

小天体柔性着陆器姿轨耦合智能控制

孙一勇¹, 郑鹤鸣², 翟光¹, 李杰¹, 王妍欣²

(1. 北京理工大学 宇航学院, 北京 100081; 2. 中国科学院 空天信息创新研究院, 北京 100094)

摘要: 针对小天体复杂摄动环境以及着陆器柔性变形作用力建模不准确对小天体着陆造成的不利影响, 提出一种基于最大熵强化学习的柔性小天体着陆器姿轨耦合智能控制方法。建立考虑柔性变形作用力等效轨道的动力学模型, 并采用基准面法表征具有复杂形变特点的柔性着陆器姿态, 从而构建用于智能控制器训练的姿轨耦合动力学环境。根据最大熵强化学习理论的软动作-评价 (Soft Actor-Critic, SAC) 算法, 设计了采用深度神经网络架构的姿轨耦合智能控制器, 各个推力器通过自适应输出推力, 在保持着陆器姿态稳定的同时高精度跟踪导航轨迹。对控制器部署实际任务后的着陆过程进行了仿真, 结果表明, 与经典的PD控制方法相比, 提出的姿轨耦合智能控制方法具有更强的鲁棒性。

关键词: 小天体着陆; 柔性着陆器; 深度强化学习; 姿轨耦合控制

中图分类号: V448.2

文献标识码: A

文章编号: 2096-9287(2024)03-0265-09

DOI: 10.15982/j.issn.2096-9287.2024.20230171

引用格式: 孙一勇, 郑鹤鸣, 翟光, 等. 小天体柔性着陆器姿轨耦合智能控制[J]. 深空探测学报(中英文), 2024, 11(3): 265-273.

Reference format: SUN Y Y, ZHENG H M, ZHAI G, et al. Attitude-orbit coupling intelligent control of flexible asteroid lander[J]. Journal of Deep Space Exploration, 2024, 11(3): 265-273.

引言

传统刚性着陆器在着陆小天体时, 极易与小天体表面碰撞发生倾覆或反弹逃逸, 而采用柔性材料的着陆器能够利用柔性材料的缓冲吸能作用, 有效避免碰撞后发生弹跳, 从而实现稳定地软着陆^[1-2]。例如, “火星探路者”(Mars Path Finder, MPF)着陆器采用了一个独特的气囊子系统, 以减弱着陆过程中冲击的能量^[3]。然而, 由于“火星探路者”没有采用主动轨迹控制, 在气囊子系统的作用下反弹了多次, 并最终停留在偏离目标一定距离的位置。不同于在火星上的软着陆, 小天体不均匀的弱引力环境以及柔性材料形变引起的复杂柔性作用力都给柔性着陆器的稳定控制带来了巨大的挑战。

很多学者已针对小天体着陆问题提出了有效的着陆器主动控制方法。Furfaro等^[4]采用多滑模面制导 (Multiple Sliding Surface Guidance, MSSG) 方法, 在小天体着陆过程中实现了实时制导。基于非奇异终端滑模控制 (Nonsingular Terminal Sliding Mode, NTSM) 技术, Lan等^[5]为小天体着陆器着陆系统设计有限时间控制律。Liu等^[6]提出了一种改进的

NTSM算法, 其使用了具有动态边界层的特定饱和函数来代替符号函数, 以抑制在对小天体实施软着陆期间产生的抖振现象。上述控制方法能够有效应对小天体不规则的引力环境, 但均没有对柔性体的控制问题展开研究。Zhai等^[7]设计了应用于小天体着陆的柔性分布式探测器结构并设计了有效的柔性体导航和控制方法, 但是该研究只针对着陆轨迹设计了控制方法。考虑到柔性着陆器具有姿轨高度耦合的特征, 即着陆器位置和姿态同时受到推力和柔性变形作用力影响, 着陆器各智能体无法单独实现有效控制; 此外柔性变形作用复杂, 具有极强不确定性, 对控制器的鲁棒性要求极高。而现有文献中针对小行星弱引力场中的柔性着陆器附着控制研究极少, 因此有必要针对柔性着陆器附着小行星的姿轨耦合控制方法开展研究。

基于深度强化学习的智能控制方法通过反复与控制对象进行交互训练来优化控制策略从而达到最优的控制效果^[8-10], 可有效应对小天体复杂引力环境建模进度不足的问题。基于最大熵原理的软动作-评价 (Soft Actor-Critic, SAC) 算法具有较强的鲁棒性^[11], 能够有效解决在复杂柔性变形作用力条件下姿态和轨道高度耦合的柔性着陆器控制问题。文献^[12~15]验证了

SAC算法在应对路面动态障碍物、空中复杂气动环境以及液体晃动不确定扰动等高随机环境的稳定控制能力,表明SAC算法具有很强的鲁棒性。

本文基于最大熵强化学习原理设计了柔性小天体着陆器的姿轨耦合智能控制方法。主要的创新点有:引入小天体的弱引力作用以及柔性材料的柔性相互作用,建立了柔性着陆器的姿轨耦合动力学环境,用于对智能控制器的训练。根据最大熵强化学习理论,设计了神经网络结构的智能控制器对柔性着陆器的姿态和轨道同时进行控制。相较于传统的PD控制器,本文设计的智能控制器在应对不均匀引力环境和复杂柔性形变产生的不确定扰动时具有更强的鲁棒性。

1 柔性小天体着陆器姿轨耦合动力学

1.1 柔性小天体着陆器结构模型

本文采用的柔性小天体着陆器结构如图1所示,该柔性着陆器由柔性材料包覆3个刚性智能体构成^[16],刚性智能体配置用于着陆控制的推力器。柔性材料设计为薄壁结构,将着陆器上下完全包覆;较薄的尺寸可减小结构重量,同时能应用薄膜制造工艺,在其表面集成柔性太阳能电池、柔性传感器等。柔性结构在与小天体表面接触时可耗散碰撞冲击、提供缓冲,但在附着过程中会发生较大的变形。着陆器采用对称结构,柔性材料完全展开后呈圆柱体,且3个智能体呈中心对称分布。柔性着陆器的着陆阶段包含着陆器释放、受控下降、与小天体表面接触三个过程,本文主要针对柔性着陆器受控下降的过程,研究其姿轨耦合智能控制方法。

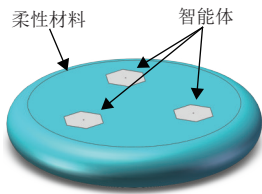


图1 柔性小天体着陆器结构示意图^[14]
Fig. 1 Diagram of the flexible lander^[14]

1.2 引入柔性变形作用力的轨道动力学模型

柔性着陆器第*i*个刚性智能体在小天体固连坐标系中的轨道动力学方程可以表示为

$$\ddot{\boldsymbol{\rho}}_i = \mathbf{g}_{ai} + \mathbf{a}_{ei} + \mathbf{u}_{ci} - [2\boldsymbol{\omega}_o \dot{\boldsymbol{\rho}}_i + \boldsymbol{\omega}_o (\boldsymbol{\omega}_o \boldsymbol{\rho}_i)] \quad (1)$$

其中: $\boldsymbol{\rho}_i = [x_i \ y_i \ z_i]^T$ 为小天体固连坐标系中第*i*个智能体的位置矢量; $\mathbf{g}_{ai} = [g_{ai}^x \ g_{ai}^y \ g_{ai}^z]^T$ 为第*i*个智能体受到的小天体引力加速度; $\mathbf{a}_{ei} = [a_{ei}^x \ a_{ei}^y \ a_{ei}^z]^T$

为第*i*个智能体受到的柔性相互作用加速度; $\mathbf{u}_{ci} = [u_{ci}^x \ u_{ci}^y \ u_{ci}^z]^T$ 为第*i*个智能体受到的控制推力加速度; $\boldsymbol{\omega}_o = [0 \ 0 \ \omega_o]^T$ 为小天体自转角速度。

将式(1)进一步展开可得

$$\begin{cases} \ddot{x}_i = 2\omega_o \dot{y}_i + \omega_o^2 x_i + g_{ai}^x + a_{ei}^x + u_{ci}^x \\ \ddot{y}_i = -2\omega_o \dot{x}_i + \omega_o^2 y_i + g_{ai}^y + a_{ei}^y + u_{ci}^y \\ \ddot{z}_i = g_{ai}^z + a_{ei}^z + u_{ci}^z \end{cases} \quad (2)$$

为降低智能控制器训练时的计算量、提高智能控制器的训练效率,将小天体视为匀质天体,采用二阶球谐函数描述小天体引力势函数,第*i*个智能体受到小天体的引力加速度分量形式为

$$\begin{cases} g_{ai}^x = -\frac{\mu_a x_i}{\rho_i^3} - \frac{3C_{20}\mu_a x_i R_a^2}{2\rho_i^5} \left(\frac{5z_i^2}{\rho_i^2} - 1 \right) - \frac{3C_{22}\mu_a x_i R_a^2}{\rho_i^5} \left[\frac{5(x_i^2 - y_i^2)}{\rho_i^2} - 2 \right] \\ g_{ai}^y = -\frac{\mu_a y_i}{\rho_i^3} - \frac{3C_{20}\mu_a y_i R_a^2}{2\rho_i^5} \left(\frac{5z_i^2}{\rho_i^2} - 1 \right) - \frac{3C_{22}\mu_a y_i R_a^2}{\rho_i^5} \left[\frac{5(x_i^2 - y_i^2)}{\rho_i^2} + 2 \right] \\ g_{ai}^z = -\frac{\mu_a z_i}{\rho_i^3} - \frac{3C_{20}\mu_a z_i R_a^2}{2\rho_i^5} \left(\frac{5z_i^2}{\rho_i^2} - 1 \right) - \frac{3C_{22}\mu_a z_i R_a^2}{\rho_i^5} \frac{5(x_i^2 - y_i^2)}{\rho_i^2} \end{cases} \quad (3)$$

其中: r 为着陆器到小天体质心的距离; μ_a 为小天体引力常数; R_a 为小天体的参考半径,且通常选取小天体近似椭球的最大半长轴; C_{20} 和 C_{22} 为小天体的二阶球谐系数。

此外,柔性着陆器的3个智能体并非完全独立,两两之间存在由柔性材料变形带来的柔性作用力。为完成对智能体轨道动力学的递推计算,需对单个智能体所受柔性变形作用力产生的加速度进行表征。本文采用二阶弹簧阻尼尼模型描述柔性材料对智能体的柔性变形作用力,则作用在第*i*个智能体上的柔性变形作用力加速度的三轴分量为

$$\begin{cases} a_{ei}^x = \frac{(x_i - x_j) F_{ij}}{m_i l_{ij}} + \frac{(x_i - x_k) F_{ik}}{m_i l_{ik}} \\ a_{ei}^y = \frac{(y_i - y_j) F_{ij}}{m_i l_{ij}} + \frac{(y_i - y_k) F_{ik}}{m_i l_{ik}} \\ a_{ei}^z = \frac{(z_i - z_j) F_{ij}}{m_i l_{ij}} + \frac{(z_i - z_k) F_{ik}}{m_i l_{ik}} \end{cases} \quad (4)$$

从式(4)可以看出,求解某一智能体在附着过程中柔性变形作用力加速度时需引入其它智能体的状态量,这意味着智能体间存在状态耦合。柔性着陆器整

体的轨道运动可用柔性着陆器质心的运动描述, 柔性着陆器质心的位置可根据3个智能体的位置实时获得。

$$\rho_m = \frac{\sum_{i=1}^3 \rho_i m_i}{\sum_{i=1}^3 m_i} \quad (5)$$

其中: ρ_m 为质心的位置; m_i 为第 i 个智能体的质量。

质心的速度可通过质心位置对时间求导获得, 即

$$\mathbf{v}_m = \frac{d\rho_m}{dt} \quad (6)$$

1.3 复杂形变着陆器的姿态表征

柔性材料在着陆过程中会产生复杂的柔性变形, 无法利用描述刚体着陆器姿态的方法来描述柔性着陆器的姿态, 定义柔性着陆器三个智能体质心的连线构成的平面为柔性着陆器的姿态基准面。定义基准面由着陆器质心指向着陆器上表面的法向矢量与小天体固连坐标系 oz 轴的夹角 ϕ 为倾斜角, 一般认为柔性着陆器附着初始时刻朝向太空的面为上表面; 定义基准面绕着基准面的正法向矢量转过的角度 θ 为旋转角。姿态基准面和姿态角的关系如图2所示。

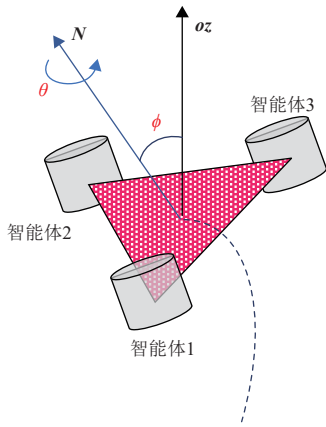


图2 姿态基准面和姿态角
Fig. 2 Attitude datum plane and attitude angle

两个姿态角也可以通过3个智能体质心位置实时计算得到, 计算公式为

$$\begin{cases} \theta = \arccos \frac{\mathbf{N} \cdot \mathbf{oz}}{|\mathbf{N}| |\mathbf{oz}|}, \mathbf{N} = \frac{(\rho_2 - \rho_m)(\rho_1 - \rho_m)}{|(\rho_2 - \rho_m)(\rho_1 - \rho_m)|} \\ \phi = \arccos \frac{(\rho_1 - \rho_m) \cdot [\rho_1(t_0) - \rho_m(t_0)]}{|(\rho_1 - \rho_m)| |[\rho_1(t_0) - \rho_m(t_0)]|} \end{cases} \quad (7)$$

其中: $\rho_1(t_0)$ 和 $\rho_m(t_0)$ 为附着初始时刻智能体1和着陆器质心的位置矢量; $\mathbf{oz} = [0; 0; 1]$ 为小天体固连坐标系的 z 轴单位矢量。

由此, 式(2)和(7)共同构成了柔性着陆器的姿轨耦合动力学模型。

2 基于最大熵强化学习的姿轨耦合智能控制器设计

2.1 最大熵强化学习基本原理

最大熵强化学习的SAC算法是一种基于随机性策略的算法, 其基本原理是在智能控制器训练的过程中, 不仅使长期的累加奖励值最大, 并且令策略的信息熵最大。该算法累加奖励的表达式为^[9]

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} + \alpha H(\Pi(\cdot|S_t)) \quad (8)$$

其中: $H(\Pi(\cdot|S_t))$ 为策略的信息熵; α 为熵的权重, 定义为熵权重系数。

通过最大化策略的信息熵, 增强训练过程中策略的调整能力, 从而提高控制的鲁棒性。SAC算法的Q函数采用柔性贝尔曼期望方程 (Soft-Bellman Equation)^[9] 为

$$Q_{\text{soft}}(S_t, A_t) = r(S_t, A_t) + \gamma E_{S_{t+1}, A_{t+1}} [Q_{\text{soft}}(S_{t+1}, A_{t+1}) - \alpha \lg \Pi(S_{t+1}, A_{t+1})] \quad (9)$$

训练过程中, 分别采用Critic神经网络和Actor神经网络来拟合评估和策略。Critic神经网络的代价函数为^[9]

$$J_Q(\theta) = E_{(S_t, A_t, S_{t+1}) \sim D, A_t \sim \Pi_\phi} \left[\frac{1}{\alpha} Q_\theta(S_t, A_t) - (r(S_t, A_t) + \gamma (Q_{\bar{\theta}}(S_{t+1}, A_{t+1}) - \alpha \lg \Pi_\phi(A_t|S_t)))^2 \right] \quad (10)$$

Actor神经网络的代价函数为^[9]

$$\begin{aligned} J_\Pi(\phi) = & D_{\text{KL}} \left[\Pi_\phi(\cdot|S_t) \parallel \exp \left(\frac{1}{\alpha} Q_\theta(S_t, \cdot) \right) - \right. \\ & \left. \lg \int \exp \left(\frac{1}{\alpha} Q_{\text{soft}}(S_t, A) \right) dA \right] = \\ & E_{S_t \sim D, A_t \sim \pi_\phi} \left[\lg \Pi_\phi(A_t|S_t) - \frac{1}{\alpha} Q_\theta(S_t, A_t) + \right. \\ & \left. \lg \int \exp \left(\frac{1}{\alpha} Q_{\text{soft}}(S_t, A) \right) dA \right] \quad (11) \end{aligned}$$

训练过程中, 奖励值在不断变化, 所以需要调整熵权重系数进行实时的调节。熵权重系数调节过程中使用的代价函数为^[9]

$$J_\alpha(\theta) = E_{A_t \sim \Pi_\phi} [-\alpha \lg \Pi_\phi(A_t|S_t) - \alpha H_\theta] \quad (12)$$

2.2 智能控制器设计

小天体在着陆的过程中, 柔性着陆器利用携带的

推力器通过喷出工质产生推力跟踪规划得到的期望轨迹,同时保持姿态稳定,从而实现稳定附着。假设智能体 i 配置两个推力器;一个推力器的推力方向朝着着陆器的上表面,并可在一定大小的空间锥内摆动,可利用推力矢量与基准面指向上表面的法向矢量的夹角 α_i 和推力矢量在基准面内投影与着陆器质心指向智能体质心的矢量的夹角 β_i 来描述推力器产生推力相对于基准面的方向,推力大小为 T_{i1} ;另外一个推力器方向固定,沿基准面法向朝着着陆器的下表面,推力大小为 T_{i2} 。智能体推力器的配置如图3所示。

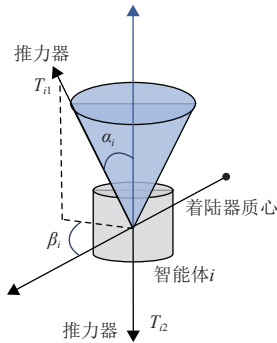


图3 智能体推力器配置

Fig. 3 Thruster configuration of the agent

智能体 i 的两个推力器产生的推力合力在小天体固连坐标系下可表示为

$$T_{ci} = - \begin{bmatrix} \cos(\alpha_i + \theta) \cos(\beta_i + \phi) \\ \cos(\alpha_i + \theta) \sin(\beta_i + \phi) \\ \sin(\alpha_i + \theta) \end{bmatrix} T_{i1} - \begin{bmatrix} \cos(-\theta) \cos(-\phi) \\ \cos(-\theta) \sin(-\phi) \\ -\sin(\pi/2 - \theta) \end{bmatrix} T_{i2} \quad (13)$$

由式(13)可以看出,智能体产生推力的方向受到着陆器整体姿态的影响,同时推力也对着陆器轨迹以及姿态产生控制,所以柔性着陆器着陆小天体的控制是一个姿轨控复杂耦合过程。最终,推力对智能体产生的控制加速度可以表示为 $\mathbf{a}_{ci} = \mathbf{T}_{ci}/m_i$ 。

在柔性着陆器执行实际任务前,根据建立的姿轨耦合动力学在地面仿真环境中对智能控制器进行训练。训练结束后,将训练好的智能控制器部署到柔性着陆器智能体的星载计算机,并在实际执行任务时对各个智能体起到控制作用。对柔性着陆器进行姿轨耦合智能控制器进行训练的系统结构如图4所示。

其中,智能控制器的输入为状态变量,包含着着陆器实际质心位置速度与期望位置速度的偏差以及实时姿态与期望姿态之间的偏差,表示为 $\mathbf{S} = [\Delta\rho_m^T \Delta v_m^T \Delta\phi \Delta\theta]^T$;输出为控制变量,是各个推力器的指令推力值和上表面推力器的偏角,记为 $\mathbf{A} = [T_{11}, T_{12}, T_{21}, T_{22}, T_{31}, T_{32}, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3]^T$ 。对智能控制器进行训

练的过程中,各智能体的推力器在接到智能控制器的指令后参照式(13)产生推力;在推力作用下各智能体按式(2)的动力学规律运动。通过各智能体的位置速度可实时计算得到着陆器质心的位置、速度以及整体的姿态,实际任务中各智能体的位置和速度是通过敏感器测量得到的;将着陆器实际状态和规划的期望状态做差得到智能控制器的输入,同时奖励值赋予智能控制器。在智能控制器的训练过程中,智能控制器输入输出的参数需要进行标准化从而提高训练过程的稳定性。将每一维输入参数的变化范围标准化为 $[-1, 1]$,每一维输出参数的变化范围为 $[-0, 1]$,标准化参数和原参数的关系可以表示为

$$\begin{cases} \mathbf{S}_N = [\Delta\rho_{mN}^T \Delta v_{mN}^T \Delta\phi_N \Delta\theta_N]^T = \\ \quad 2(\mathbf{S} - \mathbf{S}_{\min}) / (\mathbf{S}_{\max} - \mathbf{S}_{\min}) - 1 \\ \mathbf{A}_N = [T_{11N}, \dots, T_{32N}, \alpha_{1N}, \dots, \alpha_{3N}, \beta_{1N}, \dots, \beta_{3N}]^T = \\ \quad (\mathbf{A} - \mathbf{A}_{\min}) / (\mathbf{A}_{\max} - \mathbf{A}_{\min}) \end{cases} \quad (14)$$

其中: \mathbf{S}_{\min} 、 \mathbf{S}_{\max} 、 \mathbf{A}_{\min} 、 \mathbf{A}_{\max} 分别是输入状态和输出动作的最小和最大的界限值。

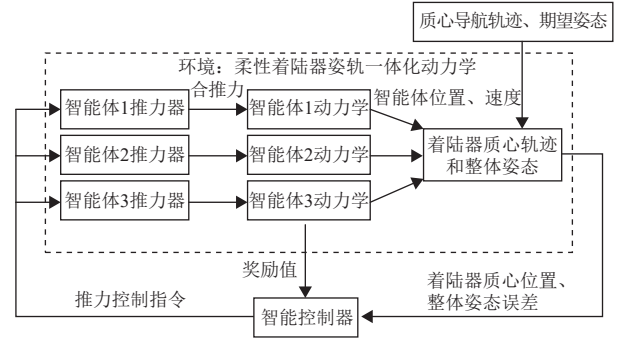


图4 柔性着陆器姿轨耦合智能控制系统示意图

Fig. 4 Control system of the attitude-orbit coupling intelligent control of the flexible lander

智能控制器的奖励函数设计为

$$R = r - 0.05|\mathbf{T}_{cN}|^2 - 100B \quad (15)$$

其中

$$r = \begin{cases} -0.1|\Delta\rho_{mN}|^2 - 0.1|\Delta\phi_N|^2, & |\Delta\rho_{mN}|^2 \leq \varepsilon_1, |\Delta\phi_N|^2 \leq \varepsilon_2 \\ -0.2, & |\Delta\rho_{mN}|^2 > \varepsilon_1, |\Delta\rho_{mN}|^2(t) < |\Delta\rho_{mN}|^2(t-1) \\ -1, & \text{else} \end{cases}$$

$$B = \begin{cases} 1, & |\Delta\rho_{mN}|^2 > N_{\max} \text{ or } |\rho_i - \rho_j|^2 > L_{\text{nex}}(i, j = 1, 2, 3, i \neq j) \\ 0, & \text{else} \end{cases}$$

奖励函数主要分为状态相关的奖励值、动作相关的奖励值和惩罚项3部分。状态相关的奖励值 r 采用分段奖励的方式, $\varepsilon_1 = \varepsilon_2 = 1 \times 10^{-3}$ 为质心位置偏差和姿

态偏差的精度值, 当质心位置偏差和姿态偏差均在精度范围内时, 给予其负值的权重并求和作为奖励值以促进偏差的进一步收敛; 当质心位置偏差在精度范围以外但当前时刻值小于前一时刻值时给予较小的负值奖励, 否则给予较大的负值奖励; 分段式的奖励方式有利于状态偏差的稳定收敛。动作奖励值设计为 $-0.05|T_{cN}|^2$, 其中 -0.05 为动作奖励值权重系数项, $T_c = T_{c1} + T_{c2} + T_{c3}$ 为所有推力器的合推力。该奖励值有利于减小推力器输出的推力值从而节省星上推进剂的消耗, 其中权重系数项可根据期望节省燃料的程度进行调整。惩罚项 B 的含义是当质心位置偏差的平方和大于三轴最大质心位置偏差平方和的 N_{max} 倍时或者两个智能体相对距离大于柔性材料最大的拉伸长度 L_{flex} 时, 给予很大的负值奖励作为惩罚, 并退出此轮训练。

智能控制器所设计的Actor和Critic神经网络结构如图5所示。其中: 每个隐含层均具有64个神经元, 输入层和输出层具有单个神经元, 神经网络采用全连接的形式。

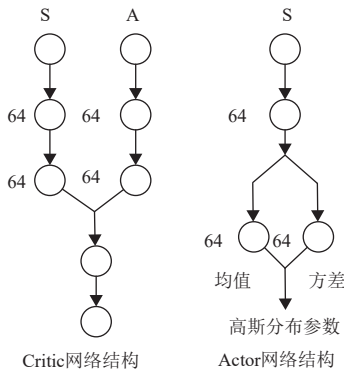


图5 智能体神经网络结构示意图
Fig. 5 Structure of the agent neural network

3 仿真验证

对上述姿轨耦合智能控制器进行训练以及部署后的仿真, 选择柔性着陆器的目标小天体为Itokawa小行星, 并将期望附着点设置于其表面的起伏区域。Itokawa小行星的相关参数如表1所示, 柔性着陆器的主要参数如表2所示, 智能控制器的训练参数如表3所示。

训练时柔性着陆器在该小天体着陆的质心导航轨迹如图6所示, 智能控制器的训练结果如图7所示, 智能控制器总计进行了580轮训练。在前500轮训练的过程中, 智能控制器获得的奖励值有大范围的变化, 说明智能控制器在不断调整策略, 进行探索和试错。在500轮后, 奖励值逐渐增长, 说明智能控制器的探索已到了可以达到较好控制效果的方向, 从而不断优化控

制效果, 得到更高的分数。当训练次数为560次左右时, 奖励值已达到设定的期望值-50, 说明控制器达到了期望的控制效果, 训练已收敛。

表1 Itokawa小行星参数
Table 1 Parameters of the small celestial body

小行星参数	数值
小行星质量/kg	3.147×10^{10}
小行星自转角速度/(rad·s ⁻¹)	1.4424×10^{-4}
小行星包络参考半径/m	300
二阶球谐系数C ₂₀	-0.073 3
二阶球谐系数C ₂₂	0.031 1

表2 柔性着陆器参数
Table 2 Parameters of the flexible lander

柔性着陆器参数	数值
各智能体质量/kg	各智能体: 166
柔性材料	3
推力器推力范围/N	[0, 30]
推力器推力脉冲间隔/s	0.1
着陆器质心始末位置/m	$\rho_m(t_0) = [28.71 \quad -60.50 \quad 115.00]$ $\rho_m(t_f) = [31.90 \quad -63.00 \quad 101.10]$
期望姿态值/(°)	$\phi_r = 0; \theta_r = 0$
初始位置偏差值/m	$\Delta\rho_m = 0$
初始速度偏差值/(m·s ⁻¹)	$\Delta v_m = 0$
初始姿态偏差值/(°)	$\Delta\phi = 0; \Delta\theta = 0$
着陆任务时长/s	150

表3 智能控制器训练参数

智能控制器训练参数	数值
学习率	Critic: 1×10^{-3} , Actor: 1×10^{-4} , 温度系数: 1×10^{-4}
经验池	5×10^{-6}
初始温度系数	0.5
目标熵	-12

Table 3 Training parameters of the intelligent controller

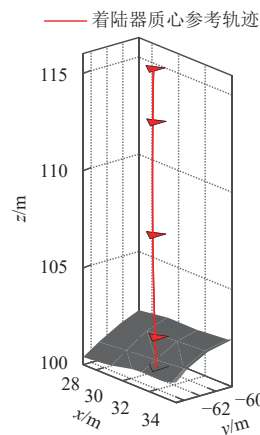


图6 柔性着陆器质心导航轨迹示意图
Fig. 6 Navigation trajectory of the mass center of the flexible lander

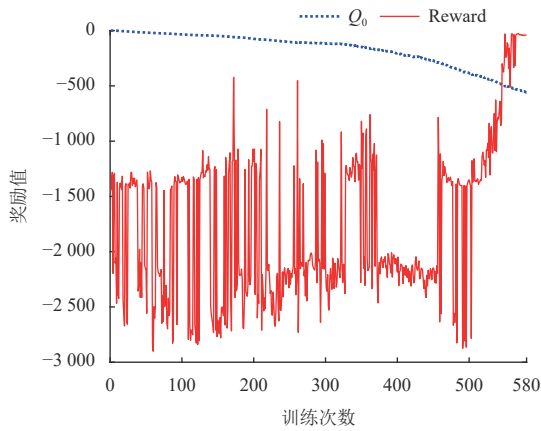


图7 智能控制器训练奖励值

Fig. 7 Training reward value of the intelligent controller

在智能控制器部署后执行实际着陆任务的过程中, 真实的小天体引力环境会存在未知的扰动, 实际的柔性变形作用力也会与动力学模型存在偏差。同时, 着陆过程中有很大可能性需要根据对小天体表面的探测结果而对导航轨迹进行实时调整。为验证部署后智能控制器的控制效果, 对每个智能体引入随机的扰动加速度来模拟引力环境和柔性变形作用力的未建模扰动为

$$|\alpha_i| = 10^{-4} \sin(0.2\pi t) \text{rand}(1)/m_i \quad (16)$$

其中: $\text{rand}(1)$ 为在区间(0, 1)内均匀分布的随机数。

选取训练过程中得分最高训练轮的控制策略, 对上述智能控制器部署到实际任务后的算例进行仿真, 并采用经典的PD控制器作为对比, 验证智能控制器在应对未建模扰动时的鲁棒性, 仿真结果如图8~12所示。

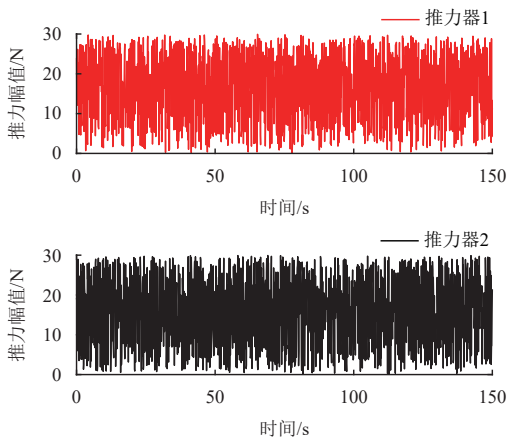


图8 单智能体两推力器推力矢量

Fig. 8 Two thruster thrust vectors of a single agent

单个智能体着陆过程中两个推力器在每个推力脉冲时刻产生的推力幅值曲线如图8所示。经过整个训练过程, 智能控制器在每个智能体上产生的控制推力是一致的, 即 $T_{11} = T_{21} = T_{31}$, $T_{12} = T_{22} = T_{32}$, $\alpha_1 = \alpha_2 =$

α_3 和 $\beta_1 = \beta_2 = \beta_3$, 从而实时保持了着陆器整体姿态的稳定。在着陆任务过程中, 单个智能体处于上表面的推力器不断调整推力脉冲的大小和方向, 下表面的推力器配合调整推力脉冲的大小, 两推力器的推力脉冲均在0~30 N的有效约束范围内; 3个智能体协同输出控制推力, 使柔性着陆器质心实时跟踪导航轨迹。

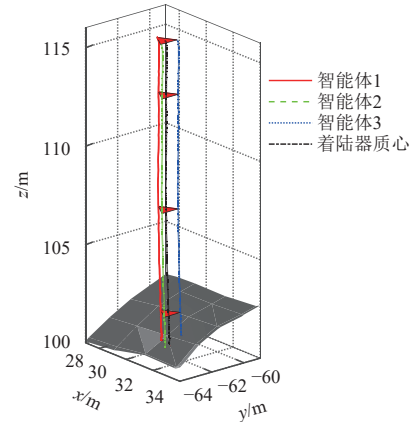


图9 着陆器附着小天体实际轨迹

Fig. 9 Actual trajectory of the flexible lander

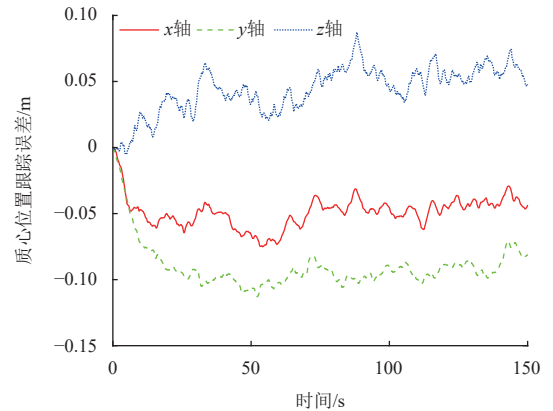


图10 着陆器质心位置偏差

Fig. 10 Position error of the mass center of the lander

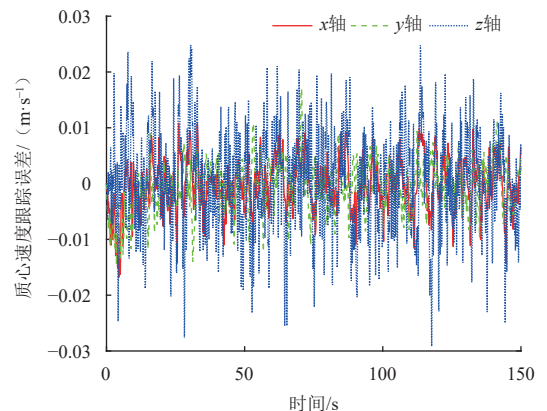


图11 着陆器质心速度偏差

Fig. 11 Velocity error of the mass center of the lander

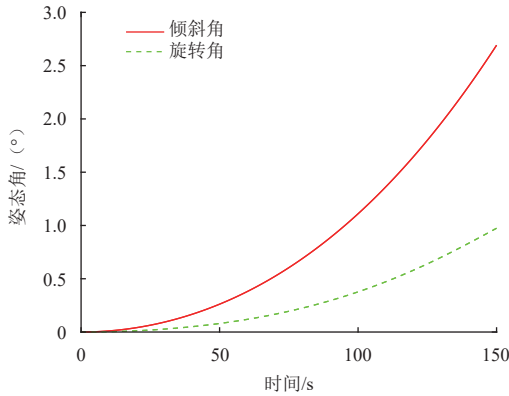


图 12 着陆器姿态角偏差

Fig. 12 Attitude angle error of the mass center of the lander

柔性着陆器的实际着陆轨迹如图9所示。在智能控制器作用下, 柔性着陆器质心的实际轨迹稳定地跟踪导航轨迹, 没有发生大范围的偏离, 3个智能体的轨迹几乎和质心轨迹保持平行, 柔性着陆器的姿态始终保持稳定。

着陆器质心位置偏差、质心速度偏差以及姿态角偏差如图10~12所示。三轴质心位置偏差始终保持在0.1 m以内, 且终端时刻三轴位置偏差为 $[-0.044, 0.081, -0.048]$ m, 终端位置误差幅值为0.103 9 m, 实现了对导航轨迹高精度的跟踪。终端三轴速度误差为 $[0.007, 0.002, 0.005]$ m/s, 误差幅值为0.008 8 m/s, 比终端位置误差幅值低一个量级, 说明着陆器在接触小行星表面时的速度很低, 不会与小行星表面产生剧烈的碰撞, 保证了附着稳定和安全的。附着过程中, 着陆器姿态基准面法向与小天体固连坐标系的z轴的夹角偏转在 3° 以内, 且着陆器绕基准面法向旋转角度在 1° 以内, 说明着陆器在附着过程中姿态始终保持稳定, 这样保证了最终的安全稳定附着, 也保证了附着过程中传感器稳定地对小行星表面进行探测。

采用经典PD控制器进行对比, PD控制器的比例系数选为 $k_p = 0.03$, 微分系数选为 $k_d = 0.05$, 该系数的选取以使部署前地面仿真环境中的动力学系统保持稳定为依据。仿真结果如图13~16所示。

PD控制下的实际附着轨迹如图13所示, 可见柔性着陆器虽然可以沿导航轨迹运动到期望落点附近, 但附着过程伴随着一定幅度的振动, 对终端着陆误差存在不利影响。

PD控制下, 位置跟踪误差曲线和速度跟踪误差曲线如图14~15。位置跟踪误差曲线在不断波动中逐渐发散, 终端时刻三轴位置误差为 $[0.035, -0.020, -0.114]$ m, 终端位置误差幅值为0.121 m。速度误差曲线逐渐收敛, 但仍存在一定幅度的波动, 终端三轴速度误差为

$[0.000 4, -0.000 1, -0.001]$ m/s, 误差幅值为0.001 1 m/s。姿态角误差曲线如图16所示, 可见旋转角存在持续震荡, 与图13中不断震动的实际附着轨迹相对应, 该状态下无法保证附着过程中传感器稳定地对小行星表面进行探测。由此可以得出结论, PD控制器无法有效应对实际任务过程中引力环境和柔性相互作用带来的不确定扰动。

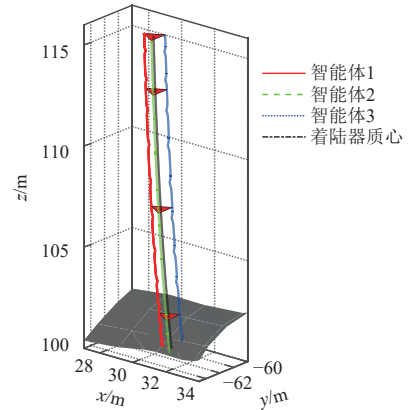


图 13 着陆器附着小天体实际轨迹 (PD控制)

Fig. 13 Actual trajectory of the flexible lander (PD control)

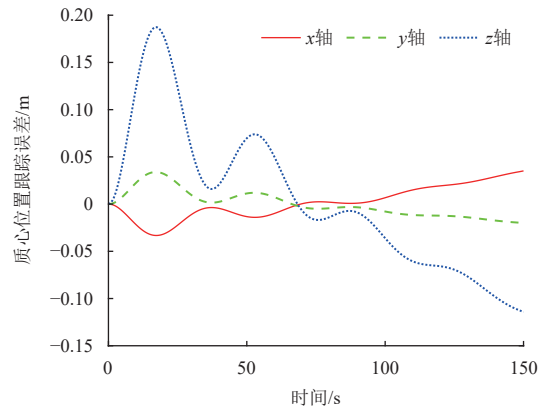


图 14 着陆器质心位置偏差 (PD控制)

Fig. 14 Position error of the mass center of the lander (PD control)

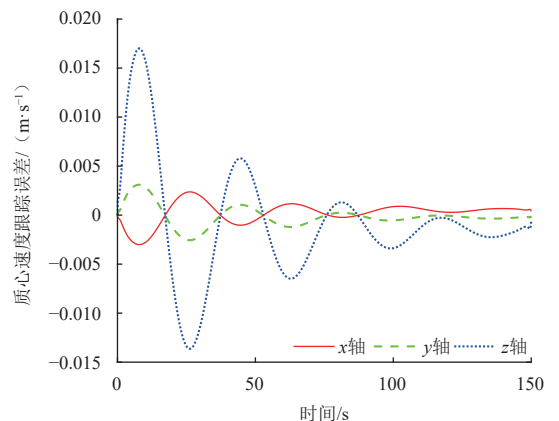


图 15 着陆器质心速度偏差 (PD控制)

Fig. 15 Velocity error of the mass center of the lander (PD control)

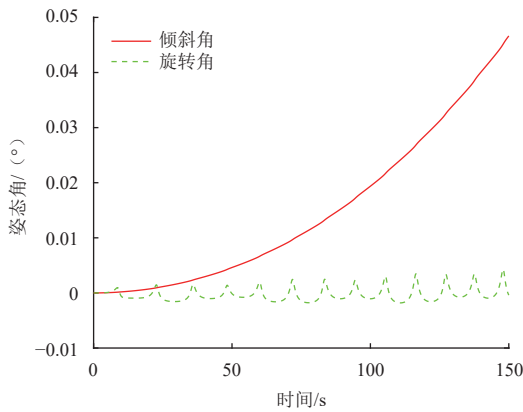


图 16 着陆器姿态角偏差 (PD控制)

Fig. 16 Attitude angle error of the mass center of the lander (PD control)

综上所述,存在不确定扰动时,PD控制器已经失效,而基于最大熵强化学习的智能控制器依然可以实现高精度的跟踪控制,验证了该智能控制器在柔性着陆器着陆小天体的任务中具有较强的鲁棒性。

4 结论

本文针对柔性着陆器着陆小天体过程中,由小天体引力摄动和复杂柔性变形引起不确定扰动的问题,基于最大熵强化学习原理设计了姿轨耦合智能控制器。通过仿真算例的对比验证得出结论,本文设计的智能控制器在应对不确定扰动时具有很强的鲁棒性,可以同时保持姿态稳定并确保着陆器平稳着陆。后续将考虑在实际工程任务的仿真环境中对所设计智能控制器进行验证。

参 考 文 献

- [1] 崔平远,张成宇,朱圣英,等. 小天体柔性附着技术[J]. 宇航学报, 2023, 44(6): 805-816.
CUI P Y, ZHANG C Y, ZHU S Y, et al. Technologies for flexible landing on small celestial bodies[J]. Journal of Astronautics, 2023, 44(6): 805-816.
- [2] 崔平远,陆晓萱,朱圣英,等. 小天体柔性附着状态协同估计方法[J]. 宇航学报, 2022, 43(9): 1219-1226.
CUI P Y, LU X X, ZHU S Y, et al. Cooperative state estimation method for small celestial body flexible landing[J]. Journal of Astronautics, 2022, 43(9): 1219-1226.
- [3] SANDY C. Development of the Mars Pathfinder inflatable airbag subsystem[C]//Proceedings of 14th Aerodynamic Decelerator Systems Technology Conference. Frederica: AIAA, 1997.
- [4] FURFARO R, CERSOSIMO D, WIBBEN D R. Asteroid precision landing via multiple sliding surfaces guidance techniques[J]. Journal of Guidance, Control, and Dynamics, 2013, 36(4): 1075-1092.
- [5] LAN Q, LI S, YANG J, et al. Finite-time soft landing on asteroids using nonsingular terminal sliding mode control[J]. Transactions of the Institute of Measurement and Control, 2013, 36(2): 216-223.
- [6] LIU X, SHAN Z, LI Y. Dynamic boundary layer based neural network quasi-sliding mode control for soft touching down on asteroid[J]. Advances in Space Research, 2017, 59(8): 2173-2185.
- [7] ZHAI G, LI J, SUN Y Y, et al. Research on asteroid landing with a new flexible spacecraft[J]. Journal of Aerospace Engineering, 2022, 35(5).
- [8] PHILIP S T, BRUNSKILL E. Policy gradient methods for reinforcement learning with function approximation and action-dependent baselines[EB/OL]. (2017-6-20)[2024-03-31]. <https://arxiv.org/abs/1706.06643>.
- [9] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[EB/OL]. (2013-12-19)[2024-03-31]. <https://arxiv.org/abs/1312.5602>.
- [10] MNIH V, KAVUKCUOGLU V, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [11] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[EB/OL]. (2018-1-4)[2024-03-31]. <https://arxiv.org/abs/1801.01290>.
- [12] 杨来义,毕敬,苑海涛. 基于SAC算法的移动机器人智能路径规划[J]. 系统仿真学报, 2023, 35(8): 1726-1736.
YANG L Y, BI J, YUAN H T. Intelligent path planning for mobile robots based on SAC algorithm[J]. Journal of System Simulation, 2023, 35(8): 1726-1736.
- [13] 唐腾,任双印,王春江. 基于SAC算法的四旋翼无人机姿态控制[C]//第三十四届中国仿真大会暨第二十一届中国仿真会议. 长沙:中国仿真大会, 2022.
TANG T, REN S Y, WANG C J. Attitude control of quadrotor UAV based on soft actor critic[C]//Proceedings of China Simulation Federation. The 34th China Simulation Conference and the 21st Asian Simulation Conference. Changsha: China Simulation Federation, 2022.
- [14] 李波,白双霞,孟波波,等. 基于SAC算法的无人机自主空战决策算[J]. 指挥控制与仿真, 2022, 44(5): 24-30.
LI B, BAI S X, MENG B B, et al. Autonomous air combat decision-making algorithm of UAVs based on SAC algorithm[J]. Command Control & Simulation, 2022, 44(5): 24-30.
- [15] 郑鹤鸣,翟光,孙一勇. 面向在轨加注的组合体姿态SAC智能控制[J]. 宇航学报, 2023, 44(7): 1020-1033.
ZHENG H M, ZHAI G, SUN Y Y. SAC intelligent attitude control method for on-orbit refueling combination[J]. Journal of Astronautics, 2023, 44(7): 1020-1033.
- [16] YAN W F, BAOYIN H. Position-attitude coupling guidance and control for asteroid landing with a flexible lander[J]. Aerospace Science and Technology, 2023, 141: 108567.

作者简介:

孙一勇(1987-),男,助理教授,硕士生导师,主要研究方向:无人系统与智能控制。

通信地址:北京市海淀区中关村南大街5号院(100081)

E-mail: sunyy@bit.edu.cn

郑鹤鸣(1995-),男,助理研究员,主要研究方向:航天器轨道和姿态动力学建模和控制方法。本文通信作者。

通信地址:北京市海淀区中关村北一条9号(100190)

E-mail: zhenghm@aircas.ac.cn

Attitude-Orbit Coupling Intelligent Control of Flexible Asteroid Lander

SUN Yiyong¹, ZHENG Heming², ZHAI Guang¹, LI Jie¹, WANG Yanxin²

(1. School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China;

2. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China)

Abstract: A method for attitude-orbit coupling intelligent control of flexible lander based on maximum entropy reinforcement learning is proposed in this paper, aiming at solve the adverse effects of the complex perturbation environment and the inaccurate flexible deformation force. Firstly, the orbital dynamics model of the equivalent agent is established by introducing the internal flexible force of the lander. The datum plane method is used to characterize the attitude of the flexible lander with complex deformation. The attitude-orbit coupling dynamic environment of the lander is constructed to train the intelligent controller. Then, an intelligent controller with deep neural network architecture is designed according to the soft actor-critic (SAC) algorithm of maximum entropy reinforcement learning theory. Each thruster can keep the lander attitude stable and track the navigation trajectory with high precision by self-adapting the output thrust. Finally, the landing process with the controller deployed is simulated. The simulation results show that compared with the classic PD control method, the intelligent control method proposed in this paper has stronger robustness.

Keywords: small celestial landing; flexible lander; deep reinforcement learning; attitude-orbit coupling control

Highlights:

- Introducing the internal flexibility of the lander, the equivalent agent orbital dynamics model is established.
- The datum plane method is used to characterize the attitude of the flexible lander with complex deformation characteristics.
- The attitude-orbit coupling dynamic environment of the lander is constructed for the training of the intelligent controller.
- According to the maximum entropy reinforcement learning theory, an intelligent controller with deep neural network structure is designed to control the attitude and orbit of the flexible lander simultaneously.
- Compared with the PD controller, the intelligent controller designed in this paper has stronger robustness when dealing with uncertain disturbance caused by uneven gravitational environment and complex flexible deformation.

[责任编辑: 杨晓燕, 英文审校: 宋利辉]