Reinforcement Learning Control with Self-Regulated Prescribed Performance for Input-Saturated Systems under External Disturbances

Kaixin Li, Ke Xiao, Yu Xia, *Member*, *IEEE*, Zsófia Lendek, *Member*, *IEEE*, Radu-Emil Precup, *Fellow*, *IEEE*, and Imre J. Rudas, *Life Fellow*, *IEEE*

***Abstract*—This paper presents an optimal control scheme for a class of nonlinear systems subject to input saturation and external disturbances, achieved through the integration of reinforcement learning (RL) and a self-regulated prescribed performance control (SRPPC) algorithm. The proposed architecture employs an Identifier-Critic-Actor RL structure, implemented via interval type-2 fuzzy logic systems (IT2FLSs), to accurately approximate unknown nonlinear dynamics while optimizing control performance. To circumvent the singularity issues inherent in conventional PPC, the SRPPC strategy is developed to dynamically initialize the prescribed performance bounds (PPBs) and autonomously relax them during severe disturbances or actuator saturation, thereby maintaining the tracking error within a feasible envelope. By synergizing RL-based optimization with the SRPPC safety mechanism, the resulting RL-SRPPC controller not only minimizes operational costs but also guarantees transient safety and strict adherence to performance specifications. Numerical simulations on a one-link manipulator demonstrate the robustness and superiority of the proposed scheme compared to existing methodologies.**

***Index Terms*—Optimal control, reinforcement learning (RL), self-regulated prescribed performance control (SRPPC), input-saturation systems, external disturbances.**

I. Introduction

This work was supported in part by the National Science and Technology Major Project under Grant 2026ZD1609300; in part by the National Natural Science Foundation of China under Grant 52375039; in part by the Postdoctoral Fellowship Program of CPSF under Grant GZC20250940; in part by the Grant of the Romanian Ministry of Research, Innovation and Digitization, CNCS/CCCDI - UEFISCDI under Project ERANET-ENUAC-e-MATS, within PNCDI IV; and in part by the Project Romanian Hub for Artificial Intelligence-HRIA, Smart Growth, Digitization, and Financial Instruments Program, MySMIS under Grant 334906. (Corresponding authors: Ke Xiao and Yu Xia.)

K. Li and K. Xiao are with the State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China (e-mails: 202007021036@cqu.edu.cn; xiaoke993@cqu.edu.cn).

Yu Xia is with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: rainyx@sjtu.edu.cn).

Zsófia Lendek is with the Department of Automation, Technical University of Cluj-Napoca, 400114 Cluj-Napoca, Romania (e-mail: zsofia.lendek@aut.utcluj.ro).

Radu-Emil Precup is with the Department of Automation and Applied Informatics, Politehnica University of Timisoara, 300223 Timisoara, Romania (e-mail: radu.precup@aut.upt.ro).

Imre J. Rudas is with Antal Bejczy Center for Intelligent Robotics, Óbuda University, 1034 Budapest, Hungary. (e-mail: rudas@uni-obuda.hu)

IN advanced fields such as industrial automation [1], aerospace [2], precision manufacturing [3], and robotics [4], high-precision position control of nonlinear systems is fundamental to ensuring overall performance and operational safety. Such systems typically exhibit complex nonlinear dynamics, posing ongoing challenges for the design of high-precision controllers. Since its inception, prescribed performance control (PPC) has offered a promising technical avenue to address the aforementioned challenges, garnering substantial research interest [5], [6], [7]. The core concept involves designing a monotonically decreasing time-varying performance function (typically of exponential decay type) to predefine transient and steady-state performance boundaries—such as overshoot, convergence rate, and steady-state error band—for the system response [8]. The control objective is then transformed into ensuring that the tracking error remains confined within a "funnel" boundary constructed by this performance function [9]. However, while pursuing strict performance constraints, classical PPC exhibits two limitations [8], [9], [10]: 1) The initial boundaries cannot adapt to different error conditions, namely that once the performance parameters are set, the initial performance boundaries lack adaptability to different initial error conditions during system startup, which can prevent the controller from functioning properly; 2) Symmetric performance boundaries lack sufficient capability to handle transient performance, meaning that efforts to increase the error convergence speed may also induce significant overshoot or even system oscillation.

To address these issues, extensive research has been conducted. In [11], an approach mapping the initial performance boundary to infinity was introduced to accommodate arbitrary initial errors. However, this infinite initial boundary sacrifices part of the transient performance and may cause input oscillation. In [12], [13], [14], Zhao et al. developed a unified PPC framework that improves both convergence speed and transient overshoot. Nevertheless, this framework does not account for convergence time, meaning that the system error might take an infinitely long duration to converge to the prescribed boundary—an impractical outcome. To remedy this, the work [15] proposed a fixed-time asymmetric PPC technique that directly prescribes the convergence time while eliminating initial feasibility conditions. Addressing high-order nonlinear systems with state constraints, Yao et al.

introduced a novel prescribed-time control methodology that guarantees prescribed-time stability [16].

However, in practical applications such as robotic manipulators, unmanned vehicles, and industrial servo systems, two types of constraints are inevitable: external bounded disturbances arising from measurement noise, model uncertainties, and environmental interactions [17], and input saturation resulting from physical actuator limits (e.g., maximum torque, voltage, or current) [18]. Disturbances can significantly degrade tracking accuracy and system stability, while input saturation leads to a mismatch between the controller's commanded signal and the actual actuator output, potentially causing performance deterioration, oscillations, or even instability. In conventional PPC [8], [9], [10], [11], [12], [13], [14], [15], [16], [19], the performance boundaries are usually fixed or monotonically decaying throughout the control process, lacking the "flexibility" to handle temporary violations caused by severe saturation or strong disturbances. This may lead to singularities in the error transformation function and subsequent controller failure, thereby limiting the applicability and robustness of the system under extreme conditions. To address this, a series of flexible PPC methods is designed with adaptive performance boundaries. In [20], an integral-based auxiliary term was incorporated to link performance boundaries with input saturation, enabling automatic boundary relaxation during saturation at the cost of slowed relaxation dynamics. In [21], Bu et al. transformed rigid PPC constraints into flexible ones by relating system states to performance boundaries, improving robustness without considering the influence of different initial errors. While a sliding adaptive-boundary PPC scheme was introduced in [22], the work does not distinguish that the singularity originates from boundary violation, not saturation itself, causing ill-conditioned control terms. However, the design of prescribed performance functions in existing studies [20], [21], [22], [23], [24] often relies on empirical or simplified assumptions, consequently limiting their flexibility and generality, which is a key challenge and important direction for future research.

On the other hand, for optimal control of nonlinear systems, achieving desired performance with minimal control cost while ensuring operational safety is of significant theoretical value and practical relevance. Methods based on solving the Hamilton-Jacobi-Bellman (HJB) equation provide a theoretical framework for this problem [25]. However, due to the strong nonlinearity and partial differential nature of the HJB equation, its analytical solution is generally difficult to obtain directly [26]. Consequently, reinforcement learning (RL)-based control approaches have been introduced to approximate the optimal solution through continuous training [27], thereby circumventing the direct solution of the HJB equation. Building on this foundation, RL-based optimal control approaches have been extended and deepened in multiple directions [28], [29], [30]. For instance, the work [31] proposed a critic-actor framework-based RL algorithm enabling simultaneous online training of control policies and performance indices. For multi-agent systems with unknown dynamics, the work [32] developed a control scheme combining backstepping design and optimization. In recent years, researchers have further applied RL to more complex systems: in [33], an identifier-critic-actor structured control framework is presented for stochastic nonlinear systems, using an adaptive identifier to approximate unknown dynamics online. Inspired by this, predefined-time RL control methods have subsequently been proposed for nonlinear systems [19], [34] and flexible-joint robots [35]. Moreover, progress has been made in integrating RL with PPC, with corresponding schemes designed for Markov jump systems [36], stochastic nonlinear systems [37], and nonlinear systems [38], [39], [40], [41], [42]. Nevertheless, for nonlinear saturated systems simultaneously subject to external disturbances and prescribed performance constraints, RL-based PPC remains challenging, and related research is still relatively scarce, warranting further in-depth exploration.

Inspired by the aforementioned research, this paper focuses on designing a reinforcement learning control scheme with self-regulated prescribed performance for the optimal control of nonlinear saturated systems subject to external disturbances. The main contributions of this work are summarized as follows:

- This paper proposes a novel self-regulated prescribed performance control (SRPPC) algorithm. This algorithm not only adaptively adjusts the performance boundaries according to the initial system error but also dynamically reconstructs these boundaries when the tracking error is about to violate the prescribed constraints due to strong disturbances and input saturation, thereby effectively preventing potential singularities during the control process. Compared to existing monotonic PPC methods [8], [9], [10], [11], [12], [13], [14], [15], [19], [38], [41] and flexible PPC methods [17], [20], [21], [22], [23], [24], the proposed SRPPC strategy eliminates the need for repeated trial-and-error parameter tuning and does not sacrifice the initial transient performance to accommodate different initial conditions. Moreover, unlike existing PPC approaches [17], [20], [21], [22], [23], [24], [40], [42], [43], the proposed control strategy establishes a selective dynamic adjustment mechanism of the performance boundaries based on the error-boundary relationship, which can avoid unnecessary relaxation and control performance loss.
- A novel reinforcement learning-based SRPPC (RL-SRPPC) optimal control algorithm is developed for systems subjected to external disturbances and input saturation constraints. By embedding the SRPPC-based transformed error directly into the performance index, the proposed scheme ensures prescribed tracking accuracy while concomitantly minimizing the control cost, thereby enhancing its universality and practical applicability. In contrast to existing RL-based PPC optimal control algorithms [19], [38], [39], [40], [41], [42]—which typically overlook the synergistic impact of intense disturbances and actuator saturation on the prescribed performance boundaries (PPBs)—the proposed RL-SRPPC optimal control algorithm ensures safer system operation and superior transient/steady-state performance metrics with lower control costs.

Notation: Throughout this paper, $\mathbb{R}$, $\mathbb{R}^+$, and $\mathbb{R}^n$ denote the sets of real numbers, positive real numbers, and n-dimensional real vectors, respectively, while $\mathbb{R}^{m\times n}$ represents the space of $m \times n$ real matrices. For a given matrix or vector A, A^T and A^{-1} (if it exists) designate its transpose and inverse, respectively. The absolute value of a scalar $a \in \mathbb{R}$ is denoted by $|a|$, and $\operatorname{sgn}(\cdot)$ refers to the standard signum function, defined as: $\operatorname{sgn}(a) = 1$ if $a > 0$, $\operatorname{sgn}(a) = -1$ if $a < 0$, and $\operatorname{sgn}(a) = 0$ if $a = 0$.

II. Preliminaries

A. Problem Statement

Consider the following input-saturation system with external disturbances:

$$\begin{cases} \dot{x}_i = x_{i+1} + f_i(\overline{x}_i) + d_i(t),\ i = 1, ..., n-1 \\ \dot{x}_n = \operatorname{sat}(u) + f_n(\overline{x}_n) + d_n(t), \\ y = x_1, \end{cases} \tag{1}$$

where $\overline{x}_i = [x_1, \ldots, x_i]^T \in \mathbb{R}^i$ is the system state vector with $\overline{x}_n = [x_1, ..., x_n]^T \in \mathbb{R}^n$, $y \in \mathbb{R}$ represents the system output, and $f_i(\overline{x}_i) \in \mathbb{R}$ is the unknown and continuous nonlinear function. $\operatorname{sat}(u) \in \mathbb{R}$ is the saturation input, which should be described as

$$\operatorname{sat}(u) = \begin{cases} u, & |u| \le u_d, \\ u_d \operatorname{sgn}(u), & |u| > u_d, \end{cases} \tag{2}$$

where u_d stands for the saturation threshold.

To mitigate the discontinuous transitions inherent in $\operatorname{sat}(u)$, the following approximation is employed:

$$\operatorname{sat}(u) = d(u) + g(u), \tag{3}$$

where $d(u) = u_d \tanh(u/u_d)$ constitutes a smooth, differentiable component, and $g(u) = \operatorname{sat}(u) - d(u)$ represents the associated residual. The magnitude of this residual is bounded such that $|g(u)| \le u_d(1 - \tanh(1)) = \bar{g}$, where $\bar{g}$ is a known positive constant.

Control Objective: Design the RL-SRPPC algorithm for the optimal control of the input-saturated system (1) with external disturbances so that: 1) all signals in the closed-loop system are semi-globally bounded and 2) the output state y follows the reference signal y_d with the prescribed control accuracy, where y_d is a smooth and bounded function.

Assumption 1 ([41]): The time-varying disturbance $d_i(t)$ possesses boundedness, specifically satisfying $|d_i(t)| < \bar{d}$, with $\bar{d} > 0$ being a constant.

B. Interval Type-2 Fuzzy Logic System (IT2FLS)

Lemma 1 ([3]): The IT2FLS possesses superior uncertainty modeling capabilities, increased disturbance resistance, and a reduced number of membership functions in comparison to the type-1 FLS. Consequently, IT2FLS is employed to approximate the unknown nonlinear functions $\mathcal{F}(X)$, which can be defined as follows:

$$\mathcal{F}(X) = W^{*T} S(X) + \varepsilon(X), \quad (|\varepsilon(X)| \le \bar{\varepsilon}, \bar{\varepsilon} \in \mathbb{R}^+) \tag{4}$$

where W^* is the ideal weight matrix, $\varepsilon(X)$ is the approximation error, $X = [\chi_1, ..., \chi_n]$ is the input variable, and $S(x) = [\frac{1}{2}S_U^T, \frac{1}{2}S_L^T]^T$ is the fuzzy basis vector. $S_U = [\overline{s}_1, ..., \overline{s}_m]^T$ and $S_L = [\underline{s}_1, ..., \underline{s}_m]^T$, respectively, are the upper fuzzy basis function and the lower fuzzy basis function, which can be formulated as

$$\begin{aligned} \overline{s}_i &= \frac{\overline{\wp}_i}{\sum_{i=1}^m \overline{\wp}_i}, \\ \underline{s}_i &= \frac{\underline{\wp}_i}{\sum_{i=1}^m \underline{\wp}_i}, \end{aligned} \tag{5}$$

where $\overline{\wp}_i$ and $\underline{\wp}_i$ are the interval type-2 fuzzy rule firing strengths, which can be calculated by the following multiplication method:

$$\begin{aligned} \overline{\wp}_i &= \textstyle\prod_{j=1}^n \overline{\mu}_{ij}, \\ \underline{\wp}_i &= \textstyle\prod_{j=1}^n \underline{\mu}_{ij}, j = 1, ..., n. \end{aligned} \tag{6}$$

The Gaussian function is utilized as the membership function (MF), wherein $\overline{\mu}_{ij}$ and $\underline{\mu}_{ij}$ can be described as

$$\begin{aligned} \overline{\mu}_{ij} &= \exp\left(-\frac{(\chi_j - c_{ij})^2}{2\bar{\sigma}_{ij}^2}\right), \\ \underline{\mu}_{ij} &= \exp\left(-\frac{(\chi_j - c_{ij})^2}{2\underline{\sigma}_{ij}^2}\right), \end{aligned} \tag{7}$$

where c_{ij} is the center, $\bar{\sigma}_{ij}$ and $\underline{\sigma}_{ij}$ are the upper and lower widths, respectively.

III. Main Results

A. Self-Regulated Prescribed Performance

To constrain the convergence performance of the system, an appointed-time prescribed performance function (PPF) can be defined as

$$\kappa(t) = \begin{cases} (\varpi_0 - \varpi_f)\cos^{a_\kappa}\left(\frac{\pi t}{2T_f}\right) + \varpi_f, t \in [0, T_f) \\ \varpi_f, t \in [T_f, +\infty) \end{cases} \tag{8}$$

where $T_f > 0$ stands for the appointed convergence time, and a_κ, ϖ_0, $\varpi_f > 0$ are the design parameters with $\varpi_0 > \varpi_f$. The PPF will be maintained at a fixed value ϖ_f when the system reaches the appointed time T_f.

To deal with the fragility problem of the classical PPC, the classical monotonous prescribed performance boundaries (PPBs) are transferred into the self-regulated PPBs, which are defined as

$$\underline{\mathcal{K}}(t) < e(t) < \overline{\mathcal{K}}(t) \tag{9}$$

with

$$\begin{aligned} \overline{\mathcal{K}}(t) &= \overline{\varkappa}(t) + \overline{\vartheta}\overline{\varrho}(t), \\ \underline{\mathcal{K}}(t) &= \underline{\varkappa}(t) - \underline{\vartheta}\underline{\varrho}(t), \end{aligned} \tag{10}$$

where $e(t) = y - y_d$ stands for the tracking error.

To remove the initial feasibility condition, the sliding PPBs can be designed as

$$\begin{aligned} \overline{\varkappa}(t) &= \tfrac{1}{2}\kappa(t)\left[1 + \flat + \aleph(e_0)(1 - \flat)\right] + e_0\mathcal{I}(t), \\ \underline{\varkappa}(t) &= -\tfrac{1}{2}\kappa(t)\left[1 + \flat - \aleph(e_0)(1 - \flat)\right] + e_0\mathcal{I}(t), \end{aligned} \tag{11}$$

with

$$\mathcal{I}(t) = \begin{cases} \cos^{a_s}\left(\frac{\pi t}{2T_f}\right), t \in [0, T_f) \\ 0, t \in [T_f, \infty) \end{cases} \tag{12}$$

where e_0 stands for the initial error value of the control system, $\aleph(\cdot) = 1$ if $\cdot \ge 0$, $\aleph(\cdot) = -1$ if $\cdot < 0$, $0 < \flat \le 1$ and $a_s \ge 1$ are the designed parameters.

Design an auxiliary variable as

$$\hbar(t) = e(t) - e_0\mathcal{I}(t) \tag{13}$$

Based on (9) and (13), one can obtain

$$\underline{\mathfrak{K}}(t) < \hbar(t) < \overline{\mathfrak{K}}(t) \tag{14}$$

where $\overline{\mathfrak{K}}(t) = \frac{1}{2}\kappa(t)\left[1+\flat+\aleph(e_0)(1-\flat)\right] + \overline{\vartheta}\overline{\varrho}(t)$ and $\underline{\mathfrak{K}}(t) = -\frac{1}{2}\kappa(t)\left[1+\flat-\aleph(e_0)(1-\flat)\right] - \underline{\vartheta}\underline{\varrho}(t)$.

To realize the dynamic relaxation of the PPBs, the flexible terms $\overline{\vartheta}\overline{\varrho}(t)$, $\underline{\vartheta}\underline{\varrho}(t)$ are designed as

$$\begin{aligned} \overline{\vartheta} &= \begin{cases} 1, & \text{if } 0 < a_u \frac{\overline{\varkappa}(t)}{\hbar(t)} \le 1, \\ 0, & \text{otherwise}, \end{cases} \\ \underline{\vartheta} &= \begin{cases} 1, & \text{if } 0 < a_l \frac{\underline{\varkappa}(t)}{\hbar(t)} \le 1, \\ 0, & \text{otherwise}, \end{cases} \end{aligned} \tag{15}$$

$$\begin{aligned} \overline{\varrho}(t) &= \begin{cases} \left(\frac{T_f-\Lambda_u(t)t}{T_f}\right)\frac{(\varpi_0-\varpi_f)\Delta\Lambda_u((t)}{1+(\Lambda_u(t)t)^2}, t<T_f, \\ (1-\Lambda_u(t))\frac{(\varpi_0-\varpi_f)\Delta\Lambda_u(t)}{1+(\Lambda_u(t)T_f)^2}, t \ge T_f, \end{cases} \\ \underline{\varrho}(t) &= \begin{cases} \left(\frac{T_f-\Lambda_l(t)t}{T_f}\right)\frac{(\varpi_0-\varpi_f)\Delta\Lambda_l(t)}{1+(\Lambda_l(t)t)^2}, t<T_f, \\ (1-\Lambda_l(t))\frac{(\varpi_0-\varpi_f)\Delta\Lambda_l(t)}{1+(\Lambda_l(t)T_f)^2}, t \ge T_f, \end{cases} \end{aligned} \tag{16}$$

with

$$\Lambda_u(t) = a_u\frac{\overline{\varkappa}(t)}{|\hbar(t)|}, \Lambda_l(t) = a_l\frac{\underline{\varkappa}(t)}{|\hbar(t)|},$$

$$\Delta\Lambda_u(t) = \tanh\left(\frac{b_u}{\Lambda_u(t)} - b_u\right)\tanh\left(\frac{\frac{a_u}{\Lambda_u(t)} - a_u}{c_u(1-a_u)}\right),$$

$$\Delta\Lambda_l(t) = \tanh\left(\frac{b_l}{\Lambda_l(t)} - b_l\right)\tanh\left(\frac{\frac{a_l}{\Lambda_l(t)} - a_l}{c_l(1-a_l)}\right),$$

where a_u, a_l, b_u, b_l, c_u, $c_l > 0$ are design parameters. These design parameters work in concert to regulate the relaxed PPBs.

Remark 1: As indicated by (10), the designed SRPPC is composed of two distinct components: the sliding terms $\overline{\varkappa}$, $\underline{\varkappa}$ and the flexible terms $\overline{\vartheta}\overline{\varrho}(t), \underline{\vartheta}\underline{\varrho}(t)$. The primary role of the sliding terms is to decouple the system's performance from the initial error condition, a limitation inherent in conventional monotonous PPC approaches. The flexible terms are introduced to mediate between the demands of input safety and tracking performance. Specifically, from (15) and (16), $\overline{\vartheta}$ and $\underline{\vartheta}$ govern the activation logic of the relaxation, while $\overline{\varrho}$ and $\underline{\varrho}$ control the expansion magnitude. These terms are modulated by proximity indices $\Lambda_u(t)$ and $\Lambda_l(t)$, which quantify how closely the error approaches the boundaries. This mechanism ensures the boundary adaptively reconfigures when output constraints conflict with prescribed performance specifications, preventing control singularities under input saturation. The conceptual diagram of the proposed SRPPC is illustrated in subfigures (a) and (b) of Fig. 1.

Remark 2: A common drawback of mainstream PPC methods—both monotonic [8], [9], [10], [11], [15], [19], [38], [41], [42] and flexible [20], [21], [23], [24]—is the static nature of their initial PPBs. Once parameters are chosen, the initial

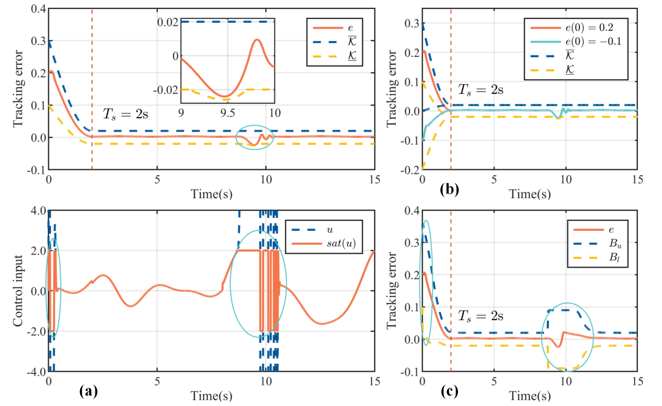


Fig. 1. (a) and (b) represent the conceptual graph of the designed SRPPC; (c) exhibits the existing flexible PPC based on the input saturation.

boundaries are fixed. Consequently, the controller cannot accommodate different initial errors $e(0)$ unless its parameters are manually resynthesized. Additionally, the infinity initial boundary method [12], [13], [14], [43] is developed to remove the limitation of initial conditions, while sacrificing transient response. Obviously, this often leads to a trade-off: a large $e(0)$ requires a loose initial bound, which deteriorates early transient response. However, the SRPPC technique (10) proposed herein fundamentally addresses this rigidity. According to (10)–(14), the proposed boundary is constructed by integrating the initial tracking error $e(0)$, the shifting function (12), and the auxiliary variable (13). This formulation (11) decouples the control design from a specific $e(0)$, enabling application to arbitrary bounded initial conditions without sacrificing initial performance. Intuitively, the designed SRPPB possesses the ability to self-adjust, always containing the initial error within a prescribed range and sliding adaptively as $e(0)$ changes, as shown in subfigure (b) of Fig. 1.

Remark 3: According to (15) and (16), the flexible components $\overline{\vartheta}\overline{\varrho}(t)$ and $\underline{\vartheta}\underline{\varrho}(t)$ remain zero, provided that the tracking error stays within the user-specified PPBs. When the input saturation or severe disturbances cause the system error to surpass the preset performance boundaries, the SRPPC method is triggered to autonomously relax the PPBs to avoid the singularity. Compared with the existing flexible PPC schemes [17], [20], [21], [22], [23], [24], [40], [43], the proposed SRPPC scheme dynamically regulates the PPBs according to the instantaneous distance between the system error and its allowable limits, as shown in subfigures (a) and (c) of Fig. 1. This adjustment mechanism, established directly on the error-boundary relationship, resolves the time-delay problem inherent in integral-based methods and prevents unnecessary PPB relaxation when saturation occurs without an actual boundary violation. Moreover, another core mechanism of the SRPPC method is the selective relaxation of only the specific PPB that is exceeded, as opposed to simultaneously relaxing both the upper and lower bounds. This selective dynamic regulation mechanism of the PPBs provides more precise control, which can obtain better system performance.

B. Optimized Backstepping Control Design

Given the prescribed performance bounds $\underline{\mathcal{K}}(t) < e(t) < \overline{\mathcal{K}}(t)$, $\forall t \geq 0$, a transformation function $\mathcal{T}(t)$ is defined as

$$\mathcal{T}(t) = \ln\left(\frac{\mho(t)}{1-\mho(t)}\right), \tag{17}$$

where $\mho(t) = \frac{\hbar(t)-\underline{\mathfrak{K}}(t)}{\overline{\mathfrak{K}}(t)-\underline{\mathfrak{K}}(t)}$.

Calculate the derivative of $\mathcal{T}(t)$ as

$$\dot{\mathcal{T}}(t) = \eth(\dot{e} + \wp), \tag{18}$$

with

$$\eth = \frac{\overline{\mathfrak{K}}(t) - \underline{\mathfrak{K}}(t)}{\left(\hbar(t) - \underline{\mathfrak{K}}(t)\right)\left(\overline{\mathfrak{K}}(t) - \hbar(t)\right)},$$

$$\wp = -e_0\dot{\mathcal{I}} + \frac{\dot{\underline{\mathfrak{K}}}(t)(\hbar(t) - \overline{\mathfrak{K}}) - \dot{\overline{\mathfrak{K}}}(t)(\hbar(t) - \underline{\mathfrak{K}}(t))}{\overline{\mathfrak{K}}(t) - \underline{\mathfrak{K}}(t)}.$$

Step 1: From (18) and the tracking error $e = x_1 - y_d$, define $z_1 = \mathcal{T}$. Then, differentiating z_1 yields

$$\dot{z}_1 = \eth(x_2 + f_1 + d_1 - \dot{y}_d + \wp). \tag{19}$$

Consider α_1 as the virtual control law and α_1^* as its optimal counterpart, and one can design the corresponding optimal value function as

$$\begin{aligned} J_1^*(z_1) &= \min_{\alpha_1 \in \Omega}\left(\int_t^\infty h_1(z_1(s), \alpha_1(z_1))ds\right) \\ &= \int_t^\infty h_1(z_1(s), \alpha_1^*(z_1))ds, \end{aligned} \tag{20}$$

where Ω stands for a compact set containing origin, and $h_1(z_1, \alpha_1) = z_1^2 + \alpha_1^2$ is the cost function.

Treating x_2 as the optimal virtual control α_1^* and differentiating (20) to obtain the HJB equation, one has

$$\begin{aligned} H_1\left(z_1, \alpha_1^*, \tfrac{dJ_1^*}{dz_1}\right) &= z_1^2 + \alpha_1^{*2} + \tfrac{dJ_1^*(z_1)}{dz_1} \\ &\quad \times \left(\eth(\alpha_1^* + f_1 + d_1 - \dot{y}_d + \wp)\right) \\ &= 0. \end{aligned} \tag{21}$$

By solving $\partial H_1 / \partial \alpha_1^* = 0$, the optimal control α_1^* can be obtained as

$$\alpha_1^* = -\frac{1}{2}\eth\frac{dJ_1^*(z_1)}{dz_1}. \tag{22}$$

For achieving control objective, the term $dJ_1^*(z_1)/dz_1$ is decomposed as

$$\frac{dJ_1^*(z_1)}{dz_1} = \frac{2c_1}{\eth^2}z_1 - 2\eth\dot{y}_d + \frac{2}{\eth^2}\mathcal{F}_1(X_1) + \frac{1}{\eth^2}J_1^o(X_1), \tag{23}$$

where $c_1 > 0$ is a design parameter and $J_1^o(X_1) = -2c_1z_1 - 2\mathcal{F}_1(X_1) + \eth^2 dJ_1^*(z_1)/dz_1$, and $\mathcal{F}_1(X_1) = \eth(\frac{\eth}{2}z_1 + f_1 + d_1 + \wp) + \frac{1}{4}z_1$ are continuous functions with $X_1 = [x_1, z_1]^T$.

Substituting (23) into (22) yields

$$\alpha_1^* = -\tfrac{c_1}{\eth}z_1 - \tfrac{1}{\eth}\mathcal{F}_1(X_1) - \tfrac{1}{2\eth}J_1^o(X_1) + \dot{y}_d. \tag{24}$$

Utilizing IT2FLSs to approximate the unknown functions $\mathcal{F}_1(X_1)$ and $J_1^o(X_1)$ over the compact domain Ω yields

$$\mathcal{F}_1(X_1) = W_{\mathcal{F}1}^{*T}S_{\mathcal{F}1}(X_1) + \varepsilon_{\mathcal{F}1}(X_1), \tag{25}$$

$$J_1^o(X_1) = W_{J1}^{*T}S_{J1}(X_1) + \varepsilon_{J1}(X_1), \tag{26}$$

From (23)-(26), one can obtain

$$\begin{aligned} \tfrac{dJ_1^*(z_1)}{dz_1} &= \tfrac{2c_1}{\eth^2}z_1 + \tfrac{2}{\eth^2}W_{\mathcal{F}1}^{*T}S_{\mathcal{F}1}(X_1) - 2\eth\dot{y}_d \\ &\quad + \tfrac{1}{\eth^2}W_{J1}^{*T}S_{J1}(X_1) + \varepsilon_1, \end{aligned} \tag{27}$$

$$\begin{aligned} \alpha_1^* &= -\tfrac{c_1}{\eth}z_1 - \tfrac{1}{\eth}W_{\mathcal{F}1}^{*T}S_{\mathcal{F}1}(X_1) + \dot{y}_d \\ &\quad - \tfrac{1}{2\eth}W_{J1}^{*T}S_{J1}(X_1) - \tfrac{1}{2}\varepsilon_1, \end{aligned} \tag{28}$$

where $\varepsilon_1 = 2\varepsilon_{\mathcal{F}1} + \varepsilon_{J1}$.

For unknown terms $W_{\mathcal{F}1}^*$ and W_{J1}^*, the designed controller is not directly applied. Hence, according to (25), (27), and (28), the RL strategy with the identifier-critic-actor IT2FLSs can be applied to obtain the valid optimized virtual control. Therefore, define the identifier IT2FLS as

$$\hat{\mathcal{F}}_1(X_1) = \hat{W}_{\mathcal{F}1}^TS_{\mathcal{F}1}(X_1), \tag{29}$$

where $\hat{\mathcal{F}}_1(X_1)$, $\hat{W}_{\mathcal{F}1}$, and $S_{\mathcal{F}1}(X_1)$, respectively, sthand for the identifier output, weight vector and fuzzy basis vector. And design the adaptive law of the weight $\hat{W}_{\mathcal{F}1}$ as

$$\dot{\hat{W}}_{\mathcal{F}1} = \Gamma_1\left(S_{\mathcal{F}1}(X_1)z_1 - \sigma_1\hat{W}_{\mathcal{F}1}\right), \tag{30}$$

where $\sigma_1 > 0$ and Γ_1, respectively, are the designed constant and positive-definite matrix.

To evaluate the control performance, design the critic IT2FLS as

$$\begin{aligned} \tfrac{d\hat{J}_1^*(z_1)}{dz_1} &= \tfrac{2c_1}{\eth^2}z_1 + \tfrac{2}{\eth^2}\hat{W}_{\mathcal{F}1}^TS_{\mathcal{F}1}(X_1) \\ &\quad + \tfrac{1}{\eth^2}\hat{W}_{c1}^TS_{J1}(X_1) - 2\eth\dot{y}_d, \end{aligned} \tag{31}$$

where $\frac{d\hat{J}_1^*(z_1)}{dz_1}$ is the approximated value of $\frac{dJ_1^*(z_1)}{dz_1}$, and $\hat{W}_{c1}$ is the critic weight, which can be formulated as

$$\dot{\hat{W}}_{c1} = -\gamma_{c1}S_{J1}(X_1)S_{J1}^T(X_1)\hat{W}_{c1}, \tag{32}$$

where $\gamma_{c1} > 0$ is the designed parameter.

To execute the control action, develop the actor IT2FLS as

$$\hat{\alpha}_1^* = -\tfrac{c_1}{\eth}z_1 - \tfrac{1}{\eth}\hat{W}_{\mathcal{F}1}^TS_{\mathcal{F}1} - \tfrac{1}{2\eth}\hat{W}_{a1}^TS_{J1} + \dot{y}_d, \tag{33}$$

where α_1^* denotes the approximated value of the optimal virtual control law (28), and $\hat{W}_{a1}$ stands for the actor weight, which can be designed as

$$\begin{aligned} \dot{\hat{W}}_{a1} &= -S_{J1}(X_1)S_{J1}^T(X_1) \\ &\quad \times \left(\gamma_{a1}\left(\hat{W}_{a1} - \hat{W}_{c1}\right) + \gamma_{c1}\hat{W}_{c1}\right), \end{aligned} \tag{34}$$

where $\gamma_{a1} > 0$ is the designed parameters.

Step i $(i = 2, \ldots, n-1)$*:* In accordance with the backstepping design framework, depicting the subsystem error variable as $z_i = x_i - \hat{\alpha}_{i-1}^*$ yields

$$\dot{z}_i = x_{i+1} + f_i + d_i - \dot{\hat{\alpha}}_{i-1}^*. \tag{35}$$

Consider α_i as the virtual control and α_i^* as its optimal counterpart, respectively. Then, similar to (20) and describing $h_i(z_i, \alpha_i) = z_i^2 + \alpha_i^2$ as the cost function, one has

$$\begin{aligned} J_i^*(z_i) &= \min_{\alpha_i \in \Omega}\left(\int_t^\infty h_i(z_i, \alpha_i)ds\right) \\ &= \int_t^\infty h_i(z_i, \alpha_i^*)ds. \end{aligned} \tag{36}$$

Letting $x_{i+1} \triangleq \alpha_i^*$, one can obtain the HJB equation for (35) as

$$H_i\left(z_i, \alpha_i^*, \frac{dJ_i^*}{dz_i}\right) = z_i^2 + \alpha_i^{*2} + \frac{dJ_i^*(z_i)}{dz_i} \times \left(f_i + d_i + \alpha_i^* - \dot{\hat{\alpha}}_{i-1}^*\right) = 0. \tag{37}$$

Calculating $\partial H_i/(\partial \alpha_i^*) = 0$, one can acquire the optimal virtual control α_i^* as

$$\alpha_i^* = -\frac{1}{2}\frac{dJ_i^*(z_i)}{dz_i}. \tag{38}$$

Decompose the gradient term $dJ_i^*(z_i)/dz_i$ as

$$\frac{dJ_i^*(z_i)}{dz_i} = 2c_i z_i + 2\mathcal{F}_i(X_i) + J_i^o(X_i), \tag{39}$$

where $c_i > 0$ is the designed parameter, $\mathcal{F}_i(X_i) = f_i + d_i - \dot{\hat{\alpha}}_{i-1}^* + \frac{5}{4}z_i$ and $J_i^o(X_i) = -2c_i z_i - 2\mathcal{F}_i(X_i) + dJ_i^*(z_i)/dz_i$ are unknown continuous functions with $X_i = [x_i, z_i]^T$.

Substituting (39) into (38), one has

$$\alpha_i^* = -c_i z_i - \mathcal{F}_i(X_i) - \frac{1}{2}J_i^o(X_i). \tag{40}$$

Using IT2FLSs to estimate the unknown terms $\mathcal{F}_i(X_i)$ and $J_i^o(X_i)$ yields

$$\mathcal{F}_i(X_i) = W_{\mathcal{F}i}^{*T} S_{\mathcal{F}i}(X_i) + \varepsilon_{\mathcal{F}i}(X_i), \tag{41}$$

$$J_i^o(X_i) = W_{Ji}^{*T} S_{Ji}(X_i) + \varepsilon_{Ji}(X_i). \tag{42}$$

According to (39)-(42), one has

$$\frac{dJ_i^*(z_i)}{dz_i} = 2c_i z_i + 2W_{\mathcal{F}i}^{*T} S_{\mathcal{F}i}(X_i) + W_{Ji}^{*T} S_{Ji}(X_i) + \varepsilon_i, \tag{43}$$

$$\alpha_i^* = -c_i z_i - W_{\mathcal{F}i}^{*T} S_{\mathcal{F}i}(X_i) - \tfrac{1}{2}W_{Ji}^{*T} S_{Ji}(X_i) - \tfrac{1}{2}\varepsilon_i, \tag{44}$$

where $\varepsilon_i = 2\varepsilon_{\mathcal{F}i} + \varepsilon_{Ji}$.

Given that the unattainable optimal virtual control law (44), implement RL via the identifier–critic–actor architecture to obtain the optimized control as

$$\hat{\mathcal{F}}_i(X_i) = \hat{W}_{\mathcal{F}i}^T S_{\mathcal{F}i}(X_i), \tag{45}$$

$$\frac{d\hat{J}_i^*(z_i)}{dz_i} = 2c_i z_i + 2\hat{W}_{\mathcal{F}i}^T S_{\mathcal{F}i}(X_i) + \hat{W}_{ci}^T S_{Ji}(X_i), \tag{46}$$

$$\hat{\alpha}_i^* = -c_i z_i - \hat{W}_{\mathcal{F}i}^T S_{\mathcal{F}i}(X_i) - \tfrac{1}{2}\hat{W}_{ai}^T S_{Ji}(X_i), \tag{47}$$

where $\hat{\mathcal{F}}_i(X_i)$, $\frac{d\hat{J}_i^*(z_i)}{dz_i}$, and $\hat{\alpha}_i^*$ stand for the approximated value of $\mathcal{F}_i(X_i)$, $\frac{dJ_i^*(z_i)}{dz_i}$, and α_i^*, respectively. $\hat{W}_{\mathcal{F}i}$, $\hat{W}_{ci}$, and $\hat{W}_{ai}$ stand for the available weights of the RL with the identifier-critic-actor structure.

Following the derivation of (30), (32) and (34), one can obtain

$$\dot{\hat{W}}_{\mathcal{F}i} = \Gamma_i\left(S_{\mathcal{F}i}(X_i)z_i - \sigma_i \hat{W}_{\mathcal{F}i}\right), \tag{48}$$

$$\dot{\hat{W}}_{ci} = -\gamma_{ci} S_{Ji}(X_i) S_{Ji}^T(X_i) \hat{W}_{ci}, \tag{49}$$

$$\dot{\hat{W}}_{ai} = -S_{Ji}(X_i) S_{Ji}^T(X_i) \times \left(\gamma_{ai}\left(\hat{W}_{ai} - \hat{W}_{ci}\right) + \gamma_{ci}\hat{W}_{ci}\right), \tag{50}$$

where σ_i, γ_{ci}, $\gamma_{ai} > 0$ and Γ_i are the designed parameters and the positive-definite matrix, respectively.

Step n: The backstepping control design culminates in obtaining the actual control law u. Concurrently, to mitigate the effects of input saturation, an auxiliary dynamic system is incorporated, which is defined as

$$\dot{\Psi} = -\Psi + (d(u) - u). \tag{51}$$

Letting $z_n = x_n - \hat{\alpha}_{n-1}^* - \Psi$, differentiating z_n yields

$$\dot{z}_n = u + f_n + d_n - \dot{\hat{\alpha}}_{n-1}^* + \Psi + g(u). \tag{52}$$

Depicting u^* as the optimal control input and $h_n(z_n, u) = z_n^2 + u^2$ as the cost function, one can obtain the optimal value function as

$$J_n^*(z_n) = \min_{u \in \Omega}\left(\int_t^\infty h_n(z_n, u)ds\right) = \int_t^\infty h_n(z_n, u^*)ds. \tag{53}$$

Then, calculate the HJB equation for (52) as

$$H_n\left(z_n, u^*, \frac{dJ_n^*}{dz_n}\right) = z_n^2 + u^{*2} + \frac{dJ_n^*}{dz_n} \times \left(u + f_n + d_n - \dot{\hat{\alpha}}_{n-1}^* + \Psi + g(u)\right) = 0. \tag{54}$$

Similar to previous steps, solving $\frac{\partial H_n}{\partial u^*} = 0$ yields

$$u^* = -\frac{1}{2}\frac{dJ_n^*(z_n)}{dz_n}. \tag{55}$$

Rewriting the term $\frac{dJ_n^*(z_n)}{dz_n}$ as

$$\frac{dJ_n^*(z_n)}{dz_n} = 2c_n z_n + \Psi + 2\mathcal{F}_n(X_n) + J_n^o(X_n), \tag{56}$$

where $c_n > 0$ is the designed parameter and $J_n^o(X_n) = -2c_n z_n - \Psi - 2\mathcal{F}_n(X_n) + \frac{dJ_n^*(z_n)}{dz_n}$, and $\mathcal{F}_n(X_n) = f_n + d_n - \dot{\hat{\alpha}}_{n-1}^* + \frac{5}{4}z_n$ are unknown with $X_n = [x_n, z_n]^T$.

Substituting (56) into (55) yields

$$u^* = -c_n z_n - \Psi - \mathcal{F}_n - \frac{1}{2}J_n^o(X_n). \tag{57}$$

Utilizing IT2FLSs to approximate $\mathcal{F}_n(X_n)$ and $J_n^o(X_n)$ yields

$$\mathcal{F}_n(X_n) = W_{\mathcal{F}n}^{*T} S_{\mathcal{F}n}(X_n) + \varepsilon_{\mathcal{F}n}(X_n), \tag{58}$$

$$J_n^o(X_n) = W_{Jn}^{*T} S_{Jn}(X_n) + \varepsilon_{Jn}(X_n). \tag{59}$$

According to (56)-(59), one has

$$\frac{dJ_n^*(z_n)}{dz_n} = 2c_n z_n + \Psi + 2W_{\mathcal{F}n}^{*T} S_{\mathcal{F}n}(X_n) + W_{Jn}^{*T} S_{Jn}(X_n) + \varepsilon_n, \tag{60}$$

$$u^* = -c_n z_n - \Psi - W_{\mathcal{F}n}^{*T} S_{\mathcal{F}n}(X_n) - \tfrac{1}{2}W_{Jn}^{*T} S_{Jn}(X_n) - \tfrac{1}{2}\varepsilon_n, \tag{61}$$

where $\varepsilon_n = 2\varepsilon_{\mathcal{F}n} + \varepsilon_{Jn}$.

Owing to the unavailability of the optimal control (61), the RL is therefore enacted via the construction of identifier, critic, and actor IT2FLSs, to acquire the optimized control as

$$\hat{\mathcal{F}}_n(X_n) = \hat{W}_{\mathcal{F}n}^T S_{\mathcal{F}n}(X_n), \tag{62}$$

$$\frac{d\hat{J}_n^*(z_n)}{dz_n} = 2c_n z_n + \Psi + 2\hat{W}_{\mathcal{F}n}^T S_{\mathcal{F}n}(X_n) + \hat{W}_{cn}^T S_{Jn}(X_n), \tag{63}$$

$$u = -c_n z_n - \Psi - \hat{W}_{\mathcal{F}n}^T S_{\mathcal{F}n}(X_n) - \frac{1}{2}\hat{W}_{an}^T S_{Jn}(X_n), \tag{64}$$

where $\hat{\mathcal{F}}_n(X_n)$ and $\frac{d\hat{J}_n^*(z_n)}{dz_n}$, respectively, stand for the approximated value of $\mathcal{F}_n(X_n)$ and $\frac{dJ_n^*(z_n)}{dz_n}$. $\hat{W}_{\mathcal{F}n}$, $\hat{W}_{cn}$, and $\hat{W}_{an}$ stand for the available weights of the RL with the identifier-critic-actor structure.

Analogous to the previous steps, one has

$$\dot{\hat{W}}_{\mathcal{F}n} = \Gamma_n\left(S_{\mathcal{F}n}(X_n)z_n - \sigma_n \hat{W}_{\mathcal{F}n}\right), \tag{65}$$

$$\dot{\hat{W}}_{cn} = -\gamma_{cn} S_{Jn}(X_n) S_{Jn}^T(X_n)\hat{W}_{cn}, \tag{66}$$

$$\dot{\hat{W}}_{an} = -S_{Jn}(X_n)S_{Jn}^T(X_n) \times \left(\gamma_{an}\left(\hat{W}_{an} - \hat{W}_{cn}\right) + \gamma_{cn}\hat{W}_{cn}\right), \tag{67}$$

where σ_n, γ_{cn}, $\gamma_{an} > 0$ and Γ_n are the designed parameters and positive-definite matrix.

Remark 4: Inspired by [36], [37], [38], [39], [40], [41], the design details of $\dot{\hat{W}}_{cj}$ and $\dot{\hat{W}}_{aj}$ $(j = 1, ..., n)$ are presented as follows:

Take the deriving details of $\dot{\hat{W}}_{c1}$ and $\dot{\hat{W}}_{a1}$, for example. From (33), the Bellman residual error Φ_1 can be defined as

$$\begin{aligned}\Phi_1 &= H_1\left(z_1, \hat{\alpha}_1^*, \frac{d\hat{J}_1^*(z_1)}{dz_1}\right) - H_1\left(z_1, \alpha_1^*, \frac{dJ_1^*}{dz_1}\right) \\ &= H_1\left(z_1, \hat{\alpha}_1^*, \frac{d\hat{J}_1^*(z_1)}{dz_1}\right).\end{aligned} \tag{68}$$

Based on the previous analysis, $\hat{\alpha}_1^*$ is designed to achieve $\Phi_1 = H_1(z_1, \hat{\alpha}_1^*, d\hat{J}_1^*(z_1)/dz_1) \to 0$. Assuming the condition $H_1(z_1, \hat{\alpha}_1^*, d\hat{J}_1^*(z_1)/dz_1) = 0$ is satisfied with solution uniqueness, which implies the following equivalence:

$$\frac{\partial H_1\left(z_1, \hat{\alpha}_1^*, \frac{d\hat{J}_1^*(z_1)}{dz_1}\right)}{\partial \hat{W}_{a1}} = \frac{1}{2\eth^2} S_{J1}(X_1)S_{J1}^T(X_1) \times \left(\hat{W}_{a1} - \hat{W}_{c1}\right) = 0. \tag{69}$$

In accordance with ensuring (69) through the formulation of RL adaptive laws, define the positive function as

$$P_1 = \left(\hat{W}_{a1} - \hat{W}_{c1}\right)^T \left(\hat{W}_{a1} - \hat{W}_{c1}\right). \tag{70}$$

It is evident that the condition $P_1 = 0$ corresponds to (69). Consequently, the adaptive laws (32) and (34) are derived from the following considerations.

Since $\partial P_1/\partial \hat{W}_{a1} = -\partial P_1/\partial \hat{W}_{c1} = 2(\hat{W}_{a1} - \hat{W}_{c1})$, differentiating P_1 along (32) and (34) has

$$\begin{aligned}&\frac{dP_1}{dt} = \frac{\partial P_1}{\partial \hat{W}_{c1}^T}\dot{\hat{W}}_{c1} + \frac{\partial P_1}{\partial \hat{W}_{a1}^T}\dot{\hat{W}}_{a1} \\ &= -\gamma_{c1}\frac{\partial P_1}{\partial \hat{W}_{c1}^T} S_{J1}(X_1)S_{J1}^T(X_1)\hat{W}_{c1} \\ &-\frac{\partial P_1}{\partial \hat{W}_{c1}^T} S_{J1}(X_1)S_{J1}^T(X_1)\left(\gamma_{a1}\left(\hat{W}_{a1} - \hat{W}_{c1}\right) + \gamma_{c1}\hat{W}_{c1}\right) \\ &= -\frac{\gamma_{a1}}{2}\frac{\partial P_1}{\partial \hat{W}_{a1}^T} S_{J1}(X_1)S_{J1}^T(X_1)\frac{\partial P_1}{\partial \hat{W}_{a1}} \le 0.\end{aligned} \tag{71}$$

The inequality (71) indicates that the adaptive laws (32) and (34) can enable that $P_1 = 0$ is true, thereby ensuring (69).

Remark 5: The key distinction between the formulated HJB equation and that of classical identifier-critic-actor RL lies in the term z_1. Unlike the standard assignment $z_1 = y - y_d$, the variable is redefined here via the SRPPC technique (9) and the transformation function (18), prompting a consequent adaptation of the first HJB equation (20). This is not merely a formal change but a substantive one, setting the proposed controller apart from conventional RL methods and enabling enhanced performance. Given the severe analytical challenges posed by the nonlinear partial differential characteristics of HJB equations, the adopted identifier-critic-actor RL strategy provides a practical alternative, acquiring near-optimal solutions through sustained training without directly solving the equation. Furthermore, the incorporation of an identifier equipped with IT2FLSs allows the actor-critic framework to explicitly account for unknown dynamics and actuator constraints. By leveraging the superior uncertainty-modeling capability of IT2FLSs, the proposed scheme effectively maintains the prescribed performance envelopes and achieves a favorable trade-off between tracking precision and optimal control expenditure.

Remark 6: The implementation of the identifier-critic-actor structure involves the online adaptation of three distinct weight sets. However, the real-time computational burden is effectively mitigated by the linearly weighted structure of the IT2FLS, which transforms the learning process into a series of efficient algebraic operations. Specifically, the fuzzy basis functions $S(\cdot)$ are computed once per time step and shared where applicable, while the weight update laws follow recursive gradients that do not require matrix inversions. For a system with N fuzzy rules, the per-step complexity is $O(N)$, which is well within the processing capabilities of modern embedded controllers. Furthermore, the decentralized architecture of the proposed scheme allows for subsystem-level parallel processing, ensuring that the computational demand does not scale exponentially with the system's degrees of freedom, thereby making it highly suitable for high-bandwidth control tasks like high-speed robots.

C. Stability Analysis

Theorem 1: Consider the input-saturated nonlinear system (1) subjected to unknown external disturbances under Assumption 1. If the SRPPC-RL control algorithm is executed via integrating the identifier–critic–actor RL into the backstepping control framework, including the identifiers (30), (48), (65), the critics (32), (49), (66), the actors (34), (50) and (67), and the control laws (33), (47), (64) with the auxiliary dynamic system (51), then the optimal control can ensure that:

1) all signals in the closed-loop system are semi-globally bounded;
2) the system output can track the reference trajectory with the desired accuracy, and the tracking error is always kept in the designed dynamic boundaries.

Proof:

Choose the Lyapunov candidate function as

$$V = \frac{1}{2}\sum_{i=1}^{n}\left(z_i^2 + \tilde{W}_{\mathcal{F}i}^T\Gamma_i^{-1}\tilde{W}_{\mathcal{F}i} + \tilde{W}_{ci}^T\tilde{W}_{ci} + \tilde{W}_{ai}^T\tilde{W}_{ai}\right), \tag{72}$$

where $\tilde{W}_{\mathcal{F}i} = \hat{W}_{\mathcal{F}i} - W_{\mathcal{F}i}^*$, $\tilde{W}_{ci} = \hat{W}_{ci} - W_{Ji}^*$ and $\tilde{W}_{ai} = \hat{W}_{ai} - W_{Ji}^*$ stand for the weight estimation errors.

Then, differentiating V along with time yields

$$\begin{aligned}\dot{V} &= z_1(\eth(z_2+\hat{\alpha}_1^*+f_1+d_1-\dot{y}_d+\wp)) \\ &\quad+\sum_{i=2}^{n-1}\left(z_i(z_{i+1}+\hat{\alpha}_i^*+f_i+d_i-\dot{\hat{\alpha}}_{i-1}^*)\right) \\ &\quad+z_n(u+f_n+d_n-\dot{\hat{\alpha}}_{n-1}^*+\Psi+g(u)) \\ &\quad+\sum_{i=1}^{n}\left(\tilde{W}_{\mathcal{F}i}^T\Gamma_i^{-1}\dot{\hat{W}}_{\mathcal{F}i}+\tilde{W}_{ci}^T\dot{\hat{W}}_{ci}+\tilde{W}_{ai}^T\dot{\hat{W}}_{ai}\right).\end{aligned} \tag{73}$$

Substituting (33), (47) and (64) into (73) yields

$$\begin{aligned}\dot{V} &= -\sum_{i=1}^{n}\left(c_iz_i^2+z_i(\hat{W}_{\mathcal{F}i}^TS_{\mathcal{F}i}(X_i)+\tfrac{1}{2}\hat{W}_{ai}^TS_{Ji}(X_i))\right) \\ &\quad+z_1\eth(z_2+f_1+d_1+\wp)+\sum_{i=2}^{n}z_i(f_i+d_i) \\ &\quad+\sum_{i=2}^{n-1}z_i(z_{i+1}-\dot{\hat{\alpha}}_{i-1}^*)+z_n(g(u)-\dot{\hat{\alpha}}_{n-1}^*) \\ &\quad+\sum_{i=1}^{n}\left(\tilde{W}_{\mathcal{F}i}^T\Gamma_i^{-1}\dot{\hat{W}}_{\mathcal{F}i}+\tilde{W}_{ci}^T\dot{\hat{W}}_{ci}+\tilde{W}_{ai}^T\dot{\hat{W}}_{ai}\right).\end{aligned} \tag{74}$$

Utilizing Young's inequality, one can obtain

$$\begin{aligned}\eth z_1z_2 &\le \tfrac{\eth^2z_1^2}{2}+\tfrac{z_2^2}{2}, \\ z_iz_{i+1} &\le \tfrac{z_i^2}{2}+\tfrac{z_{i+1}^2}{2}, \\ z_ng(u) &\le \tfrac{1}{2}z_n^2+\tfrac{1}{2}\bar{g}^2, \\ -\tfrac{1}{2}z_i\hat{W}_{ai}^TS_{Ji}(X_i) &\le \tfrac{1}{4}z_i^2+\tfrac{1}{4}\left(\hat{W}_{ai}^TS_{Ji}\right)^2.\end{aligned} \tag{75}$$

Then, utilizing the definition of $\mathcal{F}_i(X_i)$, one has

$$\begin{aligned}\dot{V} &\le -\sum_{i=1}^{n}\left(c_iz_i^2+z_i\tilde{W}_{\mathcal{F}i}^TS_{\mathcal{F}i}-\tfrac{1}{4}\left(\hat{W}_{ai}^TS_{Ji}\right)^2\right) \\ &\quad+\sum_{i=1}^{n}\left(\tilde{W}_{\mathcal{F}i}^T\Gamma_i^{-1}\dot{\hat{W}}_{\mathcal{F}i}+\tilde{W}_{ci}^T\dot{\hat{W}}_{ci}+\tilde{W}_{ai}^T\dot{\hat{W}}_{ai}\right) \\ &\quad+\sum_{i=1}^{n}z_i\varepsilon_{\mathcal{F}i}+\tfrac{1}{2}\bar{g}^2.\end{aligned} \tag{76}$$

Substituting (30), (32), (34), (48)–(50) and (65)–(67) into (76) yields

$$\begin{aligned}\dot{V} &\le -\sum_{i=1}^{n}\left(c_iz_i^2-\tfrac{1}{4}\left(\hat{W}_{ai}^TS_{Ji}\right)^2-z_i\varepsilon_{\mathcal{F}i}\right)+\tfrac{1}{2}\bar{g}^2 \\ &\quad-\sum_{i=1}^{n}\sigma_i\tilde{W}_{\mathcal{F}i}^T\hat{W}_{\mathcal{F}i}-\sum_{i=1}^{n}\left(\gamma_{ci}\tilde{W}_{ci}^TS_{Ji}S_{Ji}^T\hat{W}_{ci}\right) \\ &\quad+\sum_{i=1}^{n}(\gamma_{ai}-\gamma_{ci})\tilde{W}_{ai}^TS_{Ji}S_{Ji}^T\hat{W}_{ci} \\ &\quad-\sum_{i=1}^{n}\gamma_{ai}\tilde{W}_{ai}^TS_{Ji}S_{Ji}^T\hat{W}_{ai}.\end{aligned} \tag{77}$$

According to the definition of $\tilde{W}_{\mathcal{F}i}$, $\tilde{W}_{ci}$ and $\tilde{W}_{ai}$ and Young's inequality, one can obtain

$$\begin{aligned}\tilde{W}_{\mathcal{F}i}^T\hat{W}_{\mathcal{F}i} &= \tfrac{1}{2}\tilde{W}_{\mathcal{F}i}^T\tilde{W}_{\mathcal{F}i}+\tfrac{1}{2}\hat{W}_{\mathcal{F}i}^T\hat{W}_{\mathcal{F}i}-\tfrac{1}{2}W_{\mathcal{F}i}^{*T}W_{\mathcal{F}i}^* \\ \tilde{W}_{ci}^TS_{Ji}S_{Ji}^T\hat{W}_{ci} &= \tfrac{1}{2}\left(\tilde{W}_{ci}^TS_{Ji}\right)^2+\tfrac{1}{2}\left(\hat{W}_{ci}^TS_{Ji}\right)^2 \\ &\quad-\tfrac{1}{2}\left(W_{Ji}^{*T}S_{Ji}\right)^2 \\ \tilde{W}_{ai}^TS_{Ji}S_{Ji}^T\hat{W}_{ai} &= \tfrac{1}{2}\left(\tilde{W}_{ai}^TS_{Ji}\right)^2+\tfrac{1}{2}\left(\hat{W}_{ai}^TS_{Ji}\right)^2 \\ &\quad-\tfrac{1}{2}\left(W_{Ji}^{*T}S_{Ji}\right)^2 \\ \tilde{W}_{ai}^TS_{Ji}S_{Ji}^T\hat{W}_{ci} &\le \tfrac{1}{2}\left(\tilde{W}_{ai}^TS_{Ji}\right)^2+\tfrac{1}{2}\left(\hat{W}_{ci}^TS_{Ji}\right)^2 \\ z_i\varepsilon_{\mathcal{F}i} &\le \tfrac{1}{2}z_i^2+\tfrac{1}{2}\bar{\varepsilon}_{\mathcal{F}i}^2\end{aligned} \tag{78}$$

Let $\gamma_{ai}>\frac{1}{2}$, $\gamma_{a1}>\gamma_{c1}>\frac{\gamma_{a1}}{2}$ and substitute (78) into (77), one has

$$\begin{aligned}\dot{V} &\le -\sum_{i=1}^{n}\left((c_i-\tfrac{1}{2})z_i^2+\tfrac{1}{2}\sigma_i\tilde{W}_{\mathcal{F}i}^T\tilde{W}_{\mathcal{F}i}-\tfrac{1}{2}\bar{\varepsilon}_{\mathcal{F}i}^2\right) \\ &\quad-\tfrac{1}{2}\sum_{i=1}^{n}\gamma_{ci}\left(\tilde{W}_{ci}^TS_{Ji}\right)^2-\tfrac{1}{2}\sum_{i=1}^{n}\gamma_{ci}\left(\tilde{W}_{ai}^TS_{Ji}\right)^2 \\ &\quad+\tfrac{1}{2}\sum_{i=1}^{n}(\gamma_{ai}+\gamma_{ci})\left(W_{Ji}^{*T}S_{Ji}\right)^2+\tfrac{1}{2}\bar{g}^2 \\ &\quad+\tfrac{1}{2}\sum_{i=1}^{n}\sigma_i(W_{\mathcal{F}i}^{*T}W_{\mathcal{F}i}^*).\end{aligned} \tag{79}$$

Letting $\lambda_{\Gamma_i^{-1}}^{\max}$ be the maximal eigenvalue of Γ_i^{-1} and $\lambda_{S_{Ji}}^{\min}$ be the minimal eigenvalue of $S_{Ji}S_{Ji}^T$, one has

$$\begin{aligned}\dot{V} &\le -\sum_{i=1}^{n}\left((c_i-\tfrac{1}{2})z_i^2+\tfrac{1}{2}\sigma_i\tilde{W}_{\mathcal{F}i}^T\tilde{W}_{\mathcal{F}i}-\tfrac{1}{2}\bar{\varepsilon}_{\mathcal{F}i}^2\right) \\ &\quad-\tfrac{1}{2}\sum_{i=1}^{n}\gamma_{ci}\left(\tilde{W}_{ci}^TS_{Ji}S_{Ji}^T\tilde{W}_{ci}+\tilde{W}_{ai}^TS_{Ji}S_{Ji}^T\tilde{W}_{ai}\right) \\ &\quad+\tfrac{1}{2}\sum_{i=1}^{n}(\gamma_{ai}+\gamma_{ci})\left(W_{Ji}^{*T}S_{Ji}\right)^2+\tfrac{1}{2}\bar{g}^2 \\ &\quad+\tfrac{1}{2}\sum_{i=1}^{n}\sigma_i(W_{\mathcal{F}i}^{*T}W_{\mathcal{F}i}^*) \\ &\le -\sum_{i=1}^{n}\left((c_i-\tfrac{1}{2})z_i^2+\tfrac{1}{2\lambda_{\Gamma_i^{-1}}^{\max}}\sigma_i\tilde{W}_{\mathcal{F}i}^T\tilde{W}_{\mathcal{F}i}\right) \\ &\quad-\tfrac{1}{2}\lambda_{S_{Ji}}^{\min}\sum_{i=1}^{n}\gamma_{ci}\left(\tilde{W}_{ci}^T\tilde{W}_{ci}+\tilde{W}_{ai}^T\tilde{W}_{ai}\right) \\ &\quad+\sum_{i=1}^{n}\Delta_i+\tfrac{1}{2}\bar{g}^2,\end{aligned} \tag{80}$$

where $\Delta_i=\frac{1}{2}(\gamma_{ci}+\gamma_{ai})\left(W_{Ji}^{*T}S_{Ji}\right)^2+\frac{1}{2}\sigma_iW_{\mathcal{F}i}^{*T}W_{\mathcal{F}i}^*+\frac{1}{2}\bar{\varepsilon}_{\mathcal{F}i}^2$.

From (80), one can obtain

$$\dot{V}\le-\sigma V+\Delta, \tag{81}$$

where $\sigma=-\min\{2c_i-1,\frac{\sigma_i}{\lambda_{\Gamma_i^{-1}}^{\max}},\gamma_{ci}\lambda_{S_{Ji}}^{\min}\}$ and $\Delta=\sum_{i=1}^{n}\Delta_i+\frac{1}{2}\bar{g}$.

Calculate the integral of (81), one can obtain

$$\begin{aligned}V &\le V(0)e^{-ct}+\tfrac{\Delta}{c} \\ &\le V(0)+\tfrac{\Delta}{c}.\end{aligned} \tag{82}$$

Based on the analysis provided above, it is evident that the tracking error z_i, and the weight error signals $\tilde{W}_{\mathcal{F}i}$, $\tilde{W}_{ci}$, $\tilde{W}_{ai}$ are bounded. Then, from (2), (30), (32)–(34), (47)–(50) and (64)–(67), it can be further inferred that $\hat{\alpha}_i^*$, $\hat{W}_{\mathcal{F}i}$, $\hat{W}_{ci}$, $\hat{W}_{ai}$ and u are bounded. According to the definition of z_i, it can be concluded that z_i is bounded. Consequently, the trajectories of all closed-loop signals in the control system are maintained within bounded regions in the semi-global sense.

According to (82), we have $|z_1| = |\mathcal{T}| \leq \sqrt{2\left(V(0)+\frac{\Delta}{c}\right)} = \diamondsuit$. From the inverse transformation of (17), we have $e^{\mathcal{T}} = \mho/(1-\mho)$. This leads to $\mho = e^{\mathcal{T}}/(1+e^{\mathcal{T}})$. Given $\mathcal{T} \leq \diamondsuit$, it follows that $0 < e^{-\mathcal{T}}/(1+e^{-\mathcal{T}}) \leq \mho \leq e^{\mathcal{T}}/(1+e^{\mathcal{T}}) < 1$. Using the relationship $\mho = \frac{\hbar-\underline{\mathfrak{K}}}{\overline{\mathfrak{K}}-\underline{\mathfrak{K}}}$, we can deduce $0 < \frac{\hbar-\underline{\mathfrak{K}}}{\overline{\mathfrak{K}}-\underline{\mathfrak{K}}} < 1$. Accordingly, $\underline{\mathfrak{K}} < \hbar < \overline{\mathfrak{K}}$, $\forall t \geq 0$. Then, from (11), (13) and (14), it can be concluded that $\underline{\mathcal{K}} < e < \overline{\mathcal{K}}$. Moreover, based on the definition of $\underline{\mathcal{K}}$, $\overline{\mathcal{K}}$, we know that $\underline{\mathcal{K}}$ and $\overline{\mathcal{K}}$ incorporate time-setting characteristics. This guarantees the system output is capable of tracking the reference signal before a predefined time, while the tracking error remains consistently confined within dynamic boundaries.

∎

IV. Simulation Validation

To validate the efficacy and superiority of the presented control algorithm, a one-link manipulator driven by an electric motor is modeled as

$$\begin{cases} D\ddot{\theta} + B\dot{\theta} + N\sin(\theta) = I, \\ M\dot{I} = -\mathcal{H}I - K_m\dot{\theta} + \mathrm{sat}(u), \end{cases} \tag{83}$$

where θ and $\dot{\theta}$ represent the angular position and velocity of the joint, respectively. I represents the motor current, and $\mathrm{sat}(u)$ denotes the saturated input, which is the system voltage.

By rewriting the variables as $x_1 = \theta$, $x_2 = \dot{\theta}$, $x_3 = I$, and considering the influence of external disturbances, the one-link manipulator system can be reformulated as follows:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = \frac{1}{D}x_3 - \frac{B}{D}x_2 - \frac{N}{D}\sin(x_1) + d_2(t), \\ \dot{x}_3 = \frac{1}{M}\mathrm{sat}(u) - \frac{\mathcal{H}}{M}x_3 - \frac{K_m}{M}x_2 + d_3(t), \end{cases} \tag{84}$$

where $D = 1$, $B = 1$, $M = 0.04$, $H = 0.5$, $N = 10$, and $K_m = 10$ with the specific units detailed in [44]. The saturation threshold is set as $u_d = 24$V.

The reference signal is chosen as $y_d = 1.2\sin(t) + 0.7\cos(\frac{1}{2}t)$, and the controller parameters are selected as $c_1 = 41$, $c_2 = 61$, $c_3 = 50$, $\Gamma_1 = 390I_{50\times50}$,$\sigma_1 = 1.0$, $\gamma_{c1} = 87$, $\gamma_{a1} = 70$, $\Gamma_2 = 900I_{50\times50}$, $\sigma_2 = 1.0$, $\gamma_{c2} = 93$, $\gamma_{a2} = 90$, $\Gamma_3 = 170I_{50\times50}$, $\sigma_3 = 1.0$, $\gamma_{c3} = 75$ and $\gamma_{a3} = 60$ with initial conditions $x_1 = 1.0$, $x_2 = 0$, $x_3 = 0$, $\hat{W}_{\mathcal{F}1}(0) = \hat{W}_{\mathcal{F}2}(0) = \hat{W}_{\mathcal{F}3}(0) = 0$, $\hat{W}_{c1} = [0.5, ..., 0.5] \in \mathbb{R}^{50\times1}$, $\hat{W}_{a1} = [6.2, ..., 6.2] \in \mathbb{R}^{50\times1}$, $\hat{W}_{c2} = [3.1, ..., 3.1] \in \mathbb{R}^{50\times1}$, $\hat{W}_{a2} = [8.3, ..., 8.3] \in \mathbb{R}^{50\times1}$, $\hat{W}_{c3} = [1.4, ..., 1.4] \in \mathbb{R}^{50\times1}$, $\hat{W}_{a3} = [7.0, ..., 7.0] \in \mathbb{R}^{50\times1}$. The membership functions of IT2FLSs are selected as $\overline{\mu}_{ij} = \exp\left(\frac{-(\chi_j-c_{ij})^2)}{2*7^2}\right)$, $\underline{\mu}_{ij} = \exp\left(\frac{-(\chi_j-c_{ij})^2)}{2*3^2}\right)$ with $c_{ij} = -3+i$, $i = 1, ..., 5$ and $j = 1, 2$. Moreover, the parameters of the SRPPC method are $\varpi_0 = 0.1$, $\varpi_f = 0.02$, $T_f = 2.0$, $a_\kappa = 1.0$, $\flat = 1.0$, $a_s = 1, 0$, $a_u = 0.14$, $a_l = 0.1$, $b_u = 0.2$, $b_l = 0.2$, $c_u = 1.5$, $c_l = 1.5$. For manipulators operating in application scenarios such as handling and sorting, the disturbance acting on the joint end is set as

$$d_2 = \begin{cases} 4.1\sin(t)\cos(0.7t) - 16.0, & 9 < t \leq 14 \\ 4.1\sin(t)\cos(0.7t), & t > 14. \end{cases}$$

And the disturbance of the motor end is set as $d_3 = 5.0\sin(t)$.

To testify the validity of the presented RL-SRPPC strategy, simulation experiments are carried out, whose simulation results are displayed in Figs. 2–6. Figs. 2 and 3 show the tracking trajectory and tracking error curves of the one-link manipulator system, respectively. It can be seen from Fig. 2 that the system output is capable of tracking the reference signal for the input-saturation-constrained system subjected to the external disturbances. Fig. 4 exhibits the control input of the one-link manipulator system. According to Figs. 3 and 4, it can be found that when the one-link manipulator system is subjected to the input saturation and disturbances, the proposed RL-SRPPC method self-regulates the PPBs to ensure that the

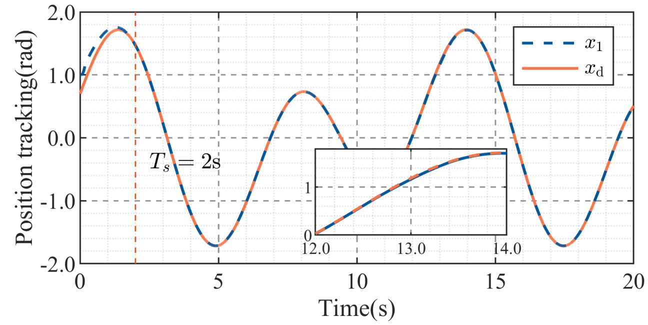


Fig. 2. The tracking trajectory of the system output.

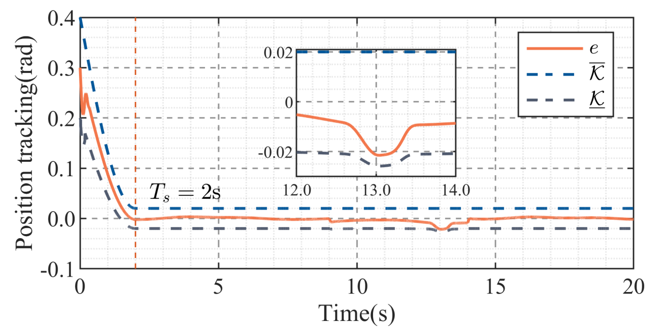


Fig. 3. The tracking error curve of the system output.

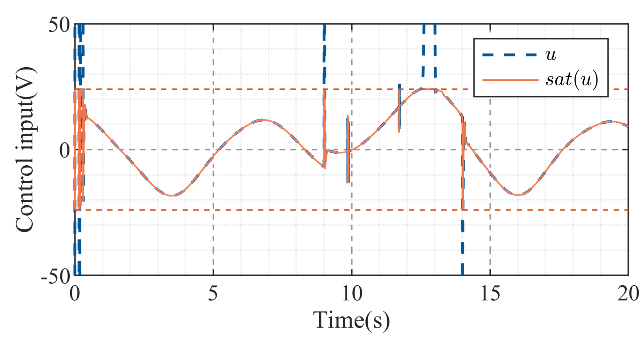


Fig. 4. The control input of the system.

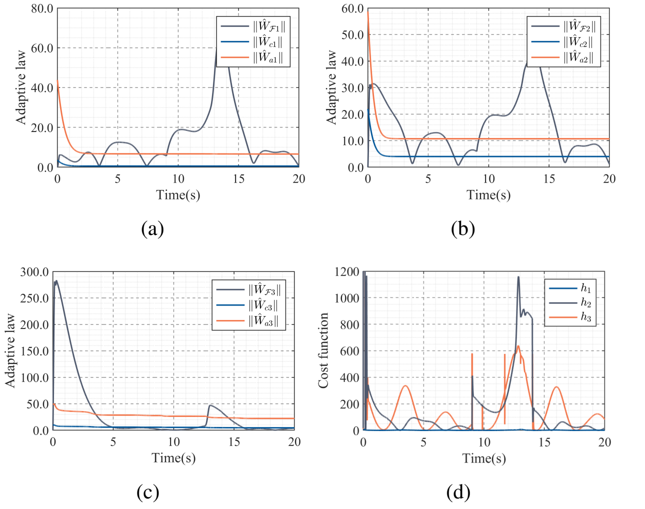


Fig. 5. (a), (b) and (c) represent the adaptive laws of the system; (d) represents the cost functions of the system.

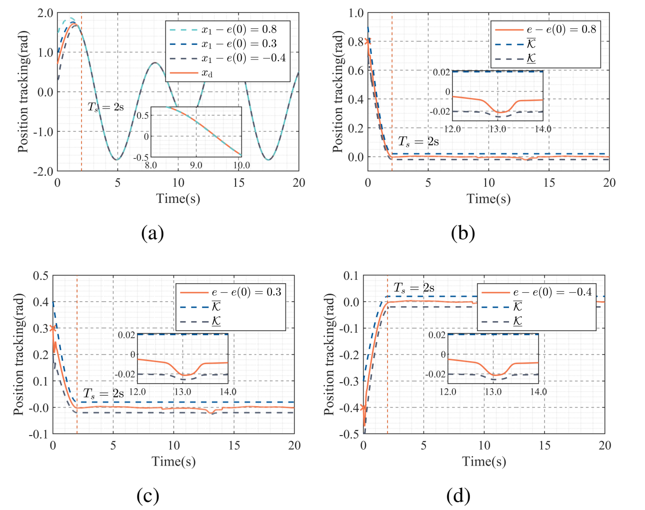


Fig. 6. (a) represents the tracking trajectories of the system output under distinct initial errors: $e(0) = 0.8$, $e(0) = -0.3$ and $e(0) = -0.4$; (b), (c) and (d) represent the error variations of the system output under distinct initial errors $e(0) = 0.8$, $e(0) = -0.3$ and $e(0) = -0.4$, respectively.

tracking error is always kept in the designed PPBs with no overshoot. Subfigures (a)-(c) in Fig. 5 illustrate the 2-norm curves of the adaptive laws $\|\hat{W}_{\mathcal{F}1}\|$, $\|\hat{W}_{c1}\|$, $\|\hat{W}_{a1}\|$, $\|\hat{W}_{\mathcal{F}2}\|$, $\|\hat{W}_{c2}\|$, $\|\hat{W}_{a2}\|$, $\|\hat{W}_{\mathcal{F}3}\|$, $\|\hat{W}_{c3}\|$ and $\|\hat{W}_{a3}\|$, indicating that the adaptive laws are bounded. The cost functions h_1, h_2, h_3 are depicted in Subfigure (d) in Fig. 5. Subfigures (a)-(d) in Fig. 6 illustrate the tracking trajectories and the evolution processes of the tracking error of the one-link manipulator system under different initial conditions: $e(0) = 0.8$, $e(0) = -0.3$ and $e(0) = -0.4$. Subfigure (a) in Fig. 6 shows that the system output can track the desired trajectory well under different initial errors. Subfigures (b)-(d) in Fig. 6 indicate that the tracking error remains in the PPBs under distinct initial error conditions. It is proved from Fig. 6 that the proposed RL-SRPPC method possesses a capacity for self-regulation, adaptively adjusting the initial PPBs to accommodate different initial conditions, and ensure the designed PPBs are still

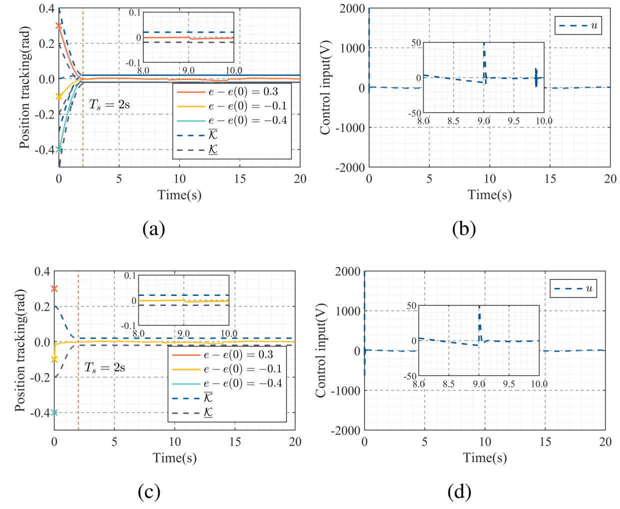


Fig. 7. (a) and (c), respectively, represent the error variation situations of the system output under the proposed method and the RL-based PPC method in [19] without input saturation; (b) and (d) represent the control input under the proposed method and the RL-based PPC method in [19] without input saturation.

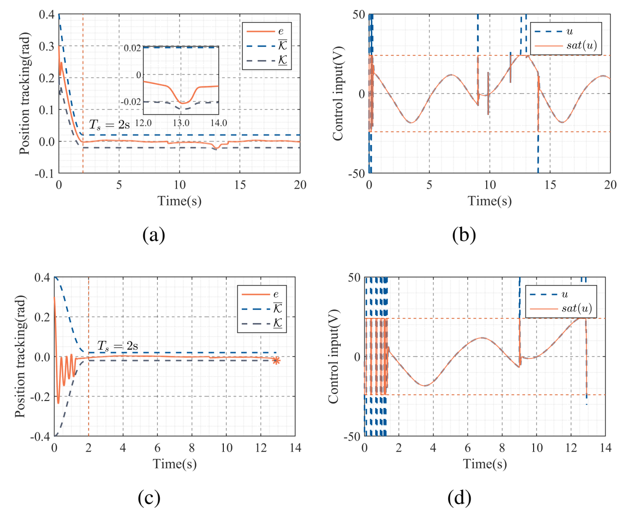


Fig. 8. (a) and (c), respectively, represent the error variation situations of the system output under the proposed method and the RL-based PPC method in [19] with considering input saturation; (b) and (d) represent the control input under the proposed method and the RL-based PPC method in [19] with considering input saturation.

capable of enveloping the tracking error when the initial error of the one-link manipulator varies.

To demonstrate the superiority of the proposed method, the existing RL-based PPC method [19] is utilized for making comparison, and the comparison results are shown in Figs. 7 and 8. Subgraphs (a), (c) and (b), (d) in Fig. 7 illustrate variation situations of the tracking error and the corresponding control inputs of the proposed method and the existing RL-based PPC method [19] under different initial conditions without input physical constraint. According to subfigures (a) and (c) in Fig. 7, it can be found that the existing RL-based PPC method [19] only adapts the initial error ($e(0) = -0.1$). In contrast, the proposed method performs self-tuning of the initial PPBs to distinct initial error values ($e(0) = 0.3$,

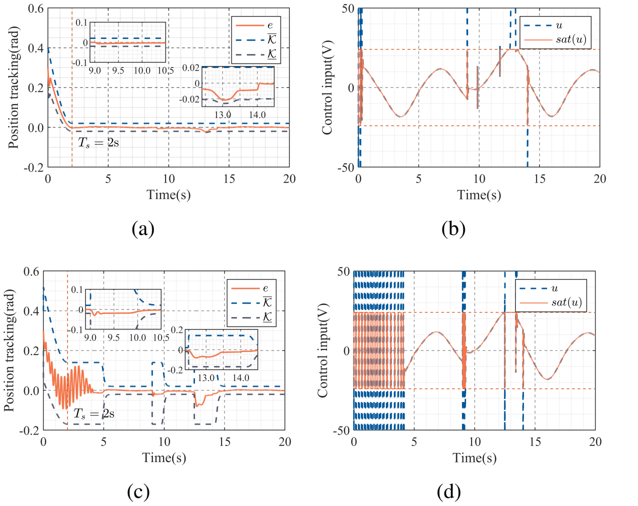


Fig. 9. (a) and (c), respectively, represent the error variation situations of the system output under the proposed method and the RL-based SFPPC method in [40]; (b) and (d), respectively, represent the control input under the proposed method and the RL-based SFPPC method in [40].

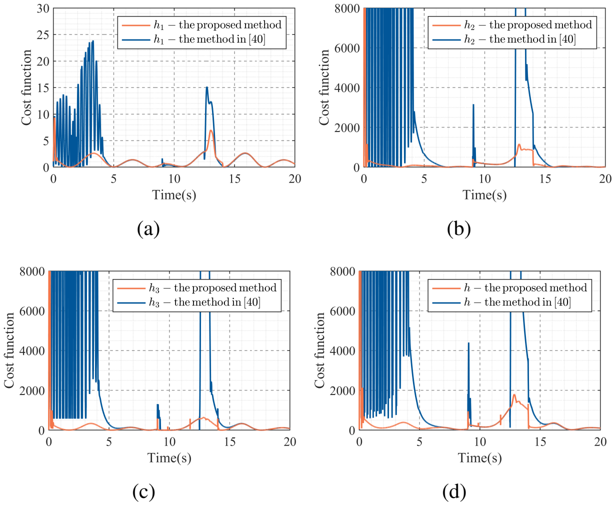


Fig. 10. (a), (b) and (c) stand for the cost functions of the subsystem under the proposed method and the RL-based SFPPC method in [40]; (d) stands for the total cost functions ($h = h_1 + h_2 + h_3$) of the one-link manipulator system under the proposed method and the RL-based SFPPC method in [40].

$e(0) = -0.1$, $e(0) = -0.4$) and ensures the running of the control system. Additionally, when the step disturbance occurs in 9s, the proposed method is capable of perceiving that the tracking error is going to violate the PPBs, quickly regulates the PPBs to envelop the tracking error and ensures the normal working of the one-link manipulator. While the control system remains functional under the existing RL-based PPC method [19], assuming no input saturation constraints in a practical system is unrealistic. Given the bounded nature of real-world control inputs, subfigures (a) and (c) in Fig. 8 exhibit the tracking error variation under the proposed method and the existing RL-based PPC method with considering the input magnitude constraint. The control input is shown in subfigures (b) and (d) in Fig. 8, which shows the occurrence of the input saturation. Obviously, although the initial tracking error meets the initial PPBs, the existing RL-based PPC method [19] fails when the input saturation appears, resulting in the operation termination of the one-link manipulator system. However, it can be seen from subfigures (a) and (b) in Fig. 8 that the presented strategy is still valid and can ensure the normal operation of the one-link manipulator system subjected to the disturbances and input saturation.

To further evaluate the optimization efficacy of the developed scheme, a comparative study is conducted between the proposed RL-SRPPC and the RL-based SFPPC method from [40]. The comparative results, depicted in Figs. 9 and 10, demonstrate that while both strategies can adaptively reconfigure the PPBs to maintain system stability under input saturation and external disturbances, the SRPPC mechanism provides a more precise and judicious boundary adjustment. This prevents the unwarranted boundary relaxation observed in SFPPC, thereby mitigating aggressive control input oscillations. Furthermore, Fig. 10 illustrates the evolution of the cost functions for the single-link manipulator, revealing that the proposed strategy achieves superior steady-state performance with significantly lower cumulative costs. To provide a rigorous quantitative assessment, metrics including the Integral of Absolute Error (IAE), Integral of Time-weighted Absolute Error (ITAE), Integral of Absolute Control Input (IACI), and Integral of Cost Function (ICF) are employed, defined as follows:

$$\begin{aligned}
\mathrm{IAE} &= \textstyle\int_0^t |e(\tau)| d\tau \\
\mathrm{ITAE} &= \textstyle\int_0^t t|e(\tau)| d\tau \\
\mathrm{IACI} &= \textstyle\int_0^t |\operatorname{sat}(u)| d\tau \\
\mathrm{ICF} &= \textstyle\int_0^t h(\tau) d\tau
\end{aligned}$$

where $h(t) = \sum_{i=1}^{n} h_i(t)$ stands for the total cost function with $n = 3$.

As shown in Table I, the quantitative comparison results in-

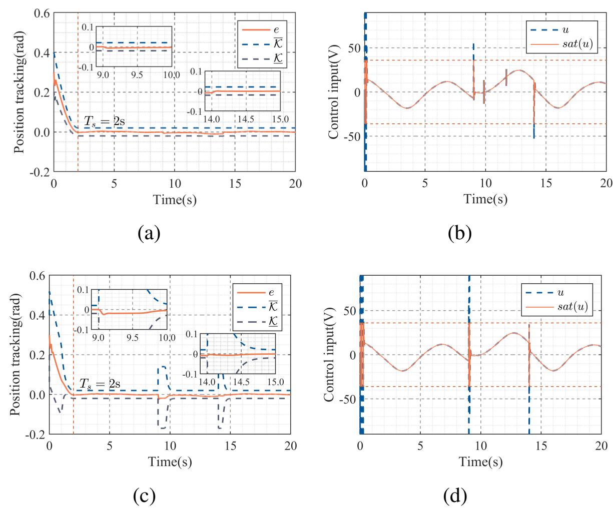


Fig. 11. (a) and (c) represent the error variation situations of the system output under the proposed method and the RL-based SFPPC method in [40], respectively, with input saturation $u_d = 36$V; (b) and (d) represent the control input under the proposed method and the RL-based SFPPC method in [40], respectively, with input saturation $u_d = 36$V.

TABLE I
THE QUANTITATIVE PERFORMANCE OF THE PROPOSED RL-SRPPC METHOD AND THE RL-BASED SFPPC METHOD IN [40]

	IAE	ITAE	IACI	ICF
The proposed method	0.2637	0.742	901.5	8931
The RL-based SFPPC method	0.4238	1.986	2.234×10^4	1.901×10^5

dicate that the presented RL-SRPPC strategy can acquire better optimization control performance than the RL-based SFPPC method. In other words, the proposed RL-SRPPC method is capable of acquiring prescribed tracking performance with a small control cost for the input-saturation system subjected to external disturbances. To further examine the robustness and versatility of the developed RL-SRPPC strategy, an alternative input saturation intensity ($u_d = 36$V) is implemented in addition to the baseline 24V scenario. As illustrated in the comparative results in Fig. 11, when input saturation occurs, the RL-based SFPPC method exhibits a premature and unwarranted expansion of the performance envelopes, leading to intensified system oscillations and performance degradation. Conversely, the proposed SRPPC mechanism successfully circumvents redundant boundary relaxation by explicitly considering the time-varying nature of disturbances and the physical constraints of the actuator. This consistency across varying saturation levels underscores the algorithm's capability to balance stringent performance specifications with limited control authority, providing a more reliable solution for practical robotic applications with different hardware limits.

Remark 7: The selection of control parameters follows a systematic tuning procedure that balances stability requirements with operational performance. To facilitate practical implementation, a "hierarchical tuning" rule-of-thumb is suggested: 1) Initialize the feedback gains c_i to establish basic tracking stability, where larger c_i accelerate convergence but increase sensitivity to noise; 2) Adjust the identifier gains Γ_i and drift correction terms σ_i to ensure robust dynamics estimation; these should be calibrated to capture rapid nonlinearities while suppressing high-frequency oscillations; 3) Fine-tune the Critic gains γ_{ci} and Actor gains γ_{ai} according to the "time-scale separation" principle, ensuring $\gamma_{ci} > \gamma_{ai}$ so the Critic can reliably evaluate the cost-to-go before the Actor refines the policy. Furthermore, the SRPPC framework enables intuitive parameterization; for instance, a target accuracy of ± 0.02mm within 2s can be straightforwardly achieved by setting $\varpi_0 = 0.02$ and $T_f = 2$s. All selections are strictly constrained by the stability regions in Theorem 1 to ensure uniform ultimate boundedness. Notably, the decentralized architecture ensures the scalability of the scheme to higher-dimensional MIMO systems by treating inter-subsystem couplings as augmented dynamics. While the current parameters are manually calibrated within feasible regions, the framework's structured nature allows for the future integration of meta-heuristic algorithms (e.g., Particle Swarm Optimization) to automate the optimization of these hyper-parameters across diverse operational scenarios.

V. CONCLUSION

In this work, an optimal control scheme is devised through the synergistic fusion of the RL technique with the SRPPC algorithm for a class of input-saturation systems subjected to external disturbances. The RL method based on the identifier–critic–actor architecture, cooperating with the IT2FLSs, is utilized to approximate the unknown nonlinear functions and optimize the control performance. To deal with the singularity issue of the traditional PPC technique, the SRPPC is designed to achieve the adjustment of the initial PPBs for different initial errors, and automatically extend the PPBs to encompass the tracking error when the input-saturation system is subjected to drastic external disturbances. Moreover, by incorporating reinforcement learning with the proposed SRPPC method, an RL-SRPPC optimal control scheme is devised, which simultaneously minimizes the control cost and guarantees both operational safety and adherence to prescribed performance specifications. Additionally, the numerical simulation using the one-link manipulator is carried out to prove the effectiveness and advantages of the presented strategy. Furthermore, considering the evolving landscape of intelligent control, future research will focus on extending the proposed RL-SRPPC framework by incorporating data-driven non-cooperative game theory and Q-learning mechanisms. Such an integration is expected to enhance the strategic resilience of the controller against sophisticated uncertainties and facilitate more efficient resource management in complex multi-player or adversarial environments.

REFERENCES

[1] P. Yang, S. Zhang, X. Yu, and W. He, "Reinforcement-learning-based finite time fault tolerant control for a manipulator with actuator faults," *IEEE Trans. Cybern.*, vol. 55, no. 6, pp. 2621–2632, Jun. 2025.

[2] K. Li, Y. Xia, R. K. Agarwal, and I. J. Rudas, "Error-triggered flexible prescribed performance control for free-flying space robot stochastic systems with actuator faults," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 5, pp. 13 748–13 758, Oct. 2025.

[3] Z. Liu, C. Wu, X. Shen, W. Yao, J. Liu, and L. Wu, "Adaptive interval type-2 fuzzy neural network-based novel fixed-time backstepping control for uncertain euler–lagrange systems," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 5, pp. 2966–2975, May 2024.

[4] Y. Xia, J. He, K. Li, F. Gao, and R. K. Agarwal, "Anti-saturation prescribed-time control for stochastic systems of free-flying space robots using a self-adapting non-monotonic approach," *Aerosp. Sci. Technol.*, vol. 162, p. 110231, Jul. 2025.

[5] X. Bu, "Prescribed performance control approaches, applications and challenges: A comprehensive survey," *Asian J. Control*, vol. 25, no. 1, pp. 241–261, Jan. 2023.

[6] Y. Xia, Z. Lendek, R. E. Precup, R. K. Agarwal, and I. J. Rudas, "Suction cup-type prescribed performance fault-tolerant fuzzy control for nonlinear systems considering actuator power," *IEEE Trans. Cybern.*, vol. 56, no. 1, pp. 347–356, Jan. 2026.

[7] Y. Xia, K. Xiao, R. E. Precup, I. J. Rudas, and R. K. Agarwal, "Power-considered fault-tolerant control for nonlinear systems with nonfragile prescribed performance," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 12, pp. 4394–4404, Dec. 2025.

[8] C. P. Bechlioulis and G. A. Rovithakis, "Robust adaptive control of feedback linearizable mimo nonlinear systems with prescribed performance," *IEEE Trans. Autom. Control*, vol. 53, no. 9, pp. 2090–2099, Oct. 2008.

[9] Y. Xia, K. Xiao, Y. Yao, Z. Geng, and Z. Lendek, "Fixed-time fuzzy vibration reduction for stochastic mems gyroscopes with low communication resources," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 8, pp. 4220–4233, Aug. 2024.

[10] Y. Xia, K. Xiao, J. Cao, R.-E. Precup, Y. Arya, H.-K. Lam, and L. Rutkowski, “Stochastic neural network control for stochastic nonlinear systems with quadratic local asymmetric prescribed performance,” *IEEE Trans. Cybern.*, vol. 55, no. 2, pp. 867–879, Feb. 2025.

[11] Y. Xia, C. Liu, Y. Tuo, and J. Li, “Command filter-based event-triggered control for stochastic mems gyroscopes with finite-time prescribed performance,” *ISA Trans.*, vol. 148, pp. 212–223, May 2024.

[12] K. Zhao, Y. Song, C. L. P. Chen, and L. Chen, “Adaptive asymptotic tracking with global performance for nonlinear systems with unknown control directions,” *IEEE Trans. Autom. Control*, vol. 67, no. 3, pp. 1566–1573, Mar. 2022.

[13] K. Zhao, F. L. Lewis, and L. Zhao, “Unifying performance specifications in tracking control of mimo nonlinear systems with actuation faults,” *Automatica*, vol. 155, p. 111102, Sep. 2023.

[14] Q. Bai, K. Zhao, and Y. D. Song, “Unified adaptive performance control of mimo input-quantized nonlinear systems,” *IEEE Transactions on Circuits and Systems I-Regular Papers*, vol. 71, no. 7, pp. 3331–3342, Jul. 2024.

[15] Y. Xia, K. Xiao, and Z. Geng, “Event-based adaptive fuzzy control for stochastic nonlinear systems with prescribed performance,” *Chaos, Solitons Fractals*, vol. 180, p. 114501, Mar. 2024.

[16] Y. Yao, Y. Kang, Y. Zhao, P. Li, and J. Tan, “A novel prescribed-time control approach of state-constrained high-order nonlinear systems,” *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 54, no. 5, pp. 2941–2951, Jan. 2024.

[17] Y. Zheng, Y. Xia, K. Li, H. Zhou, and R.-E. Precup, “Boundary-aware sliding prescribed performance control for pmsms under compound uncertainties,” *Expert Syst. Appl.*, vol. 304, p. 130786, Apr. 2026.

[18] Y. Xia, J. He, H.-K. Lam, L. Rutkowski, and R.-E. Precup, “Non-fragile fuzzy control of input-saturated systems with global prescribed performance via an error-triggered mechanism,” *Inf. Sci.*, vol. 711, p. 122111, Sep. 2025.

[19] 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 Trans. Cybern.*, vol. 54, no. 11, pp. 6345–6357, Nov. 2024.

[20] Y. Yao, Y. Kang, Y. Zhao, P. Li, and J. Tan, “Flexible prescribed performance output feedback control for nonlinear systems with input saturation,” *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 11, pp. 6012–6022, Nov. 2024.

[21] R. Luo, G. He, X. Bu, and J. Wu, “Waverider vehicles flight control using flexible prescribed envelopes,” *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 10, pp. 14 722–14 734, Oct. 2025.

[22] Y. Yao, Y. Kang, Y. Zhao, J. Tan, L. Gu, and G. Shi, “Sliding flexible prescribed performance control for input saturated nonlinear systems,” *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 4, pp. 1241–1252, Apr. 2025.

[23] X. Bu, J. H. Park, and H. Lei, “Discrete-time neural control with flexible prescribed performance for constrained systems with fragility relaxation,” *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 2386–2394, Jan. 2025.

[24] K. Yong, M. Chen, Y. Shi, and Q. Wu, “Flexible performance-based robust control for a class of nonlinear systems with input saturation,” *Automatica*, vol. 122, p. 109268, Dec. 2020.

[25] G. Wen, S. S. Ge, and F. Tu, “Optimized backstepping for tracking control of strict-feedback systems,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 8, pp. 3850–3862, Aug. 2018.

[26] D. Zhang, Q. Yuan, L. Meng, R. Xia, W. Liu, and C. Qin, “Reinforcement learning for single-agent to multi-agent systems: from basic theory to industrial application progress, a survey,” *Artif. Intell. Rev.*, vol. 59, no. 2, p. 46, Nov. 2025.

[27] S. Cao, L. Sun, J. Jiang, and Z. Zuo, “Reinforcement learning-based fixed-time trajectory tracking control for uncertain robotic manipulators with input saturation,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 8, pp. 4584–4595, Aug. 2023.

[28] C. Qin, M. Pang, Z. Wang, S. Hou, and D. Zhang, “Observer based fault tolerant control design for saturated nonlinear systems with full state constraints via a novel event-triggered mechanism,” *Eng. Appl. Artif. Intell.*, vol. 161, p. 112221, Dec. 2025.

[29] C. Qin, S. Hou, M. Pang, Z. Wang, and D. Zhang, “Reinforcement learning-based secure tracking control for nonlinear interconnected systems: An event-triggered solution approach,” *Eng. Appl. Artif. Intell.*, vol. 161, p. 112243, Dec. 2025.

[30] C. Qin, M. Pang, Z. Wang, S. Hou, and D. Zhang, “Event-triggered zero-sum game for input saturated nonlinear systems with state constraints,” *IEEE Trans. Consum. Electron.*, vol. 72, no. 1, pp. 455–465, Feb. 2026.

[31] G. Wen, C. L. P. Chen, and S. S. Ge, “Simplified optimized backstepping control for a class of nonlinear strict-feedback systems with unknown dynamic functions,” *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4567–4580, Sep. 2021.

[32] G. Wen and C. L. P. Chen, “Optimized backstepping consensus control using reinforcement learning for a class of nonlinear strict-feedback-dynamic multi-agent systems,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 3, pp. 1524–1536, Mar. 2023.

[33] H.-Y. Zhu, Y.-X. Li, and S. Tong, “Dynamic event-triggered reinforcement learning control of stochastic nonlinear systems,” *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 9, pp. 2917–2928, Sep. 2023.

[34] L. Gao, Z. Wang, X. Huang, and H. Shen, “Predefined-time event-triggered adaptive tracking control for uncertain time delay nonlinear systems,” *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 15 129–15 138, May 2025.

[35] S. Xie, W. Sun, Y. Sun, and S.-F. Su, “Adaptive prescribed-time optimal control for flexible-joint robots via reinforcement learning,” *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 55, no. 4, pp. 2633–2643, Apr. 2025.

[36] 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 Trans. Cybern.*, vol. 53, no. 11, pp. 7380–7391, Nov. 2023.

[37] C. Yan, J. Xia, J. H. Park, and X. Xie, “Reinforcement learning-based adaptive event-triggered fuzzy control for cyclic switched stochastic nonlinear systems with actuator faults,” *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 3, pp. 1131–1143, Mar. 2024.

[38] Y. Li, K. Li, and S. Tong, “Reinforcement learning-based adaptive finite-time performance constraint control for nonlinear systems,” *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 54, no. 2, pp. 1335–1344, Feb. 2024.

[39] Q. Wan, T. Fan, Z. Li, and Z. Ding, “Reinforcement learning-based fixed-time tracking control for nonlinear systems with asymmetrical guaranteed performance,” *Neurocomputing*, vol. 644, p. 130410, Sep. 2025.

[40] Y. Yao, Z. Liu, Y. Kang, Y. Zhao, J. Tan, L. Gu, Q. Li, and J. Wang, “Sliding flexible prescribed performance boundary-guided reinforcement learning control for input-constrained nonlinear systems,” *IEEE Trans. Cybern.*, vol. 56, no. 2, pp. 968–979, Feb. 2025.

[41] W. Zhao, J. Wang, S. Lv, L. Su, and H. Shen, “Reinforcement learning-based predefined-time tracking control for input-saturated nonlinear systems with performance guarantees,” *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 19 876–19 888, Aug. 2025.

[42] X. Zhao, Y. Sun, W. Sun, J. Xu, and G. Lin, “Safe reinforcement learning control for maglev train levitation system with stability guarantee and safety constraint,” *IEEE Trans. Cybern.*, vol. 12, no. 1, pp. 285–296, Feb. 2026.

[43] Y. Yao, X. Liang, Y. Kang, Y. Zhao, J. Tan, and L. Gu, “Dual flexible prescribed performance control of input saturated high-order nonlinear systems,” *IEEE Trans. Cybern.*, vol. 55, no. 3, pp. 1147–1158, Mar. 2025.

[44] T. S. Li, D. Wang, G. Feng, and S. C. Tong, “A dsc approach to robust adaptive nn tracking control for strict-feedback nonlinear systems,” *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 40, no. 3, pp. 915–927, Nov. 2010.

Kaixin Li received the B.S. degree in Mechanical Design, Manufacturing, and Automation from University of South China, Hengyang, China, in 2020, and the M.S. degree in Mechanical Engineering from Chongqing University, Chongqing, China, in 2023. He is currently pursuing the Ph.D. degree in Mechanical Engineering with Chongqing University, Chongqing, China. His research interests include robotic dynamic control, prescribed performance control, stochastic system control, and reinforcement learning control.

Ke Xiao received the Ph.D. degree in mechanical engineering from Chongqing University, Chongqing, China, in 2013. He is currently an Associate Professor with the College of Mechanical and Vehicle Engineering, Chongqing University. His research interests include special robots and electromechanical drive systems.

Yu Xia (Member IEEE) received his Ph.D. degree in mechanical engineering from Chongqing University, Chongqing, China, in 2024. He is currently a Research Fellow in the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include constrained control, optimal control, reinforcement learning control, and the application of advanced control algorithms in electromechanical systems.

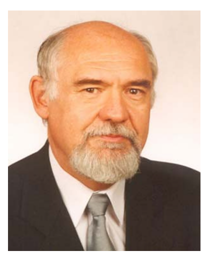
Zsófia Lendek (Member, IEEE) received the M.Sc. degree in control engineering from the Technical University of Cluj-Napoca, Cluj-Napoca, Romania, in 2003, the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2009, and the habilitation degree from the Technical University of Cluj-Napoca, in 2019.

She is currently a Full Professor with the Technical University of Cluj-Napoca. She has previously held research positions in the Netherlands and in France. Her research interests include observer and controller design for nonlinear systems, in particular Takagi–Sugeno fuzzy systems, which adapt the advantages of linear time-invariant system-based design to nonlinear systems, with applications in several fields.

Prof. Lendek is an Associate Editor for IEEE TRANSACTIONS ON FUZZY SYSTEMS and *Engineering Applications of Artificial Intelligence* and an Editorial Board Member of *Fuzzy Sets and Systems*.

Radu-Emil Precup (Fellow, IEEE) received the Dipl.Ing. (Hons.) degree in automation and computers from the Traian Vuia Polytechnic Institute of Timisoara, Timisoara, Romania, in 1987, the Diploma degree in mathematics from the West University of Timisoara, Timisoara, in 1993, and the Ph.D. degree in automatic systems from the Politehnica University of Timisoara (UPT), Timisoara, in 1996. He is currently with UPT, where he became a Professor with the Department of Automation and Applied Informatics in 2000. From 2022, he is also a Senior Researcher (CS I) and the Head of the Data Science and Engineering Laboratory, Center for Fundamental and Advanced Technical Research, Romanian Academy–Timisoara Branch, Timisoara. From 2016 to 2022, he was an Adjunct Professor with the School of Engineering, Edith Cowan University, Joondalup, WA, Australia. He is the author or co-author of more than 300 papers. His current research interests include intelligent and data-driven control systems. Prof. Precup is a corresponding member of the Romanian Academy.

He is a member of the Technical Committees on Data-Driven Control and Monitoring, and Control, Robotics and Mechatronics of the IEEE Industrial Electronics Society. He is the Editor-in-Chief of Romanian Journal of Information Science and Technology, a Senior Editor of IEEE Open Journal of the Computer Society, and an Associate Editor of IEEE TRANSACTIONS ON CYBERNETICS and IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.



Imre J. Rudas (Life Fellow, IEEE) received the bachelor's degree in mechanical engineering from Bànki Donát Polytechnic College, Budapest, Hungary, in 1971. He received the master's degree in mathematics from Eötvös Loránd University, Budapest, Hungary, in 1977, the Ph.D. degree in robotics and the Doctor of Science degree in computer science from the Hungarian Academy of Sciences, Budapest, in 1987 and 2004, respectively, and the Doctor Honoris Causa degrees in computer science from the Technical University of Košice, Košice, Slovakia, in 2001; the Politehnica University of Timișoara, Timișoara, Romania, in 2005; Óbuda University, Budapest, Hungary, in 2014; and the Slovak University of Technology in Bratislava, Bratislava, Slovakia, in 2016. He is a Rudolf Kalman Distinguished Professor, a Rector Emeritus, and a Professor Emeritus with Óbuda University, Budapest. He has edited and/or published 22 books, published more than 890 papers in international scientific journals, conference proceedings, and book chapters, and received more than 8000 citations. His present areas of research activities are Computational Cybernetics, Cyber Physical Control, Robotics, and Systems of Systems.

Dr. Rudas was awarded the Honorary Professor Title in 2013 and the Ambassador Title by the Wrocław University of Science and Technology. He is the Past President of the IEEE Systems, Man, and Cybernetics Society. He is a Fellow of the International Fuzzy Systems Association.