Prescribed Performance Robust Approximate Optimal Tracking Control Via Stackelberg Game

Junkai Tan, Shuangsi Xue, Huan Li, Zihang Guo, Hui Cao and Dongyu Li

***Abstract*—Real-world applications of nonlinear systems tracking control are always challenging due to the existence of uncertainties and disturbances. To design a robust optimal tracking controller for uncertain nonlinear systems with disturbances and actuator saturation, this paper investigates the prescribed performance robust optimal tracking control problem. A prescribed performance mechanism is constructed to convert the dynamics of tracking error into transformed error dynamics, which keeps the system's operating states within specific bounds, ensuring tracking with predefined error constraints. For the optimal tracking controller design, an optimal index is established to optimize the performance of tracking control, and a robust optimal index is established to optimize the disturbance effect on the tracking error. To achieve robust optimal tracking control that minimizes both optimal and robust optimal indexes, a Stackelberg game is constructed, which provides a hierarchical game structure for the optimal controller and the worst disturbance. The robust optimal controller is approximated online using reinforcement learning techniques. An actor-critic-identifier algorithm is designed to approximate the optimal value function, optimal controller, and drifted system parameters. Lyapunov theory is utilized to analyze the closed-loop system's stability. To demonstrate the effectiveness of the proposed robust optimal control method, two numerical simulations and a hardware experiment on a quadcopter system are conducted. The experiment results demonstrate that our method successfully achieves prescribed performance tracking control when actuators are saturated and disturbances are present.**

***Note to Practitioners*—In this paper, the probelm of mixed H_2/H_∞ prescribed-performance optimal tracking control for nonlinear systems with input saturation is investigated. To constrain the operating states of the system within certain bounds, the prescribed performance transformation is designed to achieve tracking with predefined error constraints. For the optimal controller design, the H_2 index is established to minimize the optimal tracking performance, and the H_∞ index is designed to minimize the disturbance effect on the tracking error. A Stackelberg-based non-zero sum game between the optimal controller and the worst disturbance is established to design the mixed H_2/H_∞ optimal tracking controller. The designed optimal controller is approximated online using reinforcement learning. Effectiveness of the proposed method is demonstrated by two numerical simulations and a hardware experiment on a quadcopter system. Based on the proposed high-performance controller, engineers can design a high-performance robust optimal tracking controller for uncertain nonlinear systems with extreme conditions of disturbances and actuator saturation.**

***Index Terms*—Prescribed performance control, Stackelberg game, approximate optimal control, actor-critic, system identification**

This work was supported in part by the National Natural Science Foundation of China under Grant 62103028 and 62473023, Beijing Natural Science Foundation under Grant 4242043, and China Postdoctoral Science Foundation (Project ID: 2024M762602).

Junkai Tan, Shuangsi Xue, Huan Li, Zihang Guo and Hui Cao are with the Shaanxi Key Laboratory of Smart Grid, Xi'an Jiaotong University, Xi'an 710049, China, and also with the State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: tanjk@stu.xjtu.edu.cn; xssxjtu@xjtu.edu.cn; lh2000dami@stu.xjtu.edu.cn; guozihang@stu.xjtu.edu.cn; huicao@mail.xjtu.edu.cn) (*Corresponding author: Shuangsi Xue*).

Dongyu Li is with the School of Cyber Science and Technology, Beihang University, Beijing 100191, China (e-mail: dongyuli@buaa.edu.cn).

I. Introduction

In real-world applications, tracking controller design for nonlinear systems faces significant challenges due to inherent uncertainties [1] and external disturbances [2]. A crucial challenge in designing effective tracking controllers lies in achieving an optimal balance between robustness against disturbances and control performance [3], [4]. This creates a fundamental trade-off between system response speed and disturbance rejection capabilities [5], as faster response speeds often compromise robustness against disturbances [6]. Two prominent control methodologies have been extensively studied to address these challenges: the H_2 optimal performance control [7] and the H_∞ optimal robust control [8]. The H_2 index aims to minimize the tracking error between system output and desired trajectory [9], while the H_∞ index focuses on minimizing the impact of disturbances on tracking performance [10]. However, most existing methods tend to prioritize either robustness or optimality, but not both simultaneously, which limits their practical effectiveness in real-world applications where both aspects are crucial [11].

To effectively balance control performance and robustness in tracking controller design, mixed H_2/H_∞ control has emerged as a powerful approach [12], [13]. This method integrates both H_2 and H_∞ performance indices through a game-theoretic framework, where the optimal controller minimizes the H_2 index while the worst-case disturbance maximizes the H_∞ index [14]. A hierarchical Stackelberg game structure has been incorporated into this framework [15], [16], establishing a leader-follower decision-making process that effectively handles the competing objectives. Robust game-theoretic approaches using Nash equilibrium to handle model uncertainties is investigated in [17]. The Hamiltonian-driven methods for robust optimal stabilization of Stackelberg game is discussed in [18]. For partially unknown stochastic systems, mixed H_2/H_∞ learning-based control algorithms is proposed in [19], [20]; While literature [21], [22] investigate data-driven optimized Stackelberg game for nonlinear system control. For the discrete-time formulations incorporating both performance indices, the mixed H_2/H_∞ control is investigated in [23]. However, most existing mixed H_2/H_∞ methods focus primarily on stabilization and optimality without explicitly

This article has been accepted for publication in IEEE Transactions on Automation Science and Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TASE.2025.3549114

considering performance constraints. This limitation motivates our research to develop an integrated solution that combines mixed H_2/H_∞ control with prescribed performance bounds for nonlinear tracking systems.

Uncertainties in real-world applications of nonlinear system tracking control pose significant safety challenges, as unforeseeable disturbances may drive operating states beyond acceptable limits. To address this critical issue, prescribed performance control (PPC) has emerged as an effective approach to constrain tracking errors within predefined state bounds. Significant developments in PPC include robust methods for spacecraft attitude control under external disturbances [24], and quadcopter trajectory tracking with specific operating constraints [25], [26]. To enhance control performance under temporal constraints, finite-time PPC methods have been developed [27]–[29]. For extreme operating conditions involving actuator saturation and faults, adaptive PPC schemes have been proposed to maintain safe system operation [30], [31]. To optimize computational and communication resources, event-triggered PPC mechanisms have been investigated [32], [33]. These developments demonstrate that PPC provides a systematic framework for guaranteeing bounded tracking performance in nonlinear systems, effectively addressing both safety constraints and control objectives simultaneously.

Another common issue in the real-world application of tracking control is the uncertainty or incompleteness of the system model [34], [35]. To address this problem, the reinforcement learning (RL) method is proposed to learn the optimal value function and controller. Among the RL methods, one of them is the model-free RL [36]–[38]. Q-learning algorithm is implemented to learn the optimal stabilizing controller without the prior knowledge of system dynamics in [39]. Integral RL is investigated to approximate the optimal controller for partially unknown nonlinear systems in [40], [41]. The above-mentioned methods are operated in an off-policy offline manner, which requires a large amount of data to train the optimal controller. Another category of RL methods is the model-based RL [42], [43], which mostly requires the pre-known system model to train the optimal controller online. An actor-critic-identifier (ACI) structure is investigated to obtain the optimal value function, controller and drifted system parameters online in [44], [45]. In [46], an ACI-based control algorithm is established to approximate the optimal tracking controller for a nonlinear system with prescribed state constraints. To summarize, the ACI algorithm is capable of learning optimal controllers and identifying the uncertain system parameters online simultaneously, which is suitable for the tracking control of uncertain nonlinear systems.

Motivated by the above discussions, this paper investigates a mixed H_2/H_∞ optimal prescribed performance tracking control problem for uncertain nonlinear systems. A Stackelberg game-based mixed H_2/H_∞-PPC-ACI controller is designed to track the desired trajectories under prescribed state constraints with the existence of disturbances and actuator saturation. Contributions of the paper are summarized in the following points:

1) A high-performance robust optimal tracking controller is designed for uncertain nonlinear tracking systems. To maintain system states within safe operating bounds, a PPC mechanism is incorporated to guarantee predefined tracking performance constraints. Through the design of an H_∞ index for robustness and an H_2 index for optimality, a Stackelberg game framework is established to synthesize a mixed H_2/H_∞ optimal controller. This integrated approach achieves superior performance in balancing robustness and optimality while enforcing state constraints compared to existing methods [21], [41], [46].
2) The optimal control and worst-case disturbance policies derived from the Stackelberg game are approximated using reinforcement learning techniques. An actor-critic-identifier (ACI) structure is developed to simultaneously approximate the optimal value functions and controllers for both H_2 and H_∞ objectives, while an online system identifier estimates uncertain drift parameters in the system dynamics. Through this unified ACI framework, a mixed H_2/H_∞-PPC-ACI controller is synthesized that achieves prescribed-performance robust optimal tracking.
3) Extensive validation of the proposed mixed H_2/H_∞-PPC-ACI controller is conducted through two comprehensive numerical simulations and hardware experiments on a quadcopter platform. The experimental results demonstrate that our method successfully achieves prescribed-performance tracking control even in the presence of significant external disturbances and actuator saturation, while maintaining robust and optimal performance metrics.

The rest of the paper is organized as follows: Section II introduces the nonlinear tracking system and PPC mechanism. Section III formulates the problem of mixed H_2/H_∞ optimal tracking control via the Stackelberg game. Section IV presents the ACI algorithm for approximation of optimal value functions and tracking controller. Section V-VI demonstrates the effectiveness of our proposed mixed H_2/H_∞-PPC-ACI control scheme by two numerical simulations and a hardware experiment.

Notation: Throughout this paper, the following notations are adopted: $\mathbb{R}^n$ denotes the n-dimensional Euclidean space; $\|\cdot\|$ represents the Euclidean norm for vectors or the induced matrix norm; For any matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its minimum and maximum eigenvalues, respectively; For a function $f(x)$, $\nabla f(x)$ denotes its gradient; $\mathcal{I}_n$ represents the $n \times n$ identity matrix; For a set χ, $\overline{\chi}$ denotes its closure; $\emptyset$ represents the empty set.

II. Preliminaries

A. Nonlinear tracking system

Consider the following continuous-time nonlinear-affine systems with disturbance:

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) + k(x(t))\omega(t) \tag{1}$$

where $x(t) = [x_1, ..., x_n] \in \mathbb{R}^n$ is the state of the system in a vector form, $f : \mathbb{R}^n \to \mathbb{R}^n$ denotes the nonlinear drift dynamics matrix, $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the input dynamics matrix of the control input, $u(t) \in \mathbb{R}^m$ denote the control input to the above system, $k : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the input dynamics matrix of the disturbance input, $\omega(t) \in \mathbb{R}^m$ is the disturbance input to the above system, and the drift dynamics matrix $f(x)$, input dynamics matrix $g(x)$ and $k(x)$ are all locally Lipschitz continuous functions. Consider the following dynamics for the state $x_d(t) \in \mathbb{R}^n$ of desired trajectory:

$$\dot{x}_d(t) = f_d(x_d(t)) \tag{2}$$

where $f_d : \mathbb{R}^n \to \mathbb{R}^n$ is the drift dynamics matrix of the desired trajectory's state. Define the tracking error of the desired trajectory as $e(t) = x(t) - x_d(t)$. Subtracting the nonlinear system dynamics (1) and desired trajectory's dynamics (2), we can obtain the following tracking error dynamics:

$$\begin{aligned} \dot{e}(t) &= \dot{x}(t) - \dot{x}_d(t) \\ &= [f(x) - f_d(x_d)] + g(x)u + k(x)\omega \end{aligned} \tag{3}$$

To track the desired trajectory within certain constraints, PPC method is utilized in the next subsection to transform the tracking error dynamics into prescribed performance tracking error dynamics.

B. PPC-based tracking control

In this subsection, a transformation of the tracking error dynamics is established to achieve PPC. First, we define the performance bounds for the tracking error.

Definition 1. *A smooth decreasing function $\varrho_i(t)$ is called a performance bound if it satisfies:*

$$\begin{aligned} &\lim_{t\to\infty} \varrho_i(t) = \varrho_{i\infty} \\ &\lim_{t\to 0} \varrho_i(t) = \varrho_{i0} \end{aligned} \tag{4}$$

where ϱ_{i0}, $\varrho_{i\infty}$ are positive constants for the i-th performance bound, which satisfies $\varrho_{i0} > \varrho_{i\infty} > 0$, $(i = 1, 2, \cdots, n)$.

To ensure the tracking error satisfies specific constraints $\mathcal{X}_{l,i} \le e_i \le \mathcal{X}_{u,i}$, the performance bounds are designed as exponentially decreasing functions:

$$\varrho_i = (\varrho_{i0} - \varrho_{i\infty})e^{-\lambda_i t} + \varrho_{i\infty}, \quad i = 1, 2, \cdots, n \tag{5}$$

where $\lambda_i > 0$ is the decay rate of the i-th performance bound. The lower and upper bounds of the i-th tracking error are designed as $\mathcal{X}_{l,i} = -\zeta_{l,i}\varrho_i$ and $\mathcal{X}_{u,i} = \zeta_{u,i}\varrho_i$, where $\zeta_{l,i}, \zeta_{u,i} \in (0, \infty)$ are the user-specified parameters that determine the performance bounds with asymmetric tracking error constraints. The tracking error constrained by these bounds satisfies $\mathcal{X}_{l,i} < e_i < \mathcal{X}_{u,i}$ for $i = 1, 2, \cdots, n$. To transform the tracking error within these bounds, the following transformation is utilized:

$$\begin{cases} \xi_i = \tan\left(\dfrac{2\pi e_i - \pi\mathcal{X}_{l,i} - \pi\mathcal{X}_{u,i}}{2\mathcal{X}_{u,i} - 2\mathcal{X}_{l,i}}\right) \\ e_i = \dfrac{\mathcal{X}_{l,i} + \mathcal{X}_{u,i}}{2} + \dfrac{\mathcal{X}_{u,i} - \mathcal{X}_{l,i}}{\pi} \arctan(\xi_i) \end{cases} \tag{6}$$

where $\xi_i, i = 1, 2, \cdots, n$ is the transformed tracking error. With this transformation, when ξ_i is bounded, the original tracking error e_i will always stay within the prescribed bounds $[\mathcal{X}_{l,i}, \mathcal{X}_{u,i}]$. Take the derivative of eq. (6), the transformed dynamics of tracking error ξ_i could be obtained as:

$$\dot{\xi}_i = \frac{\partial \xi_i}{\partial e_i}\dot{e}_i + \frac{\partial \xi_i}{\partial \mathcal{X}_{l,i}}\dot{\mathcal{X}}_{l,i} + \frac{\partial \xi_i}{\partial \mathcal{X}_{u,i}}\dot{\mathcal{X}}_{u,i} = \frac{\partial \xi_i}{\partial e_i}\dot{e}_i + \Omega_i \tag{7}$$

where $\frac{\partial \xi_i}{\partial e_i} = \pi \sec^2\left(\frac{2\pi e_i - \pi\mathcal{X}_{l,i} - \pi\mathcal{X}_{u,i}}{2\mathcal{X}_{u,i} - 2\mathcal{X}_{l,i}}\right)/(2\mathcal{X}_{u,i} - 2\mathcal{X}_{l,i})$, and $\Omega_i = \frac{\partial \xi_i}{\partial \mathcal{X}_{l,i}}\dot{\mathcal{X}}_{l,i} + \frac{\partial \xi_i}{\partial \mathcal{X}_{u,i}}\dot{\mathcal{X}}_{u,i}, i = 1, 2, \cdots, n$. Then the tracking error dynamics is summarized in the following form:

$$\dot{\xi} = \mathcal{H}\dot{e} + \Omega \tag{8}$$

where $\mathcal{H} = \text{diag}([\partial\xi_1/\partial e_1, ..., \partial\xi_n/\partial e_n]) \in \mathcal{R}^n$, $\Omega = [\Omega_1, ..., \Omega_n]^\top$. Accordingly, the original system (1), dynamics of the transformed tracking error (8) and dynamics of desired trajectory (2) could be augmented as the following dynamics:

$$\begin{cases} \dot{X} = F(X) + G(X)U + K(X)\omega \\ Y = H(X) \end{cases} \tag{9}$$

where $X = [x^\top, \xi^\top, x_d^\top]^\top \in \mathbb{R}^{3\times n}$ is the augmented state of above augmented system, $U = [u^\top, u^\top, 0_{1\times m}]^\top \in \mathbb{R}^{3\times m}$ is the augmented control input, Y is the output of system performance, and dynamics F, G, K, H are defined as:

$$F = \begin{bmatrix} f(x(t)) \\ \mathcal{H}\left(f(x(t)) - f_d(x_d(t))\right) + \Omega \\ f_d(x_d(t)) \end{bmatrix}, \; K = \begin{bmatrix} k(x(t)) \\ \mathcal{H}k(x(t)) \\ 0_{n\times m} \end{bmatrix},$$

$$G = \begin{bmatrix} g(x(t)) & 0_{n\times n} & 0_{n\times m} \\ 0_{n\times n} & \mathcal{H}g(x(t)) & 0_{n\times m} \\ 0_{n\times m} & 0_{n\times m} & 0_{n\times m} \end{bmatrix}, \quad H = \begin{bmatrix} \sqrt{Q}\xi(t) \\ \sqrt{\Psi(u)} \end{bmatrix}$$

where $\Psi(u) = 2R\int_0^U \left(\mu \tanh^{-1}(\zeta_U/\mu)\right) \mathrm{d}\zeta_U$ is the saturated control input penalty term, Q and R are positive definite penalty matrices. Therefore, the problem of optimal tracking control under prescribed performance could be formulated. The objective of this paper is to design an optimal control input $U(t)$ to ensure that the tracking error $\xi(t)$ could always be kept within the performance bounds $\mathcal{X}_{l,i} \le e_i \le \mathcal{X}_{u,i}$, $i = 1, 2, \cdots, n$. To solve this problem, next section will introduce the design of the optimal tracking controller that mixes H_2 and H_∞ index via the Stackelberg game.

Remark 1. *The PPC-based tracking control is designed to maintain that the tracking error is remained in the defined performance bounds. In real-world applications of tracking control, uncertainties [25], disturbances [27], and actuator faults [47] are inevitable. Compared with the other tracking control methods, such as robust tracking control [14], [39], approximate optimal tracking control [34], and sliding mode control [48], PPC is designed to achieve tracking with pre-defined error constraints, which could achieves tracking the desired trajectory effectively and restrain the tracking error within defined performance bounds.*

III. Problem Formulation: Mixed H_2/H_∞ Optimal Tracking Control with Input Saturation

Considering the augmented dynamics system (9), a Stackelberg game-based mixed H_2/H_∞ optimal tracking controller with input saturation is designed in this section.

This article has been accepted for publication in IEEE Transactions on Automation Science and Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TASE.2025.3549114

To obtain H_2 performance feedback control law $u(x)$ under H_∞ performance disturbance, the definition of finite L_2 gain stable is given first.

Definition 2. *(Finite L_2-gain stable) Considering the nonlinear system in (9), the system is finite L_2-gain stable if there exists a positive constant γ, which satisfies that for any bounded input $\omega(t)$, the output $y(t)$ is bounded and satisfies:*

$$\int_0^\infty \|Y(t)\|^2 dt \le \gamma^2 \int_0^\infty \|\omega\|^2 d\tau \tag{10}$$

To design the optimal controller that mixes H_2 index and H_∞ index, the following quadratic form H_2 and H_∞ indexes for nonlinear system (9) are defined:

$$J_1(X_0,U,\omega) = \int_t^\infty \|Y\|^2 d\tau = \int_t^\infty \left(\xi^\top Q\xi + \Psi(U)\right) d\tau \tag{11}$$

$$\begin{aligned} J_2(X_0,U,\omega) &= \int_t^\infty \|Y\|^2 d\tau - \gamma^2\|\omega\|^2 \\ &= \int_t^\infty \left(\xi^\top Q\xi + \Psi(U) - \gamma^2\|\omega\|^2\right) d\tau \end{aligned} \tag{12}$$

where γ is the disturbance attenuation level, J_1 in eq. (11) is the performance index that reflects the H_2 performance, and J_2 in eq. (12) is the index for the H_∞ performance. With the establishment of the performance indexes, the problem of mixed H_2/H_∞ could be formulated.

Problem 1. *(Mixed H_2/H_∞ optimal tracking control) Given the nonlinear system (9), find the robust optimal tracking control input $U^*(\cdot)$ that satisifies:*

1) *Ensures the tracking error $\xi(t)$ remains in the prescribed performance bounds $e_i \in (\mathcal{X}_{l,i}, \mathcal{X}_{u,i})$, $i = 1, 2, \cdots, n$ when the initial states satisfies $e_i(0) = 0$.*
2) *The optimal tracking controller U^* minimizes the performance index J_1 in eq. (11) to achieve the H_2 performance, and the worst-case disturbance $\omega^*(t)$ maximizes the performance index J_2 in eq. (12) to achieve the H_∞ performance, that is,*

$$\begin{aligned} J_1(X_0,U^*(t),\omega^*(t)) &= \min_{U\in\Omega_U} J_1(X_0,U,\omega^*(t)) \\ J_2(X_0,U(t),\omega^*(t)) &= \max_{\omega\in\Omega_W} J_2(X_0,U(t),\omega(t)) \end{aligned}$$

3) *The closed-loop system under the control of U^* is finite L_2-gain stable, satisfying the condition in eq. (10).*

To solve the mixed H_2/H_∞ optimal tracking control problem given in Problem 1, Stackelberg game is established to derive the robust optimal tracking control input. First, the definition of the Stackelberg game is given.

Definition 3. *(Stackelberg game [17]) Given two players: the leader $\mathbb{L}$ with the performance index (11), and the follower $\mathbb{F}$ with the performance index (12). For the leader $\mathbb{L}$, the control strategy $U \in \Omega_U$ is chosen based on the information collected without the follower's control strategy. However, the follower $\mathbb{F}$ chooses the control strategy $\omega \in \Omega_\omega$ based on the information including leader's control strategy U, which means, for any strategy U chosen by the leader $\mathbb{L}$, the follower $\mathbb{F}$ will choose an optimal control strategy $\omega^*(t) = \mathbf{w}(U,t)$ to maximize the performance index J_2 in eq. (12):*

$$J_2(X_0,U(t),\mathbf{w}(U,t)) = \max_{\mathbf{w}\in\Omega_W} J_2(X_0,U(t),\omega(t))$$

where $\mathbf{w} : U \to \omega$ is a mapping. By considering the strategy $\omega^* = \mathbf{w}(U,t)$ chosen by the follower $\mathbb{F}$, the leader $\mathbb{L}$ chooses the optimal control strategy $U^*(t)$ to minimize the performance index J_1 in eq. (11):

$$J_1(X_0,U^*(t),\mathbf{w}(U^*,t)) = \min_{U\in\Omega_U} J_1(X_0,U(t),\mathbf{w}(U^*,t))$$

Then the optimal strategy U^* is called Stackelberg policy of the leader $\mathbb{L}$ and $\omega^* = \mathbf{w}(U^*,t)$ is called Stackelberg policy of the follower $\mathbb{F}$.

Remark 2. *In the Stackelberg game, the leader $\mathbb{L}$ derives the optimal tracking control input that optimizes the H_2 index J_1 in eq. (11), the follower $\mathbb{F}$ is the worst disturbance that minimizes the H_∞ performance index J_2 in eq. (12). For the sequence of the leader and follower, the follower is assumed to derive a worst disturbance input ω^* to minimize the H_∞ performance index J_2 with a regular control input U chosen by the leader. Then the leader chooses the optimal tracking control input U^* subsequently, which minimizes the H_2 index J_1 with the worst disturbance input ω^* chosen by the follower.*

Defining Stackelberg game, and the follower is going to pursue the H_∞ performance, the optimal value function J_2^* of the follower $\mathbb{F}$ could be obtained:

$$\begin{aligned} J_2^* &= \max_\omega J_2(X_0,U,\omega) \\ &= \max_\omega \int_t^\infty \left(\gamma^2\|\omega\|^2 - X^\top QX - \Psi(U)\right) d\tau \end{aligned} \tag{13}$$

Then the Hamiltonian of the follower is derived as:

$$\begin{aligned} H_\mathbb{F}\left(X,U,\omega,\nabla J_2^*\right) =& (\nabla J_2^*)^\top (F + GU + K\omega) \\ &+ \left(\gamma^2\|\omega\|^2 - X^\top QX - \Psi(U)\right) \end{aligned} \tag{14}$$

Take the derivative of Hamiltonian $H_\mathbb{F}$ to disturbance input ω, and solve the obtained derivative equation, the optimal control input of the follower $\mathbb{F}$ could be obtained as:

$$\begin{aligned} \omega^*(U) &= \arg\max_\omega H_\mathbb{F}\left(X,U,\omega,\nabla J_2^*\right) \\ &= -\frac{1}{2\gamma^2} K^\top \nabla J_2^* \end{aligned} \tag{15}$$

To obtain the optimal control input from the leader $\mathbb{L}$ subsequently, a costate λ_2 for the follower is defined as the negative gradient of Hamiltonian $H_\mathbb{F}$ to state X:

$$\begin{aligned} \dot{\lambda}_2 =& -\frac{\partial H_\mathbb{F}}{\partial X} \\ =& -\left(\frac{\partial F}{\partial X} + \frac{\partial G}{\partial X}U + \frac{\partial K}{\partial X}\omega^* + G\frac{\partial U^\top}{\partial X}\right)^\top \nabla J_2^* \\ &+ 2QX + \frac{\partial \Psi}{\partial U}\frac{\partial U}{\partial X} \end{aligned} \tag{16}$$

Taking the consideration of the follower's H_2 performance, leader $\mathbb{L}$'s optimal value function J_1^* is minimized with costate λ_2 in eq. (16) as:

$$\begin{aligned} J_1^* =& \min_U J_1(X_0, U, \omega^*) \\ =& \min_U \int_t^\infty \left(X^\top Q X + \Psi(U) + \eta^\top \dot{\lambda}_2 \right) d\tau \end{aligned} \quad (17)$$

where η denotes control input's Lagrange multiplier. Then leader's Hamiltonian $H_{\mathbb{L}}$ is constructed as the following form:

$$\begin{aligned} H_{\mathbb{L}}\left(X, U, \omega^*, \nabla J_1^*, \eta\right) =& \left(\nabla J_1^*\right)^\top \left(F + GU + K\omega\right) + \eta^\top \dot{\lambda}_2 \\ &+ \left(X^\top Q X + \Psi(U)\right) \end{aligned} \quad (18)$$

The optimal control input U^* of leader $\mathbb{L}$ is derived by minimizing Hamiltonian $H_{\mathbb{L}}$ with respect to control input U:

$$\begin{aligned} U^*(\omega^*) =& \arg\min_U H_{\mathbb{L}}\left(X, U, \omega^*, \nabla J_1^*, \eta\right) \\ =& -\mu \tanh\left(\frac{R^{-1}}{2\mu} \left(G^\top \nabla J_1^* - \nabla J_2^{*\top} \frac{\partial G}{\partial X} \eta \right) \right) \end{aligned}$$

The dynamics of the Lagrange multiplier y is derived as:

$$\begin{aligned} \dot{\eta}(t) =& -\frac{\partial H_{\mathbb{L}}}{\partial \nabla J_2^*} \\ =& -\left[\left(K \frac{\partial \omega^*}{\partial \nabla J_2^*} \right)^\top \nabla J_1^* - \sum_{i=1}^n \eta_i \left(\frac{\partial K}{\partial X_i} \frac{\partial \omega^*}{\partial \nabla J_2^*} \right)^\top \nabla J_2^* \right. \\ & - \left(\nabla F + \nabla G U + \nabla K \omega^* + G \nabla U\right) \eta] \end{aligned}$$

where ∇F, ∇G, ∇K are the gradient of F, G, K to state X.

To summarize the optimal control input of the leader and the follower, by minimizing the Hamiltonian H_1 and H_2, the optimal control policy of the leader and the follower are obtained as:

$$\begin{cases} U^* = -\mu \tanh\left(\frac{R^{-1}}{2\mu} \left(G^\top \nabla J_1^* - \nabla J_2^{*\top} \frac{\partial G}{\partial X} \eta \right) \right) \\ \omega^* = -\frac{1}{2\gamma^2} K^\top \nabla J_2^* \end{cases} \quad (19)$$

Due to the complexity of the nonlinear dynamics and the mixed H_2/H_∞ performance indexes, it is difficult to obtain the concrete expression of the optimal tracking control input of leader $\mathbb{L}$ and the follower $\mathbb{F}$ from eq. (19). To obtain the optimal control input U^*, we employed the reinforcement learning-based approximation method in the next section. An actor-critic-identifier structure is constructed to approximate the optimal value function and the optimal control policy, with uncertain drifted system parameters identified online.

IV. Design of actor-critic-identifier

An actor-critic-identifier structure is designed to solve the mixed H_2/H_∞ optimal tracking control problem through approximations in this section. First, the above-mentioned optimal value functions and control inputs are reconstructed by the actor-critic neural network (NN). Then, uncertain drift system parameters are identified online using an identifier. With the identified system parameters and the reconstructed optimal value functions and control inputs, corresponded bellman error is established. By minimizing the bellman error, the actor-critic NNs are trained to obtain the approximated optimal value functions and control inputs.

A. Approximation of value function via actor-critic

To approximate optimal value functions of leader and follower, two actor-critic NNs are developed. Optimal value functions are reconstructed as:

$$J_i^*(X) = W_{ci}^\top \varphi_{ci}(X) + \delta_{ci}(X), \ i = 1, 2 \quad (20)$$

$$\nabla J_i^*(X) = \nabla \varphi_{ci}^\top(X) W_{ci} + \nabla \delta_{ci}^\top(X), \ i = 1, 2 \quad (21)$$

where $W_{ci} \in \mathbb{R}^{n_{\varphi_{ci}} \times 1}, i = 1, 2$ are the ideal weights of the critic NNs for leader and follower, δ_{ci} and $\delta_{ai}, i = 1, 2$ are the construction errors of the actor-critic NNs. To obtain the optimal control input, the actor NNs are constructed for the approximation of optimal control inputs:

$$\begin{aligned} U^*(X) =& -\mu \tanh\left(\frac{1}{2\mu} \left(R^{-1} G^\top \left(\nabla \varphi_{a1}^\top(X) W_{a1} + \nabla \delta_{a1}^\top \right) \right. \right. \\ & \left. \left. - \left(W_{a2}^\top \nabla \varphi_{a2}(X) + \nabla \delta_{a2} \right) \frac{\partial G}{\partial X} \eta \right) \right) \end{aligned} \quad (22)$$

$$\omega^*(X) = -\frac{K^\top}{2\gamma^2} \left(\nabla \varphi_{a2}^\top(X) W_{a2} + \nabla \delta_{a2}^\top(X) \right) \quad (23)$$

where $W_{ai} \in \mathbb{R}^{n_{\varphi_{ai}} \times 1}, i = 1, 2$ are the ideal weights of the actor NNs for leader and follower. In the practice, the ideal weights W_{ci} and W_{ai} are unknown, estimation of NN weights are proposed for the approximation

$$\hat{J}_i(X) = \hat{W}_{ci}^\top \varphi_{ci}(X), \ i = 1, 2 \quad (24)$$

$$\begin{aligned} \hat{U}(X) =& -\mu \tanh\left(\frac{1}{2\mu} \left(R^{-1} G^\top \nabla \varphi_{a1}^\top(X) \hat{W}_{a1} \right. \right. \\ & \left. \left. - \hat{W}_{a2}^\top \nabla \varphi_{a2}(X) \frac{\partial G}{\partial X} \eta \right) \right) \end{aligned} \quad (25)$$

$$\hat{\omega}(X) = -\frac{K^\top}{2\gamma^2} \nabla \varphi_{a2}^\top(X) \hat{W}_{a2} \quad (26)$$

where $\hat{W}_{ci} \in \mathbb{R}^{n_\varphi \times 1}, i = 1, 2$ are the estimated weights of the critic NNs for the leader and the follower. $\hat{W}_{ai}, i = 1, 2$ are the estimated weights of the actor NNs.

By inserting the estimated value functions and control inputs into the Hamilton function, the Bellman errors under un-drifted system parameters are obtained:

$$\begin{aligned} \varepsilon_i(X, \hat{W}_{ci}, \hat{W}_{ai}) =& r(X, \hat{U}, \hat{\omega}) + \left(\nabla \varphi_{c1}^\top \hat{W}_{c1} \right)^\top \\ & \times \left(F + G\hat{U}(X) + K\hat{\omega}(X) \right) \end{aligned} \quad (27)$$

where $\varepsilon_i(x, \hat{W}_{ci}, \hat{W}_{ai})$ is the Bellman error of the i-th agent, $i = 1, 2$. The structure of the proposed mixed H_2/H_∞ prescribed-performance optimal control via Stackelberg game is shown in Fig. 1, in which critic NNs approximate optimal value functions, actor NNs approximate optimal control inputs, and the identifier online identifies the uncertain drift system parameters. Training of the actor-critic-identifier is accelerated by the concurrent learning-based gradient-descent update law with an experience stack (or replay buffer) to store the historical data samples. The outputs of the actor NNs are mixed with Stackelberg game to obtain the mixed H_2/H_∞-PPC optimal control inputs. To learn the actor-critic NNs, Bellman errors are constructed to update actor-critic NN weights. However, in uncertain systems with drifted parameters, Bellman errors

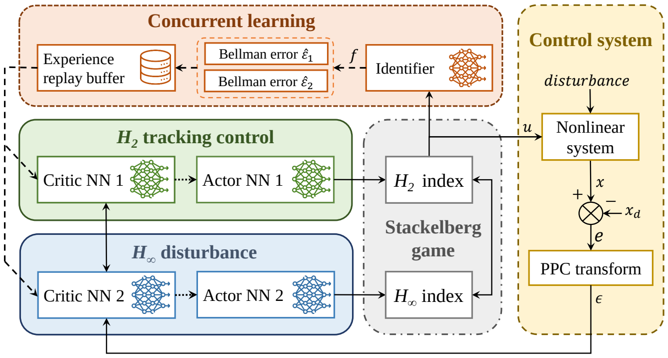


Fig. 1. Structure of the proposed mixed H_2/H_∞-PPC-ACI control scheme.

are usually not available. To obtain the Bellman errors, an identifier is designed to online identify the uncertain drift system parameters in the next subsection.

B. System identification via identifier

To identify the uncertain drift system parameters, the following identifier of the drift dynamics is designed:

$$F\left(X\right)=Y\left(X\right)\theta+\delta_\theta\left(X\right) \tag{28}$$

where $Y\in\mathbb{R}^{n\times p}$ is the basis element of the identifier and $\theta\in\mathbb{R}^p$ is the unknown drift sytem parameters, which is not available, and δ_θ is the approximation error of the identifier. To obtain the unknown drift system parameters θ, estimate the unknown parameter with an estimation $\hat{\theta}\in\mathbb{R}^p$, the estimated drift dynamics is obtained as:

$$\hat{F}\left(X\right)=Y\left(X\right)\hat{\theta} \tag{29}$$

The estimated drift dynamics $\hat{F}\left(X\right)$ is used to approximate the optimal value functions and control inputs, and update the estimated parameter $\hat{\theta}$ online by the follwing concurent learning-based gradient-descent update law:

$$\dot{\hat{\theta}}=\frac{\Gamma_\theta k_\theta}{M}\sum_{j=1}^{M}Y^\top\left(X^j\right)\left(F\left(X^j\right)-Y\left(X^j\right)\hat{\theta}\right) \tag{30}$$

where $\Gamma_\theta\in\mathbb{R}^{p\times p}$ is an positive definite matrix for tunning, k_θ is the learning rate, and M is the amount of historical data in the experience replay buffer. Updating the parameters of the identifier, the unknown drift system parameters θ are identified online, and the drift dynamics $F(X)$ is obtained. With the identified drift dynamics $F(X)$, the estimated Bellman errors with uncertain system parameters could be obtained:

$$\begin{aligned}\hat{\varepsilon}_1(X,\hat{U},\hat{\omega})=&(\nabla J_1^*)^\top(Y\left(X\right)\hat{\theta}+G\hat{U}+K\hat{\omega})\\&+\left(X^\top QX+\Psi(\hat{U})\right)+\eta^\top\dot{\lambda}_2\end{aligned} \tag{31}$$

$$\begin{aligned}\hat{\varepsilon}_2(X,\hat{U},\hat{\omega})=&(\nabla J_2^*)^\top(Y\left(X\right)\hat{\theta}+G\hat{U}+K\hat{\omega})\\&+\left(\gamma^2\|\hat{\omega}\|^2-X^\top QX-\Psi(\hat{U})\right)\end{aligned} \tag{32}$$

where $\hat{\varepsilon}_1(X,\hat{U}(X),\hat{\omega}(X))$ and $\hat{\varepsilon}_2(X,\hat{U}(X),\hat{\omega}(X))$ are the estimated Bellman errors of the leader and the follower with uncertain system parameters, respectively.

C. Online value function approximation

In this subsection, weights of actor-critic NNs are updated online by minimizing the Bellman errors. For the update of the leader agent, Then the historical stack data set of leader agent $\{\hat{U}(t),\hat{\varepsilon}_1(t),\{\hat{U}^j(t),\hat{\varepsilon}_1^j(t)\}_{j=1}^N\}$ is collected without extrapolation but stored as a stack, where $\{\hat{U}^j(t),\hat{\varepsilon}_1^j(t)\}$ is the j-th historical stored data collection. Same as the leader agent, the follower agent's trajectories data set is collected without extrapolation but stored as a a stack, i.e. $\{\hat{\omega}(t),\hat{\varepsilon}_2(t),\{\hat{\omega}^j(t),\hat{\varepsilon}_2^j(t)\}_{j=1}^N\}$. The weights of the leader-follower agents' actor-critic NNs are updated by minimizing the following squared loss function:

$$E_i=\hat{\varepsilon}_i^\top\hat{\varepsilon}_i+\sum_{k=1\cdots N}\hat{\varepsilon}_i^{k\top}\hat{\varepsilon}_i^k,\ i=1,2 \tag{33}$$

Then a concurrent learning-based gradient descent update law is utilized to train the critic NNs weights:

$$\dot{\hat{W}}_{ci}=\frac{-k_{ci,1}\hat{\varepsilon}_i\sigma_i}{\left(\sigma_i^\top\sigma_i+1\right)^2}-\sum_{k=1}^{N}\frac{k_{ci,2}\hat{\varepsilon}_i^k\sigma_i^k}{N\left(\sigma_i^{k\top}\sigma_i^k+1\right)^2},\ i=1,2 \tag{34}$$

where $k_{ci,j}>0,\ i=l,f,\ j=1,2$ are the learning rates for the critic NNs. The regression vectors $\sigma_i=\nabla\varphi_{ci}^\top(X)(F+G\hat{U}+K\hat{\omega})$, $\sigma_i^k=\nabla\varphi_{ci}^\top(X^k)(F+G\hat{U}^k+K\hat{\omega}^k)$, where X^k is the k-th historical data sample. For the update of the actor NNs, the weights are updated based on a projection-based gradient descent method to maintain numerical stability and prevent parameter divergence [49], [50]. The weights update law takes the following form:

$$\dot{\hat{W}}_{ai}=\Gamma\left(F_ik_{ai}\left(\hat{W}_{ci}-\hat{W}_{ai}\right)\right),\ i=1,2 \tag{35}$$

where $k_{ai}>0, i=1,2$ are the learning rates, $F_i\in\mathbb{R}^{n_\varphi\times n_\varphi}$, $i=1,2$ are positive definite matrices for weight updates, and $\Gamma(\cdot)$ is a projection operator that ensures the weights stay within predefined bounds. which is defined for each agent's weights $\hat{W}_{ai}, i=1,2$ as:

$$\Gamma(\dot{\hat{W}}_{ai})=\begin{cases}-F_ik_{ai}\left(\hat{W}_{ai}-\hat{W}_{ci}\right), & \text{if } \alpha_i<0\\ -F_ik_{ai}\left(\hat{W}_{ai}-\hat{W}_{ci}\right)+\beta_i, & \text{otherwise}\end{cases} \tag{36}$$

where $\gamma_i(\hat{W}_{ai})$ are smooth convex functions defining feasible weight regions, $\alpha_i=(\hat{W}_{ci}-\hat{W}_{ai})^\top\nabla\gamma_i$, $\beta_i=k_{ai}\nabla\gamma_i\nabla\gamma_i^\top/(\nabla\gamma_i^\top k_{ai}\nabla\gamma_i)F_i(\hat{W}_{ai}-\hat{W}_{ci})$ is a correction term, and $\nabla\gamma_i$ represents the gradient of γ_i. The projection mechanism ensures stable learning behavior by keeping weights within predefined feasible regions, maintaining smooth parameter adaptation, and preventing numerical instabilities and divergence through constrained updates.

Then the online learning approximation of optimal value functions and control inputs is achieved by the actor-critic structure. The detailed procedure of online learning algorithm for mixed H_2/H_∞-PPC-ACI control scheme is shown in Algorithm 1.

D. Stability analysis

In this subsection, with the utilization of the Lyapunov stability theory, closed-loop system states and the estimated errors of actor-critic NNs are proved to be ultimate uniform

Algorithm 1 Mixed H_2/H_∞-PPC-ACI Online Learning Algorithm

Initialize:
- Actor-critic-identifier NN weights: $\hat{W}_{ci}, \hat{W}_{ai}, i = 1, 2$
- Learning rates: $k_{ci,j}, k_{ai}, i = 1, 2, j = 1, 2$
- Projection matrices: $F_i, i = 1, 2$
- Performance bounds: $\varrho_{i0}, \varrho_{i\infty}, \lambda_i, \zeta_{l,i}, \zeta_{u,i}$
- Experience stacks: $\{\hat{U}(t), \hat{\varepsilon}_1(t), \{\hat{U}^j, \hat{\varepsilon}_1^j\}_{j=1}^N\} \leftarrow \emptyset$, $\{\hat{\omega}(t), \hat{\varepsilon}_2(t), \{\hat{\omega}^j, \hat{\varepsilon}_2^j\}_{j=1}^N\} \leftarrow \emptyset$

while $t < T_{end}$ **do**
  $\xi_i \leftarrow$ Transform tracking error e_i using (6)
  $X \leftarrow$ Construct augmented system state via (9)
  $\hat{U}(X) \leftarrow$ Approximate optimal control via (25)
  $\hat{\omega}(X) \leftarrow$ Approximate optimal disturbance via (26)
  $\hat{F}(X) \leftarrow$ Identify uncertain drift dynamics via (29)
  $\hat{\varepsilon}_1, \hat{\varepsilon}_2 \leftarrow$ Calculate Bellman errors via (31) and (32)
  Update experience stack:
  - Leader: $\{\hat{U}(t), \hat{\varepsilon}_1(t), \{\hat{U}^j(t), \hat{\varepsilon}_1^j(t)\}_{j=1}^N\}$
  - Follower: $\{\hat{\omega}(t), \hat{\varepsilon}_2(t), \{\hat{\omega}^j(t), \hat{\varepsilon}_2^j(t)\}_{j=1}^N\}$

  Update parameters:
  - $\hat{W}_{ci} \leftarrow$ Update critic weights via (34)
  - $\hat{W}_{ai} \leftarrow$ Update actor weights via (35)
  - $\hat{\theta} \leftarrow$ Update system parameters via (30)

  Apply control input $\hat{U}(X)$ to system
end while

bounded (UUB) under the proposed mixed H_2/H_∞ approximate optimal tracking control scheme. First, three assumptions are given here for the proof.

Assumption 1. *The following assumptions are given for the stability analysis:*

1) *On a tight set $X \in \chi \in \mathbb{R}^n$, both $F(X)$ and $G(X)$ are Lipschitz continuous with $F(0) = 0$, and $G(X)$ satisfied bounded condition $\|G(X)\| \leq G_H$ for all $X \in \chi$.*
2) *Cost matrix Q and R are bounded, such that $\underline{\lambda}_Q \leq \|Q\| \leq \bar{\lambda}_Q$, $\underline{\lambda}_R \leq \|R\| \leq \bar{\lambda}_R$, where constants $\underline{\lambda}_Q, \underline{\lambda}_R \geq 0$ and $\bar{\lambda}_q, \bar{\lambda}_R > 0$.*

Assumption 2. *Assuming that the following parameters and operators are bounded: $\|\hat{W}_{ci}\| \leq W_{Hi}$, $\|\sigma_i(X)\| \leq \sigma_{Hi}$, $\|\nabla\sigma_i(X)\| \leq \sigma_{D,Hi}$, $\|\varphi_i(X)\| \leq \varphi_{Hi}$, $\|\nabla\varphi_i(X)\| \leq \varphi_{D,Hi}$, $\|\delta_i(X)\| \leq \delta_{Hi}$, $\|\nabla\delta_i(X)\| \leq \delta_{D,Hi}$,*

Assumption 3. *Assuming that the online collected and extrapolated data set for the weights update law satisfies the following excitation condition for the i-th agent $(i = 1, 2)$:*

$$\Lambda_{1,i}\mathcal{I}_{m,i} \leqslant \int_t^{t+T} \left(\sigma_i(\tau)\sigma_i(\tau)^\top/\rho_i(\tau)\right) \mathrm{d}\tau,$$

$$\Lambda_{2,i}\mathcal{I}_{m,i} \leqslant \inf_{t \in \mathbf{R}_{t \geq t_0}} \frac{1}{N}\left(\sum_{k=1}^N \sigma_i^k(t)\sigma_i^k(t)^\top/\rho_i^k(t)\right),$$

$$\Lambda_{3,i}\mathcal{I}_{m,i} \leqslant \int_t^{t+T} \frac{1}{N}\left(\sum_{k=1}^N \sigma_i^k(\tau)\sigma_i^k(\tau)^\top/\rho_i^k(\tau)\right) \mathrm{d}\tau,$$

where $\rho_i(t) = \left(\sigma_i^\top\sigma_i + 1\right)^2$, $\rho_i^k(t) = \left(\sigma_i^{k\top}\sigma_i^k + 1\right)^2$, $\mathcal{I}_{m,i}$ is the identity matrix of size m, and at least one of the non-negative constants $\Lambda_{1,i}$,$\Lambda_{2,i}$,$\Lambda_{3,i}$ is positive.

Based on the design of the controller (25) and disturbance (26), the following inequality could be obtained:

$$\left\|U^*(X) - \hat{U}(X)\right\|^2 \leq \Sigma_1\tilde{W}_{a1}^\top\tilde{W}_{a1} + \Pi_1 \tag{37}$$

$$\|\omega^*(X) - \hat{\omega}(X)\|^2 \leq \Sigma_2\tilde{W}_{a2}^\top\tilde{W}_{a2} + \Pi_2 \tag{38}$$

where Σ_i is a upper bound related with $\varphi_{H,i}$, $\varphi_{D,Hi}$, σ_{Hi}, $\sigma_{D,Hi}$, Π_{u_i} is a upper bound related with $\delta_{D,Hi}$. The stability analysis of closed-loop system state and the estimation errors of actor-critic NNs is given in the following theoretical result.

Theorem 1. *Considering the augmented system dynamics (9) and the proposed mixed H_2/H_∞ prescribed-performance approximate optimal tracking control scheme in Algorithm 1, Assumption 1,2 and 3 are satisfied, The actor-critic NNs are updated by the adaptive update law (34) and (35). The control input and disturbance are estimated by (25) and (26). Then the close-loop system states X and weights errors $[\tilde{W}_{c1}^\top, \tilde{W}_{c2}^\top, \tilde{W}_{a1}^\top, \tilde{W}_{a2}^\top]^\top$ will be UUB provided that:*

$$\|Z\| \geq \sqrt{\frac{\Phi_{\text{res}}}{\lambda_{\min}(\mathcal{H})}} \tag{39}$$

where $Z = \left[X^\top, \tilde{W}_{c1}^\top, \tilde{W}_{c2}^\top, \tilde{W}_{a1}^\top, \tilde{W}_{a2}^\top\right]^\top$.

Proof. Theoretical result of Theorem 1 is analyzed utilizing the Lyapunov stability theory. Choosing the Lyapunov function in the following form:

$$\mathcal{V} = \sum_{i=1}^2 \left(J_i^* + \frac{1}{2}\tilde{W}_{ci}^\top\tilde{W}_{ci} + \frac{1}{2}\tilde{W}_{ai}^\top\tilde{W}_{ai}\right) \tag{40}$$

To simplify the analysis, the approximated Hamiltonian error $\hat{\varepsilon}_i$, $i = 1, 2$, or Bellman error, is abbreviated to the following form:

$$\hat{\varepsilon}_1 = -\sigma_1^\top\tilde{W}_{c1} + \frac{1}{4}\tilde{W}_{a1}G_\sigma\tilde{W}_{a1} + \Delta_1(X) + \xi_{H1}, \tag{41a}$$

$$\hat{\varepsilon}_2 = -\sigma_2^\top\tilde{W}_{c2} + \frac{1}{4}\left(\tilde{W}_{a2}K_\sigma\tilde{W}_{a2} - \tilde{W}_{a1}G_\sigma\tilde{W}_{a1}\right) + \Delta_2(X) + \xi_{H2}, \tag{41b}$$

$$\hat{\varepsilon}_1^k = -(\sigma_1^k)^\top\tilde{W}_{c1} + \frac{1}{4}\tilde{W}_{a1}G_\sigma^k\tilde{W}_{a1} + \Delta_1^k(X), \tag{41c}$$

$$\hat{\varepsilon}_2^k = -(\sigma_2^k)^\top\tilde{W}_{c2} + \frac{1}{4}\left(\tilde{W}_{a2}K_\sigma^k\tilde{W}_{a2} - \tilde{W}_{a1}G_\sigma^k\tilde{W}_{a1}\right) + \Delta_2^k(X) \tag{41d}$$

where the $G_\sigma(X) = \nabla\varphi_{a1}^\top G(X)R_1^{-1}G^\top\nabla\varphi_{a1}^\top(X)$, $K_\sigma(X) = \nabla\varphi_{a2}^\top(X)KR_2^{-1}K^\top\nabla\varphi_{a1}^\top(X)/\left(2\gamma^2\right)$, $G_\sigma^k = G_\sigma(X^K)$, $K_\sigma^k = K_\sigma(X^K)$, and $\Delta, \Delta^k : \mathbb{R}^{n_i} \to \mathbb{R}$ are uniformly bounded on χ, $\overline{\|\Delta\|}$ and $\overline{\|\Delta^k\|}$ decrease as $\overline{\|\nabla\delta\|}$ and $\overline{\|\nabla W\|}$ decrease.

Then the derivative of the Lyapunov function $\mathcal{V}$ is

$$\dot{\mathcal{V}} = \sum_{i=1}^2 \left[\nabla J_i^*\left(F + GU + K\omega\right) + \tilde{W}_{ci}^\top\dot{\tilde{W}}_{ci}^\top + \tilde{W}_{ai}^\top\dot{\tilde{W}}_{ai}^\top\right] \tag{42}$$

Substituting the $(\nabla J_i^*)^\top F(X)$ term from (18) and (14) into (42), and employing the Bellman errors from (41), the derivative can be rewritten as:

$$
\begin{aligned}
\dot{\mathcal{V}} &= -X^{\top}\left(Q_1 - Q_2\right)X - \Psi(U) - \gamma^2\|\omega\|^2 - \eta^{\top}\dot{\lambda}_2 \\
&- \tilde{W}_{c1}^{\top}\left(-k_{c1,1}\frac{\sigma_1}{\rho_1}\left(-\sigma_1^{\top}\tilde{W}_{c1} + \frac{1}{4}\tilde{W}_{a1}^{\top}G_{\sigma}\tilde{W}_{a1} + \Delta_1\right)\right) \\
&- \tilde{W}_{c2}^{\top}\left(-k_{c2,1}\frac{\sigma_2}{\rho_2}\left(-\sigma_2^{\top}\tilde{W}_{c2} + \frac{1}{4}\tilde{W}_{a2}K_{\sigma}\tilde{W}_{a2}\right.\right. \\
&\left.\left.-\frac{1}{4}\tilde{W}_{a1}G_{\sigma}\tilde{W}_{a1} + \Delta_2\right)\right) + \tilde{W}_{a1}^{\top}\left(-k_{a1}F_1\left(\hat{W}_{a1} - \hat{W}_{c1}\right)\right) \\
&+ \tilde{W}_{a2}^{\top}\left(-k_{a2}F_2\left(\hat{W}_{a2} - \hat{W}_{c2}\right)\right) \\
&- \tilde{W}_{c1}^{\top}\left(-\frac{k_{c1,2}}{N}\sum_{k=1}^{N}\frac{\sigma_1^k}{\rho_1^k}\frac{1}{4}\tilde{W}_{a1}^{\top}G_{\sigma}^k\tilde{W}_{a1}\right) \\
&- \tilde{W}_{c2}^{\top}\left(-\frac{k_{c2,2}}{N}\sum_{k=1}^{N}\frac{\sigma_2^k}{\rho_2^k}\frac{1}{4}\left(\tilde{W}_{a2}^{\top}K_{\sigma}^k\tilde{W}_{a2} - \tilde{W}_{a1}^{\top}G_{\sigma}^k\tilde{W}_{a1}\right)\right) \\
&-\tilde{W}_{c1}^{\top}\left(-\frac{k_{c12}}{N}\sum_{k=1}^{N}\frac{\sigma_1^k}{\rho_1^k}\left(-(\sigma_1^k)^{\top}\tilde{W}_{a1} + \Delta_1^k\right)\right) \\
&-\tilde{W}_{c2}^{\top}\left(-\frac{k_{c22}}{N}\sum_{k=1}^{N}\frac{\sigma_2^k}{\rho_2^k}\left(-(\sigma_2^k)^{\top}\tilde{W}_{a2} + \Delta_2^k\right)\right)
\end{aligned} \tag{43}
$$

Substitute the inequalities (37) and (38) and employing Young's inequality and conditions from assumptions 1-3, the derivative $\dot{\mathcal{V}}$ could be rewritten as:

$$
\begin{aligned}
\dot{\mathcal{V}} \leq & -Z^T\begin{bmatrix} h_1 & 0 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 & 0 \\ 0 & h_3 & h_4 & 0 & 0 \\ 0 & h_5 & 0 & h_6 & 0 \\ 0 & 0 & h_7 & 0 & h_8 \end{bmatrix} Z + \Phi_{\text{res}} \\
= & -Z^T\mathcal{H}Z + \Phi_{\text{res}}
\end{aligned}
$$

where $h_1 = \underline{\lambda}_{Q1} - \underline{\lambda}_{Q2}$, $h_2 = \frac{1}{2}k_{c1,1}\sigma_1\sigma_1^{\mathrm{T}} + \frac{1}{2}k_{c1,2}\Lambda_{2,1}\mathcal{I}_{m,1}$, $h_3 = (k_{c1,1} + k_{c2,1})\sigma_1(t)\sigma_2^{\mathrm{T}}(t)$, $h_4 = \frac{1}{2}k_{c2,1}\sigma_2(t)\sigma_2^{\mathrm{T}}(t) + \frac{1}{2}k_{c2,2}\Lambda_{2,2}\mathcal{I}_{m,2}$, $h_5 = -F_1\mathcal{I}_{m,1}$, $h_6 = F_1\mathcal{I}_{m,1} - \bar{\lambda}_{R,1}\Sigma_{u_1}\mathcal{I}_{m,1}$, $h_7 = -F_2\mathcal{I}_{m,2}$, $h_8 = F_2\mathcal{I}_{m,2} + \gamma^2\Sigma_{u_2}\mathcal{I}_{m,2}$, and

$$
\begin{aligned}
\Phi_{\text{res}} = & \frac{1}{2}k_{c1,1}\left(\frac{1}{4}\tilde{W}_{a1}^{T}G_{\sigma}\tilde{W}_{a1} + \xi_{H1} + \Delta_1\right)^2 + \gamma^2\Pi_{u_2} \\
& + \frac{1}{2}k_{c2,1}\left(\frac{1}{4}\tilde{W}_{a2}^{T}K_{\sigma}\tilde{W}_{a2} - \frac{1}{4}\tilde{W}_{a1}^{T}G_{\sigma}\tilde{W}_{a1} + \Delta_2\right)^2 \\
& + \frac{1}{2}k_{c1,2}\left(\frac{1}{4}\tilde{W}_{a1}^{T}G_{\sigma,k}\tilde{W}_{a1} + \Delta_1^k\right)^2 + \bar{\lambda}_{R,1}\Pi_{u_1} \\
& + \frac{1}{2}k_{c2,2}\left(\frac{1}{4}\tilde{W}_{a2}^{T}K_{\sigma,k}\tilde{W}_{a2} - \frac{1}{4}\tilde{W}_{a1}^{T}G_{\sigma,k}\tilde{W}_{a1} + \Delta_2^k\right)^2
\end{aligned}
$$

When a suitable positive definite matrix $\mathcal{H}$ is chosen, the closed-loop system state X and the estimation errors of actor-critic NNs $\left[\tilde{W}_{c1}^{\top}, \tilde{W}_{a1}^{\top}, \tilde{W}_{c2}^{\top}, \tilde{W}_{a2}^{\top}\right]^{\top}$ will end up being UUB when the condition (39) is satisfied. The proof is completed. □

V. SIMULATIONS

In this section, two simulation cases of the proposed mixed H_2/H_∞-PPC-ACI control scheme is provided.

A. Case 1: Verification of the PPC mechanism

Simulation setup: In this case, performance of the proposed mixed H_2/H_∞-PPC-ACI scheme is demonstrated by an uncertain drift nonlinear system with the following dynamics:

$$
f = \begin{bmatrix} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_1 & x_2(\cos(2x_1)+2) \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \tag{44}
$$

$$
g = \begin{bmatrix} \sin(2x_1+1)+2 & 0 \\ 0 & \cos(2x_1)+2 \end{bmatrix} \tag{45}
$$

where $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]^{\top}$ is the drifted parameter. Its actual value is selected as $\theta = [-1, 1, -0.5, -0.5]^{\top}$. The basis of NNs if selected as $\phi = \left[\xi_1^2, \xi_1\xi_2, \xi_2^2, \xi_1^2\xi_2, \xi_1\xi_2^2, \xi_1^2\xi_2^2\right]$, The initial weights W_{cij} and W_{aij} are all set to 1. The control objective is to track the reference signal $\dot{x}_d(t) = [-1, 1; -2, 1]x_d(t)$ with the initial condition $x_d(0) = [1, 1.5]^{\top}$. Simulations are conducted using MATLAB R2023b Simulink on a PC equipped with an Intel Core i3-12100F CPU (3.3 GHz) and 24 GB of RAM. The ODEs are solved using the fourth-order Runge-Kutta method with a fixed step size of $T = 0.001$ s. The simulation runs for a total duration of $t_{\text{end}} = 5$ s. The detailed simulation parameters are shown in Table I.

TABLE I
PARAMETERS OF SIMULATION CASE 1 AND CASE 2.

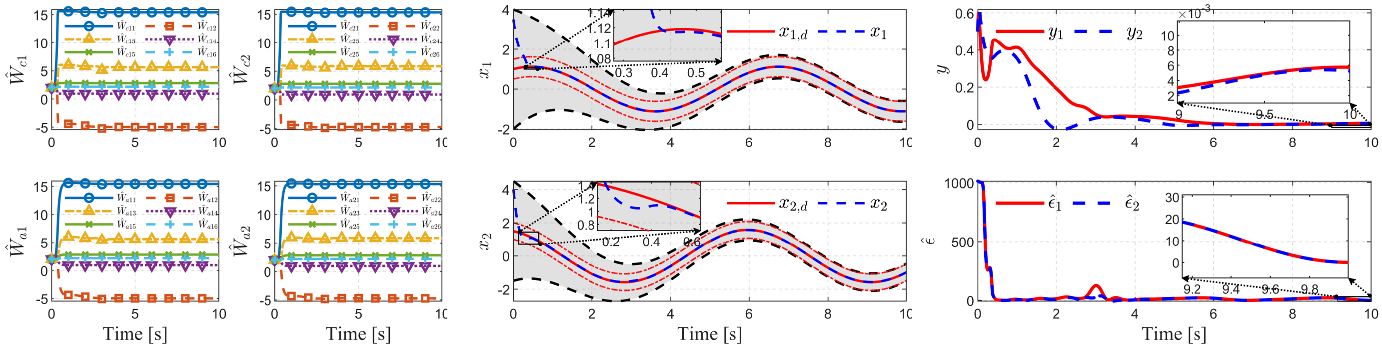
Mixed H_2/H_∞	**PPC transform**	**Update law**
$R_1 = R_2 = \mathcal{I}_2$	$\varrho_{i0} = 0.6$	$k_{1,c1} = k_{2,c1} = 0.5$
$Q_1 = \mathcal{I}_3$	$\varrho_{i\infty} = 0.01$	$k_{1,c2} = k_{2,c2} = 0.1$
$Q_2 = 20\mathcal{I}_2$	$\lambda = 0.9$	$k_{1,a} = k_{2,a} = 1$
$\mu_{sat} = 0.5$	$\zeta_{l,i} = \zeta_{u,i} = 1$	$F_1 = \mathcal{I}_3, F_2 = \mathcal{I}_6$

Results: The main results of this case are shown in Fig. 2-3. Revolution of the actor-critic NNs weights is shown in Fig 2(a). Tracking control's performance is illustrated in Fig 2(b), which shows that the system states are constrained within the pre-defined performance bounds. The value of the lagrange multiplier y and the Bellman error are shown in Fig 2(c). Fig 3 shows the estimation of the uncertain drift system parameters.

To further test the effectiveness of proposed mixed H_2/H_∞-PPC-ACI scheme, comparison simulations are tested with: 1. approximate optimal controller (AOC); 2. approximate optimal controller with fixed weights (AOC-fixed); 3. prescribed-performance controller with fixed weights (PPC-fixed); 4. the proposed mixed H_2/H_∞-PPC-ACI (PPC). As shown in Fig 5(a), the tracking error of the proposed scheme is smaller than the other controllers, which is always maintained within the prescribed performance bounds. Comparison of the cumulated cost is shown in Fig 4, which indicates that the proposed scheme take the least resources to achieve the tracking control. The detailed controller performance comparison results are shown in Table II.

B. Case 2: Verification of the mixed H_2/H_∞ mechanism

To verify the anti-disturbance effectiveness of proposed mixed H_2/H_∞ mechanism, experiments are conducted with



(a) Revolution weights of the neural networks. (b) The tracking performance of the system. (c) Value of function y and the Bellman error.

Fig. 2. Simulation results of the mixed H_2/H_∞-PPC-ACI scheme.

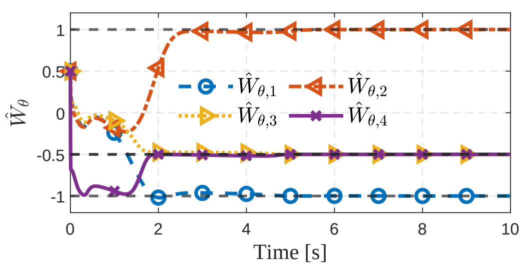


Fig. 3. The estimation of the uncertain drift system parameters.

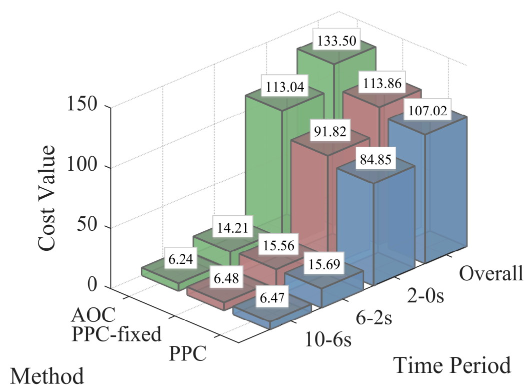


Fig. 4. Comparison of costs with different methods in several periods.

TABLE II
QUANTITATIVE PERFORMANCE COMPARISON OF CONTROL METHODS

Performance Metrics	Proposed	AOC	AOC-fixed	PPC-fixed
Overall Cost	**107.02***	133.50	134.66	113.86
Initial Phase (0-2s)	**84.85***	113.04	113.87	91.82
Middle Phase (2-6s)	15.69	**14.21***	14.50	15.56
Final Phase (6-10s)	6.47	**6.24***	6.28	6.48
Control Cost	**106.88***	133.06	134.24	113.73

*Best performance for each metric

the same system dynamics (44) and (45) from the previous case. H_∞ approximate optimal controller and H_2 approximate optimal controller are designed for comparison.

The disturbance imposed on the system is formulated as:

$$d(t)=15e^{-\frac{t}{3}}\tanh(0.1t)\sum_{i=1}^{4}[\sin(2it)+\cos((2i+1)t)] \quad (46)$$

The comparison results are shown in Fig 5(b)-5(c). The tracking performance of the system under disturbance (46) is shown in Fig 5(b), in which the proposed scheme performs faster convergence than the H_∞ method. Fig 5(c) demonstrates the comparison of normalized tracking errors from H_2/H_∞ and H_∞ under large disturbance inputs, which shows that the proposed scheme achieves smaller tracking errors under the same disturbance. The results show that the proposed control scheme performs great in the presence of disturbances.

VI. HARDWARE EXPERIMENTS

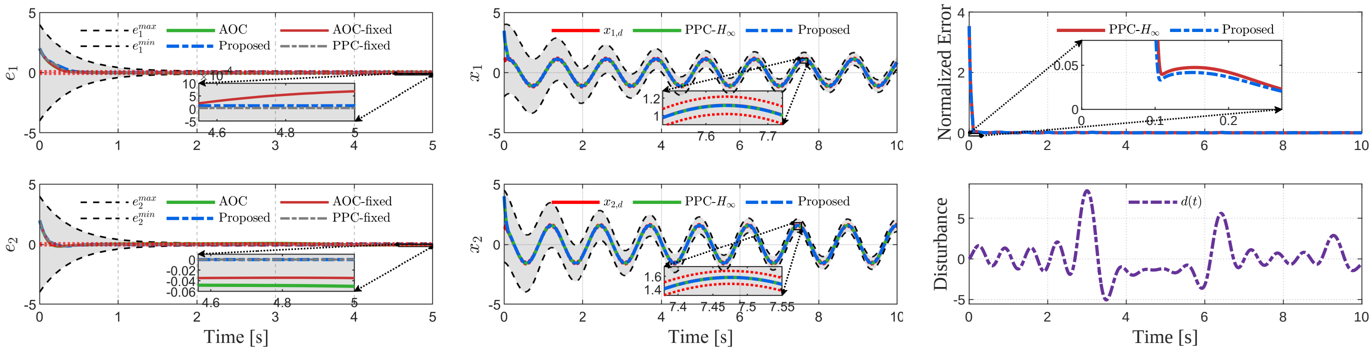
In this section, a quadcopter-based physical experiments are conducted to further verify the effectiveness of the proposed mixed H_2/H_∞-PPC-ACI control scheme.

A. Experiment Setup

The experiment is performed on a quadcopter tracking control case. and the follower quadcopter is an X150 quadcopter equipped with an RK3566 processor and 4-GB RAM. The real-time position of the quadcopter are obtained by an 8-cameras OptiTrack motion capture system. The developed controller (25) is updated and calculated online by a workstation computer equipped with an Intel i7-12700 processor (@3.60 GHz) and 32-GB RAM. The control frequency is selected as 30 Hz, and the fixed time step is $\Delta t = 1/30\ s$. The control input is formulated as a velocity command transmit to the quadcopter through 5GHz WI-FI channel. The detailed hardware equipment used in this experiment is shown in Fig. 6. To reduce the computation of online learning Algorithm 1 and accelerate the convergence of the neural network, the state-following neural network from [44], [45] is utilized for the actor-critic structure. The initial weights of the leader's critic NN are set as $\hat{W}_1(i) = 10,\ i = 1, 2, 3$. The initial weights of the follower's critic NN are chosen randomly from a uniform distribution in the range of $\hat{W}_2(i) \in [0.5,\ 2.5],\ i = 1, 2, 3$.

Note that due to the wind, the aerodynamic forces, and the sensor noise, there exist unknown disturbances in the real-world quadcopter tracking control. To verify the prescribed performance of the proposed control scheme, the desired trajectory is designed as a circle with a radius of $r = 1.5\ m$ and a period of $T = 63$ seconds, and a comparison experiment with the approximate optimal controller (AOC) is conducted.

This article has been accepted for publication in IEEE Transactions on Automation Science and Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TASE.2025.3549114



(a) Comparison of tracking error (b) Comparison of system states (c) H_2/H_∞ performance index and disturbance

Fig. 5. Performance comparison of the proposed mixed H_2/H_∞-PPC-ACI scheme and other controllers.

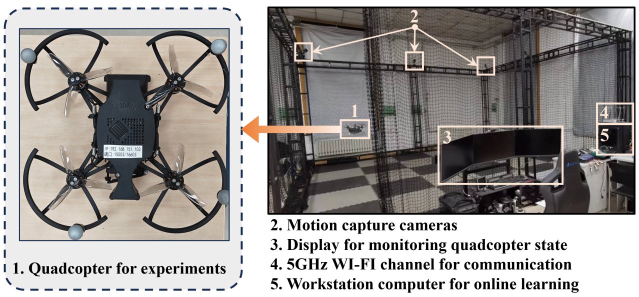


Fig. 6. Quadcopter and motion capture system for hardware experiment.

B. Experiment Results

The experimental results are presented in Figs. 7-9. Fig. 7 illustrates the key phases of autonomous navigation, demonstrating precise trajectory tracking capabilities. The planar (X-Y) position tracking shown in Fig. 8(a) confirms that the quadcopter maintains motion within the prescribed performance bounds, with maximum deviation of 0.25m. Fig. 8(b) presents a comparative analysis between the proposed scheme and AOC, where our approach achieves 88.08% reduction in mean squared error (MSE) for the X-axis and 88.33% reduction for the Y-axis. The critic neural network weight evolution in Fig. 8(c) demonstrates stable online learning convergence after 85s. The three-dimensional trajectory visualization in Fig. 9 further validates the tracking performance, with RMS position error below 0.1m throughout the flight test.

The effectiveness of the proposed control scheme is further evaluated by calculating various performance metrics:

- **Tracking Performance**
  - Mean squared error (MSE) for both X and Y axes
  - Total MSE for overall tracking accuracy
  - Maximum absolute tracking error
- **Control Input Efficiency**
  - Total energy consumption
  - Mean control input values for X and Y axes
  - Peak control input magnitude
- **Overall Performance**
  - Root mean square (RMS) values for X and Y axes
  - Total RMS across both axes
  - Individual cost metrics for X and Y axes
  - Aggregate total control cost

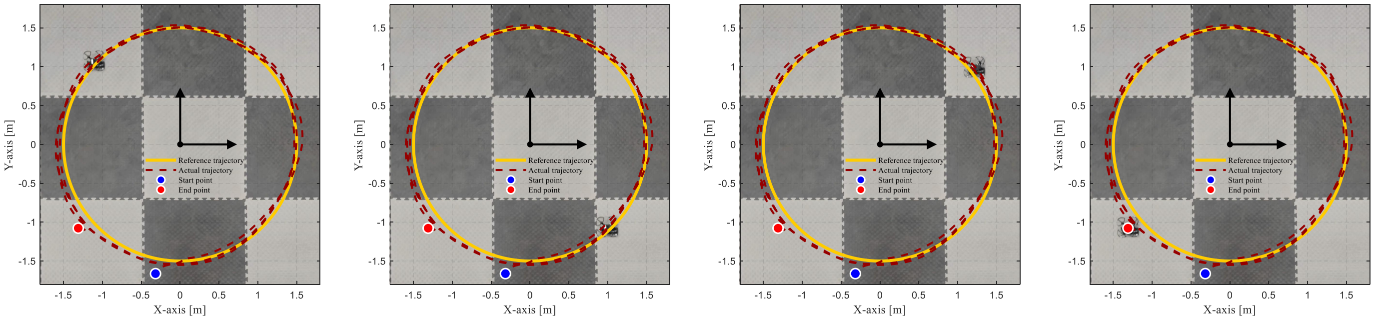
The detailed performance comparison results are shown in Table III. The proposed scheme has a better tracking performance and anti-disturbance performance than the approximate optimal controller.

TABLE III
PERFORMANCE COMPARISON BETWEEN AOC AND PROPOSED CONTROLLER

Metrics	AOC	Proposed PPC
Tracking Performance		
MSE (X-axis)	0.0688	**0.0082** ↓
MSE (Y-axis)	0.0644	**0.0075** ↓
Total MSE	0.0666	**0.0078** ↓
Max Error	0.5324	**0.2540** ↓
Control Input		
Total Energy	5.5964	**5.3174** ↓
Mean Input (X)	0.1001	0.1054
Mean Input (Y)	0.1016	**0.0995** ↓
Peak Input	**0.3299** ↓	0.3445
Overall Performance		
RMS (X)	**0.1193** ↓	0.1241
RMS (Y)	0.1253	**0.1142** ↓
Total RMS	0.1224	**0.1193** ↓
Cost (X)	**2.6598** ↓	2.8789
Cost (Y)	2.9352	**2.4380** ↓
Total Cost	5.5951	**5.3169** ↓

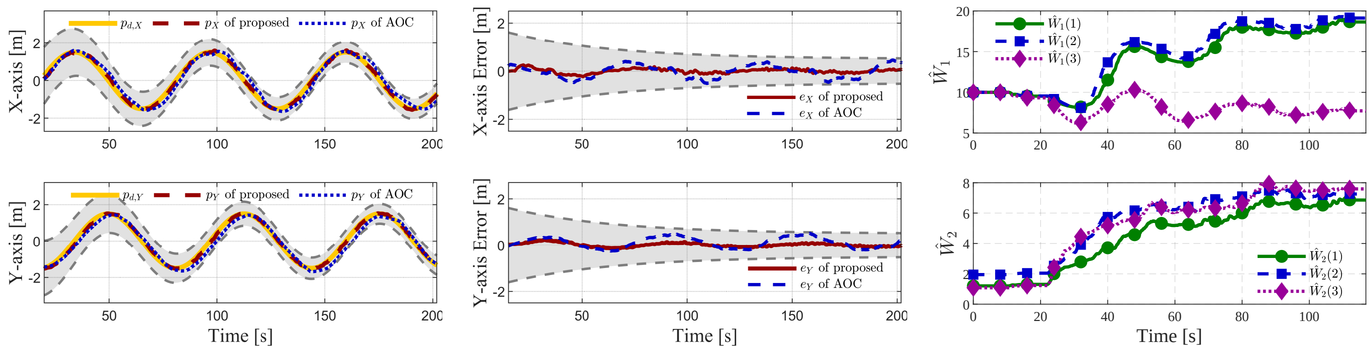
VII. Conclusion

To achieve the tracking control of nonlinear systems with unknown disturbances and system uncertainties, a novel mixed H_2/H_∞ prescribed performance approximate optimal tracking control scheme is proposed in this paper. The robust optimal controller is designed by optimizing the H_2 and the H_∞ performance index in a Stackelberg game framework. The designed optimal tracking controller is then approximated by an actor-critic-identifier structure with prescribed performance bounds. The stability of the closed-loop system is analyzed using the Lyapunov theory. The effectiveness of the proposed scheme is verified by two simulations and a quadcopter



(a) Northwest flight ($t = 57$s) (b) Southeast flight ($t = 90$s) (c) Northeast flight ($t = 104$s) (d) Landing ($t = 201$s)

Fig. 7. Key phases of quadcopter autonomous navigation and trajectory tracking experiment.



(a) Planar position tracking (X-Y coordinates) (b) Position tracking errors analysis (c) Critic neural network weight evolution

Fig. 8. Experimental results showing quadcopter control performance and learning dynamics.

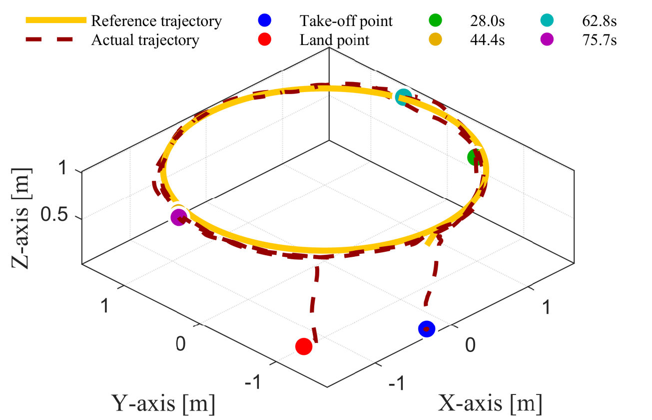


Fig. 9. Three-dimensional visualization of quadcopter autonomous navigation trajectory showing position tracking performance.

hardware experiment. The results show that the proposed scheme has a better tracking performance and anti-disturbance performance than approximate optimal controller. For future research, we will extend the proposed control scheme to stochastic systems and multi-agent systems.

References

[1] Y. Zhu, D. Zhao, and Z. Zhong, "Adaptive Optimal Control of Heterogeneous CACC System With Uncertain Dynamics," *IEEE Transactions on Control Systems Technology*, vol. 27, pp. 1772–1779, July 2019.

[2] H. Kebriaei and L. Iannelli, "Discrete-Time Robust Hierarchical Linear-Quadratic Dynamic Games," *IEEE Transactions on Automatic Control*, vol. 63, pp. 902–909, Mar. 2018.

[3] P. Tooranjipour and B. Kiumarsi, "Designing Safety Certificates for H_∞ Control of Unknown Linear Systems," *IEEE Control Systems Letters*, vol. 6, pp. 2990–2995, 2022.

[4] H. Cao, Y. Li, C. Liu, and S. Zhao, "ESO-Based Robust and High-Precision Tracking Control for Aerial Manipulation," *IEEE Transactions on Automation Science and Engineering*, vol. 21, pp. 2139–2155, Apr. 2024.

[5] J. Wang, D. Wang, H. Yan, and H. Shen, "Composite Antidisturbance \mathcal H_\infty Control for Hidden Markov Jump Systems With Multi-Sensor Against Replay Attacks," *IEEE Transactions on Automatic Control*, vol. 69, pp. 1760–1766, Mar. 2024.

[6] J. Zhang, D. Yang, H. Zhang, Y. Wang, and B. Zhou, "Dynamic Event-Based Tracking Control of Boiler Turbine Systems With Guaranteed Performance," *IEEE Transactions on Automation Science and Engineering*, pp. 1–11, 2024.

[7] Z. Zhang, Y. Shi, and W. Yan, "A Novel Attitude-Tracking Control for Spacecraft Networks With Input Delays," *IEEE Transactions on Control Systems Technology*, vol. 29, pp. 1035–1047, May 2021.

[8] J. Wang, J. Wu, H. Shen, J. Cao, and L. Rutkowski, "Fuzzy H∞ Control of Discrete-Time Nonlinear Markov Jump Systems via a Novel Hybrid Reinforcement Q-Learning Method," *IEEE Transactions on Cybernetics*, vol. 53, pp. 7380–7391, Nov. 2023.

[9] D. Huang, F. Xiao, Q. Feng, and C. Zhang, "$L2$ gain tracking control of linear completely unknown discrete-time networked control systems with dropout," *International Journal of Robust and Nonlinear Control*, p. rnc.7451, May 2024.

[10] J. Liu, H. Shen, J. Wang, J. Cao, and L. Rutkowski, "H_{infty} Control for Interconnected Systems With Unknown System Dynamics: A Two-Stage Reinforcement Learning Method," *IEEE Transactions on Automation Science and Engineering*, pp. 1–10, 2024.

[11] R. Muradore and G. Picci, "Mixed control: The discrete-time case," *Systems & Control Letters*, vol. 54, pp. 1–13, Jan. 2005.

[12] R. Yang, L. Xie, and C. Zhang, "H_2 and mixed H_2/H_∞ control of two-dimensional systems in Roesser model," *Automatica*, vol. 42, pp. 1507–1514, Sept. 2006.

[13] C. W. Scherer, "Mixed H_2/H_∞ control for time-varying and linear parametrically-varying systems," *International Journal of Robust and Nonlinear Control*, vol. 6, no. 9-10, pp. 929–952, 1996.

[14] P. Shukla, L. An, A. Chakrabortty, and A. Duel-Hallen, "A Robust Stackelberg Game for Cyber-Security Investment in Networked Control Systems," *IEEE Transactions on Control Systems Technology*, vol. 31, pp. 856–871, Mar. 2023.

[15] X. Na and D. J. Cole, "Application of Open-Loop Stackelberg Equilibrium to Modeling a Driver's Interaction with Vehicle Active Steering Control in Obstacle Avoidance," *IEEE Transactions on Human-Machine Systems*, vol. 47, pp. 673–685, Oct. 2017.

[16] B. Chen, G. Hu, W.-A. Zhang, and L. Yu, "Distributed Mixed H_2/H_ınfty Fusion Estimation With Limited Communication Capacity," *IEEE Transactions on Automatic Control*, vol. 61, pp. 805–810, Mar. 2016.

[17] J. Huang, S. Wang, and Z. Wu, "Robust Stackelberg Differential Game With Model Uncertainty," *IEEE Transactions on Automatic Control*, vol. 67, pp. 3363–3380, July 2022.

[18] Y. Yang, M. Mazouchi, and H. Modares, "Hamiltonian-driven adaptive dynamic programming for mixed H_2/H_∞ performance using sum-of-squares," *International Journal of Robust and Nonlinear Control*, vol. 31, pp. 1941–1963, Apr. 2021.

[19] Z. Ming, H. Zhang, X. Tong, and Y. Yan, "Mixed H_2/H_∞ Control With Dynamic Event-Triggered Mechanism for Partially Unknown Nonlinear Stochastic Systems," *IEEE Transactions on Automation Science and Engineering*, vol. 20, pp. 1934–1944, July 2023.

[20] Z. Ming, H. Zhang, Q. Li, and X. Tong, "Mixed H_2/H_∞ Control for Nonlinear Stochastic Systems With Cooperative and Non-Cooperative Differential Game," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, pp. 4874–4878, Dec. 2022.

[21] Z. Ming, H. Zhang, Y. Li, and Y. Liang, "Mixed H_2/H_∞ Control for Nonlinear Closed-Loop Stackelberg Games With Application to Power Systems," *IEEE Transactions on Automation Science and Engineering*, vol. 21, pp. 69–77, Jan. 2024.

[22] S. Yu, H. Zhang, Z. Ming, and J. Sun, "Adaptive Optimal Control via Continuous-Time Q-Learning for Stackelberg–Nash Games of Uncertain Nonlinear Systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, pp. 4461–4470, July 2024.

[23] Y. Ren, Q. Wang, and Z. Duan, "Output-feedback Q-learning for discrete-time linear H_∞ tracking control: A Stackelberg game approach," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 12, pp. 6805–6828, 2022.

[24] Y. Hu, Y. Geng, B. Wu, and D. Wang, "Model-Free Prescribed Performance Control for Spacecraft Attitude Tracking," *IEEE Transactions on Control Systems Technology*, vol. 29, pp. 165–179, Jan. 2021.

[25] Z. Shen, F. Li, X. Cao, and C. Guo, "Prescribed performance dynamic surface control for trajectory tracking of quadrotor UAV with uncertainties and input constraints," *International Journal of Control*, vol. 94, pp. 2945–2955, Nov. 2021.

[26] L. Chen, H. Liang, Y. Pan, and T. Li, "Human-in-the-Loop Consensus Tracking Control for UAV Systems via an Improved Prescribed Performance Approach," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, pp. 8380–8391, Dec. 2023.

[27] Y. Zhou, Y. Chen, L. Zhang, and C. Pan, "Distributed Finite-Time Prescribed Performance for Multiple Unmanned Aerial Vehicle With Time-Varying External Disturbance," *IEEE Internet of Things Journal*, vol. 11, pp. 16969–16980, May 2024.

[28] J. Xu, L. Wang, Y. Liu, and H. Xue, "Finite-time prescribed performance optimal attitude control for quadrotor UAV," *Applied Mathematical Modelling*, vol. 120, pp. 752–768, Aug. 2023.

[29] Y. Qin, L. Wang, Y. Liu, and L. Cao, "Finite-time optimal tracking control for a class of nonlinear systems with prescribed performance," *Asian Journal of Control*, vol. 25, no. 6, pp. 4785–4795, 2023.

[30] H. Yang, Q. Hu, H. Dong, X. Zhao, and D. Li, "Optimized Data-Driven Prescribed Performance Attitude Control for Actuator Saturated Spacecraft," *IEEE/ASME Transactions on Mechatronics*, vol. 28, pp. 1616–1626, June 2023.

[31] H. Yang, H. Dong, and X. Zhao, "ADP-Based Spacecraft Attitude Control Under Actuator Misalignment and Pointing Constraints," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 9, 2022.

[32] Y. Wu, M. Chen, H. Li, and M. Chadli, "Event-Triggered-Based Adaptive NN Cooperative Control of Six-Rotor UAVs With Finite-Time Prescribed Performance," *IEEE Transactions on Automation Science and Engineering*, vol. 21, pp. 1867–1877, Apr. 2024.

[33] L. Zhang, W.-W. Che, C. Deng, and Z.-G. Wu, "Prescribed Performance Control for Multiagent Systems via Fuzzy Adaptive Event-Triggered Strategy," *IEEE Transactions on Fuzzy Systems*, vol. 30, pp. 5078–5090, Dec. 2022.

[34] Q. Ma, P. Jin, and F. L. Lewis, "Guaranteed Cost Attitude Tracking Control for Uncertain Quadrotor Unmanned Aerial Vehicle Under Safety Constraints," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, pp. 1447–1457, June 2024.

[35] J. Tan, S. Xue, H. Cao, and S. S. Ge, "Human–AI interactive optimized shared control," *Journal of Automation and Intelligence*, p. S2949855425000024, Jan. 2025.

[36] B. Pang and Z.-P. Jiang, "Adaptive Optimal Control of Linear Periodic Systems: An Off-Policy Value Iteration Approach," *IEEE Transactions on Automatic Control*, vol. 66, pp. 888–894, Feb. 2021.

[37] Y. Jiang and Z.-P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, pp. 2699–2704, Oct. 2012.

[38] L. Cui, B. Pang, and Z.-P. Jiang, "Learning-Based Adaptive Optimal Control of Linear Time-Delay Systems: A Policy Iteration Approach," *IEEE Transactions on Automatic Control*, vol. 69, pp. 629–636, Jan. 2024.

[39] Y. Peng, Q. Chen, and W. Sun, "Reinforcement Q-Learning Algorithm for H_∞ Tracking Control of Unknown Discrete-Time Linear Systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, pp. 4109–4122, Nov. 2020.

[40] S. Xue, B. Luo, D. Liu, and Y. Gao, "Event-Triggered ADP for Tracking Control of Partially Unknown Constrained Uncertain Systems," *IEEE Transactions on Cybernetics*, vol. 52, pp. 9001–9012, Sept. 2022.

[41] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica*, vol. 50, pp. 1780–1792, July 2014.

[42] J. Tan, S. Xue, Z. Guo, H. Li, H. Cao, and B. Chen, "Data-driven optimal shared control of unmanned aerial vehicles," *Neurocomputing*, vol. 622, p. 129428, Mar. 2025.

[43] J. Tan, S. Xue, Q. Guan, T. Niu, H. Cao, and B. Chen, "Unmanned aerial-ground vehicle finite-time docking control via pursuit-evasion games," *Nonlinear Dynamics*, Mar. 2025.

[44] R. Kamalapurkar, J. A. Rosenfeld, and W. E. Dixon, "Efficient model-based reinforcement learning for approximate online optimal control," *Automatica*, vol. 74, pp. 247–258, Dec. 2016.

[45] P. Deptula, H.-Y. Chen, R. A. Licitra, J. A. Rosenfeld, and W. E. Dixon, "Approximate Optimal Motion Planning to Avoid Unknown Moving Avoidance Regions," *IEEE Transactions on Robotics*, vol. 36, pp. 414–430, Apr. 2020.

[46] H. Dong, X. Zhao, and B. Luo, "Optimal Tracking Control for Uncertain Nonlinear Systems With Prescribed Performance via Critic-Only ADP," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, pp. 561–573, Jan. 2022.

[47] X. Shao, Q. Hu, Y. Shi, and B. Jiang, "Fault-Tolerant Prescribed Performance Attitude Tracking Control for Spacecraft Under Input Saturation," *IEEE Transactions on Control Systems Technology*, vol. 28, pp. 574–582, Mar. 2020.

[48] D. Huang, T. Huang, N. Qin, Y. Li, and Y. Yang, "Finite-time control for a UAV system based on finite-time disturbance observer," *Aerospace Science and Technology*, vol. 129, p. 107825, Oct. 2022.

[49] J. Wang, W. Zhao, J. Cao, J. H. Park, and H. Shen, "Reinforcement Learning-Based Predefined-Time Tracking Control for Nonlinear Systems Under Identifier-Critic–Actor Structure," *IEEE Transactions on Cybernetics*, vol. 54, pp. 6345–6357, Nov. 2024.

[50] K. Wang, C. Mu, Z. Ni, and D. Liu, "Safe Reinforcement Learning and Adaptive Optimal Control With Applications to Obstacle Avoidance Problem," *IEEE Transactions on Automation Science and Engineering*, pp. 1–14, 2023.

Junkai Tan received the B.E. degree in electrical engineering at the School of Electrical Engineering in Xi'an Jiaotong University, Xi'an, China. He is currently working toward the M.E. degree in electrical engineering at the School of Electrical Engineering, Xi'an Jiaotong University.

His current research interest includes adaptive dynamic programming and inverse reinforcement learning.

Shuangsi Xue (M'24) received the B.E. degree in electrical engineering and automation from Hunan University, Changsha, China, in 2014 and the M.E. degree in electrical engineering from Xian Jiaotong University, Xian, China, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering at the School of Electrical Engineering, Xian Jiaotong University.

His current research interest includes consensus control and formation control of multi-agent systems.

Huan Li received the B.E. degree in electrical engineering at the School of Electrical Engineering in Xi'an Jiaotong University, Xi'an, China. She is currently working toward the M.E. degree in electrical engineering at the School of Electrical Engineering, Xi'an Jiaotong University.

Her current research interest includes consensus control of multi-agent systems and sliding-mode control.

Zihang Guo received the B.E. degree in electrical engineering at the School of Electrical Engineering in Xi'an Jiaotong University, Xi'an, China. He is currently working toward the M.E. degree in electrical engineering at the School of Electrical Engineering, Xi'an Jiaotong University.

His current research interest includes neural network and sliding mode-based path planning and tracking methods.

Hui Cao (M'11) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2000, 2004, and 2009, respectively.

He is a Professor at the School of Electrical Engineering, Xi'an Jiaotong University. He was a Postdoctoral Research Fellow at the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, from 2014 to 2015. He has authored or coauthored over 30 scientific and technical papers in recent years. His current research interest includes knowledge representation and discovery. Dr. Cao was a recipient of the Second Prize of National Technical Invention Award.

Dongyu Li (S'16-M'19) received the B.S. and Ph.D. degree from Control Science and Engineering, Harbin Institute of Technology, China, in 2016 and 2020. He was a joint Ph.D. student with the Department of Electrical and Computer Engineering, National University of Singapore from 2017 to 2019, and a research fellow with the Department of Biomedical Engineering, National University of Singapore, from 2019 to 2021.

He is currently a Professor with the School of Cyber Science and Technology, Beihang University, China. His research interests include networked system cooperation, adaptive systems, and spacecraft engineering.