Reinforcement Learning based Optimal Control for Constrained Nonlinear System via A Novel State-Dependent Transformation

Lei Yan · Zhi Liu · C. L. Philip Chen · Yun Zhang · Zongze Wu

Abstract

This paper focus on developing an optimal controller for the strict-feedback nonlinear systems with or without asymmetric time-varying full state constraints. A novel nonlinear state-dependent transformation function is presented, by which the strict-feedback nonlinear systems with state constraints is transformed into a new strict-feedback where the state constraints is implicit in. Optimized backstepping technique is utilized to develop the optimal controller for the new strict-feedback system to track the desired reference signal without the feasibility conditions. Reinforcement learning (RL) is exploited to implement the optimal control in every step, where identifier, critic and action network are used to estimate the unknown system dynamics and generate the control output, respectively. It is theoretically proved that all the signals in the close loop system are bounded and the proposed optimal controller can track the desired signal with or without time-varying asymmetric full state constraints. Two simulation ex-

Lei Yan · Zhi Liu · Yun Zhang · Zongze Wu
School of Automation, Guangdong University of Technology, Guangzhou , Guangdong
E-mail: lz@gdut.edu.cn

C. L. Philip Chen
Faculty of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China
E-mail: philip.chen@ieee.org

Lei Yan
yanleirex@163.com
Yun Zhang
yz@gdut.edu.cn
Zongze Wu
zzwu@gdut.edu.cn

amples are presented demonstrating the efficacy of the proposed scheme.

1 Introduction

Most practical systems are subject to various forms of constraints, due to physical limitation, system performance requirement or security consideration, such as autonomous vehicles [1,2] and robotic systems[3,4], which makes handling constraints an important area of research in the domain of control design[5,6]. If not properly accommodated those constraints, it may lead to the inaccuracy of control, system instability, and sometimes unexpected accidents, making the elementary constrained control issues of nonlinear dynamic systems extremely crucial and competitive..

There have been various approaches in the literature to address the state constraints, such as reference governors [7], set invariance[8] and model predictive control[9]. It is noted that, there have been a large array of improvements achieved in the last few years by utilizing Barrier Lyapunov Function(BLF) or integral Barrier Lyapunov Function(iBLF) in addressing output or state constraint of nonlinear system, see [10–14]. In [10] and [15] by Tee et al., the definition of BLF is given and both symmetric and asymmetric Barrier Lyapunov Function is used to address state constraints for nonlinear systems, which state constraint is time-invariant. Full state constraints in strict-feedback system is considered by proposed an adaptive neural nwtwork control scheme by Liu and Tong et al. in [16]. Then, nonlinear pure-feedback system and stochastic with full state

constraints also considered using adaptive control technique by Liu and Tong in [17,12]. In [13], Li et al. designed a time-varying asymmetric BLF candidate to cope with the time-varying full-state constraints. By developing a adaptive fuzzy controller, Sun et al. tackled a class of nontriangular structural stochastic switched nonlinear systems with full state constraints based on Barrier Lyapunov Function[18]. In [14], the Barrier Lyapunov functions are constructed to ensure the constraints are not transgressed for a class of uncertain nonlinear systems with full-state constraints. A high-order tan-type barrier Lyapunov function (BLF) is constructed to handle the full-state constraints by Sun et al. in [19]. However, the current adaptive backstepping method based on BLF (or iBLF) has to fulfill the feasibility condition of no violation by the virtual controller, which means that the virtual controller must meet a prespecified constraint interval[20]. Feasibility condition make it more difficulty for strict-feedback and pure-feedback nonlinear systems with state constraints. An offline parameter optimization method is employed to satisfy the feasibility conditions of the virtual controller at each step, bringing computational costs and complex design procedures[21].

Therefore, new techniques that do not require the employment of BLF need to be formulated to overcome the constraint problem and thus prevent the feasibility condition. Fortunately, some state mapping or transformation based methodologies have been proposed to solve the output or state state constraint issues[22,23,20,24–26]. In [22] and [23], nonlinear mapping and one-to-one nonlinear mapping were proposed to transformed the original system into a new system without constraint, respectively. However, the drawback of this type of method is that when the constraint is asymmetric, the mapped variable is not zero when the state is zero. That is, this type of method is not able to handle asymmetric constraints well. Zhao and Song constructed a nonlinear state-dependent function which depends only on constrained states and addressed asymmetric full state constraints directly in [20]. In [27], Zhao and Song et al. expanded their work to handle on constraints without any additional work. Cao and Song et al. [28] use the same technique proposed a robust control scheme to deal with full states asymmetric and time-varying constraints for pure-feedback systems. Li and Liu et al. [29] using a natural logarithmic type nonlinear mapping removed feasibility conditions for nonlinear stochastic system. Liu and Zhang et al. [25] introduced a nonlinear state-dependent function to prevent asymmetric time-varying full state constraints for nonstrict-feedback nonlinear systems. A new general constraint function is introduced for uncertain pure

feedback systems with uniform consideration of the case with or without state constraints by Cao and Wen in [21]. Yao and Tan et al. developed an adaptive fuzzy control for constrained stochastic nonlinear systems by using a nonlinear state-dependent transformation in [30] However, none of the above mentioned methods considering the optimization performance of the controller, especially in using adaptive backstepping techniques.

Optimal control, a much talked philosophy in control theory and engineering in recently years, focuses on optimizing costs to achieve maximum control performance[31]. In theoretical view, it can be obtained the optimal controller by solving the Hamilton–Jacobi–Bellman (HJB) equation, however, this nonlinear equation is very difficult to solve due to the inherent nonlinearity [32,33]. Adaptive Dynamic Programming (ADP) or Reinforcement Learning (RL) is a prospective means of addressing the solution fo HJB equation [32,33,31,34]. Adaptive backstepping control combined with ADP or RL is a way to develop optimal controller for strict-feedback system, which can be divided into two categories. For the first category, equivalent optimal regulation problem is transformed by modifying the standard backstepping technique to address the optimal problem [35–37]. Recently, a new optimized backstepping method was developed in [38–41], whose core idea is to use ADP or RL for optimization at each step of the backstepping design procedure. In [38], Wang and Liu et al. proposed an equivalent optimal controller, which is obtained by Sontag feedback formula, for strict feedback system and borden the backstepping technique. Wen et al. [39] implemented reinforcement learning algorithm of the identifier–actor–critic architecture based on fuzzy logic system (FLS) approximators for multi-agent system with unknown nonlinear dynamics. Then, Wen developed optimized backstepping scheme for a class of strict-feedback systems in [42]. In [43,40], Wen and Liu et al. developed a simplified optimized backstepping control scheme for a class of nonlinear strict-feedback system with unknown dynamic and perturbed nonlinear systems , respectively. In [44,45], optimized backstepping controll scheme have been extended into multi-agent system. Nevertheless, none of the above mentioned literature considered state constraints, not to mention time-varying asymmetric state constraints. Although [46,47] considered state constraints, they assume that the state constraints are known constants and use a barrier optimization performance function to avoid these constraints, which measn that those methods have no ability to address time-varying asymmetric full state constraints.

Motivated by the above literature and discussion, this paper concentrates on the issue of optimized back-

stepping control of strict-feedback systems with full-state constraints which is time-varying asymmetric. First, a novel nonlinear state-dependent function (NSDF) is developed and used to transform the original strict-feedback system with time-varying asymmetric state constraints into a new system without constraints. Then, simplified optimized backstepping control scheme is utilized to developed tracking controller for the new system, in which reinforcement learning algorithms containing identifiers, critic networks and action networks are deployed to acquire optimal controllers. Our contributions are outlined as follows

1. In contrast to the typical BLF (iBLF) based approach addressing the state constraints, which exists a feasibility condition that the virtual controller should satisfy, whereby a novel NSDF is formulated to directly address the time-varying asymmetric full-state constraints without the feasibility condition. Moreover, it has the ability to cope with the cases with or without state constraints simultaneously and the steady state tracking error is not affected by the proposed novel NSDF.
2. Compared to the existing literature addressing state constraints, in which backstepping technique is utilized, in this paper, an optimal controller is developed for the system with time-varying asymmetric full-state constraints based on optimized backstepping technique. In addition, reinforcement learning algorithms are used to yield optimal controller, and the restriction of the persistence excitation condition is released.

The remainder of the article is organized as follows. In Section 2, formulation of the problem and a brief fundamentals is given. The novel nonlinear state-dependent function and the optimal controller design procedure is presented in Section 3. The stability and performance analysis of the proposed scheme is showed in Section 4. In Section 5, two simulation examples is demonstrated. Finally, the conclusion is summarized in Section 6.

2 Problem Description and Preliminaries

2.1 Problem Description

A class of nonlinear strict-feedback system with time-varying asymmetric full-state constraints is considered.

The nonlinear system is described as:

$$
\begin{aligned}
\dot{x}_1 &= f_1(\bar{x}_1) + x_2 \\
\dot{x}_2 &= f_2(\bar{x}_2) + x_3 \\
&\cdots \\
\dot{x}_n &= f_n(\bar{x}_n) + u \\
y_1 &= x_1
\end{aligned} \tag{1}
$$

where $\bar{x}_i = (x_1, \ldots, x_i)^T \in \mathcal{R}^i, \bar{x}_n = (x_1, \ldots, x_n) \in \mathcal{R}^n$ are the state vectors of system. u is system input and $y = x_1$ is output of the system. And $f_i(\bar{x}_i)$ is assumed to be unknown. The state variable x_i is subjected to a asymmetric time-varying constraints and the constraint boundaries $F_{i1}(t)$ and $F_{i2}(t)$, i.e. $F_{i1}(t) < x_i(t) < F_{i2}$. And we assume that $F_{i1}(0) < x_i(0) < F_{i2}(0)$.

The control objectives of this article is to design a optimal controller for the system (1) with asymmetric time-varying state constraints such that:

1. All the signals in the closed-loop systems are bounded.
2. System output y can track the desired signal y_r and under the conditions that all states are subjected to an asymmetric time-varying constraints.

The following assumptions are made in order to realize the above control objectives, .

Assumption 1 *The desired signal y_d is continuous and its first time derivative $\dot{y}_d$ is bounded and available. Moreover, reference signal y_r should satisfy that $F_{r_1} < y_r < F_{r_2}$, where $F_{r_1} = F_{11} + \underline{F}_0$ and $F_{r_2} = F_{12} - \overline{F}_0$, $\underline{F}_0$ and $\overline{F}_0$ are positive constant, F_{11} and F_{12} are the lower and upper boundary of x_1.*

Assumption 2 *The time varying boundary $F_{i1}(t)$ and $F_{i2}(t)$ are smooth and its derivatives are bounded and continuous.*

Remark 1 Assumption 1 and 2 have been commonly employed in the literature for handling full-state constraints, such as [21, 25, 48]. Assumption 1 indicates that the upper and lower bounds of the desired reference signal are slightly smaller than the constraint on the system state x, that is to say, the desired signal cannot be surpassed the range of the state constraints. The purpose of assumption 2 is to transform the desired signal into a new reference variable with the proposed state-dependent transformation function and applied to the controller design, as detailed in section 3.2.

2.2 Neural Network

The radial basis function neural network (RBFNN) $\mathbf{w}^{*T}\mathcal{S}(\mathcal{X})$ is used to approximate uncertainty, like [49]. Suppose $f(\mathcal{X})$ is a continuous function defined on a

compact set, for any given constant ε_0, there exists a constant vector $\mathbf{w}$, such that the following equation holds

$$f(\mathcal{X}) = \mathbf{w}^{*T} S(\mathcal{X}) + \varepsilon_0 \tag{2}$$

where $\mathbf{w}$ is the ideal weight, ε_0 denotes the approximation error which satisfy $\varepsilon_0 \leq \bar{\epsilon}$, the input vector of RBFNN is denoted by $\mathcal{X} = [\mathcal{X}_1, \ldots, \mathcal{X}_m]^T \in \mathcal{R}^\ell$ and the dimension of the input denoted as ℓ. The ideal weight vector is described as

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathcal{R}^m} \left\{ \sup_{X \in \mathcal{R}^\ell} \left| \mathbf{w}^T \mathcal{S}(\mathcal{X}) - f(\mathcal{X}) \right| \right\} \tag{3}$$

where $S(\mathcal{X}) = [s_1(\mathcal{X}), \ldots, s_m(\mathcal{X})]^T \in \mathcal{R}^m$. We chose Gaussian function as the basis function, defined as

$$s_j(\mathcal{X}) = \exp \left[\frac{-(\mathcal{X} - \pi_j)^T (\mathcal{X} - \pi_j)}{\bar{\gamma}_j^2} \right] \tag{4}$$

where $\pi_j \in \mathcal{R}^\ell, j = 1, \ldots, m$ is the center of the basis function and m is the number of the hidden layers. $\bar{\gamma}_j$ is the width of Gaussian function.

2.3 Optimal Control Formulation

Consider a class of affine nonlinear continuous-time systems

$$\dot{x} = f(x) + g(x)u(x) \tag{5}$$

where $x \in \mathcal{R}^n$ is the state vector and $u(x) \in \mathcal{R}^m$ is the control input of system (5), respectively. System function $f(x) \in \mathcal{R}^n$ and $g(x) \in \mathcal{R}^{n \times m}$ are continuous function and $f(0) = 0$. And $f(x) + g(x)u(x)$ is Lipschitz continuous on the set Ω belonging to R_n and containing the origin. We assume that system (5) is stabilizable.

For the optimal control problem on finite time domain, define utility function as

$$r(x(t), u(x)) = Q(x(t)) + u^T(x)Ru(x) \tag{6}$$

where $Q(x(t)) \geq 0$, and $R = R^T > 0$ is a square matrix with dimension m. The cost function is defined as

$$\mathcal{J}(x(t), u(x)) = \int_t^\infty r(x(\tau), u(x(\tau))d\tau \tag{7}$$

Definition 1 (Admissible Control): The control $u(t)$ is said admissible with respect to the cost function on a compact set $\Omega \in \mathcal{R}^n$ if $u(x)$ is continuous on Ω, $u(0) = 0$, $u(x)$ stabilizes the systems on Ω and $\forall x_0 \in \Omega, \mathcal{J}(x_0)$ is finite.

Definte $u^*(x)$ as the optimal controller, then the cost function is described as

$$\mathcal{J}^*(x) = \min_{u \in \Psi(\Omega)} \left(\int_t^\infty r(x(\tau), u(x)) d\tau \right)$$
$$= \int_t^\infty r(x(\tau), u^*(x)) d\tau \tag{8}$$

Tacking the time derivative of (8) on both side, yielding HJB equation described as

$$\mathcal{H}(x, u^*, \mathcal{J}_x^* = u^{*T}Ru^* + Q(x) + \mathcal{J}_x^*(x)^T \dot{x} = 0 \tag{9}$$

where $\mathcal{J}_x^*(x) = \partial \mathcal{J}^*(x)/\partial x$.

The optimal state feedback control law can be obtained

$$u^*(x) = -\frac{1}{2}R^{-1}g^T(x)\mathcal{J}_x^*(x) \tag{10}$$

Substituting (10) into (9), we obtain

$$\mathcal{H}(x, u^*, \mathcal{J}_x^*) = Q(x) + \mathcal{J}_x^{*T}(x)f(x)$$
$$- \frac{1}{4}\mathcal{J}_x^{*T}(x)g(x)R^{-1}g^T(x)\mathcal{J}_x^{*T}(x) \tag{11}$$
$$= 0$$

Due to the inherent nonlinear and uncertain terms of HJB equation, it is hard to acquire a exclusive optimized controller by seeking a solution directly, and ADP or RL will be used to solve the equation.

3 Novel State-dependent Function and Optimal Backstepping Controller Design

3.1 Novel State-dependent Function

A novel NSDF is introduced to tackle the time-varying state constraints.

Firstly, we construct the following functions to denote the difference of state $x(t)$ and boundaries $F_1(t)$ and $F_2(t)$ of constraint,

$$d(x, F_1, F_2) = e^{-\frac{1}{\delta(F_2 - x)(x - F_1)}} \tag{12}$$

where $\delta > 0$ is a scale parameter. For simplicity, define $d_1(x, F_1) = \frac{1}{(x - F_1)}, d_2(x, F_2) = \frac{1}{(F_2 - x)}$, then $d(x, F_1, F_2)$ is described as

$$d(x, F_1, F_2) = e^{-\frac{d_1 d_2}{\delta}} \tag{13}$$

where $\dot{d}_1(x, F_1) = \frac{\dot{F}_1 - \dot{x}}{d_1^2}, \dot{d}_2(x, F_2) = \frac{\dot{x} - \dot{F}_2}{d_2^2}, d(x, F_1, F_2) \in (0, e^{-\frac{4}{\delta(F_2 - F_1)^2}})$ and $F_2 - F_1 \in (0, +\infty)$.

Now, to tackle the all-state constraint, we define a new constrained variable $\xi(x)$ and the novel state-dependent transformation is described as

$$\xi(x) = \frac{x}{d(x, F_1, F_2)} \tag{14}$$

(a) State-dependent transformation function. (A:Proposed, B:[27], C:[21], D:[23, 29], E:[20, 48, 25])

(b) $d(x, F_1, F_2)$ and its lower and upper boundary, (Black: $\delta = 0.05$, Blue: $\delta = 0.1$, Red: $\delta = 2$,)

Fig. 1 Schematic illustration of the relationship between the new state $\xi(x)$ and the state x as well as the boundary of $d(x, F_1, F_2)$.

Then, the time derivative of the new constrained variable is described as

$$\dot{\xi}(x) = \mu \dot{x} + \nu \tag{15}$$

where

$$
\begin{aligned}
\mu &= \frac{1}{d(x, F_1, F_2)} - \frac{x}{\delta d(x, F_1, F_2)}\left(\frac{d_2}{d_1^2} - \frac{d_1}{d_2^2}\right) \\
\nu &= \frac{x}{\delta d(x, F_1, F_2)}\left(\frac{d_2}{d_1^2}\dot{F}_1 - \frac{d_1}{d_2^2}\dot{F}_2\right)
\end{aligned}
\tag{16}
$$

The novel state-dependent variable has some properties as follows:

1. $\xi(x) = 0$ if and only if $x = 0$;
2. $x(t) \to F_1(t)$ or $x(t) \to F_2(t)$, $\xi(x) \to \infty$;
3. $F_1(t) \to -\infty$ and $F_2(t) \to \infty$, $d(x, F_1, F_2) = 1$, $\xi(x) = x$.
4. $d(x, F_1, F_2)_{\max} \to 1$ as $\delta \to \infty$, $d(x, F_1, F_2)_{\max} \to 0$ as $\delta \to 0$.

Remark 2 It should be noted that the four properties of the new state-dependent variables encompass all cases of asymmetric state constraints.

– *Case 1:* When the state x is 0, the new state-dependent variable is also 0, while when x is not equal to 0, it is impossible for the new state dependent variable to be 0.
– *Case 2:* As the state variables approach the lower or upper boundary, the new state-dependent variables rapidly converge to infinity, which enforces the controller to pull the state back away from the constraint boundary.
– *Case 3:* Specifically, the proposed new state-dependent transformation function has the ability to handle not only the state constraint but also the case of no state constraint, which indicates that the new state-dependent variables return to the original state when the state constraint boundary is at infinity.

– *Case 4:* Parameter δ is used to scale the distance between the state and the boundary. Increasing the value of δ, $\xi(x)$ will be more closer to x. On the contrary, $\xi(x)$ will be enlarged. See Fig.1(b).

Remark 3 In order to compare with the state-dependent transformations in the existing literature, the majority of approaches, which are marked as A-E, are shown schematically, see Fig.1. It can be seen that all methods can handle asymmetric state constraints, but the approach proposed in [23, 29], marked by D, when the state is 0, the new state is not equal to 0, unless the state boundary is symmetric. When the state is away from the constraint boundary, method A (Proposed), C see [21], E see [20, 48, 25] have better linearity than method B[27]. However, the slope of the linear part of method E is much smaller than that of A and C, indicating that A and C have a better ability to restore the original state. Nevertheless, the state-dependent transformation function proposed by method C is a segmented function with respect to the state and constraint boundaries, and its derivative as well depends on the segmented function.

Remark 4 Most of the literature addressing state constraints using BLF usually constructs such a Lyapunov function $V = 1/2 log(F^2/(F^2 - \zeta^2))$. Similarly, in [20], a state-dependent transformation function is constructed as $\xi(x) = x/((F_1 + x)(F_2 - x))$. Noted that, when there is no state constraint, i.e. F, F_1, F_2 tends to ∞, V or $\xi(x)$ tends to 0. That is, these methods mentioned above cannot cope with the situation where there are no state constraints or time varying state constraints tend to infinity. As the remarkable work in [21], the proposed novel NSDF has the ability to handle this situation, see Case 3 in Remark 2.

Now, reconsider the system (1) with asymmetric time-varying full state constraints, by utilizing the novel state-dependent function, the following transformations can be made on the original system

$$
\begin{aligned}
\dot{\xi}_1 &= \mu_1 \dot{x}_1 + \nu_1 \\
&= \mu_1(f_1(\bar{x}_1(t)) + x_2(t)) + \nu_1 \\
\dot{\xi}_i &= \mu_i \dot{x}_i + \nu_i \\
&= \mu_i(f_i(\bar{x}_i(t)) + x_{i+1}(t)) + \nu_i \\
\dot{\xi}_n &= \mu_n \dot{x}_n + \nu_n \\
&= \mu_n(f_n(\bar{x}_n(t)) + u) + \nu_n
\end{aligned}
\tag{17}
$$

Defining $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^T$, $\mathcal{F}_i(\xi) = \mu_i(f_i(\bar{x}_i(t)) + x_{i+1}(t)) + \nu_i - \xi_{i+1}$, $\mathcal{F}_n(\xi) = \mu_n(f_n(\bar{x}_n(t)) + u) + \nu_n - u$, we obtain

$$
\begin{aligned}
\dot{\xi}_i &= \mathcal{F}_i(\xi) + \xi_{i+1} \\
\dot{\xi}_n &= \mathcal{F}_n(\xi) + u
\end{aligned}
\tag{18}
$$

where $i = 1, 2, \ldots, n - 1$. The new state variable vector ξ and system (18), in which state constraints are incorporated, will be used to design optimal controller.

Remark 5 Under the proposed novel nonlinear state-dependent transformation function, the original time-varying asymmetric state constraint is implicitly incorporated into a new strict feedback system, see (18). It is worth mentioning that $\mathcal{F}_i(\xi)$ is a continuous dynamic function and when $x = 0$, $\mathcal{F}_i(0) = 0$. It means that the new system (18) is stabilizable, i.e., there exists a continuous input u that stabilizes the system asymptotically.

3.2 Optimal Backstepping Controller Design

An optimized controller will be designed in this section for system (18).Before designing the controller, a new tracking signal variable needs to be introduced. The new tracking signal variable ξ_r is defined as

$$\xi_r = \frac{y_r}{d_r(y_r, F_{1_r}, F_{2_r})} \tag{19}$$

where $d_r(y_r, F_{1_r}, F_{2_r}) = e^{-\frac{d_{1_r} d_{2_r}}{\delta}}$, $d_{1_r}(y_r, F_{1_r}) = \frac{1}{\delta(y_r - F_{1_r})}$, $d_{2_r}(y_r, F_{2_r}) = \frac{1}{\delta(F_{2_r} - y_r)}$.

Similarly, we can get $\dot{\xi}_r$ as

$$\dot{\xi}_r = \mu_r \dot{y}_r + \nu_r \tag{20}$$

where

$$\mu_r = \frac{1}{d_r(y_r, F_{1_r}, F_{2_r})} - \frac{x}{\delta d(y_r, F_{1_r}, F_{2_r})} \left(\frac{d_{2_r}}{d_{1_r}^2} - \frac{d_{1_r}}{d_{2_r}^2} \right)$$

$$\nu_r = \frac{y_r}{\delta d(y_r, F_{1_r}, F_{2_r})} \left(\frac{d_{2_r}}{d_{1_r}^2} \dot{F}_{1_r} - \frac{d_{1_r}}{d_{2_r}^2} \dot{F}_{2_r} \right) \tag{21}$$

In order to design the optimal controller by utilizing backstepping technique, a new coordinate transformation is defined as

$$\begin{aligned} \zeta_1 &= \xi_1 - \xi_r \\ \zeta_i &= \xi_i - \hat{\alpha}_{i-1}^* \end{aligned} \tag{22}$$

where $\hat{\alpha}_i$ is the designed virtual controller in ith step. The optimal controller is designed as follows:

$$\begin{aligned} \hat{\alpha}_i^* &= -\bar{c}_i \zeta_i - \hat{\mathcal{W}}_{f_i}^T \Phi_{f_i}(\xi) - \frac{1}{2} \hat{\mathcal{W}}_{a_i}^T \Phi_{c_i}(\xi, \zeta_i), \\ &\quad i = 1, \ldots n - 1 \\ u &= -\bar{c}_n \zeta_n - \hat{\mathcal{W}}_{f_n}^T \Phi_{f_n}(\xi) - \frac{1}{2} \hat{\mathcal{W}}_{a_n}^T \Phi_{c_n}(\xi, \zeta_n) \end{aligned} \tag{23}$$

$$\begin{aligned} \dot{\hat{\mathcal{W}}}_{f_i} &= \Gamma_i \left(\Phi_{fi}(\xi) \zeta_i(t) - \sigma_i \hat{W}_{f_i}(t) \right) \\ \dot{\hat{\mathcal{W}}}_{c_i} &= -\gamma_{c_i} \Phi_{c_i}(\xi, \zeta_i) \Phi_{c_i}(\xi, \zeta_i)^T \hat{\mathcal{W}}_{c_i} \\ \dot{\hat{\mathcal{W}}}_{a_i} &= -\Phi_{c_i}(\xi, \zeta_i) \Phi_{c_i}(\xi, \zeta_i)^T \\ &\quad \times \left(\gamma_{a_i} \left(\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i} \right) + \gamma_{c_i} \hat{\mathcal{W}}_{c_i} \right) \\ &\quad i = 1, \ldots n \end{aligned} \tag{24}$$

where, $\bar{c}_i$ is the design parameter of the ith step optimal virtual controller, and u is the optimal controller. $\hat{\mathcal{W}}_{f_i} \in \mathcal{R}^{p_i}$, $\hat{\mathcal{W}}_{c_i}^T \in \mathcal{R}^{q_i}$, $\hat{\mathcal{W}}_{a_i}^T \in \mathcal{R}^{q_i}$ is the estimated weight matrix of identifier, critic network and actor network if the ith step, respectively. $\Phi_{f_i} \in p_i$, $\Phi_{c_i} \in q_i$ are the basis function vectors of identifier and critic network, respectively. p_i and q_i are the dimension of the weight and basis function of step i. The parameters should satisfy the below conditions

$$\bar{c}_i > 3, \gamma_{a_i} > \frac{1}{2}, \gamma_{a_i} > \gamma_{c_i} > \frac{\gamma_{a_i}}{2} \tag{25}$$

Next, the detailed design procedures are presented.
Step:1 According to (19) and (22)

$$\begin{aligned} \dot{\zeta}_1 &= \dot{\xi}_1 - \dot{\xi}_r \\ &= \mathcal{F}_1(\xi) + \xi_2 - \dot{\xi}_r \end{aligned} \tag{26}$$

where ξ_2 is viewed as the intermediate controller. Define the virtual controller α_1 and the optimal virtual controller α_1^*. Then we define the optimal performance index function as

$$\begin{aligned} \mathcal{J}_1^*(\zeta_1) &= \min_{\alpha_1 \in \Psi(\Omega)} \left(\int_t^\infty h_1(\zeta_1(\tau), \alpha_1(\zeta_1)) d\tau \right) \\ &= \int_t^\infty h_1(\zeta_1(\tau), \alpha_1^*(\zeta_1)) d\tau \end{aligned} \tag{27}$$

where Ω is the admissible set of α_1, and the cost function $h_1(\zeta_1, \alpha_1) = \zeta_1^2(t) + \alpha_1^2(\zeta_1)$.

According section 2.3, define $\mathcal{J}_{\zeta_1}^* = \frac{\partial \mathcal{J}_1^*}{\partial \zeta_1}$, by using $\partial \mathcal{H}_1 \left(\zeta_1, \alpha_1^*, \mathcal{J}_{\zeta_1}^* \right) / \partial \alpha_1^* = 0$, we can obtain the optimal controller α_1^*

$$\alpha_1^* = -\frac{1}{2} \frac{\partial \mathcal{J}_1^*}{\partial \zeta_1} \tag{28}$$

$$\frac{\partial \mathcal{J}_1^*}{\partial \zeta_1} = 2\bar{c}_1 \zeta_1 + 2\mathcal{F}_1(\xi) + \mathcal{J}_1^o(\xi, \zeta_1) \tag{29}$$

where $\bar{c}_1$ is a positive design parameter and $\mathcal{J}_1^o(\xi, \zeta_1) = -2\bar{c}_1 \zeta_1(t) - 2\mathcal{F}_1(\xi) + \partial \mathcal{J}_1^*(\zeta_1)/\partial \zeta_1 \in \mathcal{R}$.

Since $\mathcal{F}_1(\xi)$ and $\mathcal{J}_1^o(\xi, \zeta_1)$ are unknown but continuous, we use a neural network to approximate them. Then $\mathcal{F}_1(\xi)$ and $\mathcal{J}_1^o(\xi, \zeta_1)$ are re-described as

$$\begin{aligned} \mathcal{F}_1(\xi) &= \mathcal{W}_{f_1}^{*T} \Phi_{f_1}(\xi) + \varepsilon_{f_1}(\xi) \\ \mathcal{J}_1^o(\xi, \zeta_1) &= \mathcal{W}_{c_1}^{*T} \Phi_{c_1}(\xi, \zeta_1) + \varepsilon_{c_1}(\xi, \zeta_1) \end{aligned} \tag{30}$$

where $\mathcal{W}_{f_1}^* \in \mathcal{R}^{p_1}$ and $\mathcal{W}_{c_1}^* \in \mathcal{R}^{q_1}$ are ideal weights, $\Phi_{f_1}(\xi) \in \mathcal{R}^{p_1}$ and $\Phi_{c_1}(\xi,\zeta_1) \in \mathcal{R}^{q_1}$ are the basis function vectors, $\varepsilon_{f_1}(\xi)$ and $\varepsilon_{J_1}(\xi,\zeta_1) \in \mathcal{R}$ are the approximation errors. Yielding that

$$
\begin{aligned}
\frac{\partial \mathcal{J}_1^*}{\partial \zeta_1} &= 2\bar{c}_1\zeta_1 + 2\mathcal{W}_{f_1}^{*T}\Phi_{f_1}(\xi) + \mathcal{W}_{c_1}^{*T}\Phi_{c_1}(\xi,\zeta_1) + 2\varepsilon_{f_1}(\xi) \\
&\quad + \varepsilon_{J_1}(\xi,\zeta_1) \\
\alpha_1^* &= -\bar{c}_1\zeta_1 - \mathcal{W}_{f_1}^{*T}\Phi_{f_1}(\xi) - \frac{1}{2}\mathcal{W}_{c_1}^{*T}\Phi_{c_1}(\xi,\zeta_1) - \\
&\quad \varepsilon_{f_1}(\xi) - \frac{1}{2}\varepsilon_{c_1}(\xi,\zeta_1)
\end{aligned}
\tag{31}
$$

Since $\mathcal{W}_{f_1}^*$ and $\mathcal{W}_{c_1}^*$ are not directly available, reinforcement learning-based identifier, critic and action networks are constructed to obtain optimal controller. The identifier, critic and action neural network are designed as follows.

The identifier $\hat{\mathcal{F}}_1(\xi)$ with updating law $\dot{\hat{\mathcal{W}}}_{f_1}$ is designed as

$$
\hat{\mathcal{F}}_1(\xi) = \hat{\mathcal{W}}_{f_1}^T \Phi_{f_1}(\xi)
\tag{32}
$$

where $\hat{\mathcal{F}}_1(\xi) \in \mathcal{R}$ is the output of identifier, $\hat{\mathcal{W}}_{f_1} \in \mathcal{R}^{p_1}$ is the weight of identifier neural network, and $\Phi_{f_1}(\xi) \in R^{p_1}$ is the basis vector. Updating law is described as

$$
\dot{\hat{\mathcal{W}}}_{f_1} = \Gamma_1 \left(\Phi_{f1}(\xi)\zeta_1(t) - \sigma_1 \hat{W}_{f1}(t) \right)
\tag{33}
$$

where Γ_1 is a positive definite matrix, $\sigma_1 > 0$ is a design parameter.

The critic and action network are designed as

$$
\frac{\partial \hat{\mathcal{J}}_1^*}{\partial \zeta_1} = 2\bar{c}_1\zeta_1 + 2\hat{\mathcal{W}}_{f_1}^T\Phi_{f_1}(\xi) + \hat{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi,\zeta_1)
\tag{34}
$$

where $\partial \hat{\mathcal{J}}_1^*/\partial \zeta_1 \in \mathcal{R}$ is the estimated $\partial \mathcal{J}_1^*/\partial \zeta_1$ and $\hat{\mathcal{W}}_{c_1} \in \mathcal{R}^{q_1}$ is weight of the critic NN, which updating law is

$$
\dot{\hat{\mathcal{W}}}_{c_1} = -\gamma_{c_1}\Phi_{c_1}(\xi,\zeta_1)\Phi_{c_1}(\xi,\zeta_1)^T\hat{\mathcal{W}}_{c_1}
\tag{35}
$$

where $\gamma_{c_1} > 0$ is the critic designed parameter.

$$
\hat{\alpha}_1^* = -\bar{c}_1\zeta_1 - \hat{\mathcal{W}}_{f_1}^T\Phi_{f_1}(\xi) - \frac{1}{2}\hat{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi,\zeta_1)
\tag{36}
$$

$$
\begin{aligned}
\dot{\hat{\mathcal{W}}}_{a_1} = &-\Phi_{c_1}(\xi,\zeta_1)\Phi_{c_1}(\xi,\zeta_1)^T \\
&\times \left(\gamma_{a_1}\left(\hat{\mathcal{W}}_{a_1} - \hat{\mathcal{W}}_{c_1}\right) + \gamma_{c_1}\hat{\mathcal{W}}_{c_1} \right)
\end{aligned}
\tag{37}
$$

where $\gamma_{a_1} > 0$ is the actor designed parameter.

Step i(i=2,...,n-1): According to (22), we can get

$$
\begin{aligned}
\dot{\zeta}_i &= \dot{\xi}_i - \dot{\hat{\alpha}}_{i-1}^* \\
&= \xi_{i+1} + \mathcal{F}_i(\xi) - \dot{\hat{\alpha}}_{i-1}^*
\end{aligned}
\tag{38}
$$

Similarly, the cost of the system i is described as

$$
\begin{aligned}
\mathcal{J}_i^*(\zeta_i) &= \min_{\alpha_i \in \Psi(\Omega)} \left(\int_t^\infty h_i\left(\zeta_i(\tau), \alpha_i(\zeta_i)\right) d\tau \right) \\
&= \int_t^\infty h_i\left(\zeta_i(\tau), \alpha_i^*(\zeta_i)\right) d\tau
\end{aligned}
\tag{39}
$$

where $h_i(\zeta_i, \alpha_i) = \zeta_i^2 + \alpha_i^2$. Treat ξ_{i+1} as the optimal virtual controller α_i^*, then

$$
\alpha_i^* = -\frac{1}{2}\frac{\partial \mathcal{J}_i^*}{\partial \zeta_i}
\tag{40}
$$

$$
\frac{\partial \mathcal{J}_i^*}{\partial \zeta_i} = 2\bar{c}_i\zeta_i + 2\mathcal{F}_i(\xi) + \mathcal{J}_i^o(\xi,\zeta_i)
\tag{41}
$$

where $\bar{c}_i$ is a positive design parameter and $\mathcal{J}_i^o(\xi,\zeta_i) = -2\bar{c}_i\zeta_i(t) - 2\mathcal{F}_i(\xi) + \partial \mathcal{J}_i^*(\zeta_i)/\partial \zeta_i \in \mathcal{R}$. The identifier, critic and action network for the subsystem i with updating law is designed as

$$
\hat{\mathcal{F}}_i(\xi) = \hat{\mathcal{W}}_{f_i}^T \Phi_{f_i}(\xi)
\tag{42}
$$

$$
\frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i} = 2\bar{c}_i\zeta_i + 2\hat{\mathcal{W}}_{f_i}^T\Phi_{f_i}(\xi) + \hat{\mathcal{W}}_{c_i}^T\Phi_{c_i}(\xi,\zeta_i)
\tag{43}
$$

$$
\hat{\alpha}_i^* = -\bar{c}_i\zeta_i - \hat{\mathcal{W}}_{f_i}^T\Phi_{f_i}(\xi) - \frac{1}{2}\hat{\mathcal{W}}_{a_i}^T\Phi_{c_i}(\xi,\zeta_i)
\tag{44}
$$

where $\hat{\mathcal{F}}_i(\xi) \in \mathcal{R}$ is the output of the identifier , $\hat{\mathcal{W}}_{f_i} \in \mathcal{R}^{p_i}$ is the weight if identifier NN, and $\Phi_{f_i}(\xi) \in R^{p_i}$ is the basis. $\partial \hat{\mathcal{J}}_i^*/\partial \zeta_i \in \mathcal{R}$ is the estimated $\partial \mathcal{J}_i^*/\partial \zeta_i$ and $\hat{\mathcal{W}}_{c_i} \in \mathcal{R}^{q_i}$ is weight of the critic. Their tuning law are described as follows

$$
\begin{aligned}
\dot{\hat{\mathcal{W}}}_{f_i} &= \Gamma_i \left(\Phi_{fi}(\xi)\zeta_i(t) - \sigma_i\hat{W}_{f_i}(t) \right) \\
\dot{\hat{\mathcal{W}}}_{c_i} &= -\gamma_{c_i}\Phi_{c_i}(\xi,\zeta_i)\Phi_{c_i}(\xi,\zeta_i)^T\hat{\mathcal{W}}_{c_i} \\
\dot{\hat{\mathcal{W}}}_{a_i} &= -\Phi_{c_i}(\xi,\zeta_i)\Phi_{c_i}(\xi,\zeta_i)^T \left(\gamma_{a_i}\left(\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i}\right) + \gamma_{c_i}\hat{\mathcal{W}}_{c_i} \right)
\end{aligned}
\tag{45}
$$

where Γ_i is a positive definite matrix, $\sigma_i > 0$ is a design parameter. $\gamma_{c_i} > 0, \gamma_{a_i} > 0$ are the designed parameter of critic and actor networks.

Step n: In the final step, the actual controller u is designed. According to (22), we have

$$
\begin{aligned}
\dot{\zeta}_n &= \dot{\xi}_n - \dot{\hat{\alpha}}_{n-1}^* \\
&= u + \mathcal{F}_n(\xi) - \dot{\hat{\alpha}}_{i-1}^*
\end{aligned}
\tag{46}
$$

The performance index function is described as

$$\mathcal{J}_n^*(\zeta_n) = \min_{u \in \Psi(\Omega)} \left(\int_t^\infty h_n\left(\zeta_n(\tau), u\left(\zeta_n\right)\right) d\tau \right)$$
$$= \int_t^\infty h_n\left(\zeta_n(\tau), u^*\left(\zeta_n\right)\right) d\tau \qquad (47)$$

where $h_n(\zeta_n, u) = \zeta_n^2 + u^2$. Similarly, we have

$$u^* = -\frac{1}{2}\frac{\partial \mathcal{J}_n^*}{\partial \zeta_n} \qquad (48)$$

$$\frac{\partial \mathcal{J}_n^*}{\partial \zeta_n} = 2\bar{c}_n \zeta_n + 2\mathcal{F}_n(\xi) + \mathcal{J}_n^o(\xi, \zeta_n) \qquad (49)$$

where $\bar{c}_n$ is a positive design parameter and $\mathcal{J}_n^o(\xi, \zeta_n) = -2\bar{c}_n \zeta_n(t) - 2\mathcal{F}_n(\xi) + \partial \mathcal{J}_n^*(\zeta_n)/\partial \zeta_n \in \mathcal{R}$. Yielding that

$$u^* = -\bar{c}_n \zeta_n(t) - \mathcal{F}_n(\xi) - \frac{1}{2}\mathcal{J}_u^o(\xi, \zeta_n) \qquad (50)$$

Since, $\mathcal{F}_n(\xi)$ and $\mathcal{J}_i^o(\xi, \zeta_n)$ are unknown, using neural networks to approximate it

$$\mathcal{F}_n(\xi) = \mathcal{W}_{f_n}^{*T}\Phi_{f_n}(\xi) + \varepsilon_{f_n}(\xi)$$
$$\mathcal{J}_n^o(\xi, \zeta_n) = \mathcal{W}_{c_n}^{*T}\Phi_{c_n}(\xi, \zeta_n) + \varepsilon_{c_n}(\xi, \zeta_n) \qquad (51)$$

where $\mathcal{W}_{f_n}^* \in \mathcal{R}^{p_n}$ and $\mathcal{W}_{c_n}^* \in \mathcal{R}^{q_n}$ are ideal weights, $\Phi_{f_n}(\xi) \in \mathcal{R}^{p_n}$ and $\Phi_{c_n}(\xi, \zeta_n) \in \mathcal{R}^{q_n}$ are the activation function vectors, $\varepsilon_{f_n}(\xi)$ and $\varepsilon_{J_n}(\xi, \zeta_n) \in \mathcal{R}$ are the approximation errors.

The identifier, critic network and actor network defined as follows:

$$\hat{\mathcal{F}}_n(\xi) = \hat{\mathcal{W}}_{f_n}^T\Phi_{f_n}(\xi) \qquad (52)$$

$$\frac{\partial \hat{\mathcal{J}}_n^*}{\partial \zeta_n} = 2\bar{c}_n \zeta_n + 2\hat{\mathcal{W}}_{f_n}^T\Phi_{f_n}(\xi) + \hat{\mathcal{W}}_{c_n}^T\Phi_{c_n}(\xi, \zeta_n) \qquad (53)$$

$$u = -\bar{c}_n \zeta_n - \hat{\mathcal{W}}_{f_n}^T\Phi_{f_n}(\xi) - \frac{1}{2}\hat{\mathcal{W}}_{a_n}^T\Phi_{c_n}(\xi, \zeta_n) \qquad (54)$$

where $\hat{\mathcal{F}}_n(\xi) \in \mathcal{R}$ is the output of identifier, $\hat{\mathcal{W}}_{f_n} \in \mathcal{R}^{p_n}$ is the weight of identifier, and $\Phi_{f_n}(\xi) \in R^{p_n}$ is the activation function vector. $\partial \hat{\mathcal{J}}_n^*/\partial \zeta_n \in \mathcal{R}$ is the estimated $\partial \mathcal{J}_n^*/\partial \zeta_n$ and $\hat{\mathcal{W}}_{c_n} \in \mathcal{R}^{q_n}$ is weight of the critic NN. The updating law as follows

$$\dot{\hat{\mathcal{W}}}_{f_n} = \Gamma_n \left(\Phi_{fn}(\xi)\zeta_n(t) - \sigma_n \hat{W}_{f_n}(t) \right)$$
$$\dot{\hat{\mathcal{W}}}_{c_n} = -\gamma_{c_n}\Phi_{c_n}(\xi, \zeta_n)\Phi_{c_n}(\xi, \zeta_n)^T\hat{\mathcal{W}}_{c_n}$$
$$\dot{\hat{\mathcal{W}}}_{a_n} = -\Phi_{c_n}(\xi, \zeta_n)\Phi_{c_n}(\xi, \zeta_n)^T$$
$$\left(\gamma_{a_n}\left(\hat{\mathcal{W}}_{a_n} - \hat{\mathcal{W}}_{c_n}\right) + \gamma_{c_n}\hat{\mathcal{W}}_{c_n} \right) \qquad (55)$$

where Γ_n is a positive definite matrix, $\sigma_n > 0$ is a design parameter. $\gamma_{c_n} > 0, \gamma_{a_n} > 0$ are the designed parameter with the conditions as follows

$$\bar{c}_n > 3, \gamma_{a_n} > \frac{1}{2}, \gamma_{a_n} > \gamma_{c_n} > \frac{\gamma_{a_n}}{2} \qquad (56)$$

4 Main Results and Stability Analysis

The main results and proofs are given as follows.

Theorem 1 *Take account of the strict-feedback uncertain nonlinear system (1) with full state constraints which is asymmetric and time-varying $F_{i1}(t) < x_i(t) < F_{i2}$ under the assumptions 1 and 2, by utilizing the optimized virtual and actual controller (23) with the updating law (24), and the design parameters conditions (25), thereafter the proposed scheme can ensure the followings*

1. *All the signals ζ_i, $\tilde{\mathcal{W}}_{f_i}$, $\tilde{\mathcal{W}}_{c_i}$ and $\tilde{\mathcal{W}}_{a_i}$ are bounded;*
2. *The output of system y_1 can track the desired signal y_r;*
3. *All states do not violate the asymmetric time-varying constraints with no dependence on feasibility conditions.*

Proof Before giving the stability analysis, we show that the designed weight updating laws (24) can minimize the approximation error of the HJB equation.

Recalling the HJB equation (11), by utilizing (43) and (44), we have

$$\mathcal{H}_i\left(\zeta_i, \hat{\alpha}_i^*, \frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i}\right)$$
$$= \zeta_i^2(t) + \left(-\bar{c}_i\zeta_i - \hat{\mathcal{W}}_{f_i}^T\Phi_{f_i}(\xi) - \frac{1}{2}\hat{\mathcal{W}}_{a_i}^T\Phi_{c_i}(\xi, \zeta_i)\right)^2$$
$$+ \left(2\bar{c}_i\zeta_i + 2\hat{\mathcal{W}}_{f_i}^T\Phi_{f_i}(\xi) + \hat{\mathcal{W}}_{c_i}^T\Phi_{c_i}(\xi, \zeta_i)\right)$$
$$\times (\mathcal{F}_i(\xi) - \bar{c}_i\zeta_i - \hat{\mathcal{W}}_{f_i}^T\Phi_{f_i}(\xi)$$
$$- \frac{1}{2}\hat{\mathcal{W}}_{a_1}^T\Phi_{c_i}(\xi, \zeta_i) - \dot{\alpha}_{i-1}^*) \qquad (57)$$

Bellman residual error $e_i(t)$ is defined as

$$e_i = \mathcal{H}_i\left(\zeta_i, \hat{\alpha}_i^*, \frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i}\right) - \mathcal{H}_i\left(\zeta_i, \alpha_i^*, \frac{\partial \mathcal{J}_i^*}{\partial \zeta_i}\right)$$
$$= \mathcal{H}_i\left(\zeta_i, \hat{\alpha}_i^*, \frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i}\right) \qquad (58)$$

where

$$\mathcal{H}_i\left(\zeta_i, \alpha_i^*, \frac{\partial \mathcal{J}_i^*}{\partial \zeta_i}\right) = h_i(\zeta_i, \alpha_i^*) + \frac{\partial \mathcal{J}_i^*}{\partial \zeta_i}\dot{\zeta}_i$$
$$= \zeta_i^2(t) + \alpha_i^{*2}(\zeta_i)$$
$$+ \frac{\partial \mathcal{J}_i^*}{\partial \zeta_i}\left(\mathcal{F}_i(\xi) + \alpha_i^* - \dot{\alpha}_{i-1}^*\right)$$
$$= 0 \qquad (59)$$

From equation (58), it can be seen that the optimal solution $\hat{\alpha}_i^*$ is obtained when $e_i(t) = \mathcal{H}_i\left(\zeta_i, \hat{\alpha}_i^*, \frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i}\right) \to 0$ is satisfied. In other words, when $\mathcal{H}_i\left(\zeta_i, \hat{\alpha}_i^*, \frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i}\right) = 0$ has a unique solution. Define $E_i = \frac{1}{2}e_i^2$, $\partial E_i/\partial \hat{\mathcal{W}}_{a_i} = \mathcal{H}_i(\zeta_i, \hat{\alpha}_i^*, \frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i})\frac{\partial \mathcal{H}_i}{\partial \hat{\mathcal{W}}_{a_i}} = 0$ i.e.

$$\frac{\partial \mathcal{H}_i\left(\zeta_i, \alpha_i^*, \frac{\partial \hat{\mathcal{J}}_i^*}{\partial \zeta_i}\right)}{\partial \hat{\mathcal{W}}_{a_i}} = \frac{1}{2}\Phi_{c_i}(\xi, \zeta_i)\Phi_{c_i}^T(\xi, \zeta_i)\left(\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i}\right)$$
$$= 0$$

(60)

To ensure that the updating laws satisfy the above equation, we define a positive function $\mathcal{P}_i = (\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i})^T(\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i})$ with the facts that $\partial \mathcal{P}_i/\partial \hat{\mathcal{W}}_{a_i} = -\partial \mathcal{P}_i/\partial \hat{\mathcal{W}}_{c_i} = 2\left(\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i}\right)$. It can be seen that when $\mathcal{P}_i$ is equal to 0, it means that the equation (60) is satisfied. Therefore, the weight updating laws can be structured by $\dot{\mathcal{P}}_i \leq 0$. Recalling the updating laws (24), we have

$$\begin{aligned}\frac{d\mathcal{P}_i}{dt} =& \frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{a_i}^T}\dot{\hat{\mathcal{W}}}_{a_i} + \frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{c_i}^T}\dot{\hat{\mathcal{W}}}_{c_i}\\ =& \frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{a_i}^T}\Big(-\Phi_{c_i}(\xi, \zeta_i)\Phi_{c_i}(\xi, \zeta_i)^T\\ & \left(\gamma_{a_i}\left(\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i}\right) + \gamma_{c_i}\hat{\mathcal{W}}_{c_i}\right)\Big)\\ & + \frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{c_i}^T}\left(-\gamma_{c_i}\Phi_{c_i}(\xi, \zeta_i)\Phi_{c_i}(\xi, \zeta_i)^T\hat{\mathcal{W}}_{c_i}\right)\\ =& -\gamma_{c_i}\frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{c_i}^T}\left(\Phi_{c_i}(\xi, \zeta_i)\Phi_{c_i}(\xi, \zeta_i)^T\hat{\mathcal{W}}_{c_i}\right)\\ & -\gamma_{c_i}\frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{a_i}^T}\left(\Phi_{c_i}(\xi, \zeta_i)\Phi_{c_i}(\xi, \zeta_i)^T\hat{\mathcal{W}}_{c_i}\right)\\ & -\frac{\gamma_{a_i}}{2}\frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{a_i}^T}\Phi_{c_i}(\xi, \zeta_i)\Phi_{c_i}(\xi, \zeta_i)^T\frac{\partial \mathcal{P}_i}{\partial \hat{\mathcal{W}}_{c_i}^T}\\ \leq& 0\end{aligned}$$

(61)

The above inequality means that the updating laws (24) can minimize the bellman residual error $e_i(t)$.

Remark 6 It should be noted that in the procedure of giving the proof of the neural network weight updating law that can guarantee that the Bellman residual error converge to 0, we use the design result of step i without using steps 1 and n. Actually, the proof process of steps 1 and n is fundamentally the same as that of step i, only the partial notation is different.

Next, the system stability proof is presented.

Step:1 Considering a Lyapunov function of ζ_1 subsystem as

$$\mathcal{V}_1 = \frac{1}{2}\zeta_1^2 + \frac{1}{2}\tilde{\mathcal{W}}_{f_1}^T\Gamma_1^{-1}\tilde{\mathcal{W}}_{f_1} + \frac{1}{2}\tilde{\mathcal{W}}_{c_1}^T\tilde{\mathcal{W}}_{c_1} + \frac{1}{2}\tilde{\mathcal{W}}_{a_1}^T\tilde{\mathcal{W}}_{a_1}$$

(62)

where $\tilde{\mathcal{W}}_{f_1} = \hat{\mathcal{W}}_{f_1} - \mathcal{W}_{f_1}^*$, $\tilde{\mathcal{W}}_{c_1} = \hat{\mathcal{W}}_{c_1} - \mathcal{W}_{c_1}^*$, $\tilde{\mathcal{W}}_{a_1} = \hat{\mathcal{W}}_{a_1} - \mathcal{W}_{a_1}^*$. Recalling (26), (30) and (36), we have

$$\begin{aligned}\dot{\mathcal{V}}_1 =& -\bar{c}_1\zeta_1^2 + \zeta_1\varepsilon_{f_1}(\xi) - \zeta_1\dot{\xi}_d + \zeta_1\zeta_2\\ & -\frac{1}{2}\zeta_1\hat{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1) - \sigma_1\tilde{\mathcal{W}}_{f_1}^T\hat{\mathcal{W}}_{f_1}\\ & -\gamma_{c_1}\tilde{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{c_1}\\ & -\tilde{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\left(\gamma_{a_1}\left(\hat{\mathcal{W}}_{a_1} - \hat{\mathcal{W}}_{c_1}\right) + \gamma_{c_1}\hat{\mathcal{W}}_{c_1}\right)\end{aligned}$$

(63)

By using Young's inequality

$$\begin{aligned}\zeta_1\varepsilon_{f_1} &\leq \frac{1}{2}\zeta_1^2 + \frac{1}{2}\varepsilon_{f_1}\\ \zeta_1\zeta_2 &\leq \frac{1}{2}\zeta_1^2 + \frac{1}{2}\zeta_2^2\\ -\zeta_1\dot{\xi}_d &\leq \frac{1}{2}\zeta_1^2 + \frac{1}{2}\dot{\xi}_d^2\\ -\frac{1}{2}\zeta_1\hat{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1) &\leq \frac{1}{4}\zeta_1^2 + \frac{1}{4}\hat{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{a_1}\end{aligned}$$

(64)

And the following equations

$$\begin{aligned}\tilde{\mathcal{W}}_{f_1}^T\hat{\mathcal{W}}_{f_1} =& \frac{1}{2}\tilde{\mathcal{W}}_{f_1}^T\tilde{\mathcal{W}}_{f_1} + \frac{1}{2}\hat{\mathcal{W}}_{f_1}^T\hat{\mathcal{W}}_{f_1}\\ & -\frac{1}{2}\mathcal{W}_{f_1}^{*T}\mathcal{W}_{f_1}^*\\ \tilde{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{c_1} =& \frac{1}{2}\tilde{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\tilde{\mathcal{W}}_{c_1}\\ & +\frac{1}{2}\hat{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{c_1}\\ & -\frac{1}{2}\mathcal{W}_{c_1}^{*T}\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\mathcal{W}_{c_1}^*\\ \tilde{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{a_1} =& \frac{1}{2}\tilde{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\tilde{\mathcal{W}}_{a_1}\\ & +\frac{1}{2}\hat{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{a_1}\\ & -\frac{1}{2}\mathcal{W}_{a_1}^{*T}\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\mathcal{W}_{a_1}^*\end{aligned}$$

(65)

And with the condition

$$\bar{c}_1 > 3, \gamma_{a_1} > \frac{1}{2}, \gamma_{a_1} > \gamma_{c_1} > \frac{\gamma_{a_1}}{2}$$

(66)

Then we have

$$\begin{aligned}(\gamma_{a_1} &- \gamma_{c_1})\tilde{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{c_1}\\ &\leq \frac{\gamma_{a_1} - \gamma_{c_1}}{2}\tilde{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\tilde{\mathcal{W}}_{a_1}\\ &\quad + \frac{\gamma_{a_1} - \gamma_{c_1}}{2}\hat{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi, \zeta_1)\Phi_{c_1}^T(\xi, \zeta_1)\hat{\mathcal{W}}_{c_1}\end{aligned}$$

(67)

According to the above equations, $\dot{\mathcal{V}}_1$ satisfy that

$$
\begin{aligned}
\dot{\mathcal{V}}_1 \leq & -(\bar{c}_1 - 2)\zeta_1^2 \\
& - \frac{\sigma_1}{2}\tilde{\mathcal{W}}_{f_1}^T\tilde{\mathcal{W}}_{f_1} \\
& - \frac{\gamma_{c_1}}{2}\tilde{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi,\zeta_1)\Phi_{c_1}^T(\xi,\zeta_1)\tilde{\mathcal{W}}_{c_1} \\
& - \frac{\gamma_{c_1}}{2}\tilde{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi,\zeta_1)\Phi_{c_1}^T(\xi,\zeta_1)\tilde{\mathcal{W}}_{a_1} \\
& + \mathcal{C}_1 + \frac{1}{2}\zeta_2^2
\end{aligned}
\tag{68}
$$

where $\mathcal{C}_1 = (\gamma_{a_1}/2 + \gamma_{c_1}/2)(\mathcal{W}_{c_1}^{*T}\Phi_{c_1}(\xi,\zeta_1))^2 + (\sigma_1/2)\mathcal{W}_{f_1}^{*T}\mathcal{W}_{f_1}^* + (1/2)\varepsilon_{f_1}^2 + (1/2)\dot{\xi}_d^2$, which satisfy $|\mathcal{C}_1| \leq \beta_1$, β_1 is a constant.

According to the following equations:

$$
-\tilde{\mathcal{W}}_{f_1}^T\tilde{\mathcal{W}}_{f_1} \leq -\frac{1}{\lambda_{\max}(\Gamma_1^{-1})}\tilde{\mathcal{W}}_{f_1}^T\Gamma_1^{-1}\tilde{\mathcal{W}}_{f_1}
$$

$$
-\tilde{\mathcal{W}}_{c_1}^T\Phi_{c_1}(\xi,\zeta_1)\Phi_{c_1}^T(\xi,\zeta_1)\tilde{\mathcal{W}}_{c_1} \leq -\lambda_{\min}(\Phi_{c_1}\Phi_{c_1}^T)\tilde{\mathcal{W}}_{c_1}^T\tilde{\mathcal{W}}_{c_1}
$$

$$
-\tilde{\mathcal{W}}_{a_1}^T\Phi_{c_1}(\xi,\zeta_1)\Phi_{c_1}^T(\xi,\zeta_1)\tilde{\mathcal{W}}_{a_1} \leq -\lambda_{\min}(\Phi_{c_1}\Phi_{c_1}^T)\tilde{\mathcal{W}}_{a_1}^T\tilde{\mathcal{W}}_{a_1}
\tag{69}
$$

where $\lambda_{\max}(\Gamma_1^{-1})$ denotes the largest eigenvalue of Γ_1 and $\lambda_{\min}(\Phi_{c_1}\Phi_{c_1}^T)$ denotes the smallest eigenvalue of $\Phi_{c_1}(\xi,\zeta_1)\Phi_{c_1}^T(\xi,\zeta_1)$. Then we have

$$
\begin{aligned}
\dot{\mathcal{V}}_1 \leq & -(\bar{c}_1 - 2)\zeta_1^2 \\
& - \frac{\sigma_1}{\lambda_{\max}(\Gamma_1^{-1})}\tilde{\mathcal{W}}_{f_1}^T\Gamma_1^{-1}\tilde{\mathcal{W}}_{f_1} \\
& - \frac{\gamma_{c_1}}{2}\lambda_{\min}(\Phi_{c_1}\Phi_{c_1}^T)\tilde{\mathcal{W}}_{c_1}^T\tilde{\mathcal{W}}_{c_1} \\
& - \frac{\gamma_{c_1}}{2}\lambda_{\min}(\Phi_{c_1}\Phi_{c_1}^T)\tilde{\mathcal{W}}_{a_1}^T\tilde{\mathcal{W}}_{a_1} \\
& + \beta_1 + \frac{1}{2}\zeta_2^2
\end{aligned}
\tag{70}
$$

i.e.

$$
\dot{\mathcal{V}}_1 \leq -\bar{a}_1\mathcal{V}_1 + \beta_1 + \frac{1}{2}\zeta_2^2
\tag{71}
$$

where $\bar{a}_1 = \min\{2(\bar{c}_1-2), (\sigma_1/\lambda_{max}(\Gamma_1^{-1}), \gamma_{c_1}\lambda_{\min}(\Phi_{c_1}\Phi_{c_1}^T)\}$.

Step i: Considering the Lyapunov function candidate of step i as

$$
\mathcal{V}_i = \sum_{k=1}^{i-1}\mathcal{V}_k + \frac{1}{2}\zeta_i^2 + \frac{1}{2}\tilde{\mathcal{W}}_{f_i}^T\Gamma_i^{-1}\tilde{\mathcal{W}}_{f_i} + \frac{1}{2}\tilde{\mathcal{W}}_{c_i}^T\tilde{\mathcal{W}}_{c_i} + \frac{1}{2}\tilde{\mathcal{W}}_{a_i}^T\tilde{\mathcal{W}}_{a_i}
\tag{72}
$$

where $\tilde{\mathcal{W}}_{f_i} = \hat{\mathcal{W}}_{f_i} - \mathcal{W}_{f_i}^*$, $\tilde{\mathcal{W}}_{c_i} = \hat{\mathcal{W}}_{c_i} - \mathcal{W}_{c_i}^*$, $\tilde{\mathcal{W}}_{a_i} = \hat{\mathcal{W}}_{a_i} - \mathcal{W}_{a_i}^*$. Recalling (38), (42), (43) and (44), it's time derivative

$$
\begin{aligned}
\dot{\mathcal{V}}_i = & \sum_{k=1}^{i-1}\mathcal{V}_k + \zeta_i\Big(-\bar{c}_i\zeta_i - \frac{1}{2}\hat{\mathcal{W}}_{a_i}^T\Phi_{c_i}(\xi,\zeta_i) - \tilde{\mathcal{W}}_{f_i}^T\Phi_{f_i}(\xi) \\
& + \varepsilon_{f_i} - \dot{\hat{\alpha}}_{i-1}^* + \zeta_{i+1}\Big) \\
& + \tilde{\mathcal{W}}_{f_i}^T\Big(\Phi_{fi}(\xi)\zeta_i(t) - \sigma_i\hat{\mathcal{W}}_{f_i}(t)\Big) \\
& - \gamma_{c_i}\tilde{\mathcal{W}}_{c_i}^T\Phi_{c_i}(\xi,\zeta_i)\Phi_{c_i}^T(\xi,\zeta_i)\hat{\mathcal{W}}_{c_i} \\
& - \tilde{\mathcal{W}}_{a_i}^T\Phi_{c_i}(\xi,\zeta_i)\Phi_{c_i}^T(\xi,\zeta_i)\Big(\gamma_{a_i}\Big(\hat{\mathcal{W}}_{a_i} - \hat{\mathcal{W}}_{c_i}\Big) \\
& + \gamma_{c_i}\hat{\mathcal{W}}_{c_i}\Big)
\end{aligned}
\tag{73}
$$

Similar to step 1, we have

$$
\begin{aligned}
\dot{\mathcal{V}}_i \leq & \sum_{k=1}^{i-1}(-\bar{a}_k\mathcal{V}_k + \beta_j) - (\bar{c}_i - 2)\zeta_i^2 \\
& - \frac{\sigma_i}{\lambda_{\max}(\Gamma_i^{-1})}\tilde{\mathcal{W}}_{f_i}^T\Gamma_i^{-1}\tilde{\mathcal{W}}_{f_i} \\
& - \frac{\gamma_{c_i}}{2}\lambda_{\min}(\Phi_{c_i}\Phi_{c_i}^T)\tilde{\mathcal{W}}_{c_i}^T\tilde{\mathcal{W}}_{c_i} \\
& - \frac{\gamma_{c_i}}{2}\lambda_{\min}(\Phi_{c_i}\Phi_{c_i}^T)\tilde{\mathcal{W}}_{a_i}^T\tilde{\mathcal{W}}_{a_i} \\
& + \mathcal{C}_i + \frac{1}{2}\zeta_{i+1}^2
\end{aligned}
\tag{74}
$$

where $\mathcal{C}_i = (\gamma_{a_i}/2 + \gamma_{c_i}/2)(\mathcal{W}_{c_i}^{*T}\Phi_{c_i}(\xi,\zeta_i))^2 + (\sigma_1/2)\mathcal{W}_{f_i}^{*T}\mathcal{W}_{f_i}^* + (1/2)\varepsilon_{f_i}^2 + (1/2)\dot{\hat{\alpha}}_{i-1}^2$, which satisfy $|\mathcal{C}_i| \leq \beta_i$, β_i is a constant. Then we have

$$
\begin{aligned}
\dot{\mathcal{V}}_i \leq & \sum_{k=1}^{i-1}(-\bar{a}_k\mathcal{V}_k + \beta_j) - (\bar{c}_i - 2)\zeta_i^2 \\
& - \frac{\sigma_i}{\lambda_{\max}(\Gamma_i^{-1})}\tilde{\mathcal{W}}_{f_i}^T\Gamma_i^{-1}\tilde{\mathcal{W}}_{f_i} \\
& - \frac{\gamma_{c_i}}{2}\lambda_{\min}(\Phi_{c_i}\Phi_{c_i}^T)\tilde{\mathcal{W}}_{c_i}^T\tilde{\mathcal{W}}_{c_i} \\
& - \frac{\gamma_{c_i}}{2}\lambda_{\min}(\Phi_{c_i}\Phi_{c_i}^T)\tilde{\mathcal{W}}_{a_i}^T\tilde{\mathcal{W}}_{a_i} \\
& + \beta_i + \frac{1}{2}\zeta_{i+1}^2
\end{aligned}
\tag{75}
$$

where $\lambda_{\max}(\Gamma_i^{-1})$ denotes the largest eigenvalue of Γ_i and $\lambda_{\min}(\Phi_{c_i}\Phi_{c_i}^T)$ denotes the smallest eigenvalue of $\Phi_{c_i}\Phi_{c_i}^T$. Let $\bar{a}_i = \min\{2(\bar{c}_i-2), (\sigma_i/\lambda_{max}(\Gamma_i^{-1}), \gamma_{c_i}\lambda_{\min}(\Phi_{c_i}\Phi_{c_i}^T)\}$, then we have

$$
\dot{\mathcal{V}}_i \leq \sum_{k=1}^{i}(-\bar{a}_k\mathcal{V}_k + \beta_k) + \frac{1}{2}\zeta_{i+1}^2
\tag{76}
$$

Step n: Lyapunov function candidate of the final step is chosen as

$$
\begin{aligned}
\mathcal{V}_n = & \sum_{k=1}^{n-1}\mathcal{V}_k + \frac{1}{2}\zeta_n^2 + \frac{1}{2}\tilde{\mathcal{W}}_{f_n}^T\Gamma_n^{-1}\tilde{\mathcal{W}}_{f_n} + \frac{1}{2}\tilde{\mathcal{W}}_{c_n}^T\tilde{\mathcal{W}}_{c_n} \\
& + \frac{1}{2}\tilde{\mathcal{W}}_{a_n}^T\tilde{\mathcal{W}}_{a_n}
\end{aligned}
\tag{77}
$$

where $\tilde{\mathcal{W}}_{f_n} = \hat{\mathcal{W}}_{f_n} - \mathcal{W}_{f_n}^*$, $\tilde{\mathcal{W}}_{c_n} = \hat{\mathcal{W}}_{c_n} - \mathcal{W}_{c_n}^*$, $\tilde{\mathcal{W}}_{a_n} = \hat{\mathcal{W}}_{a_n} - \mathcal{W}_{a_n}^*$. Recalling (46), (52), (53) and (54), it's time derivative

$$
\begin{aligned}
\dot{\mathcal{V}}_n =& \sum_{k=1}^{n-1} \mathcal{V}_k \\
&+ \zeta_n \Big(-\bar{c}_n \zeta_n - \frac{1}{2}\hat{\mathcal{W}}_{a_n}^T \Phi_{c_n}(\xi, \zeta_n) - \\
&\tilde{\mathcal{W}}_{f_n}^T \Phi_{f_n}(\xi) + \varepsilon_{f_n} - \dot{\hat{\alpha}}_{n-1}^* \Big) \\
&+ \tilde{\mathcal{W}}_{f_n}^T \Big(\Phi_{f_n}(\xi)\,\zeta_n(t) - \sigma_n \hat{W}_{f_n}(t) \Big) \\
&- \gamma_{c_n} \tilde{\mathcal{W}}_{c_n}^T \Phi_{c_n}(\xi, \zeta_n)\Phi_{c_n}^T(\xi, \zeta_n)\hat{\mathcal{W}}_{c_n} \\
&- \tilde{\mathcal{W}}_{a_n}^T \Phi_{c_n}(\xi, \zeta_n)\Phi_{c_n}^T(\xi, \zeta_n) \\
&\times \Big(\gamma_{a_n}\left(\hat{\mathcal{W}}_{a_n} - \hat{\mathcal{W}}_{c_n} \right) + \gamma_{c_n}\hat{\mathcal{W}}_{c_n} \Big)
\end{aligned} \tag{78}
$$

Similarly, we obtain

$$
\begin{aligned}
\dot{\mathcal{V}}_n \le& \sum_{k=1}^{n-1}(-\bar{a}_k \mathcal{V}_k + \beta_j) - (\bar{c}_n - 2)\zeta_n^2 \\
&- \frac{\sigma_n}{\lambda_{\max}(\Gamma_n^{-1})}\tilde{\mathcal{W}}_{f_n}^T \Gamma_n^{-1}\tilde{\mathcal{W}}_{f_n} \\
&- \frac{\gamma_{c_n}}{2}\lambda_{\min}(\Phi_{c_n}\Phi_{c_n}^T)\tilde{\mathcal{W}}_{c_n}^T \tilde{\mathcal{W}}_{c_n} \\
&- \frac{\gamma_{c_n}}{2}\lambda_{\min}(\Phi_{c_n}\Phi_{c_n}^T)\tilde{\mathcal{W}}_{a_n}^T \tilde{\mathcal{W}}_{a_n} \\
&+ \mathcal{C}_n
\end{aligned} \tag{79}
$$

where $\mathcal{C}_n = (\gamma_{a_n}/2 + \gamma_{c_n}/2)(\mathcal{W}_{c_n}^{*T}\Phi_{c_n}(\xi, \zeta_n))^2 + (\sigma_1/2)\mathcal{W}_{f_n}^{*T}\mathcal{W}_{f_n}^* + (1/2)\varepsilon_{f_n}^2 + (1/2)\dot{\alpha}_{n-1}^2$, which satisfy $|\mathcal{C}_n| \le \beta_n$, β_n is a constant.

Then we have

$$
\begin{aligned}
\dot{\mathcal{V}}_n \le& \sum_{k=1}^{n-1}(-\bar{a}_k \mathcal{V}_k + \beta_j) - (\bar{c}_n - 2)\zeta_n^2 \\
&- \frac{\sigma_n}{\lambda_{\max}(\Gamma_n^{-1})}\tilde{\mathcal{W}}_{f_n}^T \Gamma_n^{-1}\tilde{\mathcal{W}}_{f_n} \\
&- \frac{\gamma_{c_n}}{2}\lambda_{\min}(\Phi_{c_n}\Phi_{c_n}^T)\tilde{\mathcal{W}}_{c_n}^T \tilde{\mathcal{W}}_{c_n} \\
&- \frac{\gamma_{c_n}}{2}\lambda_{\min}(\Phi_{c_n}\Phi_{c_n}^T)\tilde{\mathcal{W}}_{a_n}^T \tilde{\mathcal{W}}_{a_n} \\
&+ \beta_n
\end{aligned} \tag{80}
$$

where $\lambda_{\max}(\Gamma_n^{-1})$ denotes the largest eigenvalue of Γ_n and $\lambda_{\min}(\Phi_{c_n}\Phi_{c_n}^T)$ denotes the minimal eigenvalue of $\Phi_{c_n}(\xi, \zeta_n)\Phi_{c_n}^T(\xi, \zeta_n)$.

Let $\bar{a}_n = \min\{2(\bar{c}_n - 2), (\sigma_n/\lambda_{max}(\Gamma_n^{-1}), \gamma_{c_n}\lambda_{\min}(\Phi_{c_n}\Phi_{c_n}^T)\}$, then we have

$$
\dot{\mathcal{V}}_n \le \sum_{k=1}^{n}(-\bar{a}_k \mathcal{V}_k + \beta_k) \tag{81}
$$

Based on the above inequality, we directly have

$$
\dot{\mathcal{V}} \le -\bar{a}\mathcal{V} + \varrho \tag{82}
$$

where $\bar{a} = \min\{\bar{a}_1, \ldots, \bar{a}_n\}$, and $\varrho = \sum_{k=1}^{n} \beta_k$. By multiplying (82) with $e^{\bar{a}t}$ and integrating over $[0, t]$, we have

$$
0 \le \mathcal{V}(t) \le \frac{\varrho}{\bar{a}} + \left[\mathcal{V}(0) - \frac{\varrho}{\bar{a}} \right] e^{-\bar{a}t} \tag{83}
$$

The equation (83) indicates that the signals $\zeta_i(t)$, $\tilde{\mathcal{W}}_{f_i}$, $\tilde{\mathcal{W}}_{a_i}$ and $\tilde{\mathcal{W}}_{c_i}$, $i = 1, \ldots, n$, are bounded. Moreover, $|\zeta_1| \le \sqrt{2(\mathcal{V}(0)e^{-\bar{a}t} + \varrho/\bar{a})}$.

Now, we prove that the system output signal can track the desired reference signal. The tracking error of the transformed system (18) is described as

$$
\begin{aligned}
\zeta_1 &= \xi_1 - \xi_r \\
&= \frac{x_1}{d(x_1, F_{1_1}, F_{2_1})} - \frac{y_r}{d(y_r, F_{1_r}, F_{2_r})}
\end{aligned} \tag{84}
$$

The tracking error of the original system is defined as $\bar{e} = x_1 - y_r$, based on assuption (1), then the above equation can be rewrite as

$$
\begin{aligned}
\bar{e} &= d(x_1, F_{1_1}, F_{2_1})\zeta_1 - \frac{d(y_r, F_{1_r}, F_{2_r}) - d(x_1, F_{1_1}, F_{2_1})}{d(y_r, F_{1_r}, F_{2_r})}y_r \\
|\bar{e}| &\le |d(x_1, F_{1_1}, F_{2_1})||\zeta_1| + |1 - \frac{d(x_1, F_{1_1}, F_{2_1})}{d(y_r, F_{1_r}, F_{2_r})}||y_r|
\end{aligned} \tag{85}
$$

According the definition of $d(x, F_1, F_2)$ in (12), we have $d(x_1, F_{1_1}, F_{2_1}) \in (0, 1)$, and $\frac{d(x_1, F_{1_1}, F_{2_1})}{d(y_r, F_{1_r}, F_{2_r})} \in (0, 1)$. Since ζ_1 and y_r is bound, so that the actual tracking error $\bar{e}$ is bound.

Remark 7 The domain of the actual tracking error $\bar{e}$ is given by (85). As can be seen, the actual tracking error depends on ζ_1, y_r, $d(x_1, F_{1_1}, F_{2_1})$ and $d(y_r, F_{1_r}, F_{2_r})$. What's more, under assumption 1, when x_1 well track the desired reference signal y_r, $\frac{d(x_1, F_{1_1}, F_{2_1})}{d(y_r, F_{1_r}, F_{2_r})} \to 1$, $|1 - \frac{d(x_1, F_{1_1}, F_{2_1})}{d(y_r, F_{1_r}, F_{2_r})}||y_r| \to 0$, moreover $|d(x_1, F_{1_1}, F_{2_1})| < 1$, which means the proposed state-dependent transformation function does not affect the actual tracking performance, and even significantly reduces the steady-state tracking error by decreasing δ. However, as argued in [21], the tracking error $\bar{e} = \varpi_1/\varpi_2\zeta_1$ in [25,20](see equation (55) in [20]), the tracking performance will be affected when $\varpi_2 \to 0$.

Remark 8 It is also important to note that by employing the proposed state-dependent transformation function, as argued in [27,29,20,24,48,25,21], the controller design procedure does not require the feasibility condition to be satisfied and the asymmetric time-varying full-state constraint is not violated.

5 Simulation Examples

In this section, the effectiveness of the proposed optimal control scheme are validated by some simulation examples.

5.1 Example 1:

Consider a numerical simulation model as follows

$$\begin{cases} \dot{x}_1 = x_2 + 0.1x_1x_2 \\ \dot{x}_2 = (1 + x_1^2)u + 0.1x_1x_2 - 0.2x_1 & (86) \\ y = x_1 \end{cases}$$

The asymmetric time-varying constraint boundaries are described as $F_{11}(t) = 0.5 - 0.4cos(t)$, $F_{12}(t) = 0.7 + 0.2cos(t)$, $F_{21}(t) = 1.5 - 0.5cos(t)$, $F_{22}(t) = 1.5 + 0.1cos(t)$. The desired reference signal is $y_r = 0.5cos(t)$. The initial states are $x(0) = (0.3, 0)$.

The virtual controller and actual controller of step 1 and step 2, as well as the tuning laws of identifier, critic and actor network are designed according to (23) and (24), respectively. In each step, the structure of neural networks are designed to be same, the centers are split on average in the range $[-8, 8]$ as well. For the first step, there are 24 hidden layer, i.e. $\Phi_{c_1} = [\Phi_{c_1}^1, \ldots, \Phi_{c_1}^{24}]$, the centers are split on average in the range $[-8, 8]$ and the width $\bar{\gamma}_i = 2, i = 1, \ldots, 24$. For the second step, there are 32 hidden layer, i.e. $\Phi_{c_1} = [\Phi_{c_1}^1, \ldots, \Phi_{c_1}^{32}]$, the centers are split on average in the range $[-8, 8]$ and the width $\bar{\gamma}_i = 2, i = 1, \ldots, 32$.

The controller parameters and weight initialization are chosen as $\bar{c}_1 = 8.5, \bar{c}_2 = 8, \hat{\mathcal{W}}_{f_1}(0) = [0.4]_{24\times1}$, $\hat{\mathcal{W}}_{f_2}(0) = [0.8]_{32\times1}, \hat{\mathcal{W}}_{c_1}(0) = [0.4]_{24\times1}, \hat{\mathcal{W}}_{c_2}(0) = [0.6]_{32\times1}$, $\hat{\mathcal{W}}_{a_1}(0) = [1.2]_{24\times1}, \hat{\mathcal{W}}_{a_2}(0) = [1.3]_{32\times1}. \Gamma_1 = 1.8I_{24\times24}, \Gamma_2 = 1.4I_{32\times32}. \sigma_1 = 0.36, \sigma = 0.28. \gamma_{a_1} = 1.8, \gamma_{a_2} = 1.5, \gamma_{c_1} = 1.4, \gamma_{c_2} = 1.3$.

The time-varying state constraint boundary function are $F_{1_1} = -0.5 - 0.2 \cdot 2^{-0.3t} + y_r$, $F_{2_1} = 0.7 + 0.3 \cdot 2^{-0.3t} + y_r$, $F_{2_1} = -0.9 - 0.1sin(t)$, $F_{2_2} = 0.9 + \frac{1}{t}$.
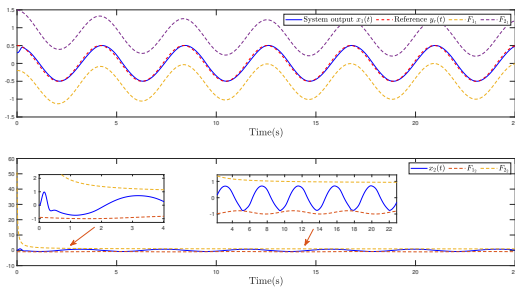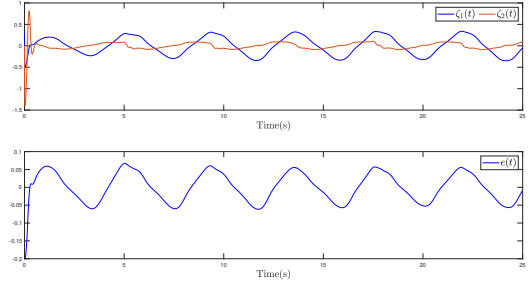

Fig. 3 Tracking errors. (a) Tracking error of transformed system. (b) Actual tracking of the original system.
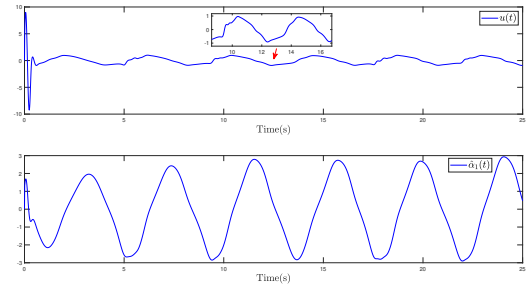
Fig. 4 Control inputs. (a) Actual controller $u(t)$, (b) Virtual controller $\hat{\alpha}_1$
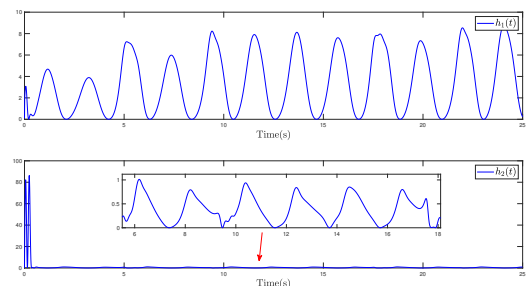

Fig. 2 Trajectories of the state x_1, x_2 and reference signal y_r. (a) Tracking performance of system output, (b) Trajectory of the state x_2.

Fig. 5 Cost functions of the proposed scheme. (a) Cost function of subsystem 1, (b) Cost function of subsystem 2.
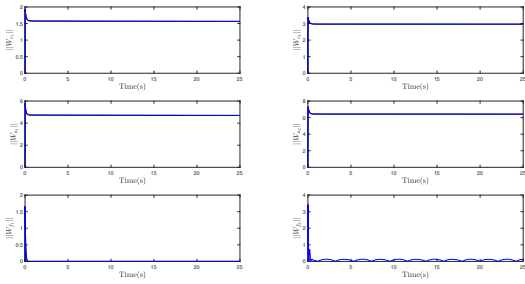
Fig. 6 Weight of the neural networks. (a) $||\hat{\mathcal{W}}_{c_1}||$, (b) $||\hat{\mathcal{W}}_{c_2}||$, (c) $||\hat{\mathcal{W}}_{a_1}||$, (d) $||\hat{\mathcal{W}}_{a_2}||$, (e) $||\hat{\mathcal{W}}_{f_1}||$, (f) $||\hat{\mathcal{W}}_{f_2}||$,

Figs 2-6 graphically illustrate the simulation results. Fig.2 demonstrates how the system output tracking the desired reference signal and the trajectory of state x_2 under the time-varying asymmetric state constraints. The tracking error of the system, including the transformed system and the original system, is given in Fig.3, showing that satisfactory tracking results are obtained. Both virtual controller and actual controller input is presented in Fig.4, which indicates that the feasibility condition is not required to be met for virtual control. The effectiveness of the proposed optimized backstepping controller is showed in Fig.5 and Fig.6, in which we can see that both cost functions and estimated weight vectors can converged rapidly.

5.2 Example 2:

To further verify the efficacy of the proposed scheme, a second set of simulation was conducted on an electromechanical system, which is formed in Table 1 and 2. The electromechanical system is described as

$$
\begin{cases}
\dot{x}_1 = x_2 + x_1^2 \sin(x_2 x_3) \\
\dot{x}_2 = \frac{x_3}{\mathcal{M}} - \frac{\mathcal{N}}{\mathcal{M}} \sin x_1 - \frac{\mathcal{B}}{\mathcal{M}} x_2 + \frac{\mathcal{B}}{\mathcal{M}} \cos x_2 \sin x_3 \\
\dot{x}_3 = \frac{u}{\mathcal{L}} - \frac{\mathcal{K}}{\mathcal{L}} x_2 - \frac{\mathcal{R}}{\mathcal{L}} x_3 \\
y = x_1
\end{cases}
\tag{87}
$$

Table 1 Notation and expressions of the electromechanical system

Notation	Expressions
$\mathcal{M}$	$J/K_T + m/L_0^2/3K_T + M_0 L_0^2/K_T + 2M_0 R_0^2/5K_T$
$\mathcal{B}$	B_0/K_T
$\mathcal{N}$	$mL_0 G/2K_T + M_0 L_0 G/K_T$

Table 2 Parameter and value of the electromechanical system

Parameter	Value	Parameter	Value
J	0.001625	m	0.506
M_0	0.434	L_0	0.305
R_0	0.023	B_0	0.01625
$\mathcal{L}$	0.025	K_T	0.9
$\mathcal{K}$	0.9	R	0.5
G	9.8		

The desired signal is $y_r = sin(0.5t) + 0.5(sin(t))$. The neural networks have the same structure as Example 1, as well as the initial value of weight updating laws. The parameters of the controllers are chosen as $\bar{a}_1 = 4$, $\bar{a}_2 = 3.1$, $\bar{a}_3 = 3.6$, $\sigma_1 = 0.36$, $\sigma_2 = 0.28$, $\sigma_1 = 0.24$, $\gamma_{c_1} = 1.4$, $\gamma_{c_2} = 1.3$, $\gamma_{c_3} = 1.3$, $\gamma_{a_1} = 1.8$, $\gamma_{a_2} = 1.5$, $\gamma_{a_3} = 1.5$. The initialization state is $x(0) = 0$.
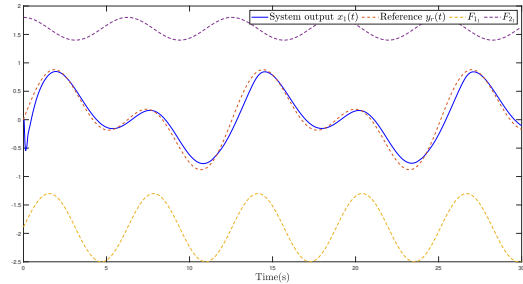

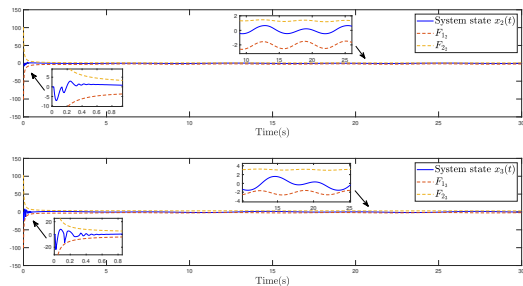**Fig. 7** Tracking performance of system output x_1

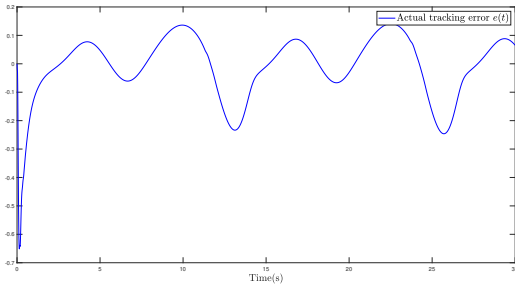**Fig. 8** System states. (a) State x_2, (b) State x_3.

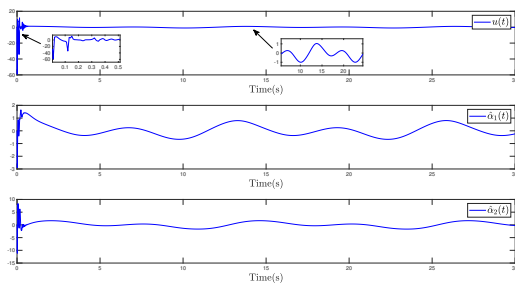**Fig. 9** Actual tracking error of the original system.

Fig. 10 Control inputs of the proposed scheme. (a) Actual control input u, (b) Virtual controller $\hat{\alpha}_1$, (c) Virtual controller $\hat{\alpha}_2$.

Simulation results are graphically illustrated in Fig.7-10. From Fig.7 and Fig.8 can be figured out that system output x_1 can track the desired signal y_r, and system state x_2 and x_3 are subjected in a predefined time-varying boundary. The actual tracking error is presented in Fig.9. And Fig.10 shows the output of the actual controller and virtual controllers.

The simulation results explicitly demonstrate that the proposed optimized back-stepping controller not only can track the desired reference signal well under the time-varying asymmetric state constraint, but also all the closed-loop signals are bounded. Alternatively, the virtual controller in the proposed scheme does not have to fulfill the feasibility condition.

6 Conclusion

This paper investigates the optimal control of nonlinear strict feedback systems subject to time-varying asymmetric state constraints. A novel state-dependent transformation function is proposed, and based on it, the original system is transformed into a new system with the state constraint incorporated. An optimized backstepping controller is designed to track the desired reference signal, and a reinforcement learning algorithm

is used to implement the optimal control, where identifier, critic network and action network are utilized to estimate the uncertain system dynamics, critic the performance and yield the controller. The proposed novel state-dependent transformation function not only avoids feasibility conditions, but also has the ability to simultaneously handle cases with or without state constraints. Simulation examples verify the effectiveness of the proposed transformation function and the optimized controller. In future, the reinforcement learning algorithm of the optimized backstepping controller without relying on the identifier is a worthwhile research problem.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Zhouhua Peng, Jun Wang, and Dan Wang. Distributed Maneuvering of Autonomous Surface Vehicles Based on Neurodynamic Optimization and Fuzzy Approximation. *IEEE Transactions on Control Systems Technology*, 26(3):1083–1090, May 2018.
2. Zhouhua Peng, Jiasen Wang, and Jun Wang. Constrained Control of Autonomous Underwater Vehicles Based on Command Optimization and Disturbance Estimation. *IEEE Transactions on Industrial Electronics*, 66(5):3627–3635, May 2019.
3. Wei He, Yuhao Chen, and Zhao Yin. Adaptive Neural Network Control of an Uncertain Robot With Full-State Constraints. *IEEE Transactions on Cybernetics*, 46(3):620–629, March 2016.

4. Kai Zhao, Yongduan Song, Tiedong Ma, and Liu He. Prescribed Performance Control of Uncertain Euler–Lagrange Systems Subject to Full-State Constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3478–3489, August 2018.

5. A.S. Shiriaev, L.B. Freidovich, and S.V. Gusev. Transverse Linearization for Controlled Mechanical Systems With Several Passive Degrees of Freedom. *IEEE Transactions on Automatic Control*, 55(4):893–906, April 2010.

6. Changyun Wen, Jing Zhou, Zhitao Liu, and Hongye Su. Robust Adaptive Control of Uncertain Nonlinear Systems in the Presence of Input Saturation and External Disturbance. *Ieee Transactions on Automatic Control*, 56(7):1672–1678, July 2011.

7. A. Bemporad. Reference governor for constrained nonlinear systems. *IEEE Transactions on Automatic Control*, 43(3):415–419, March 1998.

8. Mathias Burger and Martin Guay. Robust Constraint Satisfaction for Continuous-Time Nonlinear Systems in Strict Feedback Form. *IEEE Transactions on Automatic Control*, 55(11):2597–2601, November 2010.

9. D Q Mayne, J B Rawlings, C V Rao, and P O M Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, page 26, 2000.

10. Keng Peng Tee, Shuzhi Sam Ge, and Eng Hock Tay. Barrier Lyapunov Functions for the control of output-constrained nonlinear systems. *Automatica*, 45(4):918–927, April 2009.

11. Yan-Jun Liu and Shaocheng Tong. Barrier Lyapunov functions for Nussbaum gain adaptive control of full state constrained nonlinear systems. *Automatica*, 76:143–152, February 2017.

12. Yan-Jun Liu, Shumin Lu, Shaocheng Tong, Xinkai Chen, C.L. Philip Chen, and Dong-Juan Li. Adaptive control-based Barrier Lyapunov Functions for a class of stochastic nonlinear systems with full state constraints. *Automatica*, 87:83–93, January 2018.

13. Dapeng Li, C. L. Philip Chen, Yan-Jun Liu, and Shaocheng Tong. Neural Network Controller Design for a Class of Nonlinear Delayed Systems With Time-Varying Full-State Constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9):2625–2636, September 2019.

14. Jinpeng Yu, Lin Zhao, Haisheng Yu, and Chong Lin. Barrier Lyapunov functions-based command filtered output feedback control for full-state constrained nonlinear systems. *Automatica*, 105:71–79, July 2019.

15. Keng Peng Tee, Shuzhi Sam Ge, and Francis Eng Hock Tay. Adaptive Control of Electrostatic Microactuators With Bidirectional Drive. *IEEE Transactions on Control Systems Technology*, 17(2):340–352, March 2009.

16. Yan-Jun Liu, Jing Li, Shaocheng Tong, and C. L. Philip Chen. Neural Network Control-Based Adaptive Learning Design for Nonlinear Systems With Full-State Constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 27(7):1562–1571, July 2016.

17. Yan-Jun Liu and Shaocheng Tong. Barrier Lyapunov Functions-based adaptive control for a class of nonlinear pure-feedback systems with full state constraints. *Automatica*, 64:70–75, February 2016.

18. Kangkang Sun, Shaoshuai Mou, Jianbin Qiu, Tong Wang, and Huijun Gao. Adaptive Fuzzy Control for Nontriangular Structural Stochastic Switched Nonlinear Systems With Full State Constraints. *IEEE Transactions on Fuzzy Systems*, 27(8):1587–1601, August 2019.

19. Wei Sun, Shun-Feng Su, Yuqiang Wu, Jianwei Xia, and Van-Truong Nguyen. Adaptive Fuzzy Control With High-Order Barrier Lyapunov Functions for High-Order Uncertain Nonlinear Systems With Full-State Constraints. *IEEE Transactions on Cybernetics*, 50(8):3424–3432, August 2020.

20. Kai Zhao and Yongduan Song. Removing the Feasibility Conditions Imposed on Tracking Control Designs for State-Constrained Strict-Feedback Systems. *IEEE Transactions on Automatic Control*, 64(3):1265–1272, March 2019.

21. Ye Cao, Changyun Wen, and Yongduan Song. A Unified Event-Triggered Control Approach for Uncertain Pure-Feedback Systems With or Without State Constraints. *IEEE Transactions on Cybernetics*, 51(3):1262–1271, March 2021.

22. Tao Guo and Xiaowei Wu. Backstepping control for output-constrained nonlinear systems based on nonlinear mapping. *Neural Computing and Applications*, 25(7-8):1665–1674, December 2014.

23. Tianping Zhang, Meizhen Xia, Yang Yi, and Qikun Shen. Adaptive Neural Dynamic Surface Control of Pure-Feedback Nonlinear Systems With Full State Constraints and Dynamic Uncertainties. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(8):2378–2387, August 2017.

24. You Wu, Ruiming Xie, and Xue-Jun Xie. Adaptive finite-time fuzzy control of full-state constrained high-order nonlinear systems without feasibility conditions and its application. *Neurocomputing*, 399:86–95, July 2020.

25. Yang Liu, Huaguang Zhang, Yingchun Wang, and Shaoxin Sun. Adaptive Fuzzy Control for Nonstrict-Feedback Systems Under Asymmetric Time-Varying Full State Constraints Without Feasibility Condition. *IEEE Transactions on Fuzzy Systems*, 29(5):976–985, May 2021.

26. Ting Lei, Wenchao Meng, Kai Zhao, and Long Chen. Adaptive asymptotic tracking control of constrained multi-input multi-output nonlinear systems via event-triggered strategy. *International Journal of Robust and Nonlinear Control*, 31(5):1479–1496, 2021.

27. Kai Zhao, Yongduan Song, and Zhirong Zhang. Tracking control of MIMO nonlinear systems under full state constraints: A Single-parameter adaptation approach free from feasibility conditions. *Automatica*, 107:52–60, September 2019.

28. Ye Cao, Yongduan Song, and Changyun Wen. Practical tracking control of perturbed uncertain nonaffine systems with full state constraints. *Automatica*, 110:108608, December 2019.

29. Dapeng Li, Lei Liu, Yan-Jun Liu, Shaocheng Tong, and C. L. Philip Chen. Adaptive NN Control Without Feasibility Conditions for Nonlinear State Constrained Stochastic Systems With Unknown Time Delays. *IEEE Transactions on Cybernetics*, 49(12):4485–4494, December 2019.

30. Yangang Yao, Jieqing Tan, Jian Wu, and Xu Zhang. Event-triggered fixed-time adaptive fuzzy control for state-constrained stochastic nonlinear systems without feasibility conditions. *Nonlinear Dynamics*, 105(1):403–416, July 2021.

31. Frank L. Lewis, Draguna Vrabie, and Kyriakos G. Vamvoudakis. Reinforcement Learning and Feedback Control: Using Natural Decision Methods to Design Optimal Adaptive Controllers. *IEEE Control Systems Magazine*, 32(6):76–105, December 2012.

32. Frank L. Lewis and Draguna Vrabie. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits and Systems Magazine*, 9(3):32–50, 23.

33. Fei-Yue Wang, Huaguang Zhang, and Derong Liu. Adaptive Dynamic Programming: An Introduction. *IEEE Computational Intelligence Magazine*, 4(2):39–47, May 2009.

34. Hua-Guang Zhang, Xin Zhang, Yan-Hong Luo, and Jun Yang. An Overview of Research on Adaptive Dynamic Programming: An Overview of Research on Adaptive Dynamic Programming. *Acta Automatica Sinica*, 39(4):303–311, March 2014.

35. Kangkang Sun, Shuai Sui, and Shaocheng Tong. Optimal adaptive fuzzy FTC design for strict-feedback nonlinear uncertain systems with actuator faults. *Fuzzy Sets and Systems*, 316:20–34, June 2017.

36. Jingliang Sun and Chunsheng Liu. Backstepping-based adaptive dynamic programming for missile-target guidance systems with state and input constraints. *Journal of the Franklin Institute*, 355(17):8412–8440, November 2018.

37. Shaohua Luo, Frank L. Lewis, Yongduan Song, and Hassen M. Ouakad. Accelerated Adaptive Fuzzy Optimal Control of Three Coupled Fractional-Order Chaotic Electromechanical Transducers. *IEEE Transactions on Fuzzy Systems*, 29(7):1701–1714, July 2021.

38. Zheng Wang, Xiaoping Liu, Kefu Liu, Shuai Li, and Huanqing Wang. Backstepping-Based Lyapunov Function Construction Using Approximate Dynamic Programming and Sum of Square Techniques. *IEEE Transactions on Cybernetics*, 47(10):3393–3403, October 2017.

39. Guoxing Wen, C. L. Philip Chen, Jun Feng, and Ning Zhou. Optimized Multi-Agent Formation Control Based on an Identifier–Actor–Critic Reinforcement Learning Algorithm. *IEEE Transactions on Fuzzy Systems*, 26(5):2719–2731, October 2018.

40. Yongchao Liu, Qidan Zhu, and Guoxing Wen. Adaptive Tracking Control for Perturbed Strict-Feedback Nonlinear Systems Based on Optimized Backstepping Technique. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2020.

41. Guoxing Wen, Liguang Xu, and Bin Li. Optimized Backstepping Tracking Control Using Reinforcement Learning for a Class of Stochastic Nonlinear Strict-Feedback Systems. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2021.

42. Guoxing Wen, Shuzhi Sam Ge, and Fangwen Tu. Optimized Backstepping for Tracking Control of Strict-Feedback Systems. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8):3850–3862, August 2018.

43. Guoxing Wen, C. L. Philip Chen, and Shuzhi Sam Ge. Simplified Optimized Backstepping Control for a Class of Nonlinear Strict-Feedback Systems With Unknown Dynamic Functions. *IEEE Transactions on Cybernetics*, pages 1–14, 2020.

44. Guoxing Wen, C. L. Philip Chen, and Bin Li. Optimized Formation Control Using Simplified Reinforcement Learning for a Class of Multiagent Systems With Unknown Dynamics. *IEEE Transactions on Industrial Electronics*, 67(9):7879–7888, September 2020.

45. Guoxing Wen and C. L. Philip Chen. Optimized Backstepping Consensus Control Using Reinforcement Learning for a Class of Nonlinear Strict-Feedback-Dynamic Multi-Agent Systems. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2021.

46. Yongming Li, Yanli Fan, Kewen Li, Wei Liu, and Shaocheng Tong. Adaptive Optimized Backstepping Control-Based RL Algorithm for Stochastic Nonlinear Systems With State Constraints and Its Application. *IEEE Transactions on Cybernetics*, pages 1–14, 2021.

47. Jiaxin Zhang, Kewen Li, and Yongming Li. Output-Feedback Based Simplified Optimized Backstepping Control for Strict-Feedback Systems with Input and State Constraints. *IEEE/CAA Journal of Automatica Sinica*, 8(6):1119–1132, June 2021.

48. Kai Zhao and Yongduan Song. Neuroadaptive Robotic Control Under Time-Varying Asymmetric Motion Constraints: A Feasibility-Condition-Free Approach. *IEEE Transactions on Cybernetics*, 50(1):15–24, January 2020.

49. Zhijun Li, Junjun Li, Suna Zhao, Yuxia Yuan, Yu Kang, and C. L. Philip Chen. Adaptive Neural Control of a Kinematically Redundant Exoskeleton Robot Using Brain–Machine Interfaces. *IEEE Transactions on Neural Networks and Learning Systems*, 30(12):3558–3571, December 2019.