

文章编号: 2097-1974(2025)01-0057-08

DOI: 10.7654/j.issn.2097-1974.20250108

# 基于DDPG-LQR的高超声速飞行器时间协同再入制导

宋志飞<sup>1,2</sup>, 吉月辉<sup>1,2</sup>, 宋雨<sup>1,2</sup>, 刘俊杰<sup>1,2</sup>, 高强<sup>1,2</sup>

(1. 天津理工大学电气工程与自动化学院, 天津, 300384; 2. 天津市复杂系统控制理论与应用重点实验室, 天津, 300384)

**摘要:** 针对多高超声速飞行器协同作战的特点, 提出一种基于深度策略性梯度和线性二次型调节器 (Deep Deterministic Policy Gradient-Linear Quadratic Regulator, DDPG-LQR) 的时间协同再入制导方案。首先, 采用序列凸优化方法生成满足多个约束的时间协同再入轨迹及其相应的稳态控制量, 并且采用Radau伪谱法离散运动学方程, 以提高轨迹优化离散精度。其次, 采用线性二次型调节器 (Linear Quadratic Regulator, LQR) 跟踪时间协同再入轨迹。为了提高协同制导精度和制导效果, 采用深度策略性梯度 (Deep Deterministic Policy Gradient, DDPG) 在线优化LQR的权重矩阵系数。在DDPG算法中, 通过引入合适的奖励函数来提高算法的优化性能。仿真结果表明, 在初始状态误差和不确定性的情况下, 通过传统的LQR控制器相比, 本文所提出的协同制导方案具有更好的协同制导精度和制导效果。

**关键词:** 多高超声速飞行器; 协同制导; 序列凸优化; 深度策略性梯度; 线性二次型调节器

中图分类号: V412

文献标识码: A

## Time Cooperative Reentry Guidance for Hypersonic Vehicle based on DDPG-LQR

SONG Zhifei<sup>1,2</sup>, JI Yuehui<sup>1,2</sup>, SONG Yu<sup>1,2</sup>, LIU Junjie<sup>1,2</sup>, GAO Qiang<sup>1,2</sup>

(1. School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin, 300384;

2. Tianjin Key Laboratory of Control Theory and Application for Complex Systems, Tianjin, 300384)

**Abstract:** Aiming at the characteristics of multiple hypersonic vehicles cooperative combat, a time cooperative reentry guidance scheme based on deep deterministic policy gradient and linear quadratic regulator (DDPG-LQR) is proposed. Firstly, the sequential convex programming method is used to generate the time cooperative reentry trajectory satisfying multiple constraints and its corresponding steady-state control quantity. The Radau pseudospectral method is used to discretize the motion equations to improve the discretization accuracy of trajectory optimization. Secondly, the linear quadratic regulator (LQR) is used to track the time cooperative reentry trajectory. In order to improve the cooperative guidance accuracy and guidance effect, the deep deterministic policy gradient (DDPG) is used to optimize the weight matrix coefficients of the LQR online. In the DDPG algorithm, the optimization performance of the algorithm is improved by introducing an appropriate reward function. The simulation results show that the cooperative guidance scheme proposed has better cooperative guidance effect and guidance accuracy than the traditional LQR controller in the case of initial state error and uncertainty.

**Keywords:** multiple hypersonic vehicles; cooperative guidance; sequential convex programming; deep deterministic policy gradient; linear quadratic regulator

## 0 引言

随着反制导武器防御系统的不断完善和发展, 单个飞行器的生存能力和作战能力下降。因此, 为了抵御威胁和提高打击效率, 多高超声速飞行器协同制导引起了国内外学者的广泛研究<sup>[1-3]</sup>。对于高超声速飞行器的整个飞行阶段而言, 再入段飞行时间要远大于

末制导段, 是决定最终饱和和打击效果的关键阶段<sup>[4]</sup>。在多飞行器协同再入制导问题的研究中, 目前的主要研究方法有预测校正制导、标称轨迹跟踪制导等<sup>[5]</sup>。预测校正制导方法是指采用数值积分和解析表达式来预测给定控制变量下的终端状态, 并通过计算预测终端状态值与设定终端状态值的偏差来调整控制变

收稿日期: 2024-05-14; 修回日期: 2024-08-21

基金项目: 国家自然科学基金(No.62203331); 天津理工大学研究生教育教学研究与改革项目(ZDXM2202, YBXM2204)

量<sup>[6]</sup>。标称轨迹跟踪制导是指采用优化方法生成满足多种约束条件下的标称轨迹,然后通过制导控制系统跟踪标称轨迹,以减少实际飞行轨迹与标称轨迹的误差<sup>[7]</sup>。标称轨迹跟踪制导计算负担小且具有较好的稳定性,在时间较长的飞行任务中具有较高的可靠性,能够为飞行器提供稳定的控制量。协同轨迹优化问题的研究,主要采用分布式方法生成协同轨迹<sup>[8]</sup>。Yu等<sup>[9]</sup>针对再入滑翔阶段,提出了一种两阶段协同轨迹优化方法,在第一阶段优化倾侧角指令实现攻角协同,第二阶段优化攻角指令实现时间协同。Zhang等<sup>[10]</sup>和Li等<sup>[11]</sup>针对高超声速飞行器的时间协同问题,采用深度神经网络来训练轨迹数据集,进而获得协同打击时间。最后,将神经网络训练得到的协同打击时间作为轨迹优化问题的约束条件,采用轨迹优化算法生成时间协同再入轨迹。

近些年出现了诸多轨迹跟踪制导方法的相关研究。Vepa<sup>[12]</sup>针对航天器的轨迹跟踪问题,设计了一种反馈跟踪控制律,通过将非线性运动跟踪控制律简化为线性反馈控制律来跟踪航天器的参考轨迹。Ou等<sup>[13]</sup>针对跟踪系统中存在的不确定性,将扩展卡尔曼滤波参数校正策略与模型预测静态规划算法相结合,提出了一种自适应轨迹跟踪制导方案。杨美等<sup>[14]</sup>针对存在风扰情况时,实际轨迹和标称轨迹的偏差过大,超出LQR控制器所能控制的范围的情况,提出了一种在线再入制导方法。针对协同制导问题,李征等<sup>[15]</sup>对可重复使用航天器的时间协同再入制导进行了深入研究,通过采用伪谱法和滚动时域控制策略有效提高了协同制导效果。Xu等<sup>[16]</sup>提出一种基于参数优化和时间解析的协同制导方案,该方法可以快速求解制导指令,且满足制导要求。目前,强化学习得到了广泛的应用,可以在系统状态和控制器参数之间建立直接联系,具有较好的优化效果。Zhao等<sup>[17]</sup>采用DDPG算法对四旋翼无人机的自抗扰控制器进行参数优化,该算法在调节自抗扰控制器的控制精度和收敛速度方面具有优势。Li等<sup>[18]</sup>针对控制器参数繁琐问题,提出了一种改进的DDPG算法在线调整控制器参数,提高控制的鲁棒性和精度。

针对上述分析,本文提出了一种基于深度策略性梯度和线性二次型调节器(Deep Deterministic Policy Gradient-Linear Quadratic Regulator, DDPG-LQR)的多高超声速飞行器时间协同再入轨迹跟踪制导方案。首先,建立多高超声速飞行器再入运动学模型,采用Radau伪谱法和序列凸优化方法(Radau Sequential

Convex Programming, RSCP)生成时间协同再入轨迹。然后,利用线性二次型调节器(Linear Quadratic Regulator, LQR)轨迹跟踪控制器生成最优反馈控制量,并结合轨迹优化阶段的参考控制量实现协同轨迹跟踪制导。最后,为了提高协同制导精度和制导效果,采用DDPG算法在线优化调节多个LQR轨迹跟踪控制器的状态权重矩阵和控制权重矩阵系数。最后仿真结果验证了本文所提方案的有效性。

## 1 问题描述

### 1.1 再入运动学模型

考虑地球是旋转的球体,忽略地球自转的影响,以时间为自变量的多高超声速飞行器三自由度无量纲运动学方程为

$$\begin{cases} \dot{r}_i = V_i \sin \gamma_i \\ \dot{\theta}_i = V_i \cos \gamma_i \sin \psi_i / r_i \cos \phi_i \\ \dot{\phi}_i = V_i \cos \gamma_i \cos \psi_i / r_i \\ \dot{V}_i = -D_i - \sin \gamma_i / r_i^2 \\ \dot{\gamma}_i = L_i \cos \sigma_i / V_i - \cos \gamma_i / r_i^2 V_i + V_i \cos \gamma_i / r_i \\ \dot{\psi}_i = L_i \sin \sigma_i / V_i \cos \gamma_i + V_i \cos \gamma_i \sin \psi_i \tan \phi_i / r_i \end{cases} \quad (1)$$

式中 下标*i*代表第*i*个飞行器;*r*为飞行器到地心的径向距离,无量纲化参数为地球半径 $R_0$ ;  $\theta$ 和 $\phi$ 分别为经度和纬度;*V*为飞行器的飞行速度,其无量纲化参数为 $V_c = \sqrt{g_0 R_0}$ ;  $\gamma$ 为航迹倾角;  $\psi$ 为航迹偏角;  $\sigma$ 为倾侧角。*L*和*D*分别为无量纲化后的升力和阻力加速度,表达式为

$$\begin{cases} L_i = R_0 \rho_i V_i^2 A_{\text{ref},i} C_{L,i} / 2m_i \\ D_i = R_0 \rho_i V_i^2 A_{\text{ref},i} C_{D,i} / 2m_i \end{cases} \quad (2)$$

式中  $\rho$ 为大气密度; *m*,  $A_{\text{ref}}$ 分别为飞行器质量和特征参考面积;  $C_L$ ,  $C_D$ 分别为升力系数和阻力系数。

### 1.2 再入约束

对于高超声速飞行器再入阶段,需要严格满足多种约束条件来保证飞行器的稳定飞行,再入过程中约束主要包括路径约束、边界约束以及初始和终端状态约束。其中,典型的路径约束主要包括热流密度 $\dot{Q}$ 、动压*q*和过载*n*,表达式为

$$\begin{cases} \dot{Q}_i = k_q \rho_i^{0.5} (V_c V_i)^{3.15} \leq \dot{Q}_{\text{max}} \\ q_i = \rho_i (V_c V_i)^2 / 2 \leq q_{\text{max}} \\ n_i = \sqrt{L_i^2 + D_i^2} \leq n_{\text{max}} \end{cases} \quad (3)$$

式中  $k_q$ 为与飞行器结构相关的热流密度系数;  $V_i$ 为第*i*个飞行器的速度。为了抑制倾侧角 $\sigma$ 的高频抖动,本文选择倾侧角速率 $\dot{\sigma}$ 作为轨迹优化环节的控制变量。

倾侧角和倾侧角速率满足的约束为

$$\sigma_{i,\min} \leq \sigma_i \leq \sigma_{i,\max}, \dot{\sigma}_{i,\min} \leq \dot{\sigma}_i \leq \dot{\sigma}_{i,\max} \quad (4)$$

另外，飞行器的初始状态和终端状态约束为

$$x_i(t_{0,i}) = x_{0,i}, x_i(t_{f,i}) = x_{f,i} \quad (5)$$

式中  $t_{0,i}$ ,  $t_{f,i}$  分别为每个飞行器的初始飞行时间和到达期望目标点的时间。为了实现时间协同再入制导，设置每个飞行器的终端时间约束为

$$t_{f,1} = t_{f,2} = \dots = t_{f,i} = t_f \quad (6)$$

式中  $t_f$  为协同飞行时间。

## 2 DDPG-LQR 协同再入轨迹跟踪制导

### 2.1 基于RSCP的时间协同再入轨迹优化

DDPG-LQR 时间协同再入制导的整体框架如图1所示。

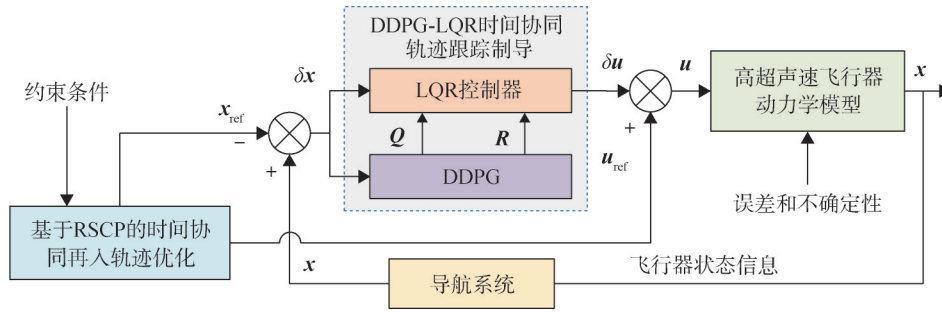


图1 基于DDPG-LQR的时间协同再入制导整体框架

Fig.1 The overall framework of time cooperative reentry guidance based on DDPG-LQR

在接下来的描述中为了方便表述省略了下标  $i$ 。在多高超声速飞行器轨迹优化的问题研究中，以时间为自变量的飞行器运动学方程，需要提前给定末端飞行时间，一般为末端时间固定的轨迹优化问题。然而，在求解多个飞行器的公共飞行时间时，目标函数需要设置为最小飞行时间或最大飞行时间，导致问题复杂且求解难度较大。因此，为了方便求解多个飞行器公共飞行时间，将飞行器运动学方程中的自变量转变为纵向航程角  $s$ ，将时间变量扩展为状态变量，进而将时间固定的轨迹优化问题转变为纵向航程角固定的轨迹优化问题，实现自由的末端飞行时间。纵向航程角的初始值  $s'_0 = 0$ ，末端值表达式为

$$s'_f = \arccos(\sin \phi_0 \sin \phi_f + \cos \phi_0 \cos \phi_0 \cos(\theta_f - \theta_0)) \quad (7)$$

式中  $\theta_f$ ,  $\phi_f$  分别为期望目标点的经度和纬度； $\theta_0$  为初始位置经度； $\phi_0$  为初始位置纬度。接下来，将倾侧角扩展为状态变量，则扩展后的运动学方程的控制仿射形式为

$$\dot{x} = F(x) + B(x)u \quad (8)$$

式中  $x$  为扩展后的状态变量， $x = [r, \theta, \phi, V, \gamma, \psi, \sigma, t]^T$ ； $u$  为控制变量， $u = \dot{\sigma}$ ；状态矩阵  $F(x)$  和控制矩阵  $B(x)$  分别为

$$F(x) = \begin{bmatrix} r \tan \gamma \\ \sin \psi / \cos \phi \\ \cos \psi \\ -rD/V \cos \gamma - \tan \gamma / rV \\ rL \cos \sigma / V^2 \cos \gamma - 1/r^2 V + 1 \\ rL \sin \sigma / V^2 \cos^2 \gamma + \sin \psi \tan \phi \\ 0 \\ r/V \cos \gamma \end{bmatrix} \quad (9)$$

$$B(x) = [0, 0, 0, 0, 0, 0, r/V \cos \gamma, 0]^T \quad (10)$$

在凸优化算法求解中通常将非凸约束进行线性化，然后将线性化后的凸子问题迭代求解。高超声速飞行器的再入轨迹优化问题中，非线性约束主要是指运动学方程和路径约束的非线性。首先，采用泰勒展开方法将运动学方程在参考轨迹  $x^k$  处线性化：

$$\begin{cases} \dot{x} = A(x^k)x + B(x^k)u + C(x^k) \\ A(x^k) = \frac{\partial F(x)}{\partial x} \Big|_{x=x^k} \\ C(x^k) = F(x^k) - A(x^k)x^k - B(x^k)u^k \end{cases} \quad (11)$$

同理，路径约束线性化后的表达式为

$$\begin{cases} \dot{Q}(r^k, V^k) + \frac{\partial \dot{Q}}{\partial r}(r_{x=x^k} r^k) + \frac{\partial \dot{Q}}{\partial V}(V_{x=x^k} V^k) \leq \dot{Q}_{\max} \\ q(r^k, V^k) + \frac{\partial q}{\partial r}(r_{x=x^k} r^k) + \frac{\partial q}{\partial V}(V_{x=x^k} V^k) \leq q_{\max} \\ n(r^k, V^k) + \frac{\partial n}{\partial r}(r_{x=x^k} r^k) + \frac{\partial n}{\partial V}(V_{x=x^k} V^k) \leq n_{\max} \end{cases} \quad (12)$$

为了提高轨迹优化问题的离散精度，采用Radau伪谱法进行离散化。由于该方法相对成熟，具体离散

过程见参考文献 [19]。Radau 伪谱法离散后的飞行器状态方程等式约束为

$$\sum_{i=0}^{N-1} \mathbf{D}_{ki} \mathbf{x} = \frac{s_f}{2} [\mathbf{A}(\mathbf{x}^k) \mathbf{x} + \mathbf{B}(\mathbf{x}^k) \mathbf{u} + \mathbf{C}(\mathbf{x}^k) + \mathbf{v}] \quad (13)$$

式中  $\mathbf{D}_{ki}$  为伪谱微分矩阵;  $\mathbf{v}$  为补偿项, 用来补偿线性化时所带来的误差。在轨迹优化阶段, 定义目标函数为末端点和期望目标点的误差, 为了提高算法的收敛速度并且使补偿项趋近于零, 在目标函数中引入正则化项和惩罚项, 表达式为

$$J = p_1 |(\theta_n - \theta_f) + (\phi_n - \phi_f)| + p_2 |r_n - r_f| + w \|\mathbf{v}\|_1 \quad (14)$$

式中  $p_1$  和  $p_2$  为正则化项系数;  $w$  为惩罚项系数。此外, 本文采用连续迭代求解思想, 逐次求解时间协同再入轨迹优化问题。首先采用四阶龙格库塔法求解初始解, 将初始解带入 RSCP 算法中连续迭代求解, 当求解结果满足所设定的收敛条件时, 算法循环迭代结束, 输出最优解。

在高超声速飞行器的再入滑翔阶段, 飞行器处于无动力滑翔状态, 飞行器的时间调节范围有限, 为了保证时间协同的有效性, 必须确定多个飞行器可行的飞行时间范围。因此, 采用 RSCP 算法首先计算出每个飞行器抵达期望目标点的最长飞行时间和最短飞行时间, 将最长飞行时间 ( $J = -t_f$ ) 或最短飞行时间 ( $J = t_f$ ) 作为目标函数, 期望目标点的位置信息作为终端状态约束。最后, 计算出所有飞行器的抵达期望目标点可行公共飞行时间范围, 根据实际飞行任务需求设计协同飞行时间。

## 2.2 LQR 轨迹跟踪设计

为了获得更好的轨迹跟踪制导效果, 在轨迹跟踪制导环节, 选择攻角  $\alpha$  和倾侧角  $\sigma$  作为控制量。首先, 将飞行器的运动学方程关于标称轨迹进行线性化, 所得到的误差方程为

$$\delta \dot{\mathbf{x}} = \mathbf{A}'(s') \delta \mathbf{x} + \mathbf{B}'(s') \delta \mathbf{u} \quad (15)$$

$$\mathbf{y} = \mathbf{C}'(s') \delta \mathbf{x} \quad (16)$$

式中 状态量误差  $\delta \mathbf{x} = [\delta r, \delta \theta, \delta \phi, \delta V, \delta \gamma, \delta \psi]^T$ ; 反馈控制量  $\delta \mathbf{u} = [\delta \alpha, \delta \sigma]^T$ ;  $\mathbf{A}'(s')$  为系统状态矩阵;  $\mathbf{B}'(s')$  为控制矩阵;  $\mathbf{C}'(s')$  为输出矩阵。轨迹跟踪控制系统的二次型性能指标为

$$J = \frac{1}{2} \int_{s_0'}^{s_f'} [\delta \mathbf{x}^T \mathbf{Q} \delta \mathbf{x} + \delta \mathbf{u}^T \mathbf{R} \delta \mathbf{u}] ds' \quad (17)$$

式中 状态权重矩阵  $\mathbf{Q} = \text{diag}(Q_1, Q_2, Q_3, Q_4, Q_5, Q_6)$ ; 控制权重矩阵  $\mathbf{R} = \text{diag}(R_1', R_2')$ 。权重矩阵系数的选择会很大程度影响轨迹跟踪精度和跟踪效果。在传统的 LQR 控制器设计中通常采用 “Bryson's Rule” 方法来

设置权重矩阵系数, 然而该方法提供的权重矩阵系数和控制矩阵系数在飞行器的整个飞行控制过程中为恒定值, 实际再入飞行任务往往具有不确定性、非线性或时变性, 这些因素可能导致 “Bryson's Rule” 设定的权重矩阵使得控制器无法达到最优控制精度。因此, 本文提出一种 DDPG 算法在线优化  $\mathbf{Q}$  矩阵和  $\mathbf{R}$  矩阵的权重系数, 以获得最优的权重矩阵, 进而提高控制器的跟踪效果和精度。DDPG 算法的设计过程将在 2.3 节给出。满足最小性能指标的反馈控制量设计如下:

$$\delta \mathbf{u} = -\mathbf{K}(s) \delta \mathbf{x} \quad (18)$$

式中 反馈增益矩阵  $\mathbf{K}(s) = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{P}$ ; 其中  $\mathbf{P}$  通过求解黎卡提方程获得, 如下所示:

$$\dot{\mathbf{P}} = \mathbf{A}^T \mathbf{P} - \mathbf{P} \mathbf{A} - \mathbf{Q} + \mathbf{P} \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{P} \quad (19)$$

因此, 增加反馈控制量后的控制变量为

$$\mathbf{u}' = \mathbf{u}_{\text{ref}} + \delta \mathbf{u} \quad (20)$$

式中  $\mathbf{u}_{\text{ref}}$  为轨迹优化阶段的参考控制量。

## 2.3 深度策略性梯度设计

DDPG 算法是一种用于处理连续动作空间问题的算法。可以将 DDPG 对 LQR 控制器的权重矩阵系数的优化过程看作马尔可夫决策过程。这个过程可以被视为  $(S, A, P', \gamma, R)$ , 其中  $S$  为状态空间,  $A$  为行动空间,  $P'$  为状态转移概率,  $\gamma$  为折现因子,  $R$  为奖励函数。强化学习目的是通过智能体与环境之间的连续交互策略, 最大化累计折扣奖励  $G_t$ , 表达式如下:

$$G_t = \sum_{i=t}^T \gamma^i r'_i, \quad 0 < \gamma < 1 \quad (21)$$

强化学习框架如图 2 所示。在每个时间步, 智能体基于当前状态  $s_t$  做出决策, 根据最优策略  $\pi^*$  执行动作  $a_t$ 。通过智能体和环境的交互, 获得下一个奖励  $r'_{t+1}$  和状态  $s_{t+1}$ 。将得到奖励  $r'$  累加累计奖励  $G_t$ , 通过连续训练以最大化累计奖励  $G_t$ 。

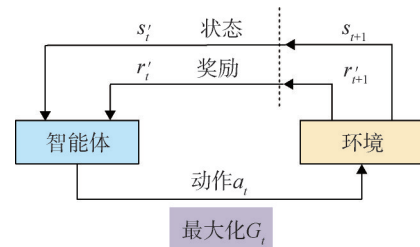


图 2 强化学习框架

Fig.2 Reinforcement learning framework

强化学习中, 奖励函数的设计是关键因素, 对控制效果有很大的影响。在本设计中, 奖励函数设计为

$$R = \begin{cases} 10, & |\delta r| \leq 300\text{m}, |\delta \theta| \leq 0.02^\circ, |\delta \phi| \leq 0.02^\circ, |\delta t| \leq 5\text{s} \\ -50, & |\delta r| > 2000\text{m}, |\delta \theta| > 2^\circ, |\delta \phi| > 2^\circ, |\delta t| > 50\text{s} \\ -2, & \text{其他} \end{cases} \quad (22)$$

式中  $|\delta r|$ 为高度误差绝对值； $|\delta \theta|$ 为经度误差绝对值； $|\delta \phi|$ 为纬度误差绝对值； $|\delta t|$ 为时间误差绝对值。这样设计是为了获得更高的再入制导精度和较小的协同时间偏差，没有引入航迹倾角和航机偏角以及飞行速度是因为再入飞行过程中的不确定性因素会引起这些实际状态量与参考状态量偏差较大，进而导致DDPG训练奖励不收敛。强化学习中动作  $a_t$  包括每个LQR控制器的  $Q$ 和  $R$ 的权重矩阵系数。同时为了保证权重矩阵的正定性，权重矩阵系数的取值范围均大于零。

损失函数  $L$  设计如下：

$$\begin{cases} L = \frac{1}{N} \sum_i (Q(s_t, a_t | \theta^e) - y_i)^2 \\ y_i = r_i + \gamma Q(s_{t+1}, \mu(s_{t+1} | \theta^\mu) | \theta^e) \end{cases} \quad (23)$$

式中  $\theta^e$ 为评价网络权重； $\theta^\mu$ 为目标评价网络权重； $\mu(s_{t+1} | \theta^\mu)$ 为策略； $\theta^\mu$ 为目标执行网络权重。通过最小化损失函数，训练评价网络权重。通过梯度，执行网络的权值被更新。策略梯度设计如下：

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a | \theta^e) \Big|_{s=s_t, a=\mu(s_t)} \nabla_{\theta^\mu} \mu(s | \theta^\mu) \Big|_{s_t} \quad (24)$$

式中  $\theta^\mu$ 为执行网络权重。为了提高算法的稳定性，引入了目标评价网络和目标执行网络，分别与评价网络和执行网络具有相同的网络结构。目标网络权重更新设计如下：

$$\begin{cases} \theta^e \leftarrow \tau \theta^e + (1 - \tau) \theta^{e'} \\ \theta^\mu \leftarrow \tau \theta^\mu + (1 - \tau) \theta^{\mu'} \end{cases} \quad (25)$$

式中  $\tau$ 为学习率。执行网络和评价网络的激活函数设置为  $\tanh$ ，即双曲正切函数。每个隐藏层神经单元设置为256个。

### 3 数值仿真

本文选取3个高超声速飞行器进行再入协同轨迹跟踪制导仿真。假设飞行器具有相同的结构和气动参数，飞行器质量  $m=104\ 305\ \text{kg}$ ，特征参考面积  $A_{\text{ref}}=391.22\ \text{m}^2$ ，期望目标点的经纬度为  $(151^\circ, 41^\circ)$ ，期望目标高度为  $35\ \text{km}$ 。升力和阻力系数表示为

$$\begin{cases} C_L = -0.041065 + 0.016292\alpha + 0.00026024\alpha^2 \\ C_D = 0.080505 - 0.03026C_L + 0.86495C_L^2 \end{cases} \quad (26)$$

#### 3.1 基于RSCP时间协同再入轨迹优化

本节设置3个飞行器的初始速度均为  $7\ 450\ \text{m/s}$ ，

初始高度均为  $100\ \text{km}$ 。根据时间协同策略，将协同时间约束设置为  $1\ 616\ \text{s}$ 。飞行器其他仿真初始条件如表1所示。仿真环境为MATLAB 2022b，Intel Core i5处理器。采用CVX工具箱的MOSEK求解器，利用原始对偶内点法求解。图3和图4为基于RSCP算法的时间协同再入轨迹仿真结果。由图3可知，3个飞行器在不同的初始状态下均能抵达期望的目标点，满足设定的初始状态和终端状态约束。由图4可知，倾侧角在所设定的约束范围内变化，并且满足协同飞行时间约束。同时，轨迹优化阶段得到的倾侧角变量为协同制导阶段提供了参考控制变量。

表1 轨迹优化仿真初始条件

Tab.1 Initial conditions of trajectory optimization simulation

单位:  $^\circ$

飞行器编号	经度	纬度	航迹偏角
编号1	68	18	51
编号2	73	12	50
编号3	81	3	48

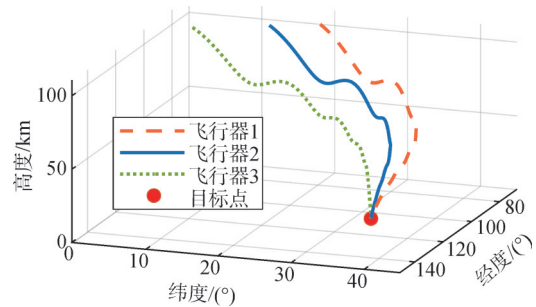


图3 三维时间协同再入参考轨迹

Fig.3 Three-dimensional time-cooperative reentry reference trajectory

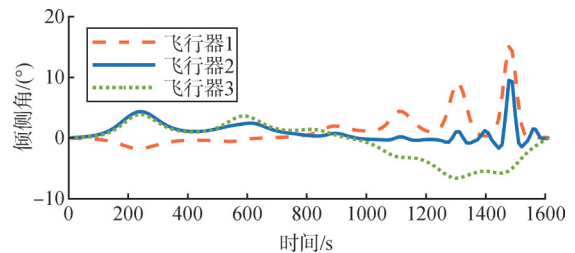


图4 倾侧角-时间剖面

Fig.4 Bank angle-time profile

#### 3.2 基于DDPG-LQR的轨迹跟踪制导

DDPG算法适用于解决连续动作空间的问题。在整个协同轨迹跟踪过程中，DDPG通过与轨迹跟踪的状态误差（环境）和状态权重矩阵系数、控制权重矩阵系数以及奖励的连续交互（智能体），根据系统状态输出最优的LQR控制器的状态权重矩阵系数和控

制权重矩阵系数,进而实现对LQR控制器的实时调参,保证状态权重矩阵系数和控制权重矩阵系数在整个控制过程中是时变的。本文DDPG算法的采样时间为10 ms。同时,为了验证DDPG-LQR协同制导方案的有效性,引入初始状态误差、气动系数误差、大气密度误差以及飞行器参数误差等多个误差项,并且与传统LQR轨迹跟踪控制器进行仿真对比。同时,设计了2组仿真条件,第1组仿真为正向扰动情况下的协同轨迹跟踪制导,第2组仿真为负向扰动情况下的协同轨迹跟踪制导。设置正向扰动误差条件分别为:飞行器初始高度均增加500 m,初始速度均增加50 m/s,升力系数、阻力系数、大气密度以及飞行器质量分别增加5%。设置负向扰动误差条件分别为:飞行器初始高度均减少500 m,初始速度均减少50 m/s,升力系数、阻力系数、大气密度以及飞行器质量分别减少5%。正向扰动和负向扰动情况下的DDPG-LQR协同轨迹跟踪制导仿真结果如图5和图6所示。

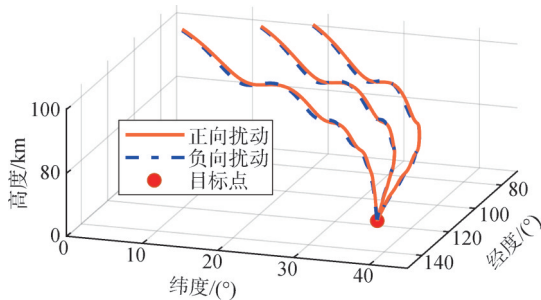


图5 正、负扰动下的DDPG-LQR三维协同再入轨迹

Fig.5 DDPG-LQR three-dimensional cooperative reentry trajectory under positive and negative disturbances

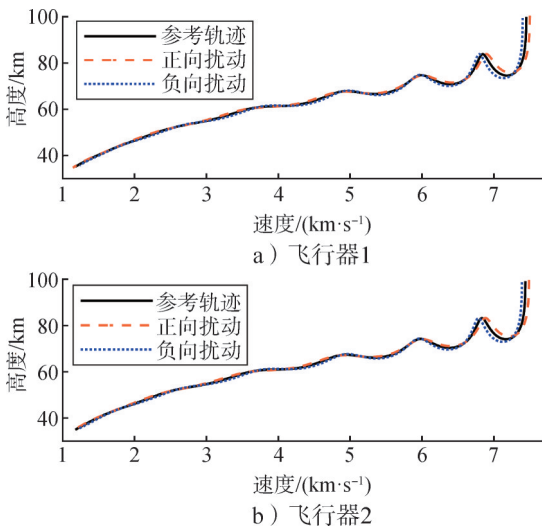
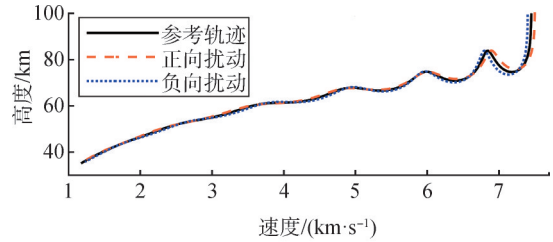


图6 DDPG-LQR协同再入轨迹速度-高度剖面

Fig.6 DDPG-LQR cooperative reentry trajectory velocity-height profile



c) 飞行器3

续图6

由图5和图6可知,在存在初始状态误差和不确定性的情况下,轨迹起始阶段误差较大。在控制器反馈控制的作用下,轨迹末阶段趋于相似。图7为在正向扰动和负向扰动情况下,LQR控制器的协同轨迹跟踪制导三维图。表2为正向扰动情况下协同制导的终端状态误差。表3为负向扰动情况下再入协同制导的终端状态误差。由表2和表3可知,在两种扰动的情况下,通过与传统LQR控制器对比,DDPG-LQR再入协同制导策略的制导精度明显高于传统LQR控制器。其中,正向扰动和负向扰动情况下经、纬度误差均减少约60%~85%;高度误差分别减少约70%和50%;两种扰动情况下的协同制导时间误差均在5 s以内。由于传统LQR控制器中的状态权重矩阵系数和控制权重矩阵系数为恒定值,因此跟踪精度较差。在DDPG-LQR控制器中通过DDPG算法与状态误差的连续交互,状态权重矩阵系数和控制权重矩阵系数根据实际系统状态进行改变,其系数是时变的,进而使得控制器获得更高的轨迹跟踪精度和协同制导效果。

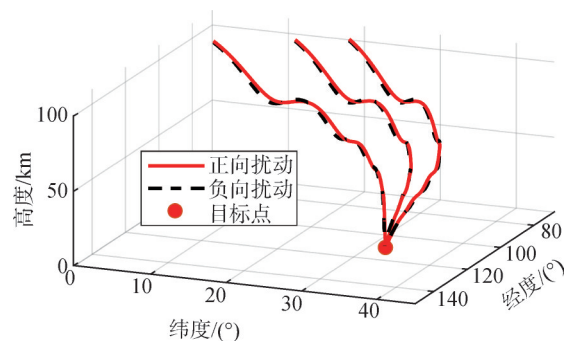


图7 正、负扰动下的LQR三维协同再入轨迹

Fig.7 LQR three-dimensional cooperative reentry trajectory under positive and negative disturbances

表2 正向扰动情况下各飞行器终端状态误差

Tab.2 The terminal state error of each vehicle under positive disturbance

飞行器编号	控制器	经度/(°)	纬度/(°)	高度/m	时间/s
1	DDPG-LQR	0.002	0.006	271	3.2
	LQR	0.012	0.035	1186	25.2
2	DDPG-LQR	0.002	0.019	377	2.7
	LQR	0.005	0.036	1255	28.7
3	DDPG-LQR	0.037	0.025	339	3.3
	LQR	0.103	0.064	1209	21.6

表3 负向扰动情况下各飞行器终端状态误差

Tab.3 Terminal state error of each vehicle under negative disturbance

飞行器编号	控制器	经度/(°)	纬度/(°)	高度/m	时间/s
1	DDPG-LQR	0.003	0.002	679	4.5
	LQR	0.008	0.007	1419	22.1
2	DDPG-LQR	0.003	0.031	661	3.1
	LQR	0.017	0.131	1497	19.7
3	DDPG-LQR	0.041	0.013	597	4.1
	LQR	0.135	0.037	1255	29.9

由于本文采用3个飞行器进行协同制导仿真，单个飞行器的LQR控制器的状态权重矩阵系数为6个，控制权重矩阵系数为2个。因此，DDPG算法在线调节的参数为24个。图8为DDPG训练奖励值，由图8可以看出，正向扰动情况下DDPG实际奖励曲线在经过约180次迭代训练后趋于收敛；负向扰动情况下DDPG实际奖励曲线在经过约300次迭代训练后趋于收敛，所提出的方法可以在训练的后期获得较高的奖励值，得到较好的轨迹跟踪效果。

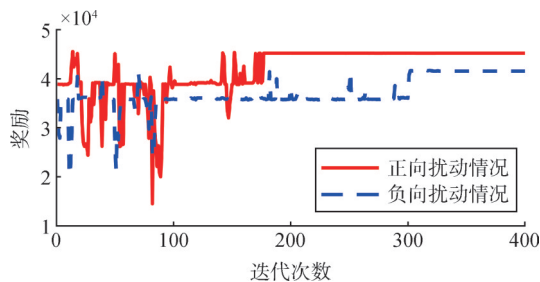


图8 DDPG训练奖励值

Fig.8 DDPG training reward value

### 4 结论

本文针对多高超声速飞行器在存在初始误差和不确定性情况下的协同再入制导问题，提出了一种

基于DDPG-LQR的时间协同再入制导策略。首先，采用RSCP算法生成多个约束条件下的时间协同再入轨迹。通过将飞行器运动方程的自变量转变为纵向航程角，同时将时间引入为状态变量，实现自由的末端飞行时间。然后，采用DDPG算法对多个LQR控制器的权重矩阵系数进行在线优化调节，通过设计合适的奖励函数提高控制器的跟踪制导精度。在考虑初始状态误差和不确定性的情况下，通过与传统LQR控制器相比较，本文提出的协同制导策略具有更好的制导精度和制导效果，且协同制导时间误差较小。在未来的研究中，将考虑轨迹重构下的协同制导方案。

### 参考文献

[1] LI Z H, HE B, WANG M H, et al. Time-coordination entry guidance for multi-hypersonic vehicles[J]. Aerospace Science and Technology, 2019(89): 123-135.

[2] 李曦, 赵长见, 高峰, 等. 高速飞行器编队通信保持控制与协同制导技术研究综述[J]. 导弹与航天运载技术(中英文), 2024(1): 71-76.

LI Xi, ZHAO Changjian, GAO Feng, et al. Research progress of formation communication holding control and collaborative guidance for hypersonic glide vehicles[J]. Missiles and Space Vehicles, 2024(1): 71-76.

[3] 刘旭, 李响, 王晓鹏. 高超声速滑翔飞行器解析协同再入制导[J]. 宇航学报, 2023, 44(5): 731-742.

LIU Xu, LI Xiang, WANG Xiaopeng. Analytical cooperative reentry guidance for hypersonic glide vehicles[J]. Journal of Astronautics, 2023, 44(5): 731-742.

[4] XU H, CAI G B, MU C X, et al. Analytical reentry guidance framework based on swarm intelligence optimization and altitude-energy profile[J]. Chinese Journal of Aeronautics, 2023, 36(12): 336-348.

[5] 郭杰, 郑金库, 王浩凝, 等. 高超声速滑翔飞行器再入制导方法及热点问题研究综述[J]. 空天技术, 2022(1): 54-63.

GUO Jie, ZHENG Jinku, WANG Haoning, et al. Review of research on reentry guidance methods and hot issues of hypersonic gliding vehicle[J]. Aerospace Technology, 2022(1): 54-63.

[6] 魏昀鹏, 都延丽, 王文凯, 等. 高超声速飞行器时间协同预测校正鲁棒制导[J]. 宇航学报, 2024, 45(3): 421-432.

WEI Yunpeng, DU Yanli, WANG Wenkai, et al. Time-cooperative predictor-corrector robust guidance for hypersonic vehicles[J]. Journal of Astronautics, 2024, 45(3): 421-432.

[7] WANG X Y, LI Y F, QUAN Z Y, et al. Optimal trajectory-tracking guidance for reusable launch vehicle based on adaptive dynamic programming[J]. Engineering Applications of Artificial Intelligence: The International Journal of Intelligent Real-Time Automation, 2023 (117): 12.

[8] 陈丹鹤, 何敬源, 刘幸川, 等. 航天器编队飞行相对运动轨迹优化方法综述[J]. 宇航学报, 2024, 45(3): 325-340.

CHEN Danhe, HE Jingyuan, LIU Xingchuan, et al. Review on relative trajectory optimization methods for spacecraft formation

- flying[J]. Journal of Astronautics, 2024, 45(3): 325-340.
- [9] YU J L, DONG X W, LI Q D, et al. Cooperative guidance strategy for multiple hypersonic gliding vehicles system[J]. Chinese Journal of Aeronautics, 2020, 33(3): 990-1005.
- [10] ZHANG X, LIU S, YAN J, et al. Fixed-time cooperative trajectory optimisation strategy for multiple hypersonic gliding vehicles based on neural network and ABC algorithm[J]. Aeronautical Journal, 2023, 127(1316): 1737-1751.
- [11] LI Y X, LIU X F, HE X H, et al. Cooperative optimal guidance of hypersonic glide vehicles by real-time optimization and deep learning[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2023, 237(10): 2266-2283.
- [12] VEPA R. Feedback tracking control of optimal reference trajectories for spacecraft relative motion[J]. Advances in Space Research: The Official Journal of the Committee on Space Research (COSPAR), 2022, 69(9): 3478-3489.
- [13] OU C, SHAN C J, CHENG Z T, et al. Adaptive trajectory tracking algorithm for the aerospace vehicle based on improved T-MPSP[J]. Mathematics, 2023, 11(9): 16.
- [14] 杨美, 苏自立, 赵吉松. 高超声速滑翔飞行器在线轨迹优化与再入制导方法研究[J]. 飞行力学, 2023, 41(6): 51-58.  
YANG Mei, SU Zili, ZHAO Jisong. A reentry guidance method with online trajectory optimization for hypersonic glide vehicle[J]. Flight Dynamics, 2023, 41(6): 51-58.
- [15] 李征, 陈海东, 彭博, 等. 可重复使用航天器再入协同制导研究[J]. 导弹与航天运载技术, 2021(3): 71-77.  
LI Zheng, CHEN Haidong, PENG Bo, et al. Coordinated reentry guidance law for reusable launch vehicle[J]. Missiles and Space Vehicles, 2021(3): 71-77.
- [16] XU H, CAI G B, FAN Y H, et al. Analytic time reentry cooperative guidance for multi-hypersonic glide vehicles[J]. Applied Sciences, 2023, 13(8): 24.
- [17] ZHAO K, SONG J, HU Y L, et al. Deep deterministic policy gradient-based active disturbance rejection controller for quad-rotor UAVs[J]. Mathematics, 2022, 10(15): 15.
- [18] LI X, JI Y H, SONG Y, et al. Modified deep deterministic policy gradient based on active disturbance rejection control for hypersonic vehicles[J]. Neural Computing & Applications, 2024, 36(8): 4071-4081.
- [19] 唐胜景, 王肖, 郭杰. 基于hp伪谱凸优化的高超声速滑翔飞行器轨迹优化与制导[J]. 战术导弹技术, 2020(5): 66-75.  
TANG Shengjing, WANG Xiao, GUO Jie. Trajectory optimization and guidance for hypersonic gliding vehicles based on hp pseudospectral convex programming[J]. Tactical Missile Technology, 2020(5): 66-75.

### 作者简介

- 宋志飞 (2000—), 男, 硕士研究生, 主要研究方向为飞行器轨迹优化、强化学习理论与应用。
- 吉月辉 (1987—), 女, 博士, 副教授, 主要研究方向为飞行器轨迹优化、制导与控制技术、故障诊断。
- 宋雨 (1989—), 男, 博士, 教授, 主要研究方向为飞行器编队控制、强化学习理论及应用。
- 刘俊杰 (1991—), 男, 博士, 副教授, 主要研究方向为飞行器轨迹优化、故障诊断、强化学习理论及应用。
- 高强 (1968—), 男, 教授, 主要研究方向为飞行器故障诊断、强化学习理论及应用。

(上接第50页)

- [11] 段飞飞, 王田宇, 温业堃. 机载产品可靠性鉴定试验定时截尾统计方案设计[J]. 中国工程机械学报, 2022, 20(6): 561-564.  
DUAN Feifei, WANG Tianyu, WEN Yekun. Design of timing censoring statistical scheme for airborne product reliability qualification test[J]. Transactions of China Engineering Machinery, 2022, 20(6): 561-564.
- [12] 金永明, 吴作栋, 王鹏飞. 基于AMSAA模型的复杂舰炮武器系统可靠性增长方法[J]. 火炮发射与控制学报, 2023, 44(2): 70-74+80.  
JIN Yongming, WU Zuodong, WANG Pengfei. Reliability growth method of a complex naval gun weapon system based on AMSAA model[J]. Journal of Artillery Launch and Control, 2023, 44(2): 70-74+80.
- [13] 鲁雪峰, 刘婧成, 江思杰, 等. 工程上环境折合系数的确定方法及其应用[J]. 环境技术, 2019, 37(2): 89-93.  
LU Xuefeng, LIU Jingcheng, JIANG Sijie, et al. The method of determining environmental conversion coefficient in engineering and its application[J]. Environmental Technology, 2019, 37(2): 89-93.
- [14] 杜振华, 赵宇, 陈金盾, 等. 产品设计定型时MTBF置信区间的确定[J]. 数学的实践与认识, 2003(2): 38-43.  
DU Zhenhua, ZHAO Yu, CHEN Jindun, et al. The determination of MTBF confidence interval at product designing finalization[J]. Practice and Understanding of Mathematics, 2003(2): 38-43.
- [15] 徐静, 李海波, 洪亮, 等. 可靠性强化试验及在航天产品中的应用研究[J]. 强度与环境, 2018, 45(6): 43-50.  
XU Jing, LI Haibo, HONG Liang, et al. Study on reliability enhancement testing and application in aerospace products[J]. Strength and Environment, 2018, 45(6): 43-50.
- [16] 商乾, 姜盛鑫, 陆宏伟, 等. 面向试验鉴定的航天产品试验标准适用性评估模型与方法研究[J]. 航天工业管理, 2023(9): 14-18.  
SHANG Qian, JIANG Shengxin, LU Hongwei, et al. Research on the applicability evaluation model and method of aerospace product test standard for test identification[J]. Aerospace Industry Management, 2023(9): 14-18.
- [17] 李迪凡, 景超杰, 李景育, 等. 一种温度、湿度、振动、盐雾四因素动态模拟加速环境试验系统的研制[J]. 环境技术, 2020, 38(5): 201-206.  
LI Difan, JING Chaojie, LI Jingyu, et al. Development of a four integrated dynamic simulation accelerated environmental test system for temperature, humidity, vibration and salt spray[J]. Environmental Technology, 2020, 38(5): 201-206.
- [18] ANWAR D A, ABTTAN R. Simultaneous fault detection and diagnosis in electric power system using hybrid method[J]. International Journal of Engineering & Technology, 2018. DOI: 10.14419/ijet.v7i4.14146.

### 作者简介

- 胡琳 (1994—), 女, 助教, 主要研究方向为可靠性评估。
- 吴和成 (1963—), 男, 教授, 博士生导师, 主要研究方向为可靠性评估。