

# 基于强化学习的多能源动态滑翔航迹优化方法

张云飞<sup>1,2</sup>, 王宏伦<sup>1,2</sup>, 张梦华<sup>1,2</sup>, 巩轶男<sup>3</sup>

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.北京航空航天大学 飞行器控制一体化技术国防科技重点实验室, 北京 100191;  
3.海鹰航空通用装备有限责任公司, 北京 100074

**摘要:**针对无人机动态滑翔问题,提出了一种基于深度强化学习的航迹优化方法。该方法综合利用梯度风能和太阳能,引入了障碍物约束以模拟复杂障碍环境。使用神经网络近似逼近高斯伪谱方法求解航迹的策略,在训练得到的策略基础上利用双延迟深度确定性策略梯度算法进行策略改进,在大幅度提升推理实时性的同时解决了传统最优控制算法在动态滑翔领域难以应对变化风场的问题。实验针对动态滑翔2种经典模式进行仿真验证,之后在考虑多种能量源的情况下进行蒙特卡洛仿真。结果表明,基于深度强化学习的动态滑翔航迹优化方法在单个滑翔周期内获能与最优结果相近,而实时推理决策时间减少了91%。在变化风场环境下,文中方法相较于传统方法具有更强的适应性。

**关键词:**动态滑翔;强化学习;高斯伪谱;航迹优化

中图分类号:V249.1

文献标志码:A

文章编号:1000-2758(2025)01-0128-12

动态滑翔是信天翁使用的一种近似无动力的飞行模式,采用该模式可以使它们在消耗最少能量的前提下进行长距离飞行。在一个存在梯度风场的环境中,信天翁通过“逆风爬升,顺风下滑”机制在风切变层之间穿梭获取能量从而实现长时间滞空<sup>[1]</sup>。由于信天翁在动态滑翔过程中其翅膀始终保持伸展状态,与小型固定翼无人机类似,因此研究人员尝试将这一机制迁移到后者从而有效地提升飞行器的航时和航程<sup>[2]</sup>。作为动态滑翔的核心问题,航迹优化需要综合考虑风场信息和无人机动力学特性,并在此基础上规划出一条可飞的周期性航迹。它是探索动态滑翔机理和进一步应用的关键技术和难点问题。另一方面,环境中不仅存在风能,还存在其他能量例如太阳能,如何在考虑多种能源的前提下进行航迹优化具有广阔的研究前景。

固定翼无人机的动态滑翔航迹优化问题可以归结为一个最优控制问题,即在满足无人机动力学约束和环境约束的前提下求解代价函数的最值。根据

任务目标不同,优化目标可以是最小化参考风速、最大化无人机获能、最大化航时等<sup>[3]</sup>。在解决这类问题时,大多数学者采用了多重打靶法、解析法、直接配点法、微分平坦法和高斯伪谱(Gaussian pseudospectral, GP)等<sup>[4-7]</sup>的最优控制方法。但在使用这类方法时存在求解复杂度较高、实时性较差的问题。文献[8]对比了上述常见的用于动态滑翔航迹优化的最优控制方法,结果表明在设置51个配点的情况下,使用直接配点法的优化耗时不足2s,相较于相同条件下的微分平坦法和高斯伪谱法优化速度更快,然而高斯伪谱法却具有更高的求解精度。文献[9]考虑到传统方法耗时较长,无法实时根据风场信息修正优化航迹,提出了一种基于凸二次规划的动态滑翔航迹优化方法。但在推导过程中使用到了较多的假设,过程较复杂,当风场模型不一致时需要重新对问题进行凸化,普适性存在一定欠缺。也有一部分学者尝试采用随机树算法进行航迹优化<sup>[10]</sup>,有效地满足了局部规划实时性的要求,但当规划航迹较长时其耗时依旧不满足要求。更为关键的是,随机树无法保证解的唯一性和最优性。

强化学习(reinforcement learning, RL)是一种机器学习范式,旨在通过智能体与环境的大量交互来

使智能体学习某种期望具有泛化性的策略,具有一次训练、多处使用的特点<sup>[11]</sup>。传统方法是在滑翔前直接优化控制量序列,若在滑翔过程中环境变化将导致优化航迹不可用。强化学习的出现为实时动态滑翔优化给出了解决思路:动作网络可以根据当前环境给出合理的控制量且每一时刻都会针对环境做出变化。Li 等<sup>[12]</sup>讨论了如何使用近端策略优化(proximal policy optimization, PPO)求解圆形航迹的动态滑翔问题。文献[13]采用无模型(model free, MF)框架的强化学习算法并取得了初步结果,证实了此方向的可行性。然而 Montella<sup>[14]</sup>指出动态滑翔问题解的搜索空间过大,直接采用强化学习算法难以使智能体学习到最优策略,并提出了基于模仿传统算法决策的“示教控制器”进行强化学习训练的解决思路,很好地解决了动态滑翔在训练初始阶段难以搜索到可行解的问题。但 Montella 给出的方案中求解场景较为单一且使用到的强化学习算法较为简单。

由于环境中不仅存在风能,还存在其他能源例如太阳能等,动态滑翔过程中可以考虑更多能量来源以进一步延长航时。现有大多文献仅考虑了从各种环境风场中获能,少数学者探索了太阳能和风能结合的情况,但缺少综合分析<sup>[15-16]</sup>。考虑将太阳能纳入动态滑翔优化范围内会使优化情况变得更加复杂,此时无人机还需要倾向于将太阳能板垂直于光线入射方向,因此优化结果可能不再满足“逆风爬升,顺风下滑”。

现有关于动态滑翔航迹优化问题研究取得了丰硕的成果,但存在以下问题:①传统最优控制方法需要离线生成航迹,如果环境风场在后续跟踪控制过程中变化,无法针对变化对航迹进行在线修正。采用强化学习的方法可以解决实时性的问题,但是存在训练初期难以收敛的情况。②现有针对多能源结合的动态滑翔航迹优化研究较少,大部分仅考虑了梯度风场获能。③现有方案在求解动态滑翔问题时没有考虑到障碍物约束,而在实际应用场景下可能存在部分以建筑物为代表的障碍。

针对以上问题,本文针对多能源综合利用下的固定翼无人机动态滑翔航迹优化问题,借助双延迟深度确定性策略梯度算法(twin delayed deep deterministic policy gradient, TD3)<sup>[17]</sup>,提出了一种基于强化学习的航迹优化算法。该算法主要创新点为:①引入了动态滑翔障碍物约束用以模拟具有障碍物

的环境;②给出了多能源综合利用下动态滑翔问题建模并对获能机理进行了分析,同时针对该问题设计了相应的强化学习交互模型;③利用深度神经网络学习最优控制算法根据环境信息规划航迹的策略,作为后续训练的“指导者”,为提升实时性奠定基础;④给出了基于“指导者-执行者”的强化学习两阶段训练步骤,用以解决动态滑翔问题搜索空间大,难以训练的问题。

## 1 风场环境及无人机运动学建模

### 1.1 梯度风场模型

固定翼无人机动态滑翔主要从海平面、地面上方数十米位置的侧向梯度风场中获取额外能量,因此本文使用近地表面(海面,地面或者山脊)风场常用的指数梯度风场模型,表达式为<sup>[18-19]</sup>:

$$\begin{cases} V_{w,x}(h) = V_R \left( \frac{h}{H_R} \right)^p \cos e_w \\ V_{w,y}(h) = V_R \left( \frac{h}{H_R} \right)^p \sin e_w \end{cases} \quad (1)$$

式中:  $H_R$  表示某一给定高度;  $V_R$  表示在给定高度下对应的实际风速,  $V_R$  决定了风剖面内的整体梯度和平均风速的大小;  $V_{w,x}$ ,  $V_{w,y}$  指风场中侧向风速在  $x$  轴和  $y$  轴的分量; 指数  $p$  为风场强度变化指数, 表征梯度风场的变化强度;  $e_w$  为风速与  $x$  轴的夹角。

### 1.2 动态滑翔航迹优化模型

根据文献[20-21], 本文给出包含空间变化风场的无人机运动学模型。在东北天坐标系下各变量关系如图 1 所示。

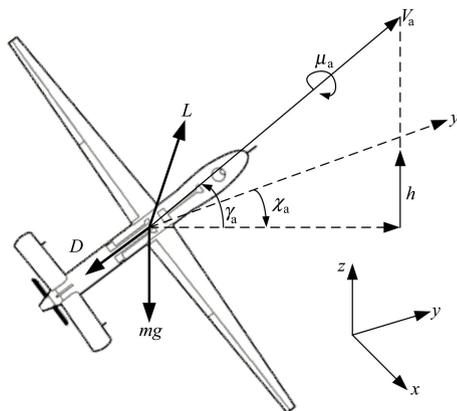


图 1 动态滑翔简化模型坐标系及变量之间关系示意图

$$\begin{cases} \dot{x} = V_a \cos\gamma_a \sin\chi_a + \dot{W}_x \\ \dot{y} = V_a \cos\gamma_a \cos\chi_a + \dot{W}_y \\ \dot{h} = V_a \sin\gamma_a + \dot{W}_h \\ m\dot{V}_a = T - D - mg\sin\gamma_a - m\dot{W}_x \cos\gamma_a \sin\chi_a - \\ \quad m\dot{W}_y \cos\gamma_a \cos\chi_a + m\dot{W}_h \sin\gamma_a \\ mV_a \dot{\gamma}_a = L \cos\mu_a - mg \cos\gamma_a + m\dot{W}_x \sin\gamma_a \sin\chi_a + \\ \quad m\dot{W}_y \sin\gamma_a \cos\chi_a + m\dot{W}_h \cos\gamma_a \\ mV_a \dot{\chi}_a \cos\gamma_a = L \sin\mu_a + m\dot{W}_x \cos\chi_a - m\dot{W}_y \sin\chi_a \end{cases} \quad (2)$$

式中

$$\begin{aligned} \dot{W}_x &= \frac{\partial W_x}{\partial t} + \frac{\partial W_x}{\partial x} \dot{x} + \frac{\partial W_x}{\partial y} \dot{y} + \frac{\partial W_x}{\partial h} \dot{h} \\ \dot{W}_y &= \frac{\partial W_y}{\partial t} + \frac{\partial W_y}{\partial x} \dot{x} + \frac{\partial W_y}{\partial y} \dot{y} + \frac{\partial W_y}{\partial h} \dot{h} \\ \dot{W}_h &= \frac{\partial W_h}{\partial t} + \frac{\partial W_h}{\partial x} \dot{x} + \frac{\partial W_h}{\partial y} \dot{y} + \frac{\partial W_h}{\partial h} \dot{h} \end{aligned} \quad (3)$$

式中,有6个状态变量和3个控制变量,其中包括东、北、天3个位置 $(x, y, h)$ ,空速 $V_a$ ,气流航迹爬升角 $\gamma_a$ 和气流航迹方位角 $\chi_a$ 。 $\mu_a$ 为气动滚转角,以描述升力 $L$ 绕空速 $V_a$ 的转动。模型假设升力系数 $C_L$ 和气动滚转角 $\mu_a$ 可以直接给定,并与飞行器推力 $T$ 一同构成模型虚拟控制输入。飞行器质量为 $m$ 。三轴风速及风加速度为 $W_x, W_y, W_h$ 和 $\dot{W}_x, \dot{W}_y, \dot{W}_h$ 。 $L$ 为升力, $D$ 为阻力。升力与阻力的计算采用简化模型(4)式表示。

$$\begin{aligned} L &= \rho S C_L V_a^2 / 2 \\ D &= \rho S C_D V_a^2 / 2 \\ C_D &= C_{D0} + K_D C_L^2 \end{aligned} \quad (4)$$

式中: $\rho$ 代表空气密度; $C_L$ 和 $C_D$ 分别是升力系数与阻力系数; $C_{D0}$ 为零升阻力系数; $K_D$ 为诱导阻力因子; $S$ 为机翼面积。

### 1.3 太阳能模型

本文除考虑无人机从梯度风场中获能外,还考虑太阳能获取,对太阳能的辐射强度进行建模。首先根据相关文献[22]可知太阳垂直照射下的强度

$$\begin{aligned} I_0 &= I \left( \frac{1 + \varepsilon \cos\alpha_s}{1 - \varepsilon^2} \right)^2 \\ \alpha_s &= 2\pi(n_d - 4) / 365 \end{aligned} \quad (5)$$

式中: $I = 1\ 367\ \text{W}$ 为太阳常数; $\varepsilon = 0.017\ 7$ 为地球偏心率; $n_d$ 为计算照射强度当天与当年1月1日的日期差,单位为天。

要计算太阳能的辐射强度,需要确定当地日地连线矢量 $n_s$ 的方向

$$\begin{aligned} \sin\beta_s &= \sin\theta_s \sin\varphi_s + \cos\theta_s \cos\varphi_s \cos(\omega_s) \\ \sin\gamma_s &= \frac{\sin\beta_s \sin\theta_s - \sin\varphi_s}{\cos\beta_s \cos\theta_s} \\ \varphi_s &= 23.45\pi \sin\left(2\pi \times \frac{284 + n_d}{365}\right) / 180 \\ \omega_s &= \pi - \pi t_{\text{mission}} / 12 \\ n_s &= (\cos\beta_s \sin\gamma_s, \cos\beta_s \cos\gamma_s, \sin\beta_s) \end{aligned} \quad (6)$$

式中: $\beta_s$ 为太阳高度角; $\gamma_s$ 为太阳方位角,光线朝向正东时方位角为0,向南偏转时取正; $\varphi_s$ 为赤纬角; $\theta_s$ 为地理纬度; $\omega_s$ 为太阳时角; $t_{\text{mission}}$ 为一天中的时刻。

容易分析,当日地连线向量 $n_s$ 与固定翼无人机机翼平面的法向量 $n_w$ 的夹角越接近于 $180^\circ$ 时,太阳能辐射吸收效率越大。根据几何关系,有太阳能原始功率 $P_{\text{sun0}}$ 为

$$P_{\text{sun0}} = \begin{cases} I_0 |\cos\langle n_s, n_w \rangle|, & \pi/2 < \langle n_s, n_w \rangle \leq \pi \\ 0, & \text{其他} \end{cases} \quad (7)$$

式中, $\langle n_s, n_w \rangle$ 表示 $n_s$ 与 $n_w$ 之间的夹角。原始功率需要经过一系列的传递单元最终才能转化为机械能,其中涉及到动力系统的能量耗散。因此,最终太阳能通过动力系统转化为机械能的功率 $P_{\text{sun}}$

$$P_{\text{sun}} = P_{\text{sun},0} S \eta_{\text{sc}} \eta_{\text{asc}} \eta_m \eta_p \quad (8)$$

式中: $\eta_{\text{sc}}$ 为光伏电池转化为电能的效率; $\eta_{\text{asc}}$ 为光伏电池在机翼上铺设的面积与机翼面积的比值; $\eta_m$ 为能源管理系统的效率; $\eta_p$ 为推进系统的效率。

当 $\pi/2 < \langle n_s, n_w \rangle \leq \pi$ 时,将(7)式重新表示为

$$\begin{aligned} P_{\text{sun}} &= I_1 |\cos\langle n_s, n_w \rangle| \\ I_1 &= I_0 S_w \eta_{\text{sc}} \eta_{\text{asc}} \eta_m \eta_p \end{aligned} \quad (9)$$

式中

$$n_s = (\cos\beta_s \sin\gamma_s, \cos\beta_s \cos\gamma_s, \sin\beta_s) = (n_{s1}, n_{s2}, n_{s3}) \quad (10)$$

$$n_w = \begin{pmatrix} \sin\varphi \cos\psi - \sin\theta \cos\varphi \sin\psi \\ -\sin\varphi \sin\psi - \sin\theta \cos\varphi \cos\psi \\ \cos\varphi \cos\theta \end{pmatrix}^T \quad (11)$$

(9)式中 $\cos\langle n_s, n_w \rangle$ 的表达式为

$$\begin{aligned} \cos\langle n_s, n_w \rangle &= (\sin\varphi \cos\psi - \sin\theta \cos\varphi \sin\psi) n_{s1} + \\ &(-\sin\varphi \sin\psi - \sin\theta \cos\varphi \cos\psi) n_{s2} + \\ &\cos\varphi \cos\theta n_{s3} = A \sin\varphi + B \cos\varphi \end{aligned} \quad (12)$$

式中

$$A = \cos\psi n_{s1} - \sin\psi n_{s2}$$

$$B = \cos\theta n_{s3} - \sin\theta \sin\psi n_{s1} - \sin\theta \cos\psi n_{s2} \quad (13)$$

式中,  $\psi, \theta, \varphi$  分别为偏航角、俯仰角和滚转角,本文中偏航角的定义以  $y$  轴为基准,向  $x$  轴转动为正。

## 2 动态滑翔最优控制问题

动态滑翔航迹优化问题的本质可归结为一个包含微分方程约束、过程约束、终端约束的最优控制问题。其中微分方程约束受到航迹优化运动学方程影响,终端状态约束受到动态滑翔模式影响,过程约束受到无人机姿态和滑翔范围等因素影响。在进行最优控制问题建模前,本文将首先分析动态滑翔问题的获能机理,为后续目标函数构建奠定基础。

### 2.1 气流系下动态滑翔获能机理分析

应用动态滑翔技术的小型太阳能飞行器飞行过程中主要的能量来自风场中获取的能量  $E_w$  和经由太阳能动力能源系统获得的太阳能  $E_{sun}$ ;主要的能量支出为空气阻力带来的能量损失  $E_D$  和动力系统做功消耗的能量  $E_T$ ,其中  $E_T$  消耗的能量会补充飞行器的机械能。

固定翼无人机的动力学往往与无人机的空速相关,因此相对于气流的能量往往能代表无人机在风场环境下的有效能量。定义气流系下固定翼无人机的总能量为相对于气流的动能与重力势能之和<sup>[23]</sup>,如(14)式所示。

$$E_a = \frac{mV_a^2}{2} + mgh \quad (14)$$

因此,能量的变化率  $\dot{E}_a$  为

$$\dot{E}_a = m\dot{V}_a V_a + mgh \quad (15)$$

进一步地,结合动态滑翔模型得

$$\dot{E}_a = TV_a - DV_a - mV_a \dot{W}_x \cos\gamma_a \sin\chi_a - mV_a \dot{W}_y \cos\gamma_a \cos\chi_a + mV_a \dot{W}_h \sin\gamma_a + mg\dot{W}_h \quad (16)$$

为了简化动态滑翔获能机理的分析,假设环境中只存在沿  $x$  轴水平的梯度风场<sup>[3]</sup>。由于指数梯度风场的大小只与高度  $h$  相关,则  $W_h = \dot{W}_h = 0, \dot{W}_y = 0$ ,

又根据  $\dot{W}_x = \frac{dW_x}{dh} h, \dot{E}_a$  可简化为

$$\dot{E}_a = \dot{E}_T - \dot{E}_D - \dot{E}_W = TV_a - DV_a - mV_a^2 \frac{dW_x}{dh} \sin\gamma_a \sin\chi_a \cos\gamma_a \quad (17)$$

式中:  $\dot{E}_T$  表示发动机推力产生的对飞行器的正功率;  $\dot{E}_D$  表示无人机在飞行过程中阻力产生的功率消耗;  $\dot{E}_W$  代表无人机从梯度风场中获取的能量功率。

### 2.2 动态滑翔问题约束建模

对于过程约束,主要考虑到动态滑翔无人机本身的一些特性,首先是空气动力特性带来的升力系数、最大爬升角、空速和气动滚转角限定为

$$\mu_{a,\min} \leq \mu_a \leq \mu_{a,\max}, 0 \leq C_L \leq C_{L,\max}$$

$$V_{a,\min} \leq V_a \leq V_{a,\max}, \gamma_{a,\min} \leq \gamma_a \leq \gamma_{a,\max} \quad (18)$$

除此之外,无人机的重心要保持一定的安全高度,即

$$h > h_{\min} \quad (19)$$

进一步,为了避免无人机机翼和水平面相碰撞,引入对翼尖间隙的约束

$$h - \frac{1}{2}b|\sin\mu_a| > h_{\min} \quad (20)$$

同时,考虑到动态滑翔所处环境中可能存在诸多障碍物,为保证无人机不与障碍物碰撞,引入对状态  $x, y, h$  的过程约束。为简化起见,障碍物以球形为代表,如(21)式所示。

$$(x - O_{ix})^2 + (y - O_{iy})^2 + (h - O_{ih})^2 > R_i^2$$

$$i = 1, 2, \dots, N \quad (21)$$

式中:  $(O_{ix}, O_{iy}, O_{ih})$  表示球形障碍物球心;  $R_i$  为第  $i$  个障碍物球半径;  $N$  为障碍物个数。

对于终端约束,不同模式的动态滑翔不尽相同。闭合模式的动态滑翔常针对环绕监视的任务场景,其要求无人机优化出来的航迹起始点和终止点状态一致,如(22)式所示。

$$\begin{cases} V_a(t_f) - V_a(t_0) = 0 \\ \chi_a(t_f) - \chi_a(t_0) = 0 \\ \gamma_a(t_f) - \gamma_a(t_0) = 0 \\ h(t_f) - h(t_0) = 0 \\ x(t_f) - x(t_0) = 0 \\ y(t_f) - y(t_0) = 0 \end{cases} \quad (22)$$

而行进模式下要求无人机在动态滑翔过程中向着某个特定方向前进。其终端约束如下

$$\begin{cases} V_a(t_f) - V_a(t_0) = 0 \\ \chi_a(t_f) - \chi_a(t_0) = 0 \\ \gamma_a(t_f) - \gamma_a(t_0) = 0 \\ h(t_f) - h(t_0) = 0 \\ \arctan \frac{x(t_f) - x(t_0)}{y(t_f) - y(t_0)} = \varepsilon \end{cases} \quad (23)$$

式中:  $\varepsilon$  表示了行进模式下无人机的前进方向。若动态滑翔模式为自由行进式, 则 (23) 式没有最后一项。

### 2.3 能量建模和代价函数构建

综合考虑同时利用多种能量构建最优控制代价函数:

1) 由无人机推力带来的能量消耗功率

$$P_T = TV_a \quad (24)$$

2) 由无人机受到的阻力产生的能量消耗功率

$$P_D = DV_a \quad (25)$$

3) 无人机动态滑翔过程中从风场获取能量的功率

$$P_W = -mV_a \dot{W}_x \cos\gamma_a \sin\chi_a - mV_a \dot{W}_y \cos\gamma_a \cos\chi_a + mV_a \dot{W}_h \sin\gamma + mgW_h \quad (26)$$

4) 无人机动态滑翔过程中获取的太阳能功率  $P_{\text{sun}}$ 。

综上所述, 衡量总体能量收支变化率的最终表达式为

$$P_{\text{sum}} = P_W + P_{\text{sun}} - P_T - P_D \quad (27)$$

在本文中, 最优滑翔的目的是在单个滑翔周期内获取最多的能量。因此, 最优控制问题代价函数建模为

$$J = \min \left\{ - \int_{t_0}^{t_f} P_{\text{sum}} dt \right\} \quad (28)$$

## 3 基于强化学习的动态滑翔优化

使用强化学习解决动态滑翔问题首先需要考虑如下问题: ①动态滑翔动作量均为连续动作, 相较于离散动作更难搜索; ②滑翔区域较大, 受各种约束, 同时还要考虑多种能量获取; ③整个训练回合可能需要上千步推理迭代。以上 3 点导致强化学习算法所需搜索的解空间范围较大。因此 Montella 等<sup>[24]</sup>指出需要优先提供一个“指导者”来模仿某一传统方法得到的航迹生成策略, 再利用强化学习在“指导者”的参考指令下进行训练学习, 获得“执行者”。根据此思路, 本文设计了基于深度强化学习的固定翼无人机动态滑翔航迹优化算法框架如图 2 所示, 其分为 3 个阶段: ①“指导者”离线训练阶段: 利用神经网络拟合传统最优控制方法在动态滑翔问题上的优化策略。本文选用了相较于直接配点法精度更高的高斯伪谱方法求解大量不同的动态滑翔问题, 并将对应的状态-控制量序列放入样本池。之后“指导者”神经网络学习从状态量到控制量的映射策略; ②“执行者”离线训练阶段: 引入强化学习, 在相同的环境下进行学习, 不同的是状态量将同时输入“指导者”网络和强化学习“执行者”网络, 之后“执行者”网络的输出将作为“指导者”网络输出的偏置项最终作用在环境中; ③在线使用阶段: 无人机在动态滑翔过程中在线使用“指导者”网络和“执行者”网络, 并将两者输出结合作用于环境获得新的滑翔方向。在这个过程中并没有耗时的最优控制算法参与, 且推理过程采用单步迭代的方式, 从而解决了实时性问题。

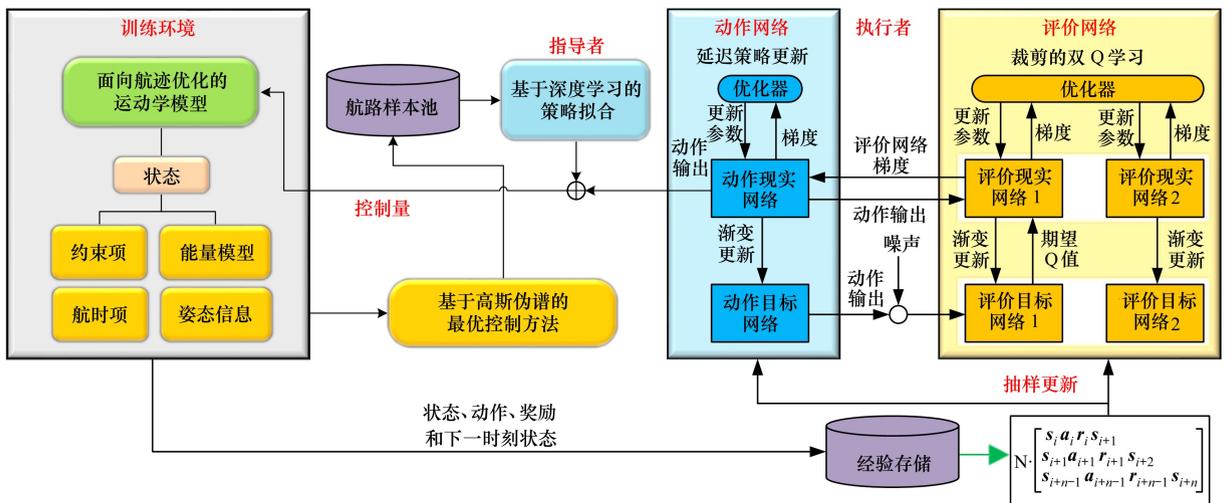


图 2 基于深度强化学习的固定翼无人机动态滑翔航迹优化算法框架

### 3.1 基于深度神经网络的“指导者”训练

“指导者”的作用在于为强化学习智能体提供训练初期的引导策略,考虑到问题的复杂性和其本身的定位,这种引导策略无需十分准确。本文设计了较为简单的多层感知机 (multi-layer perceptron, MLP) 用以拟合高斯伪谱法在动态滑翔问题上针对不同场景下的航迹生成策略。在图 2 中,首先需要在不同场景下收集大量高斯伪谱法解算得到的状态-控制轨迹,如(29)式所示。

$$\begin{cases} \mathbf{X} = \{\mathbf{x}_i | i = 1, 2, \dots, N\} \\ \mathbf{x}_i = [V_a^{(i)}, \mathcal{X}_a^{(i)}, \gamma_a^{(i)}, x^{(i)}, y^{(i)}, h^{(i)}, \\ W_x^{(i)}, W_y^{(i)}, W_h^{(i)}, \dot{W}_x^{(i)}, \dot{W}_y^{(i)}, \dot{W}_h^{(i)}]^\top \\ \mathbf{Y} = \{\mathbf{y}_i | i = 1, 2, \dots, N\} \\ \mathbf{y}_i = [\mu_a^{(i)}, C_L^{(i)}, T^{(i)}, t_c^{(i)}]^\top \end{cases} \quad (29)$$

式中:  $i$  表示样本编号;  $t_c^{(i)}$  表示控制量作用时间;  $N$  为样本总数。之后利用神经网络对样本集  $(\mathbf{X}, \mathbf{Y})$  进行拟合,网络将有能力输出风场环境中某一状态下无人机需要的虚拟控制量。训练得到的网络被用于之后深度强化学习训练中,并在其基础上进行策略的优化。

### 3.2 基于延迟确定性策略梯度算法的“执行者”训练

本文使用 TD3 作为动态滑翔问题中训练“执行者”网络的算法。TD3 算法通过引入双重评价网络、目标策略平滑正则化和延迟更新策略进一步提升了算法的收敛速度和效果。同时,本文进一步采用“价值扩展”的方式改进了算法更新过程中期望  $Q$  值的计算过程,使得动作网络学习效果更好。

#### 1) 奖励函数设计

本文对于动态滑翔问题的优化目标是单个滑翔周期内最大化能量获取。因此,强化学习奖励函数设计为

$$r = k(P_w + P_{\text{sum}} - P_T - P_D) \quad (30)$$

式中,  $k$  为大于 0 的比例系数。当  $k = 1$  时,  $r$  在数值上等于动态滑翔飞行器瞬时总功率。

#### 2) 状态量设计

强化学习中状态量设计为无人机运动学模型(2)中各状态变量

$$\mathbf{s} = (k_1 x, k_2 y, k_3 h, k_4 V_a, k_5 \gamma_a, k_6 \mathcal{X}_a, k_7 V_{w,x}, \\ k_8 V_{w,y}, k_9 V_{w,h}, k_{10} \dot{V}_{w,x}, k_{11} \dot{V}_{w,y}, k_{12} \dot{V}_{w,h}) \quad (31)$$

式中,  $k_j > 0, j = 1, \dots, 12$ 。将各个状态量转化为数量级相近的变量。

#### 3) 动作值设计

动作值选用动态滑翔简化模型(2)式中的虚拟控制量与控制量作用时间  $t_c$  的组合

$$\mathbf{a} = (\mu_a, C_L, T, t_c) \quad (32)$$

#### 4) 算法流程

算法在“指导者”网络的基础上使用 TD3 强化学习进行训练,强化学习网络将输出对于“指导者”网络基准值的偏置,用于调整其优化策略。执行流程如下。

步骤 1 动作现实网络根据无人机动态滑翔训练环境中的状态计算得到一个动作输出,并与“指导者”网络提供的动作进行叠加,最终得到动作  $a_t$  并下达给动态滑翔仿真环境执行。

步骤 2 动态滑翔仿真环境执行  $a_t$ , 返回奖励  $r_t$  和新的状态  $\mathbf{x}_{t+1}$ 。

步骤 3 将这个状态转换过程(状态  $\mathbf{x}_t$ 、动作  $a_t$ 、奖励  $r_t$  和新的状态  $\mathbf{x}_{t+1}$ ) 存入经验存储中。

步骤 4 从经验存储中采样  $N$  个状态转换序列数据,作为动作网络和评价网络训练的一个小批量数据。

步骤 5 利用动作目标网络和评价目标网络计算期望  $Q$  值。TD3 使用 2 套评价网络,从中取最小值后通过“价值扩展”的方式计算期望值,即

$$Q^* = r_t + r_{t+1} + \dots + r_{t+n-1} + \\ \gamma^n \min_{j=1,2} C_j^*(\mathbf{x}_{t+n}, A'(\mathbf{x}_{t+n} | \lambda^{A'}) + \varepsilon) \quad (33)$$

式中:  $Q^*$  表示评价现实网络的期望值,  $n$  为价值扩展的步数,  $C_j^*$  表示第  $j$  个评价目标网络。  $A'$  表示动作目标网络,  $\lambda^{A'}$  为动作目标网络的参数,  $\gamma$  是奖励衰减系数,  $\varepsilon \sim \text{clip}(N(0, \sigma), -c, c)$ ,  $c > 0$  为噪声。“价值扩展”通过考虑未来更多步的奖励值,因此计算得到的期望值  $Q$  将更接近于真实值。评价现实网络的损失函数由(34)式计算。

$$L = \frac{1}{2N} \sum_{i=1}^N (Q_i^* - C_j(\mathbf{x}_i, a_i | \lambda_j^c))^2 \quad (34)$$

式中,  $\lambda_j^c$  为第  $j$  个评价现实网络的参数。

步骤 6 使用 Adam 优化器根据损失函数的梯度对评价现实网络的参数  $\lambda_j^c$  进行更新。

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  $A$  的梯度通过评价现实网络的梯度计算,如(35)式所示。

$$\nabla_{\lambda^A} J(A) = \frac{1}{N} \sum_{i=1}^N (\nabla_u C_1(\mathbf{x}, u | \lambda_1^c) |_{\mathbf{x}=\mathbf{x}_i, u=A(\mathbf{x}_i)}) \cdot$$

$$\nabla_{\lambda^A} A(\mathbf{x} | \lambda^A) |_{\mathbf{x}=\mathbf{x}_i} \quad (35)$$

式中,  $J$  表示损失函数。由(35)式可知,  $J$  对  $\lambda^A$  的梯度由评价现实网络  $C_1$  对控制输入  $u$  的梯度点乘动作现实网络  $A$  对其参数  $\lambda^A$  的梯度得到。

步骤8 采用延迟更新的策略对动作现实网络的参数进行更新,即评价网络更新多次后,动作网络才更新一次,提高动作网络更新的准确性。

步骤9 用评价现实网络的参数软更新评价目标网络的参数。即

$$\begin{cases} \lambda^{A'} = \tau \lambda^A + (1 - \tau) \lambda^{A'} \\ \lambda_j^{C'} = \tau \lambda_j^C + (1 - \tau) \lambda_j^{C'}, j = 1, 2 \end{cases} \quad (36)$$

式中,  $\tau \in (0, 1)$  是软更新系数。

### 4 仿真验证

首先使用高斯伪谱算法求解2种经典模式下的动态滑翔问题,之后进一步引入基于强化学习的“指导者-执行者”机制并进行对比仿真实验。仿真过程中使用到的滑翔机以及环境参数如表1所示<sup>[18]</sup>。

表1 滑翔机和环境参数

参数名称	取值	参数名称	取值
$m/\text{kg}$	5.443	$e_w/\text{rad}$	0
$S/\text{m}^2$	0.957	$\beta_s/(\text{°})$	-20
$C_{D_0}$	0.017	$\gamma_s/(\text{°})$	180
$K_D$	0.019 2	$\eta_{sc}$	0.2
$H_R/\text{m}$	20	$\eta_{asc}$	0.8
$V_R/(\text{m} \cdot \text{s}^{-1})$	8	$\eta_m$	0.95
$p$	0.25	$\eta_p$	0.72
$n_d$	160	$\rho/(\text{kg} \cdot \text{m}^{-3})$	1.225

#### 4.1 高斯伪谱算法求解动态滑翔问题

本小节将使用高斯伪谱算法求解自由行进模式和闭合模式下的动态滑翔问题。无人机状态初值见表2,状态量、控制量等参数的最大、最小值限制如表3所示。

表2 动态滑翔问题无人机状态初值

$V_{a0}/(\text{m} \cdot \text{s}^{-1})$	$\psi_{a0}/(\text{°})$	$\gamma_{a0}/(\text{°})$	$x_0/\text{m}$	$y_0/\text{m}$	$h_0/\text{m}$
15	0	0	0	0	12

表3 动态滑翔最优控制问题中参数最大、最小值限制

参数名称	最小值限制	最大值限制
$t_f/\text{s}$	5	15
$V_a/(\text{m} \cdot \text{s}^{-1})$	9.54	73.2
$\psi_a/(\text{°})$	-180	180
$\gamma_a/(\text{°})$	-40	40
$x/\text{m}$	-2 000	2 000
$y/\text{m}$	-2 000	2 000
$h/\text{m}$	2	100
$\mu_a/(\text{°})$	-60	60
$C_L$	0.01	1
$T/\text{N}$	0	20

#### 1) 自由行进模式仿真

自由行进模式下,高斯伪谱算法优化结果如图3所示(其中黑线为航迹投影)。由图可知,高斯伪谱算法优化所得航迹符合动态滑翔的核心机理,即“逆风爬升,顺风下滑”。无人机通过重复图3中单个周期的航迹即可向目标方向前进。无人机从梯度风场中获取的能量功率曲线如图4所示。由图4可知,获取风能的功率呈现2个波峰形状,恰好对应逆风爬升和顺风下滑2个阶段。除此之外,图4中只有横轴上方的阴影,说明无人机在整个滑翔过程只从风场中获能而没有损失能量。

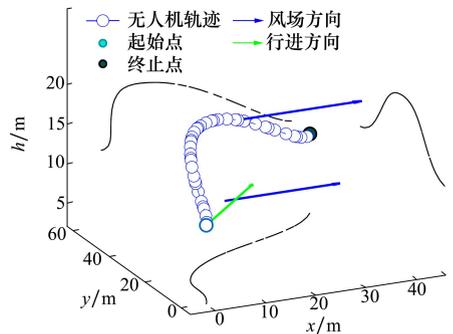


图3 利用高斯伪谱求解动态滑翔航迹优化结果

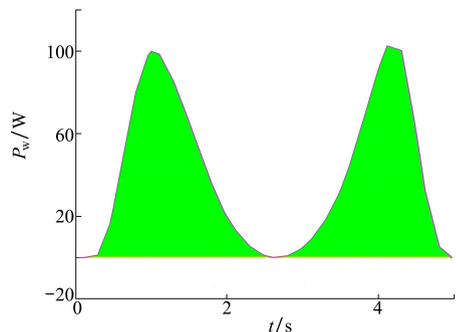


图4 无人机从梯度风场中获能功率图

### 2) 闭合模式仿真

闭合模式下,高斯伪谱算法所得优化三维航迹如图 5 所示。由图 5 可知,优化航迹形状类似“8”字,属于典型的动态滑翔航迹优化结果。无人机先逆风爬升后顺风下滑,紧接着再一次爬升,之后下滑到初始位置,完成周期运动。

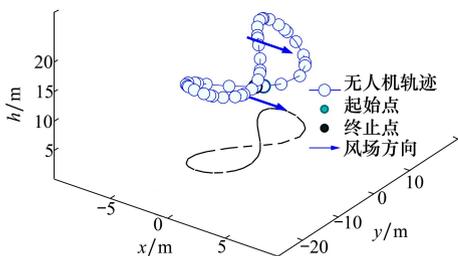


图 5 高斯伪谱算法在闭合模式下优化航迹三维图

### 3) 考虑障碍物情况下的动态滑翔

考虑多个障碍物情况下的固定翼无人机动态滑翔航迹优化求解中,障碍物用球形包络替代,2 个障碍物球心位置和半径见表 4。

表 4 动态滑翔过程中障碍物球心位置和半径

障碍物编号	球心位置/m	半径/m
1	(0,60,10)	15
2	(20,40,10)	10

行进方向  $\varepsilon$  设置为  $90^\circ$ ,其中障碍物 1 刚好处于无人机行进方向上,因此无人机需要综合考虑躲避障碍和获取能量,航迹优化结果如图 6 所示。由图 6 可知,尽管绿色箭头指向的给定行进方向上存在障碍,无人机依旧可以优化出 1 条躲避障碍的航迹,且终止位置满足行进方向。

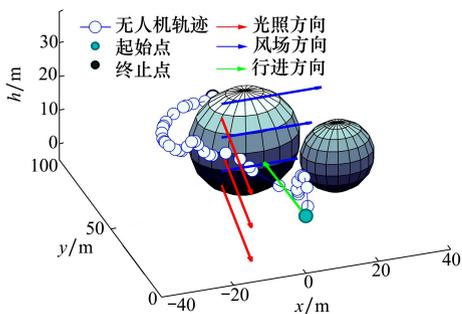


图 6 考虑障碍物情况下行进模式航迹优化结果

## 4.2 “指导者”网络训练

将由高斯伪谱算法得到的 2 000 条不同初始值

的优化航迹放入航路样本池,使用(29)式,根据航路样本池中的数据构造训练所用特征向量和标签值。训练过程中均方根误差(RMSE)指标曲线如图 7 所示。由图可知, RMSE 指标随着训练回合数增加而逐渐减小,4 000 次迭代后逐渐收敛。为评价训练效果,在相同场景下对比“指导者”神经网络与高斯伪谱算法航迹优化结果,如图 8 所示。“指导者”神经网络与高斯伪谱算法的优化结果近似。值得注意的是,“指导者”仅为之后的强化学习智能体提供粗略指导,其本身并不需要完全拟合出一致的策略。

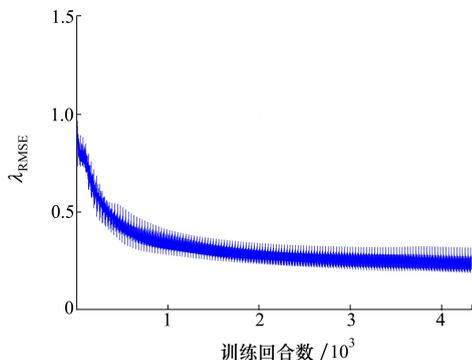


图 7 神经网络拟合高斯伪谱法策略过程中 RMSE 变化曲线

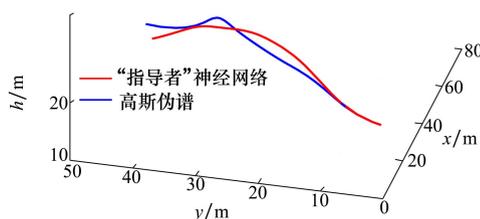


图 8 “指导者”神经网络与高斯伪谱算法航迹优化结果对比

## 4.3 “执行者”网络训练

TD3 算法中动作网络为“执行者”,训练过程中奖励值随训练回合数变化曲线如图 9 所示。由图 9 可知,奖励值逐渐增大,到 3 500 回合时,奖励值接近 3 000。图 9 中红色曲线为滑动平均后的奖励曲线,从奖励趋势来看智能体相较于最开始时有了明显的策略提升。

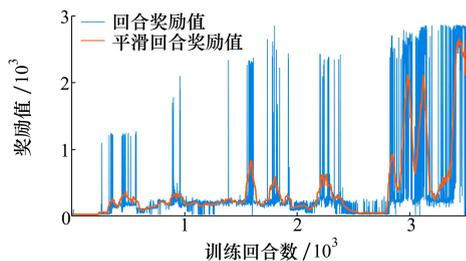


图 9 强化学习奖励值随训练回合数变化曲线

基于强化学习的动态滑翔航迹优化结果如图10所示。强化学习决策过程中产生的基于“指导者”网络的动作偏置、指导者动作和真实动作随时间变化曲线如图11所示。由图11可知,真实动作由指导者动作加上强化学习动作偏置构成,智能体通过在合适的时间点对指导者的基础动作进行修正从而更大程度地从环境中获取能量。

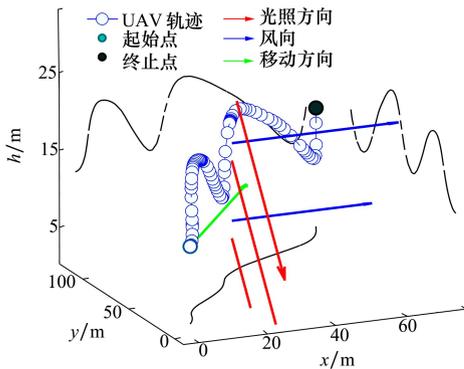


图10 基于强化学习的动态滑翔航迹优化结果

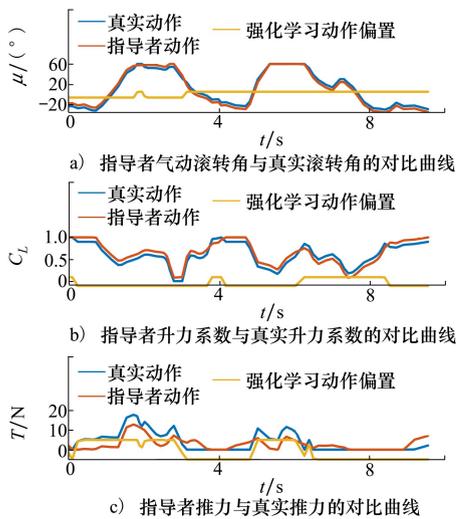


图11 指导者动作、强化学习输出动作与真实动作曲线

### 4.4 对比仿真

本文所提出的 RL 方法和传统 GP 方法在相同场景下动态滑翔平均能量获取功率数据见表5。

表5 基于强化学习的动态滑翔航迹优化方法和高斯伪谱方法平均能量功率对比

方法	$\bar{P}_T/W$	$\bar{P}_D/W$	$\bar{P}_W/W$	$\bar{P}_{sun}/W$	$\bar{P}_{sum}/W$
RL	44.2	39.4	8.4	41.6	-33.6
GP	41.5	35.0	9.7	36.8	-30

由表5可知,RL对应的太阳能平均获取功率较GP提升4.8W,总能量获取功率略小于GP(3.6W)。从能量角度而言,RL方法达到了最优结果的88%,在放弃小部分最优性的前提下,提升推理速度是可以接受的。

对比本文提出的RL方法和传统GP方法在优化相同动态滑翔问题时的决策耗时,结果如图12所示。由图12可知,RL单次优化平均耗时约0.5s,而GP为5.6s。因此本文提出的基于强化学习的优化算法大大降低了优化时长,可用于固定翼无人机动态滑翔在线航迹优化。

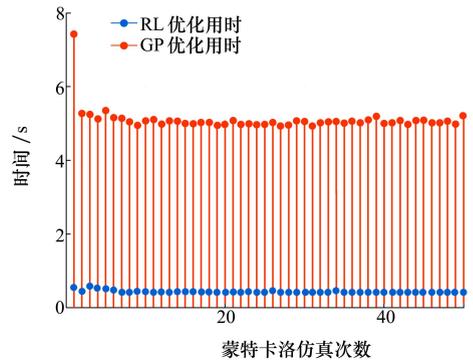


图12 RL和GP优化相同动态滑翔问题耗时蒙特卡洛仿真

为验证风场变化情况下基于RL的方法有更好的适应性,在 $t=2s$ 时将固定高度风速参考值 $V_R$ 从 $8m/s$ 增至 $10\sim 16m/s$ ,风向 $e_w$ 从初始的 $0^\circ$ 改变至 $-20^\circ\sim 20^\circ$ 。此时GP方法无法立即应对变化,只能使用之前离线规划得到的期望航迹进行滑翔,而RL方法采用单步迭代的计算方式,可以立即做出变化。风场变化情况下RL和GP方法单个滑翔周期内获取风能如表6所示。由表6可知,随着 $V_R, e_w$ 的变化,基于RL的方法能适应不同的风场强度从而获取更多的风能,而GP方法面对变化风场时效果不如RL,获能只有较小幅度提升。

表6 风场变化情况下RL和GP方法单个滑翔周期内获取的风能比较

$V_R/(m \cdot s^{-1})$	$e_w/(^\circ)$	GP方法获取风能/J	RL方法获取风能/J
8	0	27.2	24.7
10	0	82.1	455.0
12	0	116.2	592.9
14	0	85.3	903.9

续表 6

$V_R/(m \cdot s^{-1})$	$e_w/(^\circ)$	GP 方法获取 风能/J	RL 方法获取 风能/J
16	0	162.7	921.0
8	-20	53.8	51.4
8	-16	61.8	75.2
8	-12	93.5	73.4
8	-8	96.7	449.6
8	-4	92.1	435.3
8	4	86.5	495.2
8	8	76.0	543.7
8	12	103.9	558.4
8	16	103.2	546.3
8	20	111.2	554.9

## 5 结 论

本文针对多能源利用下固定翼无人机动态滑翔航迹优化问题,借助高斯伪谱法,提出了一种基于深度强化学习的在线优化方法,以应对飞行过程中由于风场变化,导致原有航迹不能充分利用环境能量的问题。仿真结果表明:①在考虑燃料耗能、飞行阻力耗能、梯度风场获能、太阳能获能、障碍物约束的情况下,本文提出的算法能有效求解动态滑翔最优航迹;②“指导者”神经网络可以有效地逼近动态滑翔航迹优化策略,为后续强化学习训练提供了基础;③在标准条件下,传统最优控制算法得到的优化结果与基于强化学习的优化结果在单个滑翔周期相当;④基于强化学习的动态滑翔航迹优化算法的实时性远优于传统最优控制算法,航迹解算耗时减少 91%;⑤应对变化风场,基于强化学习的算法具有更好的适应性,可根据环境情况动态调节策略。

## 参考文献:

- [1] MIR I, EISA S A, TAHA H, et al. A stability perspective of bioinspired unmanned aerial vehicles performing optimal dynamic soaring[J]. *Bioinspiration & Biomimetics*, 2021, 16(6): 066010
- [2] LIU S, BAI J, WANG C. Energy acquisition of a small solar UAV using dynamic soaring[J]. *The Aeronautical Journal*, 2021, 125(1283): 60-86
- [3] 刘多能. 固定翼无人机动态滑翔机理与航迹优化研究[D]. 长沙: 国防科学技术大学, 2016  
LIU Duoneng. Research on mechanism and trajectory optimization for dynamic soaring with fixed-wing unmanned aerial vehicles [D]. Changsha: National University of Defense Technology, 2016 (in Chinese)
- [4] 朱熠, 李继广, 郝向宇. 梯度风场中无人机动态滑翔飞行轨迹优化[J]. *西安航空学院学报*, 2023, 41(5): 8-16  
ZHU Yi, LI Jiguang, HAO Xiangyu. Optimization of dynamic gliding flight trajectory for UAV in gradient wind fields[J]. *Journal of Xi'an Aeronautical Institute*, 2023, 41(5): 8-16 (in Chinese)
- [5] SACHS G P. Maximum travel speed performance of albatrosses and UAVs using dynamic soaring[C]//AIAA Scitech 2019 Forum, 2019: 0568
- [6] MIR I, GUL F, EISA S, et al. On the stability of dynamic soaring: Floquet-based investigation[C]//AIAA Science and Technology Forum and Exposition, 2022: 0882
- [7] ZWENIG A, HONG H, HOLZAPFEL F. Sensitivity analysis of the energy balance of dynamic soaring[J]. *Journal of Physics*, 2023, 2514(1): 012022
- [8] BOWER G C. Boundary layer dynamic soaring for autonomous aircraft: design and validation[D]. Stanford: Stanford University, 2011
- [9] HONG H, ZHENG H, HOLZAPFEL F, et al. Dynamic soaring in unspecified wind shear: a real-time quadratic-programming approach[C]//2019 27th Mediterranean Conference on Control and Automation, 2019: 600-605
- [10] LAWRENCE N R J, SUKKARIEH S. Autonomous exploration of a wind field with a gliding aircraft[J]. *Journal of Guidance, Control, and Dynamics*, 2011, 34(3): 719-733
- [11] ARULKUMARAN K, DEISENROTH M P, BRUNDAGE M, et al. A brief survey of deep reinforcement learning[J]. *Expert Systems with Applications*, 2023, 231: 120495
- [12] LI Z, LANGELAAN J W. Parameterized trajectory planning for dynamic soaring[C]//AIAA Scitech 2020 Forum, 2020: 0856

- [13] REDDY G, WONGNG J, CELANI A, et al. Glider soaring via reinforcement learning in the field[J]. *Nature*, 2018, 562(7726): 236-239
- [14] MONTELLA C. Learning how to soar: steady state autonomous dynamic soaring through reinforcement learning[C]// *AIAA Scitech 2020 Forum*, 2020: 1848
- [15] LIU S, BAI J, WANG C. Energy acquisition of a small solar UAV using dynamic soaring[J]. *The Aeronautical Journal*, 2021, 125(1283): 60-86
- [16] BONNIN V, BÉNARD E, MOSCHETTA J M, et al. Energy-harvesting mechanisms for UAV flight by dynamic soaring[J]. *International Journal of Micro Air Vehicles*, 2015, 7(3): 213-229
- [17] FUJIMOTO S, HOOF H, MEGER D. Addressing function approximation error in actor-critic methods[C]// *International Conference on Machine Learning*, PMLR, 2018: 1587-1596
- [18] 刘思奇, 白俊强. 结合动态滑翔技术的小型太阳能无人机飞行能量变化分析[J]. *西北工业大学学报*, 2020, 38(1): 48-57  
LIU Siqi, BAI Junqiang. Analysis of flight energy variation of small solar UAVs using dynamic soaring technology[J]. *Journal of Northwestern Polytechnical University*, 2020, 38(1): 48-57 (in Chinese)
- [19] BENCATEL R, DE SOUSA J T, GIRARD A. Atmospheric flow field models applicable for aircraft endurance extension[J]. *Progress in Aerospace Sciences*, 2013, 61: 1-25
- [20] FLANZER T, BUNGE R, KROO I. Efficient six degree of freedom aircraft trajectory optimization with application to dynamic soaring[C]// *12th AIAA Aviation Technology, Integration, and Operations Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2012: 5622
- [21] 刘思奇, 白俊强. 基于六自由度模型的高空动态滑翔探究[J]. *西北工业大学学报*, 2021, 39(4): 703-711  
LIU Siqi, BAI Junqiang. Exploration of high-altitude dynamic soaring based on six-degree-of-freedom model[J]. *Journal of Northwestern Polytechnical University*, 2021, 39(4): 703-711 (in Chinese)
- [22] 马东立, 包文卓, 乔宇航. 基于重力储能的太阳能飞机飞行轨迹研究[J]. *航空学报*, 2014, 35(2): 408-416  
MA Dongli, BAO Wenzhuo, QIAO Yuhang. Study of flight path for solar-powered aircraft based on gravity energy reservation[J]. *Acta Aeronauticae Astronautica Sinica*, 2014, 35(2): 408-416 (in Chinese)
- [23] 朱炳杰. 无人机风梯度动态滑翔机理与航迹优化研究[D]. 长沙: 国防科学技术大学, 2016  
ZHU Bingjie. Research on mechanism and trajectory optimization for unmanned aerial vehicles by dynamic soaring in gradient wind[D]. Changsha: National University of Defense Technology, 2016 (in Chinese)
- [24] MONTELLA C, SPLETZER J R. Reinforcement learning for autonomous dynamic soaring in shear winds[C]// *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014: 3423-3428

# Multi energy dynamic soaring trajectory optimization method based on reinforcement learning

ZHANG Yunfei<sup>1,2</sup>, WANG Honglun<sup>1,2</sup>, ZHANG Menghua<sup>1,2</sup>, GONG Yinan<sup>3</sup>

(1.School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;  
2.The Science and Technology on Aircraft Control Laboratory, Beihang University, Beijing 100191, China;  
3.Hiwing Aviation General Equipment Co., Ltd., Beijing 100074, China)

**Abstract:** In addressing the issue of dynamic soaring in unmanned aerial vehicles, a trajectory optimization approach based on deep reinforcement learning is proposed. This method synergistically utilizes gradient wind energy and solar energy and incorporates obstacle constraints to simulate complex barrier environments. It employs neural networks to approximate the Gaussian pseudospectral method for solving trajectory policies. On the foundation of the trained policies, the method utilizes the twin delayed deep deterministic policy gradient algorithm for policy enhancement. This significantly boosts the real-time inference capabilities while addressing the challenges traditional optimal control algorithms face in dynamic soaring due to varying wind fields. The experiments initially validate the approach through simulation of two classic modes of dynamic soaring, followed by Monte Carlo simulations considering multiple energy sources. The results indicate that the dynamic soaring trajectory optimization method based on deep reinforcement learning achieves energy acquisition comparable to optimal outcomes within a single soaring cycle, with a 91% reduction in real-time inference decision time. Moreover, in changing wind field environments, this method demonstrates superior adaptability compared to traditional approaches.

**Keywords:** dynamic soaring; reinforcement learning; Gaussian pseudospectral method; trajectory optimization

**引用格式:**张云飞,王宏伦,张梦华,等.基于强化学习的多能源动态滑翔航迹优化方法[J].西北工业大学学报,2025,43(1):128-139

ZHANG Yunfei, WANG Honglun, ZHANG Menghua, et al. Multi energy dynamic soaring trajectory optimization method based on reinforcement learning[J]. Journal of Northwestern Polytechnical University, 2025,43(1):128-139 (in Chinese)