Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Asymmetric time-varying integral barrier Lyapunov function based adaptive optimal control for nonlinear systems with dynamic state constraints*#

Yan WEI, Mingshuang HAO, Xinyi YU, Linlin OU‡

College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

E-mail: weiyanok@zjut.edu.cn; haoms@zjut.edu.cn; yuxy@zjut.edu.cn; linlinou@zjut.edu.cn

Received Oct. 6, 2023; Revision accepted Jan. 21, 2024; Crosschecked May 23, 2024

Abstract: This paper investigates the issue of adaptive optimal tracking control for nonlinear systems with dynamic state constraints. An asymmetric time-varying integral barrier Lyapunov function (ATIBLF) based integral reinforcement learning (IRL) control algorithm with an actor–critic structure is first proposed. The ATIBLF items are appropriately arranged in every step of the optimized backstepping control design to ensure that the dynamic full-state constraints are never violated. Thus, optimal virtual/actual control in every backstepping subsystem is decomposed with ATIBLF items and also with an adaptive optimized item. Meanwhile, neural networks are used to approximate the gradient value functions. According to the Lyapunov stability theorem, the boundedness of all signals of the closed-loop system is proved, and the proposed control scheme ensures that the system states are within predefined compact sets. Finally, the effectiveness of the proposed control approach is validated by simulations.

Key words: State constraints; Asymmetric time-varying integral barrier Lyapunov function (ATIBLF); Adaptive optimal control; Nonlinear systems

https://doi.org/10.1631/FITEE.2300675 **CLC number:** TP13

1 Introduction

With the demand for higher-performance control systems, nonlinear control methods have been studied widely, such as backstepping control (Chen et al., 2009), sliding-mode control (Li DY et al., 2021), and finite-time control (Li DY et al., 2022). Furthermore, significant meaningful research results related to neural network (NN)/fuzzy logic system (FLS) based adaptive control methods have been achieved. In particular, NNs/FLSs are usually combined with backstepping control technology to achieve better control performance (Li Y et al., 2004; Chen et al., 2009). Nevertheless, it is quite challenging for these adaptive control strategies to achieve the best performance specifications for nonlinear systems. How to achieve the control goal at a minimum cost has become an urgent problem to be solved.

From an energy standpoint, optimization aims to reduce energy expenditure. Optimal tracking control has attracted the attention of many scholars because of this optimization performance. For nonlinear systems, it is hard to obtain the analytical solution of the Hamilton–Jacobi–Bellman (HJB) equation directly, which makes it impossible to calculate an optimal controller. For this reason, considerable effort has been devoted to developing algorithms that approximately solve this equation, such as the

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 62203392 and 62373329), the Natural Science Foundation of Zhejiang Province, China (No. LY23F030009), and the Baima Lake Laboratory Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (No. LBMHD24F030002)

Electronic supplementary materials: The online version of this article (https://doi.org/10.1631/FITEE.2300675) contains supplementary materials, which are available to authorized users

ORCID: Yan WEI, https://orcid.org/0000-0002-9818-8034; Mingshuang HAO, https://orcid.org/0000-0002-9167-7388; Linlin OU, https://orcid.org/0000-0002-8589-9961

methods of policy iteration (Vamvoudakis and Lewis, 2010) and function approximation (Bhasin et al., 2013; Mohammadi et al., 2021). To avoid solving the HJB equation directly, NNs are commonly used to approximate its optimal solution. Nevertheless, the above adaptive optimal control approaches can be applied to only affine nonlinear systems. For the n^{th}-order strict-feedback systems, a simplified NN-based optimized backstepping control method was proposed to learn the virtual control signal in each subsystem (Wen et al., 2020, 2021). An adaptive optimized control scheme was developed by incorporating the actor–critic reinforcement learning (RL) architecture into the backstepping framework (Liu YC et al., 2022). Although the above adaptive optimized control structure is suitably simple, the constraint requirement has not been considered. For some practical systems, it is important to ensure the realization of control objectives under a safety guarantee.

In addition to optimization requirements, the practical systems are normally subject to state constraints. Over the last few years, the barrier Lyapunov function (BLF) has become an effective tool for solving state constraint problems (Tee and Ge, 2012). BLF-based approaches can be employed to address the problem of constraint with online real-time immediate response. Some BLF-based adaptive control methods have been investigated considering state/output constraints (Su and Wan, 2020; Mei et al., 2022; Wang et al., 2022a), where only constant constraints were considered. Up to now, many kinds of BLFs have been investigated to handle state/output constraints, such as log-type BLFs (Tee and Ge, 2012; Liu YJ et al., 2020), tangent-type BLFs (Jin and Xu, 2014; Liu L et al., 2022), and integral-type BLFs (IBLFs) (Tee and Ge, 2012; Wei Y et al., 2021). The prescribed performance control method also enables safe control by limiting the error to a specified range so that the desired tracking error is within a bounded region (Shen et al., 2022; Xu et al., 2023). However, when the error constraint is satisfied, it does not mean that the state constraint is also satisfied. Compared with the other two forms of BLFs, the IBLF-based control approaches can handle the problem of state constraints directly, and expand the feasible range of the initial value of the constrained states. To handle asymmetric state constraints, an asymmetric IBLF-based control approach was developed (Liu BJ et al., 2020). Time-varying IBLFs were proposed to ensure the time-varying output constraints (Liu L et al., 2021). State constraints may be asymmetric and time-varying, and can be changed with six cases dynamically (Zhao et al., 2020). How to handle more general dynamic state constraints using BLF-based control methods directly is a problem worth studying. A kind of asymmetric time-varying IBLF (AT-IBLF) was constructed to deal with dynamic constraints, which can be used for only stabilization cases (Zhang LL et al., 2023). It is worth noting that the problem of tracking control with dynamic constraints has not been studied, which is of general significance.

How to ensure the optimal control performance of nonlinear systems while preventing the states from violating constraints has attracted many scholars' research interest. A neuro-adaptive optimal state-constrained control method was studied (Li YM et al., 2022a). It is more challenging to ensure safety performance in safety-critical systems of autonomous vehicles using RL algorithms. Thus, optimal backstepping methods that include the BLF have been studied to deal with the safe RL control problems (Zhang YX et al., 2024a). A log-type BLF-based optimal backstepping control method was designed to ensure the safety performance of autonomous vehicles (Zhang YX et al., 2024b). A log-type BLF-based self-learning optimal tracking control scheme was introduced for an unmanned surface vehicle by Wang et al. (2022b). A tangent-type BLF-based optimal backstepping control method was designed to prevent the nonlinear systems from being affected by state constraints (Li YM et al., 2022b). However, the above methods can neither handle state constraints directly nor deal with more general dynamic constraints. It is important for nonlinear systems, like autonomous vehicles and robots (Luo et al., 2023) which are required to perform complex and variable tasks, to have safe control performance in complex obstacle scenarios, as this can improve their utility while preventing injuries to operators or damage to equipment. The IBLF-based optimal tracking control problem for two-link robotic systems has been studied in Wei Y et al. (2023), but the n^{th}-order systems were not considered by the methodology, which motivates us to carry out the present study.

In this paper, the adaptive optimal tracking control problem for nonlinear systems with dynamic state constraints is investigated. In the framework of actor–critic networks, the ATIBLF-based adaptive neural optimal backstepping control approach is proposed to optimize the overall system control while ensuring that the system state does not violate constraints. The main contributions are outlined as follows:

1. An ATIBLF-based adaptive optimal tracking control scheme for strict-feedback nonlinear systems is first proposed. In contrast with existing BLF-based studies (Li YM et al., 2022a; Zhang YX et al., 2024b), the proposed method can handle dynamic state constraints directly without error transformation, and the feasible region of the initial state value can be expanded.

2. Different from existing traditional ATIBLF-based studies (Liu L et al., 2021; Zhang LL et al., 2023), the ATIBLF items are appropriately arranged in every step of the optimized backstepping control design. The integral-type barrier optimal cost functions are first constructed for subsystems. Under the actor–critic framework, the ATIBLF-based optimal virtual and actual controllers are presented.

Notations: $\mathbb{R}$ indicates the set of all real numbers. $\mathbb{R}^i$ and $\mathbb{R}^n$ are the spaces of real i-vectors and n-vectors, respectively. $\Omega \in \mathbb{R}^n$ denotes a compact set that contains the origin. "sup" and "inf" are the least upper bound and greatest lower bound, respectively, of a set of real numbers. $\mathbb{N}_+$ is the set of all positive integers. $\min\limits_{1\le i\le n}\{\cdot\}$ represents the minimum value of any element in a set.

2 Problem formulation

2.1 System description

Consider the following nonlinear dynamics:

$$\begin{cases} \dot{x}_i = f_i(\bar{\boldsymbol{x}}_i) + g_i(\bar{\boldsymbol{x}}_i)x_{i+1}, \quad 1 \le i \le n-1, \\ \dot{x}_n = f_n(\bar{\boldsymbol{x}}_n) + g_n(\bar{\boldsymbol{x}}_n)u, \\ y = x_1, \end{cases} \tag{1}$$

where $\bar{\boldsymbol{x}}_i = [x_1, x_2, ..., x_i]^{\mathrm{T}} \in \mathbb{R}^i$ and $\bar{\boldsymbol{x}}_n \in \mathbb{R}^n$ are the system state vectors, and $u \in \mathbb{R}$ and $y \in \mathbb{R}$ denote the system input and output respectively. $f_i(\bar{\boldsymbol{x}}_i) : \mathbb{R}^i \to \mathbb{R}$ and $g_i(\bar{\boldsymbol{x}}_i) : \mathbb{R}^i \to \mathbb{R}$ $(i = 1, 2, ..., n)$ are smooth functions; at the same time, the former is bounded. Assume that $f_n(\bar{\boldsymbol{x}}_n) + g_n(\bar{\boldsymbol{x}}_n)u$ is Lipschitz continuous on the set Ω, which contains the origin so that for system input u and bounded initial state $x_i(0)$, system (1) has a unique solution. Moreover, to satisfy the condition that the whole system is asymptotically stable, it is required that there exists continuous input u and that system (1) is also stable on Ω.

The control objectives of this study are to design an adaptive optimal controller u under dynamic state constraints and to ensure that the following propositions hold:

Proposition 1 The system output y can track the desired reference trajectory y_{d} accurately while ensuring that the tracking error is uniformly ultimately bounded, and that all signals of the closed-loop system are bounded.

Proposition 2 The dynamic full-state constraints are never violated; i.e., the states satisfy $x_i \in \mathcal{R}_i$ $(\mathcal{R}_i = \{x_i \in \mathbb{R} : -k_{\mathrm{a}i}(t) < x_i < k_{\mathrm{b}i}(t)\},\ i = 1, 2, ..., n)$, where $k_{\mathrm{a}i}(t)$ and $k_{\mathrm{b}i}(t)$ are known positive time-varying differentiable constraints, and $-k_{\mathrm{a}i}(t)$ and $k_{\mathrm{b}i}(t)$ represent the lower and upper bounds of x_i, respectively. $k_{\mathrm{a}i}(t)$ and $k_{\mathrm{b}i}(t)$ are replaced by $k_{\mathrm{a}i}$ and $k_{\mathrm{b}i}$ in the later analysis.

Assumption 1 All state variables are available for the optimal controller design, and the initial states satisfy $x_i(0) \in \mathcal{R}_i$ $(i = 1, 2, ..., n)$.

Assumption 2 (Liu BJ et al., 2020) The desired trajectory vector $(y_{\mathrm{d}}, \dot{y}_{\mathrm{d}}, \ddot{y}_{\mathrm{d}})^{\mathrm{T}} \in \Omega_{\mathrm{d}}$ is available and continuous within a known compact set $\Omega_{\mathrm{d}} = \{y_{\mathrm{d}}^2 + \dot{y}_{\mathrm{d}}^2 + \ddot{y}_{\mathrm{d}}^2 \le \mathcal{A}\} \subset \mathbb{R}^3$, where $\mathcal{A} > 0$ indicates a known constant. Furthermore, there exist two unknown functions $k_{\mathrm{A}1}(t)$ and $k_{\mathrm{B}1}(t)$ satisfying $-k_{\mathrm{a}1}(t) < -k_{\mathrm{A}1}(t) < y_{\mathrm{d}} < k_{\mathrm{B}1}(t) < k_{\mathrm{b}1}(t)$, $\forall t \ge 0$.

Assumption 3 (Zhang LL et al., 2023) The smooth function $g_i(\bar{\boldsymbol{x}}_i)$ $(i = 1, 2, ..., n)$ is known, and there exist the known continuous function $\bar{g}_i(\bar{\boldsymbol{x}}_i)$ and unknown positive constants $\underline{g}_i$ and $\bar{g}_i$ such that $0 < \underline{g}_i \le g_i(\bar{\boldsymbol{x}}_i) \le \bar{g}_i(\bar{\boldsymbol{x}}_i) \le \bar{g}_i < \infty$ for all $t \ge 0$. For generality, it is assumed that $g_i(\bar{\boldsymbol{x}}_i)$ is positive for $-k_{\mathrm{a}i} < x_i < k_{\mathrm{b}i}$.

Assumption 4 The virtual signal α_{i-1} $(i = 2, 3, \ldots, n)$ is a continuously differentiable function. There exist two unknown functions $k_{\mathrm{l}i}(t)$ and $k_{\mathrm{u}i}(t)$ satisfying $-k_{\mathrm{a}i}(t) < -k_{\mathrm{l}i}(t) < \alpha_{i-1} < k_{\mathrm{u}i}(t) < k_{\mathrm{b}i}(t)$, $\forall t \ge 0$.

Remark 1 Notice that Assumption 1 is often used in the adaptive state-constrained controller design of

strict-feedback systems. Assumption 2 is a sufficient condition that makes the control method effective. Assumption 3 indicates that the control gain functions $g_i(\bar{\boldsymbol{x}}_i)$ are bounded. Assumption 4 denotes a feasibility condition that the virtual signals should be maintained.

2.2 Radial basis function (RBF) neural networks

Define a continuous function $f(\boldsymbol{x})$ on the compact set $\mathcal{D}$. Hence, there exists $f(\boldsymbol{x}) = (\boldsymbol{W}^*)^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}) + \varepsilon$, where $\boldsymbol{x} \in \mathbb{R}^n$ denotes the input vector, $n \in \mathbb{N}_+$, ε denotes the approximation error, and $\boldsymbol{W}^*$ is regarded as the value of $\boldsymbol{W}$ that minimizes $|\varepsilon|$ (i.e., the ideal weight vector). $\boldsymbol{W} \in \mathbb{R}^r$ denotes the weight vector, and $\boldsymbol{\varphi}_i(\boldsymbol{x})$ $(i = 1, 2, ..., r)$ denotes the kernel function satisfying $\boldsymbol{\varphi}_i(\boldsymbol{x}) = \exp(-(\boldsymbol{x} - \boldsymbol{c}_i)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{c}_i)/b_i^2)$ with r denoting the number of nodes, and $\boldsymbol{c}_i \in \mathbb{R}^n$ and b_i denoting the NNs' center vector and width, respectively. The following equation can thus be derived:

$$\boldsymbol{W}^* = \arg\min_{\boldsymbol{W}\in\mathbb{R}^r} \left\{\sup_{\boldsymbol{x}\in\mathcal{D}} |f(\boldsymbol{x}) - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x})|\right\}, \quad (2)$$

where for any constant $\bar{\varepsilon} > 0$ there is $|\varepsilon| < \bar{\varepsilon}$.

3 Adaptive optimal constrained controller design

A block diagram of the designed adaptive neural optimal control scheme is shown in Fig. 1. In each step, a learning algorithm is implemented via an actor–critic network architecture, which is used to construct adaptive laws. The ATIBLF items are appropriately arranged in every subsystem of the optimized backstepping control design to ensure that all states do not violate the dynamic constraints.

Before the controller design process, the following ATIBLF is introduced to handle dynamic state constraints:

$$V_i^z = \int_0^{z_i} \frac{\delta(k_{\mathrm{a}i} + k_{\mathrm{b}i})^2}{(k_{\mathrm{a}i} + \delta + \alpha_{i-1})(k_{\mathrm{b}i} - \delta - \alpha_{i-1})}\mathrm{d}\delta, \quad (3)$$

where $\alpha_0 = y_{\mathrm{d}}$, and α_{i-1} $(i = 2, 3, \ldots, n)$ is the virtual controller to be designed. V_i^z is positive-definite and continuously differentiable within constraints. Then, the following lemma about the ATIBLF can be obtained:

Lemma 1 (Liu BJ et al., 2020) By choosing the ATIBLF candidate V_i^z, the following inequality holds:

$$\frac{1}{2}z_i^2 \le V_i^z \le \Phi_i z_i^2, \quad (4)$$

where $\Phi_i = \dfrac{(k_{\mathrm{a}i} + k_{\mathrm{b}i})^2}{(k_{\mathrm{a}i} + z_i + \alpha_{i-1})(k_{\mathrm{b}i} - z_i - \alpha_{i-1})}$.

Remark 2 In the controller design process, the optimal controller and control law are designed by introducing error variables. The error variables are aimed at obtaining better trajectory-tracking performance rather than dealing with constraints. Whether the states violate constraints is not affected by the error variables. The constraint violation depends on the system states, not the tracking errors.

Step 1: The tracking error is defined as $z_1 = x_1 - y_{\mathrm{d}}$. Then, the time derivative of z_1 is

$$\dot{z}_1 = f_1(x_1) + g_1(x_1)x_2 - \dot{y}_{\mathrm{d}}. \quad (5)$$

Define the optimal performance index function

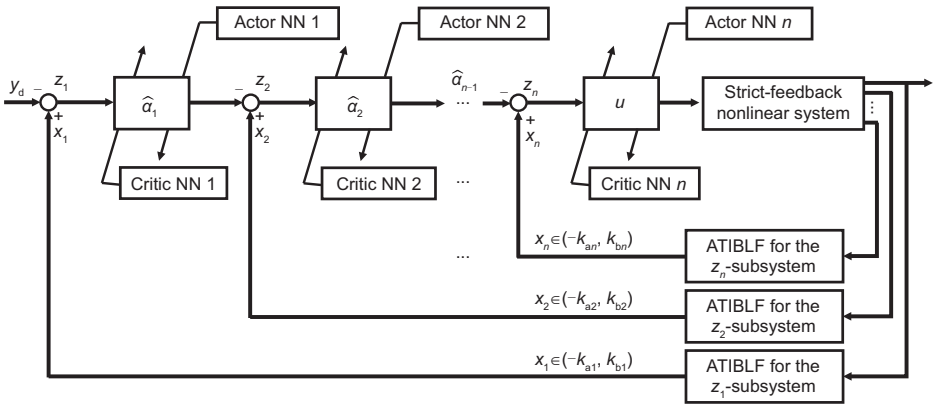

Fig. 1 Block diagram of the designed constrained control scheme (NN: neural network; ATIBLF: asymmetric time-varying integral barrier Lyapunov function)

for the z_1-subsystem as follows:

$$\begin{aligned} J_1^*(z_1) &= \min_{\alpha_1 \in \Psi(\Omega_{z_1})} \left(\int_t^\infty q_1(z_1(\tau), \alpha_1(z_1)) \mathrm{d}\tau \right) \\ &= \int_t^\infty q_1(z_1(\tau), \alpha_1^*(z_1)) \mathrm{d}\tau, \end{aligned} \tag{6}$$

where $q_1(z_1, \alpha_1^*) = \overline{q}_1 V_1^z + (\alpha_1^*(z_1))^2$ is the cost function, $\alpha_1^*(z_1)$ is the ideal virtual controller, $\overline{q}_1$ is a design parameter, Ω_{z_1} denotes a compact set containing the initial state, and $\Psi(\Omega_{z_1})$ is the admissible control set of α_1.

The function $J_1^*(z_1)$ can be split into two parts: the term $J_1^0(z_1)$ to be approximated and the other term which is already known. The term $\partial J_1^*(z_1)/\partial z_1$ is decomposed as

$$\begin{aligned} &\frac{\partial J_1^*(z_1)}{\partial z_1} \\ =&\frac{2}{g_1^2(x_1)} \Bigg[f_1(x_1) + \eta_1 \Phi_1 z_1 + k_1 z_1 + \frac{\dot{k}_{\mathrm{a}1}}{\Phi_1} \\ &\cdot \left(I_1 + \frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{a}1} + z_1 + y_{\mathrm{d}}} \right) - \frac{\dot{k}_{\mathrm{b}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{b}1} - z_1 - y_{\mathrm{d}}} \right. \\ &\left. - Q_1 \right) + \frac{\dot{y}_{\mathrm{d}}}{\Phi_1} \Psi_1 + \frac{\partial J_1^0(z_1)}{\partial z_1} \Bigg], \end{aligned} \tag{7}$$

where $k_1 > 0$ and $\eta_1 > 0$ are constants, $J_1^0(z_1)$ is a scalar-value continuous function, and the expressions of I_1, Q_1, and Ψ_1 are given in the supplementary materials.

For optimal controller design purposes, the following property is given:

Property 1 Functions I_1, Q_1, and Ψ_1 are well defined in the neighborhood of $z_1 = 0$.

The proof is given in the supplementary materials.

Based on Eqs. (5) and (7), the HJB equation of the z_1-subsystem is as follows:

$$\begin{aligned} &H_1 \left(z_1, \alpha_1^*, \frac{\partial J_1^*(z_1)}{\partial z_1} \right) \\ =&\overline{q}_1 V_1^z + (\alpha_1^*(z_1))^2 + \frac{2}{g_1^2(x_1)} \Bigg[\frac{\partial J_1^0(z_1)}{\partial z_1} + f_1(x_1) \\ &+ \eta_1 \Phi_1 z_1 + k_1 z_1 + \frac{\dot{k}_{\mathrm{a}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{a}1} + z_1 + y_{\mathrm{d}}} + I_1 \right) \\ &- \frac{\dot{k}_{\mathrm{b}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{b}1} - z_1 - y_{\mathrm{d}}} - Q_1 \right) + \frac{\dot{y}_{\mathrm{d}}}{\Phi_1} \Psi_1 \Bigg] \\ &\cdot (f_1(x_1) + g_1(x_1) \alpha_1^*(z_1) - \dot{y}_{\mathrm{d}}) \\ =&0. \end{aligned} \tag{8}$$

The ideal virtual controller α_1^* can be given by solving $\partial H_1 / \partial \alpha_1^* = 0$:

$$\begin{aligned} \alpha_1^* =& -\frac{1}{g_1(x_1)} \Bigg[\frac{\partial J_1^0(z_1)}{\partial z_1} + f_1(x_1) + \eta_1 \Phi_1 z_1 \\ &+ k_1 z_1 + \frac{\dot{k}_{\mathrm{a}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{a}1} + z_1 + y_{\mathrm{d}}} + I_1 \right) \\ &- \frac{\dot{k}_{\mathrm{b}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{b}1} - z_1 - y_{\mathrm{d}}} - Q_1 \right) + \frac{\dot{y}_{\mathrm{d}}}{\Phi_1} \Psi_1 \Bigg], \end{aligned} \tag{9}$$

where $\partial J_1^0(z_1)/\partial z_1$ is unknown and can be approximated by NNs as follows:

$$\frac{\partial J_1^0(z_1)}{\partial z_1} = (\boldsymbol{W}_1^*)^{\mathrm{T}} \boldsymbol{\varphi}_1(z_1) + \varepsilon_1(z_1), \tag{10}$$

where $\boldsymbol{W}_1^*$ denotes the ideal weight vector and $\boldsymbol{\varphi}_1(z_1)$ is the kernel function vector. $\varepsilon_1(z_1)$ indicates the approximation error, and $|\varepsilon_1(z_1)| \leq \bar{\zeta}_1$ holds with $\bar{\zeta}_1 > 0$ denoting a bounded constant.

Substituting Eq. (10) into Eq. (9), the gradient term $\partial J_1^*(z_1)/\partial z_1$ and α_1^* can be obtained as follows:

$$\begin{aligned} &\frac{\partial J_1^*(z_1)}{\partial z_1} \\ =&\frac{2}{g_1^2(x_1)} \Bigg[f_1(x_1) + \eta_1 \Phi_1 z_1 + k_1 z_1 + \frac{\dot{k}_{\mathrm{a}1}}{\Phi_1} \\ &\cdot \left(I_1 + \frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{a}1} + z_1 + y_{\mathrm{d}}} \right) - \frac{\dot{k}_{\mathrm{b}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{b}1} - z_1 - y_{\mathrm{d}}} \right. \\ &\left. - Q_1 \right) + \frac{\dot{y}_{\mathrm{d}}}{\Phi_1} \Psi_1 \Bigg] + \frac{2}{g_1^2(x_1)} \Big[(\boldsymbol{W}_1^*)^{\mathrm{T}} \boldsymbol{\varphi}_1(z_1) \\ &+ \varepsilon_1(z_1) \Big], \end{aligned} \tag{11}$$

$$\begin{aligned} \alpha_1^* =& -\frac{1}{g_1(x_1)} \Big[(\boldsymbol{W}_1^*)^{\mathrm{T}} \boldsymbol{\varphi}_1(z_1) + \varepsilon_1(z_1) + f_1(x_1) \\ &+ \eta_1 \Phi_1 z_1 + k_1 z_1 + \frac{\dot{k}_{\mathrm{a}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{a}1} + z_1 + y_{\mathrm{d}}} + I_1 \right) \\ &- \frac{\dot{k}_{\mathrm{b}1}}{\Phi_1} \left(\frac{k_{\mathrm{a}1} + k_{\mathrm{b}1}}{k_{\mathrm{b}1} - z_1 - y_{\mathrm{d}}} - Q_1 \right) + \frac{\dot{y}_{\mathrm{d}}}{\Phi_1} \Psi_1 \Big]. \end{aligned} \tag{12}$$

Note that $\boldsymbol{W}_1^*$ is unknown. Thus, it is approximated by the estimation vector $\hat{\boldsymbol{W}}_{\mathrm{c}1}$ in the critic NNs based on the RL technique, and approximated by $\hat{\boldsymbol{W}}_{\mathrm{a}1}$ in the actor NNs. The following equations can be obtained:

$$\frac{\partial \hat{J}_1^0(z_1)}{\partial z_1} = \hat{\boldsymbol{W}}_{\mathrm{c}1}^{\mathrm{T}} \boldsymbol{\varphi}_1(z_1), \tag{13}$$

$$\begin{aligned}\hat{\alpha}_1 =& -\frac{1}{g_1(x_1)}\Bigg[\hat{\boldsymbol{W}}_{\mathrm{a1}}^{\mathrm{T}}\boldsymbol{\varphi}_1(z_1)+f_1(x_1)+\eta_1\Phi_1 z_1\\&+k_1z_1+\frac{\dot{k}_{\mathrm{a1}}}{\Phi_1}\left(\frac{k_{\mathrm{a1}}+k_{\mathrm{b1}}}{k_{\mathrm{a1}}+z_1+y_{\mathrm{d}}}+I_1\right)-\frac{\dot{k}_{\mathrm{b1}}}{\Phi_1}\\&\cdot\left(\frac{k_{\mathrm{a1}}+k_{\mathrm{b1}}}{k_{\mathrm{b1}}-z_1-y_{\mathrm{d}}}-Q_1\right)+\frac{\dot{y}_{\mathrm{d}}}{\Phi_1}\Psi_1\Bigg].\end{aligned}\tag{14}$$

To reduce the complexity of variable terms in the design process, define

$$\begin{aligned}P_1 =&f_1(x_1)+\eta_1\Phi_1z_1+k_1z_1\\&+\frac{\dot{k}_{\mathrm{a1}}}{\Phi_1}\left(\frac{k_{\mathrm{a1}}+k_{\mathrm{b1}}}{k_{\mathrm{a1}}+z_1+y_{\mathrm{d}}}+I_1\right)\\&-\frac{\dot{k}_{\mathrm{b1}}}{\Phi_1}\left(\frac{k_{\mathrm{a1}}+k_{\mathrm{b1}}}{k_{\mathrm{b1}}-z_1-y_{\mathrm{d}}}-Q_1\right)+\frac{\dot{y}_{\mathrm{d}}}{\Phi_1}\Psi_1.\end{aligned}$$

The approximate HJB equation is given as

$$\begin{aligned}&H_1\left(z_1,\hat{\alpha}_1,\frac{\partial\hat{J}_1^0(z_1)}{\partial z_1}\right)\\=&\overline{q}_1V_1^z+\left[\frac{1}{g_1(x_1)}\left(\hat{\boldsymbol{W}}_{\mathrm{a1}}^{\mathrm{T}}\boldsymbol{\varphi}_1(z_1)+P_1\right)\right]^2\\&+\frac{2}{g_1^2(x_1)}\left[\hat{\boldsymbol{W}}_{\mathrm{c1}}^{\mathrm{T}}\boldsymbol{\varphi}_1(z_1)+P_1\right]\\&\cdot\left[f_1(x_1)-\hat{\boldsymbol{W}}_{\mathrm{a1}}^{\mathrm{T}}\boldsymbol{\varphi}_1(z_1)-P_1-\dot{y}_{\mathrm{d}}\right].\end{aligned}\tag{15}$$

The approximate error of Hamiltonian is defined as

$$\begin{aligned}E_1=&H_1\left(z_1,\hat{\alpha}_1,\frac{\partial\hat{J}_1^0(z_1)}{\partial z_1}\right)-H_1\left(z_1,\alpha_1^*,\frac{\partial J_1^*(z_1)}{\partial z_1}\right)\\=&H_1\left(z_1,\hat{\alpha}_1,\frac{\partial\hat{J}_1^0(z_1)}{\partial z_1}\right).\end{aligned}\tag{16}$$

To minimize E_1, its positive-definite function is defined as

$$\widetilde{E}_1(t)=\frac{1}{2(\boldsymbol{\omega}_1^{\mathrm{T}}\boldsymbol{\omega}_1+1)}E_1^2(t),\tag{17}$$

where $\boldsymbol{\omega}_1=\boldsymbol{\varphi}_1(z_1)\left[f_1(x_1)-\hat{\boldsymbol{W}}_{\mathrm{a1}}^{\mathrm{T}}\boldsymbol{\varphi}_1(z_1)-P_1-\dot{y}_{\mathrm{d}}\right]$.

The following weight-updating law of critic NNs is derived:

$$\begin{aligned}\dot{\hat{\boldsymbol{W}}}_{\mathrm{c1}}(t)=&-\xi_{\mathrm{c1}}\frac{\partial\widetilde{E}_1(t)}{\partial\hat{\boldsymbol{W}}_{\mathrm{c1}}}\\=&-\xi_{\mathrm{c1}}\frac{\boldsymbol{\omega}_1}{\boldsymbol{\omega}_1^{\mathrm{T}}\boldsymbol{\omega}_1+1}E_1(t)\\=&-\xi_{\mathrm{c1}}\frac{\boldsymbol{\omega}_1}{\boldsymbol{\omega}_1^{\mathrm{T}}\boldsymbol{\omega}_1+1}\Bigg[\overline{q}_1V_1^z+\frac{2}{g_1^2(x_1)}\boldsymbol{\omega}_1^{\mathrm{T}}\hat{\boldsymbol{W}}_{\mathrm{c1}}\\&-\frac{P_1^2}{g_1^2(x_1)}+\frac{1}{g_1^2(x_1)}\hat{\boldsymbol{W}}_{\mathrm{a1}}^{\mathrm{T}}\boldsymbol{\varphi}_1(z_1)\boldsymbol{\varphi}_1^{\mathrm{T}}(z_1)\\&\cdot\hat{\boldsymbol{W}}_{\mathrm{a1}}+\frac{2P_1}{g_1^2(x_1)}\left(f_1(x_1)-\dot{y}_{\mathrm{d}}\right)\Bigg],\end{aligned}\tag{18}$$

where $\xi_{\mathrm{c1}}>0$ is a design parameter.

Accordingly, the weight-updating law of actor NNs is designed as follows:

$$\begin{aligned}\dot{\hat{\boldsymbol{W}}}_{\mathrm{a1}}(t)=&\overline{q}_1\Phi_1z_1\boldsymbol{\varphi}_1(z_1)-\xi_{\mathrm{a1}}\boldsymbol{\varphi}_1(z_1)\boldsymbol{\varphi}_1^{\mathrm{T}}(z_1)\hat{\boldsymbol{W}}_{\mathrm{a1}}\\&+\frac{\xi_{\mathrm{c1}}}{g_1^2(x_1)(\boldsymbol{\omega}_1^{\mathrm{T}}\boldsymbol{\omega}_1+1)}\boldsymbol{\varphi}_1(z_1)\\&\cdot\boldsymbol{\varphi}_1^{\mathrm{T}}(z_1)\hat{\boldsymbol{W}}_{\mathrm{a1}}\boldsymbol{\omega}_1^{\mathrm{T}}\hat{\boldsymbol{W}}_{\mathrm{c1}},\end{aligned}\tag{19}$$

where $\xi_{\mathrm{a1}}>0$ is a parameter.

Step i ($2\le i\le n-1$): Define the tracking error as $z_i=x_i-\hat{\alpha}_{i-1}$, where $\hat{\alpha}_{i-1}$ is the virtual controller to be approximated. The error dynamics for the z_i-subsystem is

$$\dot{z}_i=f_i(x_i)+g_i(x_i)x_{i+1}-\dot{\hat{\alpha}}_{i-1}.\tag{20}$$

Define the optimal performance index function for the z_i-subsystem as follows:

$$\begin{aligned}J_i^*(z_i)=&\min_{\alpha_i\in\Psi(\Omega_{z_i})}\left(\int_t^\infty q_i\big(z_i(\tau),\alpha_i(z_i)\big)\mathrm{d}\tau\right)\\=&\int_t^\infty q_i\big(z_i(\tau),\alpha_i^*(z_i)\big)\mathrm{d}\tau,\end{aligned}\tag{21}$$

where $q_i(z_i,\alpha_i^*)=\overline{q}_iV_i^z+(\alpha_i^*(z_i))^2$ denotes the cost function, $\alpha_i^*(z_i)$ is the ideal optimal virtual controller, $\overline{q}_i>>0$ is a parameter, and Ω_{z_i} denotes a compact set containing the initial state.

The function $J_i^*(z_i)$ can be split into two parts: the term $J_i^0(z_i)$ to be approximated and the other term which is already known. Term $\partial J_i^*(z_i)/\partial z_i$ is decomposed as

$$\begin{aligned}&\frac{\partial J_i^*(z_i)}{\partial z_i}\\=&\frac{2}{g_i^2(x_i)}\Bigg[f_i(x_i)+\eta_i\Phi_iz_i+k_iz_i+\frac{\dot{k}_{\mathrm{a}i}}{\Phi_i}\Bigg(I_i\\&+\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{a}i}+z_i+\alpha_{i-1}}\Bigg)-\frac{\dot{k}_{\mathrm{b}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{b}i}-z_i-\alpha_{i-1}}-Q_i\right)\\&+\frac{\dot{\alpha}_{i-1}}{\Phi_i}\Psi_i+\frac{g_{i-1}(x_{i-1})\Phi_{i-1}}{\Phi_i}z_{i-1}+\frac{\partial J_i^0(z_i)}{\partial z_i}\Bigg],\end{aligned}\tag{22}$$

where $k_i>0$ and $\eta_i>0$ are constants, and $J_i^0(z_i)$ is a scalar-value continuous function. The expressions of I_i, Q_i, and Ψ_i are similar to those in step 1.

Based on Eqs. (20) and (22), the HJB equation of the z_i-subsystem is given as follows:

$$\begin{aligned}
&H_i\left(z_i,\alpha_i^*,\frac{\partial J_i^*(z_i)}{\partial z_i}\right)\\
=&\overline{q}_i V_i^z+(\alpha_i^*(z_i))^2+\frac{2}{g_i^2(x_i)}\Bigg[\frac{\partial J_i^0(z_i)}{\partial z_i}+f_i(x_i)\\
&+\eta_i\Phi_i z_i+k_i z_i+\frac{\dot{k}_{\mathrm{a}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{a}i}+z_i+\alpha_{i-1}}+I_i\right)\\
&-\frac{\dot{k}_{\mathrm{b}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{b}i}-z_i-\alpha_{i-1}}-Q_i\right)\\
&+\frac{\dot{\alpha}_{i-1}}{\Phi_i}\Psi_i+\frac{g_{i-1}(x_{i-1})}{\Phi_i}\\
&\cdot\Phi_{i-1}z_{i-1}\Bigg]\left(f_i(x_i)+g_i(x_i)\alpha_i^*(z_i)-\dot{\hat{\alpha}}_{i-1}\right)\\
=&0.
\end{aligned}\tag{23}$$

The optimal virtual controller α_i^* can be obtained:

$$\begin{aligned}
\alpha_i^*=&-\frac{1}{g_i(x_i)}\Bigg[\frac{\partial J_i^0(z_i)}{\partial z_i}+f_i(x_i)+\eta_i\Phi_i z_i+k_i z_i\\
&+\frac{\dot{k}_{\mathrm{a}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{a}i}+z_i+\alpha_{i-1}}+I_i\right)\\
&-\frac{\dot{k}_{\mathrm{b}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{b}i}-z_i-\alpha_{i-1}}-Q_i\right)\\
&+\frac{\dot{\alpha}_{i-1}}{\Phi_i}\Psi_i+\frac{g_{i-1}(x_{i-1})\Phi_{i-1}}{\Phi_i}z_{i-1}\Bigg],
\end{aligned}\tag{24}$$

where $\partial J_i^0(z_i)/\partial z_i$ is unknown and can be approximated by NNs as follows:

$$\frac{\partial J_i^0(z_i)}{\partial z_i}=(\boldsymbol{W}_i^*)^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)+\varepsilon_i(z_i),\tag{25}$$

where $\boldsymbol{W}_i^*$ represents the ideal weight vector and $\boldsymbol{\varphi}_i(z_i)$ denotes the kernel function vector. $\varepsilon_i(z_i)$ indicates the approximation error, and $|\varepsilon_i(z_i)|\leq\bar{\zeta}_i$ holds with $\bar{\zeta}_i>0$ denoting a bounded constant.

Substituting Eq. (25) into Eq. (24), the gradient term $\partial J_i^*(z_i)/\partial z_i$ and α_i^* can be obtained as follows:

$$\begin{aligned}
&\frac{\partial J_i^*(z_i)}{\partial z_i}\\
=&\frac{2}{g_i^2(x_i)}\Bigg[f_i(x_i)+\eta_i\Phi_i z_i+k_i z_i\\
&+\frac{\dot{k}_{\mathrm{a}i}}{\Phi_i}\left(I_i+\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{a}i}+z_i+\alpha_{i-1}}\right)\\
&-\frac{\dot{k}_{\mathrm{b}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{b}i}-z_i-\alpha_{i-1}}-Q_i\right)\\
&+\frac{\dot{\alpha}_{i-1}}{\Phi_i}\Psi_i+\frac{g_{i-1}(x_{i-1})\Phi_{i-1}}{\Phi_i}z_{i-1}\Bigg]\\
&+\frac{2}{g_i^2(x_i)}\Big[(\boldsymbol{W}_i^*)^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)+\varepsilon_i(z_i)\Big],
\end{aligned}\tag{26}$$

$$\begin{aligned}
\alpha_i^*=&-\frac{1}{g_i(x_i)}\Big[(\boldsymbol{W}_i^*)^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)+\varepsilon_i(z_i)+f_i(x_i)\\
&+\eta_i\Phi_i z_i+k_i z_i+\frac{\dot{k}_{\mathrm{a}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{a}i}+z_i+\alpha_{i-1}}+I_i\right)\\
&-\frac{\dot{k}_{\mathrm{b}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{b}i}-z_i-\alpha_{i-1}}-Q_i\right)+\frac{\dot{\alpha}_{i-1}}{\Phi_i}\Psi_i\\
&+\frac{g_{i-1}(x_{i-1})}{\Phi_i}\Phi_{i-1}z_{i-1}\Big].
\end{aligned}\tag{27}$$

Note that $\boldsymbol{W}_i^*$ is unknown. Likewise, $\boldsymbol{W}_i^*$ can be estimated by $\hat{\boldsymbol{W}}_{\mathrm{c}i}$ in the critic NNs and be estimated by $\hat{\boldsymbol{W}}_{\mathrm{a}i}$ in the actor NNs. The following equations can be given:

$$\frac{\partial\hat{J}_i^0(z_i)}{\partial z_i}=\hat{\boldsymbol{W}}_{\mathrm{c}i}^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i),\tag{28}$$

$$\begin{aligned}
\hat{\alpha}_i=&-\frac{1}{g_i(x_i)}\Big[\hat{\boldsymbol{W}}_{\mathrm{a}i}^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)+f_i(x_i)+\eta_i\Phi_i z_i\\
&+k_i z_i+\frac{\dot{k}_{\mathrm{a}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{a}i}+z_i+\alpha_{i-1}}+I_i\right)\\
&-\frac{\dot{k}_{\mathrm{b}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{b}i}-z_i-\alpha_{i-1}}-Q_i\right)+\frac{\dot{\alpha}_{i-1}}{\Phi_i}\Psi_i\\
&+\frac{g_{i-1}(x_{i-1})\Phi_{i-1}}{\Phi_i}z_{i-1}\Big].
\end{aligned}\tag{29}$$

Similar to step 1, define

$$\begin{aligned}
P_i=&f_i(x_i)+\eta_i\Phi_i z_i+k_i z_i\\
&+\frac{\dot{k}_{\mathrm{a}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{a}i}+z_i+\alpha_{i-1}}+I_i\right)\\
&-\frac{\dot{k}_{\mathrm{b}i}}{\Phi_i}\left(\frac{k_{\mathrm{a}i}+k_{\mathrm{b}i}}{k_{\mathrm{b}i}-z_i-\alpha_{i-1}}-Q_i\right)\\
&+\frac{\dot{\alpha}_{i-1}}{\Phi_i}\Psi_i+\frac{g_{i-1}(x_{i-1})\Phi_{i-1}}{\Phi_i}z_{i-1}.
\end{aligned}$$

Then, the approximate HJB equation can be simplified as

$$\begin{aligned}
&H_i\left(z_i,\hat{\alpha}_i,\frac{\partial\hat{J}_i^0(z_i)}{\partial z_i}\right)\\
=&\overline{q}_i V_i^z+\left[\frac{1}{g_i(x_i)}\left(\hat{\boldsymbol{W}}_{\mathrm{a}i}^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)+P_i\right)\right]^2
\end{aligned}$$

$$+\frac{2}{g_i^2(x_i)}\left[\hat{\boldsymbol{W}}_{ci}^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)+P_i\right]\left[f_i(x_i)-\hat{\boldsymbol{W}}_{ai}^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)-P_i-\dot{\hat{\alpha}}_{i-1}\right]. \tag{30}$$

The approximate error of Hamiltonian is defined as

$$\begin{aligned}E_i&=H_i\left(z_i,\hat{\alpha}_i,\frac{\partial\hat{J}_i^0(z_i)}{\partial z_i}\right)-H_i\left(z_i,\alpha_i^*,\frac{\partial J_i^*(z_i)}{\partial z_i}\right)\\&=H_i\left(z_i,\hat{\alpha}_i,\frac{\partial\hat{J}_i^0(z_i)}{\partial z_i}\right).\end{aligned} \tag{31}$$

To minimize E_i, its positive-definite function is defined as

$$\widetilde{E}_i(t)=\frac{1}{2(\boldsymbol{\omega}_i^{\mathrm{T}}\boldsymbol{\omega}_i+1)}E_i^2(t), \tag{32}$$

where $\boldsymbol{\omega}_i=\boldsymbol{\varphi}_i(z_i)\left[f_i(x_i)-\hat{\boldsymbol{W}}_{ai}^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)-P_i-\dot{\hat{\alpha}}_{i-1}\right]$.

Similarly, the following weight-updating law of critic NNs is derived as

$$\begin{aligned}\dot{\hat{\boldsymbol{W}}}_{ci}(t)=&-\xi_{ci}\frac{\partial\widetilde{E}_i(t)}{\partial\hat{\boldsymbol{W}}_{ci}}\\=&-\xi_{ci}\frac{\boldsymbol{\omega}_i}{\boldsymbol{\omega}_i^{\mathrm{T}}\boldsymbol{\omega}_i+1}E_i(t)\\=&-\xi_{ci}\frac{\boldsymbol{\omega}_i}{\boldsymbol{\omega}_i^{\mathrm{T}}\boldsymbol{\omega}_i+1}\left[\overline{q}_iV_i^z+\frac{2}{g_i^2(x_i)}\boldsymbol{\omega}_i^{\mathrm{T}}\hat{\boldsymbol{W}}_{ci}\right.\\&-\frac{P_i^2}{g_i^2(x_i)}+\frac{1}{g_i^2(x_i)}\hat{\boldsymbol{W}}_{ai}^{\mathrm{T}}\boldsymbol{\varphi}_i(z_i)\boldsymbol{\varphi}_i^{\mathrm{T}}(z_i)\\&\left.\cdot\hat{\boldsymbol{W}}_{ai}+\frac{2P_i}{g_i^2(x_i)}\left(f_i(x_i)-\dot{\hat{\alpha}}_{i-1}\right)\right],\end{aligned} \tag{33}$$

where $\xi_{ci}>0$ is a parameter.

Accordingly, the weight-updating law of actor NNs is designed as follows:

$$\begin{aligned}\dot{\hat{\boldsymbol{W}}}_{ai}(t)=&\overline{q}_i\Phi_iz_i\boldsymbol{\varphi}_i(z_i)-\xi_{ai}\boldsymbol{\varphi}_i(z_i)\boldsymbol{\varphi}_i^{\mathrm{T}}(z_i)\hat{\boldsymbol{W}}_{ai}\\&+\frac{\xi_{ci}}{g_i^2(x_i)(\boldsymbol{\omega}_i^{\mathrm{T}}\boldsymbol{\omega}_i+1)}\boldsymbol{\varphi}_i(z_i)\boldsymbol{\varphi}_i^{\mathrm{T}}(z_i)\\&\cdot\hat{\boldsymbol{W}}_{ai}\boldsymbol{\omega}_i^{\mathrm{T}}\hat{\boldsymbol{W}}_{ci},\end{aligned} \tag{34}$$

where $\xi_{ai}>0$ is a constant.

Step n: Define the tracking error as $z_n=x_n-\hat{\alpha}_{n-1}$. The error dynamics for the z_n-subsystem is

$$\dot{z}_n=f_n(x_n)+g_n(x_n)u-\dot{\hat{\alpha}}_{n-1}. \tag{35}$$

Define the optimal performance index function for the z_n-subsystem as follows:

$$\begin{aligned}J_n^*(z_n)=&\min_{u\in\Psi(\Omega_{z_n})}\left(\int_t^\infty q_n\left(z_n(\tau),u(z_n)\right)\mathrm{d}\tau\right)\\=&\int_t^\infty q_n\left(z_n(\tau),u^*(z_n)\right)\mathrm{d}\tau,\end{aligned} \tag{36}$$

where $q_n(z_n,u^*)=\overline{q}_nV_n^z+(u^*)^2$ denotes the cost function, u^* is the ideal optimal actual controller, $\overline{q}_n$ is a design parameter, and Ω_{z_n} denotes a compact set containing the initial state.

The function $J_n^*(z_n)$ can be split into two parts: the term $J_n^0(z_n)$ to be approximated and the other term which is already known. Term $\partial J_n^*(z_n)/\partial z_n$ is decomposed as

$$\begin{aligned}&\frac{\partial J_n^*(z_n)}{\partial z_n}\\=&\frac{2}{g_n^2(x_n)}\left[f_n(x_n)+\eta_n\Phi_nz_n+k_nz_n\right.\\&+\frac{\dot{k}_{an}}{\Phi_n}\left(I_n+\frac{k_{an}+k_{bn}}{k_{an}+z_n+\alpha_{n-1}}\right)\\&-\frac{\dot{k}_{bn}}{\Phi_n}\left(\frac{k_{an}+k_{bn}}{k_{bn}-z_n-\alpha_{n-1}}-Q_n\right)+\frac{\dot{\alpha}_{n-1}}{\Phi_n}\Psi_n\\&\left.+\frac{g_{n-1}(x_{n-1})\Phi_{n-1}}{\Phi_n}z_{n-1}+\frac{\partial J_n^0(z_n)}{\partial z_n}\right],\end{aligned} \tag{37}$$

where $k_n>0$ and $\eta_n>0$ are constants, and $J_n^0(z_n)$ is a scalar-value continuous function. The expressions of I_n, Q_n, and Ψ_n are similar to those in step 1.

The HJB equation of the z_n-subsystem is given as follows:

$$\begin{aligned}&H_n\left(z_n,u^*,\frac{\partial J_n^*(z_n)}{\partial z_n}\right)\\=&\overline{q}_nV_n^z+(u^*)^2+\frac{2}{g_n^2(x_n)}\left[\frac{\partial J_n^0(z_n)}{\partial z_n}+f_n(x_n)\right.\\&+\eta_n\Phi_nz_n+k_nz_n+\frac{\dot{k}_{an}}{\Phi_n}\left(\frac{k_{an}+k_{bn}}{k_{an}+z_n+\alpha_{n-1}}\right.\\&\left.+I_n\right)-\frac{\dot{k}_{bn}}{\Phi_n}\left(\frac{k_{an}+k_{bn}}{k_{bn}-z_n-\alpha_{n-1}}-Q_n\right)\\&\left.+\frac{\dot{\alpha}_{n-1}}{\Phi_n}\Psi_n+\frac{g_{n-1}(x_{n-1})}{\Phi_n}\Phi_{n-1}z_{n-1}\right]\\&\cdot\left(f_n(x_n)+g_n(x_n)u^*-\dot{\hat{\alpha}}_{n-1}\right)\\=&0.\end{aligned} \tag{38}$$

The ideal optimal controller u^* can be given by

solving $\partial H_n/\partial u^*=0$:

$$\begin{aligned}u^*=&-\frac{1}{g_n(x_n)}\Bigg[\frac{\partial J_n^0(z_n)}{\partial z_n}+f_n(x_n)+\eta_n\Phi_n z_n\\&+k_n z_n+\frac{\dot{k}_{\mathrm{an}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{an}}+z_n+\alpha_{n-1}}+I_n\right)\\&-\frac{\dot{k}_{\mathrm{bn}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{bn}}-z_n-\alpha_{n-1}}-Q_n\right)+\frac{\dot{\alpha}_{n-1}}{\Phi_n}\Psi_n\\&+\frac{g_{n-1}(x_{n-1})\Phi_{n-1}}{\Phi_n}z_{n-1}\Bigg],\end{aligned}\tag{39}$$

where $\partial J_n^0(z_n)/\partial z_n$ is unknown and can be approximated by NNs as follows:

$$\frac{\partial J_n^0(z_n)}{\partial z_n}=(\boldsymbol{W}_n^*)^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)+\varepsilon_n(z_n),\tag{40}$$

where $\boldsymbol{W}_n^*$ represents the ideal weight vector and $\boldsymbol{\varphi}_n(z_n)$ denotes the kernel function vector. $\varepsilon_n(z_n)$ indicates the approximation error, and $|\varepsilon_n(z_n)|\le\bar{\zeta}_n$ holds with $\bar{\zeta}_n>0$ denoting a bounded constant.

Substituting Eq. (40) into Eq. (39), the gradient term $\partial J_n^*(z_n)/\partial z_n$ and u^* can be obtained as follows:

$$\begin{aligned}&\frac{\partial J_n^*(z_n)}{\partial z_n}\\=&\frac{2}{g_n^2(x_n)}\Bigg[f_n(x_n)+\eta_n\Phi_n z_n+k_n z_n\\&+\frac{\dot{k}_{\mathrm{an}}}{\Phi_n}\left(I_n+\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{an}}+z_n+\alpha_{n-1}}\right)\\&-\frac{\dot{k}_{\mathrm{bn}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{bn}}-z_n-\alpha_{n-1}}-Q_n\right)+\frac{\dot{\alpha}_{n-1}}{\Phi_n}\Psi_n\\&+\frac{g_{n-1}(x_{n-1})\Phi_{n-1}}{\Phi_n}z_{n-1}\Bigg]+\frac{2}{g_n^2(x_n)}\\&\cdot\Big[(\boldsymbol{W}_n^*)^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)+\varepsilon_n(z_n)\Big],\end{aligned}\tag{41}$$

$$\begin{aligned}u^*=&-\frac{1}{g_n(x_n)}\Bigg[(\boldsymbol{W}_n^*)^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)+\varepsilon_n(z_n)+f_n(x_n)\\&+\eta_n\Phi_n z_n+k_n z_n+\frac{\dot{k}_{\mathrm{an}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{an}}+z_n+\alpha_{n-1}}\right.\\&\left.+I_n\right)-\frac{\dot{k}_{\mathrm{bn}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{bn}}-z_n-\alpha_{n-1}}-Q_n\right)\\&+\frac{\dot{\alpha}_{n-1}}{\Phi_n}\Psi_n+\frac{g_{n-1}(x_{n-1})\Phi_{n-1}}{\Phi_n}z_{n-1}\Bigg].\end{aligned}\tag{42}$$

Note that $\boldsymbol{W}_n^*$ is unknown. $\boldsymbol{W}_n^*$ can be estimated by $\hat{\boldsymbol{W}}_{cn}$ in the critic NNs and be estimated by $\hat{\boldsymbol{W}}_{an}$ in the actor NNs. The following equations can thus be given:

$$\frac{\partial\hat{J}_n^0(z_n)}{\partial z_n}=\hat{\boldsymbol{W}}_{\mathrm{cn}}^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n),\tag{43}$$

$$\begin{aligned}u=&-\frac{1}{g_n(x_n)}\Bigg[\hat{\boldsymbol{W}}_{\mathrm{an}}^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)+f_n(x_n)+\eta_n\Phi_n z_n\\&+k_n z_n+\frac{\dot{k}_{\mathrm{an}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{an}}+z_n+\alpha_{n-1}}+I_n\right)\\&-\frac{\dot{k}_{\mathrm{bn}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{bn}}-z_n-\alpha_{n-1}}-Q_n\right)+\frac{\dot{\alpha}_{n-1}}{\Phi_n}\Psi_n\\&+\frac{g_{n-1}(x_{n-1})\Phi_{n-1}}{\Phi_n}z_{n-1}\Bigg].\end{aligned}\tag{44}$$

Define

$$\begin{aligned}P_n=&f_n(x_n)+\eta_n\Phi_n z_n+k_n z_n\\&+\frac{\dot{k}_{\mathrm{an}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{an}}+z_n+\alpha_{n-1}}+I_n\right)\\&-\frac{\dot{k}_{\mathrm{bn}}}{\Phi_n}\left(\frac{k_{\mathrm{an}}+k_{\mathrm{bn}}}{k_{\mathrm{bn}}-z_n-\alpha_{n-1}}-Q_n\right)\\&+\frac{\dot{\alpha}_{n-1}}{\Phi_n}\Psi_n+\frac{g_{n-1}(x_{n-1})\Phi_{n-1}}{\Phi_n}z_{n-1}.\end{aligned}$$

The approximate HJB equation can be simplified as

$$\begin{aligned}&H_n\left(z_n,u,\frac{\partial\hat{J}_n^0(z_n)}{\partial z_n}\right)\\=&\overline{q}_n V_n^z+\left[\frac{1}{g_n(x_n)}\left(\hat{\boldsymbol{W}}_{\mathrm{an}}^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)+P_n\right)\right]^2\\&+\frac{2}{g_n^2(x_n)}\left[\hat{\boldsymbol{W}}_{\mathrm{cn}}^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)+P_n\right]\Big[f_n(x_n)\\&-\hat{\boldsymbol{W}}_{\mathrm{an}}^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)-P_n-\dot{\hat{\alpha}}_{n-1}\Big].\end{aligned}\tag{45}$$

The approximate error of Hamiltonian is defined as

$$\begin{aligned}E_n=&H_n\left(z_n,u,\frac{\partial\hat{J}_n^0(z_n)}{\partial z_n}\right)-H_n\left(z_n,u^*,\frac{\partial J_n^*(z_n)}{\partial z_n}\right)\\=&H_n\left(z_n,u,\frac{\partial\hat{J}_n^0(z_n)}{\partial z_n}\right).\end{aligned}\tag{46}$$

To minimize E_n, its positive-definite function is defined as

$$\widetilde{E}_n(t)=\frac{1}{2(\boldsymbol{\omega}_n^{\mathrm{T}}\boldsymbol{\omega}_n+1)}E_n^2(t),\tag{47}$$

where

$$\boldsymbol{\omega}_n=\boldsymbol{\varphi}_n(z_n)\Big[f_n(x_n)-\hat{\boldsymbol{W}}_{\mathrm{an}}^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)-P_n-\dot{\hat{\alpha}}_{n-1}\Big].$$

Similarly, the following weight-updating law of critic NNs is derived:

$$\begin{aligned}\dot{\hat{\boldsymbol{W}}}_{cn}(t) =& -\xi_{cn}\frac{\partial \widetilde{E}_n(t)}{\partial \hat{\boldsymbol{W}}_{cn}}\\ =& -\xi_{cn}\frac{\boldsymbol{\omega}_n}{\boldsymbol{\omega}_n^{\mathrm{T}}\boldsymbol{\omega}_n+1}E_n(t)\\ =& -\xi_{cn}\frac{\boldsymbol{\omega}_n}{\boldsymbol{\omega}_n^{\mathrm{T}}\boldsymbol{\omega}_n+1}\Bigg[\overline{q}_n V_n^z+\frac{2}{g_n^2(x_n)}\boldsymbol{\omega}_n^{\mathrm{T}}\hat{\boldsymbol{W}}_{cn}\\ &-\frac{P_n^2}{g_n^2(x_n)}+\frac{1}{g_n^2(x_n)}\hat{\boldsymbol{W}}_{an}^{\mathrm{T}}\boldsymbol{\varphi}_n(z_n)\boldsymbol{\varphi}_n^{\mathrm{T}}(z_n)\\ &\cdot\hat{\boldsymbol{W}}_{an}+\frac{2P_n}{g_n^2(x_n)}\left(f_n(x_n)-\dot{\hat{\alpha}}_{n-1}\right)\Bigg],\end{aligned}\tag{48}$$

where $\xi_{cn}>0$ is the learning rate.

Accordingly, the weight-updating law of actor NNs is obtained as follows:

$$\begin{aligned}\dot{\hat{\boldsymbol{W}}}_{an}(t) =&\overline{q}_n\Phi_n z_n\boldsymbol{\varphi}_n(z_n)-\xi_{an}\boldsymbol{\varphi}_n(z_n)\boldsymbol{\varphi}_n^{\mathrm{T}}(z_n)\\ &\cdot\hat{\boldsymbol{W}}_{an}+\frac{\xi_{cn}}{g_n^2(x_n)(\boldsymbol{\omega}_n^{\mathrm{T}}\boldsymbol{\omega}_n+1)}\boldsymbol{\varphi}_n(z_n)\\ &\cdot\boldsymbol{\varphi}_n^{\mathrm{T}}(z_n)\hat{\boldsymbol{W}}_{an}\boldsymbol{\omega}_n^{\mathrm{T}}\hat{\boldsymbol{W}}_{cn},\end{aligned}\tag{49}$$

where $\xi_{an}>0$ is a parameter.

Remark 3 By using ATIBLF, the dynamic state constraint can be handled directly. It can be seen from Eqs. (14), (29), and (44) that the designed ATIBLF-based optimal controller leads to a different design process from other adaptive optimal constrained control approaches (Li YM et al., 2022b; Wang et al., 2022b; Zhang YX et al., 2024b).

Remark 4 The gradient of the optimal cost function is designed and the term ATIBLF with respect to the time derivative is considered in the gradient of cost function, which gives the optimal cost function indirectly. It is difficult to design the actor–critic adaptive updating weights with IBLF. By introducing terms $\overline{q}_i V_i^z$ and P_i, the actor–critic adaptive law is designed to ensure the convergence of the closed-loop system. In the supplementary materials, we show that all signals of the closed-loop system are bounded, as well as theoretical support for the adaptive optimal control method for nonlinear systems with dynamic constraints.

4 Stability analysis

Theorem 1 Considering the nonlinear system (1) with dynamic state constraints, assuming that Assumptions 1–4 hold, then the optimal virtual controllers (14), (29) and the optimal actual controller (44), together with the weight adaptive laws of the critic NNs (18), (33), (48), and those of the actor NNs (19), (34), (49) can ensure that the following conclusions hold:

(1) all the closed-loop signals are bounded;

(2) all state variables are ensured not to violate the preselected constrained sets.

The proof is given in the supplementary materials.

Remark 5 From inequalities (S48) and (S49) in the supplementary materials, we can get some guidance on how to choose the control parameters. To have smaller tracking errors, we need to enlarge c while reducing η. Thus, we can choose large k_i and small ξ_{ai}. It is worth noting that a smaller $\overline{q}_i$ reduces η but not necessarily the tracking error, the reason being that $\overline{q}_i$ in d is weaker than the reduction in $\sum_{i=1}^n \overline{q}_i k_i\Phi_i z_i^2$ in inequality (S48), and $-\sum_{i=1}^n \overline{q}_i k_i\Phi_i z_i^2$ becomes larger despite the reduction in d. However, it can be concluded from inequalities (S19), (S32), and (S45) that by increasing η_i, a smaller tracking error can be obtained. Furthermore, the energy cost will become higher with the increase of control parameter values. Therefore, appropriate and reasonable control parameters need to be selected to obtain better control performance.

5 Simulation studies

In this section, two simulation examples are given to verify the effectiveness of the proposed control scheme. Comparative simulations of different control methods for the same control parameters are given in Example 1, and comparative simulations with different control parameters for the same control methods are given in Example 2.

5.1 Example 1

We consider a pendulum system as follows:

$$ml\ddot{\theta}=-mg\sin\theta-kl\ddot{\theta}+\frac{\tau}{l},\tag{50}$$

where θ is the angle subtended by the rod and the vertical axis through the pivot point, and τ is the torque applied to the pendulum. The values of parameters m, g, k, and l can be referred to Wei QL et al. (2016). Let $x_1=\theta$ and $x_2=\dot{\theta}$. Then, Eq. (50)

can be written as

$$\begin{cases}\dot{x}_1 = x_2,\\ \dot{x}_2 = f(x) + \dfrac{1}{ml^2}u,\end{cases} \tag{51}$$

where x_1 and x_2 denote the system states, u is the control input, and $f(x) = -(g/l)\sin x_1 - (k/m)x_2$ is a known function.

In the simulations, the initial values are set as $x_1(0) = \pi/10$, $x_2(0) = \pi/10$, $u(0) = 0$; the design parameters are selected as $k_1 = 5$, $k_2 = 5$, $\xi_{c1} = 2$, $\xi_{a1} = 20$, $\xi_{c2} = 2$, $\xi_{a2} = 20$. The desired trajectory is described by $y_d = 0.4\sin(0.4t) + 0.2\cos(0.5t)$. The control constraints are given as $k_{a1} = \pi/6 + 0.2\sin(0.5t)$, $k_{b1} = \pi/6 + 0.4\sin(0.4t)$, $k_{a2} = \pi/3 + 0.2\sin(0.5t)$, $k_{b2} = \pi/3 + 0.4\sin(0.4t)$. The initial values of the updating laws for actor–critic NNs are selected as $\hat{\boldsymbol{W}}_{c1}(0) = \hat{\boldsymbol{W}}_{a1}(0) = \hat{\boldsymbol{W}}_{c2}(0) = \hat{\boldsymbol{W}}_{a2}(0) = \boldsymbol{0}_{2000\times 201}$, and the width of the Gaussian kernel function is 8. The simulation time is $t = 20$ s.

An ATIBLF-based traditional backstepping control (ATIBLF-TBC) method with the same constraint conditions and design parameters is selected for comparison without considering optimal learning. The ATIBLF-TBC being compared is given as follows:

$$\begin{aligned}\alpha_1 = &-\left[k_1 z_1 + \frac{\dot{k}_{a1}}{\Phi_1}\left(\frac{k_{a1}+k_{b1}}{k_{a1}+z_1+y_d} + I_1\right)\right.\\ &\left. - \frac{\dot{k}_{b1}}{\Phi_1}\left(\frac{k_{a1}+k_{b1}}{k_{b1}-z_1-y_d} - Q_1\right) + \frac{\dot{y}_d}{\Phi_1}\Psi_1\right],\end{aligned} \tag{52}$$

$$\begin{aligned}u = &-ml^2\left[k_2 z_2 + \frac{z_2(k_{a1}+k_{b1})^2(k_{a2}-x_2)(k_{b2}+x_2)}{(k_{a2}+k_{b2})^2(k_{a1}-x_1)(k_{b1}+x_1)}\right.\\ &\left. + f(x) - \dot{\alpha}_1 + \frac{z_2(k_{a1}+k_{b1})^2}{(k_{a1}+x_1)(k_{b1}-x_1)}\right].\end{aligned} \tag{53}$$

The simulation results are presented in Figs. 2–5. In the figures, the designed ATIBLF-based adaptive optimal control approach is abbreviated as ATIBLF-OBC. Fig. 2 illustrates the trajectories of the system state and the tracking error. Compared with ATIBLF-TBC, the proposed method has a smaller tracking error. Figs. 2 and 3 illustrate that the dynamic constraints of the states are never violated. Fig. 4 presents the trajectory of the optimal actual control law. The trajectories of weights of actor–critic NNs are shown in Fig. 5. As shown in Figs. 2–5, all signals in the closed-loop system are bounded.

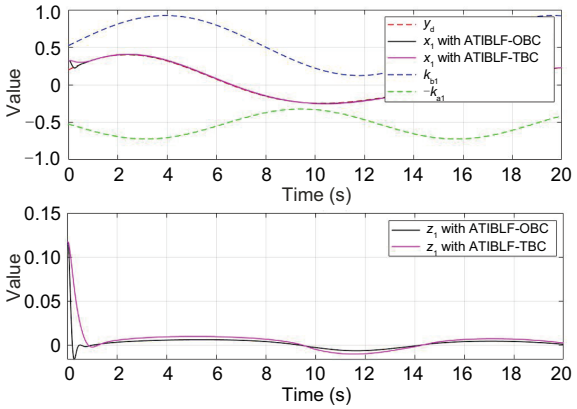

Fig. 2 Trajectories of x_1 and tracking error z_1 in Example 1 (References to color refer to the online version of this figure)

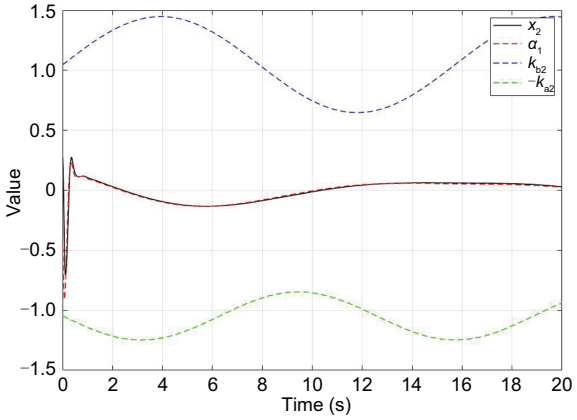

Fig. 3 Trajectories of x_2 and virtual control law α_1 in Example 1 (References to color refer to the online version of this figure)

5.2 Example 2

We consider a one-link robot system as follows (Zhang HG et al., 2021):

$$\begin{cases}D\ddot{q} + B\dot{q} + N\sin q = \tau,\\ M\dot{\tau} + H\tau = U - K_m\dot{q},\end{cases} \tag{54}$$

where q, $\dot{q}$, and $\ddot{q}$ represent the link position, velocity, and acceleration respectively, τ is the torque produced by the electrical subsystem, and U is the control input used to indicate the electromechanical torque. D, B, N, M, H, and K_m represent the mechanical inertia, the coefficient of viscous friction at the joint, a positive constant related to the mass of the load and the coefficient of gravity, the armature inductance, the armature resistance, and the back-electromotive force (back-EMF) coefficient respectively, and satisfy $D = 1\ \mathrm{kg\cdot m^2}$, $B = 1\ \mathrm{N\cdot m\cdot s/rad}$, $N = 10$, $M = 0.05$, $H = 0.5\ \Omega$, $K_m = 10\ \mathrm{N\cdot m\cdot A}$.

Let $x_1 = q$, $x_2 = \dot{q}$, and $x_3 = \tau$, system (54) can be transformed into the following form:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3 + f_2(x), \\ \dot{x}_3 = u + f_3(x), \end{cases} \tag{55}$$

where $f_2(x) = -\dfrac{B}{D}x_2 - \dfrac{N}{D}\sin x_1 + \sin x_2 \cos x_3$ and $f_3(x) = -\dfrac{K_{\mathrm{m}}}{M}x_2 - \dfrac{H}{M}x_3$.

The desired trajectory is described by $y_{\mathrm{d}} = 0.4\sin(0.4t) + 0.2\cos(0.5t)$. The control constraints are given as $k_{\mathrm{a1}} = \pi/3 + 0.2\cos(0.5t)$, $k_{\mathrm{b1}} = \pi/3 + 0.2\sin(0.8t)$, $k_{\mathrm{a2}} = \pi/3 + 0.2\sin(0.5t)$, $k_{\mathrm{b2}} = \pi/3 + 0.2\sin(0.8t)$, $k_{\mathrm{a3}} = 2\pi + 0.2\cos(0.5t)$, $k_{\mathrm{b3}} = 3\pi + 0.2\sin(0.8t)$. In the simulations, the initial values are set as $x_1(0) = \pi/20$, $x_2(0) = \pi/12$, $x_3(0) = \pi$, $u(0) = 0$; the design parameters are selected as $k_1 = 3$, $k_2 = 3$, $k_3 = 30$, $\xi_{\mathrm{c1}} = 1$, $\xi_{\mathrm{a1}} = 10$, $\xi_{\mathrm{c2}} = 1$, $\xi_{\mathrm{a2}} = 10$, $\xi_{\mathrm{c3}} = 1$, $\xi_{\mathrm{a3}} = 10$. The initial values of the updating laws for actor–critic NNs are selected as $\hat{\boldsymbol{W}}_{\mathrm{c1}}(0) = \hat{\boldsymbol{W}}_{\mathrm{a1}}(0) = \hat{\boldsymbol{W}}_{\mathrm{c2}}(0) = \hat{\boldsymbol{W}}_{\mathrm{a2}}(0) = \hat{\boldsymbol{W}}_{\mathrm{c3}}(0) = \hat{\boldsymbol{W}}_{\mathrm{a3}}(0) = \boldsymbol{0}_{2000\times 201}$, and the width of the Gaussian kernel function is 8. The simulation time is $t = 20$ s.

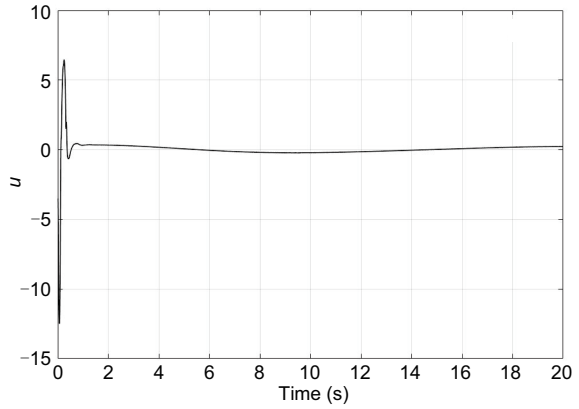

Fig. 4 Trajectory of the optimal actual control law u in Example 1

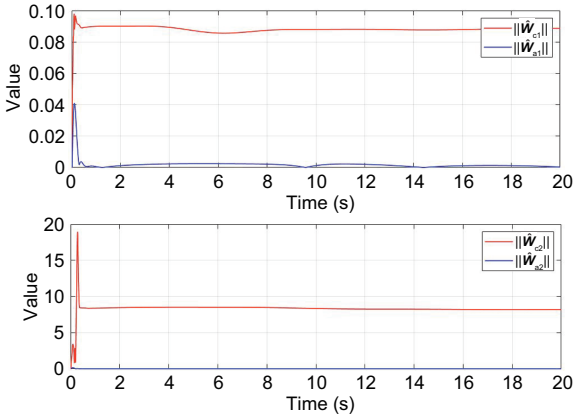

Fig. 5 Trajectories of weights of actor–critic NNs in Example 1 (References to color refer to the online version of this figure)

The simulation results are shown in Figs. 6–9. Fig. 6 displays the trajectories of the output under dynamic constraints and the tracking error. Fig. 7 presents the trajectories of x_2 and x_3 under dynamic constraints. Fig. 8 exhibits the trajectory of the optimal actual control law. The trajectories of weights of actor–critic NNs are depicted in Fig. 9. As shown in Figs. 6 and 7, the full-state constraints are never violated. It can be observed from Figs. 6–9 that all signals in the closed-loop system are bounded.

The selection of different control parameters inevitably leads to different control effects in intelligent control methods. We demonstrate the effects of different control parameters on the methods studied in this study through this example to showcase their comparative performance. We select three types of control parameters, (k_1, k_2, k_3), (ξ_{a1}, ξ_{a2}, ξ_{a3}), and (η_1, η_2, η_3), to show their different effects on the control methods in this study.

Tables 1–3 show the values taken for the three types of control parameters. The values of the control parameters that are not listed in each table are consistent with the values of the control parameters taken in the simulations of Example 2. These three comparisons are only for the parameters in the corresponding tables, and the values of the other

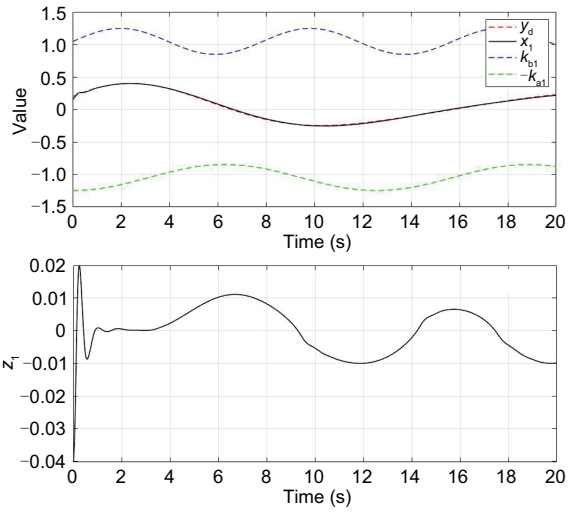

Fig. 6 Trajectories of tracking performance under constraints in Example 2 (References to color refer to the online version of this figure)

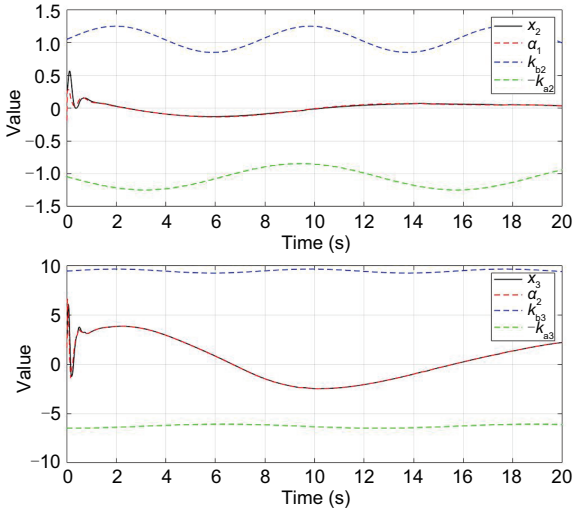

Fig. 7 Trajectories of (x_2, α_1) and (x_3, α_2) in Example 2 (References to color refer to the online version of this figure)

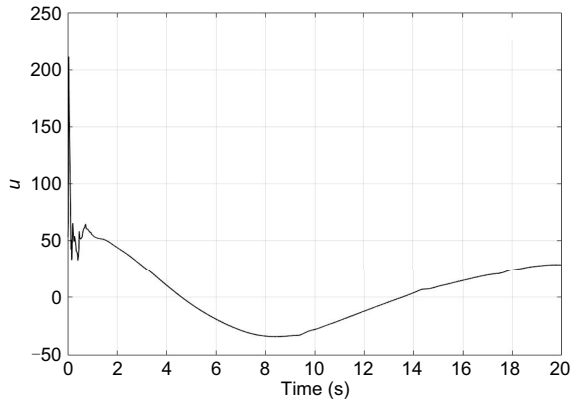

Fig. 8 Trajectory of the optimal actual control law u in Example 2

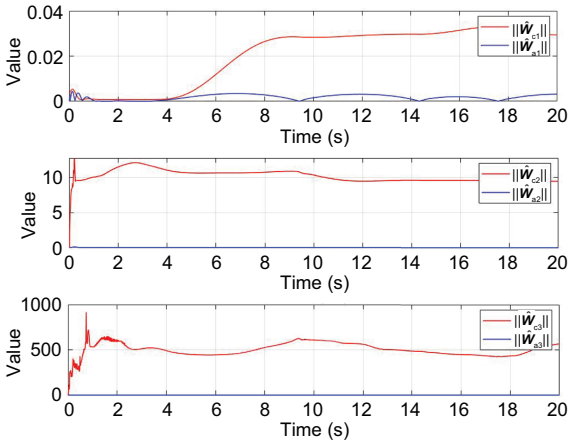

Fig. 9 Trajectories of weights of actor–critic NNs in Example 2 (References to color refer to the online version of this figure)

Table 1 Values of control parameters k_1, k_2, and k_3

Control parameter	Value	Symbol*
k_1, k_2, k_3	3, 3, 30	M1
k_1, k_2, k_3	2.5, 2.5, 29	M2
k_1, k_2, k_3	2, 2, 28	M3
k_1, k_2, k_3	1.5, 1.5, 27	M4
k_1, k_2, k_3	1, 1, 26	M5

* represents each marking of the tracking performance with the parameters k_1, k_2, and k_3

Table 2 Values of control parameters ξ_{a1}, ξ_{a2}, and ξ_{a3}

Control parameter	Value	Symbol*
ξ_{a1}, ξ_{a2}, ξ_{a3}	10, 10, 10	N1
ξ_{a1}, ξ_{a2}, ξ_{a3}	9.5, 9.5, 9.5	N2
ξ_{a1}, ξ_{a2}, ξ_{a3}	9, 9, 9	N3
ξ_{a1}, ξ_{a2}, ξ_{a3}	8.5, 8.5, 8.5	N4
ξ_{a1}, ξ_{a2}, ξ_{a3}	8, 8, 8	N5

* represents each marking of the tracking performance with the parameters ξ_{a1}, ξ_{a2}, and ξ_{a3}

Table 3 Values of control parameters η_1, η_2, and η_3

Control parameter	Value	Symbol*
η_1, η_2, η_3	0.1, 0.1, 0.1	L1
η_1, η_2, η_3	0.2, 0.2, 0.2	L2
η_1, η_2, η_3	0.3, 0.3, 0.3	L3
η_1, η_2, η_3	0.4, 0.4, 0.4	L4
η_1, η_2, η_3	0.01, 0.01, 0.01	L5

* represents each marking of the tracking performance with the parameters η_1, η_2, and η_3

parameters are the same as those in Example 2. In Table 1, symbol M1 is the mark of the control parameter value taken in the simulations of Example 2, and M2–M5 are the marks of the tracking performance difference brought about by each change of k_1, k_2, and k_3. In Table 2, symbol N1 is a marker for the value of the control parameter taken for the simulations in Example 2, and N2–N5 are markers for the tracking performance difference brought about by each change of ξ_{a1}, ξ_{a2}, and ξ_{a3}. In Table 3, symbol L1 is a marker for the value of the control parameter taken for the simulations of Example 2, and L2–L5 are markers for the difference in tracking performance brought about by each change of η_1, η_2, and η_3.

The comparative simulation results are shown in Figs. 10–12. Fig. 10 shows that the tracking error becomes smaller as k_1, k_2, and k_3 become larger. Fig. 11 shows that the tracking error becomes smaller as ξ_{a1}, ξ_{a2}, and ξ_{a3} become smaller, and vice versa.

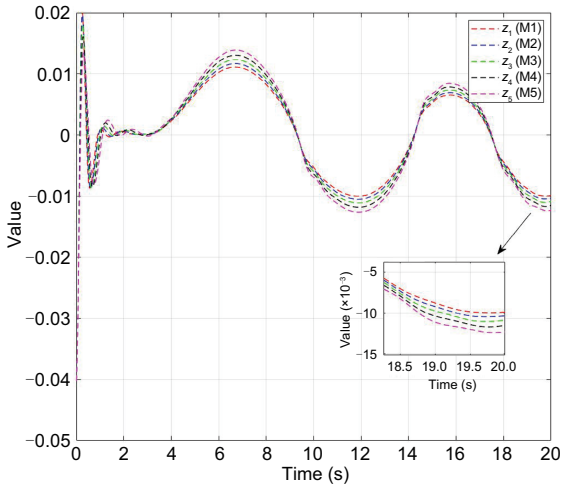

Fig. 10 Tracking errors under different values of parameters k_1, k_2, and k_3 in Example 2 (References to color refer to the online version of this figure)

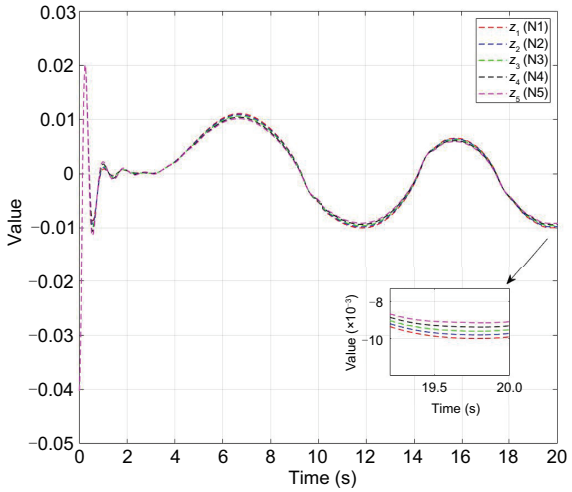

Fig. 11 Tracking errors under different values of parameters ξ_{a1}, ξ_{a2}, and ξ_{a3} in Example 2 (References to color refer to the online version of this figure)

Fig. 12 shows that the tracking error becomes smaller as η_1, η_2, and η_3 become larger, and vice versa.

6 Conclusions

In this paper, an ATIBLF-based adaptive optimal control scheme for a class of strict-feedback nonlinear systems with dynamic state constraints was proposed. Due to the advantage that ATIBLF can handle state constraints directly, the controller was designed under the optimal learning algorithm. Specifically, novel integral barrier optimal performance index functions and the actor–critic framework were employed in the controller design. Subsequently, the stability of the closed-loop system was proven via the Lyapunov stability theorem. Simulation results showed that dynamic state constraints are not violated, and that all signals of the closed-loop system remain bounded. Future work will focus on the adaptive optimal control problems for physical human–robot interaction systems with dynamic state constraints.

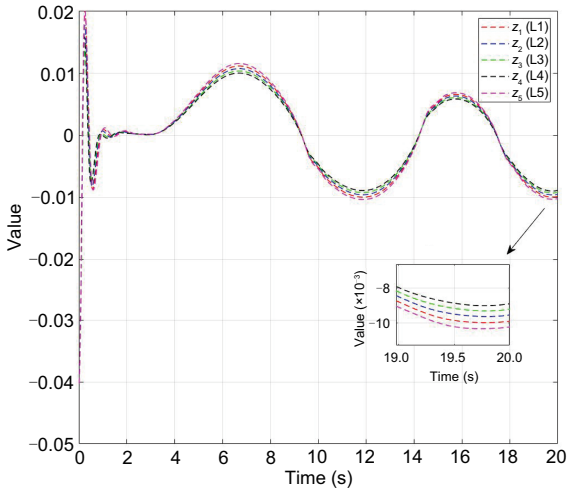

Fig. 12 Tracking errors under different values of parameters η_1, η_2, and η_3 in Example 2 (References to color refer to the online version of this figure)

Contributors

Yan WEI designed the research. Mingshuang HAO conducted the simulation verification. Xinyi YU drafted the paper. Linlin OU helped organize the paper. Yan WEI and Mingshuang HAO revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Bhasin S, Kamalapurkar R, Johnson M, et al., 2013. A novel actor–critic–identifier architecture for approximate optimal control of uncertain nonlinear systems. *Automatica*, 49(1):82-92. https://doi.org/10.1016/j.automatica.2012.09.019

Chen B, Liu XP, Liu KF, et al., 2009. Direct adaptive fuzzy control of nonlinear strict-feedback systems. *Automatica*, 45(6):1530-1535. https://doi.org/10.1016/j.automatica.2009.02.025

Jin X, Xu JX, 2014. A barrier composite energy function approach for robot manipulators under alignment condition with position constraints. *Int J Robust Nonl Contr*, 24(17):2840-2851. https://doi.org/10.1002/rnc.3028

Li DY, Ge SS, Lee TH, 2021. Simultaneous arrival to origin convergence: sliding-mode control through the norm-normalized sign function. *IEEE Trans Automat Contr*, 67(4):1966-1972. https://doi.org/10.1109/TAC.2021.3069816

Li DY, Yu HY, Tee KP, et al., 2022. On time-synchronized stability and control. *IEEE Trans Syst Man Cybern Syst*, 52(4):2450-2463. https://doi.org/10.1109/TSMC.2021.3050183

Li Y, Qiang S, Zhuang X, et al., 2004. Robust and adaptive backstepping control for nonlinear systems using RBF neural networks. *IEEE Trans Neur Netw*, 15(3):693-701. https://doi.org/10.1109/TNN.2004.826215

Li YM, Fan YL, Li KW, et al., 2022a. Adaptive optimized backstepping control-based RL algorithm for stochastic nonlinear systems with state constraints and its application. *IEEE Trans Cybern*, 52(10):10542-10555. https://doi.org/10.1109/TCYB.2021.3069587

Li YM, Zhang JX, Liu W, et al., 2022b. Observer-based adaptive optimized control for stochastic nonlinear systems with input and state constraints. *IEEE Trans Neur Netw Learn Syst*, 33(12):7791-7805. https://doi.org/10.1109/TNNLS.2021.3087796

Liu BJ, Hou MS, Ni JK, et al., 2020. Asymmetric integral barrier Lyapunov function-based adaptive tracking control considering full-state with input magnitude and rate constraint. *J Franklin Inst*, 357(14):9709-9732. https://doi.org/10.1016/j.jfranklin.2020.07.037

Liu L, Gao TT, Liu YJ, et al., 2021. Time-varying IBLFs-based adaptive control of uncertain nonlinear systems with full state constraints. *Automatica*, 129:109595. https://doi.org/10.1016/j.automatica.2021.109595

Liu L, Chen AQ, Liu YJ, 2022. Adaptive fuzzy output-feedback control for switched uncertain nonlinear systems with full-state constraints. *IEEE Trans Cybern*, 52(8):7340-7351. https://doi.org/10.1109/TCYB.2021.3050510

Liu YC, Zhu QD, Wen GX, 2022. Adaptive tracking control for perturbed strict-feedback nonlinear systems based on optimized backstepping technique. *IEEE Trans Neur Netw Learn Syst*, 33(2):853-865. https://doi.org/10.1109/TNNLS.2020.3029587

Liu YJ, Ma L, Liu L, et al., 2020. Adaptive neural network learning controller design for a class of nonlinear systems with time-varying state constraints. *IEEE Trans Neur Netw Learn Syst*, 31(1):66-75. https://doi.org/10.1109/TNNLS.2019.2899589

Luo X, Mu DR, Wang Z, et al., 2023. Adaptive full-state constrained tracking control for mobile robotic system with unknown dead-zone input. *Neurocomputing*, 524:31-42. https://doi.org/10.1016/j.neucom.2022.12.025

Mei KQ, Ding SH, Chen CC, 2022. Fixed-time stabilization for a class of output-constrained nonlinear systems. *IEEE Trans Syst Man Cybern Syst*, 52(10):6498-6510. https://doi.org/10.1109/TSMC.2022.3146011

Mohammadi M, Arefi MM, Setoodeh P, et al., 2021. Optimal tracking control based on reinforcement learning value iteration algorithm for time-delayed nonlinear systems with external disturbances and input constraints. *Inform Sci*, 554:84-98. https://doi.org/10.1016/j.ins.2020.11.057

Shen LY, Wang HQ, Yue HX, 2022. Prescribed performance adaptive fuzzy control for affine nonlinear systems with state constraints. *IEEE Trans Fuzzy Syst*, 30(12):5351-5360. https://doi.org/10.1109/TFUZZ.2022.3175606

Su QY, Wan M, 2020. Adaptive neural dynamic surface output feedback control for nonlinear full states constrained systems. *IEEE Access*, 8:131590-131600. https://doi.org/10.1109/ACCESS.2020.3010027

Tee KP, Ge SS, 2012. Control of state-constrained nonlinear systems using integral barrier Lyapunov functionals. Proc IEEE 51st IEEE Conf on Decision and Control, p.3239-3244. https://doi.org/10.1109/CDC.2012.6426196

Vamvoudakis KG, Lewis FL, 2010. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 46(5):878-888. https://doi.org/10.1016/j.automatica.2010.02.018

Wang N, Fu ZM, Song SZ, et al., 2022a. Barrier-Lyapunov-based adaptive fuzzy finite-time tracking of pure-feedback nonlinear systems with constraints. *IEEE Trans Fuzzy Syst*, 30(4):1139-1148. https://doi.org/10.1109/TFUZZ.2021.3053322

Wang N, Gao Y, Liu YJ, et al., 2022b. Self-learning-based optimal tracking control of an unmanned surface vehicle with pose and velocity constraints. *Int J Robust Nonl Contr*, 32(5):2950-2968. https://doi.org/10.1002/rnc.5978

Wei QL, Song RZ, Yan PF, 2016. Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP. *IEEE Trans Neur Netw Learn Syst*, 27(2):444-458. https://doi.org/10.1109/TNNLS.2015.2464080

Wei Y, Wang YY, Ahn CK, et al., 2021. IBLF-based finite-time adaptive fuzzy output-feedback control for uncertain MIMO nonlinear state-constrained systems. *IEEE Trans Fuzzy Syst*, 29(11):3389-3400. https://doi.org/10.1109/TFUZZ.2020.3021733

Wei Y, Hao M, Yu X, et al., 2023. Adaptive neural optimal control of a robot manipulator with time-varying state constraints. Int Annual Conf on Complex Systems and Intelligent Science, p.294-301. https://doi.org/10.1109/CSIS-IAC60628.2023.10364204

Wen GX, Chen CLP, Li WN, 2020. Simplified optimized control using reinforcement learning algorithm for a class of stochastic nonlinear systems. *Inform Sci*, 517:230-243. https://doi.org/10.1016/j.ins.2019.12.039

Wen GX, Chen CLP, Ge SS, 2021. Simplified optimized backstepping control for a class of nonlinear strict-feedback systems with unknown dynamic functions. *IEEE Trans Cybern*, 51(9):4567-4580. https://doi.org/10.1109/TCYB.2020.3002108

Xu ZB, Sun CB, Liu QY, 2023. Output-feedback prescribed performance control for the full-state constrained nonlinear systems and its application to DC motor system. *IEEE Trans Syst Man Cybern Syst*, 53(7):3898-3907. https://doi.org/10.1109/TSMC.2022.3216119

Zhang HG, Liu Y, Wang YC, 2021. Observer-based finite-time adaptive fuzzy control for nontriangular nonlinear systems with full-state constraints. *IEEE Trans Cybern*, 51(3):1110-1120. https://doi.org/10.1109/TCYB.2020.2984791

Zhang LL, Zhu LC, Hua CC, et al., 2023. Adaptive decentralized control for interconnected time-delay uncertain nonlinear systems with different unknown control directions and deferred full-state constraints. *IEEE Trans Neur Netw Learn Syst*, 34(12):10789-10801. https://doi.org/10.1109/TNNLS.2022.3171518

Zhang YX, Liang XL, Li DY, et al., 2024a. Adaptive safe reinforcement learning with full-state constraints and constrained adaptation for autonomous vehicles. *IEEE Trans Cybern*, 54(3):1907-1920. https://doi.org/10.1109/TCYB.2023.3283771

Zhang YX, Liang XL, Li DY, et al., 2024b. Barrier Lyapunov function-based safe reinforcement learning for autonomous vehicles with optimized backstepping. *IEEE Trans Neur Netw Learn Syst*, 35(2):2066-2080. https://doi.org/10.1109/TNNLS.2022.3186528

Zhao K, Song YD, Chen CLP, et al., 2020. Control of nonlinear systems under dynamic constraints: a unified barrier function-based approach. *Automatica*, 119:109102. https://doi.org/10.1016/j.automatica.2020.109102

List of supplementary materials