Research Article

Adaptive Critic Learning-Based Robust Control of Systems with Uncertain Dynamics

Jun Zhao,[1] **Qingliang Zeng** ⓘ**,**[1] **and Bin Guo** ⓘ[2]

[1]*College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, China*
[2]*College of Transportation, Shandong University of Science and Technology, Qingdao 266590, China*

Correspondence should be addressed to Bin Guo; gb_bb@126.com

Model uncertainties are usually unavoidable in the control systems, which are caused by imperfect system modeling, disturbances, and nonsmooth dynamics. This paper presents a novel method to address the robust control problem for uncertain systems. The original robust control problem of the uncertain system is first transformed into an optimal control of nominal system via selecting the appropriate cost function. Then, we develop an adaptive critic leaning algorithm to learn online the optimal control solution, where only the critic neural network (NN) is used, and the actor NN widely used in the existing methods is removed. Finally, the feasibility analysis of the control algorithm is given in the paper. Simulation results are given to show the availability of the presented control method.

1. Introduction

The basis of intelligent optimization decision-using adaptive dynamic programming (ADP) method is the optimal control design. There are many mature methods for optimal regulation control design of linear systems in the field of control theory and control engineering. However, for general nonlinear systems, Hamilton–Jacobi–Bellman (HJB) equation is yielded. The analytical solution of HJB equation is not easy since it is inherently a partial differential equation. Recently, the optimal control design of systems has attracted extensive attention. Among them, the successive approximation methods [1–3] overcome this difficulty via finding the approximate solution of HJB equation, which is closely related to the ADP method. ADP is a new method based on the idea of intelligent learning, which can provide effective optimal control solution for complex dynamic systems [4, 5]. In the past two decades, ADP has been widely used in solving adaptive optimal control problems of discrete-time and continuous-time systems [6, 7]. Now, data-driven control design has become a research hotspot in the field of control theory and control engineering [8, 9]. The ADP methods can promote the research of data-based decision-making and optimal control and is conducive to the development of artificial intelligence and computational intelligence technology.

Most of the existing results of ADP methods are obtained without considering the uncertainty of the controlled plant. However, the actual control system is always affected by model uncertainty, external disturbance, or other changes. We must consider these factors in the controller design to avoid the deterioration of the performance for the closed-loop system and improve the robustness of the controlled system. For robust control design, several alternative methodologies have been suggested in the control community. The work in [10] exploited the relationship between the robust control and the optimal control of nominal system subject to a specific value function. It indicates that one can design a robust control by solving an equivalent optimal control problem alternatively. Similarly, it was shown in [11] that the robust control design may be accomplished by addressing an H control problem. Nevertheless, online solving the derived optimal control equations was not discussed in [10]. Instead, they adopted offline schemes to seek for the solution of the derived optimal control equations. Recently, robust control design using the adaptive critic learning method has gradually become one of the research hotspots in the field of ADP, and many methods

have been proposed [12–14]. These results fully show that the ADP method is suitable for robust control design of complex nonlinear systems in uncertain environment. Since many previous ADP literatures do not focus on the robust performance of the controller, the emergence of robust adaptive critic control greatly expands the application scope of the ADP method. Generally, the controller based on robust ADP can not only stabilize the original uncertain system but also make the system optimal without dynamic uncertainty. Thus, adaptive critic learning-based robust control includes the discussion of system stability, convergence, optimality, and robustness. It plays an important role in the field of intelligent learning control of complex systems in uncertain environment.

Based on the above facts, we develop an adaptive critic learning algorithm to resolve the robust control problem of uncertain systems. To this end, we construct an equivalence between the robust control problem and the optimal control problem via selecting the appropriate cost function; then, a single critic NN is used to reformulate the cost function. To realize the optimal control solution, we design an adaptive critic leaning algorithm; since it has strong convergence, the actor NN widely used in existing ADP results is removed. Then, the feasibility analysis of the control algorithm is also given in the paper. Simulations are given to indicate the validity of the developed method.

The major contributions of this paper include

(1). To address the robust control problem, we transform the robust control problem of uncertain systems into an optimal control problem of the nominal system. It provides a new approach to address the robust control problem.

(2). Different to [13], the uncertainty in the input matrix is considered in this paper, and then, the proposed control method is used in robotic systems. This helps to apply the proposed control algorithm to the practical industrial robotic systems in the future.

(3). A novel designed adaptation algorithm driven by the NN weights' errors is used to online learn the critic NN weights. Different to [15], the convergence of the estimated NN weights to the true values can be retained.

This paper is organized as follows. In Section 2, we introduce the robust control problem and transform the robust control problem into an optimal control problem. In Section 3, a single critic NN is used to reformulate the optimal cost function, and then, an adaptive critic learning method is proposed to address the derived optimal control problem. Section 4 gives some simulation results to illustrate the effectiveness of the proposed method. Some conclusions are stated in Section 5.

2. Preliminaries and Problem Formulation

A continuous-time (CT) uncertain system can be written as

$$\dot{x}(t) = f(x) + g(x)(u + b(x)u) + g(x)d(x), \quad (1)$$

where $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$ are the system state and the control action, respectively. $f(x) \in \mathbb{R}^n$ with $f(0) = 0$ and $g(x) \in \mathbb{R}^{n \times m}$ are the nonlinear functions. $b(x)$ and $d(x)$ are the uncertainties. The purpose of this paper is designing a controller to make system (1) asymptotically stable under the uncertainties $b(x)$ and $d(x)$. To this end, we give following assumptions.

Assumption 1. $b(x) \geq 0$ is the uncertainty in the input matrix. The uncertainty $d(x)$ is bounded, i.e., there exists a nonnegative function $d_{\max}(x)$ such that $\|d(x)\| \leq d_{\max}(x)$.

To design a robust controller for the linear system, a linear matrix inequality (LMI) is proposed [16], while for nonlinear system (1), it is not easy. Inspired by [10, 12], an equivalence is built between the robust control problem of the uncertain system and the optimal control of the nominal system via selecting the appropriate cost function. Thus, we define the nominal system of the uncertain system (1) as

$$\dot{x}(t) = f(x) + g(x)u. \quad (2)$$

For system (2), a control action u should be found to minimize the following cost function [17]:

$$V(x) = \int_t^\infty \left(d_{\max}^2(x) + x^T M x + u^T N u\right) ds, \quad (3)$$

where $M \in \mathbb{R}^{n \times n}$ and $N \in \mathbb{R}^{m \times m}$ are the positive definite weight matrices. Hence, based on the optimal principle, we can obtain the Lyapunov function of the cost function (3) as

$$0 = V_x^T (f(x) + g(x)u) + d_{\max}^2(x) + x^T M x + u^T N u, \quad (4)$$

where $V_x = \partial V / \partial x$ is the derivative of V with respect to x.

Therefore, we can get the optimal cost function as

$$V^*(x) = \min V(x), \quad (5)$$

and its corresponding HJB equation can be given as

$$0 = \min H(x, u^*, V^*). \quad (6)$$

By solving (6), we have the optimal control action as

$$u^* = -\frac{1}{2} N^{-1} g^T \frac{\partial V^*(x)}{\partial x}. \quad (7)$$

Then, we will give the lemma to explain the robust control problem of system (1) which can be transformed into an optimal control problem of system (2) via constructing cost function (3).

Lemma 1 (see [11, 18]). *Assume that the solution can be solved via optimal control problem of system (2) with cost function (3) and $d_{\max}^2(x) \geq d^T(x) N \ d(x)$, and this solution can make uncertain system (1) asymptotically stable, which means that the optimal control solution is the solution of the robust control problem for system (1).*

Proof. Because $V^*(x) > 0$ for $x \neq 0$ and $V^*(0) = 0$ for $x = 0$ given in (5), then we can consider $V^*(x)$ is a Lyapunov function; based on (6) and (7), we have

$$
\begin{aligned}
\dot{V}^*(x) &= V_x^{*T} \dot{x} \\
&= V_x^{*T} \left(f(x) + g(x)(u^* + b(x)u^*) + g(x)d(x) \right) \\
&= -d_{\max}^2(x) - x^T M x - u^{*T} N u^* + V_x^{*T} g(x)b(x)u^* + V_x^{*T} g(x)d(x) \\
&= -d_{\max}^2(x) - x^T M x - u^{*T} N u^* - 2u^{*T} N \, d(x) - 2u^{*T} N b(x)u^* \\
&\leq -d_{\max}^2(x) - x^T M x - u^{*T} N u^* - 2u^{*T} N \, d(x) \\
&= -d_{\max}^2(x) - x^T M x - u^{*T} N u^* + d^T(x) N \, d(x) - d^T(x) N \, d(x) - 2u^{*T} N \, d(x) \\
&= -d_{\max}^2(x) + d^T(x) N \, d(x) - x^T M x - \left(u^* + d(x) \right)^T N \left(u^* + d(x) \right) \\
&\leq -d_{\max}^2(x) + d^T(x) N \, d(x) - x^T M x.
\end{aligned}
\tag{8}
$$

According to the condition given in Lemma 1, i.e., $d_{\max}^2(x) \geq d^T(x) N \, d(x)$, we obtain

$$
\dot{V}^*(x) \leq -x^T M x, \tag{9}
$$

i.e., $\dot{V}^*(x) < 0, x \neq 0$, and $\dot{V}^*(0) = 0$, and the uncertain system (1) is asymptotically stable for any uncertainties $b(x)$ and $d(x)$. According to the above facts, the optimal solution u^* is the robust control solution of the uncertain system (1). This completes the proof.

From Lemma 1, we have that if we select $N = I_m$, where I_m is the identity matrix with $m \times m$, then $d_{\max}^2(x) \geq d^T(x) N \, d(x) = d^T(x)d(x) = \|d(x)\|^2$ holds. $\quad\square$

Remark 1. Lemma 1 shows that the robust control problem of the original uncertain system can be equivalent to the optimal control problem of the nominal system, and then, the solution of the robust control problem can be obtained indirectly by solving the optimal control problem. Therefore, this equivalence relationship can be used to develop a new robust control design method and solve it by using ADP method, as described in the following section.

Remark 2. It is well-known that H_∞ control belongs to robust control. Although many H_∞ control design techniques have been proposed, it should be noted that, as explained in Section 8.5 in [18], the H_∞ control differs in the optimal method proposed in this paper. In the optimal control method, we start from the uncertainty bounds and then design the controller according to these bounds. Hence, if the controller exists, we can say the uncertain system is robustly stable.

3. Solving the Robust Control Problem via Adaptive Critic Learning

To obtain the optimal control solution (7), the unknown cost function (5) should be resolved. However, it is quite difficult to address the cost function (5) directly; then, a critic NN in this section will be proposed to approximate the cost function (5); this allows to develop an adaptive learning method to update online the NN weights, where the convergence of NN weights can be retained. Because its strong convergence, the actor NN widely used in the ADP schemes is removed. The proposed control system structure is given in Figure 1.

This section will propose an adaptive critic learning method to obtain the solution of the derived optimal control problem. To this end, a critic NN is trained to estimate the cost function $V^*(x)$, where the cost function $V^*(x)$ is considered as continuous; hence, we have the following NN [13],

$$
V^*(x) = W^T \sigma(x) + \varepsilon_v(x), \tag{10}
$$

where $W \in \mathbb{R}^l$ is the ideal critic NN weight, $\sigma(x) \in \mathbb{R}^l$ is the regressor vector, l is the number of neurons, and $\varepsilon_v(x)$ is the approximate error of NN.

Then, we have the partial derivative of cost function as

$$
V_x^*(x) = (\nabla \sigma(x))^T W + \nabla \varepsilon_v(x), \tag{11}
$$

where $\nabla \sigma(x) = \partial \sigma(x)/\partial x \in \mathbb{R}^{l \times n}$ is the regressor matrix and $\nabla \varepsilon_v(x) = \partial \varepsilon_v(x)/\partial x \in \mathbb{R}^n$ is the NN error.

Without loss of generality, the following assumption is given in [13].

Assumption 2. The NN weight W, the regressor vector σ, the regressor matrix $\nabla \sigma(x)$, and the approximate errors $\varepsilon_v(x)$ and $\nabla \varepsilon_v(x)$ are all bounded, i.e., $\|W\| \leq W_M, \|\sigma\| \leq \sigma_N, \|\nabla \sigma\| \leq \sigma_M$, and $\|\nabla \varepsilon_v\| \leq \sigma_\varepsilon$ for positive constants W_M, σ_N, σ_M, and σ_ε.

In fact, the ideal NN weight W is unknown; hence, we have that the practice $\widehat{W}$ can be online updated; then, the actual cost function can be written as

$$
\widehat{V}(x) = \widehat{W}^T \sigma(x). \tag{12}
$$

Hence, the practical estimated optimal solution can be solved as

$$
\widehat{V}_x(x) = (\nabla \sigma(x))^T \widehat{W}. \tag{13}
$$

According to (10) and (11), we have the ideal optimal control (7) as

$$
u^* = -\frac{1}{2} N^{-1} g^T \left[(\nabla \sigma(x))^T W + \nabla \varepsilon_v(x) \right], \tag{14}
$$

and its practical optimal control is given as

$$
\widehat{u} = -\frac{1}{2} N^{-1} g^T (\nabla \sigma(x))^T \widehat{W}. \tag{15}
$$

The problem next to be solved is solving the unknown weight $\widehat{W}$, which can guarantee the stability of the controlled system and realize the convergence to the ideal value W. Most existing ADP methods can only get the uniform ultimate boundedness (UUB) of the approximated NN weight rather
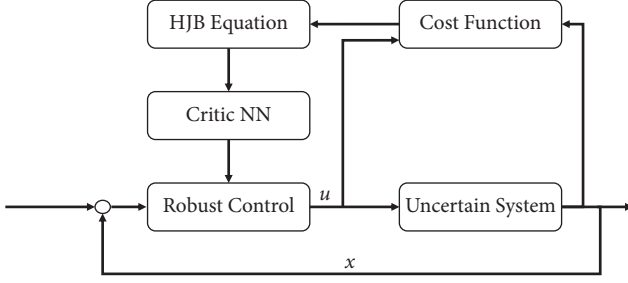
FIGURE 1: Schematic of the proposed control system.

than the convergence. In this paper, a novel adaptive critic learning method is introduced to guarantee the convergence of $\widehat{W}$ to W. This strong convergence property is conducive to avoiding the use of actor NN, and the optimal control approximated via critic NN can converge to its ideal optimal solution.

Substituting (11) into (4), we can rewrite the HJB equation as

$$
0 = W^T \nabla \sigma(x)(f(x) + g(x)u) + d_{\max}^2(x) + x^T M x \\
+ u^T N u + \varepsilon_{\mathrm{HJB}},
\tag{16}
$$

where $\varepsilon_{\mathrm{HJB}} = (\nabla \varepsilon_v(x))^T [f(x) + g(x)u]$ is the residual error determined by the approximation error $\nabla \varepsilon_v(x)$.

For developing an adaptive critic learning law to estimate the critic NN weight W, the known terms in (16) can be defined as

$$
\begin{cases}
\Xi = \nabla \sigma(x)[f(x) + g(x)u], \\
\Theta = d_{\max}^2(x) + x^T M x + u^T N u.
\end{cases}
\tag{17}
$$

Then, the HJB equation (16) with (17) can be given as

$$
\Theta = -W^T \Xi - \varepsilon_{\mathrm{HJB}}.
\tag{18}
$$

According to (18), only the NN weight W is unknown in this parameterized formulation. Hence, it can be estimated by using a recently proposed learning algorithm [19, 20], which is driven by the derived estimation error.

To this end, the filtered regressor matrices $P \in \mathbb{R}^{l \times l}$ and $Q \in \mathbb{R}^l$ can be denoted as [19, 20]

$$
\begin{cases}
\dot{P} = -\ell P + \Xi \Xi^T, & P(0) = 0, \\
\dot{Q} = -\ell Q + \Xi \Theta, & \Theta(0) = 0,
\end{cases}
\tag{19}
$$

where $\ell > 0$ is a positive parameter. Hence, its solution can be derived as

$$
\begin{cases}
P = \int_0^t e^{-\ell(t-s)} \Xi(s) \Xi^T(s) \mathrm{d}s, \\
Q = \int_0^t e^{-\ell(t-s)} \Xi(s) \Theta(s) \mathrm{d}s,
\end{cases}
\tag{20}
$$

which can be online calculated based on the system state x.

With P and Q in (20), an auxiliary vector $M_1 \in \mathbb{R}^l$ can be defined as

$$
M_1 = -P\widehat{W} + Q.
\tag{21}
$$

From (18) and (20), we have $Q = -PW + v$ with $v = -\int_0^t e^{-\ell(t-s)} \varepsilon_{\mathrm{HJB}}(s) \Xi(s) \mathrm{d}s$. A bounded variable, e.g., $\|v\| \le \varepsilon_v$, for a positive constant, $\varepsilon_v > 0$. Then, we can obtain from (19)–(21) that

$$
M_1 = -P\widetilde{W} + v,
\tag{22}
$$

with $\widetilde{W} = W - \widehat{W}$ being the NN weight estimation error.

The estimation error used in the adaptive learning algorithm can help to guarantee the convergence of the estimate, as shown in [13]. Hence, we can design the following adaptive law to online calculate $\widehat{W}$ as

$$
\dot{\widehat{W}} = -\Gamma M_1,
\tag{23}
$$

with $\Gamma > 0$ being the adaptive learning gain.

Remark 3. The adaptive law (23) is driven by the estimation error $\widetilde{W}$. The purpose of this new learning algorithm is to guarantee the convergence of estimate $\widehat{W}$ to unknown weight W. Therefore, the learning algorithm given in this paper is different from those used in the existing ADP methods, e.g., [3, 21], which employ the gradient-based methods [22] to guarantee the boundedness of $\widehat{W}$ only.

To illustrate the convergence of the proposed learning algorithm, the positive definiteness of the matrix P defined in (20) will be introduced:

Lemma 2. *When the regressor Ξ in (18) fulfills the persistent excitation (PE) condition, the matrix P defined in (20) is positive.*

The convergence of the proposed learning algorithm can be summarized as follows.

Theorem 1. *For the adopted critic NN with adaptive law (23), if the regressor vector Ξ in (18) satisfies the PE condition, the critic NN weight error $\widetilde{W}$ exponentially converges to a small bounded set around zero.*

Proof. For Lemma 2, the matrix P is positive definite when the regressor Ξ satisfies the PE condition, i.e., the minimum eigenvalue $\lambda_{\min}(P) > \delta > 0$. Hence, a Lyapunov function can be chosen as $V_1 = 1/2\widetilde{W}^T \Gamma^{-1} \widetilde{W}$, and its derivative $\dot{V}_1$ along with (23) can be derived as

$$
\dot{V}_1 = \widetilde{W}^T \Gamma^{-1} \dot{\widetilde{W}} = -\widetilde{W}^T P \widetilde{W} + \widetilde{W}^T v,
\tag{24}
$$

which further implies

$$
\dot{V}_1 = \widetilde{W}^T \Gamma^{-1} \widetilde{W} + \widetilde{W}^T v \le -\|\widetilde{W}\|(\delta\|\widetilde{W}\| - \varepsilon_v).
\tag{25}
$$

Thus, we have that the estimation errors of NN weight $\widetilde{W}$ will converge to a compact set $\Omega: \{\widetilde{W}|\|\widetilde{W}\| \le \varepsilon_v/\delta\}$, in which the size of this set depends on the approximation error ε_v and the excitation level δ, i.e., for an arbitrarily small NN approximation error (according to the NN approximation property, this error can be arbitrarily small for sufficient NN nodes, i.e., $\nabla \varepsilon_v(x) \longrightarrow 0$ with $l \longrightarrow \infty$). Therefore, $\widehat{W}$ can converge to W. In the ideal case, i.e., $\varepsilon_{\mathrm{HJB}} = 0$ and $v = 0$, then

we know the estimation errors of weights $\tilde{W}$ converge to zero exponentially. $\square$

For system (2) with practical optimal control (15) and adaptive law (23), if the regressor Ξ satisfies the PE condition, the error $\tilde{W}$ converges to a small set around zero. Moreover, the actual optimal control u in (15) converges to a region around its optimal solution u^* in (14), i.e., $\|u - u^*\| \leq \varepsilon_u$. Hence, the original robust control problem is resolved.

4. Simulation

4.1. Numerical Simulation.

Consider an uncertain system as

$$\dot{x} = f(x) + g(x)(u + b(x)u) + g(x)d(x), \tag{26}$$

where $f(x) = \begin{bmatrix} -x_1 + x_2 \\ -0.5x_1 - 0.5x_2(1 - (\cos(2x_2) + 2)^2) \end{bmatrix}$, $g(x) = \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix}$, $x = [x_1, \ x_2]^T \in \mathbb{R}^2$ is the system state, $u \in \mathbb{R}$ is the control input, and the term $k(x) = r_2 x_2^2$ with $r_2 \in [0, 1]$ and $d(x) = r_1 x_1 \sin(x_2)$ with $r_1 \in [-1, 1]$ denote the uncertainties.

Because the uncertain terms $b(x) = r_2 x_2^2 \geq 0$ and $d(x) = r_1 x_1 \sin(2x_2)$ are bounded by $\|d(x)\| = |p_1 x_1 \sin(2x_2)| \leq |x_1| = d_{\max}(x)$, then we can obtain the optimal control problem as

$$\dot{x} = f(x) + g(x)u, \tag{27}$$

with the cost function as

$$V(x) = \int_t^{\infty} \left(d_{\max}^2(x) + x^T M x + u^T N u\right) ds. \tag{28}$$

As given in [18], the optimal cost function is written as

$$V^*(x) = x_1^2 + 2x_2^2. \tag{29}$$

Then, we can obtain its optimal solution as

$$u^* = -\frac{1}{2}N^{-1}g^T V_x^* = -(\cos(2x_1) + 2)2x_2. \tag{30}$$

A critic NN will be used to approximate the cost function V; thus, the activation function $\sigma(x)$ is defined as

$$\sigma(x) = \left[x_1^2, x_1 x_2, x_2^2\right]^T. \tag{31}$$

To realize the simulation, we set the learning parameters $\Gamma = 100$ and $\ell = 10$, the initial system state is given as $x_0 = [1, \ -0.2]^T$, the initial weight $\widehat{W}(0) = 0$, and the weight matrices are set as $M = I$ and $N = I$.

Figure 2 shows the estimated value of the evaluation NN weights. It can be seen from Figure 2 that the estimated NN weight converges to a certain value. This result verifies the

convergence of Theorem 1 and the effectiveness of the proposed learning algorithm, which indicates that estimated critic NN weight $\widehat{W}$ converges to its true value, i.e., $W = [1, \ 0, \ 2]$. To better display the performance of the proposed learning method, the error between the ideal cost function V^* and practical coat function $\widehat{V}$ is given in Figure 3, where we can obtain the fairly satisfactory approximation performance. In fact, the simulation results in Figures 2 and 3 can be also found in [13]; different from [13], this paper considers the uncertainties involved in control input. Figure 4 shows the change of the state of the controlled system under the derived optimal control, which shows that the closed-loop system is asymptotically stable. The corresponding control input is shown in Figure 5, bounded and smooth.

4.2. Application to Robotic Systems.

This section will develop a simulation based on a 2-DOF robot [18, 23]. To realize the simulation, the robotic systems model can be defined as

$$\overline{M}(q)\ddot{q} + C(q, \dot{q}) + F(\dot{q}) + G(q) = \tau, \tag{32}$$

where q is the joint variables, τ is the generalized forces, $\overline{M}(q)$ denotes the inertia matrix, $C(q, \dot{q})$ is the centripetal vector, $F(q)$ is the friction vector, and $G(q)$ defines the gravity vector. In this section, we denote $\overline{W}(q, \dot{q}) = C(q, \dot{q}) + F(\dot{q}) + G(q)$. There are uncertainties in $\overline{M}(q)$ and $\overline{W}(q, \dot{q})$ due to the unknown load on the manipulator and unmodeled frictions.

The inertia matrix can be derived as

$$\overline{M}(q) = \begin{bmatrix} \overline{M}_{11} & \overline{M}_{12} \\ \overline{M}_{21} & \overline{M}_{22} \end{bmatrix}, \tag{33}$$

where $\overline{M}_{11} = J_1 + J_2 + m_1 r_1^2 + m_2 l_1^2 + m_2 r_2^2 + 2m_2 l_1 r_2 \cos(q_2) + m_L l_1^2 + m_L l_2^2 + 2m_L l_1 l_2 \cos(q_2)$, $\overline{M}_{12} = J_2 + m_2 r_2^2 + m_2 l_1 r_2 \cos(q_2) + m_L l_2^2 + m_L l_1 l_2 \cos(q_2)$, and $\overline{M}_{22} = J_2 + m_2 r_2^2 + m_L l_2^2$.

The centripetal vector is

$$C(q, \dot{q}) = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}, \tag{34}$$

where $C_1 = (m_2 l_1 r_2 + m_L l_1 l_2)(2\dot{q}_1 - \dot{q}_2)\dot{q}_2 \sin(q_2)$ and $C_2 = (m_2 l_1 r_2 + m_L l_1 l_2)\dot{q}_1^2 \sin(q_2)$.

The friction vector and gravity matrix are

$$F(\dot{q}) = \begin{bmatrix} b_1 \dot{q}_1 \\ b_2 \dot{q}_2 \end{bmatrix}, G(q) = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}, \tag{35}$$

where $G_1 = (m_1 gr_1 + m_2 gr_1 + m_L gr_1)\sin(q_1) + (m_2 gr_2 + m_L gl_2)\sin(q_1 + q_2)$ and $G_2 = (m_2 gr_2 + m_L gl_2)\sin(q_1 + q_2)$.

Some model parameters are given as $m_1 = 9.387/\text{kg}$, $m_2 = 1.729/\text{kg}$, $l_1 = l_2 = 0.250/\text{m}$, $b_1 = b_2 = 0.14$, $r_1 = 0.053/\text{m}$, $r_2 = 0.170/\text{m}$, $J_1 = 0.1190/\text{kg.m}^2$, $J_2 =$
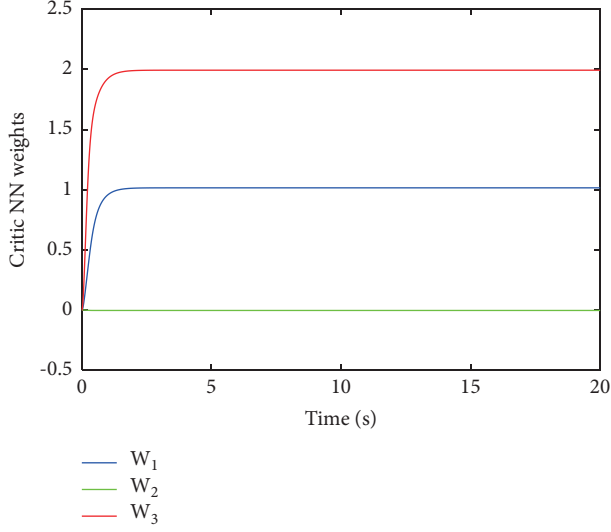
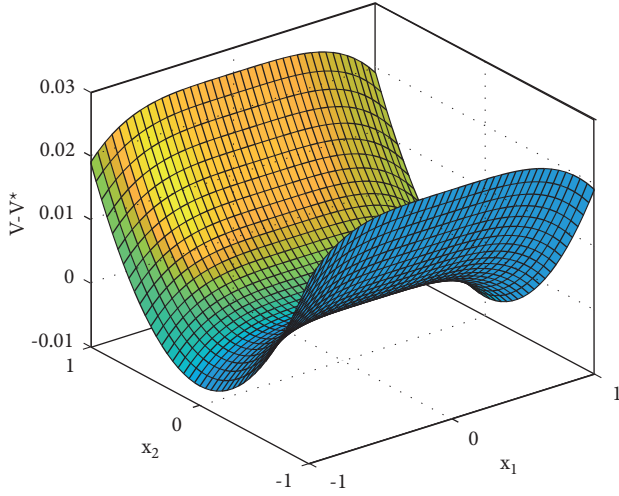FIGURE 2: Convergence of the critic NN weights $\widehat{W}$ [13].

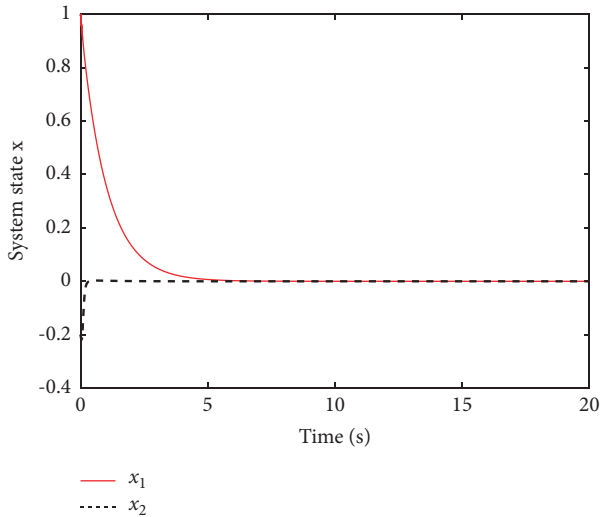FIGURE 3: Critic NN approximation error of optimal cost function [13].
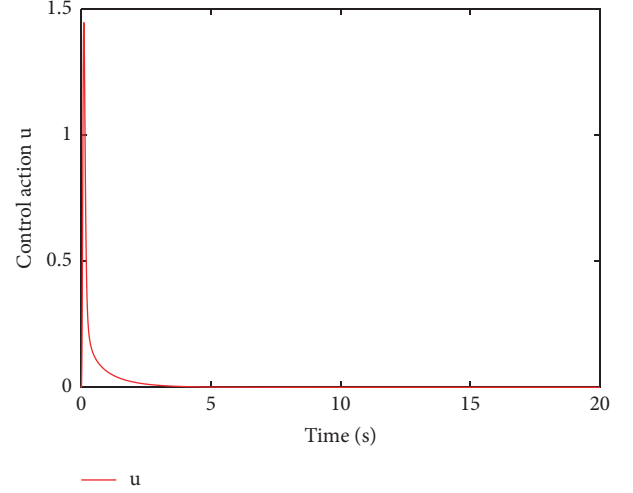

FIGURE 4: Profile of system state x.
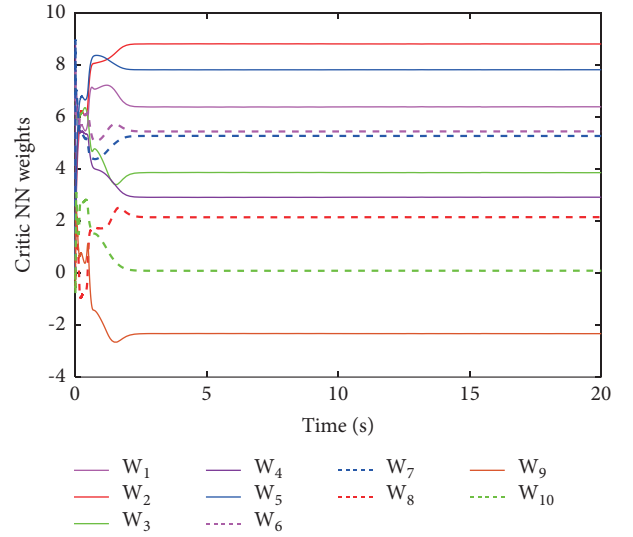

FIGURE 5: The system input for proposed control.

FIGURE 6: Convergence of the critic NN weights $\widehat{W}$.

$0.0652/kg.m^2$, $m_L = [0, 3]kg$, and $g = 9.8m/s$. With above system dynamics, the state equation of the system can be given as [18]

$$\dot{x} = f(x) + g(x)(u + b(x)u) + g(x)d(x), \qquad (36)$$

where $x = [x_1, x_2, x_3, x_4]^T = [q_1, q_2, q_3, q_4]^T$, $f(x) = \begin{bmatrix} 0 \\ x_2 \end{bmatrix}$, $g(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $b(x) = \overline{M}(x_1)^{-1}\overline{M}_0(x_1) - I \geq 0$, $\overline{M}_0$ is the value of $\overline{M}$ when the load is 0. $d(x) = \overline{M}(q)^{-1}(\overline{W}_0(q, \dot{q}) - \overline{W}(q, \dot{q}))$, and $\overline{W}_0$ is the value of $\overline{W}$ when the load 0.

In this simulation, we set the initial weight $\widehat{W}(0) = 0$; when the load $m_L = 3$ kg, we choose the leaning parameters $\ell = 21$ and $\Gamma = 0.1$ and weight matrices $M = I$ and $N = I$. The initial states are $q_1(0) = 15°$, $q_2(0) = 15°$, $\dot{q}_1(0) = 0$, and $\dot{q}_2(0) = 0$.

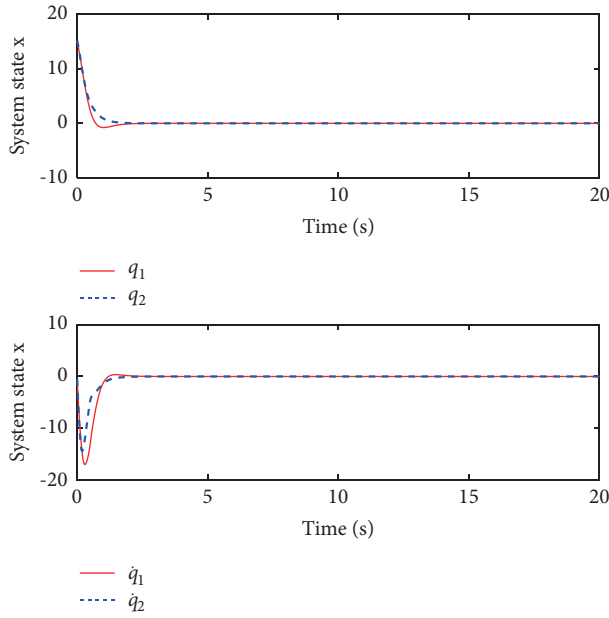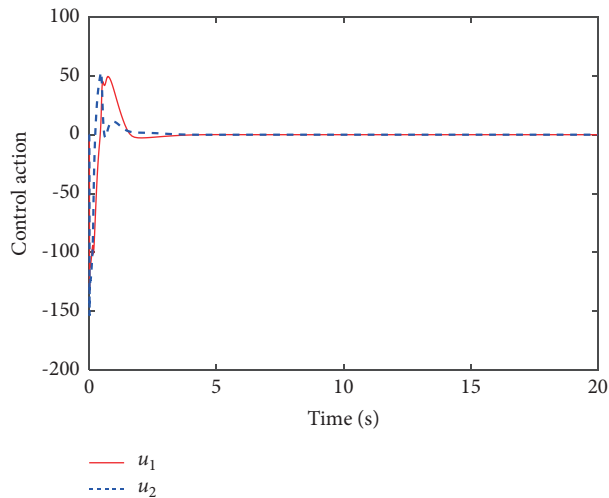FIGURE 7: Profile of system state x.

FIGURE 8: The system input for proposed control.

5. Conclusion

The purpose of this paper is to address the robust control problem of the uncertain systems via developing an adaptive critic learning method. To this end, the robust control problem of the uncertain systems is transformed into an optimal control problem of the nominal systems via selecting the cost function. Then, a single NN is used to reformulate the cost function, where the unknown cost function can be represented as a known term; then, an adaptive critic learning method based on the adaptive parameter estimation technique is presented to obtain the optimal cost function such that the optimal control problem can be solved. Simulations are given to show the effectiveness of the proposed leaning algorithm and control method. Future work will focus on the robust tracking control with uncertain systems.

Data Availability

Data were curated by the authors and are available upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] G. Saridis and F. Wang, "Suboptimal control of nonlinear stochastic systems," *Control Theory and Advanced Technology*, vol. 10, no. 4, pp. 847–871, 1994.

[2] R. W. Beard, G. N. Saridis, and J. T. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.

[3] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[4] P. J. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *General Systems Yearbook*, vol. 22, no. 6, pp. 25–38, 1977.

[5] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, Van Nostrand Reinhold, New York, NY, USA, 1992.

[6] D. Wang, D. Liu, Q. Wei, D. Zhao, and N. Jin, "Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming," *Automatica*, vol. 48, no. 8, pp. 1825–1832, 2012.

[7] B. Xu, C. Yang, and Z. Shi, "Reinforcement learning output-feedback NN control using deterministic learning technique," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 635–641, 2014.

[8] F. Wang, "Parallel control: a method for data-driven and computational control," *Acta Automatica Sinica*, vol. 39, no. 4, pp. 293–302, 2013.

[9] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: survey, classification and perspective," *Information Sciences*, vol. 235, pp. 3–35, 2013.

Figure 6 shows the estimated critic NN weights. It can be seen from Figure 6 that the estimated NN weight converges to certain value. This result verifies the convergence of Theorem 1 and the effectiveness of the proposed learning algorithm. Figure 7 shows the change of the controlled system state under the derived optimal control when the load condition is set $m_L = 3\,\mathrm{kg}$, which shows that the closed-loop system is asymptotically stable. The corresponding control input is shown in Figure 8. Although it jitters at first, it tends to be smooth when it stabilizes.

From above simulation results, we have that the proposed learning method and control technique are effective.

[10] F. Lin, R. D. Brandt, and J. Sun, "Robust control of nonlinear systems: compensating for uncertainty," *International Journal of Control*, vol. 56, no. 6, pp. 1453–1459, 1992.

[11] K. Zhou and J. Doyle, *Essentials of Robust Control*, Vol. 38, Prentice-Hall, , Hoboken, NJ, USA, 1998.

[12] D. Wang, D. Liu, C. Mu, and Y. Zhang, "Neural network learning and robust stabilization of nonlinear systems with dynamic uncertainties," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 1342–1351, 2018.

[13] J. Zhao, J. Na, and G. Gao, "Adaptive dynamic programming based robust control of nonlinear systems with unmatched uncertainties," *Neurocomputing*, vol. 395, pp. 56–65, 2020.

[14] X. Yang, H. He, and X. Zhong, "Adaptive dynamic programming for robust regulation and its application to power systems," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5722–5732, 2018.

[15] C. Yang, C. Chen, W. He, R. Cui, Z. Li, and M. Wang, "Biologically inspired motion modeling and neural control for robot learning from demonstrations," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 777–787, 2019.

[16] P. Gahinet, A. Nemirovskii, A. J. Laub, and M. Chilali, "The LMI control toolbox," in *Proceedings of the 1994 33rd IEEE Conference on Decision and Control*, vol. 3, pp. 2038–2041, Lake Buena Vista, FL, USA, December 1994.

[17] Y. Yang, K. G. Vamvoudakis, H. Modares, Y. Yin, and D. C. Wunsch, "Hamiltonian-driven hybrid adaptive dynamic programming," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 10, pp. 6423–6434, 2021.

[18] F. Lin, *Robust Control Design: An Optimal Control Approach*, John Wiley & Sons, Hoboken,, NJ, USA, 2007.

[19] J. Na, B. Wang, G. Li, S. Zhan, and W. He, "Nonlinear constrained optimal control of wave energy converters with adaptive dynamic programming," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 7904–7915, 2019.

[20] Y. Lv, J. Na, Q. Yang, X. Wu, and Y. Guo, "Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics," *International Journal of Control*, vol. 89, no. 1, pp. 99–112, 2016.

[21] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.

[22] F. Ding, X. Liu, and J. Chu, "Gradient-based and least-squares-based iterative algorithms for hammerstein systems using the hierarchical identification principle," *IET Control Theory & Applications*, vol. 7, no. 2, pp. 176–184, 2013.

[23] C. Zhang, J. Na, J. Wu, Q. Chen, and Y. Huang, "Proportional-integral approximation-free control of robotic systems with unknown dynamics," *IEEE*, vol. 26, no. 4, pp. 2226–2236, 2021.