

基于强化学习的高超声速飞行器滑模自抗扰控制

唐伟强¹, 马景泰¹, 魏子栋¹, 高海燕²

(1. 兰州理工大学, 兰州, 730050; 2. 厦门理工学院, 厦门, 361024)

摘要: 针对高超声速飞行器的不确定性问题, 提出了一种强化学习与滑模自抗扰控制协同优化的智能控制方法。首先, 将高超声速飞行器的数学模型解耦为速度和高度子系统。其次, 为速度子系统设计自抗扰控制器以保证其跟踪性能, 为高度子系统则设计滑模自抗扰控制器以提高其鲁棒性。最后, 通过训练学习扩张状态观测器参数优化的确定性策略, 并将其嵌入控制系统, 实现模型驱动与数据驱动的协同优化。结果表明, 所设计的飞行器控制系统实现了速度和高度的精准跟踪, 与一般的滑模自抗扰控制系统相比, 表现出更强的抗干扰能力, 同时在响应时间和跟踪精度方面更具优势。

关键词: 高超声速飞行器; 强化学习; 滑模控制; 自抗扰控制; 扩张状态观测器

中图分类号: TP273; V19

文献标识码: A

Sliding Mode Active Disturbance Rejection Control for Hypersonic Vehicles Based on Reinforcement Learning

TANG Weiqiang¹, MA Jingtai¹, WEI Zidong¹, GAO Haiyan²

(1. Lanzhou University of Technology, Lanzhou, 730050; 2. Xiamen University of Technology, Xiamen, 361024)

Abstract: To address the uncertainty issues in hypersonic vehicles, an intelligent control method that synergistically integrates reinforcement learning with sliding mode active disturbance rejection control is proposed. First, the mathematical model of the hypersonic vehicle is decoupled into velocity and altitude subsystems. Second, an active disturbance rejection controller is designed for the velocity subsystem to ensure tracking performance, while a sliding mode active disturbance rejection controller is developed for the altitude subsystem to enhance robustness. Finally, a deterministic policy for optimizing the parameters of the extended state observer is learned through training and embedded into the control system online, achieving a collaborative optimization of model-driven and data-driven strategies. The results demonstrate that the proposed control system achieves satisfactory tracking of velocity and altitude. Compared with conventional sliding mode active disturbance rejection control systems, the proposed method exhibits stronger anti-interference capability and superior performance in response time and tracking accuracy. It is of more significance for improving the prediction accuracy of aircraft aerodynamic performance to predict the transition position of the boundary layer accurately.

Keywords: hypersonic vehicles; reinforcement learning; sliding mode control; active disturbance rejection control; extended state observer

0 引言

高超声速飞行器作为一种集航速快、航程远、响应迅速等特点于一身的临近空间飞行器, 具有巨大的军事和民用价值^[1]。然而, 由于复杂的飞行环境和特殊的气动结构, 使其控制设计面临诸多挑战。尤其是数学模型易受参数摄动、外部干扰及未建模动态的影响, 表现为强不确定性, 极大地增加了飞行控制的难度。为应对这些不确定性, 采用了自适应控制^[2]、预

测控制^[3]、容错控制^[4]及基于干扰观测器的控制^[5]等方法并已取得一定成果。其中, 基于干扰观测器类的控制方法因其直接且有效的估计和补偿不确定性的特点而备受关注^[6]。特别是自抗扰控制(Active Disturbance Rejection Control, ADRC), 将系统所有不确定性视为总扰动, 并通过扩张状态观测器(Extended State Observer, ESO)对其进行实时估计并在控制律中进行补偿。通过估计与补偿不确定性, ADRC的抗

干扰能力和控制精度得到显著提升^[7]。

得益于对系统动力学特性依赖的较少, ADRC在高超声速飞行器控制中得到了广泛的研究。文献[8-9]分别通过基于条件扰动的ADRC策略和解耦后子系统的ADRC设计, 实现了高超声速飞行器速度、高度跟踪及姿态控制器的设计, 跟踪性能体现出良好的稳定性。尽管ADRC在高超声速飞行器控制中展现出良好的性能, 但其应对不确定性时的抗干扰能力仍受限于ESO估计能力和反馈控制结构^[10]。为此, 文献[11]提出分数阶ESO以增强抗干扰能力; 文献[12]则采用径向基函数神经网络替代ESO, 进一步提高了系统性能。此外, 滑模控制与ADRC的结合也在一定程度上弥补了反馈控制结构上的不足, 如文献[13]提出的无模型改进滑模自抗扰控制和文献[14]设计的积分滑模自抗扰控制, 均显著提升了控制器的抗干扰能力和响应速度。尽管如此, 滑模自抗扰控制仍面临控制器参数设计复杂的问题。

对于控制器的参数设计, 通常采取固定参数设计、在线调参以及增益调度的方式^[15]。然而, 对于高超声速飞行器控制而言, 固定参数设计在复杂飞行环境下会牺牲控制性能; 在线调参的方式则因实时性不足而受限; 而增益调度的方式需要根据系统状态设计相应的控制器参数, 导致设计繁琐。此外, 参数设计通常依赖工程师的经验和能力, 成本较高。当前, 基于智能算法的参数优化方法为参数设计提供了全新的思路。启发式算法已被应用于解决参数优化问题, 如文献[16]采用差分进化优化算法对自抗扰控制器参数进行优化, 以满足蒸汽压力控制的最优和高效需求; 文献[17]则将混沌灰狼算法应用于四旋翼无人机的自抗扰控制器中进行参数优化。然而, 启发式算法在处理复杂实际问题时存在局限性, 例如易陷入局部最优解、仅能寻优一组固定参数, 以及对系统不确定性缺乏自适应能力。

近年来, 强化学习(Reinforcement Learning, RL)作为一种新兴的智能优化方法, 通过构建系统状态至控制器参数之间的映射, 展现出强大的自适应能力和在多变条件下寻求最优控制策略的特性, 有效弥补了人工调参与启发式算法的不足^[18]。然而, 传统RL方法如Q学习^[19]和深Q网络^[20], 只能处理离散的动作, 在处理连续动作空间方面存在明显的不足。当用于整定高超声速飞行器的控制参数时, 难以高效覆盖所有可行参数空间, 从而影响优化效果。相比之下, 深度确定性策略梯度(Deep Deterministic Policy

Gradient, DDPG)凭借其处理连续动作空间的优点, 成为高超声速飞行器控制器参数优化更好的选择。文献[21]将DDPG与ADRC结合, 实现了电机转角控制的自适应参数优化; 文献[22]则将DDPG算法应用于高超声速飞行器姿态控制器的参数优化, 取得了良好的控制效果。

需要说明的是, 本文与文献[21-22]的核心差异体现在控制器架构与优化维度上。在架构上, 文献[22]在线性ADRC基础上引入RL辅助控制, 属于控制量补偿思路; 本文则致力于滑模自抗扰控制本身参数的深度优化, 强调滑模控制与扰动估计能力的结构性融合, RL的作用是增强复合控制器的内生适应性。在优化维度上, 文献[21-22]以优化反馈增益与带宽为主, 侧重于系统动态调整; 本文则将DDPG的优化目标对准ESO的核心参数, 从扰动估计源头提升系统的抗干扰能力。因此, 与上述方案相比, 本文更侧重于通过优化底层观测器来从根本上提升控制器的自适应能力。此外, 本文方案主要优化的是前端的观测器参数, 而非直接改变闭环控制律, 这种设计更有利于保持系统的稳定性和可分析性。

基于上述分析, 本文设计的控制系统具有如下特点:

- a) 利用DDPG算法实现了对ESO参数的自适应整定, 相比人工经验整定, 能更迅速、准确地找到使系统性能最优的参数组合, 提高了参数整定的精度和效率。
- b) 该智能控制策略使ESO更适应系统动态和外部干扰的变化, 估计能力得到提升。同时, 通过提前停止准则的引入, 提升了DDPG的训练效率和稳定性, 进一步提高了系统的控制精度。
- c) 将滑模控制与ADRC相结合, 增强了系统在应对不确定性问题时的稳定性。相较于一般的滑模自抗扰控制方法, 本文的方法在快速性和鲁棒性方面更具优势。

1 系统描述

基于文献[14], 高超声速飞行器纵向动力学模型为

$$\dot{V} = \frac{T \cos \alpha - D}{m} - \frac{\mu \sin \gamma}{r^2} \quad (1)$$

$$\dot{\gamma} = \frac{L + T \sin \alpha}{mV} - \frac{(\mu - V^2 r) \cos \gamma}{Vr^2} \quad (2)$$

$$\dot{h} = V \sin \gamma \quad (3)$$

$$\dot{\alpha} = Q - \dot{\gamma} \quad (4)$$

$$\dot{Q} = M_{yy} / I_{yy} \quad (5)$$

式中 $V, h, \alpha, \gamma, \theta$ 分别为速度、高度、航迹角、攻角和俯仰角速率; m 和 g 分别为质量和重力加速度; I_{yy} 表示转动惯量; μ 表示重力常数; L, D 和 T 分别表示推力、阻力和升力; M_{yy} 表示俯仰力矩, 具体表达式为

$$L = \frac{1}{2} \rho V^2 S C_L, D = \frac{1}{2} \rho V^2 S C_D, T = \frac{1}{2} \rho V^2 S C_T \quad (6)$$

$$M_{yy} = \frac{1}{2} \rho V^2 S \bar{c} [C_M(\alpha) + C_M(d_e) + C_M(Q)] \quad (7)$$

式中 $\rho, S, \bar{c}$ 分别为空气密度、参考面积和平均气动弦长。另外, 相关空气动力系数由式 (8) 给出。

$$\begin{cases} C_L = 0.6203\alpha \\ C_D = 0.6450\alpha^2 + 0.0043378\alpha + 0.003772 \\ C_T = \begin{cases} 0.02576\beta_c, \beta_c < 1 \\ 0.0224 + 0.00336\beta_c, \beta_c \geq 1 \end{cases} \\ C_M(\alpha) = -0.035\alpha^2 + 0.036617\alpha + 5.3261 \times 10^{-6} \\ C_M(\delta_e) = c_e(\delta_e - \alpha) \\ C_M(Q) = \frac{\bar{c}}{2V} Q(-6.796\alpha^2 + 0.3015\alpha - 0.2289) \end{cases} \quad (8)$$

式中 c_e 表示力矩系数, 控制输入为油门开度指令 β_c 和升降舵偏转角 δ_e 。

2 控制器设计

2.1 速度子系统控制器设计

根据式 (1)、(6)、(7)、(8) 可将速度子系统表示为独立的一阶子系统:

$$\begin{cases} \dot{V} = f_v + g_v \beta_c \\ y_v = V \end{cases} \quad (9)$$

其中,

$$\begin{cases} f_v = \begin{cases} -\frac{1}{mr^2}(Dr^2 - m\mu \sin r), \beta_c < 1 \\ \frac{1}{mr^2}(0.00112\rho V^2 r^2 S \cos \alpha - Dr^2 - m\mu \sin r), \beta_c \geq 1 \end{cases} \\ g_v = \begin{cases} \frac{1}{m}(0.0129\rho V^2 S \cos \alpha), \beta_c < 1 \\ \frac{1}{m}(0.00168\rho V^2 S \cos \alpha), \beta_c \geq 1 \end{cases} \end{cases}$$

速度子系统采用 ADRC 进行控制系统设计, 其参考轨迹由跟踪微分器式 (10) 得到:

$$\begin{cases} \dot{v}_{1v} = v_{2v} \\ \dot{v}_{2v} = f_{\text{han}}(v_{1v} - V_c, v_{2v}, r_v, h_v) \end{cases} \quad (10)$$

式中 v_{1v}, v_{2v} 分别表示速度参考轨迹 V_c 的滤波信号 V_d 及其微分信号 $\dot{V}_d$; $f_{\text{han}}(x_1, x_2, r, h)$ 表示最速控制综合函数^[14]。

另外, ADRC 中的 ESO 可表示为

$$\begin{cases} \dot{E}_v = z_{1v} - V \\ \dot{z}_{1v} = z_{2v} - \eta_{1v} E_v + g_v \beta_c \\ \dot{z}_{2v} = -\eta_{2v} \text{fal}(E_v, a_v, \delta_v) \end{cases} \quad (11)$$

式中 z_{1v}, z_{2v} 分别为对速度 V 和“总和干扰” f_v 的估计值; η_{1v} 和 η_{2v} 为 ESO 待优化的参数。另外,

$$\text{fal}(e, a, \delta) = \begin{cases} |e|^a \text{sign}(e), |e| > \delta \\ \frac{e}{\delta^{1-a}}, |e| \leq \delta \end{cases}, \text{ 且 } a > 0, \delta < 1.$$

根据式 (9) 至式 (11), 速度子系统非线性状态误差反馈控制律表示为

$$\begin{cases} e_v = V_d - V = v_{1v} - z_{1v} \\ V_0 = k_v \text{fal}(e_v, a_{2v}, \delta_{2v}) \\ \beta_c = V_0 - z_{2v}/g_v \end{cases} \quad (12)$$

2.2 高度子系统控制器设计

2.2.1 高度环控制器设计

根据式 (1) 和式 (3), 高度环可表示为一阶系统:

$$\begin{cases} \dot{h} = V \sin \gamma \\ y_h = h \end{cases} \quad (13)$$

定义 h_d 为期望飞行速度 h_c 的滤波信号, 即高度参考轨迹, 则高度跟踪误差为 $e_h = h_d - h$ 。选择航迹角指令与文献 [14] 相同。

由此可将高度跟踪问题转化为航迹角跟踪问题, 其中, k_{hp}, k_{hl} 为大于零的待设计参数。跟踪微分器的设计与速度子系统中相同, 此处不再赘述。

2.2.2 航迹角环控制器设计

整理式 (2) 并结合式 (8) 可将航迹角环表示为一阶系统:

$$\begin{cases} \dot{\gamma} = f_\gamma + g_\gamma \theta_c \\ y_\gamma = \gamma \end{cases} \quad (14)$$

$$\text{其中, } f_\gamma = \frac{T \sin \alpha - 0.31015\rho V^2 S \gamma}{mV} - \frac{(\mu - V^2 r \cos \gamma)}{V^2 r^2}, \quad g_\gamma = \frac{0.31015\rho V^2 S}{mV}.$$

定义 γ_d 为期望飞行速度 γ_c 的滤波信号, 即航迹角参考轨迹, 则航迹角跟踪误差为 $e_\gamma = \gamma_d - \gamma$, 对其求导得:

$$\dot{e}_\gamma = \dot{\gamma}_d - \dot{\gamma} = \dot{\gamma}_d - f_\gamma - g_\gamma \theta_c \quad (15)$$

选择滑模面为 $s_\gamma = e_\gamma$, 对其求导得:

$$\dot{s}_\gamma = \dot{e}_\gamma = \dot{\gamma}_d - f_\gamma - g_\gamma \theta_c \quad (16)$$

此处采用滑动模态趋近时间短, 且运动点到达滑模面时速度较小的指数趋近律:

$$\dot{s}_\gamma = -\varepsilon_\gamma \text{sign}(s_\gamma) - q_\gamma s_\gamma \quad (17)$$

联立式 (16) 与式 (17), 可得航迹角环控制律为

$$\theta_c = \frac{1}{g_\gamma} (\dot{\gamma}_d - f_\gamma + \varepsilon_\gamma \text{sign}(s_\gamma) + q_\gamma s_\gamma) \quad (18)$$

式中 θ_c 为航迹角子系统的虚拟控制律; ε_γ , q_γ 均大于零。 f_γ 表示“总和干扰”, 由ESO估计得到:

$$\begin{cases} E_\gamma = z_{1\gamma} - \gamma \\ \dot{z}_{1\gamma} = z_{2\gamma} - \eta_{1\gamma} E_\gamma + g_\gamma \theta_c \\ \dot{z}_{2\gamma} = -\eta_{2\gamma} \text{fal}(E_\gamma, a_{1\gamma}, \delta_{1\gamma}) \end{cases} \quad (19)$$

式中 $z_{1\gamma}$, $z_{2\gamma}$ 分别表示对航迹角 γ 和 f_γ 的估计值; $\eta_{1\gamma}$ 和 $\eta_{2\gamma}$ 为ESO待优化的参数。

另外, 航迹角环的跟踪微分器设计与速度子系统中相同, 此处不再赘述。

2.2.3 俯仰角环控制器设计

根据式 (2) 至式 (8), 俯仰角环可表示为二阶系统:

$$\begin{cases} \dot{\theta} = \dot{\alpha} + \dot{\gamma} = Q \\ \dot{Q} = f_Q + g_Q \delta_c \\ \gamma_Q = \theta \end{cases} \quad (20)$$

其中, $f_Q = \frac{\rho V^2 S \bar{c}}{2 I_{yy}} \left[\left(\frac{1}{2V} (-6.796 \bar{c} Q) - 0.035 \right) \alpha^2 + \left(\frac{1}{2V} (0.301 5 \bar{c} Q) - 0.963 383 - c_c \right) \alpha - \frac{1}{2V} (0.228 9 \bar{c} Q) + 5.326 1 \times 10^{-6} \right]$, $g_Q = \frac{1}{2 I_{yy}} (\rho V^2 S \bar{c} c_c)$ 。

定义 θ_d 为期望飞行速度 θ_c 的滤波信号, 俯仰角跟踪误差为 $e_\theta = \theta_d - \theta$, 于是: $\dot{e}_\theta = \dot{\theta}_d - \dot{\theta}$, $\ddot{e}_\theta = \ddot{\theta}_d - \ddot{\theta} = \ddot{\theta}_d - f_Q - g_Q \delta_c$ 。

选择滑模面为 $s_\theta = \dot{e}_\theta + c_\theta e_\theta$ (c_θ 为滑模面系数), 对其求导得:

$$\dot{s}_\theta = \ddot{e}_\theta + c_\theta \dot{e}_\theta = \ddot{\theta}_d - f_Q - g_Q \delta_c + c_\theta \dot{e}_\theta \quad (21)$$

同样选取指数趋近律为

$$\dot{s}_\theta = -\varepsilon_\theta \text{sign}(s_\theta) - q_\theta s_\theta \quad (22)$$

联立式 (21) 和式 (22), 整理可得控制律为

$$\delta_c = \frac{1}{g_Q} [\ddot{\theta}_d - f_Q + c_\theta \dot{e}_\theta + \varepsilon_\theta \text{sign}(s_\theta) + q_\theta s_\theta] \quad (23)$$

式中 ε_θ , q_θ 均大于零; f_Q 表示“总和干扰”, 由ESO估计得到:

$$\begin{cases} E_\theta = z_{1\theta} - \theta \\ \dot{z}_{1\theta} = z_{2\theta} - \eta_{1\theta} E_\theta \\ \dot{z}_{2\theta} = z_{3\theta} - \eta_{2\theta} \text{fal}(E_\theta, a_{2\theta}, \delta_{2\theta}) + g_\theta \delta_c \\ \dot{z}_{3\theta} = -\eta_{3\theta} \text{fal}(E_\theta, a_{3\theta}, \delta_{3\theta}) \end{cases} \quad (24)$$

式中 $z_{1\theta}$, $z_{2\theta}$, $z_{3\theta}$ 分别表示 θ 的估计信号、一阶微分信号和 f_Q 的估计信号; $\eta_{1\theta}$, $\eta_{2\theta}$ 和 $\eta_{3\theta}$ 为ESO待优化的参数。参考轨迹由跟踪微分器式 (25) 得到:

$$\begin{cases} \dot{v}_{1\theta} = v_{2\theta} \\ \dot{v}_{2\theta} = \text{fhan}(v_{1\theta} - \theta_c, v_{2\theta}, r_\theta, h_\theta) \end{cases} \quad (25)$$

式中 $v_{1\theta}$, $v_{2\theta}$ 分别表示 θ_c 的滤波信号 θ_d 和一阶微分信号 $\dot{\theta}_d$ 。

2.3 基于DDPG的ESO参数协同优化机制

DDPG是一种Actor-Critic网络架构下的RL算法, 其基本原理是: 时间 t 内, 智能体更新累积折扣奖励 G_t [22], 并根据状态 s_t 和奖励 r_t 执行动作 a_t , 通过智能体与环境的交互, 更新下一个状态 s_{t+1} , 以获得当前奖励 r_{t+1} 。Actor网络参数的更新根据目标函数 $J(\mu)$ 的梯度:

$$\nabla_{\theta^\mu} J(\mu) = \frac{1}{N} \sum_{t=1}^N \nabla_a Q(s, a | \theta^Q) \Big|_{s=s_t, a=\mu(s_t)} \nabla_{\theta^\mu} \mu(s | \theta^\mu) \Big|_{s=s_t} \quad (26)$$

式中 $\mu(s | \theta^\mu)$, θ^μ 为Actor网络表征及网络参数。而Critic网络的更新目标是其损失函数最小, 网络损失函数 $\Gamma(\theta^Q)$ 定义为

$$\Gamma(\theta^Q) = \frac{1}{N} \sum_{t=1}^N [y_t - Q(s_t, a_t | \theta^Q)]^2 \quad (27)$$

式中 $y_t = r(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_{t+1} | \theta^\mu) | \theta^Q)$, 表示目标 Q 值, 用于更新Critic网络。 $Q(s, a | \theta^Q)$, θ^Q 为Critic网络表征及网络参数。

根据对高超声速飞行器速度子系统、航迹角环和俯仰角环的分析, 可求出相应传递函数的特征方程:

$$S^2 + \eta_{1V} S + \eta_{2V} = 0 \quad (28)$$

$$S^2 + \eta_{1\gamma} S + \eta_{2\gamma} = 0 \quad (29)$$

$$S^3 + \eta_{1\theta} S^2 + \eta_{2\theta} S + \eta_{3\theta} = 0 \quad (30)$$

由式 (28) 和式 (29) 可求得特征方程对应的极点

$$p_{1V, 2V} = -\frac{\eta_{1V}}{2} \pm \sqrt{\left(\frac{\eta_{1V}}{2}\right)^2 - \eta_{2V}}, \quad p_{1\gamma, 2\gamma} = -\frac{\eta_{1\gamma}}{2} \pm \sqrt{\left(\frac{\eta_{1\gamma}}{2}\right)^2 - \eta_{2\gamma}}$$

通过调节 η_{1V} , η_{2V} , $\eta_{1\gamma}$, $\eta_{2\gamma}$ 可改变ESO扰动估计带宽和阻尼特性, 其作为DDPG的动作集选择是可行的。

另外, 为保证ESO的稳定性, 根据式 (30) 选择特征方程的三个实极点 $p_{1\theta} = p_{2\theta} = p_{3\theta} = -\omega_\theta$, 则特征方程为

$$(s + \omega_\theta)^3 = s^3 + 3\omega_\theta s^2 + 3\omega_\theta^2 s + \omega_\theta^3 = 0 \quad (31)$$

由此可得参数关系: $\eta_{1\theta} = 3\omega_\theta$, $\eta_{2\theta} = 3\omega_\theta^2$, $\eta_{3\theta} = \omega_\theta^3$ 。其中, ω_θ 为ESO带宽, 表征扰动估计的动态响应速度。DDPG通过优化参数 $\eta_{1\theta}$, $\eta_{2\theta}$, $\eta_{3\theta}$ 自适应调整极点位置, 从而提升扰动估计性能。

文献 [23] 提出基于状态和动作来塑形函数, 利用更具体的信息引导智能体, 无需智能体仅从状态奖励中自行发现这些信息。因此, 奖励函数中权重选择

可以是基于控制目标的敏感性排序。本文中航迹角 γ 的权重最大,能够直接且敏感地反映控制跟踪性能;燃油阀开度 β_c 权重适中,用于平衡飞行姿态的稳定性;而升降舵偏转角 δ_c 权重较小,旨在避免过于激进的控制动作,保证系统的平稳性。

综上所述,选取各个子系统ESO的参数作为DDPG的动作集,即 $a_n = \{\eta_{1V}, \eta_{2V}, \eta_{1\gamma}, \eta_{2\gamma}, \eta_{1\theta}, \eta_{2\theta}, \eta_{30}\}$,状态集 $S_n = \{\gamma, \beta_c, \dot{\beta}_c, \delta_c, \dot{\delta}_c\}$ 。回报函数设置为 $R =$

$-0.01|\delta_c| - 5|\gamma| - 2|\beta_c| - 5d_0$, d_0 为截至因子。一般 $d_0 = 0$,当航迹角误差 $\gamma > 1$ 时,返回值 $d_0 = 1$ 。此外,为提高训练效率,本文在DDPG算法中引入了提前停止准则:当航迹角到达预设上限时,提前终止训练。

基于以上控制器的设计,控制系统如图1所示,图中 V_c 为期望速度, h_c 为期望高度, γ_c 和 θ_c 为虚拟控制量。

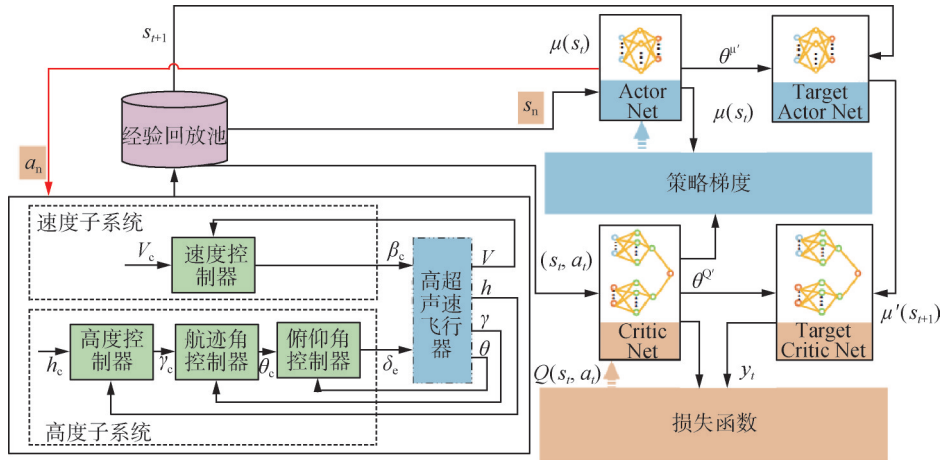


图1 基于强化学习的滑模自抗扰控制结构

Fig.1 Structure diagram of sliding mode active disturbance rejection control based on reinforcement learning

2.4 稳定性分析

本文提出的基于强化学习的控制方案,其核心架构是离线优化与在线控制相结合。DDPG算法在训练阶段学习得到一组优化的ESO参数。在线应用时,这些参数被固定并载入滑模自抗扰控制器中。因此,控制系统的实时稳定性根本上取决于基础控制器本身的稳定性。高度环、航迹角环和俯仰角环滑模面的证明与文献[14]相同,此处不再赘述。本节仅针对速度子系统进行稳定性分析。

a) ESO估计误差稳定性分析。

假设总扰动导数有界,即存在常数 $M > 0$,使得 $|\dot{f}_V| \leq M$ 。

定义ESO估计误差: $e_{1V} = z_{1V} - V$, $e_{2V} = z_{2V} - \dot{f}_V$,其动态为

$$\begin{cases} \dot{e}_{1V} = \dot{z}_{1V} - \dot{V} = e_{2V} - \eta_{1V}e_{1V} \\ \dot{e}_{2V} = \dot{z}_{2V} - \dot{\dot{f}}_V = -\eta_{2V} \text{fal}(e_{1V}, a_V, \delta_V) - \dot{f}_V \end{cases} \quad (32)$$

选取李雅普诺夫函数候选:

$$V_{\text{ESO}} = \frac{1}{2}e_{1V}^2 + \frac{1}{2\kappa}e_{2V}^2, \kappa > 0 \quad (33)$$

根据 $\text{fal}(e, a, \delta)$ 的性质,存在正常数 $k_1, k_2 > 0$,使得:

$$e_{1V} \text{fal}(e_{1V}, a_V, \delta_V) \geq k_1 e_{1V}^2 |\text{fal}(e_{1V}, a_V, \delta_V)| \leq k_2 |e_{1V}| \quad (34)$$

对时间求导并应用杨氏不等式得:

$$\begin{aligned} \dot{V}_{\text{ESO}} &= -\eta_{1V}e_{1V}^2 + e_{1V}e_{2V} - \frac{\eta_{2V}}{\kappa}e_{2V} \text{fal}(e_{1V}, a_V, \delta_V) - \frac{1}{\kappa}e_{2V}\dot{f}_V \\ &\leq -v_1e_{1V}^2 - v_2e_{2V}^2 + \frac{\epsilon_2 M^2}{2\kappa} \end{aligned} \quad (35)$$

$$\text{式中 } v_1 = \eta_{1V} - \frac{1}{2\epsilon_1} - \frac{\eta_{2V}k_2}{2\kappa\epsilon_3}, v_2 = -\frac{\epsilon_1}{2} - \frac{\eta_{2V}k_2\epsilon_3}{2\kappa} - \frac{1}{2\kappa\epsilon_2}。$$

适当选择参数 $\eta_{1V}, \eta_{2V}, \epsilon_1, \epsilon_2, \epsilon_3$ 确保 $v_1, v_2 > 0$ 。令 $v = \min(v_1, v_2\kappa) > 0$,则由比较引理,估计误差满足:

$$\limsup_{t \rightarrow \infty} \|e_V(t)\| \leq \sqrt{\frac{\epsilon_2 M^2}{v\kappa}} \quad (36)$$

b) 闭环系统跟踪误差稳定性。

在参考信号导数有界($|\dot{V}_d| \leq M_d$)和ESO估计误差有界($|e_{2V}| \leq M_e$)的条件下,速度跟踪误差是一致最终有界的。跟踪误差动态为

$$\dot{e}_v = \dot{V}_d - k_v \text{fal}(e_v, a_v, \delta_v) + e_{v2} \quad (37)$$

选取李雅普诺夫函数为

$$V_{\text{track}} = \frac{1}{2} e_v^2 \quad (38)$$

$$\dot{V}_{\text{track}} = e_v \dot{e}_v = e_v [\dot{V}_d - k_v \text{fal}(e_v, a_v, \delta_v) + e_{v2}] \quad (39)$$

当 $0 < a_v \leq 1$ 时, 存在正常数 k_3 , 使得:

$$e_v \text{fal}(e_v, a_v, \delta_v) \geq k_3 e_v^2 \quad (40)$$

则由杨氏不等式:

$$\dot{V}_{\text{track}} \leq -\left(k_v k_3 - \frac{1}{2\epsilon_3} - \frac{1}{2\epsilon_4}\right) e_v^2 + \frac{\epsilon_3}{2} \dot{V}_d^2 + \frac{\epsilon_4}{2} e_{v2}^2 \quad (41)$$

选择足够大的 k_v , 使得:

$$k_v k_3 - \frac{1}{2\epsilon_3} - \frac{1}{2\epsilon_4} > 0 \quad (42)$$

$$\dot{V}_{\text{track}} \leq -v_3 e_v^2 + v_4 \quad (43)$$

式中 $v_3 = k_v k_3 - \frac{1}{2\epsilon_3} - \frac{1}{2\epsilon_4} > 0$, $v_4 = \frac{\epsilon_3}{2} M_d^2 + \frac{\epsilon_4}{2} M_{e2}^2$ 。

由比较引理, 跟踪误差是一致最终有界的, 稳态误差满足:

$$\limsup_{t \rightarrow \infty} |e_v(t)| \leq \sqrt{\frac{v_4}{v_3}} \quad (44)$$

3 结果与分析

3.1 控制器参数

在设置 DDPG 的离线训练时, 最大训练回合为 100, 每回合最大步长为 100, 采样时间为 0.1 s, 批学习数为 64, 网络噪声统一设为 0.4。另外, 为了验证不同初始参数下 RL 的收敛稳定性, 选取不同的学习率与经验池大小。在原始参数 (学习率为 0.005, 经验池大小为 1 000 000) 基础上, 设置了以下对比组: 低学习率组 (学习率为 0.001, 经验池大小为 1 000 000)、高学习率组 (学习率为 0.01, 经验池大小为 1 000 000)、小经验池组 (学习率为 0.005, 经验池大小为 500 000), 训练结果如图 2 所示。

图 2 展示了不同参数组合下的累积奖励曲线。由图 2 可以看出, 尽管超参数有所变化, 但各曲线均在 70~100 回合内趋于收敛, 且最终性能差异小于 10%。特别是学习率在 0.001~0.01 范围内变化时, 收敛特性保持稳定, 表明本文的训练策略对超参数选择不敏感, 具有较强的鲁棒性。

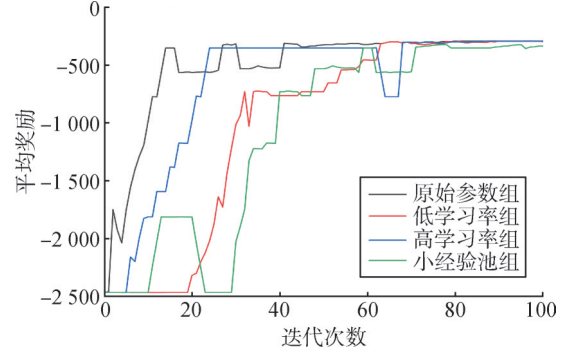


图2 累积奖励曲线

Fig.2 Average reward curve

图 3 展示了 DDPG 对 ESO 参数的优化结果。ESO 参数在迭代 70 次之后仍处于动态变化中, 但数据波动幅度较小。另外, 训练过程中 DDPG 同时在参数摄动、外部干扰和执行器故障等多种工况下进行, 其学习到的策略会尝试在一组 ESO 参数下平衡不同工况的需求, 参数移植性强, 能够泛化到所有情况。参数摄动、外部干扰和执行器故障的设计详见本文 3.2 节。

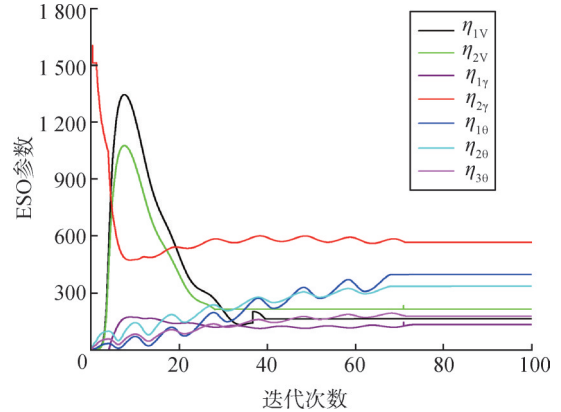


图3 ESO 参数优化结果

Fig.3 ESO parameter optimization results

3.2 仿真结果

为全面验证本文所提 DDPG-SMADRC 方法的性能优势与鲁棒性, 本节将在多种典型及极端工况下, 与传统的 SMADRC 方法进行系统性对比分析, 同时, 增加在标称、外部干扰与参数摄动情况下与文献 [24] 中基于 SESO 的 ADRC 方法的对比仿真。

3.2.1 标称情况

图 4 为标称情况系统响应。从图 4a~b 可看出三种控制方法均能使输出追踪期望信号, 但本文方法由于引入了 DDPG 对 ESO 参数的自适应整定, 在高度和速度跟踪方面整体上更具优势。控制输入方面如图 4c~d

所示,在SADRC作用下降舵偏转角出现较大振荡。其他状态最终均收敛至零,但DDPG-SMADRC作用下攻角表现更平滑,如图4e所示。总体来看,在标称情况下本文方法的飞行器控制性能更优。

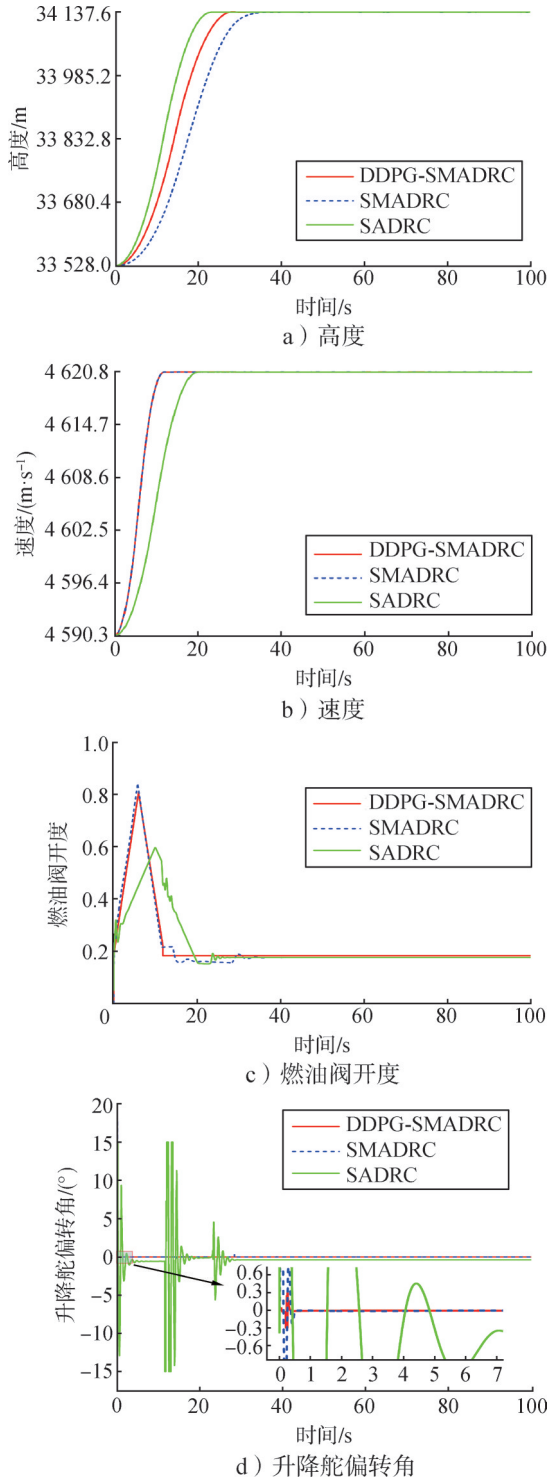
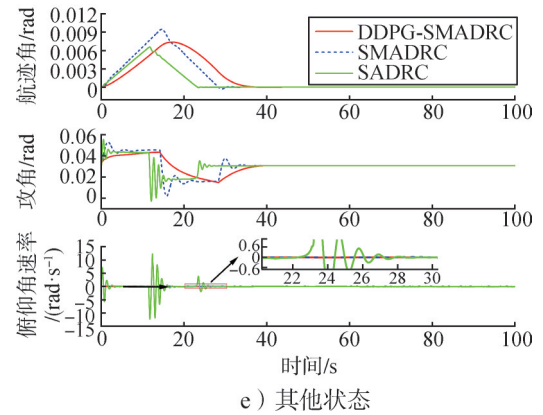


图4 标称情况系统响应

Fig.4 System response under nominal condition



续图4

3.2.2 参数摄动和外部干扰情况

外部干扰的特性,即对控制的影响通常表现为周期性平滑变化且具有对称性,本文将建模为幅值有限的有界扰动。因此,本文将外部干扰设定为正弦函数的形式,即 $d_v = 2\sin(\pi t)$, $d_h = 0.001\sin(0.2\pi t)$, $d_\gamma = 0.0005\sin(0.2\pi t)$, $d_a = 2\sin(3\pi t)$, $d_Q = 0.5\sin(0.2\pi t)$ 。参数摄动统一设置为25%^[25],参数包括 m 、 S 、 ρ 、 $\bar{c}$ 、 c_e 、 I_{yy} 、 C_L 、 C_D 、 C_T 、 $C_M(\alpha)$ 、 $C_M(\delta_e)$ 、 $C_M(Q)$ 。

图5为参数摄动与外部干扰情况下系统响应。如图5a~b显示,在高度和速度跟踪方面,SMADRC与SADRC系统作用下出现了时变误差。这一现象在控制输入方面得到了进一步印证,DDPG-SMADRC系统的高频振荡更小,如图5c~d所示。传统SMADRC方法振荡较大是滑模控制的固有现象,其本质来源于高频切换控制律的离散化执行;而DDPG通过动态优化ESO参数以调整ESO带宽,能够有效抑制抖振的频率。此外,图5e展示了其他状态的收敛情况。从以上分析可以看出,DDPG-SMADRC作用下系统的稳态误差更小,抗干扰能力更强。这是因为训练好的智能体能够通过自适应调整ESO参数,实时响应系统状态变化,从而更好地适应外部干扰和参数摄动。

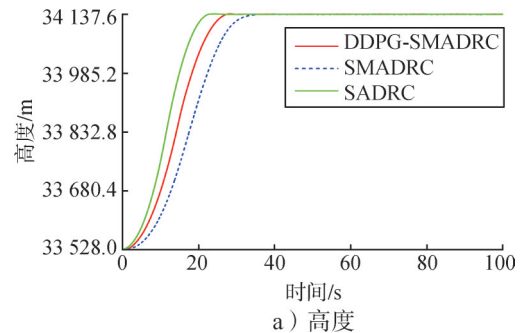
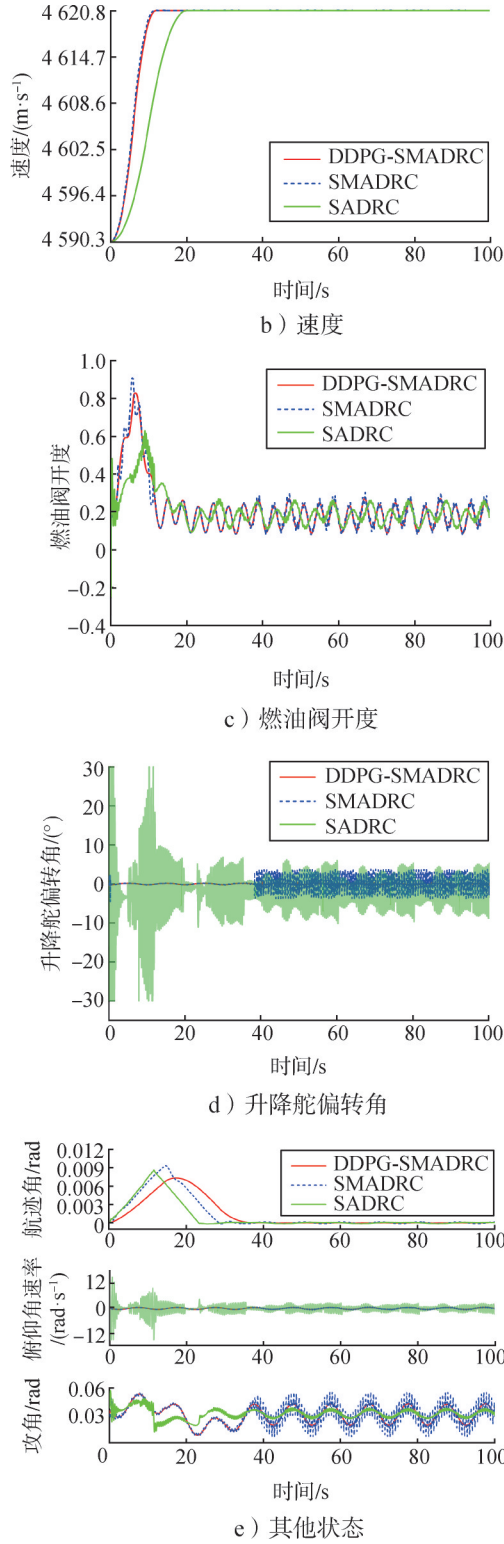


图5 参数摄动与外部干扰情况下系统响应

Fig.5 System response under parameter perturbation and external disturbance



续图5

3.2.3 执行器故障情况

为验证控制器在关键执行机构性能退化时的容错能力,采用时变振荡故障模型更真实地模拟执行器在发生故障后性能持续波动的复杂工况: $\beta_c(t) =$

$\begin{cases} \beta_c(t), & t < t_0 \\ f_1 \beta_c(t), & t \geq t_0 \end{cases}$ 和 $\delta_c(t) = \begin{cases} \delta_c(t), & t < t_0 \\ f_1 \delta_c(t), & t \geq t_0 \end{cases}$ 。其中, $f_1 = f_2 = 0.8 + 0.015 \sin(0.5\pi t)$, 发生故障时间设置为 $t_0 = 40$ s。

图6为执行器故障情况系统响应。如图6a~b所示,在DDPG-SMADRC作用下,系统跟踪情况几乎不受影响,而在SMADRC作用下系统则出现了稳态误差。在SMADRC作用下,控制输入出现明显抖振,而在本文方法作用下几乎不受影响,如图6c~d。究其原因,SMADRC作用下ESO的带宽是固定的,无法适应执行器响应延迟,从而加剧了抖振现象。此外,系统其他状态也受到不同程度的影响,见图6e。总体来看,本文方法通过对飞行器状态的学习和自适应调整ESO参数,在面对执行器故障时表现出更强的适应性。

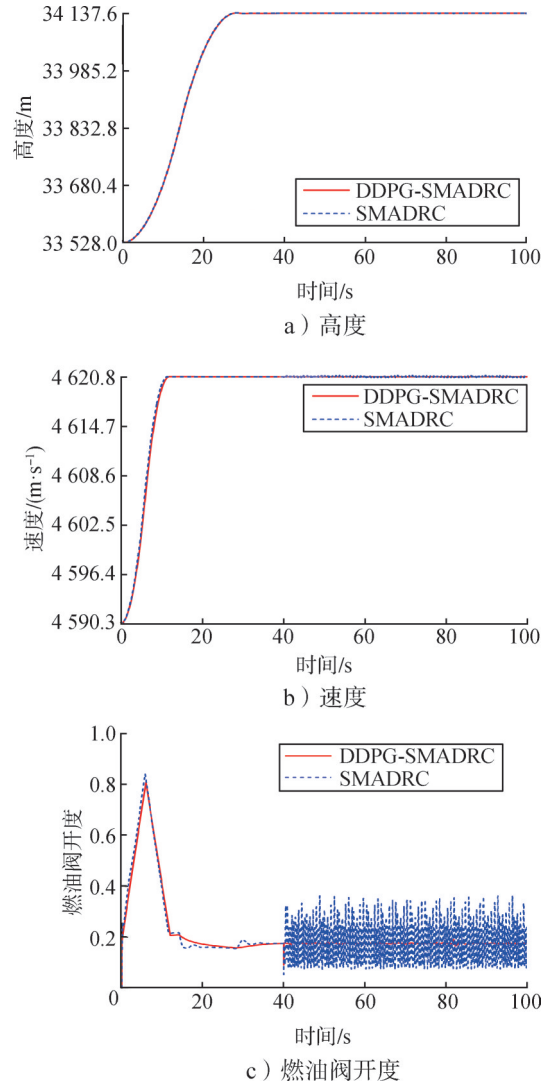
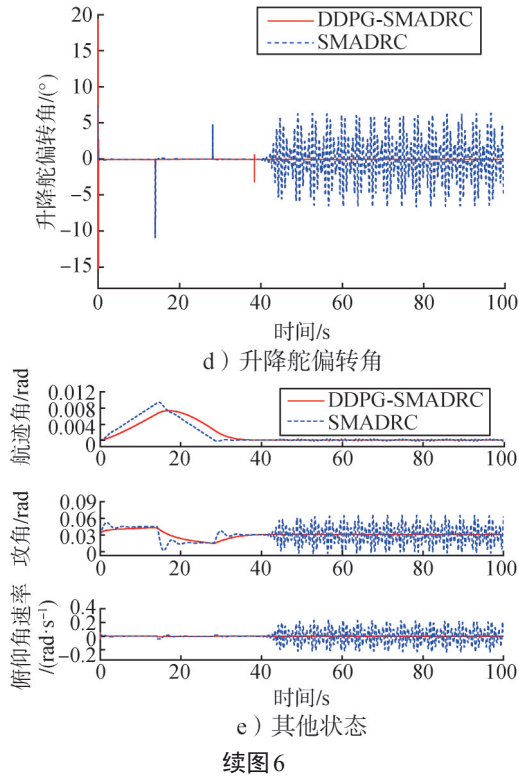


图6 执行器故障情况系统响应

Fig.6 System response under actuator failure



续图6

3.2.4 复合故障情况

为进一步检验方法的性能,测试了同时发生3种干扰下的控制系统性能。

图7为复合故障情况系统响应。如图7a~b所示,在DDPG-SMADRC作用下系统保持了良好的跟踪性能,而在SMADRC作用下系统则出现了稳态误差。在控制输入方面,DDPG-SMADRC系统的高频振荡更小,如图7c~d所示。原因在于SMADRC作用下ESO的带宽是固定的,无法适应执行器响应延迟,从而加剧了抖振现象。而DDPG通过动态优化ESO参数以调整ESO带宽,能够有效抑制抖振。此外,图7e展示了其他状态的收敛情况。综上所述,极端工况下传统SMADRC出现性能恶化,而本文方法凭借其自适应的ESO参数,仍能维持系统稳定并完成基本跟踪任务,展现了更强的适应能力。

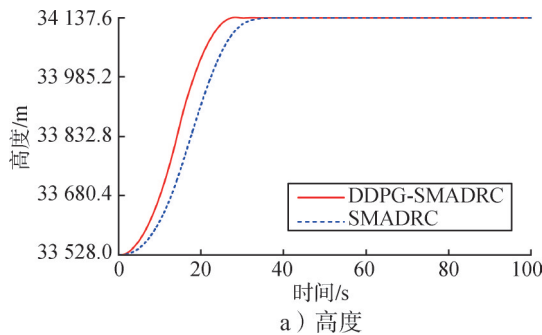
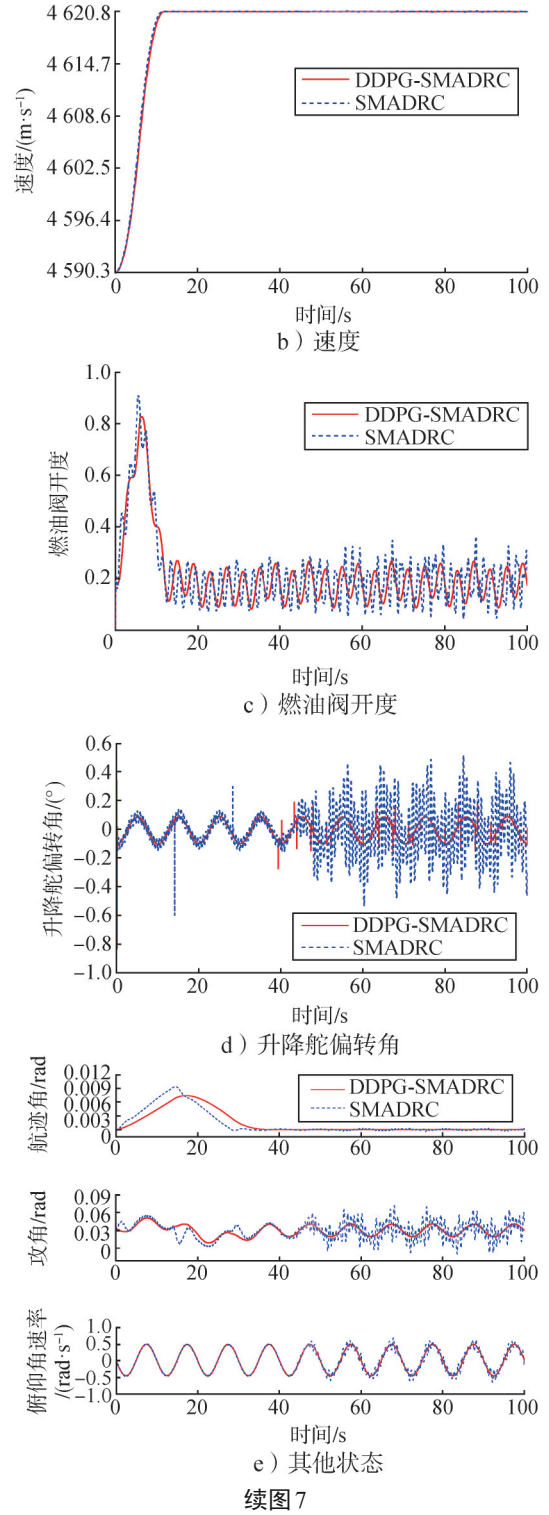


图7 复合故障情况系统响应

Fig.7 System response under multiple fault



续图7

4 结论

针对高超声速飞行器面临的不确定性问题,本文提出了一种将RL与SMADRC结合的控制策略,将RL的环境感知和优化能力与传统ADRC固有的扰动估计与补偿能力相结合,通过DDPG对ESO参数的优化,提升了系统的抗干扰能力。仿真结果表明,所提出的RL-SMADRC控制器相较于SMADRC和SADRC,

展现出了性能优势。不仅提高了速度与高度的跟踪精度,更显著地表现为动态响应过程的加速、面对参数摄动与外部干扰时鲁棒性的实质性增强,以及控制输入信号的平滑性改善。但本文的设计仍存在一些局限性,首先,奖励函数的设计尚依赖工程经验,其系统化构建方法有待探索;其次,DDPG学习过程与控制系统耦合的闭环稳定性理论证明仍是一个挑战;最后,当前方法针对的是纵向模型,将其扩展到完整的六自由度模型并考虑多目标优化将是未来的重点工作。

参 考 文 献

- [1] DING Y, YUE X, CHEN G, et al. Review of control and guidance technology on hypersonic vehicle[J]. Chinese Journal of Aeronautics, 2022, 35(7): 1-18.
- [2] AN H, GUO Z, ZHANG X, et al. Adaptive control of hypersonic vehicles using intelligent allocation[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2023, 237(10): 2383-2401.
- [3] CHENG X, SONG X, WANG L, et al. A model predictive solution to cooperative guidance of hypersonic reentry vehicle with impact angle and distance coordination[J]. Aerospace Science and Technology, 2024(145): 108855.
- [4] WANG F, ZOU Q, ZONG Q, et al. Adaptive prescribed performance fault tolerant control for a flexible air-breathing hypersonic vehicle with uncertainty[J]. IEEE Access, 2019(7): 35018-35033.
- [5] WU G, MENG X, WANG F. Improved nonlinear dynamic inversion control for a flexible air-breathing hypersonic vehicle[J]. Aerospace Science and Technology, 2018(78): 734-743.
- [6] ZHANG S, WANG Q, YANG G, et al. Anti-disturbance backstepping control for air-breathing hypersonic vehicles based on extended state observer[J]. ISA Transactions, 2019(92): 84-93.
- [7] HUANG Z, CHEN Z, ZHENG Y, et al. Optimal design of load frequency active disturbance rejection control via double-chains quantum genetic algorithm[J]. Neural Computing and Applications, 2021(33): 3325-3345.
- [8] SUN J, PU Z, YI J. Conditional disturbance negation based active disturbance rejection control for hypersonic vehicles[J]. Control Engineering Practice, 2019(84): 159-171.
- [9] 朴敏楠, 孙明玮, 黄建, 等. 基于自抗扰的高超声速飞行器姿态鲁棒控制[J]. 控制工程, 2019, 26(9): 1627-1635.
PIAO Minnan, SUN Mingwei, HUANG Jian, et al. Robust attitude control of hypersonic vehicles based on active disturbance rejection[J]. Control Engineering of China, 2019, 26(9): 1627-1635.
- [10] 吴艳, 王丽芳, 李芳. 基于滑模自抗扰的智能车路径跟踪控制[J]. 控制与决策, 2019, 34(10): 2150-2156.
WU Yan, WANG Lifang, LI Fang. Path tracking control of intelligent vehicles based on sliding mode active disturbance rejection[J]. Control and Decision, 2019, 34(10): 2150-2156.
- [11] ZHAO M, HU Y, SONG J. Improved fractional-order extended state observer-based hypersonic vehicle active disturbance rejection control[J]. Mathematics, 2022, 10(23): 4414.
- [12] LIU L, LIU Y, ZHOU L, et al. Cascade ADRC with neural network-based ESO for hypersonic vehicle[J]. Journal of the Franklin Institute, 2023, 360(12): 9115-9138.
- [13] 李争, 张梓豪, 王康涛, 等. 基于无模型的PMLSM改进自适应滑模自抗扰控制[J]. 电机与控制学报, 2024, 28(1): 142-151.
LI Zheng, ZHANG Zihao, WANG Kangtao, et al. Improved adaptive sliding mode active disturbance rejection control based on model-free PMLSM[J]. Journal of Electric Machines and Control, 2024, 28(1): 142-151.
- [14] 唐伟强, 甲成超, 石文科, 等. 高超声速飞行器积分滑模自抗扰控制研究[J]. 现代防御技术, 2024, 52(5): 40-50.
TANG Weiqiang, JIA Chengchao, SHI Wenke, et al. Research on integral sliding mode active disturbance rejection control for hypersonic vehicles[J]. Modern Defense Technology, 2024, 52(5): 40-50.
- [15] PIAO M, YANG Z, SUN M, et al. Synthesis of attitude control for statically unstable hypersonic vehicle with low-frequency aero-servo-elastic effect[J]. Aerospace Science and Technology, 2018(80): 67-77.
- [16] GUO W, SHI B, ZHANG T, et al. Research on active disturbance rejection control of small modular reactor once-through steam generator based on differential evolution optimization algorithm[J]. Nuclear Technology, 2025, 211(5): 974-993.
- [17] CAI Z, LOU J, ZHAO J, et al. Quadrotor trajectory tracking and obstacle avoidance by chaotic grey wolf optimization-based active disturbance rejection control[J]. Mechanical Systems and Signal Processing, 2019(128): 636-654.
- [18] LUONG N C, HOANG D T, GONG S, et al. Applications of deep reinforcement learning in communications and networking: a survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133-3174.
- [19] 高强, 李旭, 吉月辉, 等. 基于Q学习的高超声速飞行器自抗扰控制研究[J]. 控制工程, 2024, 31(4): 577-582.
GAO Qiang, LI Xu, JI Yuehui, et al. Research on active disturbance rejection control for hypersonic vehicles based on Q-learning[J]. Control Engineering of China, 2024, 31(4): 577-582.
- [20] ZHENG Y, SUN Q, CHEN Z, et al. Deep Q-Network based real-time active disturbance rejection controller parameter tuning for multi-area interconnected power systems[J]. Neurocomputing, 2021(460): 360-373.
- [21] CHEN G, CHEN Z, WANG L, et al. Deep deterministic policy gradient and active disturbance rejection controller based coordinated control for gearshift manipulator of driving robot[J]. Engineering Applications of Artificial Intelligence, 2023(117): 105586.
- [22] LI X, JI Y, SONG Y, et al. Modified deep deterministic policy gradient based on active disturbance rejection control for hypersonic vehicles[J]. Neural Computing and Applications, 2024, 36(8): 4071-4081.
- [23] WIEWIORA E, COTTRELL G W, ELKAN C. Principled methods for advising reinforcement learning agents[C]. Washington: Proceedings of the Twentieth International Conference on International Conference on Machine Learning, 2003.
- [24] TANG W, JIA C, SHI W, et al. Switching active disturbance rejection tracking control of hypersonic vehicles[C]. Tianjin: 2024 IEEE 25th China Conference on System Simulation Technology and its Application (CCSSTA), 2024.
- [25] XU H, IOANNOU P A, MIRMIRANI M. Adaptive sliding mode control design for a hypersonic flight vehicle[J]. Journal of Guidance Control and Dynamics, 2004, 27(5): 829-838.

作 者 简 介

唐伟强 (1978—), 男, 博士, 教授, 主要研究方向为飞行器动力学与控制。
马景泰 (1995—), 男, 硕士研究生, 主要研究方向为飞行器动力学与控制。
魏子栋 (2000—), 男, 硕士研究生, 主要研究方向为飞行器动力学与控制。
高海燕 (1987—), 女, 博士, 讲师, 主要研究方向为飞行器制导与控制, 预测控制。