D tracking control problem and achieved a better tracking control effect using dynamic surface control, but the control parameters could not be adaptively adjusted to meet the demand of long-range unpowered gliding trajectory tracking under strong disturbance conditions.

Among the model-free reinforcement learning methods, the actor-critic algorithm combines the advantages of policy-based methods in terms of continuous action space problem description and value-based methods in terms of convergence speed, and has therefore been extensively studied. In the literature (Li et al, 2018), an optimization model for hypersonic vehicle control parameters was designed in the actor-critic framework. The Reference (Zhen et al, 2019) designed a PID controller based on actor critical network which adjusted parameters online. The literature (Lillicrap et al, 2015) applied the successful concept of Deep Q-Learning to the

continuous action domain and proposed an Actor-Critic deep deterministic policy gradient (DDPG) approach based on no model dependency, which successfully solved several simulated physics tasks. The literature (Cheng et al, 2019; Gao, 2019) applied the DDPG method to complete the optimization of hypersonic re-entry flight trajectories with terminal altitude and velocity magnitude constraints using velocity inclination as the action space, but did not consider the process position tracking and terminal constraint guidance requirements under perturbation conditions.

In this paper, for the needs of high precision position tracking and terminal velocity deviation minimization of long-range hypersonic vehicles, the basic framework of the guidance system is established by adopting the path tracking idea and ADSC theory, and converting the effects of earth rotation and curvature into system modeling deviations, while introducing the DDPG reinforcement learning method to generate the optimal adjustment network of guidance parameters under perturbation conditions for online generation of guidance commands, and finally, the effectiveness of the proposed method is verified by simulation.

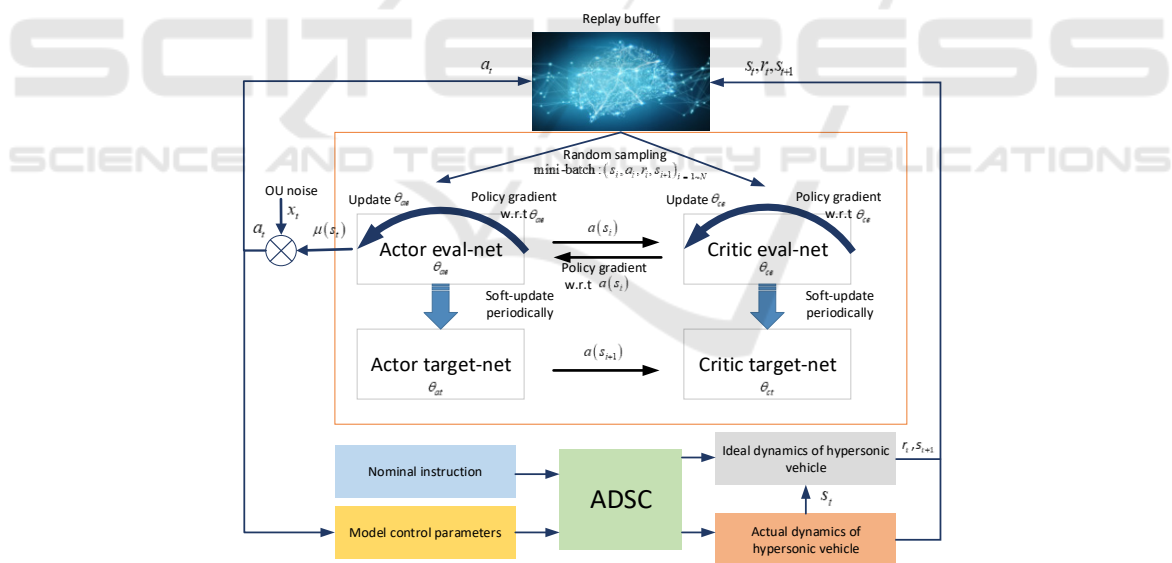


Figure 1: Guidance system training and online application framework based on DDPG.

2 OVERALL FRAMEWORK

The hypersonic vehicle guidance method based on dynamic surface control and reinforcement learning takes the adaptive dynamic surface control method (ADSC) as the basic guidance framework and adopts the angle of attack increment and velocity inclination

increment as the control commands for position tracking; meanwhile, s_i is defined as the reinforcement learning state expressed in terms of instantaneous flight state and predicted terminal state, a_i is the reinforcement learning action expressed in terms of ADSC control matrix coefficients, and r_i is

the reward function, and the control parameters of ADSC are continuously adjusted by the DDPG method to finally obtain the optimal control parameters that satisfy the path and terminal constraints. The overall framework of the guidance system is shown in Figure 1.

3 DYNAMIC SURFACE CONTROL TRACKING GUIDANCE METHODS

The reference trajectory is interpolated with the actual x position coordinates of the vehicle to obtain the

reference control program angle, which transforms the three-dimensional trajectory tracking problem into a two-dimensional path tracking problem. A simplified model of the dynamics at half speed is used in the calculation of the guidance command (Song et al, 2016). The differences between the simplified and real models are then considered in ADSC in the form of model deviations. The form of the control system used is:

$$\begin{cases} \dot{x}_1 = f_1 + w_1 = x_2 \\ \dot{x}_2 = f_2 + Bu + w_2 + \dot{w}_1 \\ \dot{x}_c = f_c + w_c \end{cases} \quad (1)$$

Where w_1 and w_2 are the model deviations, and the variables in the control equation are:

$$\begin{aligned} w_1 &= [w_{11} \quad w_{12}]^T & w_2 &= [w_{21} \quad w_{22}]^T \\ x_1 &= [y \quad z]^T & x_2 &= [y \quad z]^T & u &= [\Delta\alpha \quad \Delta\gamma_v]^T \end{aligned} \quad (2)$$

$$f_1 = [f_{11} \quad f_{12}]^T \quad f_2 = [f_{21} \quad f_{22}]^T \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$x_c = [y_{\text{ref}} \quad z_{\text{ref}}]^T \quad w_c = [w_{c1} \quad w_{c2}]^T$$

$$f_c = [v_{\text{ref}} \sin \theta_{\text{ref}} \quad -v_{\text{ref}} \cos \theta_{\text{ref}} \sin \psi_{\text{vref}}]^T$$

$$\begin{cases} f_{11} = v \sin \theta \\ f_{12} = -v \cos \theta \sin \psi_v \\ f_{21} = C_{L\text{ref}} q S_r \cos \theta \cos \gamma_{\text{reff}} / m - C_{D\text{ref}} q S_r \sin \theta / m - g \\ f_{22} = C_{L\text{ref}} q S_r (\sin \theta \sin \psi_v \cos \gamma_{\text{trf}} + \cos \psi_v \sin \gamma_{\text{veff}}) / m \\ \quad + C_{D\text{ref}} q S_r \cos \theta \sin \psi_v / m \end{cases} \quad (3)$$

$$\begin{cases} b_{11} = -C_{D\text{ref}}^\alpha q S_r \sin \theta / m + \cos \theta \cos \gamma_{\text{vref}} C_{L\text{ref}}^\alpha q S_r / m \\ b_{12} = -C_{L\text{ref}} q S_r \cos \theta \sin \gamma_{\text{vref}} / m \\ b_{21} = C_{L\text{ref}}^\alpha q S_r (\sin \theta \sin \psi_v \cos \gamma_{\text{trf}} + \cos \psi_v \sin \gamma_{\text{veff}}) / m + C_{D\text{ref}}^\alpha q S_r \cos \theta \sin \psi_v / m \\ b_{22} = C_{L\text{ref}} q S_r (\cos \psi_v \cos \gamma_{\text{vref}} - \sin \theta \sin \psi_v \sin \gamma_{\text{vref}}) / m \end{cases} \quad (4)$$

Where the subscripts c and ref denote the variables on the reference trajectory, the control quantities $\Delta\alpha$ and $\Delta\gamma_v$ are the differences from the reference angle of attack and velocity inclination, $C_{D\text{ref}}$ and $C_{L\text{ref}}$ are the aerodynamic coefficients on the reference trajectory, $C_{D\text{ref}}^\alpha$ and $C_{L\text{ref}}^\alpha$ are the aerodynamic derivatives on the reference trajectory.

Designing of guidance laws:

$$\begin{cases} e_1 = x_1 - x_c \\ x_{2c} = g_c - K_1 e_1 - \dot{w}_1 - \hat{\delta} \\ \tau \dot{x}_{2d} + x_{2d} = x_{2c}, x_{2d}(0) = x_{2c}(0) \\ e_2 = g_1 - x_{2d} \\ u = B^{-1} (\dot{x}_{2d} - g_2 - K_2 e_2 - \dot{w}_2) \end{cases} \quad (5)$$

Where τ , K_1 , K_2 are non-singular diagonal control matrices.

Define filtering error:

$$\delta = x_{2d} - x_{2c} \quad \dot{\delta} = -\tau^{-1} \delta - \dot{x}_{2c} \quad (6)$$

The adaptive equations for the model error \hat{w}_i and the filtering error $\hat{\delta}$ are:

$$\begin{cases} \dot{\hat{w}}_i = \mathbf{Q}_i e_i - \mathbf{K}_{w_i} \hat{w}_i, & \hat{w}_i(0) = 0 \\ \dot{\hat{\delta}} = \mathbf{Q}_\delta e_\delta - \mathbf{K}_\delta \hat{\delta}, & \hat{\delta}(0) = 0 \end{cases} \quad (7)$$

Where \mathbf{Q}_i , \mathbf{K}_{w_i} , \mathbf{K}_δ are non-singular diagonal matrices, and a sufficiently large \mathbf{Q}_i can make the position tracking error e_i sufficiently small, as demonstrated in the literature (Swaroop et al, 1997). However, for a long-range unpowered gliding hypersonic vehicle, too large a \mathbf{Q}_i will lead to an increase in the velocity error, which in turn will make it difficult to maintain the flight state when approaching the flight terminal due to practical constraints such as flight angle of attack, and eventually lead to a sharp increase in both position and velocity deviations. Therefore, the optimal control coefficients \mathbf{Q}_i need to be selected online for different flight states and deviation conditions in order to minimize the combined deviation of process position and terminal velocity.

4 REINFORCEMENT LEARNING BASED CONTROL PARAMETER TUNING METHOD

In this paper, an actor-critic-based DDPG reinforcement learning architecture is built with a

$$\begin{aligned} \theta_{ce}^{t+1} &= \theta_{ce}^t + \lambda_c [r_t + \kappa Q_{\theta_{ct}}(s_{t+1}, \mu_{\theta_{at}}(s_{t+1})) - Q_{\theta_{ce}}(s_t, a_t)] \\ \theta_{ae}^{t+1} &= \theta_{ae}^t + \lambda_a \nabla \mu_{\theta_{ae}}(s_t) \nabla Q_{\theta_{ce}}(s_t, a_t) \end{aligned} \quad (10)$$

Where: λ_c and λ_a are the learning rates of critic eval-net and actor eval-net, respectively, r_t is the current reward value, and κ is the discounting factor. When eval-net is updated p_t times, the DDPG algorithm periodically soft-update target-net.

$$\begin{cases} \theta_{ct}^{t+1} = (1 - \tau) \theta_{ct}^t + \tau \theta_{ce} \\ \theta_{at}^{t+1} = (1 - \tau) \theta_{at}^t + \tau \theta_{ae} \end{cases} \quad (11)$$

The actor eval-net parameters $\theta_{ae}(s)$ after the training can be used to adjust the control parameters of the hypersonic vehicle during guided flight online.

vehicle as to the agent. Actor outputs actions $a(s_t)$ based on state decisions s_t , and critic evaluates Q values based on state s_t and actions a_t . The relevant learning elements are:

$$\begin{aligned} o_t &= [x, h, v, \theta, \psi, \alpha, \sigma, h_f, v_f, h_{fpre}, v_{fpre}] \\ s_t &= [o_{t-p_e}, \dots, o_{t-1}, o_t] \\ a_t &= [Q_1, Q_2] \\ r_t &= 160 - \|h_{fpre} - h_f\| / 40 - \|v_{fpre} - v_f\| / 10 \end{aligned} \quad (8)$$

Considering the actual aerodynamic characteristics and control requirements, the guidance system needs to limit the actions and program angles:

$$\begin{aligned} Q_i &\in [0.1, 3], i = 1, 2 \\ \alpha &\in [0^\circ, 30^\circ], \dot{\alpha}_{max} = 5^\circ/s, \sigma \in [-70^\circ, 70^\circ] \end{aligned} \quad (9)$$

To explore well in physical environments that have momentum, Ornstein-Uhlenbeck (OU) noise (Uhlenbeck et al, 1930) is added to the output of the actions by the actor-network during reinforcement learning training, $[s_t, a_t, r_t, s_{t+1}]$ is obtained by interacting with the environment and stored in the replay buffer. Every time the intelligence interacts with the environment p_e step by step, a sample is randomly selected N from the replay buffer for training and updating the parameters of critic eval-net θ_{ce} and actor eval-net θ_{ae} using the Adam (Kingma et al, 2014).

5 REINFORCEMENT LEARNING BASED CONTROL PARAMETER TUNING METHOD

5.1 Simulation Conditions

The simulation uses a publicly available CAV model with a mass of 907 kg, an aerodynamic reference area of 0.48 m², and the mission parameters shown in Table 1.

The simulation step and guidance period are both taken as 100 ms, and the control parameter update and environment interaction period is 1 s. The structure of actor-net is set to $[11 \times 5, 20, 20, 10, 2]$, and the

structure of critic-net is set to $[11 \times 5 + 2, 20, 20, 10, 1]$. The training hyperparameters are shown in Table 2.

Table 1: List of vehicle mission parameters.

	Parameters	Parameter value
Initial parameters	Height/km	60
	Speed/(m/s)	5,500
	Flight path angle/(°)	-1
	Heading angle/(°)	90
	Design range/km	4300
Terminal parameters	Height/km	28
	Speed/(m/s)	1,840
	Remaining range/km	50

Table 2: Super parameter setting of RL training.

Item	Value
p_e	5
N	32
λ_a	0.000 1
λ_c	0.002
κ	0.99
p_t	5
τ	0.001
κ_0 (OU noise)	0.15
η_0 (OU noise)	0.15

Table 3: Status deviations and environmental disturbances.

Deviations	Value
Initial of height $\square h_0$ (m)	-100~100
Initial of velocity $\square v_0$ (ms ⁻¹)	-20~20
Initial of flight path angle $\square \gamma_0$ (°)	-1~1
Resistance coefficient (%)	-10~10
Lift coefficient (%)	-10~10
Atmospheric density (%)	-10~10

5.2 Reinforcement Learning Training

Introducing normally distributed state biases and environmental perturbations during reinforcement learning training. The status deviations and environmental disturbances are shown in Table 3.

With 1000 training sessions, the reward function gradually converges, as shown in Figure 2.

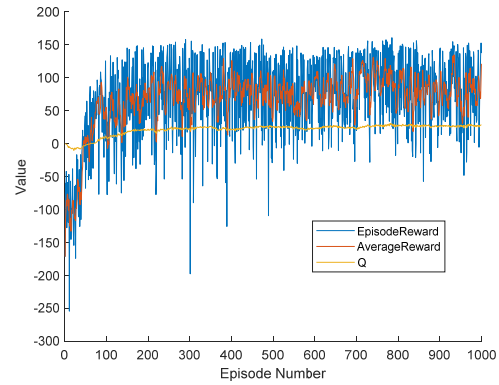


Figure 2: Curve of reward function in the process of reinforcement learning and training.

5.3 Guidance Simulation Results

To verify the adaptability of the guidance method proposed in this paper, given the limiting pull-off condition.

$$\begin{cases} \square h_0 = \pm 100\text{m}, \square v_0 = \pm 20\text{m/s} \\ \square \gamma_0 = \pm 0.5^\circ, \square C_L = \pm 10\% \\ \square C_D = \mp 10\%, \square \rho = \mp 10\% \end{cases} \quad (12)$$

The ADSC guidance law and ADSC+DDPG methods are used for guidance simulation, respectively. The simulation results are shown in Figures 3-8.

Under the two limit deviation conditions, the maximum altitude deviations of ADSC+DDPG are reduced by 28 m and 5023 m, respectively, compared with the ADSC process when jumping out x according to the terminal position; the terminal velocity deviation is reduced by 16.2 m/s and 101.1 m/s, respectively. This result proves the effectiveness of the proposed guidance method.

6 CONCLUSIONS

In this paper, the basic framework of the guidance system is established by using the idea of path tracking and ADSC theory to address the needs of high-precision position tracking and terminal velocity deviation minimization for long-range hypersonic vehicles, converting the effects of earth rotation and curvature into system modeling deviations, and introducing the DDPG reinforcement learning method to generate the optimal adjustment network of ADSC control parameters under perturbation conditions for online generation of guidance commands. Finally, the effectiveness of the proposed method is verified by simulation.

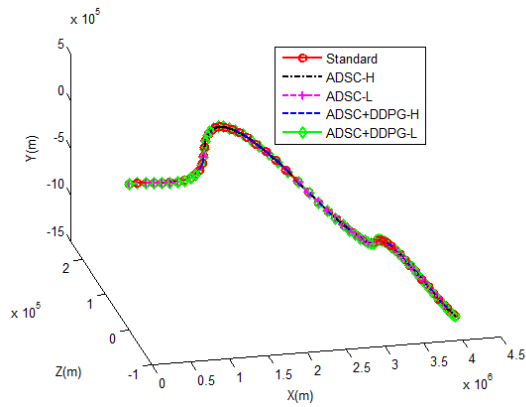


Figure 3: 3D trajectory versus time.

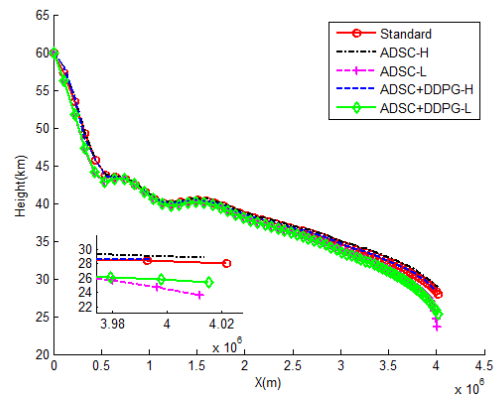


Figure 4: Height versus time.

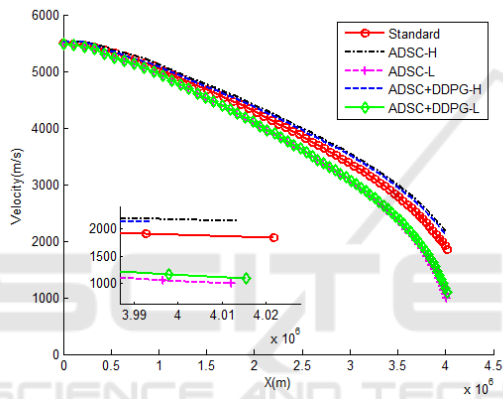


Figure 5: Velocity versus time.

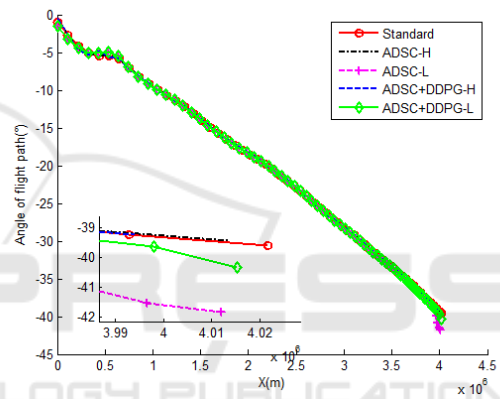


Figure 6: Flight path angle versus time.

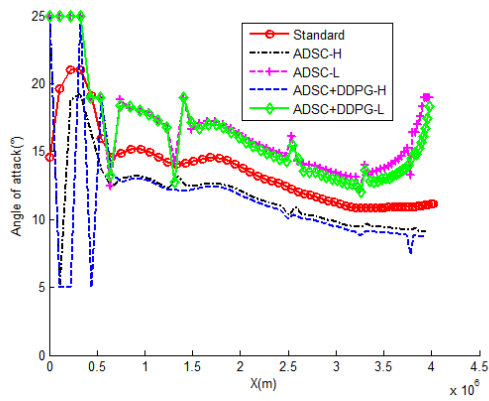


Figure 7: Attack angle versus time.

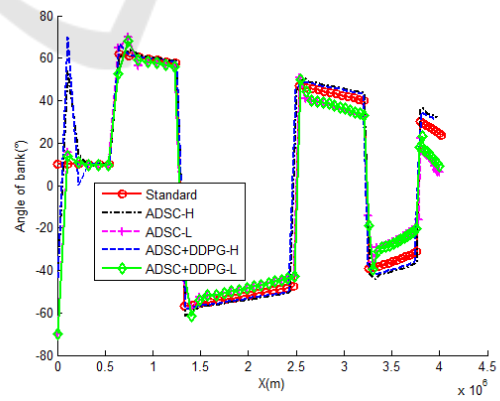


Figure 8: Bank angle versus time.

ACKNOWLEDGEMENTS

The authors would like to thank the financial supports of the open Fund of National Defense Key Discipline Laboratory of Micro-Spacecraft Technology (Grant No. HIT.KLOF.MST.2018028).

REFERENCES

- Xu B., Sun F., Wang S., et al. (2011). Adaptive hypersonic flight control via back-stepping and Kriging estimation. In *Proc. 2011 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1603-1608. IEEE.
- Swaroop D., Gerdes J. C., Yip P. P., et al. (1997). Dynamic surface control of nonlinear systems. In *Proc. The 1997 American Control Conference*, pages 3028-3034. IEEE.
- Liu X. D., Huang W. W., Yu C. M. (2015). Dynamic surface attitude control for hypersonic vehicle containing extended state observer. *Journal of Astronautics*, 36(8):916-922.
- Xu B., Sun F., Wang S., et al. (2014). Dynamic surface control of hypersonic aircraft with parameter estimation. *Advances in Intelligent Systems and Computing*, 213:667-677.
- Wu X., Zheng W., Zhou X., et al. (2021). Adaptive dynamic surface and sliding mode tracking control for uncertain QUAV with time-varying load and appointed-time prescribed performance. *Journal of the Franklin Institute*, 358(8):4178-4208.
- Butt W. A., Yan L., Kendrick A. S. (2010). Dynamic surface control for nonlinear hypersonic air vehicle using neural network. In *Proc. The 29th Chinese Control Conference*, pages 733-738. IEEE.
- Butt W. A., Amezquita, et al. (2013). Adaptive integral dynamic surface control of a hypersonic flight vehicle. *International Journal of Systems Science: The Theory and Practice of Mathematical Modelling, Simulation, Optimization and Control in Relation to Biological, Economic, Industrial and Transportation Systems*, 46(9/12):1717-1728.
- Yu J., Chen J., Wang C., et al. (2014). Near space hypersonic unmanned aerial vehicle dynamic surface backstepping control design. *Sensors & Transducers Journal*, 174(7):292-297.
- Xu B., Zhang Q., Pan Y. (2016). Neural network based dynamic surface control of hypersonic flight dynamics using small-gain theorem. *Neurocomputing*, 173(JAN.15PT.3):690-699.
- Shin J. (2017). Adaptive dynamic surface control for a hypersonic aircraft using neural networks. *IEEE Transactions on Aerospace & Electronic Systems*, 53(5):2277-2289.
- Hu C. F., Liu Y. W. (2013). Fuzzy adaptive nonlinear control of hypersonic vehicles based on dynamic surfaces. *Control and Decision Making*, 28(12):1849-1854.
- Aguiar A. P., Hespanha J. P. (2007). Trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty. *IEEE Transactions on Automatic Control*, 52(8):1362-1379.
- Li R. F., Hu L., Cai L. (2018). Adaptive tracking control of a hypersonic flight aircraft using neural networks with reinforcement synthesis. *Aero Weaponry*, 2018(6):3-10.
- Zhen Y., Hao M. R. (2019). Research on intelligent PID control method based on deep reinforcement learning. *Tactical missile technology*, 2019 (5):37-43.
- Lillicrap T. P., Hunt J. J., Pritzel A., et al. (2015). Continuous control with deep reinforcement learning. *arXiv e-prints*, arXiv:1509.02971.
- Cheng Y., Shui Z. S., Xu C., et al. (2019). Cross-cycle iterative unmanned aerial vehicle reentry guidance based on reinforcement learning. In *Proc. 2019 IEEE International Conference on Unmanned Systems*, pages 587-592.
- Gao J. S. (2019). Research on trajectory optimization and guidance method of lift reentry vehicle. Doctoral Dissertation, Huazhong University of science and technology.
- Song C., Zhou J., Guo J. G., et al. (2016). Guidance method based on path tracking for hypersonic vehicles. *Acta Astronautica*, 37(4): 435-441.
- Uhlenbeck, George E. and Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical Review*, 36(5):823.
- Kingma D., Ba J. (2014). Adam: a method for stochastic optimization. *arXiv preprint*, arXiv:1412.6980.