Information Sciences 621 (2023) 407–423

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Adaptive reinforcement learning optimal tracking control for strict-feedback nonlinear systems with prescribed performance

Zongsheng Huang [a], Weiwei Bai [a], Tieshan Li [a,b,c,*], Yue Long [a], C.L. Philip Chen [d,a], Hongjing Liang [a], Hanqing Yang [a]

[a] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China
[b] Yangtze Delta Region Institute, University of Electronic Science and Technology of China, Huzhou 313000, China
[c] Laboratory of Electromagnetic Space Cognition and Intelligent Control, Beijing 100089, China
[d] School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510641, China

ARTICLE INFO

Article history:
Received 28 June 2022
Received in revised form 22 November 2022
Accepted 24 November 2022
Available online 29 November 2022

Keywords:
Reinforcement learning
Prescribed performance control
Adaptive dynamic programming
Tracking control
Strict-feedback nonlinear systems

ABSTRACT

The reinforcement learning-based prescribed performance optimal tracking control problem is considered for a class of strict-feedback nonlinear systems in this paper. The unknown nonlinearities and cost function are approximated by radial-basis-function (RBF) neural network (NN). The overall controller consists of an adaptive controller and an optimal compensation term. Firstly, the adaptive controller is designed by backstepping control method. Subsequently, the optimal compensation term is derived via policy iteration by minimizing cost function. In addition, depending on the prescribed performance control, the tracking error can be limited in the prescribed area. Therefore, the whole control scheme can effectively guarantee that the tracking error converges to a bound with prescribed performance while the cost function is minimized. The stability analysis shows that all signals in the closed-loop system are bounded. Finally, the effectiveness and advantages of the designed control strategy are illustrated by the simulation examples.

1. Introduction

In practice, many classical systems in real physical world can be described as the strict-feedback nonlinear form, such as the underwater vehicles [1], the quadrotor unmanned aerial vehicles [2], the flexible joint robots [3], the on-orbit spacecrafts [4] and so on. Therefore, the control problems for strict-feedback nonlinear systems have become one of the hot issues, and many control methods have been explored in the last decades. Especially, backstepping control technique has been witnessed to be a powerful technique to cope with the control problems for strict-feedback nonlinear systems with parametric uncertainties [5,6]. To handle the huge challenge induced by the overall nonlinear uncertainties in the process of controller design, the neural networks (NNs) approximation-based adaptive backstepping control methodology has been developed rapidly due to its universal approximation ability of NNs [7–9]. For instance, for the nonlinear systems with strict-feedback form, Ge and Wang in [10] proposed the NNs-based adaptive control method. In [11], a robust adaptive neural control scheme was studied for the investigated system with unknown nonlinearities. Recently, in [12], based on a modified

* Corresponding author at: School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China;Yangtze Delta Region Institute, University of Electronic Science and Technology of China, Huzhou 313000, China; Laboratory of Electromagnetic Space Cognition and Intelligent Control, Beijing 100089, China.
E-mail address: tieshanli@126.com (T. Li).

https://doi.org/10.1016/j.ins.2022.11.109
0020-0255/

backstepping technique, an adaptive NN controller was designed. The adaptive control scheme on the basis of NNs had been extended to the practical control problem of quadrotor in [13]. Nevertheless, the designed controller also needs to have the ability of optimization due to the significant requirements in engineering practice.

Optimal control is a popular control method and has been applied widely [14–17]. In general, one of the classical optimal control methods is dynamic programming, which obtains the optimal solution by solving the Ricatti equation for linear systems. As for the nonlinear systems, the optimal controller can be derived by solving the Hamilton–Jacobian-Bellman (HJB) equation. However, the HJB equation is a tough partial differential equation to deal with. To overcome this limitation, many significant methods based on reinforcement learning (RL) come into being, such as the temporal-difference learning method [18], Q-learning method [19], adaptive dynamic programming (ADP) method [20], and so on. It should be noted that the RL method can interact with environment and modify control policies so that the successful control decisions can be remembered or used again by constructing a reinforcement signal [21–23]. Recently, some remarkable results on RL-based optimal control schemes have been given [24–32]. For example, in [26], the RL control scheme was presented to study the H_∞ optimization problem. The optimal control problem was solved in [27] by using RL algorithm for marine surface vessel system. In the works of [28,29], the RL-based robust control scheme was proposed for multi-agent systems and nonlinear systems with input constraints and disturbances, respectively. Among these results, the optimal performance function was approximated with the aid of critic NN, thereby getting the optimal solution. Besides, the RL-based actor-critic approach can be employed to obtain optimal solution without involving HJB equation. Deng et al. in [30] proposed a novel actor-critic method to solve the tracking control problem for autonomous underwater vehicles. A new integral RL-based actor-critic strategy was designed in [31] for the tracking control problem of multi-input multi-output (MIMO) nonlinear systems. In [32], an adaptive optimal controller based on sliding-mode surface was studied with the help of actor-critic strategy for a class of switched nonlinear systems. These results make significant contributions on the progress of optimal control based on RL method. It is worth mentioning that the aforementioned control schemes do not discuss the problem of performance constraints, such as convergence rate, overshoot and maximum steady state error, which are significant in view of engineering.

The prescribed performance (PP) technique is an effective method to restrict the tracking error within the prescribed domain and achieve performance constraints. It was firstly proposed in [33] for SISO strict-feedback nonlinear system. Subsequently, the PP technique was applied in MIMO strict-feedback nonlinear systems in [34]. Recently, the PP method has been widely employed in nonlinear systems control. In [35], the backstepping technique combined with the PP technique was utilized in the single-link flexible-joint robotic manipulator. The PP method and sliding mode control technique were integrated to guarantee the tracking performance in [36]. For the tracking control problem of nonlinear systems, the finite-time control strategy combined with PP and the fixed-time control strategy combined with PP were studied in [37,38], respectively. Nevertheless, these works pay little attention in consideration of the PP technique and optimal control problem simultaneously for strict-feedback nonlinear systems.

Inspired by above observations, in this paper, we are motivated to investigate the PP optimal tracking control problem based on RL for a class of strict-feedback nonlinear systems. The designed controller includes the adaptive controller and the optimal compensation term. The adaptive controller is derived based on the backstepping method and the optimal compensation term is established via the RL-based policy iteration algorithm. PP method is introduced to guarantee the tracking error within the predefined range. The main contributions of this paper are summarized as follows:

(1) To guarantee the convergence rate, tracking error and the maximum overshoot within the predefined range, a performance function is adopted in the controller design. Compared with [39–42], the designed controller can make a balance between cost and tracking performance with prescribed performance, thus both the transient state and steady state performance of the tracking error have been obtained simultaneously.
(2) In this paper, the backstepping method is integrated with the policy iteration strategy for the investigated system. The proposed controller consists of the adaptive controller and the optimal compensation term, which guarantees the stability of the closed-loop system and achieves the optimal tracking performance, respectively.
(3) Different from the tracking problems without considering the optimization and PP technique for the considered system, the tracking error can be limited in the bound with PP while the cost function is minimized by the designed control scheme.

The remainder of the paper is organized as follows. The problem formulation and preliminaries are detailed in Section 2. In Section 3, the adaptive controller and the optimal compensation term are designed. The stability analysis is given in Section 4. The simulation results and conclusion are shown in Sections 5 and 6, respectively.

2. Problem formulation and preliminaries

In this section, the model of controlled plant is formulated, the necessary knowledge of RBF NN and prescribed performance technique are introduced.

2.1. System model

Consider the investigated system as follows

$$\begin{aligned}\dot{x}_i &= x_{i+1} + f_i(\bar{x}_i), 1 \leqslant i \leqslant n-1\\ \dot{x}_n &= u + f_n(\bar{x}_n)\\ y &= x_1\end{aligned} \tag{1}$$

where $\bar{x}_i = [x_1, x_2, \ldots, x_i]^T \in \mathbb{R}^i, i = 1, \ldots, n$, denotes the state vector, $u \in \mathbb{R}$ and $y \in \mathbb{R}$ respectively represent the input and output of the system. $f_i(\bar{x}_i) \in \mathbb{R}, i = 1, \ldots, n$ denotes the unknown smooth continuous nonlinear function.

In this paper, it is necessary to make the following assumption to facilitate the controller design in this paper.

Assumption 1. The $y_r(t), \dot{y}_r(t)$ and $\ddot{y}_r(t)$ represent the desired signal and its derivatives, which are known and bounded.

2.2. RBF NN

RBF NN can be described with the form of $W^T S(X)$. As pointed out in [7], RBF NN has the universal approximation property in the sense that if the node number is large enough, it can approximate an unknown continuous function $F_{nn}(X)$ as follows

$$F_{nn}(X) = W^T S(X) \tag{2}$$

where $X \in R^n$ is an input vector, $W = [w_1, \cdots, w_l] \in R^l$ is weight vector, $S(X) = [s_1(X), \cdots, s_l(X)]$ is the basis function vector, and l denotes the number of nodes in hidden layer.

Generally, $s_i(X)$ is selected as Gaussian basis function, we have

$$s_i(X) = \exp\left(-(X - b_i)^T (X - b_i)/2\varsigma_i^2\right), \quad i = 1, \ldots, l$$

where b_i is the center of the receptive field, and ς_i represents the width of the Gaussian kernel.

According to [7], RBF NN can uniformly approximate any unknown continuous function $F(X)$ with any accuracy over a compact set as

$$F(X) = W^{*T} S(X) + \xi(X) \tag{3}$$

where W^* is the ideal weight vector, $\xi(X)$ denotes the bounded estimation error.

2.3. Prescribed performance (PP) technique

The aim to introduce the PP technique is to ensure the steady-state tracking performance and keep the tracking error $z_1(t)$ within the prescribed bound, and z_1 is defined as $z_1 = x_1 - y_r$.

According to [43], the performance of the system satisfies the prescribed performance requirements if the restricted tracking error $z_1(t)$ meets the condition as follows

$$-\rho_{min}\eta(t) < z_1(t) < \rho_{max}\eta(t), \forall t \geqslant 0 \tag{4}$$

where $\rho_{\min}$ and ρ_{max} are positive adjustable parameters, $\eta(t)$ denotes the performance function and is given as the following form

$$\eta(t) = (\eta_0 - \eta_\infty)e^{-kt} + \eta_\infty$$

where $k > 0, \eta_0 = \eta(0), \eta_\infty = \lim_{t\to\infty}\eta(t)$, and $\eta_0 > \eta_\infty > 0, -\rho_{min}\eta(0) < z_1(0) < \rho_{max}\eta(0)$.

To achieve the performance condition (4), a strictly monotonically increasing smooth function $\varpi(\bar{z}_1(t))$ is designed as

$$\varpi(\bar{z}_1(t)) = \frac{\rho_{max}e^{\bar{z}_1} - \rho_{min}e^{-\bar{z}_1}}{e^{\bar{z}_1} + e^{-\bar{z}_1}}$$

where $\bar{z}_1(t)$ is the unrestricted error [43].

The following equality is introduced to transform the condition (4) as

$$z_1(t) = \eta(t)\varpi(\bar{z}_1(t)), \forall t \geqslant 0 \tag{5}$$

From (5), we have

$$\bar{z}_1(t) = \varpi^{-1}\left(\frac{z_1(t)}{\eta(t)}\right) = \frac{1}{2}\ln\frac{\varpi + \rho_{min}}{\rho_{max} - \varpi} \tag{6}$$

Take the derivative of the unrestricted error $\bar{z}_1(t)$, the relationship between the restricted error $z_1(t)$ and unrestricted error $\bar{z}_1(t)$ can be obtained as

$$\dot{\bar{z}}_1(t) = p\left(\dot{z}_1(t) - \frac{\dot{\eta}(t)z_1(t)}{\eta(t)}\right) \tag{7}$$

where $p = \frac{1}{2\eta}\left(\frac{1}{\varpi+\rho_{min}} - \frac{1}{\varpi-\rho_{max}}\right)$.

3. Main results

In this section, inspired by [39], the controller design includes two parts, i.e., an adaptive controller u^f via backstepping method will be given firstly. Subsequently, an optimal compensation term u^* will be derived by the policy iteration-based RL technique. The controller consists of an adaptive controller and an optimal compensation term, i.e., $u = u^f + u^*$.

3.1. Adaptive controller design

Compared with the traditional backstepping method, we consider the optimal control problem by minimizing the cost function in this paper. Therefore, the controller consists of an adaptive controller and an optimal compensation term, where the adaptive controller will be designed by the backstepping method in this subsection, the optimal compensation term will be derived in the next subsection.

Define the following conversion of coordinates

$$\begin{aligned} z_1 &= x_1 - y_r \\ z_i &= x_i - x_{id}, \quad i = 2, \ldots, n \end{aligned} \tag{8}$$

where y_r is a desired signal, x_{id} is the virtual controller and consists of the adaptive virtual controller x_{id}^f and optimal virtual compensation term x_{id}^*, i.e., $x_{id} = x_{id}^f + x_{id}^*$. The tracking error z_1 satisfies the PP condition (4).

Step 1: Define $y_r = x_{1d}$, based on the prescribed performance transformation between the restricted error $z_1(t)$ and unrestricted error $\bar{z}_1(t)$ mentioned in (7), the derivative of $\bar{z}_1$ is described as

$$\begin{aligned} \dot{\bar{z}}_1 &= p\left(\dot{z}_1 - \tfrac{\dot{\eta}z_1}{\eta}\right) \\ &= p\left(x_2 + f_1(x_1) - \dot{x}_{1d} - \tfrac{\dot{\eta}z_1}{\eta}\right) \\ &= p\left(z_2 + x_{2d}^f + x_{2d}^* + f_1(x_{1d}) + o_1(Z_1) - \dot{x}_{1d} - \tfrac{\dot{\eta}z_1}{\eta}\right) \end{aligned} \tag{9}$$

where $o_1(Z_1) = f_1(x_1) - f_1(x_{1d}), Z_1 = \bar{z}_1$.

Due to the fact that $f_1(x_{1d})$ is the unknown continuous function, we apply the RBF NN (3) to approximate it as follows

$$f_1(x_{1d}) = W_1^{*T} S_1(x_{1d}) + \xi_1(x_{1d})$$

where $|\xi_1(x_{1d})| < \zeta_1$ and $\zeta_1 > 0$.

Define $\hat{f}_1(x_{1d}) = \widehat{W}_1^T S_1(x_{1d})$ as the estimation of $f_1(x_{1d})$, $\widehat{W}_1$ is the estimated value of ideal weight vector W_1^*, and W_1^* is the constant vector. Let $\widetilde{W}_1 = W_1^* - \widehat{W}_1$ be the parameter estimation error and one has $\dot{\widetilde{W}}_1 = -\dot{\widehat{W}}_1$.

Consider the following Lyapunov function

$$V_1 = \frac{1}{2}\bar{z}_1^2 + \frac{1}{2}\widetilde{W}_1^T\widetilde{W}_1. \tag{10}$$

Applying Young's inequality, one has

$$\bar{z}_1 p\xi_1(x_{1d}) \leqslant \frac{p^2\bar{z}_1^2}{2} + \frac{\zeta_1^2}{2}.$$

Take the derivative of V_1, one has

$$\begin{aligned} \dot{V}_1 &= \bar{z}_1\dot{\bar{z}}_1 - \widetilde{W}_1^T\dot{\widehat{W}}_1 \\ &= \bar{z}_1 p\left(z_2 + x_{2d}^f + x_{2d}^* + f_1(x_{1d}) + o_1(Z_1) - \dot{x}_{1d} - \tfrac{\dot{\eta}z_1}{\eta}\right) - \widetilde{W}_1^T\dot{\widehat{W}}_1 \\ &= \bar{z}_1 p\left(z_2 + x_{2d}^f + x_{2d}^* + \widehat{W}_1^T S_1(x_{1d}) + \widetilde{W}_1^T S_1(x_{1d}) + \xi_1(x_{1d}) + o_1(Z_1) - \dot{x}_{1d} - \tfrac{\dot{\eta}z_1}{\eta}\right) - \widetilde{W}_1^T\dot{\widehat{W}}_1 \\ &\leqslant \bar{z}_1 p\left(z_2 + x_{2d}^f + x_{2d}^* + \widehat{W}_1^T S_1(x_{1d}) + \widetilde{W}_1^T S_1(x_{1d}) + \tfrac{\bar{z}_1 p}{2} + o_1(Z_1) - \dot{x}_{1d} - \tfrac{\dot{\eta}z_1}{\eta}\right) + \tfrac{\zeta_1^2}{2} - \widetilde{W}_1^T\dot{\widehat{W}}_1. \end{aligned} \tag{11}$$

The adaptive virtual controller x_{2d}^f and the adaptive law $\dot{\widehat{W}}_1$ are respectively designed as follows

$$x_{2d}^{f} = -\frac{c_1\bar{z}_1}{p} - \frac{p\bar{z}_1}{2} - \widehat{W}_1^T S_1(x_{1d}) + \dot{x}_{1d} + \frac{\dot{\eta}z_1}{\eta} \tag{12}$$

$$\dot{\widehat{W}}_1 = p\bar{z}_1 S_1(x_{1d}) - \sigma_1 \widehat{W}_1 ||\widehat{W}_1||^2 \tag{13}$$

where $c_1 > 0$ and $\sigma_1 > 0$.

Following from (12) and (13), (11) can be rewritten as

$$\dot{V}_1 \leqslant -c_1\bar{z}_1^2 + p\bar{z}_1 z_2 + p\bar{z}_1 x_{2d}^* + p\bar{z}_1 o_1(Z_1) + \frac{\zeta_1^2}{2} + \sigma_1 \widetilde{W}_1^T \widehat{W}_1 ||\widehat{W}_1||^2. \tag{14}$$

*Step*2 : From $z_2 = x_2 - x_{2d}$, its derivative is

$$\dot{z}_2 = \dot{x}_2 - \dot{x}_{2d} = z_3 + x_{3d}^f + x_{3d}^* + f_2(\bar{x}_{2d}) + o_2(Z_2) - \dot{x}_{2d} \tag{15}$$

where $\bar{x}_{2d} = [x_{1d}, x_{2d}]^T, o_2(Z_2) = f_2(\bar{x}_2) - f_2(\bar{x}_{2d}), Z_2 = [\bar{z}_1, z_2]^T$.

Motivated by [10], from (15), the lumped uncertain parts $f_2(\bar{x}_{2d}) - \dot{x}_{2d}$ can be denoted as $F_2(\bar{x}_{2d})$ and approximated by the RBF NN as follows

$$F_2(\bar{x}_{2d}) = W_2^{*T} S_2(\bar{x}_{2d}) + \xi_2(\bar{x}_{2d})$$

where $|\xi_2(x_{2d})| < \zeta_2$ and $\zeta_2 > 0$.

Define $\widehat{F}_2(\bar{x}_{2d}) = \widehat{W}_2^T S_2(\bar{x}_{2d})$ as the estimation of $F_2(\bar{x}_{2d})$, $\widehat{W}_2$ is the estimated value of ideal weight vector W_2^* and W_2^* is the constant vector. Let $\widetilde{W}_2 = W_2^* - \widehat{W}_2$ be the parameter approximation error, and one has $\dot{\widetilde{W}}_2 = -\dot{\widehat{W}}_2$.

Consider the following Lyapunov function

$$V_2 = \frac{1}{2}z_2^2 + \frac{1}{2}\widetilde{W}_2^T\widetilde{W}_2 + V_1. \tag{16}$$

Using Young's inequality, one has

$$z_2\xi_2(\bar{x}_{2d}) \leqslant \frac{z_2^2}{2} + \frac{\zeta_2^2}{2}.$$

Take the derivative of V_2, one has

$$\begin{aligned}
\dot{V}_2 &= z_2\dot{z}_2 - \widetilde{W}_2^T\dot{\widehat{W}}_2 + \dot{V}_1 \\
&= z_2\Big(z_3 + x_{3d}^f + x_{3d}^* + \widehat{W}_2^T S_2(\bar{x}_{2d}) + \widetilde{W}_2^T S_2(\bar{x}_{2d}) + \xi_2(\bar{x}_{2d}) + o_2(Z_2)\Big) \\
&\quad -\widetilde{W}_2^T\dot{\widehat{W}}_2 + \dot{V}_1 \\
&= z_2\Big(z_3 + x_{3d}^f + x_{3d}^* + \widehat{W}_2^T S_2(\bar{x}_{2d}) + \widetilde{W}_2^T S_2(\bar{x}_{2d}) + o_2(Z_2) + \tfrac{z_2}{2}\Big) - \widetilde{W}_2^T\dot{\widehat{W}}_2 \\
&\quad -c_1\bar{z}_1^2 + p\bar{z}_1 z_2 + p\bar{z}_1 x_{2d}^* + p\bar{z}_1 o_1(Z_1) + \tfrac{\zeta_1^2}{2} + \tfrac{\zeta_2^2}{2} + \sigma_1\widetilde{W}_1^T\widehat{W}_1||\widehat{W}_1||^2.
\end{aligned} \tag{17}$$

The adaptive virtual controller x_{3d}^f and the adaptive law $\dot{\widehat{W}}_2$ are respectively given as follows

$$x_{3d}^f = -c_2 z_2 - p\bar{z}_1 - \frac{z_2}{2} - \hat{W}_2^T S_2(\bar{x}_{2d}) \tag{18}$$

$$\dot{\hat{W}}_2 = z_2 S_2(\bar{x}_{2d}) - \sigma_2\hat{W}_2\left\|\hat{W}_2\right\|^2 \tag{19}$$

where $c_2 > 0$ and $\sigma_2 > 0$.

According to (18) and (19), (17) can be rewritten as

$$\begin{aligned}
\dot{V}_2 \leqslant\ & -c_1\bar{z}_1^2 - c_2 z_2^2 + z_2 z_3 + p\bar{z}_1 o_1(Z_1) + p\bar{z}_1 x_{2d}^* + z_2 o_2(Z_2) + z_2 x_{3d}^* \\
& +\sigma_1\tilde{W}_1^T\hat{W}_1\left\|\hat{W}_1\right\|^2 + \sigma_2\tilde{W}_2^T\hat{W}_2\left\|\hat{W}_2\right\|^2 + \tfrac{\zeta_1^2}{2} + \tfrac{\zeta_2^2}{2}.
\end{aligned} \tag{20}$$

Stepi : $(i = 3, \cdots, n-1)$ For $z_i = x_i - x_{id}$, the derivative of z_i is expressed as

$$\dot{z}_i = \dot{x}_i - \dot{x}_{id} = z_{(i+1)} + x_{(i+1)d}^f + x_{(i+1)d}^* + f_i(\bar{x}_{id}) + o_i(Z_i) - \dot{x}_{id} \tag{21}$$

where $\bar{x}_{id} = [x_{1d}, \cdots, x_{id}]^T, o_i(Z_i) = f_i(\bar{x}_i) - f_i(\bar{x}_{id}), Z_i = [\bar{z}_1, \cdots, z_i]^T$.

Similarly, the lumped uncertain parts $f_i(\bar{x}_{id}) - \dot{x}_{id}$ can be denoted as $F_i(\bar{x}_{id})$ and approximated by the RBF NN as follows

$$F_i(\bar{x}_{id}) = W_i^{*T} S_i(\bar{x}_{id}) + \xi_i(\bar{x}_{id}) \tag{22}$$

where $|\xi_i(x_{id})| < \zeta_i$ and $\zeta_i > 0$.

Define $\widehat{F}_i(\bar{x}_{id}) = \widehat{W}_i^T S_i(\bar{x}_{id})$ as the estimation value of $F_i(\bar{x}_{id})$, $\widehat{W}_i$ is the estimation weight vector of W_i^* and W_i^* is the constant vector. Let $\widetilde{W}_i = W_i^* - \widehat{W}_i$ be the parameter approximation error, and one has $\dot{\widetilde{W}}_i = -\dot{\widehat{W}}_i$.

Choose the Lyapunov function as follows

$$V_i = \frac{1}{2}z_i^2 + \frac{1}{2}\widetilde{W}_i^T\widetilde{W}_i + V_{i-1}. \tag{23}$$

According to Young's inequality, one has

$$z_i\xi_i(\bar{x}_{id}) \leqslant \frac{z_i^2}{2} + \frac{\zeta_i^2}{2}.$$

The adaptive virtual controller $x_{(i+1)d}^f$ and the adaptive law $\dot{\widehat{W}}_i$ are respectively designed as follows

$$x_{(i+1)d}^f = -c_i z_i - z_{i-1} - \frac{z_i}{2} - \widehat{W}_i^T S_i(\bar{x}_{id}) \tag{24}$$

$$\dot{\hat{W}}_i = z_i S_i(\bar{x}_{id}) - \sigma_i \widehat{W}_i \left\| \hat{W}_i \right\|^2 \tag{25}$$

where $c_i > 0$ and $\sigma_i > 0$.

It follows from (24)–(25), the derivative of V_i is

$$\begin{aligned} \dot{V}_i \leqslant\ & \sum_{j=1}^{i}\left(\tfrac{\zeta_j^2}{2} + \sigma_j\widetilde{W}_j^T\hat{W}_j\|\hat{W}_j\|^2\right) + z_i z_{(i+1)} - c_1\bar{z}_1^2 + p\bar{z}_1 o_1(Z_1) + p\bar{z}_1 x_{2d}^* \\ & + \sum_{j=2}^{i}\left(-c_j z_j^2 + z_j o_j(Z_j) + z_j x_{(j+1)d}^*\right). \end{aligned} \tag{26}$$

Stepn : Similarly, the derivative of z_n is

$$\dot{z}_n = \dot{x}_n - \dot{x}_{nd} = u^f + u^* + f_n(\bar{x}_{nd}) + o_n(Z_n) - \dot{x}_{nd} \tag{27}$$

where $\bar{x}_{nd} = [x_{1d}, \cdots, x_{nd}]^T$, $o_n(Z_n) = f_n(\bar{x}_n) - f_n(\bar{x}_{nd}), Z_n = [\bar{z}_1, \cdots, z_n]^T$.

Remark 1. In this paper, $f_i(\bar{x}_i) - f_i(\bar{x}_{id})$ is defined as $o_i(Z_i), i = 1, \cdots, n$, which facilitates the design of the optimal compensation term in the next subsection.

The lumped uncertain parts $f_n(\bar{x}_{nd}) - \dot{x}_{nd}$ can be denoted as $F_n(\bar{x}_{nd})$ and approximated by the RBF NN as follows

$$F_n(\bar{x}_{nd}) = W_n^{*T} S_n(\bar{x}_{nd}) + \xi_n(\bar{x}_{nd}) \tag{28}$$

where $|\xi_n(x_{nd})| < \zeta_n$ and $\zeta_n > 0$.

Define $\widehat{F}_n(\bar{x}_{nd}) = \widehat{W}_n^T S_n(\bar{x}_{nd})$ as the estimation of $F_n(\bar{x}_{nd})$, $\widehat{W}_n$ is the estimated value of ideal weight vector W_n^* and W_n^* is the constant vector. Let $\widetilde{W}_n = W_n^* - \widehat{W}_n$ be the parameter approximation error, and one has $\dot{\widetilde{W}}_n = -\dot{\widehat{W}}_n$.

Select the Lyapunov function with following form

$$V_n = \frac{1}{2}z_n^2 + \frac{1}{2}\widetilde{W}_n^T\widetilde{W}_n + V_{n-1}. \tag{29}$$

The u^f and the adaptive law $\dot{\widehat{W}}_n$ are respectively given as follows

$$u^f = -c_n z_n - z_{n-1} - \frac{z_n}{2} - \widehat{W}_n^T S_n(\bar{x}_{nd}) \tag{30}$$

$$\dot{\hat{W}}_n = z_n S_n(\bar{x}_{nd}) - \sigma_n \hat{W}_n \left\| \hat{W}_n \right\|^2 \tag{31}$$

where $c_n > 0$ and $\sigma_n > 0$.

According to (30) and (31), the derivative of V_n is

$$\begin{aligned} \dot{V}_n \leqslant\ & \sum_{i=1}^{n}\left(\tfrac{\zeta_i^2}{2} + \sigma_i\widetilde{W}_i^T\hat{W}_i\|\hat{W}_i\|^2\right) - c_1\bar{z}_1^2 - \sum_{i=2}^{n}c_i z_i^2 + p\bar{z}_1\left(o_1(Z_1) + x_{2d}^*\right) \\ & + \sum_{i=2}^{n-1} z_i\left(o_i(Z_i) + x_{(i+1)d}^*\right) + z_n(u^* + o_n(Z_n)). \end{aligned} \tag{32}$$

Applying Young's inequality, one has

$$\sum_{i=1}^{n}\widetilde{W}_i^T\widehat{W}_i||\widehat{W}_i||^2 \leqslant -\frac{1}{10}||\widetilde{W}||^4+\frac{1}{2}||W||^4 \tag{33}$$

Policy Iteration Algorithm

Step1: Initialization. Give an initial stabilizing control protocol $U^{*0}\in\Omega$
Step2: Policy evaluation. Solve the cost function $V^l(Z)$ as follows:
$H(Z,U^*)=Z^TQZ+U^{*T}RU^*+\nabla_Z^TV(Z)P(o(\mathscr{Z})+U^*)$=0
Step3: Policy improvement. Update the control protocol $U^{*(l+1)}$ as follows:
$U^{*(l+1)}=\underset{U^*\in\Omega}{argmin}\,H(Z,U^*)=-\frac{1}{2}R^{-1}P^T\nabla_ZV^l(Z)$
Step4: If $||V^{l+1}(Z)-V^l(Z)||\leqslant\varepsilon$ with the predefined constant ε,stop the iteration; otherwise, set $l=l+1$ and go back to Step 2 and continue.

Substituting (33) into (32), we have

$$\begin{aligned}\dot{V}_n &\leqslant -\alpha_1||Z||^2-\tfrac{\sigma_{max}}{10}||\widetilde{W}||^4+\sum_{i=1}^{n}\tfrac{\zeta_i^2}{2}+\tfrac{\sigma_{max}}{2}||W||^4\\ &\quad+Z^TP\left(\begin{bmatrix}o_1(Z_1)\\o_2(Z_2)\\\vdots\\o_n(Z_n)\end{bmatrix}+\begin{bmatrix}1&&\\&1&\\&&\ddots&\\&&&1\end{bmatrix}\begin{bmatrix}x_{2d}^*\\\vdots\\x_{nd}^*\\u^*\end{bmatrix}\right)\\ &\leqslant -\alpha_1||Z||^2-\alpha_2||\widetilde{W}||^4+\Delta+Z^TP\left(\begin{bmatrix}o_1(Z_1)\\o_2(Z_2)\\\vdots\\o_n(Z_n)\end{bmatrix}+\begin{bmatrix}1&&\\&1&\\&&\ddots&\\&&&1\end{bmatrix}\begin{bmatrix}x_{2d}^*\\\vdots\\x_{nd}^*\\u^*\end{bmatrix}\right)\end{aligned} \tag{34}$$

where $Z=[\bar{z}_1,z_2,\ldots,z_n]^T;\widetilde{W}=\left[\widetilde{W}_1,\widetilde{W}_2,\ldots,\widetilde{W}\{_n\right]^T,\alpha_1=\min\{c_i-1\},\alpha_2=\min\{\frac{\sigma_i}{10}|1\leqslant i\leqslant n\},P=diag\{p,1,\ldots,1\},\Delta=\sum i=1^n\frac{\zeta_i^2}{2}+\frac{\sigma_{max}}{2}||W||^4,\sigma_{max}=\max\{\sigma_i\}$.

Remark 2. From (34), denote $\left[x_{2d}^*,\cdots,x_{nd}^*,u^*\right]^T=U^*$, due to the existence of U^*, the stability of the closed-loop system cannot be obtained in this stage. Therefore, U^* in (34) will be designed in the next stage by minimizing the cost function while guaranteeing the stability of the following system

$$\dot{Z}=P\left(\begin{bmatrix}o_1(Z_1)\\o_2(Z_2)\\\vdots\\o_n(Z_n)\end{bmatrix}+\begin{bmatrix}1&&\\&1&\\&&\ddots&\\&&&1\end{bmatrix}\begin{bmatrix}x_{2d}^*\\\vdots\\x_{nd}^*\\u^*\end{bmatrix}\right). \tag{35}$$

Then, the stability of the whole closed-loop system is thus obtained.

3.2. Optimal compensation term design

Consider the system in (34), and we can get

$$\dot{Z}=P(o(\mathscr{Z})+U^*) \tag{36}$$

where $o(\mathscr{Z})=[o_1(Z_1),o_2(Z_2),\ldots,o_n(Z_n)]^T,U^*=\left[x_{2d}^*,\ldots,x_{nd}^*,u^*\right]^T$.

The cost function of (36) is defined as

$$V(Z)=\int_0^\infty\left(Z^TQZ+U^{*T}RU^*\right)dt \tag{37}$$

where $Q=Q^T>0$ and $R=R^T>0$.

According to (37), the Hamilton function is expressed as

$$H(Z,U^*) = Z^TQZ + U^{*T}RU^* + \nabla_Z^T V(Z)P(o(\mathscr{Z}) + U^*) \tag{38}$$

where $\nabla_Z^T V(Z)$ is the partial derivative of $V(Z)$.

In order to obtain the optimal compensation term, we solve $\frac{\partial H}{\partial U} = 0$ and get

$$U^* = -\frac{1}{2}R^{-1}P^T\nabla_Z V^*(Z). \tag{39}$$

With the aid of the optimal control theory, we get the corresponding HJB equation as

$$Z^TQZ + U^{T*}RU^* + \nabla_Z^T V(Z)P(o(\mathscr{Z}) + U^*) = 0. \tag{40}$$

It can be noted that the optimal solution cannot be obtained directly. Thus, inspired by [44,45], the optimal solution is approximately solved by policy iteration (PI) method. As for the PI algorithm, the cost function $V^l(Z)$ and control policy $U^{(l+1)}$ are updated as the number of iteration l increases. The purpose of PI is to find the optimal compensation term U^*, that is, (37) is minimized under this control policy.

According to the PI algorithm, the RBF NN is used to approximate the cost function and its gradient as

$$V^l(Z) = W_c^{*T}S_c(Z) + \xi_c(Z) \tag{41}$$

$$\nabla V^l(Z) = W_c^{*T}\nabla_Z S_c(Z) + \nabla_Z\xi_c(Z) \tag{42}$$

where W_c^* and $\xi_c(Z)$ are ideal weight vector and estimation error, respectively; $\nabla_Z S_c(Z)$ and $\nabla_Z\xi_c(Z)$ are the gradients of the Gaussian basis function and the approximation error, respectively.

Substituting (42) into (38) and (39), we have

$$H(Z,W_c^*) = Z^TQZ + W_c^{*T}\nabla_Z S_c(Z)Po(\mathscr{Z}) - \tfrac{1}{4}W_c^{*T}\Psi W_c^* + e_{HJB} = 0 \tag{43}$$

$$U^* = -\frac{1}{2}R^{-1}P^T\left(W_c^{*T}\nabla_Z S_c(Z) + \nabla_Z\xi_c(Z)\right) \tag{44}$$

where $\Psi = \nabla_Z S_c(Z)PR^{-1}P^T\nabla_Z S_c^T(Z)$; e_{HJB} denotes the residual error caused by the RBF NN and $e_{HJB} = \nabla_Z\xi_c(Z)(o(\mathscr{Z}) + U^*) + \frac{1}{4}\nabla_Z\xi_c^T(Z)PR^{-1}P^T\nabla_Z\xi_c(Z)$.

Due to the ideal bounded weight W_c^* is unknown, we usually obtain the approximation value $\hat{V}^l(Z)$. (41) and (42) can be denoted as

$$\hat{V}^l(Z) = \widehat{W}_c^T S_c(Z) \tag{45}$$

$$\nabla\hat{V}^l(Z) = \widehat{W}_c^T\nabla_Z S_c(Z) \tag{46}$$

where $\widehat{W}_c \in \mathbb{R}^N$ denotes the estimation of ideal bounded weight vector. Let $\widetilde{W}_c = W_c^* - \widehat{W}_c$.

Now, recalling (43) and (44), the estimation values of the optimal compensation term and HJB function can be respectively described as

$$\hat{U}^{(l+1)} = -\frac{1}{2}R^{-1}P^T\nabla_Z^T S_c(Z)\widehat{W}_c \tag{47}$$

$$H\left(Z,\widehat{W}_c\right) = Z^TQZ + \widehat{W}_c^T\nabla_Z S_c(Z)P\hat{o}(\mathscr{Z}) - \frac{1}{4}\widehat{W}_c^T\Psi\widehat{W}_c = 0 \tag{48}$$

where $\hat{o}(\mathscr{Z}) = \left[\hat{o}_1\left(Z_1|\widehat{W}_1\right), \hat{o}_2\left(Z_2|\widehat{W}_2\right), \cdots, \hat{o}_n\left(Z_n|\widehat{W}_n\right)\right]^T$, $\hat{o}_i\left(Z_i|\widehat{W}_i\right) = \hat{f}_i\left(\bar{x}_i|\widehat{W}_i\right) - \hat{f}_i\left(\bar{x}_{id}|\widehat{W}_i\right), i = 1, \cdots, n$.

The tuning law for the weight $\widehat{W}_c$ is now provided as

$$\dot{\widehat{W}}_c = -\frac{\beta_c\hat{\rho}}{\left(\hat{\rho}^T\hat{\rho}+1\right)^2}\left(Z^TQZ + \widehat{W}_c^T\nabla_Z S_c(Z)P\hat{o}(\mathscr{Z}) - \tfrac{1}{4}\widehat{W}_c^T\Psi\widehat{W}_c\right) \tag{49}$$

where $\hat{\rho} = \nabla_{\mathscr{Z}} S_c(\mathscr{Z})P\left(\hat{o}(\mathscr{Z}) + \hat{U}^l\right)$, $\beta_c > 0$.

From (43), one has

$$Z^TQZ = -W_c^{*T}\nabla_Z S_c(Z)Po(\mathscr{Z}) + \frac{1}{4}W_c^{*T}\Psi W_c^* - e_{HJB}. \tag{50}$$

Applying $\dot{\widehat{W}}_c = -\dot{\widetilde{W}}_c$, one has

$$\begin{aligned}\dot{\widetilde{W}}_c &= -\tfrac{\beta_c}{q_1}\Big(\nabla_Z S_c(Z)\Big(\dot{Z}-P\tilde{o}(\mathscr{Z})+\tfrac{1}{2}PR^{-1}P^T\nabla_Z\xi_c(Z)+\tfrac{1}{2}\Psi\widetilde{W}_c\Big)\\ &\times\Big(\widetilde{W}_c^T\nabla_Z S_c(Z)\Big(\dot{Z}-P\tilde{o}(\mathscr{Z})+\tfrac{1}{2}PR^{-1}P^T\nabla\xi_c(Z)\Big)\\ &+\tfrac{1}{4}\widetilde{W}_c^T\Psi\widetilde{W}_c+e_{HJB}+W_c^T\nabla_Z S_c(Z)P\tilde{o}(\mathscr{Z})\Big)\end{aligned} \tag{51}$$

where $q_1=\frac{\hat{\rho}}{(\hat{\rho}^T\hat{\rho}+1)^2}$, $\tilde{o}(\mathscr{Z})=o(\mathscr{Z})-\hat{o}(\mathscr{Z})$.

4. Stability analysis

In this section, according to the Lyapunov stability theory, the stability of the whole closed-loop system is proved. The main results are given in Theorem 1.

Theorem 1. *For the system (1) under PP constraint condition(5), designing the adaptive controller (30), optimal compensation term (47) and the weight update laws (13), (25), (31) and (49) of the RBF NN control structure, the boundness of variables in the closed-loop system is guaranteed.*

Proof. We select the Lyapunov function with the following form

$$v_{HJB}=V_n+\frac{1}{2}\widetilde{W}_c^T\widetilde{W}_c. \tag{52}$$

Take the time derivative of v_{HJB}, one has

$$\dot{v}_{HJB}=\dot{V}_n+\widetilde{W}_c^T\dot{\widetilde{W}}_c. \tag{53}$$

Further, we have

$$\begin{aligned}\dot{v}_{HJB}\leqslant&-\alpha_1||Z||^2-\alpha_2||\widetilde{W}||^4+Z^TP(o(\mathscr{Z})+U^*)+\Delta\\ &-\tfrac{\beta_c}{q_1}\widetilde{W}_c^T\nabla_Z S_c(Z)\Big(\Big(\dot{Z}-P\tilde{o}(\mathscr{Z})+\tfrac{1}{2}PR^{-1}P^T\nabla\varepsilon_c(\mathscr{Z})\Big)+\tfrac{1}{2}\Psi\widetilde{W}_c\Big)\\ &\times\Big(\widetilde{W}_c^T\nabla_{\mathscr{Z}}\phi_c(\mathscr{Z})\Big(\dot{\mathscr{Z}}-P\tilde{o}(\mathscr{Z})+\tfrac{1}{2}PR^{-1}P^T\nabla\varepsilon_c(\mathscr{Z})\\ &+\tfrac{1}{4}\widetilde{W}_c^T\Psi\widetilde{W}_c+e_{HJB}+W_c^T\nabla_{\mathscr{Z}}\phi_c(\mathscr{Z})P\tilde{o}(\mathscr{Z})\Big).\end{aligned} \tag{54}$$

Assume that $\|\nabla_Z\xi_c(Z)\|\leqslant b_\xi,\|\nabla_Z S_c(Z)\|\leqslant\phi_M,\|e_{HJB}\|\leqslant\lambda_e,\|P(o(\mathscr{Z})+U^*)\|\leqslant I\sqrt{\mathscr{Z}},\|\nabla_Z\xi_c(Z)PR^{-1}P^T\nabla_Z\xi_c(Z)\|\geqslant\lambda_M$. Here, $b_\xi,\phi_M,\lambda_e,I,\lambda_M$ are positive parameters.

Applying the Young's inequality, take one term for example, we have

$$\begin{aligned}&-\tfrac{\beta_c}{q_1}\Big(\widetilde{W}_c^T\nabla_Z S_c(Z)P\tilde{o}(\mathscr{Z})\Big)\Big(\widetilde{W}_c^T\nabla_Z S_c(Z)P\tilde{o}(\mathscr{Z})\Big)\leqslant\\ &\tfrac{\beta_c}{q_1}\left[\tfrac{\pi_1||P||^4}{2}\left(8||\widetilde{W}||^4+\left(\sum_{i=1}^{n}4\varepsilon_i^2+4\varepsilon_i'^2\right)^2\right)+\tfrac{9}{2\pi_1}\phi_M^4||\widetilde{W}_c||^4\right]\end{aligned} \tag{55}$$

where $\pi_1>0,\varepsilon_i$ and ε_i' are upper bounds of $S_i(\bar{x}_i)$ and $S_i(\bar{x}_{id})$, respectively.

Other terms in (54) are solved as the same way. Then, (54) becomes

$$\dot{v}_{HJB}\leqslant\ -r_1||Z||^2+r_2||Z||-r_3||\widetilde{W}||^4-r_4||\widetilde{W}_c||^4+r_5||\widetilde{W}_c||^2+r_6 \tag{56}$$

where $r_1=\alpha_1-\Big(\frac{1}{2\pi_2}+\frac{1}{2\pi_3}+\frac{3}{16\pi_5}+\frac{1}{4\pi_8}+\frac{1}{4\pi_9}+\frac{1}{2\pi_{16}}\Big)\beta_c I^4-\frac{1}{2}$, $r_2=\frac{1}{2}I^2, r_3=\alpha_2-\beta_c||P||^4[4(\pi_1+\pi_2)+2(\pi_4+\pi_9+\pi_{11})+\frac{3\pi_6}{2}+\frac{2}{\pi_{10}}+\frac{2}{\pi_{11}}+\pi_{13}+\pi_{15}], r_4=\beta_c\Big[-\Big(\frac{9}{2\pi_1}+\frac{1}{2\pi_2}+\frac{1}{2\pi_3}+\frac{9\pi_2}{2}+\frac{9\pi_4}{4}+\frac{3}{16\pi_5}+\frac{27\pi_6}{16}+\frac{1}{4\pi_8}+\frac{1}{4\pi_9}+\frac{9}{4\pi_{10}}+\frac{9}{4\pi_{11}}+\frac{\pi_{16}}{2}\Big)\phi_M^4\Big], r_5=\beta_c\big[(\frac{\pi_3+1}{4}+\frac{1}{2\pi_4}+\frac{3}{16\pi_7}+\frac{1}{4\pi_{13}}+\frac{1}{4\pi_{12}})\lambda_M^2\big], r_6=\Delta+\beta_c\Big[\Big(\frac{\pi_1}{2}+\frac{\pi_2}{2}+\frac{\pi_4}{4}+\frac{3\pi_6}{16}+\frac{\pi_9}{4}+\frac{1}{4\pi_{10}}+\frac{\pi_{11}}{4}+\frac{\pi_{13}}{8}+\frac{\pi_{15}}{8}\Big)||P||^4\Big(\sum_{i=1}^{n}4\varepsilon_i^2+4\varepsilon_i'^2\Big)^2+\big(\frac{\pi_8}{2}+\frac{\pi_{10}}{2}+\frac{\pi_{12}}{2}+\frac{\pi_{14}}{2}\big)\lambda_e^2+\big(\frac{9\pi_9}{4}+\frac{9\pi_{11}}{4}+\frac{9\pi_{13}}{8}+\frac{9\pi_{15}}{8}\big)\phi_M^4||W_c||^4\big], \pi_i, i=1,\cdots,16$, are positive parameters. Noting that $r_1,r_3,r_4>0$ by selecting appropriate parameters.

If $||Z||>B_1$, or $||\widetilde{W}_c||>C_1$, or $||\widetilde{W}||>D_1$ holds, where $B_1=\frac{r_2+\sqrt{r_2^2+4r_1r_6}}{2r_1}, C_1=\sqrt{\frac{r_5+\sqrt{r_5^2+4r_4r_6}}{2r_4}}, D_1=\sqrt[4]{\frac{r_6}{r_3}}$, $\dot{v}_{HJB}$ is negative.

The proof is completed. □

5. Simulation

In this section, the effectiveness and advantages of the proposed control scheme are verified by the following two examples.

*A.Example*1 : The second-order investigated system is considered as

$$\begin{cases} \dot{x}_1 &= x_2 + f_1(x_1) \\ \dot{x}_2 &= u + f_2(\bar{x}_2) \\ y &= x_1 \end{cases}$$

where $f_1(x_1) = 0.5x_1^2, f_2(\bar{x}_2) = x_1^2\cos^2(x_2)$. The reference signal is $y_r = \sin(t)$. The design parameters are $c_1 = 10, c_2 = 50, \sigma_1 = 1, \sigma_2 = 1, \beta_c = 0.01, \varepsilon = 10^{-5}, Q = R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \rho_{min} = 0.6, \rho_{max} = 0.8, \eta_0 = 2.55, \eta_\infty = 0.05, k = 0.5$.

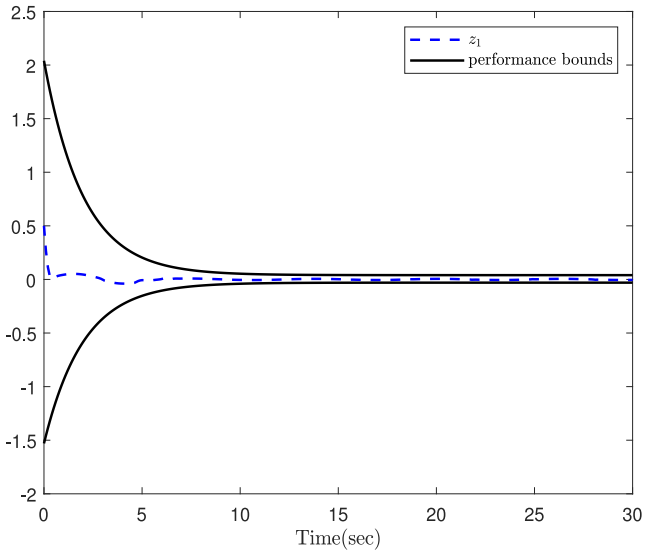

Fig. 1. Trajectory of z_1 and performance bounds of Example 1.

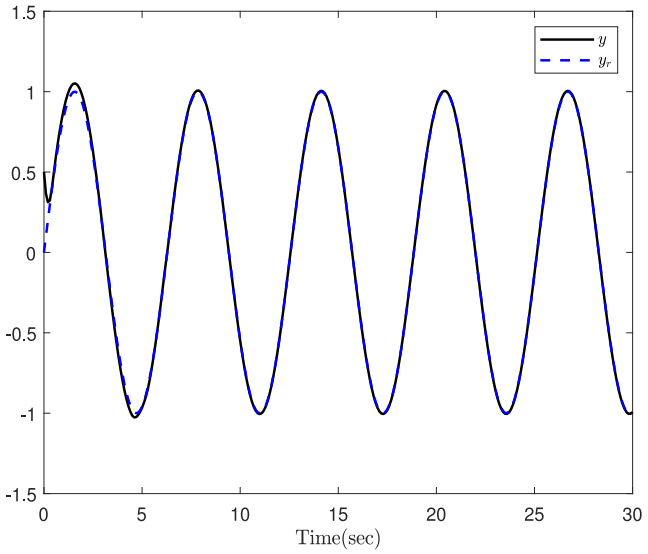

Fig. 2. Reference signal y_r and trajectory of y of Example 1.

The initial values are given as $x_0 = [0.5, 0.2]^T, \widehat{W}_1(0) = \widehat{W}_2(0) = 1, \widehat{W}_c(0) = 1.5$. The RBF NN is constructed with 9 neurons with the centers evenly spaced in $[-2 \times 2] \times [-2 \times 2] \times [-2 \times 2], \varsigma_i = 1.5$.

For Example 1, the tracking error curve and the performance bounds are shown in Fig.1, which indicate the tracking error $z_1(t)$ converges in the predefined range by the proposed method in this paper. Fig.2 shows that the y_r can be tracked by the y accurately. Fig.3 illustrates the boundness of NN weights and Hamilton function. Fig.4 depicts that the input is stable and bounded.

B.Example2 : In this example, the validity of the proposed control scheme is illustrated by the simulation of a manipulator system [46]. The dynamics equation of the manipulator system is described as

$$\begin{cases} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= f_2(\bar{x}_2) + u \\ y &= x_1 \end{cases} \tag{57}$$

with

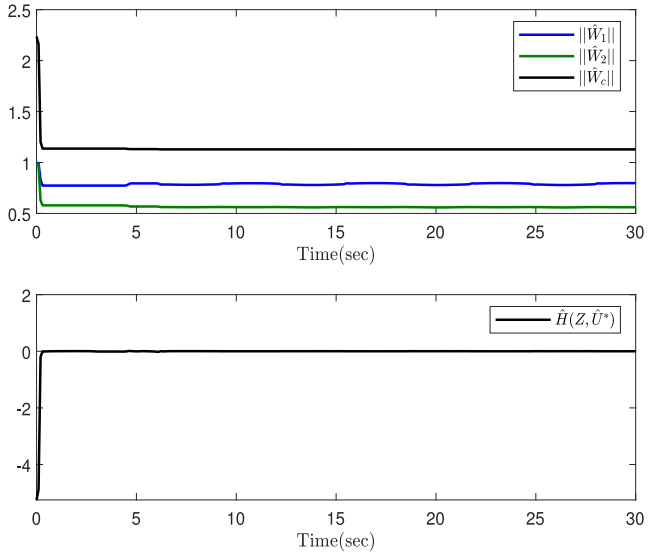

Fig. 3. The convergence process of the weights and Hamilton function of Example 1.

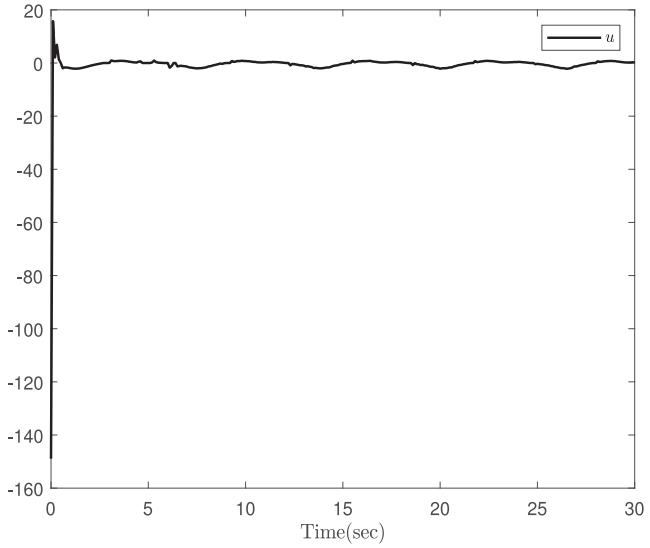

Fig. 4. The convergence process of input u of Example 1.

$$f_2(\bar{x}_2) = -\frac{Mgd}{J}\sin x_1 - \frac{D}{J}x_2$$

where x_1 and x_2 are the angular position and angular velocity of the link, respectively; M is the mass of the link and $M = 1kg; g = 9.8m/s^2, d$ is the distance between the center of mass and the rotation center of the link and $d = 1m; D$ is the friction coefficient of rotation of link and $D = 2N \cdot m \cdot s/rad, J$ is the rotational inertia and $J = 1kg \cdot m^2$.

In this paper, the RBF NN is constructed with 9 neurons with the centers evenly spaced in $[-2 \times 2] \times [-2 \times 2] \times [-2 \times 2], \varsigma_i = 2$. The initial values are given as $x_0 = [1.4, -0.2]^T, \widehat{W}_1(0) = \widehat{W}_2(0) = 1, \widehat{W}_c(0) = 0$. The desired signal is $y_r = sin(t)$. The design parameters are $c_1 = 10, c_2 = 50, \sigma_1 = 1, \sigma_2 = 1, \beta_c = 0.01, \varepsilon = 10^{-5}, Q = R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \rho_{min} = 0.6, \rho_{max} = 0.8, \eta_0 = 2.55, \eta_\infty = 0.05, k = 0.5$.

For Example 2, Fig.5 shows the tracking error curve and the performance bounds, which indicate that the tracking error $z_1(t)$ converges in the compact set of zero and satisfies the prescribed performance requirement. The tracking performance is

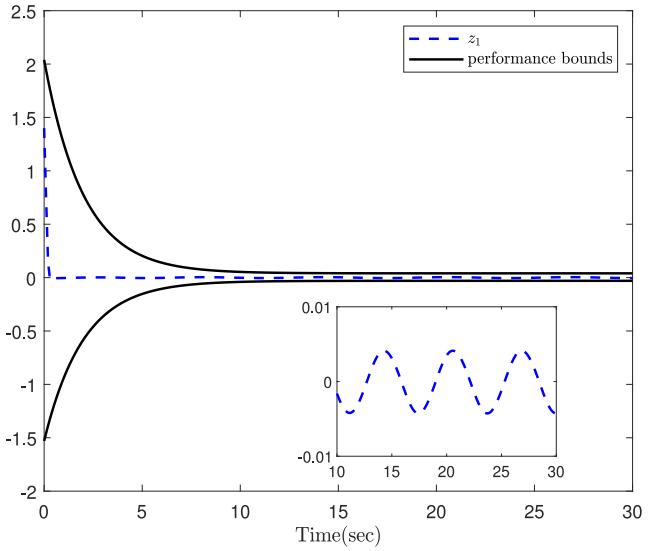

Fig. 5. Trajectory of z_1 and performance bounds of Example 2.

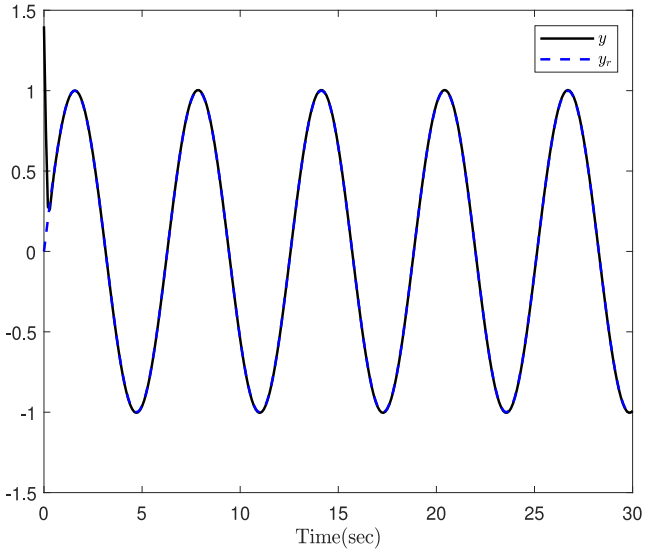

Fig. 6. Reference signal y_r and trajectory of y of Example 2.

shown in Fig.6, which indicates that the y_r can be tracked by the y precisely. Fig.7 illustrates the boundness of NN weights and Hamilton function. Fig.8 depicts the boundness of input.

The simulation results of Example 1 show that the proposed method can achieve the optimal tracking performance, and the boundness of all variables in the closed-loop system is ensured. In addition, the Example 2 shows that the proposed control scheme can be employed to an actual manipulator system, the advantage and practicability of the proposed control scheme are further verified.

C.ComparativeResults : Firstly, in the traditional adaptive tracking problem, the tracking performance can be achieved via an adaptive backstepping control scheme. In this paper, the optimization is considered by the proposed control scheme based on RL technique, which minimizes the cost function and achieves the tracking performance. Therefore, in order to verify the validity of introducing the optimal control, based on the PP technique, the comparison simulation is adopted between the RL-based optimal tracking control problem and the traditional adaptive tracking control problem.

Borrowed from [33], in the traditional adaptive tracking control problem combined with PP technique, the virtual controller x_{2d} and the actual controller u can be modified as

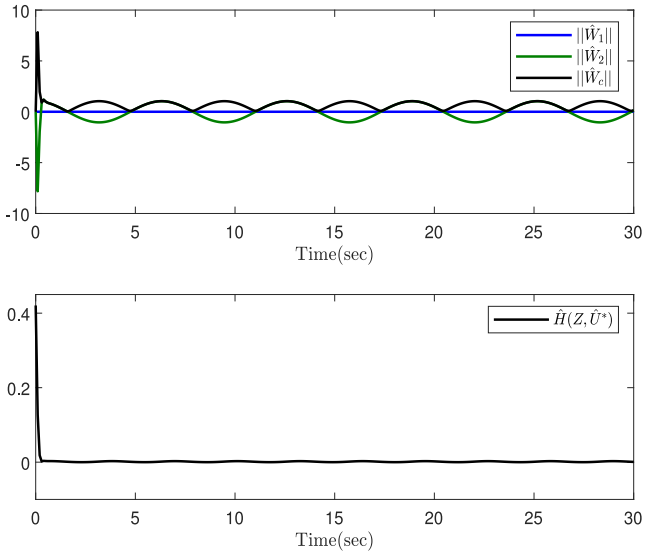

Fig. 7. The convergence process of the weights and Hamilton function of Example 2.

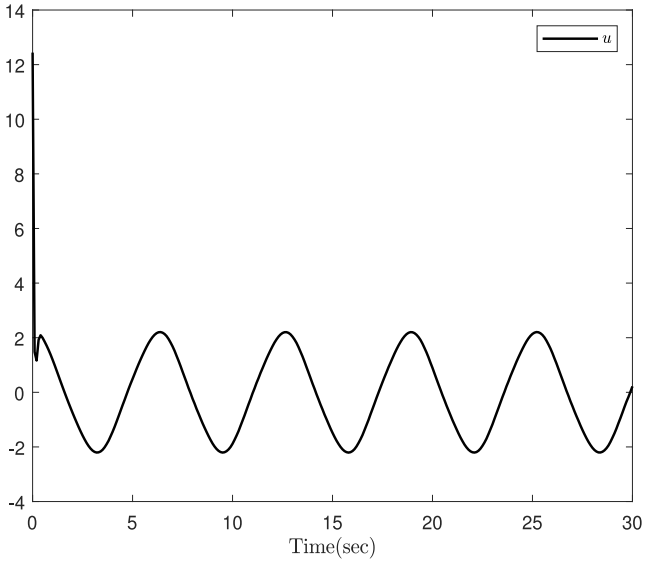

Fig. 8. The convergence process of input u of Example 2.

$$x_{2d} = -\frac{c_1\bar{z}_1}{p} - \frac{p\bar{z}_1}{2} - \widehat{W}_1^T S_1(x_1) + \dot{x}_{1d} + \frac{\dot{\eta}z_1}{\eta},\ u = -c_2 z_2 - z_1 - \frac{z_2}{2} - \widehat{W}_2^T S_2(\bar{x}_2) + \dot{x}_{2d}.$$

The comparison simulation results are shown in Figs.9 and 10.

Remark 3. For simplicity, the manipulator system (57) of Example 2 is utilized in the following comparison simulations.

Fig. 9 shows that the traditional tracking control with the aid of backstepping method can also track the reference signal. However, from Fig.10, the reference trajectory can be tracked by the output of the system optimally, which signifies that the cost function can be minimized while the tracking error can be limited in the prescribed domain by the PP optimal tracking control method based on RL technique in this article.

Next, for proving the superiority of introducing PP technique in the design of the controller, the comparative simulations are respectively carried out with and without applying the PP technique. The comparison simulation result is shown in Fig.11.

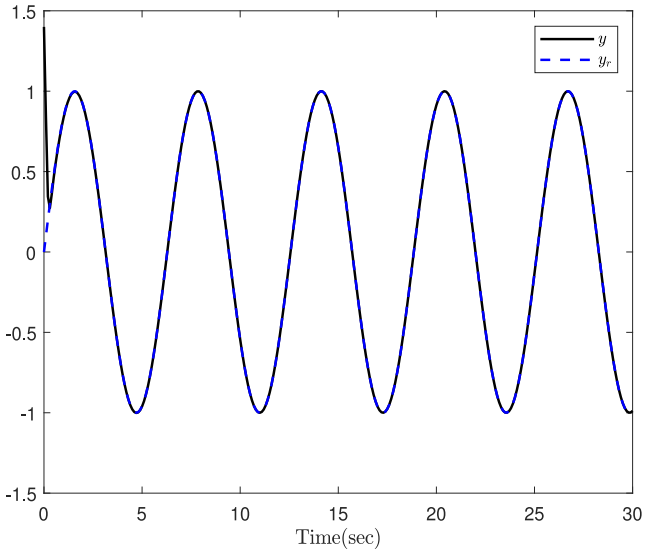

Fig. 9. Reference signal y_r and trajectory of y in the traditional tracking control.

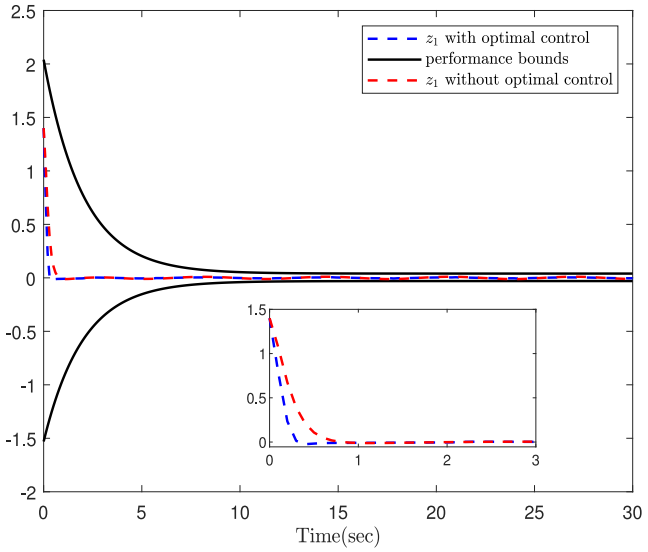

Fig. 10. The comparison results of z_1 with/without optimal control.

From Fig.11, it shows that the RL method combined with PP method has more superior tracking performance, which indicates that the PP is a valid technique to guarantee the transient and steady state tracking error performance in prescribed area.

Moreover, to illustrate that the convergence rate, tracking error and the maximum overshoot can be influenced by the parameters of the performance function in the controller design, the following case $\eta_0 = 1.35, \eta_\infty = 0.05, k = 0.5$ and $\eta_0 = 2.55, \eta_\infty = 0.05, k = 0.5$ are respectively considered. The comparison result of z_1 with different performance function parameters is presented in Fig.12.

Fig. 12 shows that the tracking performance can be achieved while ensuring the predefined transient state and steady state error bounds. Moreover, Fig.12 indicates that the PP method is an effective technique to ensure that the tracking error can converge to the predefined range.

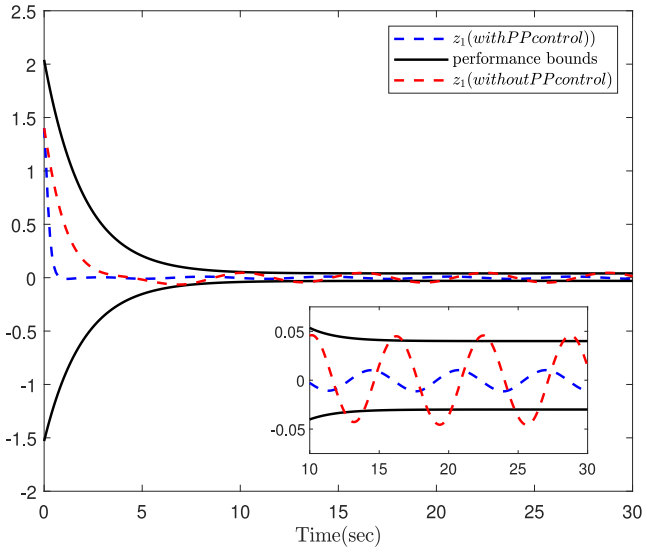

Fig. 11. The comparison results of z_1 with/without PP control.

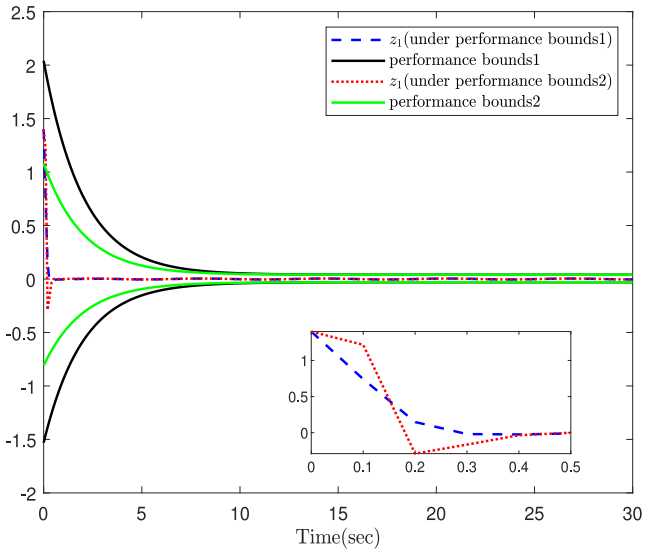

Fig. 12. The comparison results of z_1 with different performance function parameters$\eta_0 = 2.55, \eta_\infty = 0.05, k = 0.5$(performance bounds1)$\eta_0 = 1.35, \eta_\infty = 0.05, k = 0.5$(performance bounds2).

6. Conclusion

In this paper, based on RL technique, the optimal tracking control problem with PP has been considered for a class of strict-feedback nonlinear systems. The RBF NN has been utilized to approximate the unknown nonlinearities and cost function. The backstepping control method has been developed to generate the adaptive controller firstly. Subsequently, the optimal compensation term has been derived via policy iteration to minimize the cost function. Moreover, to restrict the tracking error within the predefined domain, the PP technique has been introduced. It has been proved that the boundness of the signals in the closed-loop system is guaranteed, and the effectiveness and advantages of this control scheme have been demonstrated by the simulation results. In the future, the authors will take the communication burden and the computational complexity into account in adaptive optimal tracking control for strict-feedback nonlinear systems.

CRediT authorship contribution statement

Zongsheng Huang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft. **Weiwei Bai:** Validation, Writing – review & editing, Funding acquisition. **Tieshan Li:** Writing – review & editing, Supervision, Funding acquisition. **Yue Long:** Writing – review & editing, Funding acquisition. **C.L. Philip Chen:** Project administration. **Hongjing Liang:** Writing – review & editing. **Hanqing Yang:** Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tieshan Li reports financial support was provided by National Natural Science Foundation of China. Weiwei Bai reports financial support was provided by National Natural Science Foundation of China. Yue Long reports financial support was provided by National Natural Science Foundation of China. Hanqing Yang reports financial support was provided by National Natural Science Foundation of China.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (under Grant Nos. 51939001, 61976033, 62273072, 52271360, 62203088); Natural Science Foundation of Sichuan Province (under Grant Nos. 2022NSFSC0891); the Open Project Program of Engineering Research Center of Human-Robot Intelligent Technologies and Systems, MOE of China (under Grant Nos. KP2021PY015).

References

[1] Y. Zhang, J. Gao, Y. Chen, C. Bian, F. Zhang, Q. Liang, Adaptive neural network control for visual docking of an autonomous underwater vehicle using command filtered backstepping, Int. J. Robust Nonlinear Control 32 (8) (2022) 4716–4738.
[2] G. Wen, W. Hao, W. Feng, K. Gao, Optimized backstepping tracking control using reinforcement learning for quadrotor unmanned aerial vehicle system, IEEE Trans. Syst. Man Cybern.: Syst. 52 (8) (2022) 5004–5015.
[3] S. Ling, H. Wang, P.X. Liu, Adaptive fuzzy tracking control of flexible-joint robots based on command filtering, IEEE Trans. Industr. Electron. 67 (5) (2020) 4046–4055.
[4] D. Xia, X. Yue, Y. Yin, Output-feedback asymptotic tracking control for rigid-body attitude via adaptive neural backstepping, ISA Trans. doi: 10.1016/j.isatra.2022.10.042.
[5] M. Krstić, I. Kanellakopoulos, P. Kokotović, Adaptive nonlinear control without overparametrization, Syst. Control Lett. 19 (3) (1992) 177–185.
[6] M. Krstic, P.V. Kokotovic, I. Kanellakopoulos, Nonlinear and adaptive control design, John Wiley & Sons Inc, 1995.
[7] M.M. Polycarpou, Stable adaptive neural control scheme for nonlinear systems, IEEE Trans. Autom. Control 41 (3) (1996) 447–451.
[8] C. Kwan, F. Lewis, Robust backstepping control of nonlinear systems using neural networks, IEEE Trans. Syst. Man Cybern. Part A 30 (6) (2000) 753–766.
[9] X. Zheng, X. Yang, Command filter and universal approximator based backstepping control design for strict-feedback nonlinear systems with uncertainty, IEEE Trans. Autom. Control 65 (3) (2020) 1310–1317.
[10] S.S. Ge, C. Wang, Direct adaptive NN control of a class of nonlinear systems, IEEE Trans. Neural Networks 13 (1) (2002) 214–221.
[11] S. Ge, C. Wang, Adaptive neural control of uncertain MIMO nonlinear systems, IEEE Trans. Neural Networks 15 (3) (2004) 674–692.
[12] P. Parsa, M.A. T, F. Baghbani, Command-filtered backstepping robust adaptive emotional control of strict-feedback nonlinear systems with mismatched uncertainties, Inf. Sci. 579 (2021) 434–453.
[13] N. Vafamand, M.M. Arefi, Robust neural network-based backstepping landing control of quadrotor on moving platform with stochastic noise, Int. J. Robust Nonlinear Control 32 (4) (2022) 2007–2026.
[14] D. Wang, D. Liu, H. Li, H. Ma, Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming, Inf. Sci. 282 (2014) 167–179.
[15] Y. Li, Y. Liu, S. Tong, Observer-based neuro-adaptive optimized control of strict-feedback nonlinear systems with state constraints, IEEE Trans. Neural Networks Learn. Syst. 33 (7) (2022) 3131–3145.
[16] L. Yuan, T. Li, S. Tong, Y. Xiao, Q. Shan, Broad learning system approximation-based adaptive optimal control for unknown discrete-time nonlinear systems, IEEE Trans. Syst. Man Cybern.: Syst. 52 (8) (2022) 5028–5038.
[17] G. Wen, B. Niu, Optimized tracking control based on reinforcement learning for a class of high-order unknown nonlinear dynamic systems, Inf. Sci. 606 (2022) 368–379.
[18] Y. Chen, J. Wen, X. Luan, F. Liu, Optimal control for semi-markov jump linear systems via TP-free temporal difference learning, Int. J. Robust Nonlinear Control 31 (14) (2021) 6905–6916.
[19] X. Wen, H. Shi, C. Su, X. Jiang, P. Li, J. Yu, Novel data-driven two-dimensional Q-learning for optimal tracking control of batch process with unknown dynamics, ISA Trans. 125 (2022) 10–21.

[20] Z. Chen, S. Chen, K. Chen, Y. Zhang, Constrained decoupling adaptive dynamic programming for a partially uncontrollable time-delayed model of energy systems, Inf. Sci. 608 (2022) 1352–1374.
[21] H. Ma, L. Xu, G. Yang, Multiple environment integral reinforcement learning-based fault-tolerant control for affine nonlinear systems, IEEE Trans. Cybern. 51 (4) (2019) 1913–1928.
[22] W. Bai, T. Li, Y. Long, C.L.P. Chen, Event-triggered multigradient recursive reinforcement learning tracking control for multiagent systems, IEEE Trans. Neural Networks Learn. Syst. doi:10.1109/TNNLS.2021.3094901.
[23] Z. Chen, W. Xue, N. Li, F.L. Lewis, Two-loop reinforcement learning algorithm for finite-horizon optimal control of continuous-time affine nonlinear systems, Int. J. Robust Nonlinear Control 32 (1) (2022) 393–420.
[24] T. Li, W. Bai, Q. Liu, Y. Long, C.P. Chen, Distributed fault-tolerant containment control protocols for the discrete-time multiagent systems via reinforcement learning method, IEEE Trans. Neural Networks Learn. Syst. doi:10.1109/TNNLS.2021.3121403.
[25] W. Xue, P. Kolaric, J. Fan, B. Lian, T. Chai, F.L. Lewis, Inverse reinforcement learning in tracking control based on inverse optimal control, IEEE Trans. Cybern. 52 (10) (2022) 10570–10581.
[26] S. He, H. Fang, M. Zhang, F. Liu, X. Luan, Z. Ding, Online policy iterative-based H_∞) optimization problems optimization algorithm for a class of nonlinear systems, Inf. Sci. 495 (2019) 1–13.
[27] Z. Yin, W. He, C. Yang, C. Sun, Control design of a marine vessel system using reinforcement learning, Neurocomputing 311 (2018) 353–362.
[28] S. Xue, B. Luo, D. Liu, Y. Gao, Event-triggered ADP for tracking control of partially unknown constrained uncertain systems, IEEE Trans. Cybern. 52 (9) (2022) 9001–9012.
[29] B. Yan, P. Shi, C. Lim, Z. Shi, Optimal robust formation control for heterogeneous multi-agent systems based on reinforcement learning, Int. J. Robust Nonlinear Control 32 (5) (2022) 2683–2704.
[30] Y. Deng, T. Liu, D. Zhao, Event-triggered output-feedback adaptive tracking control of autonomous underwater vehicles using reinforcement learning, Appl. Ocean Res. 113 (2021) 102676.
[31] X. Guo, W. Yan, R. Cui, Integral reinforcement learning-based adaptive NN control for continuous-time nonlinear MIMO systems with unknown control directions, IEEE Trans. Syst. Man Cybern.: Syst. 50 (11) (2020) 4068–4077.
[32] H. Zhang, H. Wang, B. Niu, L. Zhang, A.M. Ahmad, Sliding-mode surface-based adaptive actor-critic optimal control for switched nonlinear systems with average dwell time, Inf. Sci. 580 (2021) 756–774.
[33] C.P. Bechlioulis, G.A. Rovithakis, Adaptive control with guaranteed transient and steady state tracking error bounds for strict feedback systems, Automatica 45 (2) (2009) 532–538.
[34] C.P. Bechlioulis, G.A. Rovithakis, Prescribed performance adaptive control for multi-input multi-output affine in the control nonlinear systems, IEEE Trans. Autom. Control 55 (5) (2010) 1220–1226.
[35] H. Ma, Q. Zhou, H. Li, R. Lu, Adaptive prescribed performance control of a flexible-joint robotic manipulator with dynamic uncertainties, IEEE Trans. Cybern. doi:10.1109/TCYB.2021.3091531.
[36] W. Zhang, D. Xu, B. Jiang, T. Pan, Prescribed performance based model-free adaptive sliding mode constrained control for a class of nonlinear systems, Inf. Sci. 544 (2021) 97–116.
[37] S. Sui, C.L.P. Chen, S. Tong, Finite-time adaptive fuzzy prescribed performance control for high-order stochastic nonlinear systems, IEEE Trans. Fuzzy Syst. 30 (7) (2022) 2227–2240.
[38] G. Cui, W. Yang, J. Yu, Z. Li, C. Tao, Fixed-time prescribed performance adaptive trajectory tracking control for a QUAV, IEEE Trans. Circuits Syst. II Express Briefs 69 (2) (2022) 494–498.
[39] H. Zargarzadeh, T. Dierks, S. Jagannathan, Optimal control of nonlinear continuous-time systems in strict-feedback form, IEEE Trans. Neural Networks Learn. Syst. 26 (10) (2015) 2535–2549.
[40] S. Tong, K. Sun, S. Sui, Observer-based adaptive fuzzy decentralized optimal control design for strict-feedback nonlinear large-scale systems, IEEE Trans. Fuzzy Syst. 26 (2) (2018) 569–584.
[41] K. Sun, Y. Li, S. Tong, Fuzzy adaptive output feedback optimal control design for strict-feedback nonlinear systems, IEEE Trans. Syst. Man Cybern.: Syst. 47 (1) (2016) 33–44.
[42] X. Gao, W. Bai, T. Li, L. Yuan, Y. Long, Broad learning system-based adaptive optimal control design for dynamic positioning of marine vessels, Nonlinear Dyn. 105 (2) (2021) 1593–1609.
[43] L. Yan, Z. Liu, C.P. Chen, Y. Zhang, Z. Wu, Optimized adaptive consensus control for multi-agent systems with prescribed performance, Inf. Sci. 613 (2022) 649–666.
[44] J. Li, L. Ji, H. Li, Optimal consensus control for unknown second-order multi-agent systems: Using model-free reinforcement learning method, Appl. Math. Comput. 410 (2021) 126451.
[45] M. Mohammadi, M.M. Arefi, P. Setoodeh, O. Kaynak, Optimal tracking control based on reinforcement learning value iteration algorithm for time-delayed nonlinear systems with external disturbances and input constraints, Inf. Sci. 554 (2021) 84–98.
[46] Z. Guo, W. Bai, Q. Zhou, R. Lu, Adaptive optimal control for a class of nonlinear systems with dead zone input and prescribed performance, Acta Autom. Sin. 45 (11) (2019) 2128–2136.