中国科学:信息科学 2026 年 第 56 卷 第 5 期: 1279–1295
SCIENTIA SINICA Informationis
论文

基于 Stackelberg 博弈的敌对目标袭扰下无人船轨迹跟踪控制与避障

文利燕[1], 崔玉康[1], 姜斌[1*], 孟亦真[2], 马亚杰[1]

1. 南京航空航天大学自动化学院, 南京 211106
2. 上海航天控制技术研究所空间目标感知全国重点实验室, 上海 201109
* 通信作者. E-mail: binjiang@nuaa.edu.cn

收稿日期: 2025–04–11; 修回日期: 2025–07–05; 接受日期: 2025–09–16; 网络出版日期: 2026–02–14

国家自然科学基金 (批准号: 62173179, 62020106003) 和江苏省自然科学基金 (批准号: BK20222012) 资助项目

摘要 本文针对无人船在敌对目标策略性追击袭扰下的避障与轨迹跟踪问题, 提出一种基于 Stackelberg 博弈的安全跟踪控制方法. 借鉴人工势场法的思想, 将敌方动态排斥力的策略性干扰建模为博弈框架下的跟随者行为, 并基于 Stackelberg 博弈设计分层控制策略, 即: 分别为跟随者 (排斥力) 和领导者 (控制力) 设计性能指标函数, 并引入非二次型控制输入代价函数实现控制输入饱和约束设计; 进而利用极大值原理, 设计领导者和跟随者的控制策略, 最终实现避障安全性与轨迹跟踪精度的动态平衡. 针对无人船未知系统动力学和外部干扰的非线性特性, 利用积分强化学习算法, 设计神经网络直接推导 Stackelberg 博弈的近似最优解. 所设计的无人船安全轨迹跟踪控制策略可以保证所有闭环系统信号有界, 且能保证无人船在敌对目标袭扰下能够跟踪到参考轨迹. 最后, 进行仿真实验验证本文提出的无人船安全跟踪控制策略的有效性.

关键词 积分强化学习, Stackelberg 博弈, 饱和约束, 避障, 轨迹跟踪

1 引言

无人船作为一种新型的自主航行平台, 特别是在军事、海洋探测及运输领域, 已逐渐成为国内外研究的热点. 无人船具备高度自主的决策能力和强大的环境适应性, 在执行危险或重复性任务方面有着实质性的优势 [1,2]. 但由于无人船自身结构特性及其工作环境的复杂性, 其系统动力学模型具有非线性和参数不确定性等特点, 给无人船的控制设计带来诸多挑战. 对无人船运动控制的研究涉及多个方面, 包括路径规划 [3]、目标跟踪 [4] 和轨迹跟踪 [5~7] 等. 文献 [8] 研究了多变量非线性连续不确定系统的跟踪控制, 提出了线性参数化估计与矩阵分解融合技术, 解决了未知强干扰下的自适应渐近跟踪问题. 值得注意的是, 无人船在沿预定轨迹执行任务时, 除了需要有效抑制环境风浪等扰动外, 还必须考虑环境障碍物和敌对目标追击带来的避障问题. 因此, 如何确保无人船在复杂环境下的安全性与稳定性已成为无人船轨迹跟踪控制关键问题.

目前, 为了提高无人船系统的安全性能, 在无人船避障算法领域出现了诸多研究成果 [9~13], 典型的方法有基于人工势场法的方法 [9,10]、基于模型预测控制的方法 [14~17]、基于强化学习的方法 [11,18]、基于博弈论的方

引用格式: 文利燕, 崔玉康, 姜斌, 等. 基于 Stackelberg 博弈的敌对目标袭扰下无人船轨迹跟踪控制与避障. 中国科学: 信息科学, 2026, 56: 1279–1295, doi: 10.1360/SSI-2025-0144
Wen L Y, Cui Y K, Jiang B, et al. Stackelberg game-based trajectory tracking control and obstacle avoidance for USVs under hostile target interference. Sci Sin Inform, 2026, 56: 1279–1295, doi: 10.1360/SSI-2025-0144

© 2026《中国科学》杂志社

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

法 [19]、基于模糊逻辑控制的方法 [20, 21], 以及基于视觉和传感器融合的方法 [22] 等. 这些方法往往可以应用于无人船系统的避障轨迹规划和跟踪控制, 在一定程度上解决了复杂环境扰动下无人船轨迹跟踪控制的安全性问题. 人工势场法可将无人船在复杂环境下的避障问题转化为简单的受力驱动问题, 且人工势场法和其他控制算法可以有效融合以实现不同的控制目标. 文献 [23] 针对道路宽度和车辆形状约束下与障碍物间的距离保持问题, 设计基于模型预测控制自适应人工势场法, 保证车辆平滑避障. 文献 [24] 针对特殊的避障约束问题, 设计基于事件触发模型预测控制自适应人工势场法, 实现低计算量下电动车辆避障和轨迹跟踪. 势场力的引入虽实现了系统对环境中障碍物的避障, 但同时会作为扰动影响系统轨迹跟踪的性能, 尤其是考虑敌对目标针对我方策略性的追击袭扰过程, 很难快速精确地实现系统的安全可靠跟踪控制.

当涉及双方策略迭代变化问题时, Stackelberg 博弈无疑是一种有效工具. Stackelberg 博弈通过将决策者分为领导者和跟随者, 能够有效描述在不完全信息下, 双方如何通过策略调整来优化自己的目标函数. 文献 [22] 针对多车竞速问题, 设计基于 Nash-Stackelberg 嵌套博弈框架, 实现每辆车的最优轨迹并推断对手规划路径. 文献 [25] 针对自动驾驶车道变换问题, 设计基于 Stackelberg 的车道变换方法, 综合考虑车道变换决策和轨迹规划以确保车道变换的安全稳定. 值得注意的是, Stackelberg 博弈均衡解的求取非常困难, 典型的求取方法有: 基于动态规划的方法 [26, 27]、基于遗传算法的方法 [28], 以及基于强化学习的方法 [29, 30] 等. 考虑实际无人船系统执行器的输出限制, 在控制器的设计过程中应考虑饱和约束限制. 目前针对无人船的避障设计和轨迹跟踪控制研究大多仅考虑复杂环境中的动、静态障碍物, 而对于敌对目标的策略性追击袭扰下考虑执行器饱和约束的无人船轨迹安全跟踪设计鲜有研究. 本文的主要创新与贡献如下所述.

(1) 针对敌对目标的策略性追击袭扰问题, 借鉴人工势场法的思想, 在无人船系统模型中引入可策略性变化的势场排斥力, 建立敌对目标袭扰下含部分未知非线性动力学的无人船系统模型, 进而将敌对目标袭扰下的安全跟踪控制问题转化为势场排斥力策略性变化下的安全跟踪控制问题.

(2) 针对敌对目标袭扰下的无人船轨迹安全跟踪控制问题, 提出一种基于 Stackelberg 博弈的安全跟踪控制方法. 分别建立领导者 (控制力) 和跟随者 (排斥力) 的最优性能指标, 并引入非二次型能量函数, 应对执行器的饱和输入约束限制; 进而基于极大值原理, 得到领导者和跟随者的最优控制器结构和求解最优控制输入的哈密顿 – 雅可比 (Hamiltonian-Jacobi, HJ) 方程.

(3) 针对无人船未知非线性动力学和 HJ 方程强耦合特性导致的 HJ 方程难求解的问题, 设计了基于积分强化学习 (integral reinforcement learning, IRL) 的 Stackelberg 博弈均衡求解方法, 设计基于 actor-critic 的神经网络架构, 得到系统未知非线性动力学下领导者和跟随者最优控制策略, 实现了敌对目标袭扰下无人船闭环系统稳定和渐近跟踪参考轨迹; 并基于无人船模型, 开展了系统性的仿真研究, 证明了所提出方法的有效性.

本文后续内容安排如下: 第 2 节给出敌对目标袭扰下的无人船模型及相应的控制问题; 第 3 节考虑执行器的饱和约束和分层最优控制的协态约束问题, 设计基于 Stackelberg 的安全跟踪控制方法, 得到最优领导者跟随者控制器结构; 第 4 节设计 actor-critic 神经网络求解 Stackelberg 博弈均衡, 得到领导者跟随者最优控制策略, 并给出了稳定性证明; 第 5 节进行了控制器仿真分析; 第 6 节给出主要结论.

2　问题描述

本节给出含有部分未知非线性动力学的无人船系统模型, 同时给出本文所要解决的控制问题.

2.1　系统模型

如图 1 所示, 考虑如下含部分未知非线性动力学的无人船系统模型:

$$\begin{aligned} \dot{\eta} &= R(\eta)v, \\ M\dot{v} &= f(\eta, v) + u(t) + d_1(t), \end{aligned} \tag{1}$$

其中 $\eta = [x, y, \psi]^{\mathrm{T}} \in \mathbb{R}^3$ 是无人船的位姿向量, ψ 为偏航角. $v = [p, q, r]^{\mathrm{T}} \in \mathbb{R}^3$ 代表无人船的速度和偏航角速度,

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

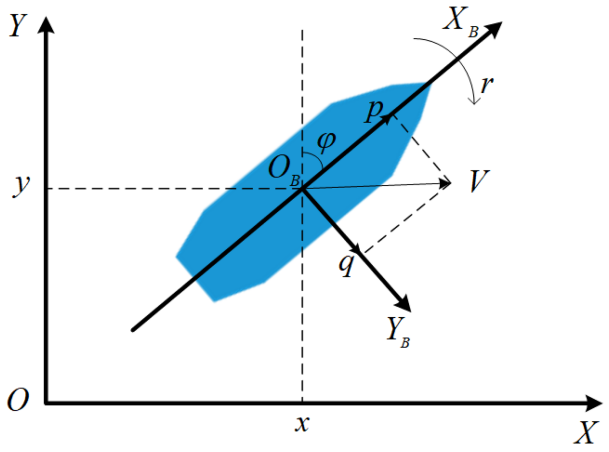


图 1　(网络版彩图) 地球坐标系 OXY 和无人船机体坐标系 $O_BX_BY_B$.

Figure 1　(Color online) Earth-fixed OXY and body-fixed $O_BX_BY_B$ coordinate frames.

r 为偏航角速度. $M \in \mathbb{R}^{3\times3}$ 代表无人船的惯性矩阵, $f(\eta, v)$ 代表无人船未知的非线性动力学, 包括未建模的流体动力学 (如涡流阻尼力、波浪漂移力和向心力等). $u(t) \in \mathbb{R}^3$ 是无人船的控制输入, $d_1(t) \in \mathbb{R}^3$ 表示外部环境扰动, $||d_1(t)|| \leqslant \bar{d} < \infty$, $\bar{d} \in \mathbb{R}^+$. $R(\eta) = [\cos\psi, -\sin\psi, 0; \sin\psi, \cos\psi, 0; 0, 0, 1]$ 是坐标变换矩阵.

避障是本文所考虑的敌对目标袭扰下的无人船安全跟踪控制设计的核心内容, 借鉴传统路径规划算法人工势场法中势函数的思想, 首先, 为保证我方无人船与敌对目标不发生碰撞, 我方无人船会受到敌对目标的排斥力, 其次, 为避免与环境本身存在的障碍物发生碰撞, 我方无人船会受到环境障碍物的排斥力, 排斥力仅取决于我方无人船与敌对目标或障碍物的相对位置. 鉴于此, 无人船系统模型可重新描述为

$$\begin{aligned} \dot{\eta} &= R(\eta)v, \\ M\dot{v} &= f(\eta, v) + u(t) + \omega(t) + d(t), \end{aligned} \tag{2}$$

其中 $d(t) = d_1(t) + d_2(t)$, $d_2(t)$ 为无人船受到环境障碍物的排斥力, $\omega(t) \in \mathbb{R}^3$ 是无人船受到敌对目标的排斥力, 排斥力 $\omega(t)$ 的引入在一定程度上会破坏无人船的跟踪性能. 需要注意的是, 与 $d_2(t)$ 不同的是, 敌对目标会根据我方无人船的实时状态信息及时调整自身控制策略, 即排斥力 $\omega(t)$ 的变化是具有一定策略性的. 现有基于人工势场法的避障研究 (如文献 [9, 10]) 多假设障碍物静态或被动响应, 而敌对目标的策略性排斥力会动态调整, 导致传统势场法无法平衡跟踪精度与避障安全性.

2.2　控制问题

本文所考虑控制问题的一个主要特征是: 在进行控制器设计时, 敌对目标排斥力的策略性变化会持续影响无人船的轨迹跟踪性能, 同时考虑执行器的饱和输入约束和未知的系统非线性动力学使得控制器的设计更加困难. 为了有效处理敌对目标袭扰下具有饱和输入约束的不确定非线性无人船系统安全跟踪控制问题, 本文将基于 Stackelberg 博弈的思想, 设计一种安全跟踪控制方法, 以实现理想的轨迹跟踪性能. 需注意的是: 本文所考虑的轨迹跟踪性能兼顾无人船的位姿和速度跟踪性能, 若分别设计控制策略, 以实现对参考位姿向量和参考速度向量的跟踪, 将大大增加控制设计的复杂程度; 为此, 将利用实际的位姿误差向量和参考位姿向量, 为无人船规划参考速度向量, 以保证无人船速度跟踪的同时实现位姿跟踪, 从而简化控制器的设计过程.

控制目标. 针对敌对目标袭扰下含部分未知非线性动力学的无人船系统 (2), 开发一种基于 Stackelberg 博弈的安全跟踪控制方案, 以保证无人船在安全控制 (与敌对目标不发生碰撞) 的前提下, 保证闭环系统中的所有信号均是有界的, 且系统位姿向量 $\eta(t)$ 和速度向量 $v(t)$ 分别渐近跟踪参考位姿向量 $\eta_r(t)$ 和参考速度向量 $v_r(t)$.

参考模型. 相对于轨迹规划, 本文更侧重于无人船的轨迹跟踪控制策略设计, 故假设参考位姿向量 $\eta_r = [x_r, y_r, \psi_r]^{\mathrm{T}}$ 为已知的期望向量. 此外, 为简化控制器的设计过程, 基于位姿误差向量 $e_\eta = \eta - \eta_r$ 和参考位姿向

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

量 η_r, 设计参考速度向量为

$$v_r = R^{-1}(\eta)(-K_1 e_\eta + \dot{\eta}_r), \tag{3}$$

其中 $K_1 \in \mathbb{R}^{3\times 3}$ 是一个选定的已知正定矩阵. 令速度误差向量为 $e_v = v - v_r$, 由式 (2) 和 (3) 得 $e_v = v - v_r = R^{-1}(\eta)(\dot{e}_\eta + K_1 e_\eta)$. 当无人船的速度误差向量 $e_v \to 0$ 时, $\dot{e}_\eta + K_1 e_\eta \to 0$. 若令 $\dot{e}_\eta + K_1 e_\eta = 0$, 则 $e_\eta(t) = \mathrm{e}^{-K_1 t} x(0)$, 其中 K_1 是正定矩阵, 故易得无人船位姿误差向量 $e_\eta \to 0$. 根据式 (2) 中关于 v 的定义, 得到势场力影响下的速度误差动力学为

$$\dot{e}_v = h(e_v) + g(u + \omega), \tag{4}$$

其中 $h(e_v) = M^{-1}[f(\eta, v) + d(t)] - \dot{v_r}$ 是未知的非线性动力学, $g = M^{-1}$ 表示输入动力学.

Stackelberg 博弈. Stackelberg 博弈是解决双方策略迭代变化的有效方法. Stackelberg 博弈的平衡是通过一组控制变量 $\{u_1^*, u_2^*(u_1^*)\}$ 实现的, 控制变量满足以下性质. (1) 对于任意 u_1, 存在一个唯一的控制变量 $u_2^*(u_1)$ 使得跟随者代价函数 $J_1(u_1, u_2)$ 最小, 即 $J_1(u_1, u_2^*(u_1)) \leqslant J_1(u_1, u_2)$. (2) 领导者旨在寻找最优控制变量 u_1^*, 使得领导者代价函数 $J_2(u_1, u_2^*(u_1))$ 最小, 对于任意控制变量组 $\{u_1, u_2^*(u_1)\}$, 即 $J_2(u_1^*, u_2^*(u_1^*)) \leqslant J_2(u_1, u_2^*(u_1))$. 迭代产生的控制变量组 $\{u_1^*, u_2^*(u_1^*)\}$ 即为 Stackelberg 博弈的最优解.

执行器饱和输入约束. 无人船在接近障碍物时, 根据人工势场法, 其所受到的排斥力会趋于无穷大. 而考虑到实际无人船的执行器输出功率限制, 无人船的控制输入不能无限增大, 此时无人船会在排斥力的作用下优先进行避障以保证我方无人船的安全. 因此, 在控制器的设计过程中必须考虑无人船执行器饱和输入约束设计.

3　基于 Stackelberg 博弈的安全跟踪控制方案

本节针对无人船系统因敌对目标袭扰导致的排斥力策略性注入问题, 在 Stackelberg 博弈框架内开发了考虑敌对目标追击袭扰的安全跟踪控制策略. 在以速度误差向量 e_v 为中心的博弈中, 敌对目标排斥力变化和我方无人船控制输入的变化可表述为一个分层顺序决策过程, 描述如下:

(1) 顺序决策过程源于无人船所采用的初始控制输入 $u(0)$;

(2) 将敌对目标产生的排斥力视为跟随者, 敌对目标通过改变空间位置, 以最优的排斥力 ω^* 最大化跟踪误差 e_η 和 e_v;

(3) 我方无人船根据跟随者产生的排斥力来主动调整控制输入, 以选择最优的控制输入最小化跟踪误差 e_η 和 e_v, 我方无人船的控制输入可视为领导者.

3.1　饱和输入约束设计

考虑敌对目标过于接近我方无人船时, 应重点考虑无人船的安全控制, 同时当敌方无人船与我方无人船相撞时, 其产生的排斥力会趋于无穷大, 而无人船的控制输入不能无限增大. 因此在构造领导者和跟随者的代价函数之前, 应优先考虑无人船饱和输入约束设计. 当敌对目标接近我方到某一程度, 执行器产生的控制输入无法消除排斥力产生的跟踪误差, 无人船只能在排斥力的作用下进行避障. 具有饱和输入约束的 Stackelberg 博弈将与不考虑饱和输入约束完全不同, 这使得分层控制更加复杂. 为了处理无人船控制输入的饱和约束问题, 在设计性能指标函数之前, 设计 u 的非二次型函数代替原来的二次型形式 $(u^{\mathrm{T}} R u)$:

$$M(u) = 2\int_0^u \lambda \tanh^{-\mathrm{T}}(v/\lambda) R dv. \tag{5}$$

记

$$\tanh(u) = \begin{bmatrix} \tanh(u_1) \\ \tanh(u_2) \\ \tanh(u_3) \end{bmatrix}, \quad \tanh^{-\mathrm{T}}(u) = \begin{bmatrix} \tanh^{-1}(u_1) \\ \tanh^{-1}(u_2) \\ \tanh^{-1}(u_3) \end{bmatrix}^{\mathrm{T}},$$

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

其中领导者权重矩阵 R 是对称正定矩阵. $\tanh(u)$ 是对应于 u 的双曲正切函数, 该函数具有单调性和有界导数. $\tanh^{-1}(u)$ 是 $\tanh(u)$ 的反函数. λ 是大于 0 的常数, 通过调整 λ 的大小可直接改变无人船控制输入的约束幅值 $u_{\max}$, 式 (5) 可用积分除法展开, 关于 u 的非二次型函数可表示如下:

$$M(u) = 2\lambda \tanh^{-\mathrm{T}}(u/\lambda) R u + \lambda^2 \bar{R} \ln(\underline{1} - (u/\lambda)^2), \tag{6}$$

其中 $\bar{R}$ 是由 R 对角元素组成的行向量, $\underline{1} = [1,1,1]^{\mathrm{T}}$.

3.2　安全跟踪控制设计

本小节针对 2.2 小节所提出的控制问题, 分别为跟随者和领导者设计其表征控制器性能的代价函数, 通过最小化代价函数可以得到领导者和跟随者的最优控制器结构. 值得注意的是, 跟随者的决策取决于领导者的控制策略, 领导者的决策取决于跟随者的理性选择, 双方都在努力使其代价函数最小化. 跟随者和领导者的代价函数可分别定义为

$$\begin{aligned} J_1(e_v^0, \omega, u) &= \int_0^\infty r_1(e_v, \omega, u) ds = \int_0^\infty (\omega^{\mathrm{T}} G \omega - e_v^{\mathrm{T}} Q e_v - M(u)) ds, \\ J_2(e_v^0, \omega, u) &= \int_0^\infty r_2(e_v, \omega, u) ds = \int_0^\infty (e_v^{\mathrm{T}} Q e_v + M(u)) ds, \end{aligned} \tag{7}$$

其中, J_1, J_2 分别为跟随者和领导者的代价函数, 跟随者权重矩阵 G, 误差权重矩阵 Q 为对称正定矩阵. 跟随者的代价函数最小化意味着增大速度误差向量 e_v, 领导者的代价函数最小化意味着减小速度误差向量 e_v. 跟随者 ω 和领导者 u 在 t 时刻的性能指标函数可表示为

$$\begin{aligned} V_1(e_v(t), \omega, u) &= \int_t^\infty r_1(e_v, \omega, u) ds, \\ V_2(e_v(t), \omega, u) &= \int_t^\infty r_2(e_v, \omega, u) ds. \end{aligned} \tag{8}$$

跟随者最优控制器. 基于当前领导者的控制策略 u, 可定义跟随者的最优性能指标函数为

$$V_1^*(e_v^0) = \min_{\omega} \int_0^\infty r_1(e_v, \omega, u) ds, \tag{9}$$

其中 $r_1(e_v, \omega, u) = \omega^{\mathrm{T}} G \omega - e_v^{\mathrm{T}} Q e_v - M(u)$. 考虑速度误差动力学 (4), 将跟随者的哈密顿函数定义为

$$H_1(e_v, \nabla V_1, \omega, u) = r_1(e_v, \omega, u) + \nabla V_1^{\mathrm{T}} (h(e_v) + g\omega + gu), \tag{10}$$

其中 $\nabla V_1 = \partial V_1 / \partial e_v$, 根据 $\partial H_1(e_v, \nabla V_1, \omega, u)/\partial \omega = 0$, 跟随者的最优控制策略可推导为

$$\omega^* = -\frac{1}{2} G^{-1} g^{\mathrm{T}} \nabla V_1. \tag{11}$$

根据 $\dot{\lambda}_1 = -\left[\frac{\partial H_1(e_v, \nabla V_1, \omega, u)}{\partial e_v}\right]^{\mathrm{T}}$, 可得跟随者的协态方程为

$$\dot{\lambda}_1 = 2Q e_v - \nabla h^{\mathrm{T}} \nabla V_1, \tag{12}$$

其中, $\nabla h = \partial h / \partial e_v$. 跟随者的协态方程中包含无人船系统未知的非线性动力学 $h(e_v)$.

领导者最优控制器. 跟随者采用最优控制策略 (11) 作为其预定策略, 跟随者的协态约束 (12) 也进一步影响领导者的控制策略, 协态约束量化了领导者和跟随者策略的交互影响. 领导者通过考虑跟随者的预定策略 (11) 和跟随者协态约束方程 (12) 制定自身控制策略, 其协态约束下最优性能指标函数可定义为

$$V_2^*(e_v^0) \quad = \min_{u} \int_0^\infty \left[r_2(e_v, \omega, u) + \beta^{\mathrm{T}} \dot{\lambda}_1 \right] ds, \tag{13}$$

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

其中, $r_2(e_v,\omega,u)=e_v^{\mathrm{T}}Qe_v+M(u)$, β 代表拉格朗日乘子. 考虑速度误差动力学 (4) 和协态约束 $\beta^{\mathrm{T}}\dot{\lambda}_1$, 将领导者的哈密顿函数定义为

$$H_2(e_v,\nabla V_2,\omega^*,u)=r_2(e_v,\omega^*,u)+\beta^{\mathrm{T}}\dot{\lambda}_1+\nabla V_2^{\mathrm{T}}(h(e_v)+gu+g\omega^*),\tag{14}$$

其中 $\nabla V_2=\partial V_2/\partial e_v$, 根据 $\partial H_2(e_v,\nabla V_2,\omega^*,u)/\partial u=0$, 领导者的最优控制策略可推导为

$$u^*=-\lambda\tanh\left(\frac{R^{-1}g^{\mathrm{T}}\nabla V_2}{2\lambda}\right).\tag{15}$$

根据协态方程 (12) 和拉格朗日乘子算式 $\dot{\beta}=-\left[\frac{\partial H_2(e_v,\nabla V_1,\omega^*,u)}{\partial\nabla V_1}\right]^{\mathrm{T}}$, 可得拉格朗日乘子的更新率

$$\dot{\beta}=\frac{1}{2}gG^{-1}g^{\mathrm{T}}\nabla V_2+\nabla h\beta.\tag{16}$$

领导者 – 跟随者耦合 HJ 方程. 考虑式 (15) 和 (11) 的最优领导者 – 跟随者控制器结构, 性能指标函数的梯度 ∇V_1 和 ∇V_2 可通过计算以下 HJ 方程得到:

$$\begin{aligned}0&=r_1(e_v,\omega^*,u)+\nabla V_1^{*\mathrm{T}}(h(e_v)+gu+g\omega^*),\\0&=r_2(e_v,\omega^*,u^*)+\nabla V_2^{*\mathrm{T}}(h(e_v)+gu^*+g\omega^*).\end{aligned}\tag{17}$$

由上述耦合 HJ 方程得到的博弈均衡解是存在且唯一的, 文献 [31] 基于类 Lyapunov 方程和最优理论分析, 对分层 Stackelberg 博弈最优解存在且唯一进行系统证明, 为此本文不再赘述. 值得注意的是, 耦合 HJ 方程具有非线性特性, 同时耦合 HJ 方程, 跟随者协态方程 (12) 和拉格朗日乘子更新率 (16) 中均存在系统未知非线性动力学 $h(e_v)$, 使得 Stackelberg 的求解非常困难, 因此需要一种有效的算法求解 Stackelberg 博弈.

4　基于 IRL 的 Stackelberg 博弈求解

由于耦合 HJ 方程的非线性特性和未知的非线性动力学 $h(e_v)$ 的存在, 推导 Stackelberg 博弈解析解是极具挑战性的. 因此, 本节针对性能指标函数梯度 ∇V_1 和 ∇V_2 的求解问题, 设计一种新的基于 actor-critic 神经网络框架的积分强化学习算法. 同时针对系统的未知非线性动力学 $h(e_v)$, 设计一个额外的神经网络进行近似求解, 得到神经网络下跟随者和领导者的最优控制策略. 通过设计贝尔曼误差迭代逼近最优控制策略下的值函数是求解非线性耦合 HJ 方程的一种有效方法. 最后通过构造 Lyapunov 函数证明了闭环系统所有信号的有界性和系统的稳定性.

4.1　IRL 神经网络设计

本小节设计 3 个神经网络分别估计跟随者和领导者的最优性能指标函数和系统未知非线性动力学, 进而得到领导者和跟随者的最优控制策略. 如图 2 所示, 由系统模型和参考模型得到的速度跟踪误差 e_v 分别输入到跟随者网络、领导者网络和未知非线性动力学网络. 跟随者网络基于领导者控制指令 $\hat{u}$ 和速度跟踪误差 e_v 构造跟随者控制指令 $\hat{\omega}$; 未知非线性动力学网络近似无人船系统未知非线性动力学 $\hat{h}(e_v)$ 并辅助计算领导者与跟随者之间的协态约束; 领导者网络基于跟随者控制指令 $\hat{\omega}$, 速度跟踪误差 e_v 和协态约束得到领导者最优控制指令 $\hat{u}$. 具体神经网络结构解释如下.

跟随者网络策略生成. 考虑使用理想最优跟随者 critic 神经网络近似跟随者的最优性能指标函数 V_1^*, 定义 V_1^* 及其梯度 ∇V_1^* 如下:

$$V_1^*=W_{1c}^{*\mathrm{T}}\varphi_{1c}+\varepsilon_{1c},\nabla V_1^*=\nabla\varphi_{1c}^{\mathrm{T}}W_{1c}^*+\nabla\varepsilon_{1c},\tag{18}$$

其中 $W_{1c}^*\in\mathbb{R}^N$ 表示跟随者 critic 神经网络有界最优权重, 即 $\parallel W_{1c}^*\parallel\leqslant w_{1c}$, N 表示该神经网络基函数数量, $\varphi_{1c}=\varphi_{1c}(e_v)$, $\varepsilon_{1c}=\varepsilon_{1c}(e_v)$ 分别表示有界基函数和近似误差, $\nabla V_1^*=\partial V_1^*/\partial e_v$ 和 $\nabla\varepsilon_{1c}=\partial\varepsilon_{1c}/\partial e_v$ 是有界的.

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

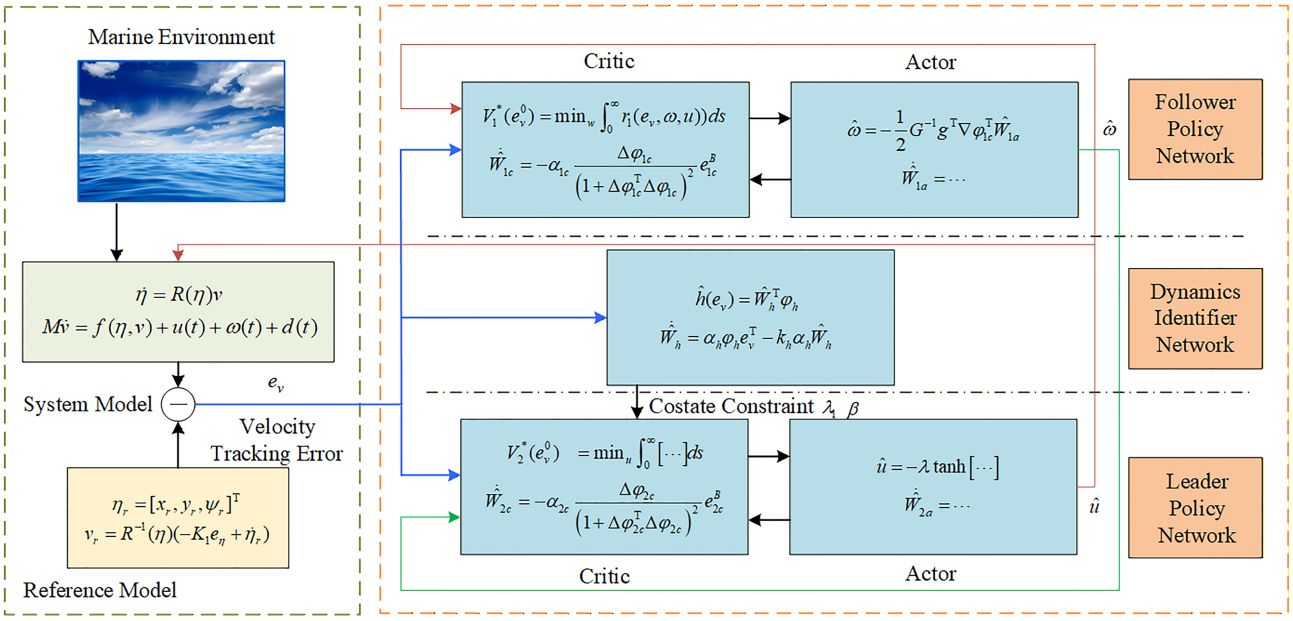


图 2　(网络版彩图) 基于 IRL 的无人船安全跟踪控制 Stackelberg 博弈求解框架.

Figure 2　(Color online) Stackelberg game solution framework for IRL-based safe tracking control of USVs.

设计实际跟随者 critic 神经网络, 驱动跟随者实际性能指标函数 $\hat{V}_1$ 和 $\nabla\hat{V}_1$ 分别向 V_1^* 和 ∇V_1^* 逼近, 其中 $\hat{V}_1$ 和 $\nabla\hat{V}_1$ 可表示如下:

$$\hat{V}_1 = \hat{W}_{1c}^{\mathrm{T}}\varphi_{1c}, \nabla\hat{V}_1 = \nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1c}, \tag{19}$$

其中 $\nabla\hat{V}_1 = \partial\hat{V}_1/\partial e_v$. 为了避免在整个学习过程中使用未知非线性动力学 $h(e_v)$, 考虑以下跟随者贝尔曼误差方程:

$$e_{1c}^B = \int_{t-h}^{t} r_1(e_v, \hat{\omega}, \hat{u})ds + \hat{W}_{1c}^{\mathrm{T}}\Delta\varphi_{1c}, \tag{20}$$

其中, h 是一次迭代学习的时间步长, $r_1(e_v, \hat{\omega}, \hat{u}) = \hat{\omega}^{\mathrm{T}}G\hat{\omega} - e_v^{\mathrm{T}}Qe_v - M(\hat{u})$, 其中含有有关待设计领导者控制策略 $\hat{u}$ 的非二次型函数 $M(\hat{u})$, $\Delta\varphi_{1c} = \varphi_{1c}(t) - \varphi_{1c}(t-h)$. 以通过调整实际跟随者 critic 神经网络权重 $\hat{W}_{1c}$ 使误差 e_{1c}^B 最小化为目的, 目标函数可设为 $E_{1c} = 1/2(e_{1c}^B)^2$, 利用梯度下降法可得 $\hat{W}_{1c}$ 的更新律为

$$\dot{\hat{W}}_{1c} = -\alpha_{1c}\frac{\Delta\varphi_{1c}}{\left(1+\Delta\varphi_{1c}^{\mathrm{T}}\Delta\varphi_{1c}\right)^2}e_{1c}^B, \tag{21}$$

其中 α_{1c} 表示跟随者 critic 神经网络的学习率. 为了确保跟随者控制策略和性能指标函数更新的稳定性, 跟随者控制策略可表示为

$$\hat{\omega} = -\frac{1}{2}G^{-1}g^{\mathrm{T}}\nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1a}, \tag{22}$$

其中跟随者 actor 神经网络权重 $\hat{W}_{1a}$ 是 W_{1c}^* 权重的估计值. 权重 $\hat{W}_{1a}$ 的更新率可设计为

$$\begin{aligned}\dot{\hat{W}}_{1a} = -\alpha_{1a}\Bigg[&\frac{h}{\left(1+\Delta\varphi_{2c}^{\mathrm{T}}\Delta\varphi_{2c}\right)^2}\nabla\varphi_{1c}\nabla\varphi_h^{\mathrm{T}}\hat{W}_h\beta\Delta\varphi_{2c}^{\mathrm{T}}\hat{W}_{2c}\\ &-\frac{h}{4(1+\Delta\varphi_{1c}^{\mathrm{T}}\Delta\varphi_{1c})^2}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1a}\Delta\varphi_{1c}^{\mathrm{T}}\hat{W}_{1c} - \frac{1}{2}D_{1c}e_v\Bigg] - k_{1a}\alpha_{1a}\hat{W}_{1a},\end{aligned} \tag{23}$$

其中 α_{1a} 为跟随者 actor 神经网络的学习率, $D_{1c} = \nabla\varphi_{1c}gG^{-1}g^{\mathrm{T}}$, D_{1c} 是有界的, 即 $\|D_{1c}\| \leqslant d_{1c}$, k_{1a} 是正常数. 考虑领导者最优性能指标函数含有协态约束项, 因此, 在设计领导者的控制优化网络之前需先对无人船系统的未知非线性动力学 $h(e_v)$ 进行近似求解.

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

未知非线性动力学网络近似. 协态约束中 $\dot{\lambda}_1$ 和 β 项均包含未知非线性系统动力学 $h(e_v)$, 因此必须引入神经网络来近似 $h(e_v)$. 由此, 未知非线性动力学 $h(e_v)$ 及其梯度 ∇h 在理想权重神经网络的估计下可表示为

$$h(e_v) = W_h^{*\mathrm{T}}\varphi_h + \varepsilon_h, \nabla h = \nabla\varphi_h^{\mathrm{T}} W_h^* + \nabla\varepsilon_h, \tag{24}$$

其中 $W_h^* \in \mathbb{R}^{N\times 3}$ 是未知非线性动力学 $h(e_v)$ 估计神经网络的有界最优权重, 即 $\| W_h^* \| \leqslant w_h$. $\varphi_h = \varphi_h(e_v)$, $\varepsilon_h = \varepsilon_h(e_v)$ 分别表示该神经网络有界基函数和近似误差, 即 $\| \varphi_h \| \leqslant b_{\varphi_h}$, $\| \varepsilon_h \| \leqslant b_{\varepsilon_h}$, $\nabla\varphi_h = \partial\varphi_h/\partial e_v$, $\nabla\varepsilon_h = \partial\varepsilon_h/\partial e_v$. 考虑实际权重神经网络下未知系统动力学 $h(e_v)$ 及其梯度 ∇h 的估计可表示为

$$\hat{h}(e_v) = \hat{W}_h^{\mathrm{T}}\varphi_h, \nabla\hat{h} = \nabla\varphi_h^{\mathrm{T}}\hat{W}_h, \tag{25}$$

其中 $\nabla\hat{h} = \partial\hat{h}(e_v)/\partial e_v$, 选取目标函数 $E_h = 1/2 e_v^{\mathrm{T}} e_v$ 驱动权重 $\hat{W}_h$ 向 W_h^* 逼近, 利用梯度下降法推导出权值 $\hat{W}_h$ 的更新率为

$$\dot{\hat{W}}_h = \alpha_h\varphi_h e_v^{\mathrm{T}} - k_h\alpha_h\hat{W}_h, \tag{26}$$

其中 α_h 为未知非线性动力学神经网络的学习率, k_h 是正常数.

领导者网络决策优化. 近似领导者的最优性能指标函数可通过构造类似跟随者 actor-critic 神经网络相似的方式. 通过构造神经网络直接近似领导者的最优性能指标函数 $\hat{V}_2$ 及其偏导数, 代入领导者最优控制器 (15) 可得到领导者的最优控制策略. 此外, 为避免完全黑箱优化, 利用值函数分解技术, 可将系统状态信息 (速度跟踪误差 e_v 和未知非线性动力学) 融入神经网络设计过程. 首先, 领导者最优性能指标函数 V_2^* 可分解为

$$V_2^* = e_v^{\mathrm{T}} K_z e_v + 2(e_\eta^{\mathrm{T}} R(\eta) + \hat{h}(e_v)^{\mathrm{T}})K_{eh}e_v + \overline{V_2}^*, \tag{27}$$

其中 $\overline{V}_2^* = V_2^* - e_v^{\mathrm{T}}K_z e_v + 2(e_\eta^{\mathrm{T}}R(\eta) + \hat{h}(e_v)^{\mathrm{T}})K_{eh}e_v$, $K_z = K_2 K_{eh}$, $K_2 \in \mathbb{R}^{3\times 3}$ 是一个正定矩阵, $K_{eh} = (g^{\mathrm{T}})^{-1}R^{\mathrm{T}}g^{-1}$. 根据式 (15) 和 (27) 可得

$$u^* = -\lambda\tanh\left[\frac{1}{\lambda}g^{-1}(K_2 e_v + R(\eta)^{\mathrm{T}}e_\eta + \hat{W}_h^{\mathrm{T}}\varphi_h) + \frac{1}{2\lambda}R^{-1}g^{\mathrm{T}}\nabla\overline{V}_2^*\right], \tag{28}$$

其中 $\nabla\overline{V_2}^* = \partial\overline{V_2}^*/\partial e_v$. 考虑使用理想最优领导者 critic 神经网络近似部分领导者的最优性能指标函数 $\overline{V_2}^*$, $\overline{V_2}^*$ 及其梯度 $\nabla\overline{V_2}^*$ 可定义如下:

$$\overline{V_2}^* = W_{2c}^{*\mathrm{T}}\varphi_{2c} + \varepsilon_{2c}, \nabla\overline{V_2}^* = \nabla\varphi_{2c}^{\mathrm{T}}W_{2c}^* + \nabla\varepsilon_{2c}, \tag{29}$$

其中 $W_{2c}^* \in \mathbb{R}^N$ 表示领导者 critic 神经网络的有界理想最优权重, 即 $\| W_{2c}^* \| \leqslant w_{2c}$, $\varphi_{2c} = \varphi_{2c}(e_v)$, $\varepsilon_{2c} = \varepsilon_{2c}(e_v)$ 分别表示该神经网络的基函数和近似误差. 同样, 利用实际领导者 critic 神经网络对 $\overline{V_2}^*$ 和 $\nabla\overline{V_2}^*$ 的估计可表示为

$$\hat{\overline{V}}_2 = \hat{W}_{2c}^{\mathrm{T}}\varphi_{2c}, \nabla\hat{\overline{V}}_2 = \nabla\varphi_{2c}^{\mathrm{T}}\hat{W}_{2c}, \tag{30}$$

其中 $\nabla\hat{\overline{V}}_2 = \partial\hat{\overline{V}}_2/\partial e_v$. 将式 (19), (25) 和 (30) 代入式 (12) 和 (16) 中, $\dot{\lambda}_1$ 和 β 项可表示如下:

$$\begin{aligned}
&\dot{\lambda}_1 = 2Qe_v - \hat{W}_h^{\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1a},\\
&\dot{\beta} = \frac{1}{2}gG^{-1}g^{\mathrm{T}}\left[2K_z e_v + 2(e_\eta^{\mathrm{T}}R(\eta) + \varphi_h^{\mathrm{T}}\hat{W}_h)K_{eh} + \nabla\varphi_{2c}^{\mathrm{T}}\hat{W}_{2a}\right] + \nabla\varphi_h^{\mathrm{T}}\hat{W}_h\beta.
\end{aligned} \tag{31}$$

考虑系统的协态约束以及避免使用系统的未知非线性动力学 $h(e_v)$, 构建领导者的贝尔曼误差方程为

$$e_{2c}^B = \int_{t-h}^{t}[r_2(e_v, \hat{\omega}, \hat{u}) + \beta^{\mathrm{T}}\dot{\lambda}_1]ds + \hat{W}_{2c}^{\mathrm{T}}\varphi_{2c}, \tag{32}$$

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

其中 $r_2(e_v,\hat{\omega},\hat{u})=e_v^{\mathrm{T}}Qe_v+M(\hat{u})$, $\Delta\varphi_{2c}=\varphi_{2c}(t)-\varphi_{2c}(t-h)$. 选取目标函数 $E_{2c}=1/2(e_{2c}^B)^2$ 驱动权重 $\hat{W}_{2c}$ 向 W_{2c}^* 逼近, 利用梯度下降法推导出权值 $\hat{W}_{2c}$ 的更新率为

$$\dot{\hat{W}}_{2c}=-\alpha_{2c}\frac{\Delta\varphi_{2c}}{\left(1+\Delta\varphi_{2c}^{\mathrm{T}}\Delta\varphi_{2c}\right)^2}e_{2c}^B, \tag{33}$$

其中 α_{2c} 为领导者 critic 神经网络的学习率. 为了确保领导者控制策略和性能指标函数更新的稳定性, 领导者控制策略可表示为

$$\hat{u}=-\lambda\tanh\left[\frac{1}{\lambda}g^{-1}(K_2e_v+R(\eta)^{\mathrm{T}}e_\eta+\hat{W}_h^{\mathrm{T}}\varphi_h)+\frac{1}{2\lambda}R^{-1}g^{\mathrm{T}}\varphi_{2c}^{\mathrm{T}}\hat{W}_{2a}\right], \tag{34}$$

其中领导者 actor 神经网络权重 $\hat{W}_{2a}$ 是 W_{2c}^* 权重的估计值. 权重 $\hat{W}_{2a}$ 的更新率可设计为

$$\dot{\hat{W}}_{2a}=\alpha_{2a}\left[\frac{h}{4(1+\Delta\varphi_{2c}^{\mathrm{T}}\Delta\varphi_{2c})^2}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}\hat{W}_{2a}\Delta\varphi_{2c}^{\mathrm{T}}\hat{W}_{2c}+\frac{1}{2}D_{2c}e_v\right]-k_{2a}\alpha_{2a}\hat{W}_{2a}, \tag{35}$$

其中 α_{2a} 为领导者 actor 神经网络的学习率, $D_{2c}=\nabla\varphi_{2c}gG^{-1}g^{\mathrm{T}}$, D_{2c} 是有界的, 即 $\|D_{_{2c}}\|\leqslant d_{_{2c}}$, k_{2a} 是正常数.

4.2　稳定性分析

本小节对无人船系统信号的有界性与稳定性进行分析.

定理1　考虑敌对目标袭扰下含未知非线性动力学无人船系统 (2), 其耦合 HJ 方程 (17) 在 Stackelberg 博弈框架下有解, 则可设计跟随者控制器 (22)、领导者控制器 (34)、协态约束项 (31) 及神经网络权重更新律 (21), (23), (26), (33) 和 (35), 保证所有闭环系统信号有界, 且系统状态 $\eta(t)$ 渐近跟踪参考信号 $\eta_r(t)$.

证明　考虑如下 Lyapunov 函数:

$$V(t)=V_1+V_2+V_3, \tag{36}$$

其中 $V_1=e_\eta^{\mathrm{T}}e_\eta/2+e_v^{\mathrm{T}}e_v/2+\mathrm{tr}(\tilde{W}_h^{\mathrm{T}}\alpha_h^{-1}\tilde{W}_h)/2$, $V_2=\tilde{W}_{1a}^{\mathrm{T}}\alpha_{1a}^{-1}\tilde{W}_{1a}/2+\tilde{W}_{1c}^{\mathrm{T}}\alpha_{1c}^{-1}\tilde{W}_{1c}/2$, $V_3=W_{2a}^{\mathrm{T}}\alpha_{2a}^{-1}\tilde{W}_{2a}/2+\tilde{W}_{2c}^{\mathrm{T}}\alpha_{2c}^{-1}\tilde{W}_{2c}/2$, $\tilde{W}_h=W_h^*-\hat{W}_h$, $\tilde{W}_{1a}=W_{1c}^*-\hat{W}_{1a}$, $\tilde{W}_{1c}=W_{1c}^*-\hat{W}_{1c}$, $\tilde{W}_{2a}=W_{2c}^*-\hat{W}_{2a}$, $\tilde{W}_{2c}=W_{2c}^*-\hat{W}_{2c}$.

根据式 (3), (22), (26) 和 (34), V_1 对时间的导数为

$$\begin{aligned}\dot{V}_1&=e_\eta^{\mathrm{T}}\dot{e}_\eta+e_v^{\mathrm{T}}\dot{e}_v-\mathrm{tr}(\tilde{W}_h^{\mathrm{T}}\alpha_h^{-1}\dot{\hat{W}}_h)\\&=e_\eta^{\mathrm{T}}[R(\eta)(e_v+v_r)-\eta_r]+e_v^{\mathrm{T}}[h(e_v)+g(\hat{u}+\hat{\omega})]-\mathrm{tr}(\tilde{W}_h^{\mathrm{T}}(\varphi_he_v^{\mathrm{T}}-k_h\hat{W}_h))\\&\leqslant-e_\eta^{\mathrm{T}}K_1e_\eta-e_v^{\mathrm{T}}K_2e_v-k_h\mathrm{tr}(\tilde{W}_h^{\mathrm{T}}\tilde{W}_h)+\frac{1}{2}e_v^{\mathrm{T}}D_{1c}^{\mathrm{T}}\tilde{W}_{1a}+\frac{1}{2}e_v^{\mathrm{T}}D_{2c}^{\mathrm{T}}\tilde{W}_{2a}\\&\quad+b_z\parallel e_v\parallel+k_hw_h\parallel\tilde{W}_h\parallel_F,\end{aligned} \tag{37}$$

其中 $b_z=b_{\varepsilon_h}+(d_{1c}w_{1c}+d_{2c}w_{2c})$, V_2 对时间的导数为 $\dot{V}_2=-\tilde{W}_{1c}^{\mathrm{T}}\alpha_{1c}^{-1}\dot{\hat{W}}_{1c}-\tilde{W}_{1a}^{\mathrm{T}}\alpha_{1a}^{-1}\dot{\hat{W}}_{1a}$, 先计算第一项 $\dot{V}_{21}=-\tilde{W}_{1c}^{\mathrm{T}}\alpha_{1c}^{-1}\dot{\hat{W}}_{1c}$, 根据式 (21) 和 $\xi_\omega^{hj}=\omega^{*\mathrm{T}}G\omega^*-e_v^{\mathrm{T}}Qe_v-\hat{u}^{\mathrm{T}}R\hat{u}+W_{1c}^{*\mathrm{T}}\nabla\varphi_{1c}[h(e_v)+g(\hat{\omega}+u^*)]$, 可得

$$\begin{aligned}\dot{V}_{21}&=\tilde{W}_{1c}^{\mathrm{T}}\frac{\Delta\varphi_{1c}}{m_{1c}}\left[\int_{t-h}^{t}(\hat{\omega}^{\mathrm{T}}G\hat{\omega}-e_v^{\mathrm{T}}Qe_v-\hat{u}^{\mathrm{T}}R\hat{u})ds+\hat{W}_{1c}^{\mathrm{T}}\Delta\varphi_{1c}\right]\\&=k_{1c}\tilde{W}_{1c}^{\mathrm{T}}\hat{W}_{1c}+\frac{\overline{h}_{1c}}{m_{1c}}\tilde{W}_{1c}^{\mathrm{T}}\Delta\varphi_{1c}+\tilde{W}_{1c}^{\mathrm{T}}\frac{\Delta\varphi_{1c}}{m_{1c}}\int_{t-h}^{t}\left(\hat{\omega}^{\mathrm{T}}G\hat{\omega}-\omega^{*\mathrm{T}}G\omega^*-\frac{1}{2}W_{1c}^{*\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}\tilde{W}_{1a}\right)ds\\&=-k_{1c}\tilde{W}_{1c}^{\mathrm{T}}\tilde{W}_{1c}+\left(k_{1c}W_{1c}^{*\mathrm{T}}+\frac{\overline{h}_{1c}}{m_{1c}}\Delta\varphi_{1c}\right)\tilde{W}_{1c}-\frac{h}{2m_{1c}}\tilde{W}_{1c}^{\mathrm{T}}\Delta\varphi_{1c}W_{1c}^{*\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}\tilde{W}_{1a}\\&\quad-\frac{h}{4m_{1c}}\tilde{W}_{1c}^{\mathrm{T}}\Delta\varphi_{1c}\tilde{W}_{1a}^{\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}W_{1c}^*+\frac{h}{4m_{1c}}\tilde{W}_{1a}^{\mathrm{T}}\Delta\varphi_{1c}\hat{W}_{1a}^{\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1c}\\&\quad+\frac{h}{4m_{1c}}\tilde{W}_{1a}^{\mathrm{T}}\Delta\varphi_{1c}\tilde{W}_{1a}^{\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}W_{1c}^*-\frac{h}{4m_{1c}}\tilde{W}_{1a}^{\mathrm{T}}\Delta\varphi_{1c}W_{1c}^{*\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}W_{1c}^*,\end{aligned} \tag{38}$$

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

其中 $m_{1c}=(1+\Delta\varphi_{1c}^{\mathrm{T}}\Delta\varphi_{1c})^2$, $\overline{h}_{1c}=\int_{t-h}^{t}\xi_{\omega}^{hj}-W_{1c}^{*\mathrm{T}}\nabla\varphi_{1c}[h(e_v)+g(u^*+\omega^*)]ds$, $k_{1c}=(\Delta\varphi_{1c}\Delta\varphi_{1c}^{\mathrm{T}})/m_{1c}$, 根据式 (23), 可得 $\dot{V}_2$ 的第二项 $\dot{V}_{22}=-\tilde{W}_{1a}^{\mathrm{T}}\alpha_{1a}^{-1}\dot{\hat{W}}_{1a}$ 为

$$\begin{aligned}\dot{V}_{22}=&-k_{1a}\tilde{W}_{1a}^{\mathrm{T}}\tilde{W}_{1a}+\frac{h}{m_{2c}}\tilde{W}_{1a}^{\mathrm{T}}\nabla\varphi_{1c}\nabla\varphi_h^{\mathrm{T}}\hat{W}_h\beta\Delta\varphi_{2c}^{\mathrm{T}}\hat{W}_{2c}\\&-\frac{h}{4m_{1c}}\tilde{W}_{1a}^{\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1a}\Delta\varphi_{1c}^{\mathrm{T}}\hat{W}_{1c}-\frac{1}{2}\tilde{W}_{1a}^{\mathrm{T}}D_{1c}e_v+k_{1a}\tilde{W}_{1a}^{\mathrm{T}}W_{1a}^{*}.\end{aligned}\tag{39}$$

V_3 对时间的导数为 $\dot{V}_3=-\tilde{W}_{2c}^{\mathrm{T}}\alpha_{2c}^{-1}\dot{\hat{W}}_{2c}-\tilde{W}_{2a}^{\mathrm{T}}\alpha_{2a}^{-1}\dot{\hat{W}}_{2a}$, 先计算 $\dot{V}_3$ 第一项 $\dot{V}_{31}=-\tilde{W}_{2c}^{\mathrm{T}}\alpha_{2c}^{-1}\dot{\hat{W}}_{2c}$, 根据式 (33) 和 $\xi_u^{hj}=e_v^{\mathrm{T}}Qe_v+u^{*\mathrm{T}}Ru^*+\beta^{\mathrm{T}}\left(2Qe_v-W_h^{*\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}W_{1c}^{*}\right)+W_{2c}^{*\mathrm{T}}\nabla\varphi_{2c}[h(e_v)+g(\omega^*+u^*)]$, 可得

$$\begin{aligned}\dot{V}_{31}=&\tilde{W}_{2c}^{\mathrm{T}}\frac{\Delta\varphi_{2c}}{m_{2c}}\left[\int_{t-h}^{t}(e_v^{\mathrm{T}}Qe_v+\hat{u}^{\mathrm{T}}R\hat{u}+\beta^{\mathrm{T}}(2Qe_v-\hat{W}_h^{\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1\alpha}))ds+\hat{W}_{2c}^{\mathrm{T}}\Delta\varphi_{2c}\right]\\=&k_{2c}\tilde{W}_{2c}^{\mathrm{T}}\hat{W}_{2c}+\frac{\overline{h}_{2c}}{m_{2c}}\tilde{W}_{2c}^{\mathrm{T}}\Delta\varphi_{2c}\\&+\tilde{W}_{2c}^{\mathrm{T}}\frac{\Delta\varphi_{2c}}{m_{2c}}\int_{t-h}^{t}(\hat{u}^{\mathrm{T}}R\hat{u}-u^{*\mathrm{T}}Ru^*+\beta^{\mathrm{T}}(W_h^{*\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}W_{1c}^{*}-\hat{W}_h^{\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}\hat{W}_{1a}))ds\\=&-k_{2c}\tilde{W}_{2c}^{\mathrm{T}}\tilde{W}_{2c}+\left(k_{2c}W_{2c}^{*\mathrm{T}}+\frac{\overline{h}_{2c}}{m_{2c}}\Delta\varphi_{2c}^{\mathrm{T}}\right)\tilde{W}_{2c}-\frac{1}{4m_{2c}}\tilde{W}_{2c}^{\mathrm{T}}\Delta\varphi_{2c}\tilde{W}_{2d}^{\mathrm{T}}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}W_{2c}^{*}\\&+\frac{1}{4m_{2c}}\tilde{W}_{2a}^{\mathrm{T}}\Delta\varphi_{2c}\hat{W}_{2a}^{\mathrm{T}}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}\hat{W}_{2c}+\frac{1}{4m_{2c}}\tilde{W}_{2a}^{\mathrm{T}}\Delta\varphi_{2c}\tilde{W}_{2a}^{\mathrm{T}}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}W_{2c}^{*}\\&-\frac{1}{4m_{2c}}\tilde{W}_{2a}^{\mathrm{T}}\Delta\varphi_{2c}W_{2c}^{*\mathrm{T}}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}W_{2c}^{*}+\frac{h}{m_{2c}}\tilde{W}_{2c}^{\mathrm{T}}\Delta\varphi_{2c}\beta^{\mathrm{T}}\tilde{W}_h^{\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}W_{1c}^{*}\\&-\frac{h}{m_{2c}}W_{2c}^{*\mathrm{T}}\Delta\varphi_{2c}\beta^{\mathrm{T}}\tilde{W}_h^{\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}\tilde{W}_{1a}+\frac{h}{m_{2c}}W_{2c}^{*\mathrm{T}}\Delta\varphi_{2c}\beta^{\mathrm{T}}W_h^{*\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}\tilde{W}_{1a}\\&-\frac{h}{m_{2c}}\hat{W}_{2c}^{\mathrm{T}}\Delta\varphi_{2c}\beta^{\mathrm{T}}\hat{W}_h^{\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}\tilde{W}_{1a},\end{aligned}\tag{40}$$

其中 $m_{2c}=(1+\Delta\varphi_{2c}^{\mathrm{T}}\Delta\varphi_{2c})^2$, $\overline{h}_{2c}=\int_{t-h}^{t}\xi_u^{hj}-W_{2c}^{*\mathrm{T}}\nabla\varphi_{2c}[h(e_v)+g(\omega^*+u^*)]ds$, $k_{2c}=(\Delta\varphi_{2c}\Delta\varphi_{2c}^{\mathrm{T}})/m_{2c}$, 根据式 (35), 可得 $\dot{V}_3$ 的第二项 $\dot{V}_{32}=-\tilde{W}_{2a}^{\mathrm{T}}\alpha_{2a}^{-1}\dot{\hat{W}}_{2a}$ 为

$$\dot{V}_{32}=-k_{2a}\tilde{W}_{2a}^{\mathrm{T}}\tilde{W}_{2a}-\frac{h}{4m_{2c}}\tilde{W}_{2a}^{\mathrm{T}}D_{2c}\Delta\varphi_{2c}^{\mathrm{T}}\hat{W}_{2a}\Delta\varphi_{2c}^{\mathrm{T}}\hat{W}_{2c}-\frac{1}{2}\tilde{W}_{2a}^{\mathrm{T}}D_{2c}e_v+k_{2a}\tilde{W}_{2a}^{\mathrm{T}}W_{2a}^{*}.\tag{41}$$

根据 Lyapunov 函数各项的推导, 可得

$$\begin{aligned}\dot{V}(t)\leqslant&-e_\eta^{\mathrm{T}}K_1e_\eta-e_v^{\mathrm{T}}K_2e_v-k_h\mathrm{tr}(\tilde{W}_h^{\mathrm{T}}\tilde{W}_h)-k_{1c}\tilde{W}_{1c}^{\mathrm{T}}\tilde{W}_{1c}-k_{1a}\tilde{W}_{1a}^{\mathrm{T}}\tilde{W}_{1a}\\&-k_{2c}\tilde{W}_{2c}^{\mathrm{T}}\tilde{W}_{2c}-k_{2\alpha}\tilde{W}_{2\alpha}^{\mathrm{T}}\tilde{W}_{2\alpha}+b_z\parallel e_v\parallel+k_hw_h\parallel\tilde{W}_h\parallel_F+M_1\tilde{W}_{1c}\\&+M_2\tilde{W}_{1a}+M_3\tilde{W}_{2c}+M_4\tilde{W}_{2a}+\tilde{W}_{1c}^{\mathrm{T}}M_5\tilde{W}_{1a}+\tilde{W}_{2c}^{\mathrm{T}}M_6\tilde{W}_{2a}^{\mathrm{T}}\\&+\tilde{W}_{1a}^{\mathrm{T}}M_7\tilde{W}_{1a}+\tilde{W}_{2a}^{\mathrm{T}}M_8\tilde{W}_{2a}+\tilde{W}_{2c}^{\mathrm{T}}M_9\tilde{W}_h+\tilde{W}_{1a}^{\mathrm{T}}M_{10}\tilde{W}_h,\end{aligned}\tag{42}$$

其中 $M_1=k_{1c}W_{1c}^{*\mathrm{T}}+\frac{\overline{h}_{1c}}{m_{1c}}\Delta\varphi_{1c}^{\mathrm{T}}$, $M_2=k_{1a}W_{1a}^{*\mathrm{T}}-\frac{h}{4m_{1c}}W_{1c}^{*\mathrm{T}}\Delta\varphi_{1c}W_{1c}^{*\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}+\frac{h}{m_{2c}}W_{2c}^{*\mathrm{T}}\Delta\varphi_{2c}\beta^{\mathrm{T}}W_h^{*\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}$, $M_3=k_{2c}W_{2c}^{*\mathrm{T}}+\frac{\overline{h}_{2c}}{m_{2c}}\Delta\varphi_{2c}^{\mathrm{T}}$, $M_4=k_{2a}W_{2a}^{*\mathrm{T}}\tilde{W}_{2a}-\frac{h}{4m_{2c}}W_{2c}^{*\mathrm{T}}\Delta\varphi_{2c}W_{2c}^{*\mathrm{T}}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}$, $M_5=-\frac{3h}{4m_{1c}}\Delta\varphi_{1c}W_{1c}^{*\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}$, $M_6=-\frac{h}{4m_{2c}}\Delta\varphi_{2c}W_{2c}^{*\mathrm{T}}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}$, $M_7=\frac{h}{4m_{1c}}\Delta\varphi_{1c}W_{1c}^{*\mathrm{T}}D_{1c}\nabla\varphi_{1c}^{\mathrm{T}}$, $M_8=\frac{h}{4m_{2c}}\tilde{W}_{2a}^{\mathrm{T}}\Delta\varphi_{2c}W_{2c}^{*\mathrm{T}}D_{2c}\nabla\varphi_{2c}^{\mathrm{T}}$, $M_9=\frac{h}{m_{2c}}\cdot\Delta\varphi_{2c}\beta^{\mathrm{T}}W_{1c}^{*\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}$, $M_{10}=-\frac{h}{m_{2c}}\Delta\varphi_{2c}\beta^{\mathrm{T}}W_{2c}^{*\mathrm{T}}\nabla\varphi_h\nabla\varphi_{1c}^{\mathrm{T}}$, 且 $\parallel M_i\parallel\leqslant m_i$.

应用杨氏不等式 [4] 可以得到以下不等式:

$$\dot{V}(t)\leqslant-a_{v_0}V(t)+b_{v_0},\tag{43}$$

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

其中 $a_{v_0}=\min\{\rho_e,\rho_z,\rho_h,\rho_{1a},\rho_{1c},\rho_{2a},\rho_{2c}\}$, $b_{v_0}=b_z^2/[2\lambda_{\min}(K_z)]+k_hw_h^2/2+k_{1c}m_1^2+k_{1\alpha}m_2^2/2+k_{2c}m_3^2$. 其中

$$\begin{aligned}&\rho_e=2\lambda_{\min}(K_1),\rho_z=\lambda_{\min}(K_z),\rho_h=\frac{k_h-2(m_9^2/k_{2c}+m_{10}^2/k_{2a})}{\lambda_{\min}(\alpha_h^{-1})},\\&\rho_{1a}=\frac{k_{1a}-2(m_5^2/k_{1c}+m_7)}{\lambda_{\min}(\alpha_{1a}^{-1})},\rho_{1c}=\frac{k_{1c}}{2\lambda_{\min}(\alpha_{1c}^{-1})},\\&\rho_{2a}=\frac{k_{2a}-2(m_6^2/k_{2c}+m_8)}{\lambda_{\min}(\alpha_{1a}^{-1})},\rho_{2c}=\frac{k_{2c}}{\lambda_{\min}(\alpha_{2c}^{-1})}.\end{aligned}\tag{44}$$

进一步可得

$$V(t)\leqslant\left(V(0)-\frac{b_{v_0}}{a_{v_0}}\right)\mathrm{e}^{-a_{v_0}t}+\frac{b_{v_0}}{a_{v_0}}\leqslant V(0)-\frac{b_{v_0}}{a_{v_0}}.\tag{45}$$

根据式 (35) Lyapunov 函数的定义, 可得无人船闭环系统中所有信号有界, 包括系统的轨迹跟踪误差 e_η, 速度跟踪误差 e_v 和神经网络权重 $\tilde{W}_h$, $\tilde{W}_{1a}$, $\tilde{W}_{1c}$, $\tilde{W}_{2a}$ 和 $\tilde{W}_{2c}$.

5　仿真实验

将本文提出的敌对目标袭扰下基于 Stackelberg 博弈的轨迹安全跟踪控制方法应用到文献 [32] 给出的 CyberShip II 无人船模型上, 并给出仿真结果验证其有效性.

5.1　无人船动力学模型

CyberShip II 无人船的动力学模型可以表示为

$$\begin{aligned}\dot{\eta}&=R(\eta)v,\\M\dot{v}&=u(t)+\omega(t)+d_1(t)-C(v)v-D(v)v,\end{aligned}\tag{46}$$

其中 M 是质量惯性矩阵, 科里奥利力和离心力项矩阵 $C(v)=-C(v)^{\mathrm{T}}$ 是反对称矩阵, 矩阵 $D(v)$ 是非线性阻尼矩阵. 矩阵具体形式可表示如下:

$$M=\begin{bmatrix}m_{11}&0&0\\0&m_{22}&m_{23}\\0&m_{32}&m_{33}\end{bmatrix},\ C(v)=\begin{bmatrix}0&0&c_{13}(v)\\0&0&c_{23}(v)\\-c_{13}(v)&-c_{23}(v)&0\end{bmatrix},\ D(v)=\begin{bmatrix}d_{11}(v)&0&0\\0&d_{22}(v)&d_{23}(v)\\0&d_{32}(v)&d_{33}(v)\end{bmatrix}.\tag{47}$$

系统时变质量惯性矩阵和未知的非线性动力学设计如下: $m_{11}(t)=20+50\sin^2(0.1\pi t)$, $m_{22}(t)=20+50\sin^2(0.2\pi t)$, $m_{23}(t)=1+\cos^2(0.1\pi t)$, $m_{32}(t)=1+\cos^2(0.1\pi t)$, $m_{33}(t)=10+20\sin^2(0.5\pi t)$, $c_{13}(v)=-24.6612q-1.0948r$, $c_{23}(v)=25.8p$, $d_{11}(v)=0.7225+1.3274|p|+5.8664p^2$, $d_{22}(v)=0.8612+36.2823|q|+8.05|r|$, $d_{23}(v)=-0.1079+0.845|q|+3.45|r|$, $d_{32}(v)=-0.1052-5.0437|q|-0.13|r|$ 以及 $d_{33}(v)=1.9-0.08|q|+0.75|r|$. 外部环境扰动设置为 $d_1(t)=[10-10\sin(0.005\pi t),10+10\sin(0.005\pi t),10-6\cos(0.005\pi t)]^{\mathrm{T}}$. 考虑环境中存在的随机性扰动对无人船轨迹跟踪精度的影响, 注入高斯白噪声下的环境扰动可表示为 $d_1'(t)=d_1(t)+\sigma W(t)$, $\sigma=200$ 为噪声强度, $W(t)$ 为高斯白噪声, 满足标准正态分布.

5.2　仿真条件

给定无人船的参考轨迹向量

$$\eta_r=\begin{bmatrix}3\sin(0.025\pi t)\\2\sin(0.05\pi t)\\\pi\cos(0.02\pi t)\end{bmatrix},\tag{48}$$

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

表 1　不同方法下的控制性能.

Table 1　Control performance of different methods.

Method	Convergence time (s)	IAE	
		$\|\|e_\eta\|\|$	$\|\|e_v\|\|$
The proposed scheme	4.76	2.96	6.02
The proposed scheme without VFD	8.25	3.70	7.41

初始位姿向量 $\eta(0)=[-1,0,0]^{\mathrm{T}}$ 和初始速度向量 $v(0)=[0.5,0,0]^{\mathrm{T}}$. 神经网络参数设计如下. (1) 领导者神经网络参数: 饱和输入约束 $\lambda=100$; $K_2=1.5I_3$, $R=0.05I_3$; critic 和 actor 神经网络学习率分别为 $\alpha_{2c}=0.8I_6$ 和 $\alpha_{2a}=0.4I_6$, $k_{2a}=1$. (2) 未知系统动力学神经网络参数: $k_h=0.5$, 学习率 $\alpha_h=0.8I_6$. (3) 跟随者神经网络参数: $G=0.001I_3$; critic 和 actor 神经网络学习率分别为 $\alpha_{1c}=0.7I_6$ 和 $\alpha_{1a}=0.5I_6$, $k_{1a}=1$. (4) 协态约束参数: $Q=20I_3$. 未知非线性系统动力学和领导者 – 跟随者 actor-critic 神经网络的初始权重限制在一定范围内并随机生成, 基函数为 $\varphi_{1c}=\varphi_h=\varphi_{2c}=[e_v(1)^2,e_v(2)^2,e_v(3)^2,e_v(1)e_v(2),e_v(1)e_v(3),e_v(2)e_v(3)]^{\mathrm{T}}$, 其关于速度跟踪误差 e_v 的梯度为

$$\nabla\varphi_{1c}=\nabla\varphi_h=\nabla\varphi_{2c}=\begin{bmatrix}2e_v(1) & 0 & 0\\ 0 & 2e_v(2) & 0\\ 0 & 0 & 2e_v(3)\\ e_v(2) & e_v(1) & 0\\ e_v(3) & 0 & e_v(1)\\ 0 & e_v(3) & e_v(2)\end{bmatrix}. \tag{49}$$

5.3　仿真结果

将本文设计领导者 – 跟随者控制信号应用到上述无人船模型上, 配置系统参数和仿真初值, 得到如下仿真结果.

图 3∼7 分别为: 仅考虑敌对目标策略性袭扰下, 我方无人船实际轨迹与参考轨迹, 无人船的轨迹跟踪误差 e_η 和速度跟踪误差 e_v, 无人船控制策略 u 和我方无人船受到敌对目标策略性排斥力 ω, 以及非线性动力学和领导者跟随者神经网络权重变化曲线. 由图可以看出, 本文设计的轨迹安全跟踪控制方法保证了无人船系统在受到敌对目标袭扰影响后: (1) 系统的所有闭环信号的有界性, 包括系统的轨迹跟踪误差, 速度跟踪误差和神经网络权重曲线; (2) 系统的渐近跟踪性能, 如图 4 所示, 系统的轨迹跟踪误差和速度跟踪误差均渐近收敛; (3) 执行器饱和输入约束, 当系统未跟踪参考轨迹或敌对目标或障碍物距离我方无人船较近时, 我方无人船会产生足够大的驱动力来实现避障和轨迹跟踪, 考虑实际执行器输出能力的限制, 具有饱和输入约束的无人船驱动力曲线如图 5 所示, 系统的饱和输入约束影响敌对目标对我方排斥力 ω 和无人船跟踪误差的收敛速度, 敌对目标对我方排斥力收敛意味着我方无人船逐渐适应敌对目标的袭扰影响, 当无人船的饱和输入约束 $\lambda/2$ 越大, 敌对目标对我方排斥力 ω 和无人船跟踪误差的收敛速度越快.

为进一步评估领导者决策优化过程中引入的值函数分解 (value function decomposition, VFD) 技术的有效性, 本文采用以下指标来区分两种方法: 神经网络的收敛时间, 位姿和速度积分绝对误差 (integral absolute error, IAE). 表 1 为相同实验条件下是否引入值函数分解技术的实验结果. 由表 1 可知, 引入值函数分解技术可有效加快神经网络收敛时间, 提高积分强化学习效率并减小无人船跟踪误差.

图 8 为考虑环境原始障碍物无人船的实际轨迹与参考轨迹图, 针对不具策略性运动的障碍物, 仅需在无人船的系统模型 (2) 中引入由人工势场法产生的排斥力 $d_2(t)$. 仿真实验场景设置两个静态障碍物和一个动态障碍物, 静态障碍物中心坐标为 $[-2.5,0]$ 和 $[2.5,0]$, 半径为 0.6, 动态障碍物起始位置中心坐标为 $[1.163,-1.154]$, 半径为 0.18, 其速度向量为 $[-0.226,0.266]$, 动态障碍物做匀速直线运动, 在仿真时间区间 $[35,37.9]$ 引入. 动态障碍

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

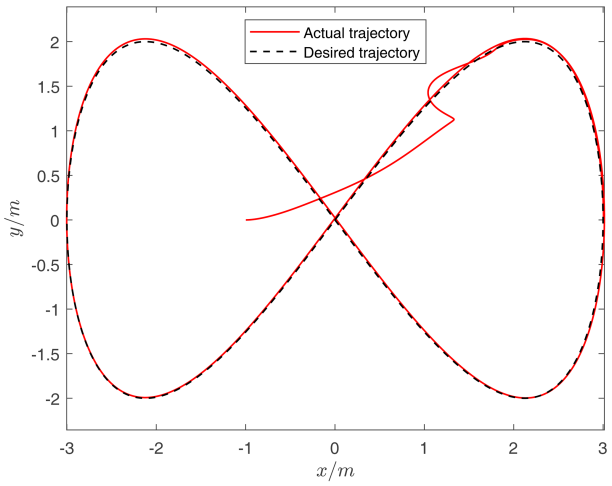


图 3　(网络版彩图) 敌对目标袭扰下无人船实际轨迹与参考轨迹.

Figure 3　(Color online) Actual and desired trajectories under hostile target interference.

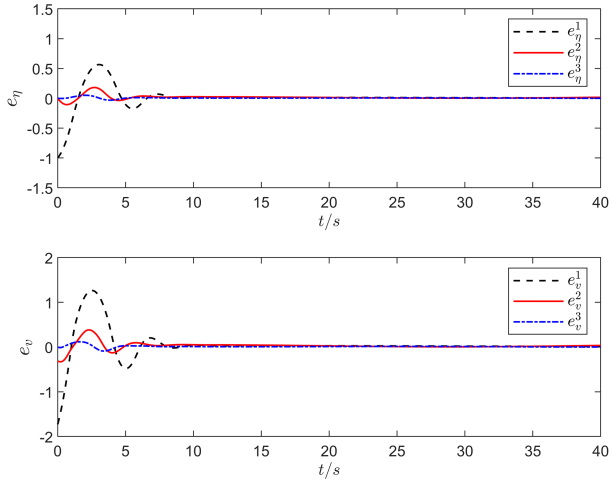


图 4　(网络版彩图) 无人船轨迹跟踪误差 e_η 和速度跟踪误差 e_v.

Figure 4　(Color online) Trajectory tracking error e_η and velocity tracking error e_v.

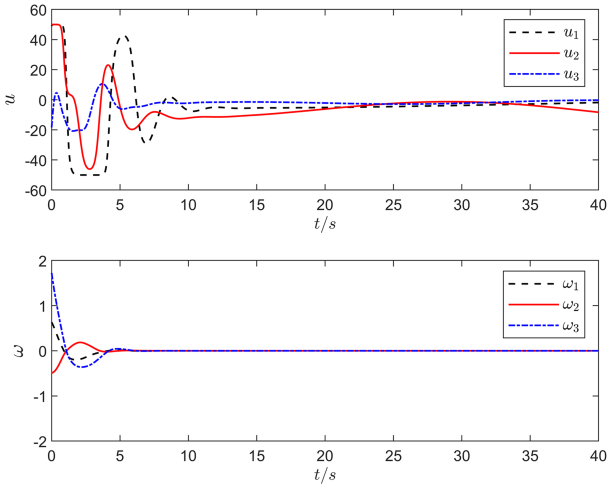


图 5　(网络版彩图) 无人船领导者控制策略 u 和跟随者控制策略 ω.

Figure 5　(Color online) Leader control strategy u and follower control strategy ω.

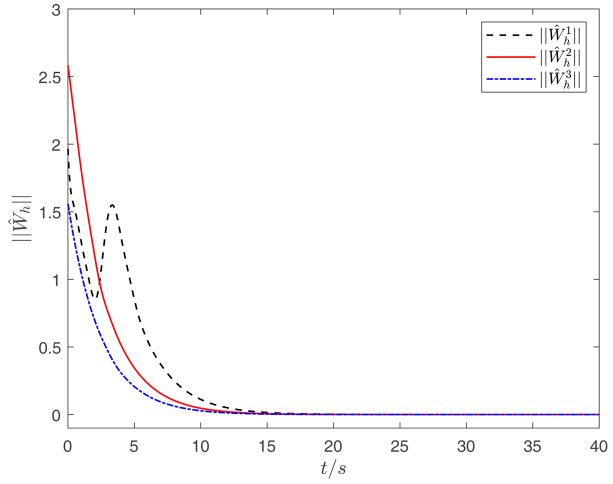


图 6　(网络版彩图) 无人船未知非线性动力学神经网络权重.

Figure 6　(Color online) Unknown nonlinear dynamics neural network weights.

物和静态障碍物对无人船的排斥力的计算方式相同, 排斥力 $d_2(t)$ 为无人船受到所有障碍排斥力之和, 每个障碍物对无人船的排斥力 $d_{2,i}(t)$ 可定义为

$$||d_{2,i}(t)|| = \begin{cases} k_{rep}\left(\dfrac{1}{d_i} - \dfrac{1}{d_{safe}}\right)\dfrac{1}{d_i^2}, & d_i < d_{safe}, \\ 0, & \text{else}, \end{cases} \tag{50}$$

其中 k_{rep} 为斥力系数, $k_{rep} > 0$, d_i 为无人船与障碍物之间的距离, d_{safe} 为安全距离. 本文设置人工势场力参数为 $k_{rep} = 0.0032$, $d_{safe} = 2$. 由图 9 可知, 所提出的无人船安全跟踪控制方法能够与人工势场法有效结合处理无人船对环境原始存在静态和动态障碍物避障和敌对目标策略性袭扰问题, 实现敌对目标袭扰下无人船的安全跟

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

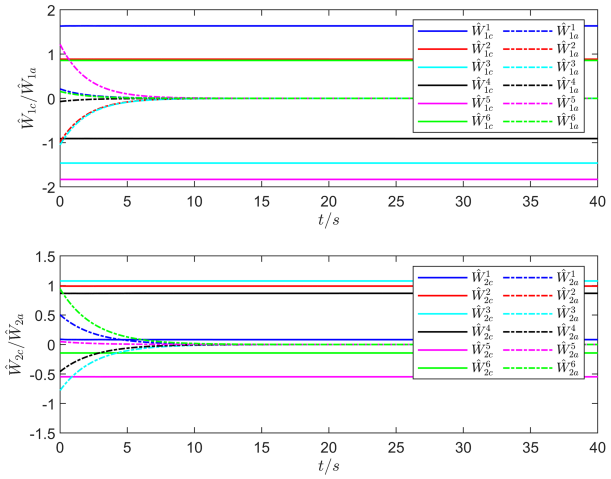


图 7　(网络版彩图) 领导者跟随者 actor-critic 神经网络权重.

Figure 7　(Color online) Leader-follower actor-critic neural network weights.

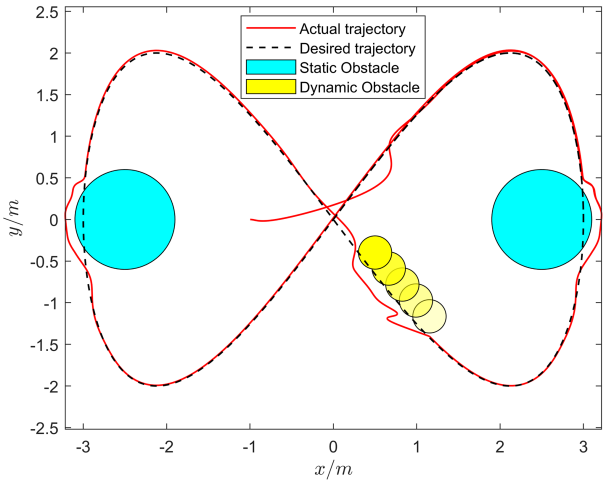


图 8　(网络版彩图) 考虑环境障碍物无人船实际轨迹与参考轨迹.

Figure 8　(Color online) Actual and desired trajectories of the USV considering environmental obstacles.

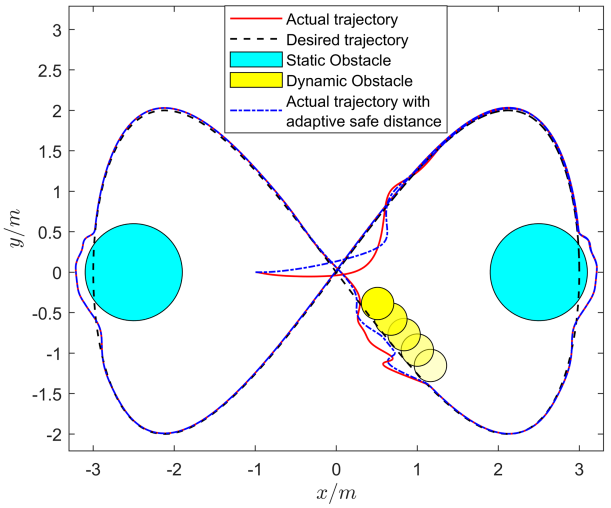


图 9　(网络版彩图) 自适应安全距离下考虑环境障碍物无人船实际轨迹与参考轨迹.

Figure 9　(Color online) Actual and desired trajectories of the USV considering environmental obstacles within an adaptive safe distance.

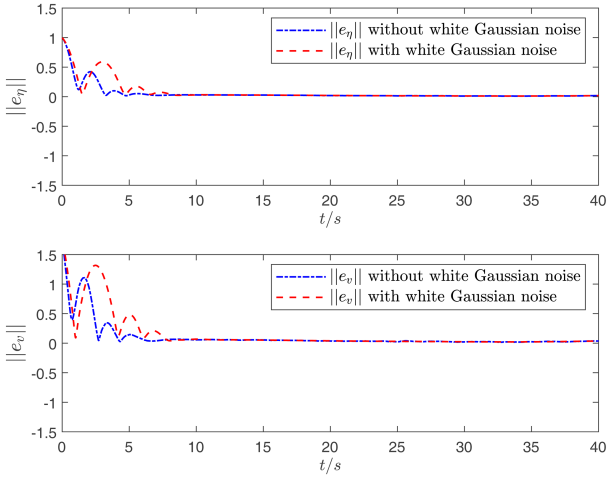


图 10　(网络版彩图) 高斯白噪声扰动下无人船轨迹跟踪误差 e_η 和速度跟踪误差 e_v.

Figure 10　(Color online) Trajectory tracking error e_η and velocity tracking error e_v with white Gaussian noises.

踪控制.

考虑到无人船快速或缓慢接近障碍物及动态障碍物快速或缓慢接近我方无人船时, 固定的安全距离 (势场排斥力影响范围) 可能会导致无人船出现紧急机动 (快速) 或过早避障 (缓慢) 的情形. 为此, 可设计自适应安全距离为

$$d_{safe}(v, v_o) = d_0 + \gamma||v - v_o||, \tag{51}$$

其中 d_0 为静态安全半径, 取决于船体尺寸, γ 为速度差敏感系数, $\gamma > 0$, v_o 为动态障碍物速度 (静态障碍物 $v_o = 0$). 图 9 为考虑自适应安全距离和未考虑情况下无人船实际轨迹和无人船参考轨迹图. 设置自适应安全距离参数为 $d_0 = 1$, $\gamma = 0.5$. 由图 9 可知, 自适应安全距离在一定程度上能减少无人船的轨迹跟踪误差, 有效提升

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

无人船的轨迹跟踪性能.

图 10 为考虑高斯白噪声扰动下无人船轨迹跟踪误差 e_η 和速度跟踪误差 e_v 与未考虑随机性噪声扰动下跟踪误差对比图. 由图 10 可知, 高斯白噪声的引入一定程度上增大了无人船跟踪误差和延长误差收敛时间, 但所提出的无人船安全跟踪控制方法仍能保证无人船的安全跟踪控制.

6　结论

本文提出了一种基于 Stackelberg 博弈的无人船轨迹安全跟踪控制方法, 以应对敌对目标策略性袭扰下的避障与轨迹跟踪问题. 所设计的控制信号可以保证含有未知非线性动力学的无人船在敌对目标策略性袭扰下的稳定性和渐近跟踪性能. 基于 Stackelberg 博弈框架, 将控制力与排斥力分别视为领导者和跟随者, 为其分别设计最优性能指标函数, 同时引入非二次型代价函数处理执行器饱和约束; 进而通过分层最优控制理论, 求解领导者 – 跟踪者的最优控制器结构. 针对系统部分动力学未知与博弈均衡难求解问题, 利用积分强化学习算法实现了未知非线性特性下 Stackelberg 博弈的近似最优解求解. 仿真结果表明, 所提方法能保证闭环系统信号有界, 使无人船在敌对目标袭扰下渐近跟踪参考轨迹, 同时避免与障碍物及敌对目标碰撞. 本文所提出的安全跟踪控制方法主要解决了敌对目标袭扰下的单个无人船的避障与轨迹跟踪问题, 而针对其在多船编队运行场景, 仍有待进一步研究. 未来工作将重点研究敌对目标袭扰下的多船协同避障与轨迹跟踪问题, 基于领导者 – 跟随着控制框架以及本文所提安全跟踪控制方法实现无人船编队的敌对目标袭扰下协同跟踪控制和局部最优避障.

参考文献

1 Zhao J, Wang Y, Cai Z, et al. Learning visual representation for autonomous drone navigation via a contrastive world model. IEEE Trans Artif Intell, 2024, 5: 1263–1276
2 Song S, Park J H, Zhang B Y, et al. Event-triggered adaptive practical fixed-time trajectory tracking control for unmanned surface vehicle. IEEE Trans Circ Syst II-Express Briefs, 2021, 68: 436–440
3 Wang N, Gao Y, Zhang X. Data-driven performance-prescribed reinforcement learning control of an unmanned surface vehicle. IEEE Trans Neural Netw Learn Syst, 2021, 32: 5456–5467
4 Zhao Y, Ma Y, Hu S. USV formation and path-following control via deep reinforcement learning with random braking. IEEE Trans Neural Netw Learn Syst, 2021, 32: 5468–5478
5 Gao S, Peng Z, Liu L, et al. Fixed-time resilient edge-triggered estimation and control of surface vehicles for cooperative target tracking under attacks. IEEE Trans Intell Veh, 2023, 8: 547–556
6 Xu Y, Zhang Y, Zhang J F. Singularity-free adaptive control of MIMO discrete-time nonlinear systems with general vector relative degrees. Automatica, 2023, 153: 111054
7 Xu Y, Zhang Y, Zhang J F. Parameter estimation and tracking control of MIMO linear systems without prior knowledge of control signs and parameter bounds. IEEE Trans Automat Contr, 2024, 70: 3695–3710
8 Li F, Zhang Y, Sun J. Matrix decomposition-based parametrization and singularity-free adaptive control of MIMO nonlinear systems. Automatica, 2024, 174: 112129
9 Li L, Wu D, Huang Y, et al. A path planning strategy unified with a COLREGS collision avoidance function based on deep reinforcement learning and artificial potential field. Appl Ocean Res, 2021, 113: 102759
10 Lyu H, Yin Y. COLREGS-constrained real-time path planning for autonomous ships using modified artificial potential fields. J Navigation, 2019, 72: 588–608
11 Wang N, Zhang Y, Ahn C K, et al. Autonomous pilot of unmanned surface vehicles: bridging path planning and tracking. IEEE Trans Veh Technol, 2022, 71: 2358–2374
12 Wang J, Wang R, Lu D, et al. USV dynamic accurate obstacle avoidance based on improved velocity obstacle method. Electronics, 2022, 11: 2720–2736
13 Wang C, Mao Y S, Du K J. Simulation on local obstacle avoidance algorithm for unmanned surface vehicle. Int J simul model, 2016, 15: 460–472
14 Ji J, Khajepour A, Melek W W, et al. Path planning and tracking for vehicle collision avoidance based on model predictive control with multiconstraints. IEEE Trans Veh Technol, 2017, 66: 952–964
15 Funke J, Brown M, Erlien S M, et al. Collision avoidance and stabilization for autonomous vehicles in emergency scenarios. IEEE Trans Contr Syst Technol, 2017, 25: 1204–1216

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

16 Rasekhipour Y, Khajepour A, Chen S K, et al. A potential field-based model predictive path-planning controller for autonomous road vehicles. IEEE Trans Intell Transp Syst, 2017, 18: 1255–1267
17 Huang Y, Ding H, Zhang Y, et al. A motion planning and tracking framework for autonomous vehicles based on artificial potential field elaborated resistance network approach. IEEE Trans Ind Electron, 2020, 67: 1376–1386
18 Wang W, Luo X, Li Y, et al. Unmanned surface vessel obstacle avoidance with prior knowledge-based reward shaping. Concurr Comput, 2020, 33: e6110
19 D'Amato E, Mattei M, Notaro I. Bi-level flight path planning of UAV formations with collision avoidance. J Intell Robot Syst, 2019, 93: 193–211
20 Xiang X, Yu C, Lapierre L, et al. Survey on fuzzy-logic-based guidance and control of marine surface vehicles and underwater vehicles. Int J Fuzzy Syst, 2018, 20: 572–586
21 Patle B K, Babu L G, Pandey A, et al. A review: on path planning strategies for navigation of mobile robot. Defence Tech, 2019, 15: 582–606
22 Arafat M Y, Alam M M, Moh S. Vision-based navigation techniques for unmanned aerial vehicles: review and challenges. Drones, 2023, 7: 89–120
23 Yang H, He Y, Xu Y, et al. Collision avoidance for autonomous vehicles based on MPC with adaptive APF. IEEE Trans Intell Veh, 2024, 9: 1559–1570
24 Cui Y, Tang J, Luo Q, et al. A Nash-Stackelberg game theoretic planner for many-to-few multi-vehicle racing. IEEE Trans Intell Veh, 2025, doi: 10.1109/TIV.2024.3478391
25 Wu D, Li Z, Du C, et al. Lane changing control of autonomous vehicle with integrated trajectory planning based on Stackelberg game. IEEE Open J Intell Transp Syst, 2024, 5: 810–825
26 Ming Z, Zhang H, Li Y, et al. Mixed H_2/H_∞ control for nonlinear closed-loop Stackelberg games with application to power systems. IEEE Trans Automat Sci Eng, 2024, 21: 69–77
27 Zhao H, Zhao N, Zong G, et al. Sliding-mode surface-based approximate optimal control for nonlinear multiplayer Stackelberg-Nash games via adaptive dynamic programming. Commun Nonlinear Sci Numer Simul, 2024, 132: 107928
28 Greiner D, Periaux J, Emperador J M, et al. Game theory based evolutionary algorithms: a review with nash applications in structural engineering optimization problems. Arch Computat Methods Eng, 2017, 24: 703–750
29 Trejo K K, Clempner J B, Poznyak A S. Adapting attackers and defenders patrolling strategies: a reinforcement learning approach for Stackelberg security games. J Comput Syst Sci, 2018, 95: 35–54
30 Bandyopadhyay S, Zhu C, Daniel P, et al. Targets in reinforcement learning to solve Stackelberg security games. 2022. ArXiv:2211.17132
31 Cui X, Wang B, Wang L. Mixed H_2/H_∞ optimal hierarchical tracking control for partially unknown dynamics. Trans Inst Measurement Control, 2025, 47: 700–712
32 Skjetne R, Fossen T I, Kokotović P V. Adaptive maneuvering, with experiments, for a model ship in a marine control laboratory. Automatica, 2005, 41: 289–298

https://www.sciengine.com/doi/10.1360/SSI-2025-0144

Stackelberg game-based trajectory tracking control and obstacle avoidance for USVs under hostile target interference

Liyan WEN[1], Yukang CUI[1], Bin JIANG[1*], Yizhen MENG[2] & Yajie MA[1]

1. *College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
2. *National Key Lab of Space Target Awareness, Shanghai Aerospace Control Technology Institute, Shanghai 201109, China*
* Corresponding author. E-mail: binjiang@nuaa.edu.cn

Abstract In this paper, a Stackelberg game-based safe tracking control method is proposed to address obstacle avoidance and trajectory tracking challenges for an unmanned surface vehicle (USV) under strategic pursuit and harassment by hostile targets. Inspired by the artificial potential field approach, strategic interference caused by dynamic repulsive forces from hostile targets is modeled as follower behavior within the game framework. A hierarchical control strategy is then designed based on the Stackelberg game. The performance index functions are formulated separately for the follower (repulsive force) and the leader (control force). A non-quadratic control-input cost function is introduced to enforce saturation constraints on control inputs. Using the maximum principle, control strategies for the leader and the follower are developed to achieve a dynamic balance between obstacle-avoidance safety and trajectory tracking accuracy. To address the nonlinearity of the unknown system dynamics and external disturbances in the USV, an integral reinforcement learning algorithm is employed to design a neural network that directly approximates the optimal solution to the Stackelberg game. The proposed safe trajectory tracking control strategy ensures the boundedness of all signals in the closed-loop system and enables the USV to track reference trajectories despite hostile target interference. Finally, simulation studies are conducted. The results demonstrate the effectiveness of the proposed safe tracking control strategy for the USV.

Keywords integral reinforcement learning, Stackelberg game, saturation constraints, obstacle avoidance, trajectory tracking

https://www.sciengine.com/doi/10.1360/SSI-2025-0144