



## 无人机反应式扰动流体路径规划

吴健发 王宏伦 王延祥 刘一恒

### UAV Reactive Interfered Fluid Path Planning

WU Jian-Fa, WANG Hong-Lun, WANG Yan-Xiang, LIU Yi-Heng

在线阅读 View online: <https://doi.org/10.16383/j.aas.c210231>

---

## 您可能感兴趣的其他文章

### 基于深度强化学习的平行企业资源计划

Parallel Enterprises Resource Planning Based on Deep Reinforcement Learning

自动化学报. 2017, 43(9): 1588–1596 <https://doi.org/10.16383/j.aas.2017.c160664>

### 舰载无人机自主着舰回收制导与控制研究进展

Research Development in Autonomous Carrier-Landing/Ship-Recovery Guidance and Control of Unmanned Aerial Vehicles

自动化学报. 2019, 45(4): 669–681 <https://doi.org/10.16383/j.aas.2018.c170261>

### 基于旋翼无人机近地面空间应急物联网节点动态协同部署

Dynamic Cooperative Deployment of Emergency Internet of Things Near Ground Space Based on Drone

自动化学报. 2021, 47(8): 2002–2015 <https://doi.org/10.16383/j.aas.c180146>

### 基于显著图融合的无人机载热红外图像目标检测方法

Object Detection Method Based on Saliency Map Fusion for UAV-borne Thermal Images

自动化学报. 2021, 47(9): 2120–2131 <https://doi.org/10.16383/j.aas.c200021>

### 多智能体深度强化学习的若干关键科学问题

Important Scientific Problems of Multi-Agent Deep Reinforcement Learning

自动化学报. 2020, 46(7): 1301–1312 <https://doi.org/10.16383/j.aas.c200159>

### 基于深度强化学习的有轨电车信号优先控制

Signal Priority Control for Trams Using Deep Reinforcement Learning

自动化学报. 2019, 45(12): 2366–2377 <https://doi.org/10.16383/j.aas.c190164>

# 无人机反应式扰动流体路径规划

吴健发<sup>1,2,3</sup> 王宏伦<sup>1,3</sup> 王延祥<sup>1,3</sup> 刘一恒<sup>1,3</sup>

**摘要** 针对复杂三维障碍环境,提出一种基于深度强化学习的无人机(Unmanned aerial vehicles, UAV)反应式扰动流体路径规划架构.该架构以一种受约束扰动流体动态系统算法作为路径规划的基本方法,根据无人机与各障碍的相对状态以及障碍物类型,通过经深度确定性策略梯度算法训练得到的动作网络在线生成对应障碍的反应系数和方向系数,继而可计算相应的总和扰动矩阵并以此修正无人机的飞行路径,实现反应式避障.此外,还研究了与所提路径规划方法相适配的深度强化学习训练环境规范性建模方法.仿真结果表明,在路径质量大致相同的情况下,该方法在实时性方面明显优于基于预测控制的在线路径规划方法.

**关键词** 无人机,反应式路径规划,受约束扰动流体动态系统,深度强化学习,训练环境

**引用格式** 吴健发,王宏伦,王延祥,刘一恒.无人机反应式扰动流体路径规划.自动化学报,2023,49(2):272-287

**DOI** 10.16383/j.aas.c210231

## UAV Reactive Interfered Fluid Path Planning

WU Jian-Fa<sup>1,2,3</sup> WANG Hong-Lun<sup>1,3</sup> WANG Yan-Xiang<sup>1,3</sup> LIU Yi-Heng<sup>1,3</sup>

**Abstract** In this paper, aiming at complex 3D obstacle environments, a reactive interfered fluid path planning framework is proposed for unmanned aerial vehicles (UAV) based on deep reinforcement learning. The constrained interfered fluid dynamical system algorithm is used as the fundamental path planning method in the framework. According to relative states between unmanned aerial vehicles and each obstacle, and categories of obstacles, the reaction and direction coefficients of the corresponding obstacle are generated online using the actor networks trained by deep deterministic policy gradient. On this basis, the total modulation matrices in constrained interfered fluid dynamical system can be resolved and the flight path is accordingly modified to realize the reactive obstacle avoidance. In addition, the normative modeling method of deep reinforcement learning training environments, which is matched with the proposed path planning method, is studied. Finally, simulation results show that the proposed method is obviously superior to the online path planning method based on predictive control in real-time performance under the condition that the path qualities are approximately the same.

**Key words** Unmanned aerial vehicle (UAV), reactive path planning, constrained interfered fluid dynamical system, deep reinforcement learning, training environments

**Citation** Wu Jian-Fa, Wang Hong-Lun, Wang Yan-Xiang, Liu Yi-Heng. UAV reactive interfered fluid path planning. *Acta Automatica Sinica*, 2023, 49(2): 272-287

目前,随着无人机(Unmanned aerial vehicles, UAV)的作业空域,由中高空向低空乃至超低空不断拓展,其所面临的障碍环境也日趋复杂,具体表现为低空障碍具有密集性、动态性和不确定性的特

点<sup>[1]</sup>.复杂障碍环境对无人机的飞行安全带来了极大的挑战,同时也对无人机的自主控制能力提出了更高要求.作为无人机自主控制能力的关键技术,在线路径规划方法受到广泛关注,从决策行为角度看,可大致分为慎思式和反应式两类方法<sup>[2-3]</sup>.

慎思式在线路径规划方法主要基于全局静态障碍信息和对动态障碍的状态预测信息进行决策,其代表性方法为基于预测控制的路径规划方法,即预测有限步长内的障碍物状态,基于此优化该时间段内的控制序列,最后执行当前时刻所需控制输入并以此类推,例如Lindqvist等<sup>[4]</sup>和茹常剑等<sup>[5]</sup>采用非线性模型预测控制方法直接产生规避机动的控制输入;Luo等<sup>[6]</sup>和Wu等<sup>[7]</sup>将势场类路径规划方法与滚动时域控制策略(Receding horizon control, RHC)相结合,通过RHC策略在线优化势场类方

收稿日期 2021-03-29 录用日期 2021-09-17

Manuscript received March 29, 2021; accepted September 17, 2021

国家自然科学基金(62173022, 61673042, 61175084)资助  
Supported by National Natural Science Foundation of China (62173022, 61673042, 61175084)

本文责任编辑 许斌

Recommended by Associate Editor XU Bin

1. 北京航空航天大学自动化科学与电气工程学院 北京 100191  
2. 北京控制工程研究所空间智能控制技术重点实验室 北京 100094  
3. 北京航空航天大学飞行器控制一体化技术重点实验室 北京 100191

1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191 2. Science and Technology on Space Intelligent Control Laboratory, Beijing Institute of Control Engineering, Beijing 100094 3. Science and Technology on Aircraft Control Laboratory, Beihang University, Beijing 100191

法的参数, 以应对复杂多变的障碍环境. 这类方法虽然能取得较好的规划效果, 但由于障碍状态预测和串行优化控制序列两大过程需要耗费较长的计算时间, 因此可能无法满足复杂环境下规划的实时性要求.

与慎思式方法相反, 反应式在线路径规划方法一般不需要对未来障碍状态进行预测, 而是基于当前或过去检测到的障碍与规划主体间相对状态进行快速决策, 例如 Steiner 等<sup>[8]</sup>提出一种基于开放扇区的无人机反应性避障路径规划方法, 该方法根据机载激光雷达的二维扫描信息和对无人机过去机动行为的短期记忆信息, 设计了一系列规避规则; 魏瑞轩等<sup>[8]</sup>借鉴生物条件反射机制, 提出基于 Skinner 理论的无人机反应式应急规避方法; Hebecker 等<sup>[9]</sup>将无人机传感器视场离散化为网格地图, 然后基于障碍在网格地图中的分布情况采用波前算法实现局部三维路径规划.

近年来, 以深度强化学习为代表的新一代人工智能方法广泛应用于各类复杂系统的优化控制问题, 此类机器学习方法具有如下优点<sup>[10-12]</sup>: 1) 不依赖于环境模型和先验知识, 仅需要通过与环境进行交互即可实现策略的升级; 2) 所引入的神经网络具有强大的非线性逼近能力, 可以有效应对高维连续状态-动作空间下的优化控制问题 (三维复杂障碍环境下无人机避障路径规划的本质); 3) 由于深度强化学习得到的策略在使用时只需进行一个神经网络的前向传播过程, 非常适用于具有高实时性需求的决策任务. 基于上述优点, 部分学者对其在反应式路径规划中的应用进行了一定的探索, 例如 Guo 等<sup>[13]</sup>提出一种面向离散动作空间的分层 Q 学习反应式路径规划方法, 可用于动态威胁环境下的无人机自主导航; Tai 等<sup>[14]</sup>、Wang 等<sup>[15-16]</sup>和 Hu 等<sup>[17]</sup>则针对连续动作空间, 基于深度确定性策略梯度算法 (Deep deterministic policy gradient, DDPG) (也是应用最为广泛的连续型深度强化学习方法之一) 及其衍生算法设计反应式路径规划方法. 这些方法均实现了良好的避障效果, 但仍有如下两个问题值得进一步进行深入研究:

1) 深度强化学习本质上属于一种通用型的决策方法, 在处理路径规划这种特定问题时可能难以兼顾安全性和路径质量. 从上述文献的仿真结果可以看出, 直接使用深度强化学习方法生成控制输入以规划路径虽然能确保无人机快速安全避障, 但路径的平滑性并不理想, 不利于底层控制器精确跟踪. 如果能将深度强化学习与经典路径规划方法有机结合, 分别发挥二者在优化速度和路径规划质量方面的优势, 则有望取得更好的规划效果. 然而, 如何设

计此类反应式路径规划架构, 使其能有效应对复杂的障碍环境 (如动静态障碍并发、多障碍、环境中存在不同形状尺寸的障碍等), 目前仍处于探索阶段.

2) 基于深度强化学习的路径规划方法需要无人机与模拟的任务环境进行交互, 并根据环境的反馈不断更新深度神经网络的权重, 最终提取训练好的深度动作网络用于实际环境下的在线规划. 因此如何设计与所用路径规划方法相适配的模拟训练环境, 对于提升训练效率并保障动作网络在复杂障碍环境下泛化性能至关重要. 遗憾的是, 上述文献并没有对训练环境的规范性建模方法进行针对性的研究.

针对上述两个问题, 本文提出一种基于深度强化学习的无人机反应式扰动流体路径规划架构, 主要贡献如下:

1) 在一种经典自然启发式路径规划方法: 扰动流体动态系统算法 (Interfered fluid dynamical system, IFDS)<sup>[7, 18-20]</sup>基础上, 进一步引入无人机运动学模型和约束条件以提升规划路径的可跟踪性, 改进算法称为受约束 IFDS 算法 (Constrained-IFDS, C-IFDS).

2) 将深度强化学习中的 DDPG 算法与 C-IFDS 算法相结合, 分别发挥二者在实时性和生成路径质量方面的优势, 构建反应式路径规划架构. 该架构以 C-IFDS 算法为路径规划的基础方法, 根据当前各障碍与无人机的相对状态、无人机自身状态和障碍包络形状, 通过 DDPG 算法在线优化对应障碍的反应系数和方向系数, 继而计算相应的总和扰动矩阵修正无人机的飞行路径, 实现反应式避障.

3) 提出一种与上述反应式路径规划架构相适配的强化学习训练环境规范性建模方法, 以提升训练效率.

## 1 问题建模

### 1.1 无人机运动学模型与约束

假设飞控系统可保证无人机姿态和速度的稳定, 可建立如下运动学模型:

$$\begin{cases} \dot{x} = V \cos \gamma \cos \chi \\ \dot{y} = V \cos \gamma \sin \chi \\ \dot{z} = V \sin \gamma \\ \dot{V} = (n_x - \sin \gamma)g \\ \dot{\gamma} = \frac{g(n_z - \cos \gamma)}{V} \\ \dot{\chi} = \frac{gn_y}{V \cos \gamma} \end{cases} \quad (1)$$

$\mathbf{P} = [x, y, z]^T$  表示无人机的三维位置;  $V$  为飞行速度;  $\gamma$  和  $\chi$  分别为航迹倾角和航迹偏角;  $g$  为重力常数; 作为控制输入的  $n_x$ 、 $n_y$ 、 $n_z$  表示沿航迹系  $x$ 、 $y$ 、 $z$  轴的过载. 该式所描述的运动学模型还须满足如下约束条件:

$$\begin{cases} n_i \in [n_{i \min}, n_{i \max}], i = x, y, z \\ \dot{\gamma} \in [-\dot{\gamma}_{\max}, \dot{\gamma}_{\max}], \dot{\chi} \in [-\dot{\chi}_{\max}, \dot{\chi}_{\max}] \\ \gamma \in [\gamma_{\min}, \gamma_{\max}] \end{cases} \quad (2)$$

## 1.2 障碍环境建模

为了避免过于精细地描述飞行环境信息, 提升路径规划效率, 可采用标准凸面体包络对地形或威胁进行等效. 对于地形或其他静态障碍可用相应凸多面体及其组合体直接等效, 例如延绵的山脉可用半球体等效, 建筑可视为平行六面体或圆柱体; 对于动态威胁 (如入侵飞行器) 可建模为具有速度的球体. 因此, 可建立如下障碍/威胁的等效标准凸面体包络方程:

$$\Gamma(\mathbf{P}) = \left( \frac{x - x_0}{a + R_A} \right)^{2p} + \left( \frac{y - y_0}{b + R_A} \right)^{2q} + \left( \frac{z - z_0}{c + R_A} \right)^{2r} \quad (3)$$

式中,  $a, b, c > 0$  和  $p, q, r > 0$  分别决定了障碍物的覆盖范围与形状, 例如: 当  $p = q = r = 1$  且  $a = b = c$  时, 障碍为圆球; 当  $p = q = 1, r > 1$  且  $a = b$  时, 障碍为圆柱;  $\mathbf{P}_0 = [x_0, y_0, z_0]^T$  表示障碍物中心; 无人机自身安全半径为  $R_A$ ;  $\Gamma(\mathbf{P}) > 1$ ,  $\Gamma(\mathbf{P}) = 1$  和  $\Gamma(\mathbf{P}) < 1$  分别表示无人机位置  $\mathbf{P}$  位于障碍物等效包络的外部、表面和内部.

## 2 受约束扰动流体动态系统路径规划方法

IFDS 路径规划方法模拟了自然界水流的宏观特征: 当无障碍物时, 水流沿直线流动; 当遇到障碍物时, 水流总会平滑地绕过该障碍并最终流向终点. 基于障碍物的位置、速度、形状等具体信息, 该方法可将障碍物对初始流线的扰动影响量化表示, 经计算得到的扰动流线即可作为规划路径. 传统 IFDS 方法的基本原理如下<sup>[18-19]</sup>.

假设无人机当前位置和目的地位置分别为  $\mathbf{P}$  和  $\mathbf{P}_d = [x_d, y_d, z_d]^T$ , 飞行速度为  $V$ . 当环境内不存在障碍物时, 初始流场 (飞行路径) 应为从  $\mathbf{P}$  到  $\mathbf{P}_d$  的直线, 惯性系下的初始流速 (飞行速度矢量)  $\mathbf{u}(\mathbf{P})$  应为:

$$\mathbf{u}(\mathbf{P}) = \left[ \frac{V(x_d - x)}{\|\mathbf{P} - \mathbf{P}_d\|} \quad \frac{V(y_d - y)}{\|\mathbf{P} - \mathbf{P}_d\|} \quad \frac{V(z_d - z)}{\|\mathbf{P} - \mathbf{P}_d\|} \right]^T \quad (4)$$

当环境中存在  $K$  个障碍物时, 障碍物对  $\mathbf{u}(\mathbf{P})$  的干扰影响可用总和扰动矩阵  $\bar{\mathbf{M}}(\mathbf{P})$  表示:

$$\bar{\mathbf{M}}(\mathbf{P}) = \sum_{k=1}^K \omega_k(\mathbf{P}) \mathbf{M}_k(\mathbf{P}) \quad (5)$$

式中,  $\omega_k(\mathbf{P})$  为第  $k$  个障碍物的权重系数, 该值取决于无人机与障碍物等效表面的距离, 距离越大权重系数越小;  $\mathbf{M}_k(\mathbf{P})$  为第  $k$  个障碍物的扰动矩阵.  $\omega_k(\mathbf{P})$  和  $\mathbf{M}_k(\mathbf{P})$  的公式如下:

$$\omega_k(\mathbf{P}) = \begin{cases} 1, & K = 1 \\ \prod_{i=1, i \neq k}^K \frac{(\Gamma_i(\mathbf{P}) - 1)}{(\Gamma_i(\mathbf{P}) - 1) + (\Gamma_k(\mathbf{P}) - 1)}, & K \neq 1 \end{cases} \quad (6)$$

$$\mathbf{M}_k(\mathbf{P}) = \mathbf{I} - \frac{\mathbf{n}_k(\mathbf{P}) \mathbf{n}_k^T(\mathbf{P})}{|\Gamma_k(\mathbf{P})|^{\frac{1}{\rho_k}} \mathbf{n}_k^T(\mathbf{P}) \mathbf{n}_k(\mathbf{P})} + \frac{\mathbf{t}_k(\mathbf{P}) \mathbf{n}_k^T(\mathbf{P})}{|\Gamma_k(\mathbf{P})|^{\frac{1}{\sigma_k}} \|\mathbf{t}_k(\mathbf{P})\| \|\mathbf{n}_k(\mathbf{P})\|} \quad (7)$$

式中,  $\Gamma(\mathbf{P})$  表示由式 (3) 定义的障碍包络方程,  $\mathbf{I}$  为三阶单位吸引矩阵. 式 (7) 等号右边第 2 项和第 3 项分别为排斥矩阵和切向矩阵;  $\rho_k$  和  $\sigma_k$  分别为对应障碍的排斥反应系数和切向反应系数, 其值决定了规划路径的形状, 值越大, 规避障碍的时机越早;  $\mathbf{n}_k(\mathbf{P})$  为径向法向量, 垂直于障碍表面向外;  $\mathbf{t}_k(\mathbf{P})$  为惯性系  $O-xyz$  下的切向矩阵, 推导过程如下.

在与  $\mathbf{n}_k(\mathbf{P})$  垂直的切平面  $S$  上定义两个相互垂直的切向量  $\mathbf{t}_{k,1}(\mathbf{P})$  和  $\mathbf{t}_{k,2}(\mathbf{P})$ :

$$\mathbf{t}_{k,1}(\mathbf{P}) = \left[ \frac{\partial \Gamma_k(\mathbf{P})}{\partial y} \quad -\frac{\partial \Gamma_k(\mathbf{P})}{\partial x} \quad 0 \right]^T \quad (8)$$

$$\mathbf{t}_{k,2}(\mathbf{P}) = \begin{bmatrix} \frac{\partial \Gamma_k(\mathbf{P})}{\partial x} \frac{\partial \Gamma_k(\mathbf{P})}{\partial z} \\ \frac{\partial \Gamma_k(\mathbf{P})}{\partial y} \frac{\partial \Gamma_k(\mathbf{P})}{\partial z} \\ -\left( \frac{\partial \Gamma_k(\mathbf{P})}{\partial x} \right)^2 - \left( \frac{\partial \Gamma_k(\mathbf{P})}{\partial y} \right)^2 \end{bmatrix} \quad (9)$$

以  $\mathbf{t}_{k,1}(\mathbf{P})$ 、 $\mathbf{t}_{k,2}(\mathbf{P})$ 、 $\mathbf{n}_k(\mathbf{P})$  为  $x'$ 、 $y'$ 、 $z'$  三轴建立坐标系  $O'-x'y'z'$ , 则切平面  $S$  内任意单位切向量在  $O'-x'y'z'$  表示为:

$$\mathbf{t}'_k(\mathbf{P}) = [\cos \theta_k \quad \sin \theta_k \quad 0]^T \quad (10)$$

式中,  $\theta_k \in [-\pi, \pi]$  为任意切向量与  $x'$  轴的夹角, 称为切向方向系数, 决定流线的方向.

通过坐标旋转矩阵  $\mathbf{R}'_k$  可将  $O'-x'y'z'$  下的  $\mathbf{t}'_k(\mathbf{P})$

转换为  $O-xyz$  下的  $t_k(\mathbf{P})$ :

$$\begin{cases} \mathbf{t}_k(\mathbf{P}) = \mathbf{R}_k^I(\mathbf{P}) \mathbf{t}'_k(\mathbf{P}) \\ \mathbf{R}_k^I(\mathbf{P}) = \begin{bmatrix} r_y & r_x r_z & r_x \\ r_2 & r_2 r_3 & r_3 \\ r_x & r_y r_z & r_y \\ r_2 & r_2 r_3 & r_3 \\ 0 & -\frac{r_2}{r_3} & \frac{r_z}{r_3} \end{bmatrix} \\ r_x = \frac{\partial \Gamma_k(\mathbf{P})}{\partial x}, r_y = \frac{\partial \Gamma_k(\mathbf{P})}{\partial y}, r_z = \frac{\partial \Gamma_k(\mathbf{P})}{\partial z} \\ r_2 = \sqrt{r_x^2 + r_y^2}, r_3 = \sqrt{r_x^2 + r_y^2 + r_z^2} \end{cases} \quad (11)$$

然后, 考虑到移动威胁的影响, 通过  $\bar{\mathbf{M}}(\mathbf{P})$  修正  $\mathbf{u}(\mathbf{P})$  可得扰动流速  $\bar{\mathbf{u}}(\mathbf{P})$ :

$$\bar{\mathbf{u}}(\mathbf{P}) = \bar{\mathbf{M}}(\mathbf{P}) (\mathbf{u}(\mathbf{P}) - \mathbf{v}(\mathbf{P})) + \mathbf{v}(\mathbf{P}) \quad (12)$$

式中,  $\mathbf{v}(\mathbf{P})$  为障碍总和和速度矢量, 定义为:

$$\mathbf{v}(\mathbf{P}) = \sum_{k=1}^K \omega_k(\mathbf{P}) e^{1-\Gamma_k(\mathbf{P})} \mathbf{v}_k \quad (13)$$

式中,  $\mathbf{v}_k$  为第  $k$  个障碍物的速度矢量.

最后, 通过  $\bar{\mathbf{u}}(\mathbf{P})$  对位置进行积分计算得到规划的下一时刻航路点.

由上述推导过程可以看出, 传统 IFDS 在规划时并未直接考虑无人机的运动模型和约束. 因此, 本文引入了如式 (1) 的模型和式 (2) 的约束对扰动流速  $\bar{\mathbf{u}}(\mathbf{P})$  进一步修正, 改进后的算法即为受约束 IFDS (C-IFDS) 算法. 假设当前时刻为  $n$ , 此时航迹倾角和航迹偏角分别为  $\gamma_n$  和  $\chi_n$ , 则修正步骤如下:

**步骤 1.** 式 (12) 所计算出的扰动流速  $\bar{\mathbf{u}}(\mathbf{P}) = [\bar{u}_x, \bar{u}_y, \bar{u}_z]^T$  为无人机规避障碍的期望速度, 据此计算期望航迹角  $\gamma_c$  和  $\chi_c$ , 以及相应期望角速率, 分别如式 (14)、式 (15) 所示:

$$\begin{cases} \gamma_c = \arcsin\left(\frac{\bar{u}_z}{\|\bar{\mathbf{u}}(\mathbf{P})\|}\right) \\ \chi_c = \arctan\left(\frac{\bar{u}_y}{\bar{u}_x}\right) \end{cases} \quad (14)$$

$$\begin{cases} \dot{\gamma}_c = \frac{\gamma_c - \gamma_n}{\Delta T} \\ \dot{\chi}_c = \frac{\chi_c - \chi_n}{\Delta T} \end{cases} \quad (15)$$

**步骤 2.** 在式 (15) 中引入式 (2) 的角速率和航迹倾角约束, 可计算出如下实际可达的角度  $\gamma_{af}$  和  $\chi_{af}$ :

$$\gamma'_{af} = \begin{cases} \gamma_c, & |\dot{\gamma}_c| < \dot{\gamma}_{\max} \\ \gamma_n + \dot{\gamma}_{\max} \Delta T, & \dot{\gamma}_c \geq \dot{\gamma}_{\max} \\ \gamma_n - \dot{\gamma}_{\max} \Delta T, & \dot{\gamma}_c \leq -\dot{\gamma}_{\max} \end{cases}$$

$$\gamma_{af} = \begin{cases} \gamma'_{af}, & \gamma_{\max} \leq \gamma'_{af} \leq \gamma_{\min} \\ \gamma_{\max}, & \gamma'_{af} > \gamma_{\max} \\ \gamma_{\min}, & \gamma'_{af} < \gamma_{\min} \end{cases} \quad (16)$$

$$\chi_{af} = \begin{cases} \chi_c, & |\dot{\chi}_c| < \dot{\chi}_{\max} \\ \chi_n + \dot{\chi}_{\max} \Delta T, & \dot{\chi}_c \geq \dot{\chi}_{\max} \\ \chi_n - \dot{\chi}_{\max} \Delta T, & \dot{\chi}_c \leq -\dot{\chi}_{\max} \end{cases} \quad (17)$$

**步骤 3.** 将  $\gamma_{af}$  和  $\chi_{af}$  代入式 (15), 得到受约束的角速率  $\dot{\gamma}_{af}$  和  $\dot{\chi}_{af}$ .

**步骤 4.** 将  $\dot{\gamma}_{af}$  和  $\dot{\chi}_{af}$  代入式 (1) 中的航迹角方程, 得到此时按规划路径飞行的需用过载  $n_x$ 、 $n_y$ 、 $n_z$ , 并根据式 (2) 对需用过载进行约束.

**步骤 5.** 将约束后的过载作为控制输入代入式 (1) 的速度和航迹角方程中, 可求解得到下一个路径点位置.

### 3 基于深度强化学习的反应式扰动流体路径规划架构

由式 (7) 可以看出, 扰动矩阵  $\mathbf{M}_k(\mathbf{P})$  除了与无人机位置  $\mathbf{P}$  和障碍方程  $\Gamma_k(\mathbf{P})$  等不可更改的因素有关外, 还与两个可调的反应系数  $\rho_k$  和  $\sigma_k$  以及一个方向系数  $\theta_k$  有关, 其对规划航路的影响如图 1 所示.

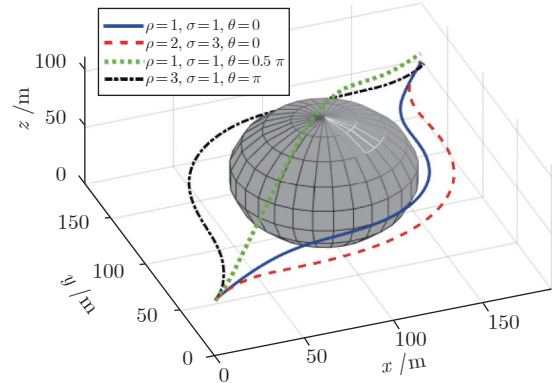


图 1 不同反应系数和方向系数组合对规划路径的影响  
Fig.1 Effects of different combinations of reaction coefficients and direction coefficients on planned paths

如图 1 所示, 不同系数的组合能够决定路径的形状和方向. 在之前的研究中<sup>[7, 18-20]</sup>, 大多采用 RHC 策略在线优化这些系数. 然而, RHC 的串行求解机制并不能很好地满足复杂障碍环境下的强实时性需求, 因此本文将强化学习中的 DDPG 算法与 C-IF-



制<sup>[22-23]</sup>, 即优先采样  $Q$  值估计误差较大的样本, 以提高训练效率, 相应样本  $i$  的时间差分误差  $\delta_i$  定义如下:

$$\delta_i = Q_i^* - Q_i \quad (22)$$

则样本  $i$  的采样概率  $P_i$  为:

$$P_i = \frac{p_i^\alpha}{\sum_{\kappa=1}^{N_S} p_\kappa^\alpha} \quad (23)$$

式中,  $\alpha \in [0, 1]$  用于调节优先程度 (当  $\alpha = 0$  时退化为均匀采样);  $p_i$  为样本  $i$  的优先级, 定义如下:

$$p_i = |\delta_i| + \varepsilon \quad (24)$$

式中,  $\varepsilon$  用于防止概率为 0.

由于基于优先级的经验回放改变了样本的采样频率, 因此需要引入重要性采样更新样本计算梯度时的误差权重  $w_i$ :

$$w_i = \frac{1}{N_S^\beta P_i^\beta} \quad (25)$$

式中,  $\beta$  用于控制校正程度.

**步骤 5.** 通过  $Q^*$  和评价现实网络输出  $Q$  值的均方差作为损失函数计算评价现实网络的梯度, 评价现实网络的损失函数  $L$  由下式计算:

$$L = \frac{1}{N_S} \sum_{\kappa=1}^{N_S} w_\kappa \delta_\kappa^2 \quad (26)$$

式中,  $C$  的梯度可由  $L$  计算.

**步骤 6.** 使用 Adam 优化器<sup>[24]</sup> 更新  $\lambda_t^C$  至  $\lambda_{t+1}^C$ .

**步骤 7.** 动作现实网络的目标是使评价网络的输出  $Q$  值增大, 得到可以获得更多奖励的策略, 所以动作现实网络的梯度通过评价现实网络的梯度计算:

$$\begin{aligned} \nabla_{\lambda_t^A} J(\lambda_t^A) = & \\ & \frac{1}{N_S} \sum_{\kappa=1}^{N_S} (\nabla_a C(o_t, a_t | \lambda_{t+1}^C) |_{o=o_\kappa, a=A(o_\kappa)}) \\ & \nabla_{\lambda^A} A(o_t | \lambda_t^A) |_{o=o_\kappa} \end{aligned} \quad (27)$$

式中,  $J$  表示给定策略的期望回报. 由式 (27) 可知,  $J$  对  $\lambda_t^A$  的梯度由  $C$  对控制输入  $a$  的梯度点乘  $A$  对其参数  $\lambda_t^A$  的梯度得到.

**步骤 8.** 使用 Adam 优化器更新  $\lambda_t^A$  至  $\lambda_{t+1}^A$ .

**步骤 9.** 用现实网络的参数渐变更新目标网络的参数:

$$\begin{cases} \lambda_{t+1}^{A'} = \tau \lambda_t^A + (1 - \tau) \lambda_t^{A'} \\ \lambda_{t+1}^{C'} = \tau \lambda_t^C + (1 - \tau) \lambda_t^{C'} \end{cases} \quad (28)$$

式中,  $\tau$  是渐变更新系数. 然后返回步骤 1.

当迭代次数达到最大值  $T$  或达到此时设定的终止条件 (例如无人机与障碍发生碰撞或无人机成功到达目的地) 时, 进入下一回合, 直至达到最大回合  $M$  结束训练. 通过上述迭代过程, DDPG 深度强化学习模型通过对象模型及环境不断学习, 调整自身网络参数, 使得自身性能不断增强.

评价网络和动作网络所采用的网络结构如图 3 所示, 其中, 评价网络包括观测量输入通路和动作量输入通路; 整个网络由输入层、全连接层 (FC)、线性整流 (ReLU) 激活函数层和添加层 (ADD) 组成; 动作网络由输入层、全连接层、ReLU 激活函数层和双曲正切 (tanh) 激活函数层组成. 全连接层节点数均为 128.

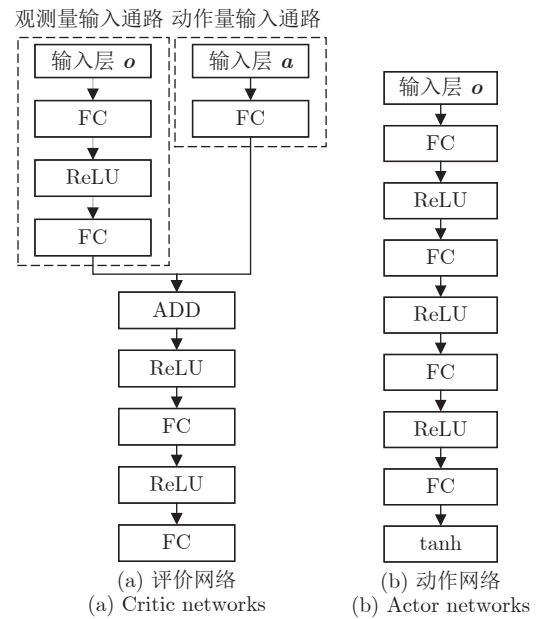


图 3 评价网络和动作网络结构  
Fig. 3 Structures of critic network and actor network

经过上述迭代训练得到的动作网络可用于对 C-IFDS 中反应系数和方向系数的优化, 该系数优化机制由多个经 DDPG 算法训练好的深度动作网络并行组成, 其数量与当前检测到的障碍物数量相同. 对于各个障碍物, 首先判断其形状 (球体、圆柱等) 和类型 (静态障碍或动态威胁), 然后选择对应的 DDPG 动作网络 (障碍形状和类型对网络选择及训练环境建模的影响详见本文第 4 节), 每个动作网络以当前无人机与对应障碍物的相对状态 (相对位置、速度、距离) 和无人机自身状态 (航迹角) 作为输入项, 以对应的反应系数和方向系数组合作为输出项, 通过式 (7) 计算生成各障碍对应的扰动矩阵  $M_k(P)$ . 最终通过加权求和的方式计算出总和扰动矩阵  $\bar{M}(P)$ , 从而实现对空间中多个障碍物的规避机动, 反应式路径规划流程如图 4 所示.

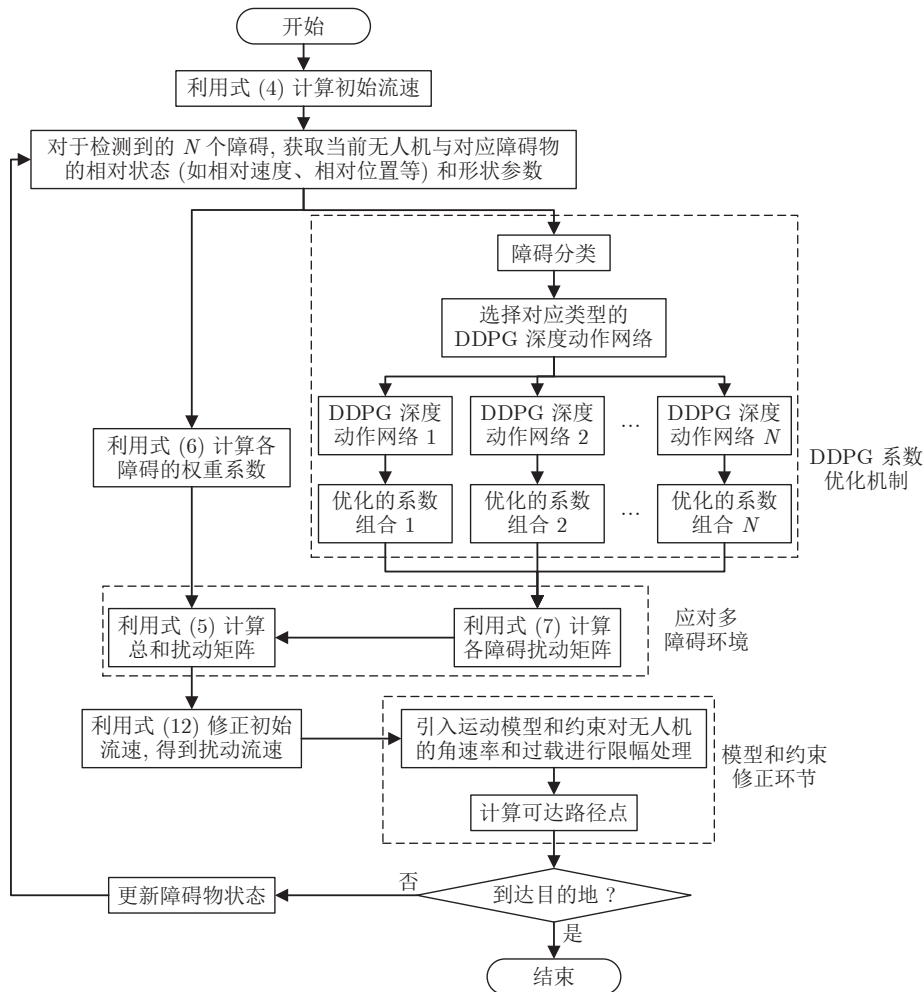


图 4 基于深度强化学习的反应式扰动流体路径规划总体流程图

Fig. 4 Overview flow chart of the DRL-based reaction interfered fluid path planning

#### 4 面向无人机反应式路径规划的强化学习训练环境建模

本文反应式路径规划方法的关键在于训练,而在训练中最为重要的部分就是对规范性模拟环境的搭建,这也是之前研究所相对忽视的.具体建模过程如下.

首先,需要根据障碍物的类型和形状精细化地设计相应的模拟环境,也就是说,针对不同类型或形状的障碍物应设计不同的模拟环境,由不同类型模拟环境训练出的动作网络将组成一个网络集合,在真实环境应用时,无人机应首先判断障碍物的类型,然后选择对应的网络优化 C-IFDS 中的系数(如图 4 所示).原因有以下两点:1) 不同形状的障碍可能对 C-IFDS 中反应系数和方向系数的选择产生影响(特别是方向系数),例如当无人机遭遇圆柱体障碍时,一般会倾向于规划使无人机沿圆柱体侧面进行规避的路径(如  $\theta = 0, \pi$  等);而当遭遇半球

体障碍时,还可规划使无人机沿球体上方越过的路径(如  $\theta = 0.5\pi$  等);2) 静态障碍和动态威胁在环境构建方面存在差异,主要体现在对环境中的相对速度幅值和相对初始位置的设定上.对于相对速度幅值设定的差异,首先,模拟环境中统一设定障碍或威胁保持静止状态,将无人机的飞行速度等效为无人机与障碍或威胁的相对速度;然后,当无人机在模拟环境中以恒定速率飞行时,其与静态障碍的相对速度幅值始终为其飞行速度幅值,因此在每次模拟中不需要改变无人机速度幅值.但对于动态威胁来说,考虑到真实任务情景中动态威胁运动的不确定性,因此在模拟环境的构建中会引入不同运动速率的动态威胁,即在每次模拟中设定的无人机速度幅值均有所不同.对于相对初始位置设定的差异,以静态半球体障碍和动态球体威胁为例,如图 5 所示,当模拟环境由静态半球体障碍(球心位于地面)组成时,无人机的初始位置只能设置在球心所在水平面之上(称为“上半球”区域,同理还有“下半球”区

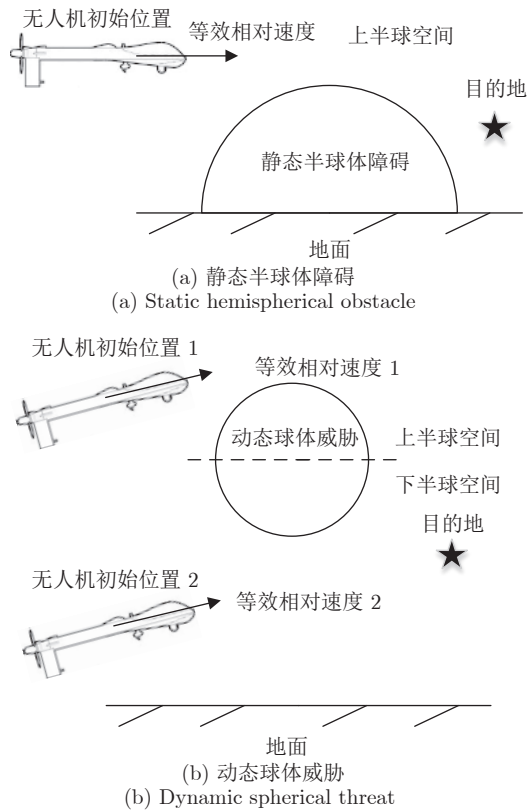


图 5 无人机相对初始位置设定的差异: 以静态半球体障碍和动态球体威胁为例  
 Fig.5 Differences in the setting of UAV initial locations: Taking the static hemispherical obstacle and the dynamic spherical threat as examples

域); 相反, 当模拟环境由空中的动态球体威胁组成时, 无人机的初始位置既可设置在威胁的上半球区域, 也可设置在下半球区域。

综上所述, 对障碍/威胁进行精细化的分类, 有助于降低训练环境的设计难度, 提升 DDPG 训练效率. 因此, 本文主要考虑静态半球体障碍、静态圆柱体障碍和动态球体威胁三类障碍/威胁。

1) 静态半球体障碍

本文设计的模拟训练环境如图 6 所示, 具体建模步骤如下:

**步骤 1.** 设定训练环境中的无人机目的地处于固定位置  $P_d = (0, 400, 150)$  m, 其在水平面的投影点为  $P_{dxy}$ , 障碍球心处于固定位置  $O_{obs} = (0, 0, 0)$  m, 障碍等效半径为  $100 \sim 300$  m 的随机数,  $R_{obs} = (100 + 200 \cdot rand)$  m ( $rand$  表示  $[0, 1]$  的随机数).

**步骤 2.** 以  $O_{obs}$  为中心, 以  $P_{dxy}O_{obs}$  的射线为轴  $O_{obs}x_{obs}$  (该轴与惯性系  $Ox$  轴平行), 建立直角坐标系  $O_{obs}-x_{obs}y_{obs}z_{obs}$  (轴  $O_{obs}y_{obs}$  和  $O_{obs}z_{obs}$  分别与惯性系  $Oy$  和  $Oz$  轴平行且相反).

**步骤 3.** 设定无人机的初始位置为  $P(0)$ , 初始航迹角为  $\gamma(0) = 0$ ,  $\chi(0) = 90^\circ$ , 速度幅值恒为  $V = 30$  m/s. 然后, 从  $P(0)$  向下引垂线, 其与水平面  $O_{obs}x_{obs}y_{obs}$  的交点为  $P_{xy}(0)$ , 此时可通过无人机的高度  $z_{UAV} = |P(0)P_{xy}(0)|$ 、水平面距离  $L_h = |O_{obs}P_{xy}(0)|$  和直线段  $O_{obs}P_{xy}(0)$  与轴  $O_{obs}x_{obs}$  的夹角  $\theta_h$  确定无人机与障碍的相对关系, 上述 3 个量应满足如下约

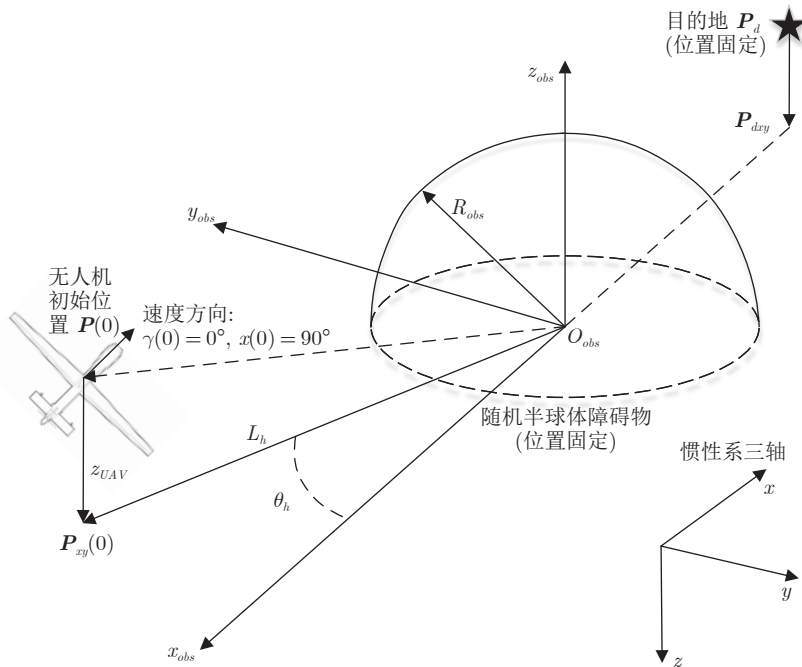


图 6 针对静态半球体障碍的无人机反应式路径规划训练环境  
 Fig.6 Training environment of UAV reaction path planning for static hemispherical obstacles

束条件:

$$\begin{cases} z_{UAV} \leq R_{obs} \\ L_h > R_{obs} + \varepsilon_{Dis} \\ |\theta_h| \leq |\theta_h|_{\max} = \arcsin \frac{R_{obs}}{L_h} + \varepsilon_{Ang} \end{cases} \quad (29)$$

式中,  $\varepsilon_{Dis} > 0$  和  $\varepsilon_{Ang} > 0$  分别表示一定的距离裕量和角度裕量;  $\theta_h$  的约束的意义为: 从俯视角度来看, 无人机的初始位置在  $O_{obs}y_{obs}$  轴向上应处于半球体半径所覆盖的  $[-R_{obs}, R_{obs}]$  范围内, 以提升训练过程中无人机与障碍的交互性 (如果  $|\theta_h|$  过大, 则可能出现无论如何调整动作量, 规划路径均不受障碍明显影响的现象). 在此基础上, 进一步引入一定的角度裕量, 从而进一步提升无人机初始位置选择的灵活性.

**步骤 4.** 根据式 (29), 首先设定  $z_{UAV} = (50 + (R_{obs} - 50) \cdot rand)$  m (即 50 m 至  $R_{obs}$  内的随机高度) 和  $L_h = 600$  m, 则当  $\varepsilon_{Ang} \approx 5^\circ$  时,  $|\theta_h|_{\max} \in [15^\circ, 35^\circ]$ . 然后, 根据常识, 在无人机的初始速度方向与轴  $O_{obs}x_{obs}$  平行且相反的情况下, 迎头障碍对无人机的威胁最大 (即  $|\theta_h|$  较小时), 因此, 随机设定的初始  $\theta_h$  应满足一定的概率分布条件, 使得随机得到越小  $|\theta_h|$  的分布概率较高, 反之分布概率越低, 从而保证无人机能与环境进行充分交互, 避免过早满足重置环境 (即更新回合) 的条件 (见步骤 7). 本文设定  $\theta_h$  满足高斯分布, 其概率分布函数为:

$$f(\theta_h) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\theta_h^2}{2\sigma^2}\right) \quad (30)$$

式中,  $\sigma^2$  为随机变量  $\theta_h$  的方差, 本文中  $\sigma^2 = 4$ . 在生成 5000 次随机初始  $\theta_h$  的条件下, 其概率分布情况如图 7 所示. 由图 7 可以看出, 尽管随机生成  $|\theta_h| > |\theta_h|_{\max}$  的情况非常罕见, 但仍存在可能性, 因此规定, 如果随机生成了  $\theta_h > |\theta_h|_{\max}$  或  $\theta_h < -|\theta_h|_{\max}$ , 则将其分别强制置为  $\theta_h = |\theta_h|_{\max}$  或  $\theta_h = -|\theta_h|_{\max}$ .

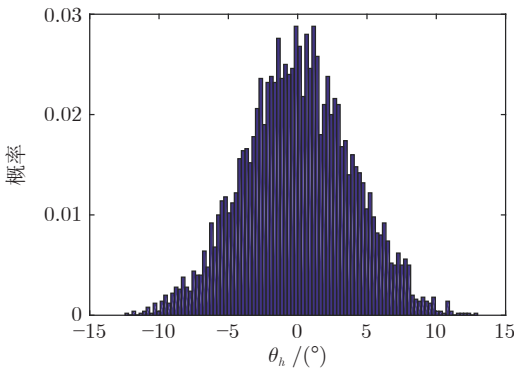


图 7 初始  $\theta_h$  的概率分布

Fig. 7 Probability distribution of the initial  $\theta_h$

**步骤 5.** 根据步骤 4 随机产生的变量, 生成无人机的初始位置:

$$\mathbf{P}(0) = [\Delta x(0), \Delta y(0), \Delta z(0)]^T = [-L_h \sin \theta_h, L_h \cos \theta_h, z_{UAV}]^T \quad (31)$$

则无人机与障碍表面的初始相对距离为  $\Delta L(0) = \sqrt{L_h^2 + z_{UAV}^2} - R_{obs}$ .

**步骤 6.** 步骤 1 ~ 5 设置好初始环境后, 应设计相应奖励函数  $\mathbf{r}$ , 无人机正是以与环境交互获得的奖励函数值为依据来更新其每一步动作.  $\mathbf{r}$  由避撞奖励项  $r_{Col}$ 、航迹角速率奖励项  $r_{Ang}$  和路径长度奖励项  $r_{Len}$  组成:

$$\mathbf{r} = w_{Col}r_{Col} + w_{Ang}r_{Ang} + w_{Len}r_{Len} \quad (32)$$

式中,  $w_{Col}$ ,  $w_{Ang}$ ,  $w_{Len}$  为相应奖励的权重.

$r_{Col}$  表征无人机到障碍等效表面的距离, 距离越远, 奖励值越大. 当无人机与等效障碍发生碰撞时, 需要给该奖励项施加一个负的惩罚值  $pen$ , 该值应与未施加惩罚值时的  $r_{Col}$  在量级上大致相等, 从而避免出现因惩罚值过大而不易收敛的情况. 因此  $r_{Col}$  设计如下:

$$r_{Col} = \begin{cases} \frac{|\mathbf{P}(t+1)O_{obs}| - R_{obs}}{R_{obs}}, & |\mathbf{P}(t+1)O_{obs}| > R_{obs} \\ \frac{|\mathbf{P}(t+1)O_{obs}| - R_{obs}}{R_{obs}} + pen, & |\mathbf{P}(t+1)O_{obs}| \leq R_{obs} \end{cases} \quad (33)$$

式中,  $\mathbf{P}(t+1)$  表示根据当前动作执行 C-IFDS 路径规划方法而更新的无人机位置.

$r_{Ang}$  表征无人机的航迹角变化量, 变化越小, 说明无人机机动幅度越小, 奖励值越大. 因此,  $r_{Ang}$  设计如下:

$$r_{Ang} = -\frac{|\dot{\gamma}(t)|}{\dot{\gamma}_{\max}} - \frac{|\dot{\chi}(t)|}{\dot{\chi}_{\max}} \quad (34)$$

式中,  $\dot{\gamma}(t)$  和  $\dot{\chi}(t)$  表示根据当前动作执行 C-IFDS 路径规划方法而引起航迹角变化时对应的角速率.

$r_{Len}$  表征无人机规划的下一个路径点到目的地的距离, 距离越小, 说明无人机存在向目的地逐渐靠拢的趋势, 对应路径长度可能越短, 奖励值越大. 因此,  $r_{Len}$  设计如下:

$$r_{Len} = \begin{cases} -\frac{\|\mathbf{P}_d - \mathbf{P}(t+1)\|}{L_{SD}}, & \|\mathbf{P}_d - \mathbf{P}(t+1)\| > R_{des} \\ -\frac{\|\mathbf{P}_d - \mathbf{P}(t+1)\|}{L_{SD}} + r_{des}, & \|\mathbf{P}_d - \mathbf{P}(t+1)\| \leq R_{des} \end{cases} \quad (35)$$

式中,  $L_{SD}$  为规划起点到目的地的粗略直线距离, 其目的在于将  $r_{Len}$  的数量级调整至与  $r_{Col}$  和  $r_{Ang}$  大致相等, 从而提升学习算法的收敛性; 当无人机在训练过程中到达目的地时, 则额外给予奖励值  $r_{des}$ .

**注 1.** 式 (33) ~ 式 (35) 设置相应分母项的目的, 在于使各奖励项在量级上大致相同.

**步骤 7.** 设置本回合的终止条件  $IsDone$ . 当无人机到达以  $\mathbf{P}_d$  为中心, 半径为  $R_{des}$  的球形区域时, 或当无人机与障碍发生碰撞时, 以及本回合已达到最大迭代次数  $T$  时, 触发终止条件结束本回合, 在进入下一回合后重新依次随机化设置  $R_{obs}$ 、 $z_{UAV}$  和  $\theta_h$ , 进行试探学习. 则  $IsDone$  的公式为:

$$IsDone = ((\|\mathbf{P}(t+1) - \mathbf{P}_d\| \leq R_{des}) \cup (|\mathbf{P}(t+1)O_{obs}| \leq R_{obs}) \cup (t+1 = T)) \quad (36)$$

## 2) 静态圆柱体障碍

针对静态圆柱体障碍的环境构建步骤与静态半球体障碍的基本相同, 区别在于除了要随机生成圆柱底面半径 (同样记为  $R_{obs}$ , 计算方法也相同) 外, 还要随机生成圆柱体的高  $H_{obs}$ , 其计算方法与  $R_{obs}$  相同, 则无人机的随机初始高度改为  $z_{UAV} = (50 + (H_{obs} - 50) \cdot rand)$  m.

## 3) 动态球体威胁

针对动态球体障碍的环境构建步骤与静态半球体障碍的也基本相同, 区别有以下两点:

a) 动态球体威胁的等效半径范围修改为  $R_{obs} = (50 + 100 \cdot rand)$  m, 球心处于固定位置  $O_{obs} = (0, 0, 150)$  m, 即与  $\mathbf{P}_d$  处于相同高度.

b) 无人机初始高度应处于 50 m 至  $(150 + R_{obs} + \varepsilon_{Dis})$  m 的范围内 (同时包含了威胁的上下半球区域), 则初始随机高度修改为  $z_{UAV} = (50 + (R_{obs} + 150) \cdot rand)$  m; 初始速度幅值修改为随机值  $V = (30 + 30 \cdot rand)$  m/s, 以模拟无人机与不同威胁的相对速度.

**注 2.** 上述训练环境中的参数可根据实际无人机性能和任务环境进行调整.

# 5 仿真实验

## 5.1 案例 1. C-IFDS 与 IFDS 的性能对比测试

仿真情景设置如下: 无人机的初始位置和目的地分别为  $(0, 0, 50)$  m 和  $(600, 600, 50)$  m, 初始速度方向为  $\gamma(0) = 0$  和  $\chi(0) = 45^\circ$ , 速度幅值恒定为 30 m/s; 无人机运动约束为:  $\dot{\gamma}_{\max} = \pi/6$  rad/s、 $\dot{\chi}_{\max} = \pi/6$  rad/s、 $\gamma \in [-\pi/3, \pi/3]$  rad、 $n_x \in [-0.5, 2]$ 、 $n_y \in [-2, 2]$ 、 $n_z \in [-1, 3]$ ; 在  $(250, 250, 0)$  m 处设置等效半径

为 200 m 的半球形障碍物 (已含无人机安全半径); 仿真步长为  $\Delta T = 1$  s. 为保证对比公平性, 两种方法中扰动矩阵参数统一设置为:  $\rho = 2$ 、 $\sigma = 4$  和  $\theta = \pi/4$ . 部分受约束的状态和规划路径对比情况如图 8 所示.

实验结果表明, 尽管 IFDS 和 C-IFDS 均可驱使无人机规避三维空间中的障碍, 但采用 C-IFDS 时无人机的角速率、航迹倾角和过载可以始终保持在约束范围内 (除图 8 所列举的, 其他状态均满足相应约束), 规划路径的可跟踪性较好 (路径能够被无人机精确跟踪的可能性较高). 相反, 采用传统 IFDS 得到的路径则表现出过大的角度和过载变化, 这与无人机的实际运动模型不符, 意味着规划路径的可跟踪性较差. 因此, C-IFDS 是一种比传统 IFDS 更合理的方法.

## 5.2 案例 2. 复杂障碍环境下路径规划性能测试

仿真情景设置如下: 无人机的初始位置和目的地分别为  $(0, 0, 400)$  m 和  $(5000, 5000, 500)$  m, 初始速度方向为  $\gamma(0) = 0$  和  $\chi(0) = 90^\circ$ , 速度幅值恒定为 30 m/s; 无人机运动约束同第 5.1 节; 任务空间内存在多个静态半球体和圆柱体障碍, 还有一个等效半径为 100 m 的动态球体威胁, 于第 222 s 时突然被无人机检测到, 检测后的运动方程为  $x(t) = 4500$  m,  $y(t) = (4900 - 20t)$  m,  $z(t) = 450$  m. DDPG 训练参数如下: 训练回合数为 5000, 回合最大迭代次数  $T$  为 50, 评价网络和动作网络的学习率分别为 0.0001 和 0.001, 批大小为 256, 奖励衰减系数  $\gamma'$  为 0.99, 渐变更新因子  $\tau$  为 0.05, 噪声方差为 0.1, 基于优先级的经验回放机制的参数分别为:  $N_S = 10^6$ 、 $\alpha = 0.6$  和  $\beta = 0.4$ . 对比项设置为基于 RHC 的 C-IFDS 在线航路规划方法, 并假设突发动态威胁的运动轨迹能够直接被精确预测 (即省略了预测轨迹的时间); 为保证对比的公平性, 其代价函数组成和各指标的权重与本文方法的奖励函数相同, 但取值相反; RHC 的解算器为经典的 PSO 算法, 其种群规模为 50, 迭代次数为 20, 滚动步长  $N$  分别取 1、3 和 5 ( $N = 1$  时, 即为贪心算法; 由文献 [18-19] 可知,  $N = 5$  时, 具有相对最佳的优化效果). 仿真计算机配置为: CPU Intel Core i5-4460 3.20 GHz; 内存 8 GB.

DDPG 训练过程中的奖励函数情况如图 9 所示, 在线规划的三维航路如图 10 所示 (图 10 中所绘障碍轮廓均为其等效表面), 与动态威胁等效表面的最近距离如图 11 所示, 规划路径平滑性对比如表 1 所示 (平滑性指标定义为各段路径三维夹角的

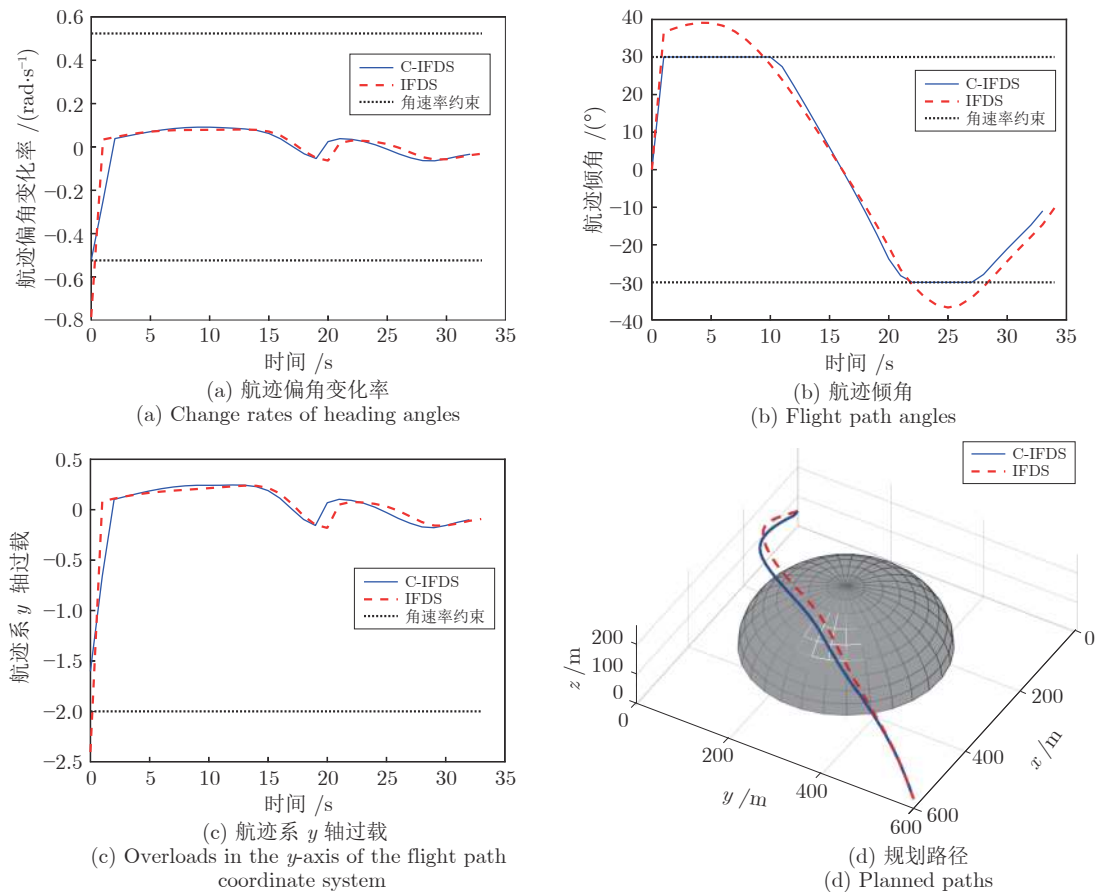


图 8 采用 IFDS 和 C-IFDS 时部分受约束的状态和规划路径的对比情况

Fig. 8 Comparisons of some constrained states and planned paths when using IFDS and C-IFDS

平方和除以总路径段数量, 值越小越平滑), 两类算法的规划时间对比如图 12 所示。

如图 9 所示, 针对三种障碍/威胁, DDPG 算法可分别在训练过程的约 2200、4500 和 3500 回合使奖励函数进入收敛状态。如图 10 ~ 12 和表 1 所示, 本文方法和传统基于 RHC 的 C-IFDS 方法均能使无人机对三维静态障碍和动态威胁进行有效的在线规避, 对比项 2 和 3 在规划路径长度与平滑性方面与本文方法的规划效果大致相近 (本文方法规划路径的长度和平滑性指标甚至更优), 但即使在忽略状态预测时间的前提下, 其单步平均运行时间也远高于本文方法; 而作为对比项 1 的贪心算法虽然相较于其他对比项在规划时间方面具有优势, 但仍为本文方法单步平均运行时间的 8 倍以上, 且其规划路径较长, 质量较低。

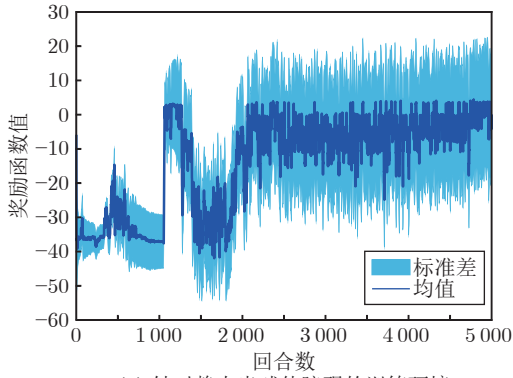
为了进一步验证本文方法的泛化能力以及训练环境规范性建模方法在训练效率方面的优势, 本文针对该仿真情景进行了 20 次蒙特卡洛对比测试。无人机的初始位置设置为如下随机值:  $(-500 + 1000 \cdot$

$\text{rand}, -500 + 1000 \cdot \text{rand}, 400) \text{ m}$ ; 对比项为仍基于本文架构但未采用本文训练环境建模方法的路径规划方法。具体地, 将式 (29)  $\theta_h$  范围扩大为  $[-90^\circ, 90^\circ]$  间的随机值, 且不满足类似于图 7 的概率分布情况, 则以静态半球体障碍为例, 对比项训练情况中奖励函数的情况如图 13 所示。

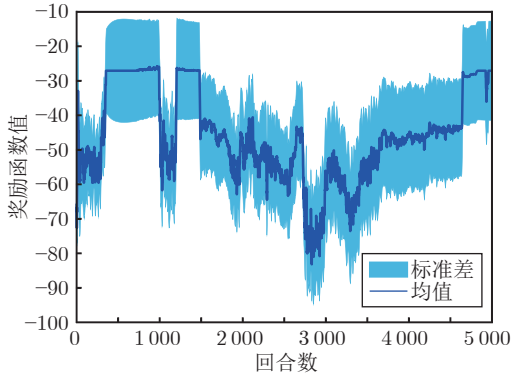
如图 13 所示, 奖励函数并没有如图 9(a) 一样产生比较明显的收敛趋势, 同时, 蒙特卡洛仿真结果也表明, 基于本文架构和训练环境建模方法时, 20 次测试中无人机成功避障并顺利到达目的地的成功率达 100%, 而对比项仅有 60%, 这一方面说明本文方法具有较好的泛化能力, 另一方面也说明通过对强化学习的训练环境进行规范性的建模, 可以显著提升动作网络的训练效率, 在回合数相同时能够取得更好的训练效果, 从而使无人机的避障成功率更高。

### 5.3 案例 3. 多动态威胁环境下路径规划性能测试

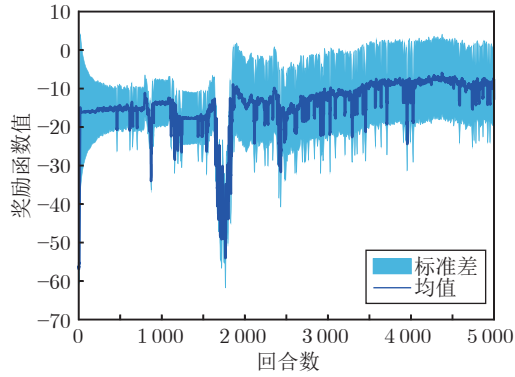
本节进一步验证本文方法在多动态威胁环境下



(a) 针对静态半球体障碍的训练环境  
(a) Reward function for the training environments with static hemispherical obstacles



(b) 针对静态圆柱体障碍的训练环境  
(b) Reward function for the training environments with static cylindrical obstacles

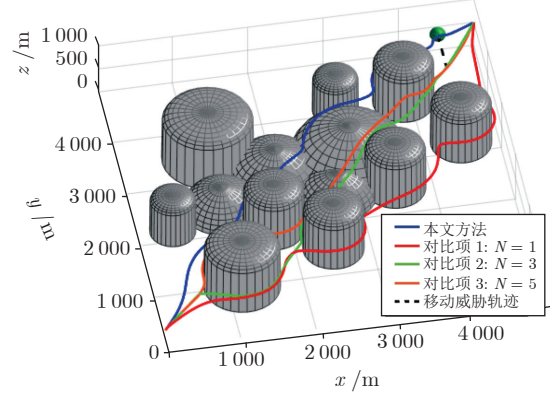


(c) 针对动态球体威胁的训练环境  
(c) Reward function for the training environments with dynamic spherical threats

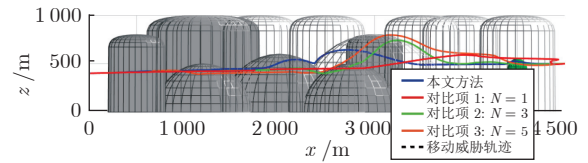
图 9 DDPG 训练过程中的奖励函数情况

Fig.9 Reward functions in the DDPG training process

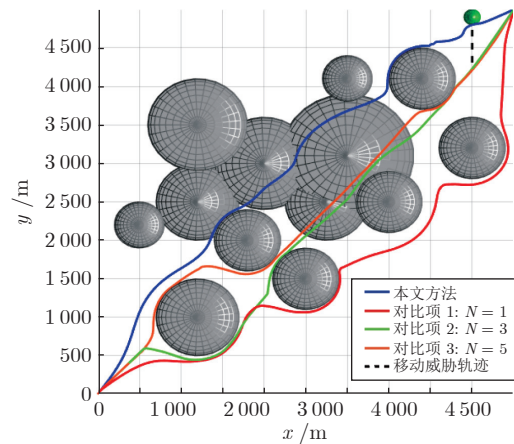
的路径规划性能, 仿真情景设置如下: 无人机的初始位置和目的地分别为  $(-1\ 200, -1\ 200, 2\ 000)$  m 和  $(4\ 000, 4\ 000, 2\ 000)$  m, 初始速度方向为  $\gamma(0) = 0$  和  $\chi(0) = 45^\circ$ , 速度幅值恒定为  $30\ \text{m/s}$ ; 其他运动学参数和 DDPG 参数同第 5.2 节; 任务空间内存在 3 个等效安全半径  $200\ \text{m}$  的动态球体威胁, 其运动模式各有不同, 具体为:



(a) 三维路径  
(a) 3D paths



(b) XY 平面路径 (部分障碍进行了透明化处理)  
(b) Paths in the XY plane (some obstacles are artificially transparent)



(c) XY 平面路径  
(c) Paths in the XY plane

图 10 案例 2 中在线规划的三维路径

Fig.10 3D online planned paths in case 2

1) 动态威胁 1. 匀速直线运动:

$$\begin{cases} x(t) = 4\ 000 - 20t \\ y(t) = 4\ 000 - 20t \\ z(t) = 2\ 000 \end{cases} \quad (37)$$

2) 动态威胁 2. 蛇形运动:

$$\begin{cases} x(t) = 1\ 200\sqrt{2} - 30t \\ y(t) = 30 \sin \frac{3\sqrt{2}\pi t}{80} \\ z(t) = 2\ 000 \end{cases} \quad (38)$$

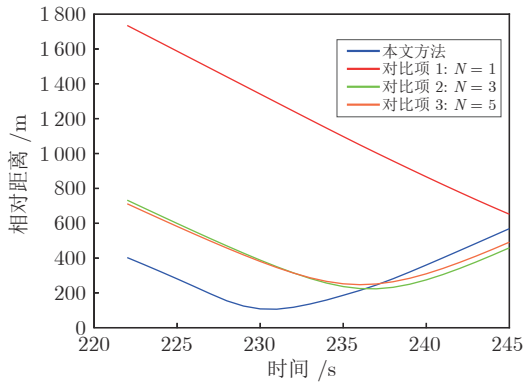
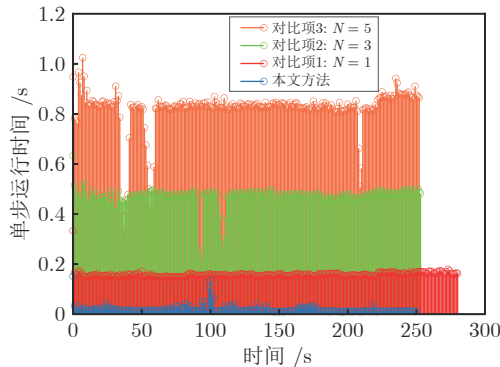


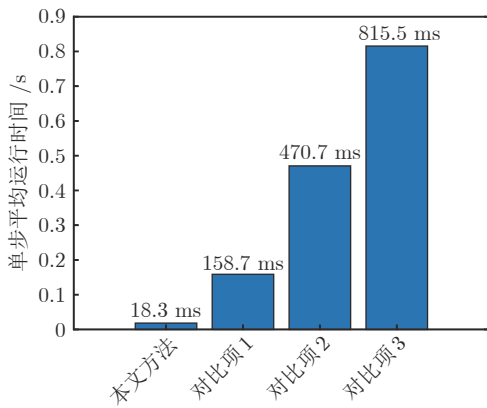
图 11 案例 2 中与动态威胁等效表面的最近距离

Fig.11 Closest distances to the equivalent surface of the dynamic threat in case 2



(a) 单步运行时间对比

(a) Comparison of the single-step runtime



(b) 单步平均运行时间对比

(b) Comparison of the single-step average runtime

图 12 案例 2 中规划时间对比

Fig.12 Comparison of the planning time in case 2

3) 动态威胁 3. 匀速圆周运动:

$$\begin{cases} x(t) = 1000 + 500 \sin(0.1t) \\ y(t) = 1000 + 500 \cos(0.1t) \\ z(t) = 2000 \end{cases} \quad (39)$$

表 1 案例 2 中规划路径长度和平滑性指标对比  
Table 1 Comparison of the length and the smooth indexes for planned paths in Case 2

指标	本文方法	对比项 1	对比项 2	对比项 3
长度 (km)	<b>7.56</b>	8.49	7.68	7.65
平滑性	<b>0.1318</b>	0.3506	0.1593	0.1528

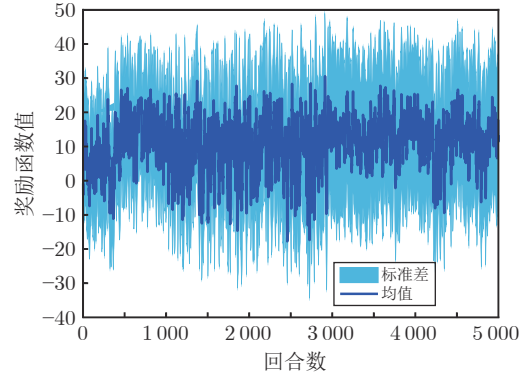


图 13 未采用所提环境建模方法时, DDPG 训练过程中的奖励函数情况: 以静态半球体障碍为例

Fig.13 Reward functions in the DDPG training process when the proposed environment modeling method is not adopted: Taking the static hemispherical obstacle as an example

则不同时刻无人机的航迹 (实线) 与规划路径 (虚线) 如图 14 所示, 无人机与各威胁等效表面的最近距离如图 15 所示. 由图 14、图 15 可见, 无人机与威胁等效表面的最近距离为 36.43 m, 可对多个具有不同运动模式的动态威胁进行有效规避.

综上所述, 本文将深度强化学习与 C-IFDS 相结合的反应式规划方法具有规划速度快、路径质量高等优点, 可用于求解复杂障碍环境下的在线三维路径规划问题.

## 6 结束语

针对复杂障碍环境, 本文提出一种基于深度强化学习的无人机反应式扰动流体路径规划架构. 首先, 在传统 IFDS 方法的基础上提出 C-IFDS 路径规划方法作为架构中的基础规划方法, 该方法引入无人机运动学模型和约束对扰动流速进行可飞性修正; 然后, 提出面向反应式扰动流体路径规划的强化学习训练环境规范性建模方法, 以提升训练效率. 最后, 采用 DDPG 算法在构造的环境中训练相应的深度网络, 并利用训练好的动作网络在线优化 C-IFDS 的反应系数和方向系数. 仿真结果表明, 在生

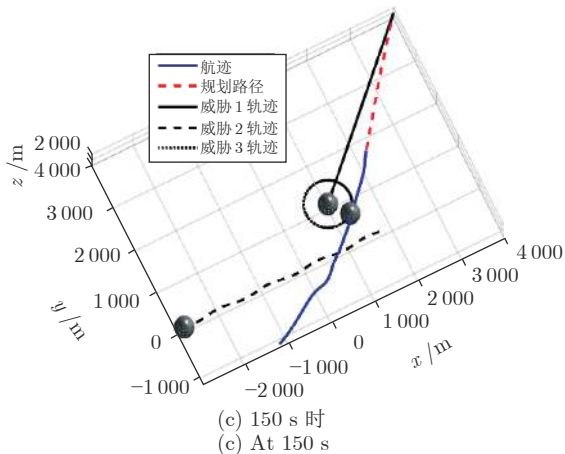
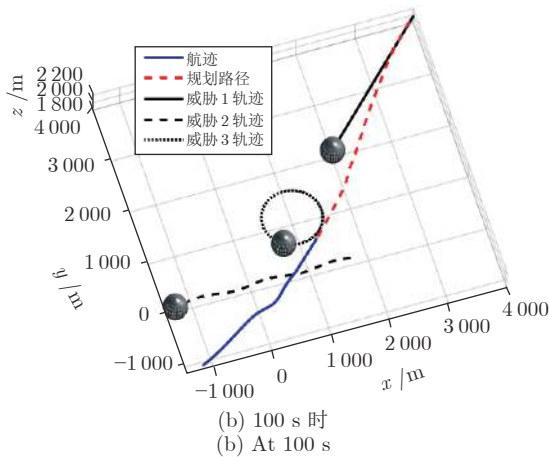
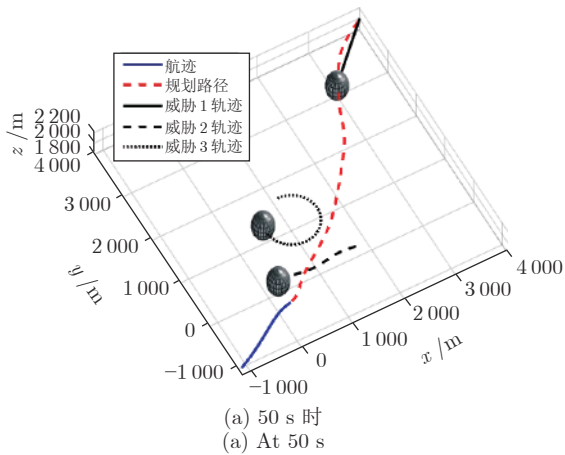


图 14 案例 3 中不同时刻无人机的航迹与规划路径

Fig.14 UAV flight paths and planned paths at different times in case 3

成路径质量大体相同的前提下, 取得了相较于传统 RHC 方法更快的规划速度。

今后的研究工作主要集中在以下几个方面:

1) 本文架构中的深度强化学习方法可以进一步从以下两个角度改进: a) 本文通过对奖励函数加权求和, 从而将路径规划问题转化为一个单目标优

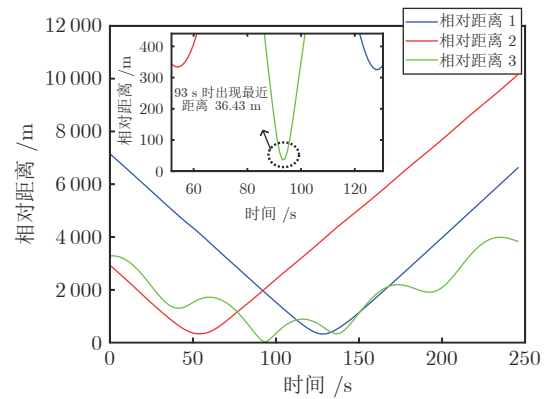


图 15 案例 3 中与各动态威胁等效表面的最近距离

Fig.15 Closest distances to the equivalent surface of each dynamic threat in case 3

化问题, 尽管这种思路比较简单直接, 但也存在着权值不易确定的缺点, 因此在未来可考虑在本文路径规划架构的基础上进一步引入多目标强化学习方法<sup>[25-26]</sup>; b) 理论上, 其他连续型深度强化学习方法亦可应用于本文架构, 因此未来可将更先进的强化学习方法 (如 SAC<sup>[27]</sup> 和 TD3<sup>[28]</sup>等) 与反应式路径规划相结合, 并与本文方法进行对比测试。

2) 将本文架构拓展应用于更多复杂飞行任务中, 例如目标跟踪<sup>[18-19]</sup>、边界监视<sup>[20]</sup>和编队避障<sup>[29]</sup>等, 同时适时开展相应的实物验证工作。

3) 与多数无人机路径规划研究<sup>[3-6, 8-9, 13, 15-19]</sup>相同, 本文架构在规划时只引入了如式 (1)、式 (2) 所示的无人机运动学模型和约束, 而并未考虑更为复杂的无人机六自由度非线性动力学模型和约束, 以及内环控制器的响应特性, 这可能存在着规划指令因无法被控制器及时精确跟踪导致无人机与密集障碍发生碰撞的风险. 因此在未来应考虑在本文路径规划架构下, 将无人机规划-控制-模型所组成的闭环系统引入所构建的强化学习训练环境中, 实现考虑控制器和动力学特性的无人机状态转移, 并据此计算相应的奖励函数。

## References

- 1 Lv Yang, Kang Tong-Na, Pan Quan, Zhao Chun-Hui, Hu Jin-Wen. UAV sense and avoidance: Concepts, technologies, and systems. *Scientia Sinica Informationis*, 2019, **49**(5): 520-537 (吕洋, 康童娜, 潘泉, 赵春晖, 胡劲文. 无人机感知与规避: 概念, 技术与系统. *中国科学: 信息科学*, 2019, **49**(5): 520-537)
- 2 Liu Lei, Gao Yan, Wu Yue-Peng. High speed obstacle avoidance control of wheeled mobile robots with non-homonymic constraint based on viability theory. *Control and Decision*, 2014, **29**(9): 1623-1628 (刘磊, 高岩, 吴越鹏. 基于生存理论的非完整约束轮式机器人高速避障控制. *控制与决策*, 2014, **29**(9): 1623-1628)

- 3 Steiner J A, He X, Bourne J R, Leang K K. Open-sector rapid-reactive collision avoidance: Application in aerial robot navigation through outdoor unstructured environments. *Robotics and Autonomous Systems*, 2019, **112**: 211–220
- 4 Lindqvist B, Mansouri S S, Agha-mohammadi A, Nikolakopoulos G. Nonlinear MPC for collision avoidance and control of UAVs with dynamic obstacles. *IEEE Robotics and Automation Letters*, 2020, **5**(4): 6001–6008
- 5 Ru Chang-Jian, Wei Rui-Xuan, Dai Jing, Shen Dong, Zhang Li-Peng. Autonomous reconfiguration control method for UAV's formation based on Nash bargain. *Acta Automatica Sinica*, 2013, **39**(8): 1349–1359  
(茹常剑, 魏瑞轩, 戴静, 沈东, 张立鹏. 基于纳什议价无人机编队自主重构控制方法. 自动化学报, 2013, **39**(8): 1349–1359)
- 6 Luo G, Yu J, Mei Y, Zhang S. UAV path planning in mixed-obstacle environment via artificial potential field method improved by additional control force. *Asian Journal of Control*, 2015, **17**(5): 1600–1610
- 7 Wu J, Wang H, Zhang M, Yu Y. On obstacle avoidance path planning in unknown 3D environments: A fluid-based framework. *ISA Transactions*, 2021, **111**: 249–264
- 8 Wei Rui-Xuan, He Ren-Ke, Zhang Qi-Rui, Xu Zhuo-Fan, Zhao Xiao-Lin. Skinner-based emergency collision avoidance mechanism for UAV. *Transactions of Beijing Institute of Technology*, 2016, **36**(6): 620–624  
(魏瑞轩, 何仁珂, 张启瑞, 许卓凡, 赵晓林. 基于 Skinner 理论的无人机应急威胁规避方法. 北京理工大学学报, 2016, **36**(6): 620–624)
- 9 Hebecker T, Buchholz R, Ortmeier F. Model-based local path planning for UAVs. *Journal of Intelligent and Robotic Systems*, 2015, **78**(1): 127–142
- 10 Li Kai-Wen, Zhang Tao, Wang Rui, Qin Wei-Jian, He Hui-Hui, Huang Hong. Research reviews of combinatorial optimization methods based on deep reinforcement learning. *Acta Automatica Sinica*, 2021, **47**(11): 1001–1028  
(李凯文, 张涛, 王锐, 覃伟健, 贺惠晖, 黄鸿. 基于深度强化学习的组合优化研究进展. 自动化学报, 2021, **47**(11): 1001–1028)
- 11 Wang Ding. Research progress on learning-based robust adaptive critic control. *Acta Automatica Sinica*, 2019, **45**(6): 1031–1043  
(王鼎. 基于学习的鲁棒自适应评判控制研究进展. 自动化学报, 2019, **45**(6): 1031–1043)
- 12 Wang D, Ha M, Qiao J. Self-learning optimal regulation for discrete-time nonlinear systems under event-driven formulation. *IEEE Transactions on Automatic Control*, 2019, **65**(3): 1272–1279
- 13 Guo T, Jiang N, Li B, Zhu X, Wang Y, Du W. UAV navigation in high dynamic environments: A deep reinforcement learning approach. *Chinese Journal of Aeronautics*, 2021, **34**(2): 479–489
- 14 Tai L, Paolo G, Liu M. Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2017. 31–36
- 15 Wang C, Wang J, Shen Y, Zhang X. Autonomous navigation of UAVs in large-scale complex environments: A deep reinforcement learning approach. *IEEE Transactions on Vehicular Technology*, 2019, **68**(3): 2124–2136
- 16 Wang C, Wang J, Wang J, Zhang X. Deep-reinforcement-learning-based autonomous UAV navigation with sparse rewards. *IEEE Internet of Things Journal*, 2020, **7**(7): 6180–6190
- 17 Hu Z, Gao X, Wang K, Zhai Y, Wang Q. Relevant experience learning: A deep reinforcement learning method for UAV autonomous motion planning in complex unknown environments. *Chinese Journal of Aeronautics*, 2021, **34**(12): 187–204
- 18 Yao P, Wang H, Su Z. Real-time path planning of unmanned aerial vehicle for target tracking and obstacle avoidance in complex dynamic environment. *Aerospace Science and Technology*, 2015, **47**: 269–279
- 19 Yao P, Wang H, Su Z. Cooperative path planning with applications to target tracking and obstacle avoidance for multi-UAVs. *Aerospace Science and Technology*, 2016, **54**: 10–22
- 20 Wu J, Wang H, Zhang M, Su Z. Cooperative dynamic fuzzy perimeter surveillance: Modeling and fluid-based framework. *IEEE Systems Journal*, 2020, **14**(4): 5210–5220
- 21 Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. arXiv preprint arXiv: 1509.02971, 2015.
- 22 Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. arXiv preprint arXiv: 1511.05952, 2016.
- 23 Hou Y, Liu L, Wei Q, Xu X, Chen C. A novel DDPG method with prioritized experience replay. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Piscataway, USA: IEEE, 2017. 316–321
- 24 Kingma D P, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980, 2017.
- 25 Wang Y, He H, Sun C. Learning to navigate through complex dynamic environment with modular deep reinforcement learning. *IEEE Transactions on Games*, 2018, **10**(4): 400–412
- 26 Sutton R S, Modayil J, Delp M, Degris T, Pilarski P M, White A. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In: Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems. Richland, USA: 2011. 761–768
- 27 Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: Proceedings of the 35th International Conference on Machine Learning. New York, USA: 2018. 1861–1870
- 28 Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods. In: Proceedings of the 35th International Conference on Machine Learning. New York, USA: 2018. 1587–1596
- 29 Wu J, Wang H, Li N, Su Z. Formation obstacle avoidance: A fluid-based solution. *IEEE Systems Journal*, 2019, **14**(1): 1479–1490



**吴健发** 北京控制工程研究所博士后. 主要研究方向为飞行器智能决策与协同控制.

E-mail: jianfa\_wu@163.com

**(WU Jian-Fa** Postdoctor at Beijing Institute of Control Engineering. His research interest covers intelligent decision-making and coordinated control of flight vehicles.)



**王宏伦** 北京航空航天大学自动化科学与电气工程学院教授. 主要研究方向为飞行器自主与智能控制, 抗扰动控制, 无人系统路径规划与精确跟踪. 本文通信作者.

E-mail: hl\_wang\_2002@126.com

**(WANG Hong-Lun** Professor at the School of Automation Science and Electrical Engineering, Beihang University. His research interest covers autonomous and intelligent control of flight vehicles, anti-disturbance control, and path planning and precise tracking control of unmanned systems. Corresponding author of this paper.)



**王延祥** 北京航空航天大学自动化科学与电气工程学院博士研究生. 主要研究方向为无人机路径规划, 空中加油精准引导与控制.

E-mail: wyxjy51968@163.com

**(WANG Yan-Xiang** Ph.D. candidate at the School of Automation Science and Electrical Engineering, Beihang University. His research interest covers UAV path planning and precision guidance, and control of air refueling.)



**刘一恒** 北京航空航天大学自动化科学与电气工程学院博士研究生. 主要研究方向为飞行控制, 轨迹规划和机器学习.

E-mail: 18810010709@163.com

**(LIU Yi-Heng** Ph.D. candidate at the School of Automation Science and Electrical Engineering, Beihang University. His research interest covers flight control, trajectory planning, and machine learning.)