Meng *et al*. *Intell. Robot.* **2025**, *5*(1), 88-104
DOI: 10.20517/ir.2025.06

Intelligence & Robotics

Research Article

Open Access

Stackelberg game-based anti-disturbance control for unmanned surface vessels via integrative reinforcement learning

Yizhen Meng[1], Chun Liu[2], Jing Zhao[3], Jing Huang[1], Guanbo Jing[2]

[1]Aerospace R&D Center, Shanghai Aerospace Control Technology Institute, and Shanghai Key Laboratory of Aerospace Intelligent Control Technology, Shanghai 201108, China.
[2]School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China.
[3]Shanghai Microsate Engineering Center, Shanghai 200240, China.

Correspondence to: Prof. Chun Liu, School of Mechatronic Engineering and Automation, Shanghai University, 99 Shangda Road, Baoshan District, Shanghai 200444, China. E-mail: Chun_Liu@shu.edu.cn; ORCID: 0000-0002-9294-6519

How to cite this article: Meng, Y.; Liu, C.; Zhao, J.; Huang, J.; Jing, G. Stackelberg game-based anti-disturbance control for unmanned surface vessels via integrative reinforcement learning. *Intell. Robot.* **2025**, *5*(1), 88-104. http://dx.doi.org/10.20517/ir.2025.06

Received: 1 Nov 2024 **First Decision:** 12 Dec 2024 **Revised:** 31 Dec 2024 **Accepted:** 10 Jan 2025 **Published:** 21 Jan 2025

Academic Editor: Simon Yang **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

In the navigation of unmanned surface vessels (USVs), external disturbances, particularly ocean waves, frequently induce deviations from the desired trajectory. To mitigate these challenges, we propose a novel disturbance rejection control strategy based on Stackelberg game theory, designed to address unmodeled system dynamics, complex environmental conditions, and other external perturbations. This approach incorporates several key innovations. First, we introduce a velocity error dynamic system coupled with a non-cooperative Stackelberg game model, where the USV's control inputs (as the leader) and external disturbances (as the follower) interact within an alternating update framework. This leader-follower interaction facilitates the joint optimization of both the disturbance rejection and performance-optimal control strategies, enhancing the USV's tracking accuracy while maximizing its disturbance rejection capacity. Second, we rigorously verify the existence of a cooperative optimal solution through an analysis of the Nash equilibrium under sequential decision-making between the leader and follower. Building on this, integral reinforcement learning and neural networks are employed to approximate the optimal Stackelberg solution. The boundedness and convergence of the proposed approach are validated using Lyapunov functions, ensuring stability and optimal performance under dynamic operating conditions. Finally, simulation results confirm the efficacy of the proposed strategy, demonstrating its ability to concurrently optimize control robustness and performance - such as minimizing tracking error and energy consumption - when confronted with unmodeled dynamics and external disturbances.

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Keywords: Unmanned surface vehicles, integral reinforcement learning, Stackelberg game, anti-disturbance control

1. INTRODUCTION

Unmanned surface vessels (USVs) offer substantial advantages for performing hazardous or repetitive tasks, owing to their autonomy and adaptability, which significantly reduce operational costs and mitigate associated risks[1–3]. In practical applications, robust disturbance-resistant control methods are crucial to ensuring both the safety and maneuverability of USVs, particularly in missions such as environmental monitoring[4] and maritime rescue[5]. The significance of these methods becomes even more pronounced in complex marine environments, where USVs are exposed to various unpredictable disturbances, including wave dynamics, unmodeled vessel behaviors, and potential cyber-physical attacks[4]. Ocean waves, in particular, continuously apply forces and torques that alter the vessel's motion, affecting its velocity and direction, thereby causing substantial deviations from the intended trajectory. Moreover, the intensity of these disturbances can sometimes overwhelm the control system's computational capacity, complicating the maintenance of stable and accurate trajectory tracking. The compounded effect of such unknown disturbances can severely undermine both the control stability and tracking precision of the USVs, leading to significant deviations from the planned route and, in some cases, putting mission success at serious risk.

Disturbance control for USVs is pivotal, not only ensuring the vessel's operational safety but also enhancing the success rate of mission execution. Early research has predominantly concentrated on disturbance estimation techniques under the assumption of bounded disturbances, such as disturbance observers, neural network observers, and adaptive disturbance observers. Leveraging these estimates, disturbance-rejection strategies, including sliding mode control and robust control, have been developed, creating an "estimation + robust control" paradigm. This framework improves the reliability and robustness of USVs, enabling them to sustain tracking performance despite inevitable disturbances[6–9]. In the context of path-tracking tasks, Xu *et al.* introduced a disturbance-resistant algorithm that combines adaptive neural network estimation with backstepping, incorporating an event-triggered mechanism to counteract input disturbances arising from unknown actuator faults[9]. Zhao *et al.* proposed a fault-tolerant tracking control strategy, integrating "estimation + adaptive sliding mode robust control" to mitigate the impacts of external disturbances and propeller faults on tracking accuracy[10]. Furthermore, Yu *et al.* designed a disturbance-resistant approach based on integral sliding mode control combined with disturbance estimation to address the challenges posed by comprehensive propeller failures and complex variations in control inputs[11]. However, while these disturbance-resistant techniques have demonstrated effectiveness, they remain fundamentally reliant on the precision of disturbance estimation. The accuracy of the disturbance estimate directly influences the performance of robust tracking control. Yet, in dynamic and complex environments, achieving high-precision estimation that can account for diverse and evolving disturbances remains a significant challenge. Moreover, the increasing complexity of USV tasks in such environments demands greater adaptability and interaction between the disturbance control strategies and the surrounding conditions. This highlights a critical limitation of the "estimation + robust control" paradigm, underscoring the need for more synchronized and adaptive coordination between the USV's control mechanisms and the dynamic environment in which it operates.

To address the aforementioned challenges, in the interaction capability between the disturbance-resistant control paradigm and the environment, as well as the disturbance-resistant strategy, Stackelberg game theory characterizes the interactive dynamics between leaders and followers in a game context[12], providing a novel perspective for enhancing the tracking performance of unmanned vessels while mitigating disturbances. In this framework, the leader first establishes its strategy, followed by the follower's selection of an optimal strategy based on the leader's choices[13]. This model forms a robust foundation for analyzing the leader-follower relationship within control systems. Consequently, a non-cooperative game is formulated within the Stackelberg framework, accounting for external disturbances, unmodeled dynamics, and control inputs, with the objective of deriving an optimal anti-disturbance control strategy that ensures the tracking performance of USVs. However, Stackelberg games introduce unique challenges that traditional optimal theory does not easily resolve. Prior research has applied a variational inequality algorithm to address synchronization issues in multi-agent systems within the Stackelberg game framework[14]. It is important to highlight that this approach depends on precise system model information, necessitating high-level accuracy in dynamic modeling. An algorithm for solving multiplayer Stackelberg–Nash games within nonlinear dynamic systems is proposed in[15]. This method allows the leader to optimize its strategy based on followers' responses, utilizing a two-tiered reinforcement learning framework that ensures convergence to equilibrium under weak coupling conditions. Consequently, within the interactive decision-making context of Stackelberg games, the leader must anticipate and integrate the follower's strategic responses into its decision-making process, while the follower adjusts based on the leader's choices. This dynamic interplay compels both the leader and follower to continuously refine their control strategies in response to an evolving environment. Utilizing the evaluation-action control mechanism in reinforcement learning, USVs autonomously evaluate the effectiveness of their control actions concerning environmental conditions, iteratively refining their strategies to maximize rewards. This adaptive approach enables USVs to identify optimal control strategies within dynamically shifting environments, presenting a novel pathway to achieving Nash equilibrium solutions in Stackelberg games through alternating and iterative optimization[2,5,16,17]. An online integral reinforcement learning algorithm is introduced to tackle Stackelberg games with unknown dynamics[18]; however, it is important to note that this method is limited to linear systems. In[19], an adaptive neural network tracking control method based on integral reinforcement learning is developed for continuous-time nonlinear systems with unknown control directions. Simulation results demonstrate the stability and boundedness of the closed-loop system while effectively managing an autonomous underwater vehicle model, thereby offering a promising strategy for addressing uncertainties in USV systems through reinforcement learning.

Inspired by the aforementioned research, this paper delves into a disturbance rejection control strategy for unmanned vessels based on Stackelberg game theory, aiming to overcome unmodeled dynamics, complex ocean waves, and other external disturbances. The strategy seeks to achieve a cooperative optimal solution with disturbance rejection robustness and optimal control performance (such as minimal energy consumption), leading to the following innovations:

1. In the anti-jamming method based on Stackelberg game theory, the interactive behavior of the USVs in the alternating update framework of non-cooperative games was elucidated. This interaction optimizes the cooperative search for both the anti-jamming strategy and the performance-optimal control strategy. The goal is to maximize the USV's anti-jamming capability while simultaneously optimizing its tracking control performance. To simplify the control strategy design, a virtual control variable is introduced to reduce the complexity of the USV's motion model, leading to the development of a velocity error dynamics model. Based on this, a control strategy that integrates robustness and optimality under the Stackelberg game framework is proposed.
2. Compared to existing Stackelberg game solutions that neglect dynamic factors, our proposed anti-jamming control strategy effectively addresses challenges posed by unknown drift dynamics and external bounded disturbances in the USVs. Furthermore, the Nash equilibrium of the anti-jamming strategy under sequential decision-making is analyzed, and the theoretical effectiveness of the proposed strategy is rigorously demonstrated.
3. Utilizing the Stackelberg Nash equilibrium and the "critic-action" control framework of reinforcement learning, an approximation method for the optimal anti-jamming strategy based on neural networks is presented. Additionally, the boundedness of the proposed method is proven using Lyapunov functions, and the convergence of the reinforcement learning algorithm under interactive performance metrics is elucidated.

This paper is structured as follows: Section 2 focuses on the formulation of the disturbance-resistant tracking problem for unmanned vessels. In Section 3, we present the derivation of the disturbance-resistant control law. Section 4 assesses the effectiveness and superiority of the proposed control scheme through numerical simulations. Finally, Section 5 provides the concluding remarks of this study.

2. PROBLEM FORMULATION

2.1. USVs dynamics

In this study, considering the unmodeled dynamics inherent to the unmanned vessel and the presence of external disturbances, its kinematics and dynamics are expressed as follows:

$$\begin{aligned}\dot{\aleph} &= \wp(\aleph)\nu,\\ \dot{\nu} &= F(\nu) + g\pi_\tau + \mathcal{D}(\aleph, \nu, t),\end{aligned} \tag{1}$$

where $\aleph = [x, y, \psi]^T$ represents the pose components. $\nu = \left[u, v, r\right]^T$ denotes the velocity vector, and $F(\nu)$ represents the known dynamic characteristics of the unmanned vessel. g signifies the control input gain. π_τ indicates the control input. $\mathcal{D}(\aleph, \nu, t)$ accounts for time-varying external disturbances and the unmodeled dynamics of the system. Furthermore, the rotation matrix $\wp(\eta)$ is given by:

$$\wp(\aleph) = \begin{bmatrix}\cos\psi & -\sin\psi & 0\\ \sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix}. \tag{2}$$

Furthermore, considering the external disturbances experienced by the unmanned vessel, along with system uncertainties and unmodeled dynamics, the control input of the unmanned vessel is given as follows:

$$\pi_\tau = \pi_{\tau c} + \pi_{\tau d}, \tag{3}$$

where $\pi_{\tau c}$ is the disturbance-resistant controller to be designed, and $\pi_{\tau d}$ represents the disturbance caused by unknown factors such as disturbances and the unmodeled dynamics of the unmanned vessel.

2.2. Control objectives

In this context, the attitude error vector $e = \aleph - \aleph_d$ is introduced to assess the interaction between the command inputs to the unmanned vessel and the disturbances, which subsequently influences its trajectory performance. To effectively utilize the inherent structural characteristics of the unmanned vessel system (1), an intermediate virtual control variable is designed to streamline the controller design process while ensuring the convergence of the attitude error, as given below:

$$\nu_d = \wp(\aleph)^{-1}\left(-\Gamma_1 e + \dot{\aleph}_d\right), \tag{4}$$

where Γ_1 is a positive definite control gain matrix. The difference between the actual velocity ν of the unmanned vessel and the desired control law ν_d plays a crucial role in determining the convergence of the pose error e. To this end, the velocity error vector $\Im_e = \nu - \nu_d$ is denoted as:

$$\dot{\Im}_e = f(\Im_e) + g\left(\pi_{\tau c} + \pi_{\tau d}\right), \tag{5}$$

where $f(\Im_e) = F(\nu) + \mathcal{D}(\aleph, \nu, t) - \dot{\nu}_d$ represents the unknown component in the dynamic system.

Control Objectives:

The control objective of this paper is to design an adaptive intelligent disturbance-resistant control scheme for unmanned vessels within the context of Stackelberg game theory. This scheme aims to achieve boundedness of the tracking error e and $\Im_e$, as well as all signals within the closed-loop system, even in the presence of unknown dynamics and external disturbances.

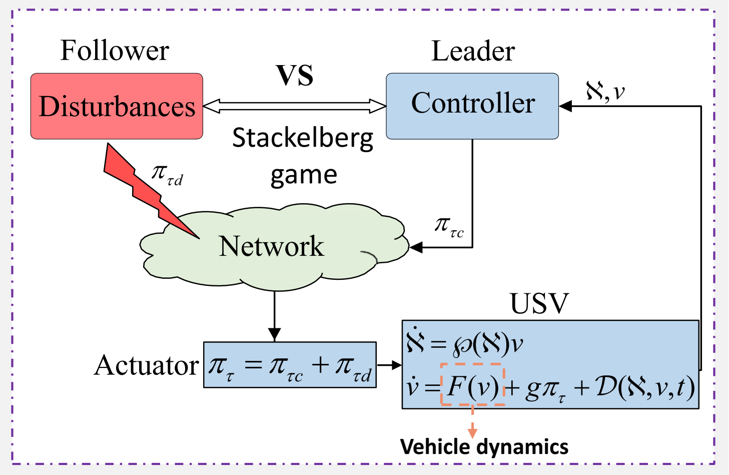

Figure 1. Stackelberg game-based framework for anti-disturbance control of unmanned aerial vehicles.

To achieve the aforementioned control objectives, the following definitions are introduced before proceeding with the design of the controller in this paper:

Definition 1: (Stackelberg Game) A unique Stackelberg equilibrium control pair $\{\pi^*_{\tau c}, \pi^*_{\tau d}(\pi^*_{\tau c})\}$ must satisfy the following properties:

1. For the control output $\pi_{\tau c}$, there exists a unique follower control output $\pi^*_{\tau d}(\pi_{\tau c})$ that minimizes the objective function J_1. Moreover, for any $\pi_{\tau c}$, it follows that $J_1(\pi_{\tau c}, \pi^*_{\tau d}(\pi_{\tau c})) \leq J_1(\pi_{\tau c}, \pi_{\tau d}(\pi_{\tau c}))$.
2. For the optimal response of the follower $\pi^*_{\tau d}(\pi_{\tau c})$, there exists an optimal leader response $\pi^*_{\tau c}$ such that $J_2(\pi^*_{\tau c}, \pi^*_{\tau d}(\pi^*_{\tau c})) \leq J_2(\pi_{\tau c}, \pi^*_{\tau d}(\pi_{\tau c}))$ is satisfied.

3. DESIGN OF THE DISTURBANCE-RESISTANT SATURATION CONTROL SCHEME BASED ON STACKELBERG GAME THEORY

3.1. Design framework for the disturbance-resistant controller based on Stackelberg game theory

In scenarios where the control inputs of unmanned vessels are subject to disturbances, it is imperative to identify the time-varying nature of these disturbances. To address this, we propose an intelligent optimal control strategy for disturbance rejection within the Stackelberg game framework, accounting for actuator input limitations. A non-cooperative game model is developed around the dynamic system of speed errors, wherein the disturbance signal $(\pi_{\tau d})$ is represented as the follower, and the disturbance rejection control strategy $(\pi_{\tau c})$ is represented as the leader in the Stackelberg game, as shown in Figure 1. This framework engenders a sequential decision-making process between the leader and follower, facilitating alternating iterative optimization, as outlined below:

1. The sequential decision-making begins with the initial output of the unmanned vessel's disturbance-resistant saturation controller $\pi^0_{\tau d}$.
2. Subsequently, the control output of the disturbances is regarded as the follower, interacting with the designed saturation controller to achieve the optimal control quantity aimed at maximizing the tracking error $e, \Im_e$ and the control input $\pi^*_{\tau d}$.
3. Conversely, the disturbance-resistant controller is regarded as the leader in this study, actively adjusting its strategy $\pi_{\tau c}$ based on the follower's control output $\pi^*_{\tau d}$, and selecting the optimal disturbance-resistant strategy $\pi^*_{\tau c}$ to minimize the tracking error $e, \Im_e$.

To establish a robust optimal disturbance rejection control strategy for unmanned vessels under alternating iterations between the leader and follower, the leader's decision-making process is contingent on the rational choices of the follower. This mutual dependence guarantees the simultaneous minimization of the cost functions for both parties. These cost functions are given as follows:

$$J_1\left(\Im_{e0}, \pi_{\tau d}, \pi_{\tau c}\right) = \int_t^{\infty} r_1\left(\Im_e, \tau_d, \tau_c\right) \mathrm{d}s = \int_0^{\infty} \left(\pi_{\tau d}^T G \pi_{\tau d} + \pi_{\tau c}^T R \pi_{\tau d} - {\Im_e}^T Q \Im_e\right) \mathrm{d}s, \tag{6}$$

$$J_2\left(\Im_{e0}, \pi_{\tau d}, \pi_{\tau c}\right) = \int_t^{\infty} r_2\left(\Im_e, \pi_{\tau d}, \pi_{\tau c}\right) \mathrm{d}s = \int_0^{\infty} \left({\Im_e}^T Q \Im_e + \pi_{\tau c}^T R \pi_{\tau c} + \Pi\left(\pi_{\tau d}\right)\right) \mathrm{d}s, \tag{7}$$

where G, Q, R represents positive definite matrices, and $\Pi(\pi_{\tau d}) = \vartheta^T \dot{\hat{\lambda}}_1$ denotes the control input influence associated with the follower.

It is noteworthy that the performance metric functions J_1 and J_2, as defined within the Stackelberg game framework, are characterized as follows:

1. J_1 encapsulates the robustness of the unmanned vessel system, particularly focusing on the degradation of tracking performance under maximum disturbance conditions.
2. J_2 governs the vessel's optimal tracking control performance, striving to minimize tracking error while minimizing control energy expenditure.
3. Through an alternating iterative optimization process of both J_1 (robustness) and J_2 (optimality), a Nash equilibrium is attained. In this equilibrium, the disturbance rejection robustness and optimal control (minimal energy consumption and tracking error) of the unmanned vessel cannot be further improved by independently altering either control strategy. Consequently, the system reaches a cooperative optimal state, thus fulfilling the design objectives of the proposed control strategy.

Furthermore, their corresponding value functions are expressed as follows:

$$\begin{aligned} V_1\left(\Im_e(t), \pi_{\tau d}, \pi_{\tau c}\right) &= \int_t^{\infty} r_1\left(\Im_e, \pi_{\tau d}, \pi_{\tau c}\right) \mathrm{d}s, \\ V_2\left(\Im_e(t), \pi_{\tau d}, \pi_{\tau c}\right) &= \int_t^{\infty} r_2\left(\Im_e, \pi_{\tau d}, \pi_{\tau c}\right) \mathrm{d}s. \end{aligned} \tag{8}$$

The optimal value function for the follower is defined as

$$V_1^*\left(\Im_{e0}\right) = \min_{\pi_{\tau d}} \int_0^{\infty} r_1\left(\Im_e, \pi_{\tau c}, \pi_{\tau d}\right) \mathrm{d}s, \tag{9}$$

In this case, considering the velocity error dynamics (5), the Hamiltonian function for the follower is defined as

$$H_1\left(\Im_e, \nabla V_1, \pi_{\tau d}, \pi_{\tau c}\right) = r_1\left(\Im_e, \pi_{\tau d}, \pi_{\tau c}\right) + \nabla V_1^T\left(f(z) + g\pi_{\tau c} + g\pi_{\tau d}\right), \tag{10}$$

where $\nabla V_1 = \partial V_1 / \partial \Im_e$ denotes the partial derivative of the Hamiltonian with respect to the variable $\Im_e$. Furthermore, based on $\partial H_1 / \partial \pi_{\tau d} = 0$, the optimal control output $\pi_{\tau d}^*$ for the follower is given by:

$$\pi_{\tau d}^* = -\frac{1}{2} G^{-1} g^T \nabla V_1^T - \frac{1}{2} G^{-1} R \pi_{\tau c}, \tag{11}$$

Furthermore, introducing the definition $\dot{\hat{\Lambda}}_1 = \nabla \dot{V}_1 = -\partial H_1 / \partial \Im_e$, the following co-state equations are obtained:

$$\dot{\hat{\Lambda}}_1 = 2Q\Im_e - \nabla f^T \nabla V_1, \tag{12}$$

where $\nabla f = \partial f\left(\Im_e\right) / \partial \Im_e$. Thus, the follower adopts the optimal strategy (11) as its predetermined decision.

Subsequently, the leader's disturbance-resistant controller formulates its strategy by considering the follower's strategy (11) and the co-state constraints (12). Therefore, the constrained optimal control problem $\Pi(\tau_d) = \vartheta^T \dot{\hat{\lambda}}_1$ for the leader is expressed as:

$$V_2^* \left(\mathfrak{I}_e\right) = \min_{\pi_{\tau c}} \int_0^{\infty} r_2 \left(\mathfrak{I}_e, \pi_{\tau d}, \pi_{\tau c}\right) \mathrm{d}s, \tag{13}$$

where ϑ is the designed Lagrange multiplier.

Thus, the Hamiltonian function for the leader can be expressed as follows:

$$H_2 \left(\mathfrak{I}_e, \nabla V_2, \pi_{\tau d}^*, \pi_{\tau c}\right) = r_2 \left(\mathfrak{I}_e, \pi_{\tau d}^*, \pi_{\tau c}\right) + \nabla V_2^T \left(f(\mathfrak{I}_e) + g\pi_{\tau c} + g\pi_{\tau d}^*\right), \tag{14}$$

where $\nabla V_2 = \partial V_2 / \partial \mathfrak{I}_e$. By utilizing the necessary conditions for optimality, the leader's optimal disturbance-resistant saturation control strategy and co-state equations can be derived, with the specific steps outlined as follows:

$$\pi_{\tau c}^* = -\frac{1}{2} R^{-1} g^T \nabla V_2^T \tag{15}$$

$$\dot{\vartheta} = -\frac{\partial H_2}{\partial \hat{\lambda}_1}^T = \frac{1}{2} g G^{-1} g \nabla V_1 + \nabla f \vartheta \tag{16}$$

To obtain the minimum cost function V_1^*, V_2^* under the optimal control input $\pi_{\tau d}^*, \pi_{\tau c}^*$, the corresponding Hamilton-Jacobi (HJ) equation can be derived after introducing the definition $\nabla V_1^* = \partial V_1^* / \partial \mathfrak{I}_e$, $\nabla V_2^* = \partial V_2^* / \partial \mathfrak{I}_e$, as follows:

$$\begin{aligned} 0 &= r_1 \left(\mathfrak{I}_e, \pi_{\tau d}^*, \pi_{\tau c}^*\right) + \nabla V_1^{*T} \left(f(\mathfrak{I}_e) + g\pi_{\tau c}^* + \pi_{\tau d}^*\right), \\ 0 &= r_2 \left(\mathfrak{I}_e, \pi_{\tau d}^*, \pi_{\tau c}^*\right) + \nabla V_2^{*T} \left(f(\mathfrak{I}_e) + g\pi_{\tau c}^* + \pi_{\tau d}^*\right). \end{aligned} \tag{17}$$

In a further step, by substituting $\pi_{\tau d}^*$ (11) and $\pi_{\tau c}^*$ (15) into (17), it can be inferred that:

$$\begin{aligned} 0 = {} & \frac{1}{4} \nabla V_1^* g G^{-1} g^T \nabla V_1^{*T} - \mathfrak{I}_e{}^T Q \mathfrak{I}_e - \frac{1}{4} \nabla V_2^* g R^{-1} g^T \nabla V_2^{*T} \\ & + \nabla V_1^{*T} f(\mathfrak{I}_e) - \frac{1}{2} g R^{-1} g^T \nabla V_2^{*T} - \frac{1}{2} g G^{-1} g^T \nabla V_1^{*T}, \\ 0 = {} & \mathfrak{I}_e{}^T Q \mathfrak{I}_e + \frac{1}{4} \nabla V_2^* g R^{-1} g^T \nabla V_2^{*T} + 2\vartheta^T Q \mathfrak{I}_e - \vartheta^T \nabla f \nabla V_1 \\ & + \nabla V_1^{*T} f(\mathfrak{I}_e) - \frac{1}{2} g R^{-1} g^T \nabla V_2^{*T} - \frac{1}{2} g G^{-1} g^T \nabla V_1^{*T}. \end{aligned} \tag{18}$$

Based on the above content, Theorem 1 is formulated as follows.

Theorem 1: Under the optimal disturbance influence strategy, the velocity error dynamics constrained by the cost function (6) can ensure stability by utilizing the optimal disturbance-resistant strategy designed in (15). Furthermore, considering the disturbance-resistant problem of the velocity error dynamics (5) with the cost function given by (7), the control for $\left\{\pi_{\tau c}^*, \pi_{\tau d}^*(\pi_{\tau c}^*)\right\}$ achieves Stackelberg equilibrium if and only if the coupled HJ equations in (17) have a solution.

Proof: To prove the stability of the tracking error dynamics, V_2^* is selected as a candidate Lyapunov function. Its derivative with respect to (18) is calculated as follows:

$$\dot{V}_2^* = \left(\frac{\partial V_2^*}{\partial \delta}\right) \dot{\mathfrak{I}}_{ei} + \left(\frac{\partial V_2^*}{\partial \nabla V_1^*}\right) \dot{V}_1^* = -\mathfrak{I}_e{}^T Q \mathfrak{I}_e - \left(\pi_{\tau c}^*\right)^T R \pi_{\tau c}^*, \tag{19}$$

Subsequently, based on $Q_i > 0$ and $R_i > 0$, the conclusion can be made through $\dot{V}_2^* < 0$, indicating that under the optimal strategy, the tracking error system (5) can achieve asymptotic stability.

To demonstrate the Stackelberg equilibrium, it is noteworthy that the performance index (6) is reformulated as follows:

$$\begin{aligned} & J_1\left(\mathfrak{I}_{e0}, \pi_{\tau d}, \pi_{\tau c}\right) \\ = & \int_0^{\infty}\left(-\mathfrak{I}_e{}^T Q \mathfrak{I}_e+\pi_{\tau d}^T G \pi_{\tau d}+\pi_{\tau c}^T R \pi_{\tau d}\right) d s+\int_0^{\infty} \dot{V}_d^* d s+V_d^*\left(\mathfrak{I}_e(0)\right)-V_d^*\left(\mathfrak{I}_e(\infty)\right), \end{aligned} \tag{20}$$

By combining expressions $\dot{V}_1^* = \left(\nabla V_1^*\right)^T \left[f\left(\mathfrak{I}_e\right) + g\pi_{\tau c} + g\pi_{\tau d}\right]$ and $-2\left(\pi_{\tau d}^*\right)^T G\pi_{\tau d} = \left(\nabla V_d^*\right)^T g\pi_{\tau d} + \pi_{\tau c} R \pi_{\tau d}^T$, and employing the complete square method, it yields:

$$\begin{aligned} & J_1\left(\mathfrak{I}_{e0}, \pi_{\tau d}, \pi_{\tau c}\right) \\ = & \int_t^{\infty}\left(\pi_{\tau d}-\pi_{\tau d}^*\right)^T G\left(\pi_{\tau d}-\pi_{\tau d}^*\right) d s+\int_t^{\infty}\left(-\mathfrak{I}_e{}^T Q \mathfrak{I}_e-\left(\pi_{\tau d}^*\right)^T G \pi_{\tau d}^*+\left(\nabla V_d^*\right)^T\right. \\ & \left.\times\left[f\left(\mathfrak{I}_e\right)+g \pi_{\tau c}\right] d s+V_d^*\left(\mathfrak{I}_e(0)\right)-V_d^*\left(\mathfrak{I}_e(\infty)\right)\right), \end{aligned} \tag{21}$$

Subsequently, starting from $-\left(\pi_{\tau d}^*\right)^T G\pi_{\tau d}^* = \left(\pi_{\tau d}^*\right)^T G\pi_{\tau d}^* + \left(\nabla V_d^*\right)^T g\pi_{\tau d}^* + \pi_{\tau c} R\pi_{\tau d}^*$, and utilizing the coupled HJ equations for the follower as given in (18), this expression can be further simplified to:

$$J_1\left(\mathfrak{I}_{e0}, \pi_{\tau d}\right)=\int_t^{\infty}\left(\pi_{\tau d}-\pi_{\tau d}^*\right)^T G\left(\pi_{\tau d}-\pi_{\tau d}^*\right) d s+V_1^*\left(\mathfrak{I}_{e0}\right)-V_1^*\left(\mathfrak{I}_e(\infty)\right), \tag{22}$$

By setting $\pi_{\tau d} = \pi_{\tau d}^*$, the optimal value of the follower's cost function is $J_1\left(\mathfrak{I}_{e0}, \pi_{\tau d}^*\right) = V_1^*\left(\mathfrak{I}_{e0}\right) - V_1^*\left(\mathfrak{I}_e(\infty)\right)$. According to condition $\dot{V}_1^* = \mathfrak{I}_e{}^T Q\mathfrak{I}_e - \left(\pi_{\tau d}^*\right)^T G\pi_{\tau d}^* - \pi_{\tau c}^T R\pi_{\tau d}^*$, it is able to conclude that when inequality $\mathfrak{I}_e{}^T Q\mathfrak{I}_e \leq -\pi_{\tau c}^T R\pi_{\tau d}^*$ holds, one has:

$$J_1(\pi_{\tau c}, \pi_{\tau d}^*(\pi_{\tau c})) \leq J_1(\pi_{\tau c}, \pi_{\tau d}(\pi_{\tau c})), \tag{23}$$

Consequently, by applying a similar derivation process as outlined for the leader's cost function, we derive:

$$J_2(\pi_{\tau c}^*, \pi_{\tau d}^*(\pi_{\tau c}^*)) \leq J_2(\pi_{\tau c}, \pi_{\tau d}^*(\pi_{\tau c})), \tag{24}$$

Therefore, combined with Figure 1, assuming that the control strategy of the follower $\pi_{\tau d}$ and the leader's disturbance-rejection control strategy $\pi_{\tau c}$ in the Stackelberg game satisfy the aforementioned conditions, such that neither the leader nor the follower can reduce their respective cost function values (J_1 or J_2) by unilaterally adjusting their strategy, the proof of the Stackelberg game Nash equilibrium, as defined in Definition 1, is thus concluded.

3.2. Stackelberg game resolution via integral reinforcement learning techniques

As is well known, deriving an analytical solution to the coupled HJ Equation (17) is highly challenging due to the current nonlinear characteristics. To address this issue, this subsection employs an action-evaluation neural network algorithm based on integral reinforcement learning to solve the Stackelberg game in the disturbance-resistant context of the USVs system. This algorithm includes an auxiliary neural network capable of effectively handling the complexities arising from unknown dynamics. The discussion regarding the solution of the Stackelberg game will involve the follower, the unknown dynamics, and the leader.

To continue, an optimized evaluation neural network is developed to approximate the follower's optimal value function V_1^*, enabling the representations of V_1^* and its gradient ∇V_1^* as follows:

$$V_1^* = W_{1c}^{*T}\Phi_{1c} + \zeta_{1c}, \nabla V_1^* = \nabla\Phi_{1c}^T W_{1c}^* + \nabla\zeta_{1c}, \tag{25}$$

where W_{1c}^* represents the optimal weights of the follower's evaluation network, $\Phi_{1c} = \Phi_{1c}(\mathfrak{I}_e)$ denotes the activation function of the evaluation network, and $\zeta_{1c} = \zeta_{1c}(\mathfrak{I}_e)$ represents the estimation error. Furthermore, let $\nabla V_1^* = \partial V_1^*/\partial\mathfrak{I}_e, \nabla\Phi_{1c} = \partial\Phi_{1c}/\partial\mathfrak{I}_e, \nabla\zeta_{1c} = \partial\zeta_{1c}/\partial\mathfrak{I}_e$.

Considering the unknown nature of the ideal weights W_{1c}^{*}, a neural network approximation to approach $V_1^{*}, \nabla V_1^{*}$ is applied, resulting in:

$$\hat{V}_1 = \hat{W}_{1c}^{T}\Phi_{1c}, \nabla\hat{V}_1 = \nabla\Phi_{1c}^{T}\hat{W}_{1c}, \tag{26}$$

where $\nabla\hat{V}_1 = \partial\hat{V}_1/\partial z$.

$$\hat{\pi}_{\tau d} = -\frac{1}{2}G^{-1}g^{T}\nabla\Phi_{\mathrm{lc}}^{T}\hat{W}_{\mathrm{la}}, \tag{27}$$

where $\hat{W}_{\mathrm{la}}$ is the estimate of the optimal weights W_1^{*} of the optimal evaluation network.

To avoid utilizing the unknown dynamics $f(\mathfrak{I}_e)$ throughout the learning process, the Bellman error equation with arbitrary time integral λ is introduced as follows:

$$e_{1c}^{J} = \hat{W}_{1c}^{T}\Delta\Phi_{1c} + \int_{t-\lambda}^{t} r_1\left(\mathfrak{I}_e, \hat{\pi}_{\tau d}, \hat{\pi}_{\tau c}\right)ds, \tag{28}$$

where $r_1\left(\mathfrak{I}_e, \hat{\pi}_{\tau d}, \hat{\pi}_{\tau c}\right) = \hat{\pi}_{\tau d}^{T}R\hat{\pi}_{\tau d}^{T} - {\mathfrak{I}_e}^{T}Q\mathfrak{I}_e - \hat{\pi}_{\tau c}^{T}G\hat{\pi}_{\tau c}$, $\hat{\pi}_{\tau c}$ is the control strategy of the leader to be designed. Additionally, $\Delta\Phi_{\mathrm{lc}} = \Phi_{\mathrm{lc}}(t) - \Phi_{\mathrm{lc}}(t-\lambda)$.

In addition, with the objective of minimizing the error $E_{1c} = \frac{1}{2}e_{1c}^{T}J^{T}e_{1c}$, the optimal weights W_{1c}^{*} of the evaluation network are adaptively adjusted, leading to:

$$\dot{\hat{W}}_{1c} = -k_{1c}\frac{\mathfrak{I}_e\Phi_{1c}}{\Phi_{1c}}\hat{W}_{1c}^{T}e_{1c}^{J}, \tag{29}$$

where $\Phi_{_{1c}} = \left(1 + \Delta\Phi_{1c}^{T}\Delta\Phi_{1c}\right)^2$. Additionally, $k_{_{1c}} > 0$ represents the learning rate for adjusting the follower's evaluation network.

By substituting the follower's control output (27) into equations (12) and (16), it can be deduced that:

$$\begin{aligned} \dot{\hat{\Lambda}}_1 &= 2Q\mathfrak{I}_e - f\left(\mathfrak{I}_e\right)\nabla\Phi_{1c}^{T}\hat{W}_{1a}, \\ \dot{\beta} &= -\frac{\partial H_2}{\partial\hat{\Lambda}_1}^{T} = \frac{1}{2}gG^{-1}g\nabla\Phi_{1c}^{T}\hat{W}_{1a} + f\left(\mathfrak{I}_e\right)\vartheta, \end{aligned} \tag{30}$$

Additionally, to fully leverage the state information for disturbance-resistant control and to enhance the multifunctionality of the unmanned vessel's behavior, the optimal value function V_2 is decomposed as follows:

$$V_2^{*} = \Gamma_{\mathfrak{I}_e}{\mathfrak{I}_e}^{T}\mathfrak{I}_e + 2\Gamma_s{\mathfrak{I}_e}^{T}\wp(\aleph)e + 2\Gamma_h{\mathfrak{I}_e}^{T}f(\mathfrak{I}_e) + E_{V_2}^{*}, \tag{31}$$

where $E_{V_2}^{*} = V_2^{*} - \Gamma_{\mathfrak{I}_e}{\mathfrak{I}_e}^{T}\mathfrak{I}_e - 2\Gamma_e{\mathfrak{I}_e}^{T}\wp(\aleph)e - 2\Gamma_h{\mathfrak{I}_e}^{T}f(z)$, $\Gamma_{\mathfrak{I}_e}, \Gamma_s, \Gamma_h$ is the positive definite controller gain. By taking the partial derivative with respect to $\mathfrak{I}_e$ and substituting (31) into (15), one obtains:

$$\pi_{\tau c}^{*} = -\frac{1}{2}R^{-1}g^{T}\left(\Gamma_{\mathfrak{I}_e}\mathfrak{I}_e + \Gamma_e\wp(\aleph)e + \Gamma_h f\left(\mathfrak{I}_e\right) + \nabla E_{V_2}^{*}\right), \tag{32}$$

where $\nabla E_{V_*}^{*} = \partial E_{V_*}^{*}/\partial\mathfrak{I}_e$.

Consequently, the evaluation network for designing the leader's (disturbance-resistant) control strategy is developed to approximate the value function $E_{V_2}^{*}$ and its gradient $\nabla E_{V_2}^{*}$ as follows:

$$E_{V_2}^{*} = W_{2c}^{*T}\Phi_{2c} + \zeta_{2c}, \nabla E_{V_2}^{*} = \nabla\Phi_{2c}^{T}W_{2c}^{*} + \nabla\zeta_{2c}, \tag{33}$$

Where W_{2c}^{*} represents the optimal weights of the leader's evaluation network, and Φ_{2c}, ζ_{2c} denote the activation function and estimation error of the leader's evaluation network, respectively.

Similarly, by estimating $E^*_{V_2}$ and $\nabla E^*_{V_2}$, one has:

$$\hat{E}_{V_2} = \hat{W}^T_{2c}\Phi_{2c}, \nabla\hat{E}_{V_2} = \nabla\Phi^T_{2c}\hat{W}_{2c}, \tag{34}$$

where $\nabla\hat{E}_{V_2} = \partial\hat{E}_{V_2}/\partial\mathfrak{I}_e$.

As a result of the preceding analysis, the leader's disturbance-resistant controller can be expressed as follows:

$$\hat{\pi}_{\tau c} = -\frac{1}{2}R^{-1}g^T\left(\Gamma_{\mathfrak{I}_e}\mathfrak{I}_e + \Gamma_s\wp(\aleph)e\ +\Gamma_h f\left(\mathfrak{I}_e\right) + \nabla\varphi^T_{2c}\hat{W}_{2a}\right), \tag{35}$$

where $\hat{W}_{2a}$ is the estimate of the optimal weights W^*_{2c} of the leader's evaluation network.

Moreover, akin to the formulation employed for the follower, the Bellman error corresponding to the leader can be expressed as:

$$e^J_{2c} = \int_{t-\lambda}^{t} r_2\left(\mathfrak{I}_e, \hat{\pi}_{\tau d}, \hat{\pi}_{\tau c}\right)ds + \hat{W}^T_{2c}\Delta\Phi_{2c}, \tag{36}$$

where $r_2\left(\mathfrak{I}_e, \hat{\pi}_{\tau d}, \hat{\pi}_{\tau c}\right) = \mathfrak{I}_e{}^T Q\mathfrak{I}_e + \hat{\pi}^T_{\tau d}G\hat{\pi}_{\tau d} + \vartheta^T\dot{\hat{\Lambda}}_1$ and $\Delta\Phi_{2c} = \Phi_{2c}\left(t\right) - \Phi_{2c}\left(t-\lambda\right)$.

Based on this, the adaptive update law for the evaluation network weights that minimizes the objective function $E_{2c} = \frac{1}{2}e^{J^T}_{2c}e^J_{2c}$ are designed as follows:

$$\dot{\hat{W}}_{2c} = -k_{2c}\frac{\Delta\Phi_{2c}}{\Phi_{2c}}\hat{W}^T_{2c}e^J_{2c}, \tag{37}$$

where $\Phi_{2c} = \left(1 + \Delta\Phi^T_{2c}\Delta\Phi_{2c}\right)^2$. Additionally, $k_{2c} > 0$ represents the learning rate for the adaptive update law of the evaluation network weights.

Additionally, to maintain the stability of the policy updates, the weight update rule for the action network is reformulated as follows:

$$\begin{aligned}
\dot{\hat{W}}_{1a} = {} & -k_{1a}\left[\frac{\lambda}{\Gamma_{2c}}\nabla\Phi_{1c}f\left(\delta\right)\vartheta\Delta\Phi^T_{2c}\hat{W}_{2c} - \frac{1}{2}D_{1c}\hat{W}_{1c}\right. \\
& \left. - \frac{\lambda}{4\Gamma_{1c}}D_{1c}\hat{W}_{1a}\Delta\Phi^T_{1c}\hat{W}_{1c} - \frac{1}{2}\nabla\Phi_{1c}gG^{-1}g^T\mathfrak{I}_e\right] - k_{1a}\hat{W}_{1a}, \\
\dot{\hat{W}}_{2a} = {} & -k_{2a}\left[\lambda D_{2c}\hat{W}_{2a}\Delta\Theta^T_{2c}\left(\hat{W}_{1c}\middle/\left(4\Gamma_{1c}\right) - \hat{W}_{2c}\middle/\left(4\Gamma^2_{2c}\right)\right)\right. \\
& \left. - \frac{1}{2}D_{1c}\hat{W}_{2c} - \frac{1}{2}\nabla\Phi_{2c}gR^{-1}g^T\mathfrak{I}_e\right] - k_{2a}\hat{W}_{2a},
\end{aligned} \tag{38}$$

where, $D_{1c} = \nabla\Phi_{1c}gR^{-1}g^T\nabla\Phi^T_{1c}, \quad D_{2c} = \nabla\Phi_{2c}gR^{-1}g^T\nabla\Phi^T_{2c}$.

Building upon the preceding analysis, we can formulate the following Theorem 2:

Theorem 2: Consider an unmanned vessel system with partially unknown dynamics, subjected to the approximately optimal disturbance strategy (27) and update rules (29) and (38). The unmanned vessel system is designed with an approximately optimal disturbance-resistant control strategy (35), which includes update rules (37) and (38), ensuring ideal tracking of the trajectory under disturbances, while keeping all signals, as well as the tracking errors e and $\mathfrak{I}_e$, bounded within the closed-loop system.

Proof: The following Lyapunov function is utilized in our analysis to assess system stability and performance:

$$L = L_1 + L_2 + L_3, \tag{39}$$

where $\mathrm{L}_1 = e^T e/2 + {\mathfrak{I}_e}^T \mathfrak{I}_e/2, \mathrm{L}_2 = \tilde{W}_{2c}^T k_{2c}^{-1} \tilde{W}_{2c}/2 + \tilde{W}_{2a}^T k_{2a}^{-1} \tilde{W}_{2a}/2, \mathrm{L}_3 = V_a^* \left(\mathfrak{I}_e\right) + \tilde{W}_{1c}^T k_{1c}^{-1} \tilde{W}_{1c}/2 + \tilde{W}_{1a}^T k_{1a}^{-1} \tilde{W}_{1a}/2.$

Step 1: Taking the derivative of L_1 with respect to time, it obtains $\dot{\mathrm{L}_1} = e^T \dot{e} + {\mathfrak{I}_e}^T \dot{\mathfrak{I}}_e$. Combining this with (4) and $e^T \dot{e} = -e^T \Gamma_1 e + e^T v_d \mathfrak{I}_e$, one derives ${\mathfrak{I}_e}^T \dot{\mathfrak{I}}_e$, which can be expressed as:

$$\dot{\mathrm{L}}_1 \leq -e^T \Gamma_1 e - {\mathfrak{I}_e}^T K_{\mathfrak{I}_e} \mathfrak{I}_e + \frac{1}{2} z_i^T D_1 \tilde{W}_{1a} + \frac{1}{2} {\mathfrak{I}_e}^T D_2 \tilde{W}_{2a} + \left[b_\varepsilon^{ih} + \frac{1}{2} \left(b_\pi^{1i} + b_\pi^{2i} \right) \right] \|\mathfrak{I}_e\|, \tag{40}$$

Step 2: By taking the time derivative of L_2, we derive $\dot{L}_2 = -\tilde{W}_{2a}^T k_{2a}^{-1} \dot{\hat{W}}_{2a} - \tilde{W}_{2c}^T k_{2c}^{-1} \dot{\hat{W}}_{2c}$. Following this, the substitution of equations (37) and (38) into $\dot{L}_2$ yields the subsequent result:

$$\begin{aligned} \dot{\mathrm{L}}_2 \leq & -k_{2a} \tilde{W}_{2a}^{\prime} \tilde{W}_{2a} - k_{2c} \tilde{W}_{2c}^{\prime} \tilde{W}_2 - \frac{1}{2} \tilde{W}_{2a}^{\prime} D_2^T \mathfrak{I}_e - \frac{\lambda}{4\Gamma_{2c}} \tilde{W}_{2a}^{\prime} D_4 \hat{W}_{2a} \mathfrak{I}_e \Phi_{2c}^{\prime} \hat{W}_{2c} + k_{2a} \left(W_{2c}^*\right)^T \tilde{W}_{2a} \\ & + k_{2c} \left(W_{2c}^*\right)^T \tilde{W}_{2c} + \frac{1}{\Gamma_{2c}} \tilde{W}_{2c}^{\prime} \Delta\Phi_{2c} \left(p_c + \Delta V_c^1 \right), \end{aligned} \tag{41}$$

where $k_{2c} = \|\Phi_{2c}\|^2 / \Gamma_{2c}$. Meanwhile, let $p_c = \int_{t-\lambda}^t \left(r_{2c}(\mathfrak{I}_e lta, \hat{\pi}_{\tau a}, \hat{\pi}_{\tau c}) + \vartheta \dot{\hat{\pi}}_{\tau c} \right) ds$. To further elaborate, combining with $\varepsilon_{l\dot{g}b}^{2c} = {\mathfrak{I}_e}^T Q \mathfrak{I}_e + \left(\pi_{\tau c}^*\right)^T R \pi_{\tau c}^* + \vartheta \dot{\lambda}_1 + \left(\nabla\Phi_{2c}^{\prime} W_{2c}^* + \nabla V_c^1 \right)^T \left[f\left(\mathfrak{I}_e\right) + g\pi_{\tau d}^* + g\pi_{\tau c}^* \right]$, it yields:

$$\begin{aligned} \frac{1}{\Gamma_{2c}} \tilde{W}_{2c}^{ic^r} \Delta\Phi_{2c} \left(p_c + \Delta V_c^1 \right) = & \frac{1}{\Gamma_{2c}} \tilde{W}_{2c}^{\prime} \Delta\Phi_{2c} \Bigg[\overline{p}_c - \int_{t-\lambda}^t \nabla V_c^1 d\left(\mathfrak{I}_e\right) + \Delta V_c^1 \\ & + \int_{t-\lambda}^t \left[\hat{\pi}_{\tau c}^T R \hat{\pi}_{\tau c} - \left(\pi_{\tau c}^*\right)^T R \pi_{\tau c}^* \right] ds \Bigg] \\ = & \frac{\overline{p}_c}{\Gamma_{2c}} \nabla\Phi_{2c}^T \tilde{W}_{2c} - \frac{\lambda}{2\Gamma_{2c}} \tilde{W}_{2c} D_{2c} \tilde{W}_{2a} + \frac{h}{4\Gamma_{2c}} \tilde{W}_{2c}^T \Delta\Phi_{2c} \tilde{W}_{2a} D_4 \tilde{W}_{2a}, \end{aligned} \tag{42}$$

where $\overline{p}_c = \int_{t-\lambda}^t \varepsilon_{hjb}^c - \left(W_{2c}^*\right)^T \nabla\Phi_{2c} \left[f\left(\mathfrak{I}_e\right) + g\pi_{\tau c}^* + g\pi_{\tau d}^* \right]$, $D_{2c} = \Delta\Phi_{2c} \left(\nabla V_c^*\right)^T \Psi^T g R^{-1} g^T \Psi \nabla\Phi_{2c}^T$. To take it a step further, $\dot{L}_2$ can be further derived as:

$$\begin{aligned} \mathrm{L}_2 \leq & -k_{2a} \tilde{W}_{2a}^T \tilde{W}_{2a} + \frac{\lambda}{4\Gamma_{2c}} \left(W_{2c}^*\right)^T \Delta\Phi_{2c} \tilde{W}_{2a}^T \Pi_4 \tilde{W}_{2a} - k_{2c} \tilde{W}_{2c}^T \tilde{W}_{2c} + \frac{\lambda}{4\Gamma_{2c}} \tilde{W}_{2c}^T \Delta\Phi_{2c} \left(W_{2c}^*\right)^T \Pi_4 \tilde{W}_{2a} \\ & - \frac{\lambda}{2\Gamma_{2c}} \tilde{W}_{2c}^T D_{2c} \tilde{W}_{2a} + k_{2a} \left(W_{2c}^*\right)^T \tilde{W}_{2a} + \frac{\overline{p}_c}{\Gamma_{2c}} \Delta\Phi_{2c}^T \tilde{W}_{2c} - \frac{\lambda}{4\Gamma_{2c}} \left(W_{2c}^*\right)^T \Delta\Phi_{2c} \left(W_{2c}^*\right)^T \Pi_4 \tilde{W}_{2a} \\ & + k_{2c} \left(W_{2c}^*\right)^T \tilde{W}_{2c} - \frac{1}{2} \tilde{W}_{2a}^T \Pi_2^T \mathfrak{I}_e, \end{aligned} \tag{43}$$

where, $\prod_1 = g G^{-1} g^T \nabla\Phi_{1c}^T, \quad \Pi_3 = \nabla\Phi_{1c}^T \prod_1$.

Step 3: By performing a time derivative of L_3, we obtain $\dot{L}_3 = -\tilde{W}_{1a}^T k_{1a}^{-1} \dot{\hat{W}}_{1a} - \tilde{W}_{1c}^T k_{1c}^{-1} \dot{\hat{W}}_{1c}$. The integration of this result, in conjunction with equations (29) and (38), leads to the following outcome:

$$\begin{aligned} \dot{\mathrm{L}}_3 \leq & -k_{1a} \tilde{W}_{1a}^T \tilde{W}_{1a} - k_{1c} \tilde{W}_{1c}^T \tilde{W}_{1c} - \frac{1}{2} \tilde{W}_{1a}^T \Pi_1^T \mathfrak{I}_e - \frac{h}{4\Gamma_{1a}} \tilde{W}_{1a}^T \Pi_3 \hat{W}_{1a} \Delta\Phi_{1c}^T \hat{W}_{1c} + k_{1a} \left(W_{1c}^*\right)^T \tilde{W}_{1a} \\ & + k_{1c} \left(W_{1c}^*\right)^T \tilde{W}_{1c} + \frac{1}{m\Gamma_{1a}} \tilde{W}_{1c}^T \Delta\Phi_{1c} p_a, \end{aligned} \tag{44}$$

where $\Gamma_{1a} = \left(1 + \Delta\Phi_{1c}^T \Delta\Phi_{1c}\right)^2$ and $k_{_{1c}} = \left\|\Phi_{_{1c}}\right\|^2 / \Gamma_{_{1a}}$.

By integrating $\varepsilon_{hjb}^{1a} = -{\mathfrak{I}_e}^T Q \mathfrak{I}_e + \left(\pi_{\tau d}^*\right)^T G \pi_{\tau d}^* + \hat{\pi}_{\tau c}^T R \pi_{\tau d}^* + \left(W_{1c}^*\right)^T \nabla\Phi_{1c} \left[f\left(\mathfrak{I}_e\right) + g\hat{\pi}_{\tau c} + g\pi_{\tau d}^* \right]$ with $p_a = \int_{t-\lambda}^t \left(r_{1a}(\mathfrak{I}_e, \hat{\pi}_{\tau d}, \hat{\pi}_{\tau c}) \right) ds$, and applying a differentiation process analogous to that employed in deriving $\dot{L}_2$, we can subsequently obtain $\dot{L}_3$ as follows:

$$\begin{aligned}\mathrm{L}_3 \leq & -k_{1a}\tilde{W}_{1a}^T\tilde{W}_{1a} + \frac{\lambda}{4\Gamma_{1a}}\left(W_{1c}^*\right)^T \Delta\Phi_{1c}\tilde{W}_{1a}^T\Pi_3\tilde{W}_{1a} - k_{1c}\tilde{W}_{1c}^T\tilde{W}_{1c} + \frac{\lambda}{4\Gamma_{1a}}\tilde{W}_{1c}^T\Delta\Phi_{1c}\left(W_{1c}^*\right)^T\Pi_3\tilde{W}_{1a} \\ & - \frac{\lambda}{2\Gamma_{1a}}\tilde{W}_{1c}^T D_{1c}\tilde{W}_{1a} + k_{1a}\left(W_{1c}^*\right)^T\tilde{W}_{1a} + \frac{\lambda}{2\Gamma_{1a}}D_{1c}\tilde{W}_{1c} - \frac{\lambda}{4\Gamma_{1a}}\left(W_{1c}^*\right)^T\Delta\Phi_{1c}\left(W_{1c}^*\right)^T\Pi_3\tilde{W}_{1a} \\ & + \frac{\overline{p}_a}{\Gamma_{1a}}\Delta\Phi_{1c}^T\tilde{W}_{1c} + k_{1c}\left(W_{1c}^*\right)^T\tilde{W}_{1c} - \frac{1}{2}\tilde{W}_{1a}^T\Pi_1^T\Im_e,\end{aligned} \tag{45}$$

where $\Pi_2 = g\Psi R^{-1}g^T\Psi\nabla\Phi_{2c}^T, \Pi_4 = \nabla\Phi_{2c}\Psi^T gR^{-1}g^T\Psi\nabla\Phi_{2c}^T$.

Step 4: In light of the extensive analysis provided above, we are now positioned to derive:

$$\begin{aligned}\mathrm{L} \leq & -\Gamma_1 e_i^T e_i - K_{\Im_e}\Im_e{}^T\Im_e - k_{2c}\tilde{W}_{2c}^T\tilde{W}_{2c} - k_{2a}\tilde{W}_{2a}^T\tilde{W}_{2a} - k_{1c}\tilde{W}_{1c}^T\tilde{W}_{1c} \\ & - k_{1a}\tilde{W}_{1a}^T\tilde{W}_{1a} + \Psi_1\tilde{W}_{2c} + \Psi_2\tilde{W}_{2a} + \Psi_3'\tilde{W}_{1c} + \Psi_4\tilde{W}_{1a} + \tilde{W}_{2c}^T\Psi_5\tilde{W}_{2a} \\ & + \tilde{W}_{1c}^T\Psi_6\tilde{W}_{1a} + \tilde{W}_a^{ic^\gamma}\frac{\lambda}{4\Gamma_{2c}}\left(W_{2c}^*\right)^T\Delta\Phi_{2c}\Pi_4\tilde{W}_a^{ic} + \tilde{W}_a^{ia^\gamma}\frac{\lambda}{4\Gamma_{1a}}\left(W_{1c}^*\right)^T\Delta\Phi_{1c}\Pi_3\tilde{W}_a^{ia},\end{aligned} \tag{46}$$

where $\Psi_1 = \frac{\overline{p}_c}{\Gamma_{2c}}\Delta\Phi_{2c}^T + k_{1c}\left(W_{1c}^*\right)^T$, $\Psi_2 = k_{2a}\left(W_{2c}^*\right)^T - \frac{\lambda}{4\Gamma_{2c}}\left(W_{2c}^*\right)^T\Delta\Phi_{2c}\left(W_{2c}^*\right)^T\Pi_4$, $\Psi_3 = \frac{\lambda}{2\Gamma_{1a}}D_{1c} + \frac{\overline{p}_a}{\Gamma_{1a}}\Delta\Phi_{1c}^T + k_{1c}\left(W_{1c}^*\right)^T$, $\Psi_4 = k_{1a}\left(W_{1c}^*\right)^T - \frac{\lambda}{4\Gamma_{1a}}\left(W_{1c}^*\right)^T\Delta\Phi_{1c}\left(W_{1c}^*\right)^T\Pi_{3i}$, $\Psi_5 = \frac{\lambda}{4\Gamma_{1a}}\Delta\Phi_{2c}\left(W_{2c}^*\right)^T\Pi_4 - \frac{\lambda}{2\Gamma_{2c}}D_{2c}$, $\Psi_6 = \frac{\lambda}{4\Gamma_{1a}}\Delta\Phi_{2c}\left(W_{1c}\right)^T\Pi_3 - \frac{\lambda}{2\Gamma_{1a}}D_{1c}$.

In addition, leveraging the principles outlined in Young's inequality, the following result can be deduced:

$$\begin{aligned} & \Psi_1\tilde{W}_{2c} \leq \frac{k_{2c}}{2}\tilde{W}_{2c}^T\tilde{W}_{2c} + \frac{(\Psi_1)^2}{2k_{2c}}, \Psi_2\tilde{W}_{2a} \leq \frac{k_{2a}}{2}\tilde{W}_{2a}^T\tilde{W}_{2a} + \frac{(\Psi_2)^2}{2k_{2a}}, \\ & \Psi_3\tilde{W}_{1c} \leq \frac{k_{1c}}{2}\tilde{W}_{1c}^T\tilde{W}_{1c} + \frac{(\Psi_3)^2}{2k_{1c}}, \Psi_4\tilde{W}_{1a} \leq \frac{k_{1a}}{2}\tilde{W}_{1a}^T\tilde{W}_{1a} + \frac{(\Psi_4)^2}{2k_{1a}}, \\ & b_{\Im_e}\left\|\Im_e\right\| \leq \frac{k_{\Im_e}}{2}\Im_e{}^T\Im_e + \frac{b_{\Im_e ta}^2}{2k_{\Im_e}}.\end{aligned}$$

In this way, it can be effectively simplified to:

$$\mathrm{L} \leq -a\mathrm{L} + b, \tag{47}$$

where $a = \min\left\{2\lambda_{\min}\left(\Gamma_1\right), \lambda_{\min}\left(K_{\Im_e}\right), \frac{k_{2c}}{\lambda_{\min}\left(\Gamma_{2c}^{-1}\right)}, \frac{k_{2a}}{\lambda_{\min}\left(\Gamma_{2a}^{-1}\right)}, \frac{k_{1c}}{\lambda_{\min}\left(\Gamma_{1c}^{-1}\right)}, \frac{k_{1a}}{\lambda_{\min}\left(\Gamma_{1a}^{-1}\right)}\right\}$, $b = \frac{(\Psi_1)^2}{2k_{2c}} + \frac{(\Psi_2)^2}{2k_{2a}} + \frac{(\Psi_3)^2}{2k_{1c}} + \frac{(\Psi_4)^2}{2k_{1a}} + \frac{b_{\Im_e}^2}{2k_{\Im_e}}$.

In addition, recognizing that:

$$\mathrm{L}(t) \leq \left(\mathrm{L}(0) - \frac{b}{a}\right)e^{-at} + \frac{b}{a} \leq \mathrm{L}(0) - \frac{b}{a}, \tag{48}$$

Consequently, leveraging the principles established by the Lyapunov stability theorem[19], it can be inferred that the variables $e, \Im_e, \tilde{W}_{2a}, \tilde{W}_{2c}, \tilde{W}_{1a}$, and $\tilde{W}_{1c}$ remain constrained within bounded limits throughout the operation of the closed-loop system.

4. SIMULATION

This section examines the efficacy of the proposed Stackelberg game-based anti-disturbance strategy for trajectory tracking of the USVs, addressing partially uncertain dynamics and externally bounded disturbances. The

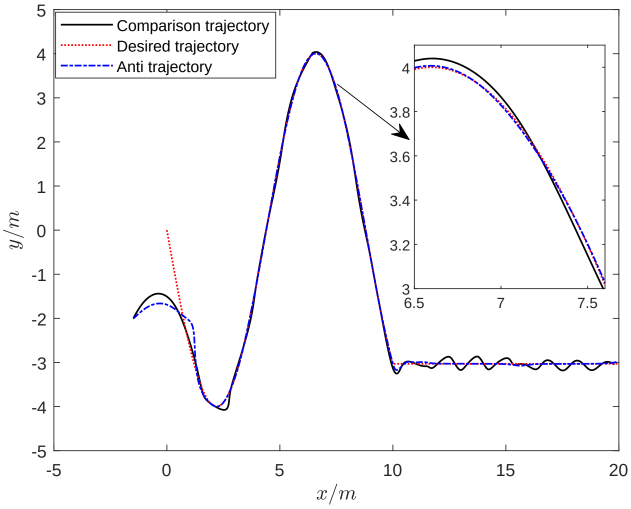

Figure 2. Comparision, desired and anti trajectory within the proposed Stackelberg game-oriented anti-disturbance framework.

dynamics of the USVs can be modeled for simulation purposes as shown in[19]. The simulation scenario parameters of the system, along with the user-defined control variables, are specified as follows: $R = 0.38I_3, Q = 20.3I_3, G = 0.89I_3, \Gamma_1 = 2.9I_3, K_{\mathfrak{I}_e} = 2.2I_3, k_{1c} = 0.94I_6, k_{1a} = 0.53I_6, k_{2c} = 0.81I_6, k_{2a} = 0.62I_6$, where I_N denotes an N-dimensional identity matrix. In addition, the initial position and velocity of the USVs are specified as follows: $\aleph = [-1.41, -1.98, 0]^T, v = [0.5, 0, 0]^T$, and the desired trajectory is as follows: $\aleph_d = \begin{cases} \left[0.23t, 4\sin\left(\frac{t}{7.5}\right), \arctan\left(\frac{4.1}{6.8}\sin\left(t/6.8\right)\right)\right]^T, & \text{if} \quad t < 50 \\ \left[0.23t, 4\sin\left(\frac{50}{7}\right), \arctan\left(\frac{4.1}{6.8}\sin\left(t/6.8\right)\right)\right]^T, & \text{if} \quad t \geq 50 \end{cases}$. Meanwhile, the external disturbances affecting the USVs and the system uncertainties are set as follows: $D = [d_1, d_2, d_3]^T$, $d_1 = -7.5\sin(t), d_2 = 5.2\sin(t)\cos(0.1t), d_3 = -3t$, and $\Delta F(v) = \begin{bmatrix} \Delta_{11} & 0 & 0 \\ 0 & \Delta_{22} & \Delta_{23} \\ 0 & \Delta_{32} & \Delta_{33} \end{bmatrix}$, $\Delta_{11} = 0.68 + 1.29\,|u| + 5.86u^2, \Delta_{22} = 0.89 + 36.2\,|v| + 8.1\,|r|\,, \Delta_{23} = -0.11 + 0.832\,|v| + 3.27\,|r|\,, \Delta_{32} = -0.11 - 5.04\,|v| - 0.13\,|r|\,, \Delta_{33} = 1.9 - 0.08\,|v| + 0.75\,|r|$. Based on the adjustments made in the simulations described above, the numerical simulation results are presented as follows:

As illustrated in Figure 2, the trajectory tracking outcomes are depicted, highlighting a comparison with the tracking performance achieved through sliding mode control supported by a disturbance observer, named as Comparison trajectory, and the proposed Stackelberg game-based anti-disturbance approach demonstrates the capability to achieve accurate and stable tracking of the desired trajectory for the USVs, even in the face of significant unknown dynamics and external bounded disturbances, named as Anti trajectory. Figure 3 illustrates the tracking errors related to both attitude and velocity, providing compelling evidence of the effectiveness of this method in achieving precise trajectory tracking of the USVs in the presence of external bounded disturbances. The conventional anti-interference approach, which combines observers with sliding mode control, faces a critical limitation: when there is a deviation in the estimation of disturbances, the robust nature of sliding mode control leads to large corrective actions aimed at driving the error to zero. While this accelerates convergence, it often results in excessive overshoot, as observed in the trajectory within the [10, 15] m interval. In contrast, this study introduces an innovative framework for estimating unknown disturbances, coupled with a control strategy grounded in reinforcement learning. Through an iterative interaction process, the framework simultaneously optimizes control strategies $\pi_{\tau d}$ (27) and $\pi_{\tau c}$ (35), ultimately achieving the Stackelberg equilibrium. At this equilibrium, neither the interference rejection strategy nor the optimal con-

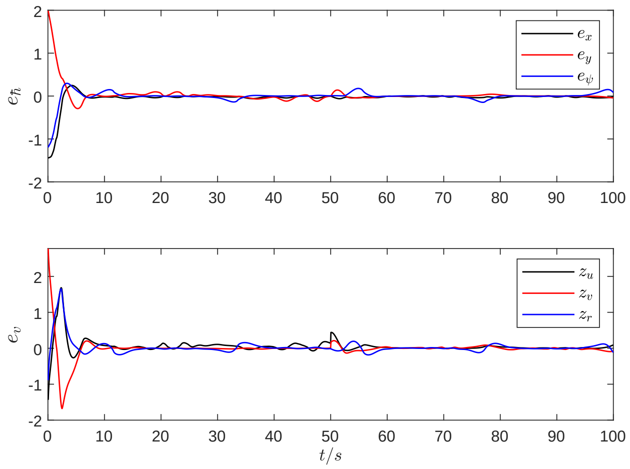

Figure 3. Tracking errors $e_{\hbar}$ and e_v within the proposed Stackelberg game-oriented anti-disturbance framework.

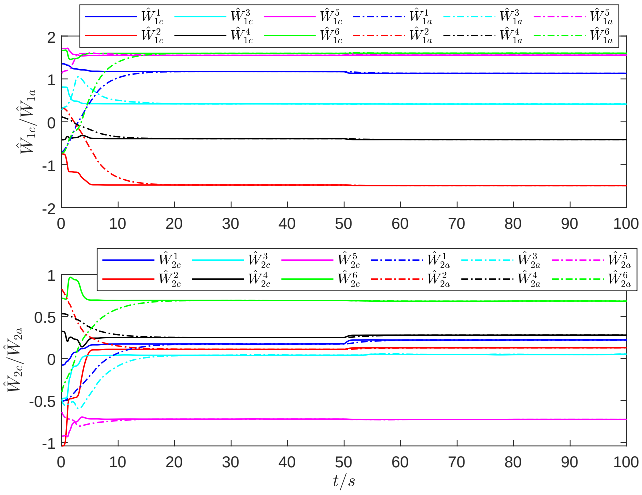

Figure 4. Convergence of weights for actor and critic NNs within the Stackelberg game-based anti-disturbance framework.

trol strategy can further reduce the cost function values $V_1(J_1)$ (8) and $V_2(J_2)$ (9) by adjusting their respective gains. This approach enables the USV to rapidly detect and respond to unknown environmental disturbances, even under highly dynamic conditions. By optimizing the control strategy in conjunction with disturbance estimation, the method ensures that the USV attains a Nash equilibrium, balancing robustness and optimal control. As a result, the trajectory demonstrates enhanced accuracy and robustness, particularly evident in the [10, 15] m range. This strategy effectively mitigates the limitations of traditional interference rejection methods, while keeping tracking errors within an acceptable threshold. In contrast, conventional approaches rely primarily on disturbance estimation via observers and robust controllers, without the coordinated interplay between estimation and control, thereby limiting their capacity to address complex, time-varying environments.

In Figure 4, the convergence trends of the weights for the actor and critic neural networks, which illustrate the disturbance-resistant control strategy and the auxiliary compensation policy for unmodeled dynamics and external disturbances, are presented. Meanwhile, Figure 5 illustrates the norm convergence curve of the weights utilized in the approximation of the unknown dynamics, which encompass several unmodeled system dynamics and bounded external disturbances. Figures 4 and 5 demonstrate that, utilizing a sequential decision-

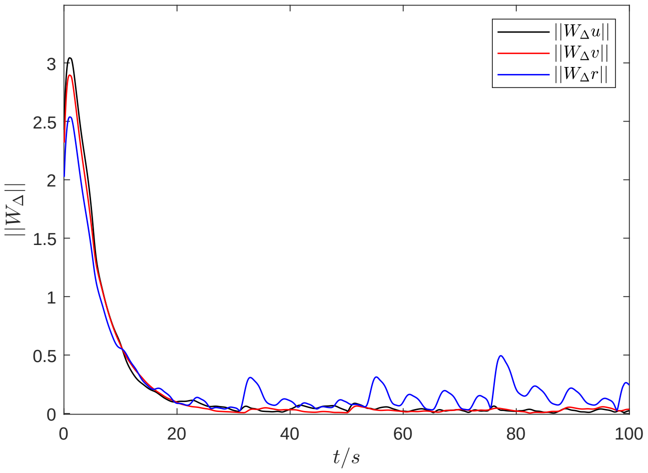

Figure 5. Norm convergence of weights for unknown information that encompasses external disturbances and uncharacterized system dynamics.

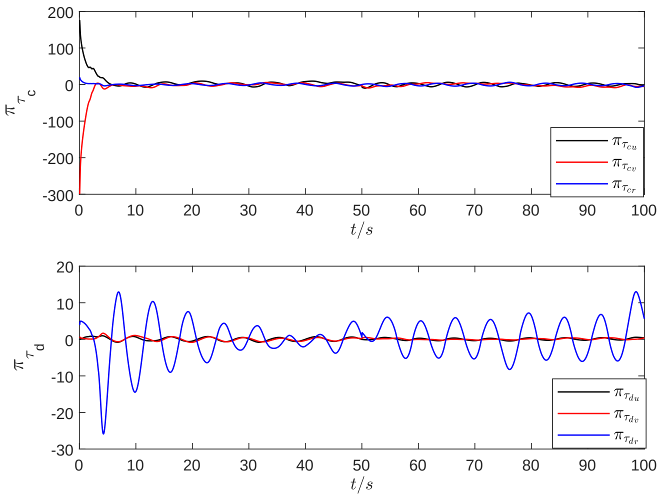

Figure 6. Control input within the proposed Stackelberg game-oriented anti-disturbance framework.

making mechanism, the weight curves of the neural networks converge rapidly to optimal values and maintain stability within a defined range. This finding offers substantial evidence that the proposed approach is capable of achieving the Nash equilibrium solution for the Stackelberg game using integral reinforcement learning. As illustrated in Figure 6, both the disturbance-assisted control signal and anti-disturbance input reveal that, when employing the optimal disturbance-assisted control strategy, the anti-disturbance mechanism achieves superior tracking accuracy, thereby enhancing the operational safety of the USVs.

5. CONCLUSION

This study explores the challenges USVs encounter during navigation and introduces an innovative anti-disturbance control strategy tailored for partially known dynamic systems, leveraging Stackelberg game theory. Within this theoretical framework, we formulate a sequential non-cooperative game that incorporates control inputs. To enhance the optimization process, we employ an action-evaluation integral reinforcement learning algorithm designed to directly minimize the Bellman error, deriving an approximately optimal solution. Moreover, auxiliary neural networks are integrated to accurately approximate the unknown dynamics and external disturbances affecting the system. Simulation results substantiate the efficacy and superiority of the proposed Stackelberg game-based integral reinforcement learning control strategy in mitigating disturbances in USVs. Future research will concentrate on the development of optimal anti-jamming, fault-tolerant, and cooperative obstacle avoidance strategies for multiple USVs, grounded in Stackelberg game theory, with a particular emphasis on scenarios involving deception attacks and complex multi-obstacle environments.

DECLARATIONS

Authors' contributions

Writing - original draft: Meng, Y.
Writing - review: Liu, C.
Writing - editing: Zhao, J.
Conceptualization: Huang, J.
Validation: Jing, G.

Availability of data and materials

Not applicable.

Financial support and sponsorship

This work was supported by the National Key R&D Program of China (2022YFB3902702); National Natural Science Foundation of China (62103250, 62273223, 62333011, and 62336005); Project of Science and Technology Commission of Shanghai Municipality, China (22JC1401401).

Conflicts of interest

Liu, C. is a Junior Editorial Board Member of the journal *Intelligence & Robotics*. He is not involved in any steps of editorial processing, notably including reviewer selection, manuscript handling, or decision-making. The other authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Ma, S.; Guo, W.; Song, R.; Liu, Y. Unsupervised learning based coordinated multi-task allocation for unmanned surface vehicles. *Neurocomputing* **2021**, *420*, 227-45. DOI
2. Wang, Q.; Liu, C.; Lan, J.; Ren, X.; Meng, Y.; Wang, X. Distributed secure surrounding control for multiple USVs against deception attacks: a Stackelberg game approach with reinforcement learning. *IEEE Trans. Intell. Veh.* **2024**, 1-12. DOI
3. Zhao, J.; Wang, Y.; Cai, Z.; Liu, N.; Wu, K.; Wang, Y. Learning visual representation for autonomous drone navigation via a contrastive world model. *IEEE Trans. Artif. Intell.* **2024**, *5*, 1263-76. DOI
4. Guo, J.; Wang, X.; Xue, W.; Zhao, Y. System identification with binary-valued observations under data tampering attacks. *IEEE Trans. Autom. Control.* **2021**, *66*, 3825-32. DOI
5. Cui, Y.; Peng, L.; Li, H. Filtered probabilistic model predictive control-based reinforcement learning for unmanned surface vehicles *IEEE. Trans. Ind. Informat.* **2022**, *18*, 6950-61. DOI

6. Cui, Y.; Li, A.; Meng, X. A fault-tolerant control method for distributed flight control system facing wing damage. *J. Syst. Eng. Electron.* **2021**, *32*, 1041-52. DOI
7. Qu, Y.; Cai, L. Nonlinear positioning control for underactuated unmanned surface vehicles in the presence of environmental disturbances. *IEEE/ASME Trans. Mechatronics.* **2022**, *27*, 5381-91. DOI
8. Peng, Z.; Wang, D.; Wang, J. Data-driven adaptive disturbance observers for model-free trajectory tracking control of maritime autonomous surface ships. *IEEE Tran. Neural. Netw. Learn. Syst.* **2021**, *32*, 5584-94. DOI
9. Xu, J.; Fang, H.; Zhang, B.; Guo, H. High-frequency square-wave signal injection based sensorless fault tolerant control for aerospace FTPMSM system in fault condition. *IEEE Trans. Transp. Electrification.* **2022**, *8*, 4560-8. DOI
10. Zhao, X.; Liu, C.; Zhao J. Adaptive sliding mode-based faulttolerant tracking control of multi-USV systems. In *2022 34th Chinese Control and Decision Conference (CCDC)*, Hefei, China, Aug 15-17, 2022; IEEE, 2022; pp 5980-5. DOI
11. Yu, X. N.; Hao, L. Y.; Wang, X. L. Fault tolerant control for an unmanned surface vessel based on integral sliding mode state feedback control. *Int. J. Control. Autom. Syst.* **2022**, *20*, 2514-22. DOI
12. Kebriaei, H.; Iannelli, L. Discrete-time robust hierarchical linear quadratic dynamic games. *IEEE Trans. Autom. Control.* **2018**, *63*, 902-9. DOI
13. Xu, Y.; Yang, H.; Jiang, B.; Polycarpou, M. M. Distributed optimal fault estimation and fault-tolerant control for interconnected systems: a Stackelberg differential graphical game approach. *IEEE Trans. Autom. Control.* **2022**, *67*, 926-33. DOI
14. Li, M.; Qin, J.; Ma, Q.; Zheng, W. X.; Kang, Y. Hierarchical optimal synchronization for linear systems via reinforcement learning: a Stackelberg–nash game perspective. *IEEE Trans. Autom. Control.* **2021**, *32*, 1600-11. DOI
15. Li, M.; Qin, J.; Freris, N. M.; Ho, D. W. C. Multiplayer stackelberg-Nash game for nonlinear system via value iteration-based integral reinforcement learning. *IEEE Trans. Neural. Netw. Learn. Syst.* **2022**, *33*, 1429-40. DOI
16. Chu, Z.; Wang, F.; Lei, T.; Luo, C. Path planning based on deep reinforcement learning for autonomous underwater vehicles under ocean current disturbance. *IEEE Trans. Intell. Veh.* **2023**, *8*, 108-20. DOI
17. 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*, 5468-78. DOI
18. Cui, X.; Wang, B.; Wang, L.; Chen, J. Online optimal learning algorithm for Stackelberg games with partially unknown dynamics and constrained inputs. *Neurocomputing* **2021**, *445*, 1-11. DOI
19. Guo, X.; Yan, W.; Cui, R. Integral reinforcement learning-based adaptive NN control for continuous-time nonlinear MIMO systems with unknown control directions. *IEEE Tran. Syst. Man. Cybern. Syst.* **2020**, *50*, 4068-77. DOI