mathematics

MDPI

Article

Fault-Tolerant Integrated Guidance and Control Design for Hypersonic Vehicle Based on PPO

Jia Song [1], Yuxie Luo [1,*], Mingfei Zhao [1], Yunlong Hu [1] and Yanxue Zhang [2]

[1] School of Astronautics, Beihang University, Beijing 100191, China
[2] Science and Technology on Space Physics Laboratory, Beijing 100076, China
[*] Correspondence: luoyuxie@buaa.edu.cn

Abstract: Aiming at the problem of the terminal guidance phase of hypersonic vehicles (HSV) under fault condition, and considering the existence of various uncertain parameters and actuator faults in the control system, a fault-tolerant integrated guidance and control design of a hypersonic vehicle based on the proximal policy optimization algorithm (PPO) is proposed. First, in view of the problem that the separate guidance and control loop design cannot make full use of the coupling relationship between the two, the relationship between the guidance loop and the control loop is considered and an integrated guidance and control system of HSV is established. Then, the integrated guidance and control problem is converted into a reinforcement learning model, the action space, state observation space, and reward function of the PPO agent are designed, and the network is initialized and designed. Simulations verify the feasibility of the proposed PPO-based IGC system.

Keywords: hypersonic vehicle; integrated guidance and control; proximal policy optimization; fault-tolerant control

MSC: 93-10

check for updates

Citation: Song, J.; Luo, Y.; Zhao, M.; Hu, Y. Fault-Tolerant Integrated Guidance and Control Design for Hypersonic Vehicle Based on PPO. *Mathematics* **2022**, *10*, 3401. https://doi.org/10.3390/math10183401

Academic Editor: António Lopes

Received: 3 August 2022
Accepted: 15 September 2022
Published: 19 September 2022

1. Introduction

Hypersonic vehicles (HSV) generally refer to aircrafts with a flying Mach number greater than 5 [1]. However, the cost of hypersonic vehicles is high, and they have the characteristics of strong nonlinearity, strong coupling, and severe uncertainty of aerodynamic parameters during the re-entry flight. At the same time, taking into account the possible fault of the actuators affected by the environment, the reliability and security of hypersonic vehicle flight is a problem that cannot be ignored [2–6]. To improve the guidance accuracy and flight stability of HSV, designing a fault-tolerant guidance and control system becomes a priority.

At present, the traditional hypersonic vehicle guidance and control system design is usually decoupled into two loops, the guidance loop and the attitude control loop, and these two loops are designed separately. However, there is a strong coupling relationship between the guidance loop and the attitude control loop. Separate designs cannot make full use of the synergistic relationship between the two, and the uncertainty and fault factors cannot be comprehensively considered in a single loop. Therefore, a design method of Integrated Guidance and Control (IGC) has been proposed [7–9]. The method of using the comprehensive information such as attitude, overload, and line-of-sight angular velocity in the feedback control design of actuator action to improve the final guidance quality is usually called the IGC method [8]. This method designs the overall system of the guidance loop and the attitude control loop, which can take into account the environmental information and feedback information in the two types of loops and improve the fault-tolerance capability and final guidance quality of HSV.

At present, many scholars have conducted a lot of research on the algorithm of integrated guidance and control, such as active disturbance rejection control [10,11], sliding mode control [12,13], adaptive control [14], and so on. For hypersonic vehicles, Chong proposed a finite-time IGC method for HSV with improved robustness-to-parameter uncertainty [15]. Reference [16] proposed an IGC method based on the L1 adaptive state feedback control to solve the problem of dynamic uncertainty in hypersonic vehicles. Zhang proposed a multi-constraints finite-time IGC method based on adaptive control [14]. Reference [17] designed a fault-tolerant IGC system under rudder surface faults based on the predictive correction and backstepping control method. Most of the current research on integrated guidance and control systems of hypersonic vehicles only considers uncertainty conditions or actuator fault, and there is little research on the design of IGC systems under complex fault conditions. Under the condition of considering both uncertainty and actuator fault, higher requirements are put forward for the fault-tolerance and real-time performance of IGC systems.

With the continuous development of intelligent algorithms, their combination with traditional control algorithms has become a new mainstream research direction. Reinforcement learning is a machine learning method that is widely used in various fields. It outputs actions according to the observation of the environment by the agent and updates the action output policy according to the reward generated by the action. Reinforcement learning is an algorithm that can find suitable real-time action policies and is often used to solve decision-making problems [18,19]. Since RL is a real-time algorithm and can learn policies according to environmental feedback, it has environmental adaptability and can be used in fault-tolerant control research. Reference [20] applied reinforcement learning to the control of fuel-transfer systems without prior information on faults and solved the problem of gradual fault tolerance. Reference [21] uses Meta-RL to solve slip-steering vehicle control under actuator failure. Reference [22] compared the fault tolerance of MPC and RL in the presence of sensor noise and slowly changing faults, and the result shows that reinforcement learning has better fault tolerance. In summary, due to the real-time characteristics of reinforcement learning, it is feasible to apply it to fault-tolerant control. Proximal Policy Optimization (PPO) is an advanced RL algorithm with the general advantages of RL and more stable training [23], it can be applied to design the IGC system of hypersonic vehicles under the condition of compound faults.

Due to the possible fault problems of hypersonic vehicles in the terminal guidance phase and the high requirements for guidance accuracy, and aiming at the fault-tolerant control of terminal guidance of hypersonic vehicles, the main contributions of this paper can be summarized as follows:

- A nonlinear IGC model of a hypersonic vehicle is established with actuator faults and parameter uncertainty.
- A PPO-based IGC system is proposed and designed. The IGC system is modeled as a reinforcement learning problem and the PPO algorithm is applied to the system. The simulation proves the passive fault tolerance of the method, and the method can complete the guidance task under complex fault conditions.

2. Problem Formulation

2.1. 3DOF Model of the Hypersonic Vehicle

This paper mainly studies the terminal guidance problem of hypersonic vehicles. Based on the Winged-Cone model with detailed relevant information published by NASA [24], it is simplified to a 3DOF vertical plane model for research. During the terminal guidance phase, HSV is only affected by gravity and aerodynamic force [3]. The aerodynamic rudder of HSV is simplified to a pitch elevator in the vertical plane δ_z. The actions of the pitch elevator will change the aerodynamic force and aerodynamic torque acting on HSV, thereby changing the flight attitude of HSV.

Assuming that the earth is a non-rotating spherical model, the launch coordinate system of HSV is an inertial coordinate system, as shown in Figure 1.

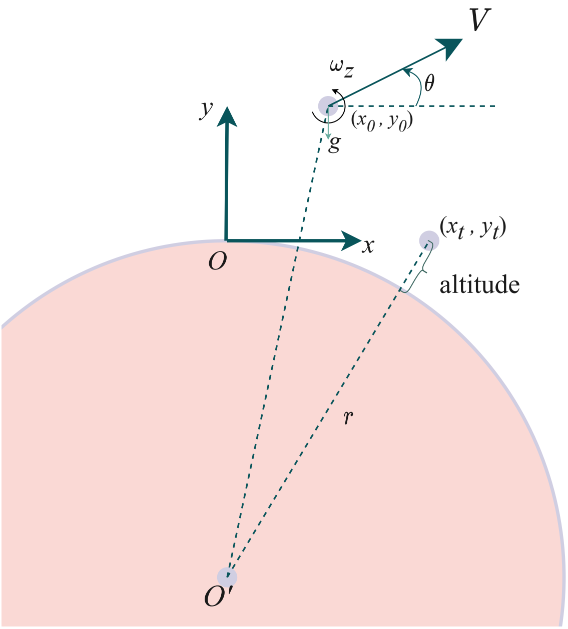

Figure 1. HSV motion diagram.

The equation of motion established in this system is:

$$\begin{cases} \dot{x} = V\cos\theta \\ \dot{y} = V\sin\theta \\ \dot{V} = -\dfrac{D}{m} - g\sin\theta \\ \dot{\omega}_z = \dfrac{M_z}{I_{zz}} \\ \dot{\theta} = \dfrac{L}{mV} - \dfrac{g\cos\theta}{V} \\ \dot{\alpha} = \omega_z - \dot{\theta} \end{cases} \tag{1}$$

where x and y represent the coordinates of HSV in the launch coordinate system; V represents the flight speed of HSV; g is the local gravitational acceleration of HSV; D and L represent the drag and lift caused by aerodynamic force; M_z represents the aerodynamic pitching moment; ω_z is the pitch angular velocity of HSV; I_{zz} is the moment of inertia of the Z axis of HSV; θ represents the flight trajectory inclination of HSV; and α represents the angles of attack. The calculation of aerodynamic force and aerodynamic moment can be obtained by (2):

$$\begin{cases} L = C_L qS \\ D = C_D qS \\ M_z = C_{mz} qS \end{cases} \tag{2}$$

In the Equation (2), $q = 0.5\rho V^2$ represents the dynamic pressure, where ρ is the atmospheric density. C_D, C_L, and C_{mz} represent the aerodynamic coefficients of lift, drag, and pitching moment. They are generally fitted as a function of angle of attack α, Mach number, and elevator δ_z, expressed as (3):

$$\begin{cases} C_L = C_{L,\alpha} + C_{L,\delta_z} \\ C_D = C_{D,\alpha} + C_{D,\delta_z} \\ C_{mz} = C_{mz,Ma} + C_{mz,\alpha} + C_{mz,\delta_z} + C_{mz,q}\frac{\omega_z c}{2V} \end{cases} \tag{3}$$

2.2. Line-of-Sight Angle Model

In this paper, the line-of-sight angle model is used as the relative motion model. It is assumed that the target line-of-sight vector of HSV is expressed as $\vec{R}$ in the launch coordinate system, and its included angle with the x-axis of the transmission system is q, as shown in Figure 2:

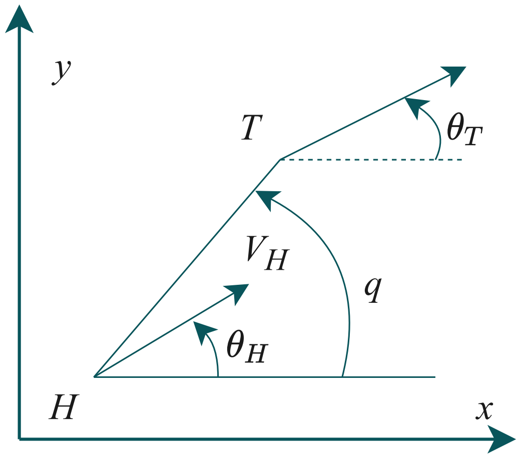

Figure 2. Line-of-Sight-Angle.

The relative motion equation of HSV in the longitudinal plane can be obtained as Equation (4):

$$\begin{cases} \|\dot{d}\| = V_T\cos(q-\theta_T) - V_H\cos(q-\theta_H) \\ d\dot{q} = -V_T\sin(q-\theta_T) - V_H\sin(q-\theta_H) \end{cases} \tag{4}$$

where d represents the relative distance; q represents the line-of-sight angle; V_T and θ_T represent the target's flight speed and flight trajectory inclination; and V_H and θ_H represent the HSV's flight speed and flight trajectory inclination.

2.3. Fault Model

The actuator of HSV works in the air flow and is easily affected by disturbances such as gusts and atmospheric turbulence. Actuators suffer from loss of effectiveness and jams. The actuator fault model is shown in (5) [3].

$$\delta_{zf} = \delta_z - E\delta_z + \bar{\delta_z} \tag{5}$$

where δ_{zf} represents the fault outputs; E represents the fault indicator; and $\bar{\delta_z}$ represents the bias fault.

3. Method

3.1. Proximal Policy Optimization Algorithm

The concept of reinforcement learning originated in the 1950s, and since Deepmind proposed deep reinforcement learning in 2013, reinforcement learning has developed rapidly. A reinforcement learning process can be described as a Markov decision process, $G = \{s, a, p, r, \gamma\}$. The basic process of the reinforcement learning algorithm is: the agent obtains the state observation s_t of the environment and outputs the action a_t. The environment is then transformed into the next state s_{t+1} by the state transition function p_t. The environment gives the agent a feedback reward r_{t+1} according to the state s_{t+1} after the environment changes; the process is shown in Figure 3. The agent continues this process until it reaches the condition to end a single training episode.

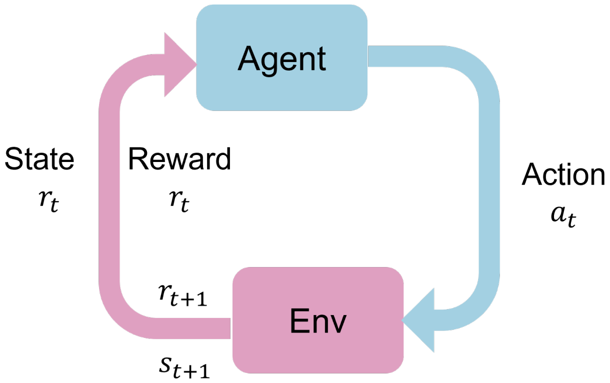

Figure 3. Markov decision process.

The purpose of RL training is to obtain the greatest action output policy function $\pi : s_t \rightarrow p(a_t|s_t)$, which represents the probability map of the observed action output from the environmental state. The policy function is generally parameterized as π_θ. The optimal policy requires that after the agent performs the current action, the total rewards obtained in the future reaches the maximum, which is defined as (6):

$$V^{\pi_\theta}(s) = \mathbb{E}_{\pi_\theta} \sum_t [R(s_t, a_t) \mid s], R(s_t, a_t) = \sum_t (\gamma^{t-1} r_t) \tag{6}$$

where $\gamma \in [0, 1]$ represents the discount factor, which determines whether the agent attaches to future rewards. In the policy gradient algorithm [25], in order to improve the learning efficiency and make learning more stable, the advantage function is introduced, as shown in (8):

$$Q^\pi(s, a) = \sum_t \mathbb{E}_{\pi_\theta}[R(s_t, a_t) \mid s, a] \tag{7}$$

$$A^\pi(s, a) = Q^\pi(s, a) - V^\pi(s) \tag{8}$$

the objective function can be defined as (9), the policy parameter θ is updated by the gradient of the objective function, as shown in (10).

$$J(\theta) = \hat{\mathbb{E}}_t\left[\log \pi_\theta(a_t \mid s_t) \hat{A}_t\right] \tag{9}$$

$$\theta_{k+1} = \theta_k + \alpha \nabla_\theta \mathbb{E} J(\theta) \tag{10}$$

The policy gradient algorithm has the problems of low sampling efficiency and unstable training. Based on the traditional policy gradient method, Schulman proposed the PPO algorithm [23]. PPO is an actor–critic algorithm. Actor refers to the policy function $\pi_\theta(a|s)$, and critic refers to the value function $V^\pi(s)$. They are both parameterized neural networks and are updated during training. The PPO algorithm introduces importance sampling and clip function in the update, its objective function is shown in (11).

$$L(\theta) = \hat{\mathbb{E}}_t\left[\min\left(\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)} \hat{A}_t, \text{clip}\left(\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}, 1 - \epsilon, 1 + \epsilon\right) \hat{A}_t\right)\right] \tag{11}$$

Importance sampling enables the PPO algorithm to reuse the sampled data when updating, and the sampling network and the actual policy network update parameters at the same time, thereby speeding up the training speed. The clip function makes the PPO algorithm easier to converge and improves the performance of the algorithm.

3.2. PPO-Based IGC System

Firstly, the original IGC problem is modeled as a reinforcement learning problem. For the IGC problem studied in this paper, the agent is the IGC system of HSV. The overall structure is shown in Figure 4.

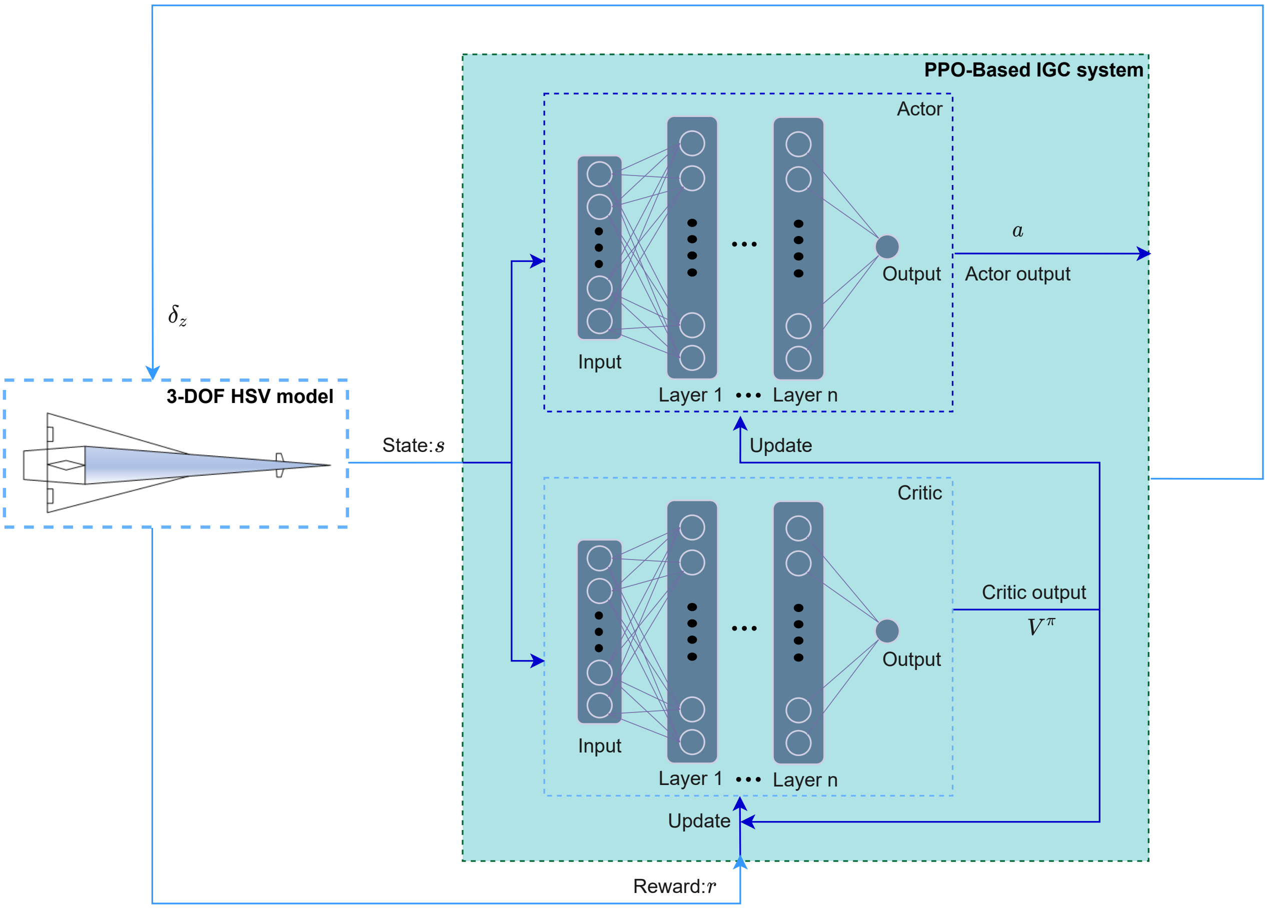

Figure 4. PPO-Based IGC system.

The input of the PPO-based system is the state space s, and the output is the action space a. The system will output actions according to the input; the process can be seen in Section 3.1. Next, the state, action, and reward functions in the system are specifically designed.

3.2.1. Action Space Design

Since HSV is only affected by aerodynamic force and gravity, the action of the agent is designed as the elevator offset of HSV, namely δ_z. Its range is $-30^{\circ} \leq \delta_z \leq 30^{\circ}$.

3.2.2. Observation Space Design

Considering the fault-tolerant performance of the algorithm, the state observation should include as many flight states as possible as feedback input. In this paper, the state observation space is designed as a collection of position, speed, angular velocity, angle of attack, and relative motion model $S = \{x, y, V, \theta, \omega_z, \alpha, R, q\}$. All state observations need to be normalized and input into the neural network.

3.2.3. Reward Function Design

The design of reward function is an important part in the PPO algorithm. The simplest design method is to give a reward when HSV reach the target. However, this reward design is too sparse, and HSV cannot obtain enough and effective rewards, resulting in slow learning or even ineffective learning. Therefore, it is necessary to design the reward function for the intermediate state variables to ensure the learning of the agent. Aiming at the problem of terminal guidance IGC system design studied in this paper, the reward functions of relative distance, relative speed, sight angle, offline distance, and elevator are designed, namely $R = f(d, \dot{d}, \dot{q}, Z, \delta_z)$ can be expressed as (12).

$$\begin{cases} R_d = \left(1 - \frac{|d|}{|d_0|}\right) \\ R_{\dot{d}} = \beta_{\dot{d}} \left(\frac{\overrightarrow{\dot{d}} \cdot \overrightarrow{d}}{\left|\overrightarrow{\dot{d}}\right| \times \left|\overrightarrow{d}\right|} \right) \\ R_q = -\dot{q} \\ R_{\delta_z} = -\frac{|\Delta\delta_z|}{|\Delta\delta_{z,max}|} \\ R_Z = -\frac{Z}{Z_0} \end{cases} \tag{12}$$

R_d gives HSV the reward on the relative distance, and d_0 represents the initial relative distance. The smaller the relative distance, the greater the reward value, which belongs to the direct reward function based on the terminal guidance task; $R_{\dot{d}}$ encourages HSV to move in the direction of decreasing relative distance. When the speed direction is the same as the relative position direction, the reward is the largest, and when the two directions are opposite, the reward is the smallest; R_q encourage HSV to move towards the line-of-sight angular rate of 0; and R_{δ_z} prevents the elevator of HSV from frequently swinging in a large range and is designed with the elevator differential term. When the swing is larger, the negative reward will be larger. R_Z gives HSV rewards according to the amount of zero effort missed, when the value is larger, the negative reward is larger. The calculation of the Z is (13):

$$Z = \frac{R^2|\dot{q}|}{\sqrt{\dot{R}^2 + R^2\dot{q}^2}} \tag{13}$$

The overall reward function is designed as (14)

$$R = \beta_d R_d + \beta_{\dot{d}} R_{\dot{d}} + \beta_q R_q + \beta_{\delta_z} R_{\delta_z} + \beta_Z R_Z \tag{14}$$

where β represents the weight of each reward, which is set before training.

3.2.4. Network Initialization

According to the PPO algorithm, a four-layer fully connected neural network is used to represent the evaluation network, as shown in Table 1:

Table 1. Critic Network.

Layers	**Size**
Input Layer	8
Hidden Layer 1	128
Hidden Layer 2	64
Hidden Layer 3	32
Output Layer	1

The network input is state space s, and the output is the estimated reward. According to the PPO algorithm, a four-layer fully connected neural network is used to represent the evaluation network, as shown in the following table: A 5-layer fully connected neural network is used to represent the execution network, as shown in Table 2:

The network input is state space s, and the output is a Gaussian distribution of action a.

Table 2. Actor network.

Layers	Size
Input Layer	8
Hidden Layer 1	128
Hidden Layer 2	64
Hidden Layer 3	32 (variance) + 32 (mean)
Output Layer	1 (variance) + 1 (mean)

The above networks together constitute the PPO-based IGC system. Through training in the environment, a network with an optimal strategy is finally obtained, thereby realizing the integrated guidance and control of hypersonic vehicles. The settings of some agent parameters are shown in Table 3:

Table 3. Hyperparameters.

Parameters	Value
Horizon	16 384
Clip Factor	0.2
Discount Factor	0.9999
Mini Batch Size	128
Sample Time	0.01
Learn Rate	1×10^{-4}

4. Simulation and Results

In this section, the performance of the proposed PPO-based IGC system is verified and validated by simulation under the conditions of no faults and actuator faults, and an IGC method based on active disturbance rejection control (ADRC) is adopted for comparison with a validation of the effectiveness and superiority of the proposed algorithm in dealing with actuator fault problems. Assume the target is stationary, the initial conditions are shown in Table 4.

Table 4. Initial Conditions.

Parameters	Value	Parameters	Value
x_0 (m)	0	y_0 (m)	30,000
V_0 (m/s)	4500	θ_0 (°)	−5
ω_{z0} (rad/s)	0	α_0 (°)	−10
x_{target} (m)	150,000	y_{target} (m)	10,000

4.1. Simulation without Actuator Faults

First, the simulation is performed when no fault occurs. The flight trajectories of HSV are shown in Figure 5.

Under the condition of no fault, both the PPO and ADRC algorithms can complete the guidance task well. The miss distance of HSV was 0.7 m and 1 m, respectively, and PPO was slightly better than ADRC The speed change in HSV during flight is shown in the Figure 6.

Since there is no thrust, HSV glide without power, their speed is continuously reduced, and the Mach numbers of PPO and ADRC when they land are 2.35 and 4.41, respectively. The attitude angle changes in HSV during flight are shown in the Figure 7. The angle of attack of PPO and ADRC both change sharply in the early stage of flight. The reason is that the atmosphere is thin and the aerodynamic control ability is weak in the early stage of flight. After the altitude drops, the aircraft gradually tends to be stable. The line-of-sight angle curve of PPO changes less than that of ADRC. The line-of-sight angle will change significantly at the end, because the aircraft has already reached the target. The elevator deflection curve of HSV during flight is shown in Figure 8.

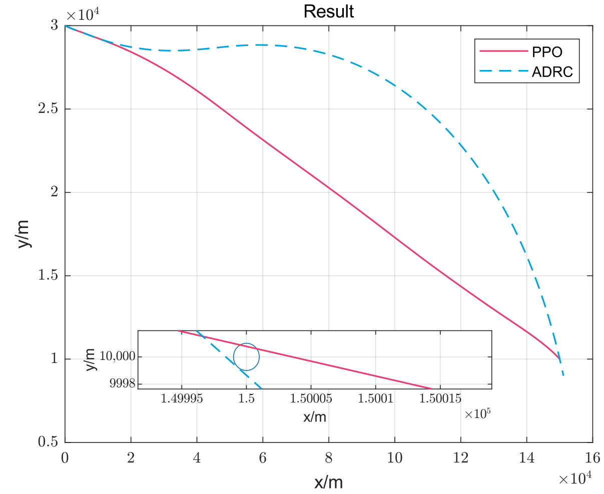

Figure 5. Trajectory when no fault occurs.

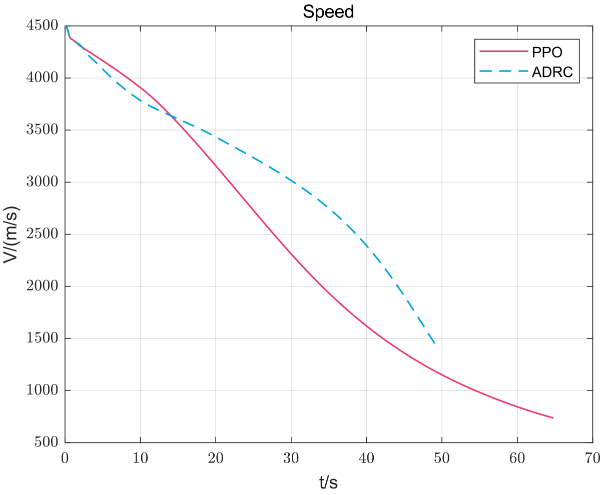

Figure 6. Speed curves when no fault occurs.

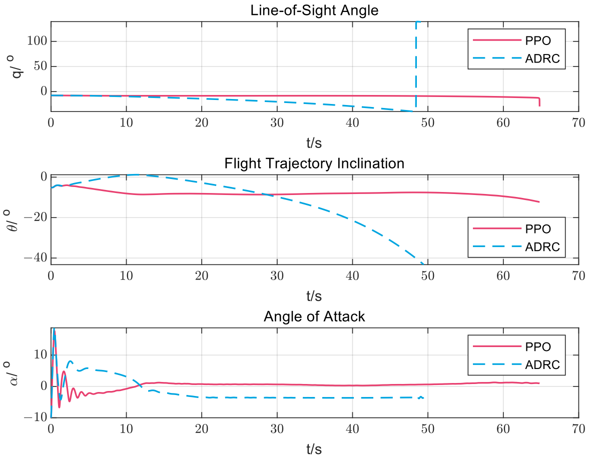

Figure 7. Angle curves when no fault occurs.

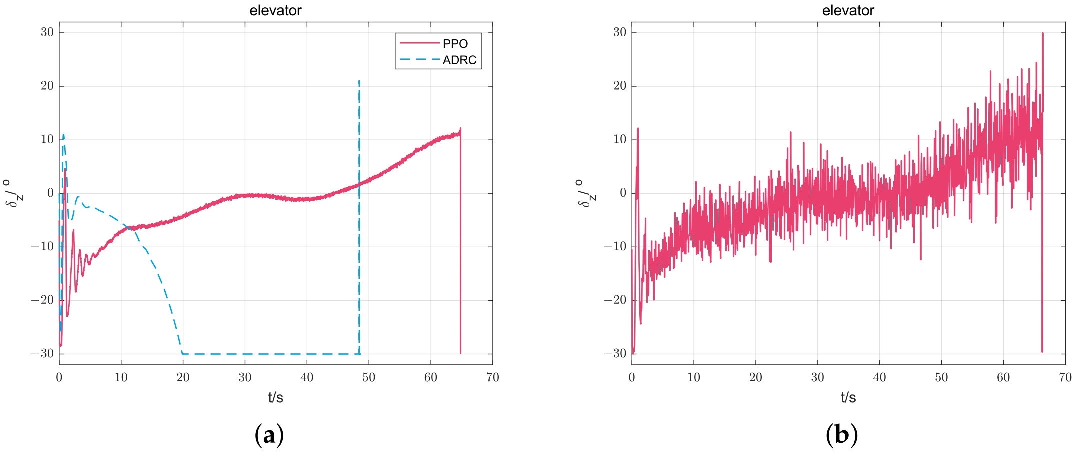

Figure 8. Elevator deflection when no fault occurs. (**a**) elevator deflection, (**b**) elevator deflection without R_{δ_z}.

From the comparison result of Figure 8a,b, it can be seen that in the absence of R_{δ_z}, the elevator deflection oscillates more violently, but when R_{δ_z} is added, the elevator deflection almost does not oscillate.

Based on the above results, it can be seen that HSV will oscillate briefly in the first 10 s. At this phase, HSV will adjust their attitude and then enter the stable flight phase until the guidance task is completed. In the absence of faults, both the PPO-based IGC and ADRC-based IGC systems can perform the guidance tasks well.

4.2. Simulation with Actuator Faults and Uncertainty

According to Section 2.3, it is assumed that HSV's fault parameter $E = 0.2$, $\bar{\delta_z} \in [-2, 2]$ is a random number. Moreover, 20% deflection is added to the aerodynamic coefficient C_L, C_D, C_z. Figure 9 is the curve of the actual elevator deflection, and the expected elevator deflection when the fault occurs, as well as the curve of the aerodynamic coefficient.

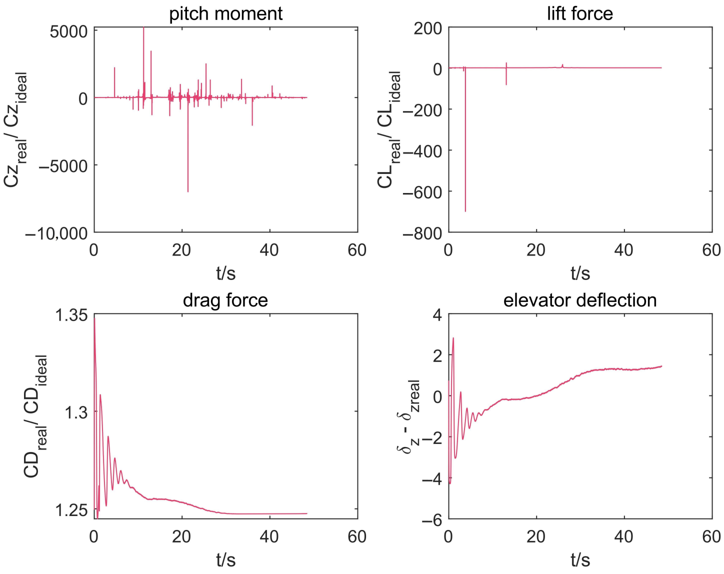

Figure 9. The effect on aerodynamics.

The pitching moment and lift are greatly affected, and the problem of moment and lift reversal will occur at certain times. The impact on resistance is relatively small, but there is also an increase of 25~35%. The elevator deflection difference is between −4 and 3, and the final difference is roughly around 1. It can be seen that the influence on HSV is relatively large. The flight curves of HSV are shown in Figure 10.

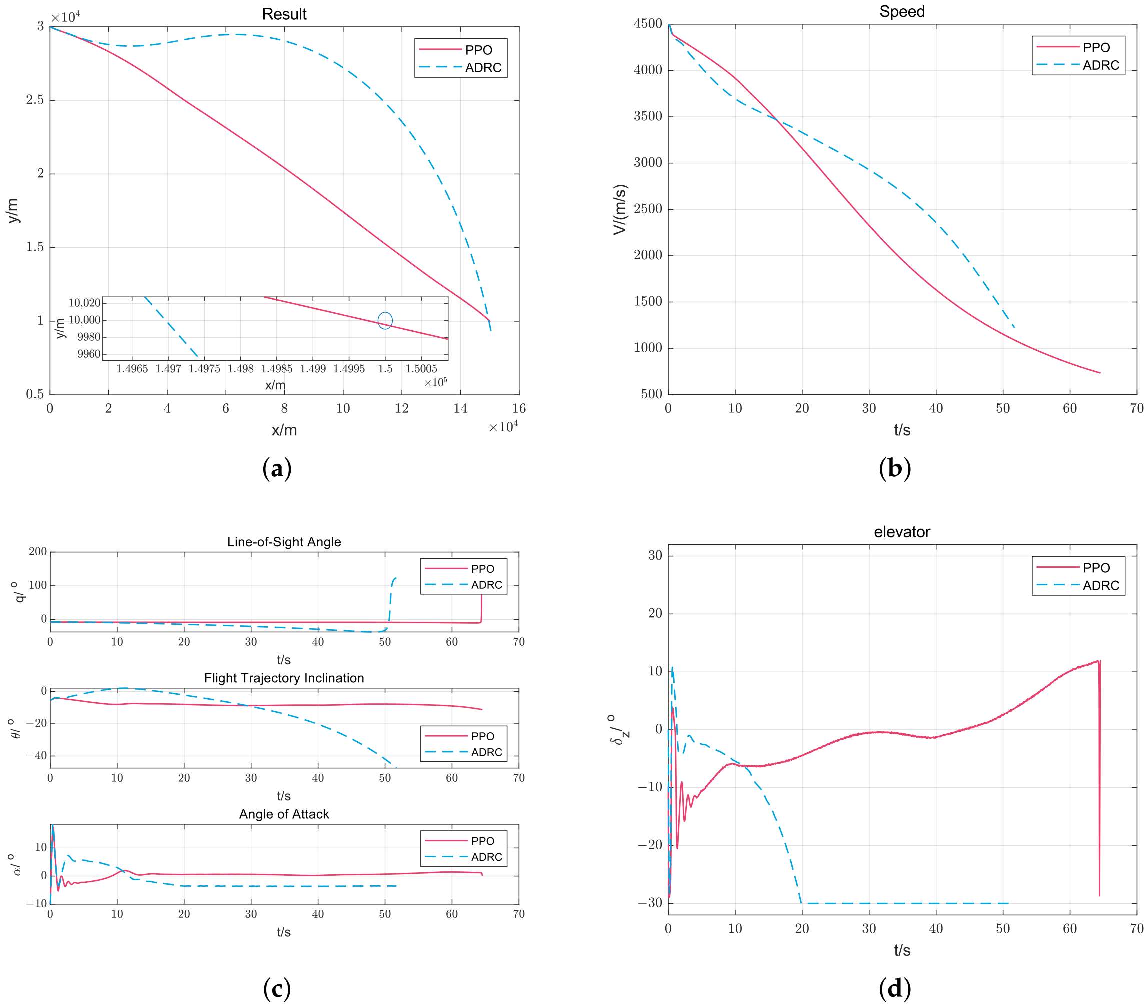

Figure 10. Simulation with fault and uncertainty. (**a**) Flight Trajectory. (**b**) Speed curve. (**c**) Angle Curve. (**d**) Elevator Deflection.

The trajectories of HSV are shown in Figure 10a. The miss distance of HSV in the PPO algorithm is 5.1 m, while that of ADRC is 217 m. This is still an acceptable range of misses considering the long range, and both of them can be regarded as completing the guidance task, but it is obvious that PPO is less affected by actuator faults The speed changes of HSV during flight are shown in Figure 10b. Under fault conditions, the speed change is still continuously reduced. The attitude angle changes of HSV during flight are shown in Figure 10c. The angle change trend of the two methods is similar. After the first few seconds of oscillation, the attitude angle gradually stabilizes. The elevator deflection curves of HSV during flight are shown in Figure 10d. From the comparison results, it can be seen that although the two methods can complete the guidance task well under no fault conditions, the PPO-based IGC has better results under fault conditions and stronger fault tolerance than the ADRC-based IGC. It can be seen that the PPO-based IGC proposed in this paper can cope with the fault conditions well and can still complete the guidance task when the fault occurs.

5. Conclusions

Based on the PPO algorithm, this paper studies the design of the fault-tolerant guidance and control system in the terminal guidance phase of hypersonic vehicles. Considering uncertain parameters and actuator faults, a 3DOF motion model of HSV in the longitudinal plane is established. For this model, the IGC system is modeled as a reinforcement learning process, and the PPO-based IGC system is designed. According to the terminal guidance task requirements, the action space, state observation space, and reward function of HSV are designed, respectively, and the IGC system is trained by the PPO algorithm. Finally, we carried out the simulation experiments of PPO-based IGC and ADRC-based IGC under the conditions of actuator faults and without faults, respectively. The simulation verification shows that the PPO-based IGC system in this paper can complete the terminal guidance task under no fault conditions and under fault conditions with small miss distance and does not rely on fault prior information, which verifies its effectiveness and robustness.

Author Contributions: Conceptualization, J.S. and Y.L.; methodology, J.S. and M.Z.; software, J.S. and Y.L.; validation, J.S., Y.H. and Y.Z.; formal analysis, J.S. and Y.H.; investigation, J.S. and Y.L.; resources, J.S. and Y.Z.; data curation, Y.L. and M.Z.; writing—original draft preparation, J.S.; writing—review and editing, J.S. and Y.L.; visualization, J.S.; supervision, J.S.; project administration, J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Laboratory of Hunan Province for Control Technology of Distributed Electric Propulsion Air Vehicle OF FUNDER grant number 62073020.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank their colleagues for their constructive suggestions and research assistance throughout this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IGC	Integrated Guidance and Control
PPO	Proximal Policy Optimization
RL	Reinforcement Learning
HSV	Hypersonic Vehicles
DOF	Degree of Freedom

References

1. Xu, B.; Shi, Z. An overview on flight dynamics and control approaches for hypersonic vehicles. *Sci. China Inf. Sci.* **2015**, *58*, 1–19. [CrossRef]
2. Shao, X.; Shi, Y.; Zhang, W. Fault-Tolerant Quantized Control for Flexible Air-Breathing Hypersonic Vehicles With Appointed-Time Tracking Performances. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *57*, 1261–1273. [CrossRef]
3. Gao, K.; Song, J.; Wang, X.; Li, H. Fractional-order proportional-integral-derivative linear active disturbance rejection control design and parameter optimization for hypersonic vehicles with actuator faults. *Tsinghua Sci. Technol.* **2020**, *26*, 9–23. [CrossRef]
4. Zhao, K.; Song, J.; Ai, S.; Xu, X.; Liu, Y. Active Fault-Tolerant Control for Near-Space Hypersonic Vehicles. *Aerospace* **2022**, *9*, 237. [CrossRef]
5. Meng, Y.; Jiang, B.; Qi, R. Adaptive fault-tolerant attitude tracking control of hypersonic vehicle subject to unexpected centroid-shift and state constraints. *Aerosp. Sci. Technol.* **2019**, *95*, 105515. [CrossRef]
6. Bu, X.; He, G.; Wang, K. Tracking control of air-breathing hypersonic vehicles with non-affine dynamics via improved neural back-stepping design. *ISA Trans.* **2018**, *75*, 88–100. [CrossRef] [PubMed]
7. Williams, D.; Richman, J.; Friedland, B. Design of an integrated strapdown guidance and control system for a tactical missile. In Proceedings of the Guidance and Control Conference, Gatlinburg, TN, USA, 15–17 August 1983; p. 2169.
8. Santoso, F.; Garratt, M.A.; Anavatti, S.G. State-of-the-art integrated guidance and control systems in unmanned vehicles: A review. *IEEE Syst. J.* **2020**, *15*, 3312–3323. [CrossRef]
9. Levy, M.; Shima, T.; Gutman, S. Integrated single vs. two loop autopilot-guidance design for dual controlled missiles. In Proceedings of the 2013 10th IEEE International Conference on Control and Automation (ICCA), Hangzhou, China, 12–14 June 2013; pp. 1730–1735.

10. Zhang, C.; Wu, Y.J. Dynamic surface control and active disturbance rejection control-based integrated guidance and control design and simulation for hypersonic reentry missile. *Int. J. Model. Simul. Sci. Comput.* **2016**, *7*, 1650025. [CrossRef]
11. Yi, K.; Tang, K.; She, S.; Li, M.; Tan, Z. Integrated guidance and control for Semi-Strapdown Missiles. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 9826–9830.
12. Ming, C.; Wang, X.; Sun, R. A novel non-singular terminal sliding mode control-based integrated missile guidance and control with impact angle constraint. *Aerosp. Sci. Technol.* **2019**, *94*, 105368. [CrossRef]
13. Xiong, Y.; Guo, J.; Zhou, J. Sliding Mode Control for Integrated Missile Guidance and Control System. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 8371–8374.
14. Zhang, D.; Ma, P.; Wang, S.; Chao, T. Multi-constraints adaptive finite-time integrated guidance and control design. *Aerosp. Sci. Technol.* **2020**, *107*, 106334. [CrossRef]
15. Chong, Z.; Guo, J.; Zhao, B.; Guo, Z.; Lu, X. Finite-time integrated guidance and control system for hypersonic vehicles. *Trans. Inst. Meas. Control* **2021**, *43*, 842–853.
16. Khankalantary, S.; Rezaee Ahvanouee, H.; Mohammadkhani, H. L1 Adaptive integrated guidance and control for flexible hypersonic flight vehicle in the presence of dynamic uncertainties. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **2021**, *235*, 1521–1531. [CrossRef]
17. Qian, J.; Qi, R.; Jiang, B. Fault-tolerant guidance and control design for reentry hypersonic flight vehicles based on control-allocation approach. In Proceedings of the 2014 IEEE Chinese Guidance, Navigation and Control Conference, Yantai, China, 8–10 August 2014; pp. 1624–1629.
18. Li, Y. Deep reinforcement learning: An overview. *arXiv* **2017**, arXiv:1701.07274.
19. François-Lavet, V.; Henderson, P.; Islam, R.; Bellemare, M.G.; Pineau, J. An introduction to deep reinforcement learning. *Found. Trends Mach. Learn.* **2018**, *11*, 219–354. [CrossRef]
20. Ahmed, I.; Quiñones-Grueiro, M.; Biswas, G. Fault-Tolerant Control of Degrading Systems with On-Policy Reinforcement Learning. *IFAC-PapersOnLine* **2020**, *53*, 13733–13738. [CrossRef]
21. Dai, H.; Chen, P.; Yang, H. Metalearning-Based Fault-Tolerant Control for Skid Steering Vehicles under Actuator Fault Conditions. *Sensors* **2022**, *22*, 845. [CrossRef] [PubMed]
22. Ahmed, I.; Khorasgani, H.; Biswas, G. Comparison of model predictive and reinforcement learning methods for fault tolerant control. *IFAC-PapersOnLine* **2018**, *51*, 233–240. [CrossRef]
23. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
24. Keshmiri, S.; Colgren, R.; Mirmirani, M. Development of an aerodynamic database for a generic hypersonic air vehicle. In Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, San Francisco, CA, USA, 15–18 August 2005; p. 6257.
25. Sutton, R.S.; McAllester, D.; Singh, S.; Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. *Adv. Neural Inf. Process. Syst.* **1999**, *12*, 1057–1063.