SCIENCE CHINA
Information Sciences

• LETTER •

July 2024, Vol. 67, Iss. 7, 179202:1–179202:2
https://doi.org/10.1007/s11432-022-4063-x

A unified intelligent control strategy synthesizing multi-constrained guidance and avoidance penetration

Sibo ZHAO[1], Jianwen ZHU[1,2*], Weimin BAO[1,3] & Xiaoping LI[1]

[1] *School of Aerospace Science and Technology, Xidian University, Xi'an* 710126*, China;*
[2] *College of Missile Engineering, Rocket Force University of Engineering, Xi'an* 710025*, China;*
[3] *China Aerospace Science and Technology Corporation, Beijing* 100048*, China*

Received 29 June 2022/Revised 26 January 2023/Accepted 10 May 2024/Published online 14 June 2024

Satisfying various terminal constraints and achieving evasive flight in no-flying-zones (NFZs) are two important flight missions undertaken by high velocity vehicles. We studied an intelligent maneuvering guidance strategy using deep reinforcement learning (DRL) methods to achieve the coordination and unification of evasive flight and multi-constrained guidance with minimal energy loss.

In a previous study [1], optimal control theory was used to solve the pure guidance problem. Adjusting the amplitude or direction of the bank angle is the primary method of avoiding NFZs; however, the terminal deviation caused by maneuvering has not yet been analyzed [2]. Mathematical optimization methods offer certain improvements in solving problems of vehicle trajectory planning [3,4], but these algorithms have the disadvantage of a long online computation time. DRL effectively solves model-free dynamic programming problems by interacting with environmental information, which enables intelligent decision-making. In addition, DRL can effectively avoid the problems of complex modeling and difficult numerical solutions. Soft actor-critic (SAC) networks have significantly improved the practicability of DRL algorithms [5].

To improve the adaptability and penetration performance of gliding guidance in complex missions and changeable environments, this study proposes a unified intelligent control strategy that synthesizes optimal guidance and DRL. Optimal guidance is introduced to satisfy the terminal latitude, longitude, altitude, and velocity constraints, and SAC is adopted to generate a lateral maneuvering command, which thereby achieves the avoidance of NFZs with minimal energy loss and satisfies the terminal constraints. The deep neural network (DNN) parameters are trained offline and utilized online. The main contributions of this study are as follows.

(1) Propose an avoidance strategy with terminal multi-constrained guidance and guarantee less energy loss during maneuvering.

(2) To enhance training efficiency, this study adopts a prediction method that calculates terminal states and adds process rewards to accelerate training.

(3) To enhance the adaptability of the avoided guidance algorithm in dynamically complex environments, flight condition deviation is considered in the training process.

Analysis of flight constraints. In the gliding flight phase, each NFZ is considered to be a cylindrical zone. The vehicle is guided to the terminal target under the constraints of NFZs to satisfy the terminal altitude h_f, longitude λ_f, and latitude φ_f. To cope with the subsequent flight mission, the penetration mission is completed with minimal energy loss.

Design of avoidance guidance scheme. An integrated guidance and avoidance scheme is proposed, as shown in Figure 1. By analyzing the flight process and terminal constraints, a guidance strategy considering minimum energy loss was solved based on optimal guidance theory. The optimal guidance overload factor $\left(n_y^*, n_z^*\right)$ is generated by the guidance strategy, and the solution process is presented in Appendix A. The avoidance of NFZ is realized by adding a lateral maneuvering command Δn_z. The avoidance mission is described as a Markov decision process (MDP) that consists of a finite-dimensional continuous flight state space, an overload command set, and a reward function for judging the command. Depending on the flight data generated by numerical integration, the SAC networks are trained based on the stochastic gradient descent method. Finally, the optimal guidance and trained SAC networks are comprehensively used to generate intelligent avoidance guidance commands.

Construction of avoidance guidance MDP. State and action space. The flight status and information on NFZs are depicted as the state space, $S = \{v, \theta, H, \Delta\sigma, \Delta L\}$. v is the velocity of the vehicle relative to the Earth, θ is the velocity slope angle, and H is the flight altitude. $\Delta\sigma$ denotes heading deviation. ΔL denotes the remaining range. To eliminate the dimensional influence on the states and improve comparability, S was normalized. Under the constraint of the total overload, the maximum maneuvering overload factor is 2, and the action space is $A \in [-2g, 2g]$.

Multiple-mission reward functions. The reward function f_r in this study has the physical meaning of avoiding penetration while minimizing energy consumption and satisfying terminal multi-constrained guidance. The formula $f_r = c_1 v_{\rm fp} - c_2 T_{\rm nfz} - c_3 \dot{\sigma}_{\rm LOSF}$ represents the reward function. The value of $v_{\rm fp}$ is larger, indicating lower energy

* Corresponding author (email: zhujianwen1117@163.com)

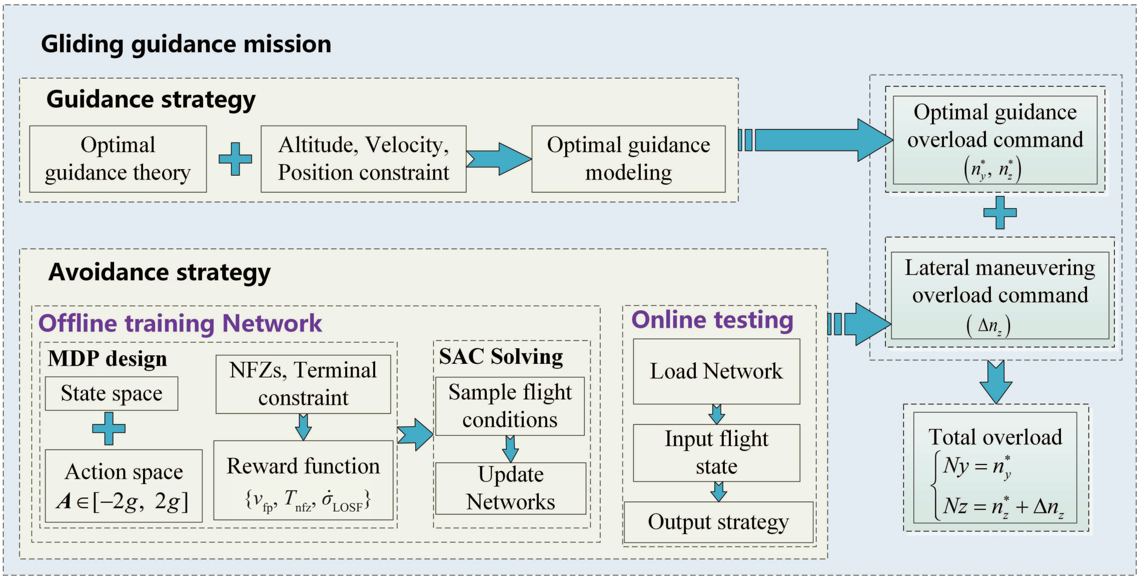

Figure 1 (Color online) Scheme of avoidance guidance.

loss caused by maneuvering. $T_{\rm nfz}$ is the encounter time for NFZs. A necessary and sufficient condition for satisfying the terminal position constraint is to control the heading error to zero, which is directly related to the line-of-sight (LOS) angular rate $\dot{\sigma}_{\rm LOSF}$. If rewards are set only in the terminal state, they result in sparse rewards, leading to a decrease in training efficiency. Therefore, this study introduces a prediction method to calculate the terminal states and add process rewards to accelerate training. In the second half of the gliding flight, the vehicle has sufficient lift to maintain quasi-equilibrium gliding (QEG) flight, while the longitudinal force is balanced. Based on QEG flight conditions, $v_{\rm fp}$ is predicted by calculating the remaining flight time and the differential of the velocity $\dot{v}$. Similarly, $\dot{\sigma}_{\rm LOSF}$ between the vehicle and target at the terminal moment is predicted. The detailed calculation process is presented in Appendix B.

The solving of avoidance guidance law via SAC. Compared to other policy-learning algorithms, the entropy $\mathcal{H}(\pi(\cdot|s_t))$ is added by the SAC, which represents the distribution of actions in strategy π at the current state s_t. Introduction of $\mathcal{H}(\pi(\cdot|s_t))$ improves the robustness of the agent to environmental disturbances and its ability to explore the environment while preventing it from falling into a local minimal solution too early. The entropy parameter τ determines the importance of entropy against the reward. To enhance the adaptability of the algorithm, the training process was improved by adding a random error to the positions of NFZs and targets when generating the flight data. The algorithm is presented in Appendix C.

Simulation verification. Taking CAV-H to verify avoidance guidance performance. We verified the correctness of the proposed scheme and whether the training efficiency improved. Offline-trained DNN parameters were utilized to generate an online avoidance command. Adaptability was validated by changing the flight conditions. The flight conditions and DNN parameters are provided in Appendix D. Specific performance analyses and comparisons are presented in Appendix D.

Conclusion. We propose an intelligent control strategy that synthesizes optimal guidance and SAC, guidance and NFZs avoidance missions are realized with high precision and low energy loss. The training efficiency is enhanced by introducing a prediction method to calculate the terminal states and adding process rewards. By improving the training process, the learned strategy has a strong generalization on problems of dynamic NFZs, indicating higher applicability and flexibility in flight missions.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 92371203).

Supporting information Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Zhu J W, Zhang H, Zhao S B, et al. Multi-constrained intelligent gliding guidance via optimal control and DQN. Sci China Inf Sci, 2023, 66: 132202

2 Yu W, Chen W, Jiang Z, et al. Analytical entry guidance for no-fly-zone avoidance. Aerospace Sci Tech, 2018, 72: 426–442

3 Li Z, Yang X, Sun X, et al. Improved artificial potential field based lateral entry guidance for waypoints passage and no-fly zones avoidance. Aerospace Sci Tech, 2019, 86: 119–131

4 Zhang H, Wang H, Li N, et al. Time-optimal memetic whale optimization algorithm for hypersonic vehicle reentry trajectory optimization with no-fly zones. Neural Comput Applic, 2020, 32: 2735–2749

5 Haarnoja T, Zhou A, Abbeel P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with stochastic actor. In: Proceedings of International conference on machine learning, Stockholm Sweden, 2018. 1861–1870