Reinforcement-Learning-Based Robust Controller Design for Continuous-Time Uncertain Nonlinear Systems Subject to Input Constraints

Derong Liu, *Fellow, IEEE*, Xiong Yang, Ding Wang, *Member, IEEE*, and Qinglai Wei, *Member, IEEE*

***Abstract*—The design of stabilizing controller for uncertain nonlinear systems with control constraints is a challenging problem. The constrained-input coupled with the inability to identify accurately the uncertainties motivates the design of stabilizing controller based on reinforcement-learning (RL) methods. In this paper, a novel RL-based robust adaptive control algorithm is developed for a class of continuous-time uncertain nonlinear systems subject to input constraints. The robust control problem is converted to the constrained optimal control problem with appropriately selecting value functions for the nominal system. Distinct from typical action-critic dual networks employed in RL, only one critic neural network (NN) is constructed to derive the approximate optimal control. Meanwhile, unlike initial stabilizing control often indispensable in RL, there is no special requirement imposed on the initial control. By utilizing Lyapunov's direct method, the closed-loop optimal control system and the estimated weights of the critic NN are proved to be uniformly ultimately bounded. In addition, the derived approximate optimal control is verified to guarantee the uncertain nonlinear system to be stable in the sense of uniform ultimate boundedness. Two simulation examples are provided to illustrate the effectiveness and applicability of the present approach.**

***Index Terms*—Approximate dynamic programming (ADP), neural networks (NNs), neuro-dynamic programming, nonlinear systems, optimal control, reinforcement learning (RL), robust control.**

I. Introduction

THE DESIGN of stabilizing controller for constrained-input uncertain nonlinear systems has always been a challenging issue. During the past several decades, considerable efforts have been made to enhance the control performance of uncertain nonlinear systems with control constraints [1]–[4]. Various methods were developed and successfully applied to this type of robust control problems. Nevertheless, most of the proposed approaches focused on designing the direct adaptive controller and tackling control constraints by employing compensator schemes [1]–[3]. It is often a rather challenging task to construct compensator schemes and Lyapunov functions for guaranteeing the stability of this kind of nonlinear systems.

In order to overcome the above difficulty, in this paper, we transform the robust control problem to a class of optimal control problems by properly selecting value functions for the nominal system. The optimal control theory has made significant progress in the past half century. The important and valuable insights into the optimal control theory were presented in [5] and [6]. Up to now, optimal control problems for nonlinear systems have attracted extensive attentions. A core challenge of obtaining the solution of nonlinear optimal control problems is that it often falls to solve the Hamilton–Jacobi–Bellman (HJB) equation. Because the HJB equation is actually a nonlinear partial differential equation (PDE), it is usually intractable to solve by analytical methods. To cope with the problem, Bellman [7] developed dynamic programming (DP) theory. Though DP is successfully applied to solve optimal control problems, it is implemented backward-in-time which often makes the computation untenable to be run with increasing dimension of nonlinear systems. Consequently, approximate DP (ADP) algorithms were introduced by Werbos [8]. The ADP methods can give approximate solutions of the HJB equation forward-in-time by employing neural networks (NNs). After that, various ADP approaches were developed [9]–[20]. Nevertheless, most of ADP algorithms are either implemented offline by utilizing iterative schemes or they require *a priori* knowledge of system dynamics. Since the exact knowledge of nonlinear dynamic systems is often unavailable, these ADP algorithms are intractable to real-time control applications.

To address the above issues, reinforcement learning (RL) is introduced. RL is a class of methods employed in machine learning to revise the actions of an agent based on responses from its environment [21], [22]. A general structure used to implement RL algorithm is the actor-critic architecture, where the actor performs actions by interacting with its environment, and the critic evaluates actions and offers feedback information to the actor, leading to the improvement in performance of the subsequent actor [23]. RL differs significantly from typical ADP methods in that there is no prescribed behavior or training model proposed.

Manuscript received October 12, 2014; revised January 28, 2015; accepted February 13, 2015. Date of publication April 8, 2015; date of current version June 12, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61034002, Grant 61233001, Grant 61273140, Grant 61304086, and Grant 61374105, in part by the Beijing Natural Science Foundation under Grant 4132078, and in part by the Early Career Development Award of the State Key Laboratory of Management and Control for Complex Systems. This paper was recommended by Associate Editor D. Wang.

The authors are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: derong.liu@ia.ac.cn; xiong.yang@ia.ac.cn; ding.wang@ia.ac.cn; qinglai.wei@ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2015.2417170

Consequently, RL is often applied to adaptive optimal controller designs [24]–[37].

During the past several years, many researchers have paid their attentions to the applications of RL methods to constrained nonlinear optimal control problems [38]–[41]. Abu-Khalaf and Lewis [38] developed an offline RL-based algorithm to solve the HJB equation of optimal control of continuous-time (CT) nonlinear systems with saturating actuators. By using the algorithm, the actor and the critic were sequentially tuned and the solution of the HJB equation was successively approximated. After that, Modares *et al.* [39] proposed a novel algorithm based on integral RL methods to synchronously tune the critic and the actor. Thereafter, Modares and Lewis [40] applied the proposed algorithm to study the constrained-input optimal tracking control problems. It should be mentioned that the knowledge of internal dynamics is not required in [39] and [40] [that is, the knowledge of $f(x)$ presented in (2) is unknown]. Different from the algorithms proposed in [39] and [40], Yang *et al.* [41] employed identifier NNs to remove the requirement of the knowledge of internal dynamics and developed a new RL-based algorithm to obtain the optimal control for nonlinear systems with unknown structures. From [39]–[41], one shall find that the uncertainties of nonlinear systems can be conquered by using integral RL algorithms or introducing identifier NNs.

A question to be asked: based on the above two approaches, could the robust control for uncertain nonlinear systems be derived from the optimal control solution with appropriate value functions for the original uncertain nonlinear systems, rather than for the nominal system? In fact, the above two methods cannot be used for the former case. By using integral RL algorithms, the system state needs to be reset at each iteration step and it gives rise to difficulties for stability analysis [42]. On the other hand, the identifier NNs might not accurately obtain the information of the uncertainties, when the uncertainty terms contain noise or immeasurable perturbation. For these reasons, Adhyaru *et al.* [43] transformed the robust control problem to the constrained optimal control problem by selecting a suitable value function for the nominal system. The algorithm developed in [43] is constructed by utilizing the least squares method and performed offline. Meanwhile, the stability analysis of the closed-loop optimal control system is not addressed.

More recently, Jiang and Jiang [44] developed a robust ADP algorithm to derive the robust control for a class of uncertain nonlinear systems. Based on the algorithm in [44], the robust control is obtained by getting the optimal control solution with the infinite horizon cost for the original uncertain nonlinear systems, which is an advantage over the algorithm given in [43]. Nevertheless, similar to the algorithms presented in [38]–[41] and [43], the algorithm developed in [44] also requires the initial stabilizing control. To the best of our knowledge, there is no general method proposed to derive such a control law. From a mathematical point of view, the initial stabilizing control is actually a suboptimal control. The suboptimal control is intractable to obtain, since it is often impossible to solve the nonlinear PDEs analytically. Accordingly, the initial stabilizing control is a rather restrictive condition. Recently, Dierks and Jagannathan [45] provided a way to relax the requirement of initial stabilizing control under a single online approximator-based framework. However, the control constraints are not taken into consideration. In real engineering applications, ignoring the actuators' limitation may give rise to undesirable transient response, and cause system instability. In addition, the developed algorithm is not utilized to derive robust control for CT nonlinear systems with unknown perturbation.

Motivated by the aforementioned work, in this paper, a novel RL-based robust adaptive control algorithm is developed for constrained-input CT nonlinear systems in the presence of unknown perturbation. The robust control problem is transformed to a constrained optimal control problem with properly selecting value functions for the nominal system. Distinct from traditional action-critic dual networks employed in RL, only one critic NN is constructed to derive the approximate optimal control. Meanwhile, unlike initial stabilizing control often indispensable in RL, there is no special requirement imposed on the initial control. By using Lyapunov's direct method, the closed-loop optimal control system and the estimated weights of the critic NN are proved to be uniformly ultimately bounded (UUB). In addition, the derived approximate optimal control is verified to guarantee the uncertain nonlinear system to be stable in the sense of uniform ultimate boundedness.

The main contributions of this paper include the following.

1) To the best of authors' knowledge, it is the first time that, by using RL methods, a critic NN is constructed to derive the robust control of constrained-input uncertain nonlinear CT systems without the requirement of the initial stabilizing control.
2) Unlike [39] ignoring the higher-order terms of Taylor series in the stability analysis, in this paper, we take these terms into consideration. The higher-order terms often have a close connection with stability analysis (see Fact 1 in subsequent section). It will be more reasonable to take them into account for stability analysis.
3) In comparison with [38]–[41] and [44], a clear advantage of the developed algorithm in this paper is that a simpler algorithm architecture is constructed, that is, only one critic NN is employed. In this sense, the complexity of the computation is reduced.

The rest of this paper is organized as follows. In Section II, we present the problem statement and preliminaries. In Section III, we provide the nominal system for uncertain nonlinear systems and show that the robust control problem can be transformed to a constrained optimal control problem. In Section IV, we design an online RL-based control scheme to derive the approximate solution of the HJB equation. In Section V, we develop the stability analysis. In Section VI, two examples are given to illustrate the theoretical results. Finally, in Section VII, we give several concluding remarks.

A. Notations

$\mathbb{R}$ represents the set of all real numbers. $\mathbb{R}^m$ denotes the Euclidean space of all real m-vectors. $\mathbb{R}^{n\times m}$ denotes the space

of all $n \times m$ real matrices. I_m represents $m \times m$ identity matrix. T is the transposition symbol. Ω is a compact set of $\mathbb{R}^n$, $C^m(\Omega)$ represents the class of functions having continuous mth derivative on Ω. When ξ is a vector, $\|\xi\|$ denotes the Euclidean norm of ξ. When A is a matrix, $\|A\|$ denotes the two-norm of A.

II. Problem Statement and Preliminaries

Consider the uncertain nonlinear CT system described by

$$\dot{x}(t) = f(x(t)) + g(x(t))u(x(t)) + \Delta f(x(t)) \tag{1}$$

with the state $x(t) \in \Omega \subset \mathbb{R}^n$ and the control $u(x) \in \mathfrak{A}$, and $\mathfrak{A} = \{u | u \in \mathbb{R}^m, |u_i| \le \kappa, i = 1, \ldots, m\}$, where $\kappa > 0$ is the saturating bound. $f(x) \in \mathbb{R}^n$ and $g(x) \in \mathbb{R}^{n \times m}$ are known functions, and $\Delta f(x) \in \mathbb{R}^n$ is an unknown perturbation. For convenience of later analysis, we provide the following assumptions, which have been used in [30], [34], [46], and [47].

Assumption 1: The perturbation term $\Delta f(x)$ satisfies the matching condition. That is, $\Delta f(x) = g(x)d(x)$, where $d(x) \in \mathbb{R}^m$ is an unknown function bounded by a known function $d_M(x)$, i.e., $\|d(x)\| \le d_M(x)$. In addition, $d(0) = 0$ and $d_M(0) = 0$.

Assumption 2: $f(x) + g(x)u$ is Lipschitz continuous on the compact set Ω containing the origin, such that system (1) is stabilizable on Ω. Moreover, $f(0) = 0$.

Assumption 3: The control matrix $g(x)$ is known and bounded, i.e., there exist constants g_m and g_M $(0 < g_m < g_M)$ such that $g_m \le \|g(x)\| \le g_M$, for every $x \in \Omega$.

For system (1), in order to successfully tackle the robust control problem, one needs to derive a feedback control $u(x) \in \mathfrak{A}$, such that the closed-loop system is stable with the unknown term $d(x)$. It is generally difficult to directly design such a controller, for the control is constrained and the uncertainty term $d(x)$ is involved.

In this paper, we shall demonstrate that the robust control problem can be converted into the constrained optimal control problem with appropriately selecting value functions for the nominal system. Then, by solving the constrained optimal control problem, we can obtain a robust controller to guarantee system (1) to be stable in the sense of uniform ultimate boundedness (as for the definition of uniform ultimate boundedness, readers can refer to [48]).

III. Nominal Systems and Problem Transformation

This section consists of two parts. First, the HJB equation for the constrained nominal system is developed. Then, we verify that the robust control for system (1) can be obtained by the optimal control for the constrained nominal system.

A. HJB Equation for Constrained Nominal Systems

The nominal system [i.e., system (1) without uncertainty] is described by

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

with $u(x) \in \mathfrak{A} \subset \mathbb{R}^m$. It is desired to obtain the control policy $u(x)$ which minimizes the infinite horizon value function

$$V(x(t)) = \int_t^\infty \left[d_M^2\big(x(s)\big) + r\big(x(s), u(s)\big)\right] ds \quad (s \ge t) \tag{3}$$

where $r(x, u) = x^\mathsf{T} Q x + \mathcal{W}(u)$, Q is a symmetric positive definite matrix and $\mathcal{W}(u)$ is positive definite. In order to overcome bounded controls in system (2), inspired by the work of [38]–[40], we define $\mathcal{W}(u)$ as

$$\begin{aligned}\mathcal{W}(u) &= 2\kappa \int_0^u \left(\boldsymbol{\psi}^{-1}(\upsilon/\kappa)\right)^\mathsf{T} R d\upsilon \\ &= 2\kappa \sum_{i=1}^m \int_0^{u_i} \psi^{-1}(\upsilon_i/\kappa) r_i d\upsilon_i\end{aligned}$$

where $\boldsymbol{\psi}^{-1}(\upsilon/\kappa) = [\psi^{-1}(\upsilon_1/\kappa), \ldots, \psi^{-1}(\upsilon_m/\kappa)]^\mathsf{T}$, $R = \mathrm{diag}\{r_1, \ldots, r_m\}$ with $r_i > 0$, $i = 1, \ldots, m$, $\boldsymbol{\psi} \in \mathbb{R}^m$, $\boldsymbol{\psi}^{-\mathsf{T}}$ denotes $(\boldsymbol{\psi}^{-1})^\mathsf{T}$, and $\psi(\cdot)$ is a bounded one-to-one function satisfying $|\psi(\cdot)| \le 1$ and belonging to $C^p (p \ge 1)$ and $L_2(\Omega)$ note that $\psi(\upsilon) \in L_2(\Omega)$ means that $\left(\int_\Omega \psi^\mathsf{T}(\nu)\psi(\nu)d\nu\right)^{1/2} < \infty$ and $\int_\Omega \psi^\mathsf{T}(\nu)\psi(\nu)d\nu$ is the Lebesgue integral on Ω [48], [49]. Meanwhile, $\psi(\cdot)$ is a monotonic odd function with its derivative bounded by a constant ψ_M, i.e., $\|d\psi(\varsigma)/d\varsigma\| \le \psi_M$, $\forall \varsigma \in \mathbb{R}$. It should be emphasized that $\mathcal{W}(u)$ is positive definite since $\psi^{-1}(\cdot)$ is a monotonic odd function and R is positive definite. Without loss of generality, in this paper, we choose $\boldsymbol{\psi}(\cdot) = \tanh(\cdot)$ and $R = I_m$.

Let $\mathscr{A}(\Omega)$ be the set of admissible control [50]. Given a control $u(x) \in \mathscr{A}(\Omega)$, if the associated value function $V(x) \in C^1(\Omega)$, then the infinitesimal version of (3) is the so-called Lyapunov equation

$$V_x^\mathsf{T}\big(f(x) + g(x)u\big) + d_M^2(x) + r(x, u) = 0 \tag{4}$$

where $V_x \in \mathbb{R}^n$ denotes the partial derivative of $V(x)$ with respect to x, and $V(0) = 0$.

Define the Hamiltonian for the control $u(x) \in \mathscr{A}(\Omega)$ and the value function $V(x)$ as

$$H(x, V_x, u) = V_x^\mathsf{T}\big(f(x) + g(x)u\big) + d_M^2(x) + r(x, u). \tag{5}$$

The optimal value function $V^*(x) \in C^1(\Omega)$ is given as

$$V^*(x(t)) = \min_{u \in \mathscr{A}(\Omega)} \int_t^\infty \left[d_M^2\big(x(s)\big) + r\big(x(s), u(s)\big)\right] ds \tag{6}$$

with $V^*(0) = 0$. Then, the optimal cost $V^*(x)$ can be obtained by solving the HJB equation

$$\min_{u \in \mathscr{A}(\Omega)} H\big(x, V_x^*, u\big) = 0. \tag{7}$$

Suppose that the minimum value on the left-hand side of (7) exists and is unique. Then, the closed-form expression for the optimal control is derived as

$$u^*(x) = -\kappa \tanh\left(\frac{1}{2\kappa} g^\mathsf{T}(x) V_x^*\right). \tag{8}$$

Substituting (8) into (7), we obtain the HJB equation for the nonlinear systems as

$$V_x^{*\mathsf{T}} f(x) - 2\kappa^2 \mathcal{A}^{\mathsf{T}}(x)\tanh(\mathcal{A}(x)) + d_M^2(x) + x^{\mathsf{T}}Qx + 2\kappa \int_0^{-\kappa\tanh(\mathcal{A}(x))} \tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon = 0 \tag{9}$$

where $\mathcal{A}(x) = (1/2\kappa)g^{\mathsf{T}}(x)V_x^*$.

Denote $\mathcal{A}(x) = [\mathcal{A}_1(x), \ldots, \mathcal{A}_m(x)]^{\mathsf{T}} \in \mathbb{R}^m$ with $\mathcal{A}_i(x) \in \mathbb{R}$, $i = 1, \ldots, m$. By [39] and [40], we know

$$2\kappa \int_0^{-\kappa\tanh(\mathcal{A}(x))} \tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon = 2\kappa^2 \mathcal{A}^{\mathsf{T}}(x)\tanh(\mathcal{A}(x)) + \kappa^2 \sum_{i=1}^m \ln\Big[1 - \tanh^2(\mathcal{A}_i(x))\Big].$$

Then, the HJB equation (9) can be rewritten as

$$V_x^{*\mathsf{T}} f(x) + d_M^2(x) + x^{\mathsf{T}}Qx + \kappa^2 \sum_{i=1}^m \ln\Big[1 - \tanh^2(\mathcal{A}_i(x))\Big] = 0. \tag{10}$$

B. Problem Transformation

In this section, we establish a theorem to show that the robust control for system (1) can be obtained by the optimal control solution for system (2) with the value function (3).

Theorem 1: Consider the nominal system described by (2) with the value function (3). Let Assumptions 1–3 hold. Then, the optimal control $u^*(x)$ developed in (8) ensures system (1) to be stable in the sense of uniform ultimate boundedness.

Proof: Let $V^*(x)$ and $u^*(x)$ be the optimal value given in (6) and the optimal control derived in (8), respectively. According to the definition of $V^*(x)$ given in (6), we can obtain that $V^*(x) > 0$ for $\forall x \neq 0$ and $V^*(x) = 0 \Leftrightarrow x = 0$. Taking the derivative of $V^*(x)$ along the system trajectory $\dot{x} = f(x) + g(x)u^* + \Delta f(x)$, we have

$$\dot{V}^*(x) = V_x^{*\mathsf{T}}\big(f(x) + g(x)u^*\big) + V_x^{*\mathsf{T}}\Delta f(x). \tag{11}$$

Using (9), we obtain

$$V_x^{*\mathsf{T}}\big(f(x) + g(x)u^*\big) = -d_M^2(x) - x^{\mathsf{T}}Qx - 2\kappa \sum_{i=1}^m \int_0^{u_i^*} \tanh^{-1}(\upsilon_i/\kappa)d\upsilon_i \tag{12}$$

where $u^* = \big[u_1^*, \ldots, u_m^*\big]^{\mathsf{T}}$ with $u_i^* \in \mathbb{R}$, $i = 1, \ldots, m$. Observe that (8) yields $V_x^{*\mathsf{T}} g(x) = -2\kappa \tanh^{-\mathsf{T}}(u^*/\kappa)$. Then, based on Assumption 1, we get

$$V_x^{*\mathsf{T}}\Delta f(x) = -2\kappa \tanh^{-\mathsf{T}}(u^*/\kappa)d(x). \tag{13}$$

Substituting (12) and (13) into (11), we derive

$$\dot{V}^*(x) = -d_M^2(x) - x^{\mathsf{T}}Qx + \pounds_1(x) - 2\kappa \tanh^{-\mathsf{T}}\big(u^*/\kappa\big)d(x) \tag{14}$$

with $\pounds_1(x) = -2\kappa \sum_{i=1}^m \int_0^{u_i^*} \tanh^{-1}(\upsilon_i/\kappa)d\upsilon_i$.

Let $\tau_i = \tanh^{-1}(\upsilon_i/\kappa)$, $i = 1, \ldots, m$. Then, by applying variable substitution methods [51] to $\pounds_1(x)$, we have

$$\begin{aligned}\pounds_1(x) &= -2\kappa^2 \sum_{i=1}^m \int_0^{\tanh^{-1}(u_i^*/\kappa)} \tau_i\big(1 - \tanh^2(\tau_i)\big)d\tau_i \\ &= 2\kappa^2 \sum_{i=1}^m \int_0^{\tanh^{-1}(u_i^*/\kappa)} \tau_i \tanh^2(\tau_i)d\tau_i \\ &\quad - \kappa^2 \sum_{i=1}^m \big(\tanh^{-1}\big(u_i^*/\kappa\big)\big)^2.\end{aligned} \tag{15}$$

Note that

$$\sum_{i=1}^m \big(\tanh^{-1}(u_i^*/\kappa)\big)^2 = \tanh^{-\mathsf{T}}\big(u^*/\kappa\big)\tanh^{-1}\big(u^*/\kappa\big). \tag{16}$$

Then, by using (15) and (16), (14) can be represented as

$$\begin{aligned}\dot{V}^*(x) &= -d_M^2(x) - x^{\mathsf{T}}Qx + d^{\mathsf{T}}(x)d(x) \\ &\quad - \big[d(x) + \kappa\tanh^{-1}(u^*/\kappa)\big]^{\mathsf{T}} \\ &\quad \times \big[d(x) + \kappa\tanh^{-1}(u^*/\kappa)\big] + \pounds_2\end{aligned} \tag{17}$$

with $\pounds_2(x) = 2\kappa^2 \sum_{i=1}^m \int_0^{\tanh^{-1}(u_i^*/\kappa)} \tau_i \tanh^2(\tau_i)d\tau_i$.

Using the integral mean-value theorem [51], we have

$$\pounds_2(x) = 2\kappa^2 \sum_{i=1}^m \tanh^{-1}\big(u_i^*/\kappa\big)\theta_i \tanh^2(\theta_i) \tag{18}$$

where $\theta_i \in \mathbb{R}$ is selected between 0 and $\tanh^{-1}(u_i^*/\kappa)$, $i = 1, \ldots, m$. From the expression of $\pounds_2(x)$ given in (18), one can easily derive that $\pounds_2(x) > 0$.

Because u^* is an admissible control for nominal system (2) with the value function (3), by using the definition of admissible control [50], one can derive that $V^*(x)$ is finite for arbitrary $x \in \Omega$. Moreover, one can conclude that V_x^* is bounded. Without loss of generality, we denote that V_x^* is bounded by $\delta_M > 0$, i.e., $\|V_x^*\| \leq \delta_M$. Accordingly, by using Assumption 3 and (16), and observing that $0 < \tanh^2(\theta_i) \leq 1$, we obtain

$$\begin{aligned}\pounds_2(x) &\leq 2\kappa^2 \sum_{i=1}^m \tanh^{-1}(u_i^*/\kappa)\theta_i \\ &\leq 2\kappa^2 \tanh^{-\mathsf{T}}\big(u^*/\kappa\big)\tanh^{-1}\big(u^*/\kappa\big) \\ &= \frac{1}{2}V_x^{*\mathsf{T}} g(x)g^{\mathsf{T}}(x)V_x^* \leq \frac{1}{2}g_M^2\delta_M^2.\end{aligned} \tag{19}$$

Combining (17) with (19) and using Assumption 1, we derive

$$\begin{aligned}\dot{V}^*(x) &\leq -d_M^2(x) - x^{\mathsf{T}}Qx + d^{\mathsf{T}}(x)d(x) \\ &\quad - \big[d(x) + \kappa\tanh^{-1}\big(u^*/\kappa\big)\big]^{\mathsf{T}} \\ &\quad \times \big[d(x) + \kappa\tanh^{-1}\big(u^*/\kappa\big)\big] + \frac{1}{2}g_M^2\delta_M^2 \\ &\leq -\lambda_{\min}(Q)\|x\|^2 + \frac{1}{2}g_M^2\delta_M^2\end{aligned}$$

where $\lambda_{\min}(Q)$ denotes the minimum eigenvalue of the matrix Q. Noting that Q is positive definite, we obtain $\lambda_{\min}(Q) > 0$.

Consequently, $\dot{V}^*(x) < 0$ as long as the state $x(t)$ is out of the compact set

$$\Omega_x = \left\{ x\colon \|x\| \le \frac{g_M \delta_M}{\sqrt{2\lambda_{\min}(Q)}} \right\}.$$

This shows that $V^*(x)$ is a Lyapunov function for system (1) with the control u^*, whenever $x(t)$ lies outside the compact set Ω_x. Therefore, the optimal control u^* developed in (8) guarantees the trajectory of system (1) to be UUB. ■

Remark 1: The optimal value $V^*(x)$ is often considered to be a smooth function [31]–[41]. It implies that $V^*(x) \in C^1(\Omega)$. Therefore, by functional analysis [49], we obtain that V_x^* is bounded on Ω. This verifies that there exists a constant $\delta_M > 0$ such that $\|V_x^*\| \le \delta_M$. In addition, Ω should be selected large enough to make $\max\{\Omega_x, \Omega_{\tilde{x}}\} \subseteq \Omega$, where $\Omega_{\tilde{x}}$ is given in subsequent (60). In this sense, x will remain in Ω.

According to Theorem 1, the robust control for system (1) can be obtained by solving the optimal control problem (2) and (3). In other words, we need to get the solution of the HJB equation (10). Nevertheless, one shall find that (10) is actually a nonlinear PDE with respect to $V^*(x)$, which is difficult to solve by analytical methods. To overcome the difficulty, an online RL-based optimal control scheme shall be developed.

IV. RL-Based Optimal Control Scheme

Two subsections are embodied in this section, including the introduction of policy iteration algorithm and the design of online NN-based optimal control scheme.

A. Policy Iteration Algorithm

Step 1: Select a computation accuracy $\epsilon > 0$. Let $j = 0$ with $V^{(0)}(x) = 0$. Then, begin with an initial admissible control policy $u^{(0)}(x)$.

Step 2: Get the value $V^{(j+1)}(x)$ by solving the equation

$$\left(V_x^{(j+1)}\right)^{\mathsf{T}}\left(f(x) + g(x)u^{(j)}\right) + d_M^2(x) + x^{\mathsf{T}}Qx - 2\kappa \int_0^{u^{(j)}} \tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon = 0.$$

Step 3: Update the control policy using

$$u^{(j+1)}(x) = -\kappa \tanh\left(\frac{1}{2\kappa} g^{\mathsf{T}}(x) V_x^{(j+1)}\right).$$

Step 4: If $\|V^{(j+1)}(x) - V^{(j)}(x)\| \le \epsilon$ for every $x \in \Omega$, then stop and derive the approximate optimal control; otherwise, let $j = j + 1$ and go back to step 2.

Based on the present algorithm, one can obtain that, for $i \to \infty$, there exist $V^{(i)}(x) \to V^*(x)$ and $u^{(i)}(x) \to u^*(x)$. The convergence of the algorithm was shown in [38].

Though the present policy iteration algorithm can be applied to solve (10), it is often implemented offline (see [38]). On the other hand, it needs the initial admissible control. As mentioned before, it is a rather restrictive condition. To handle the above two problems, a novel online NN-based control algorithm shall be developed to solve (10).

B. Online NN-Based Control Design

In this section, a critic NN is constructed to approximate the value function. According to the universal approximation property of NNs, $V^*(x)$ given in (6) can be represented by a single-layer NN on a compact set Ω as

$$V^*(x) = W_c^{\mathsf{T}}\sigma(x) + \varepsilon(x) \tag{20}$$

where $W_c \in \mathbb{R}^{N_0}$ is the ideal NN weight vector, $\sigma(x) = [\sigma_1(x), \sigma_2(x), \ldots, \sigma_{N_0}(x)]^{\mathsf{T}} \in \mathbb{R}^{N_0}$ is the activation function with $\sigma_j(x) \in C^1(\Omega)$ and $\sigma_j(0) = 0$, the set $\{\sigma_j(x)\}_1^{N_0}$ is often selected to be linearly independent, N_0 is the number of the neurons, and $\varepsilon(x)$ is the NN function reconstruction error.

The derivative of $V^*(x)$ with respect to x is given as

$$V_x^* = \nabla\sigma^{\mathsf{T}}(x)W_c + \nabla\varepsilon \tag{21}$$

with $\nabla\sigma(x) = \partial\sigma(x)/\partial x$ and $\nabla\sigma(0) = 0$.

Substituting (21) into (10), we have

$$d_M^2(x) + x^{\mathsf{T}}Qx + W_c^{\mathsf{T}}\nabla\sigma f(x) + \nabla\varepsilon^{\mathsf{T}} f(x) + \kappa^2 \sum_{i=1}^{m} \ln\left[1 - \tanh^2(\Phi_{1i}(x) + \Psi_i(x))\right] = 0 \tag{22}$$

where $\Phi_1(x) = (1/2\kappa)g^{\mathsf{T}}(x)\nabla\sigma^{\mathsf{T}}W_c$, $\Psi(x) = (1/2\kappa)g^{\mathsf{T}}(x)\nabla\varepsilon$, and $\Phi_1(x) = [\Phi_{11}(x), \ldots, \Phi_{1m}(x)]^{\mathsf{T}}$ with $\Phi_{1i}(x) \in \mathbb{R}$, and $\Psi(x) = [\Psi_1(x), \ldots, \Psi_m(x)]^{\mathsf{T}}$ with $\Psi_i(x) \in \mathbb{R}$, $i = 1, \ldots, m$.

Using the mean-value theorem [51], (22) is represented as

$$d_M^2(x) + x^{\mathsf{T}}Qx + W_c^{\mathsf{T}}\nabla\sigma f(x) + \kappa^2 \sum_{i=1}^{m} \ln\left[1 - \tanh^2\left(\Phi_{1i}(x)\right)\right] + \varepsilon_{\mathrm{HJB}} = 0 \tag{23}$$

where $\varepsilon_{\mathrm{HJB}}$ is the HJB approximation error [38], [39], and the expression is given as

$$\varepsilon_{\mathrm{HJB}} = \nabla\varepsilon^{\mathsf{T}} f(x) + \sum_{i=1}^{m} \frac{2\kappa^2}{\zeta_{1i}} \tanh(\zeta_{2i})\left(\tanh^2(\zeta_{2i}) - 1\right)\Psi_i(x)$$

with $\zeta_{1i} \in \mathbb{R}$ selected between $1 - \tanh^2(\mathcal{A}_i(x)))$ and $1 - \tanh^2(\Phi_{1i}(x))$, and $\zeta_{2i} \in \mathbb{R}$ chosen between $\mathcal{A}_i(x)$ and $\Phi_{1i}(x)$.

Remark 2: It was shown in [38] that $\varepsilon_{\mathrm{HJB}}$ converges to zero as the number of neurons N_0 increases. In other words, for $\forall\varepsilon_h > 0$, there exists a positive N_h (depending only on ε_h) such that $N_0 > N_h$ implies $\|\varepsilon_{\mathrm{HJB}}\| \le \varepsilon_h$.

Similarly, by using (21) and the mean-value theorem, the optimal control (8) can be rewritten as

$$u^*(x) = -\kappa \tanh\left(\Phi_1(x)\right) + \varepsilon_{u^*} \tag{24}$$

where $\varepsilon_{u^*} = -1/2\left(\underline{\mathbf{1}} - \tanh^2(\xi)\right)g^{\mathsf{T}}\nabla\varepsilon$ with $\xi \in \mathbb{R}^m$ selected between $\Phi_1(x)$ and $\mathcal{A}(x)$ and $\underline{\mathbf{1}} = [1, \ldots, 1]^{\mathsf{T}} \in \mathbb{R}^m$.

Because the ideal NN weight W_c is typically unknown, (24) cannot be implemented in real-control process. Hence, we use a critic NN to approximate the value function given in (6) as

$$\hat{V}(x) = \hat{W}_c^{\mathsf{T}}\sigma(x) \tag{25}$$

where $\hat{W}_c$ is the estimate of W_c. Meanwhile, the estimation error for the weight is defined as

$$\tilde{W}_c = W_c - \hat{W}_c. \tag{26}$$

Utilizing (25), the estimate of (8) is derived as

$$\hat{u}(x) = -\kappa \tanh\Big(\frac{1}{2\kappa} g^{\mathsf{T}}(x)\nabla\sigma^{\mathsf{T}}\hat{W}_c\Big). \tag{27}$$

Combining (5), (25), and (27), we derive the approximate Hamiltonian as

$$H(x,\hat{W}_c) = d_M^2(x) + x^{\mathsf{T}}Qx + \hat{W}_c^{\mathsf{T}}\nabla\sigma f(x) + \kappa^2\sum_{i=1}^{m}\ln\big[1-\tanh^2(\Phi_{2i}(x))\big] \triangleq e \tag{28}$$

where $\Phi_2(x) = (1/2\kappa)g^{\mathsf{T}}(x)\nabla\sigma^{\mathsf{T}}\hat{W}_c$, and $\Phi_2(x) = [\Phi_{21}(x),\ldots,\Phi_{2m}(x)]^{\mathsf{T}}$ with $\Phi_{2i}(x)\in\mathbb{R}$, $i=1,\ldots,m$.

From (23) and (28), we have

$$e = -\tilde{W}_c^{\mathsf{T}}\nabla\sigma f(x) + \sum_{i=1}^{m}\kappa^2\big[\Gamma(\Phi_{2i}) - \Gamma(\Phi_{1i})\big] - \varepsilon_{\mathrm{HJB}} \tag{29}$$

with $\Gamma(\Phi_{\iota i}) = \ln\big[1-\tanh^2(\Phi_{\iota i}(x))\big]$, $\iota = 1, 2$. Observe that, for $\forall\Phi_{\iota i}(x)\in\mathbb{R}$, $\Gamma(\Phi_{\iota i})$ can be represented as [41]

$$\Gamma(\Phi_{\iota i}) = -2\ln\big[1+\exp\big(-2\Phi_{\iota i}(x)\mathrm{sgn}(\Phi_{\iota i}(x))\big)\big] - 2\Phi_{\iota i}(x)\mathrm{sgn}(\Phi_{\iota i}(x)) + \ln 4$$

where $\mathrm{sgn}(\Phi_{\iota i}(x))\in\mathbb{R}$ is a sign function [51]. Note that

$$\sum_{i=1}^{m}\Gamma(\Phi_{\iota i}) = -2\sum_{i=1}^{m}\ln\big[1+\exp\big(-2\Phi_{\iota i}(x)\mathrm{sgn}(\Phi_{\iota i}(x))\big)\big] - 2\Phi_{\iota}^{\mathsf{T}}(x)\mathrm{sgn}(\Phi_{\iota}(x)) + m\ln 4. \tag{30}$$

Therefore, using (29) and (30), we get

$$\begin{aligned} e =& \ 2\kappa^2\Big[\Phi_1^{\mathsf{T}}(x)\mathrm{sgn}(\Phi_1(x)) - \Phi_2^{\mathsf{T}}(x)\mathrm{sgn}(\Phi_2(x))\Big] - \tilde{W}_c^{\mathsf{T}}\nabla\sigma f(x) + \kappa^2\Delta_\Phi - \varepsilon_{\mathrm{HJB}} \\ =& \ \kappa\Big[W_c^{\mathsf{T}}\nabla\sigma g(x)\mathrm{sgn}(\Phi_1(x)) - \hat{W}_c^{\mathsf{T}}\nabla\sigma g(x)\mathrm{sgn}(\Phi_2(x))\Big] - \tilde{W}_c^{\mathsf{T}}\nabla\sigma f(x) + \kappa^2\Delta_\Phi - \varepsilon_{\mathrm{HJB}} \\ =& -\tilde{W}_c^{\mathsf{T}}\big[\nabla\sigma f(x) - \kappa\nabla\sigma g(x)\mathrm{sgn}(\Phi_2(x))\big] + \rho(x) \end{aligned} \tag{31}$$

where

$$\begin{aligned} \Delta_\Phi =& \ 2\sum_{i=1}^{m}\ln\frac{1+\exp\big[-2\Phi_{1i}(x)\mathrm{sgn}(\Phi_{1i}(x))\big]}{1+\exp\big[-2\Phi_{2i}(x)\mathrm{sgn}(\Phi_{2i}(x))\big]} \\ \rho(x) =& \ \kappa W_c^{\mathsf{T}}\nabla\sigma g(x)\big[\mathrm{sgn}(\Phi_1(x)) - \mathrm{sgn}(\Phi_2(x))\big] + \kappa^2\Delta_\Phi - \varepsilon_{\mathrm{HJB}}. \end{aligned}$$

To derive the minimum value of e, it is desired to choose $\hat{W}_c$ to minimize the squared residual error $E = (1/2)e^{\mathsf{T}}e$. By utilizing the gradient descent algorithm, the weight tuning law for the critic NN is generally given as [31]–[34], [40], and [41]

$$\dot{\hat{W}}_c = -\frac{\gamma}{(1+\phi^{\mathsf{T}}\phi)^2}\frac{\partial E}{\partial\hat{W}_c} = -\frac{\gamma\phi}{(1+\phi^{\mathsf{T}}\phi)^2}e \tag{32}$$

where $\phi = \nabla\sigma\big(f(x)+g(x)\hat{u}\big)$, $\gamma>0$ is a design constant, and the term $(1+\phi^{\mathsf{T}}\phi)^2$ is employed for normalization.

However, there exist two issues about the tuning rule (32).

1) Based on (32), the initial admissible control for systems (2) and (3) is required, for guaranteeing the validity of policy iteration algorithms presented in aforementioned literature. As stated before, the initial admissible control is actually a suboptimal control of system (2) with (3). The suboptimal control is intractable to obtain because it needs to get the analytical solution of the PDE (10).
2) By utilizing (32), if the initial control is not admissible, then tuning the critic NN alone might not guarantee the stability of the closed-loop system during the learning process of NNs.

To tackle the above two issues, the weight update law for the critic NN should be redefined. Prior to proceeding, we provide a assumption as follows. The assumption is a common technique, which has been used in [15], [25], [45], and [52].

Assumption 4: $J(x(t))$ is a continuously differentiable radially unbounded Lyapunov function candidate such that $\dot{J}(x(t)) = J_x^{\mathsf{T}}\big(f(x)+g(x)u^*\big) < 0$ with J_x the partial derivative of $J(x)$ with respect to x. Moreover, there exists a symmetric positive definite matrix $\mathcal{B}(x)\in\mathbb{R}^{n\times n}$ defined on Ω such that

$$J_x^{\mathsf{T}}\big(f(x)+g(x)u^*\big) = -J_x^{\mathsf{T}}\mathcal{B}(x)J_x. \tag{33}$$

Remark 3: $f(x)+g(x)u^*$ is often assumed to be bounded by a positive constant on the compact set Ω [30]–[34]. That is, for every $x\in\Omega$, there exists a constant $\varrho>0$ such that $\|f(x)+g(x)u^*\|\le\varrho$. To relax the condition, we assume that $f(x)+g(x)u^*$ is bounded by a function with respect to x. Because J_x is a function with respect to x, without loss of generality, we assume that $\|f(x)+g(x)u^*\|\le\eta\|J_x\|(\eta>0)$. In this sense, we derive that $\|J_x^{\mathsf{T}}\big(f(x)+g(x)u^*\big)\|\le\eta\|J_x\|^2$. Observing that $J_x^{\mathsf{T}}\big(f(x)+g(x)u^*\big)<0$, one shall find that (33) defined as in Assumption 4 is reasonable. In addition, $J(x(t))$ is usually derived through properly selecting functions, such as polynomials.

Based on Assumption 4 and aforementioned analyzes, we develop a novel weight update law for the critic NN as

$$\begin{aligned} \dot{\hat{W}}_c =& -\gamma\bar{\phi}\Big(d_M^2(x) + x^{\mathsf{T}}Qx + \hat{W}_c^{\mathsf{T}}\nabla\sigma f(x) + \kappa^2\sum_{i=1}^{m}\ln\Big[1-\tanh^2(\Phi_{2i}(x))\Big]\Big) \\ &+ \frac{\gamma}{2}\Pi(x,\hat{u})\nabla\sigma g(x)\big[I_m - \mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)J_x \\ &+ \gamma\Big(\kappa\nabla\sigma g(x)\big[\tanh(\Phi_2(x)) - \mathrm{sgn}(\Phi_2(x))\big]\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c \\ &- \big(P_2 - P_1\varphi^{\mathsf{T}}\big)\hat{W}_c\Big) \end{aligned} \tag{34}$$

where $\bar{\phi} = \phi/m_s^2$, $\varphi = \phi/m_s$, $m_s = 1+\phi^{\mathsf{T}}\phi$, and $\mathscr{B}(\Phi_2(x)) = \mathrm{diag}\big\{\tanh^2(\Phi_{2i}(x))\big\}$, $i=1,\ldots,m$, J_x is defined as in Assumption 4, P_1 and P_2 are tuning parameters with suitable dimensions, and $\Pi(x,\hat{u})$ is a sign function given as

$$\Pi(x,\hat{u}) = \begin{cases} 0, & \text{if } J_x^{\mathsf{T}}\big(f(x)\big)+g(x)\hat{u}\big) < 0 \\ 1, & \text{otherwise.} \end{cases} \tag{35}$$

Remark 4: Several notes about (34) are listed as follows.

1) The first term given in (34) shares the same feature with (32), which is employed to minimize the objective function $E = (1/2)e^{\mathsf{T}}e$.
2) The second term provided in (34) is used to guarantee the stability of the closed-loop system during the NN learning process. We denote the derivative of the Lyapunov function candidate for system (2) with the control (27) as

$$\Theta = J_x^{\mathsf{T}}\Big(f(x) - \kappa g(x)\tanh(\Phi_2(x))\Big).$$

If the closed-loop system is unstable, then we can obtain $\Theta > 0$. In order to keep the closed-loop system stable, we just need to make $\Theta < 0$. Using the gradient descent method, we have

$$\begin{aligned} -\gamma\frac{\partial\Theta}{\partial\hat{W}_c} &= -\gamma\frac{\partial\big[J_x^{\mathsf{T}}\big(f(x) - \kappa g(x)\tanh(\Phi_2(x))\big)\big]}{\partial\hat{W}_c} \\ &= \gamma\left(\frac{\partial\Phi_2}{\partial\hat{W}_c}\right)^{\mathsf{T}}\cdot\frac{\partial\big[\kappa J_x^{\mathsf{T}}g(x)\tanh(\Phi_2(x))\big]}{\partial\Phi_2(x)} \\ &= \frac{\gamma}{2}\nabla\sigma g(x)\big[I_m - \mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)J_x \quad (36) \end{aligned}$$

where $\mathscr{B}(\Phi_2(x)) = \mathrm{diag}\big\{\tanh^2(\Phi_{2i}(x))\big\}$, $i = 1,\ldots,m$. Equation (36) indicates the reason that we employ the second term given in (34). Actually, by the definition of $\Pi(x,\hat{u})$ given in (35), we find that if there exists $\Theta < 0$ (that is, the closed-loop system is stable), then $\Pi(x,\hat{u}) = 0$ and the second term in (34) disappears. If the closed-loop system is unstable, then $\Pi(x,\hat{u}) = 1$ and the second term in (34) [that is (36)] works. Based on (34), it makes no requirement of the initial stabilizing control for system (2). This property will be illustrated in numerical simulation.
3) The last term given in (34) is a robust term, which is used for stability analysis in the subsequent discussion.
4) If selecting proper P_i $(i = 1, 2)$ such that $P_2 = P_1\varphi^{\mathsf{T}}$, then, by (34), we have $\dot{\hat{W}}_c = 0$ when $x = 0$. In this case, $\hat{V}(x)$ will no longer be updated. However, the optimal control might not be achieved at finite time t_f which makes $x(t_f) = 0$. To avoid this case, persistency of excitation (PE) condition is required.

Observing the expression of ϕ given in (32) and using (27), we obtain that $\nabla\sigma f(x) = \phi + \kappa\nabla\sigma g(x)\tanh(\Phi_2(x))$. Then, based on (26), (28), (31), and (34), we derive

$$\begin{aligned} \dot{\tilde{W}}_c = {} & \gamma\frac{\varphi}{m_s}\Big[-\tilde{W}_c^{\mathsf{T}}\phi + \kappa\tilde{W}_c^{\mathsf{T}}\nabla\sigma g(x)\mathcal{F}(x) + \rho(x)\Big] \\ & - \frac{\gamma}{2}\Pi(x,\hat{u})\nabla\sigma g(x)\big[I_m - \mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)J_x \\ & + \gamma\Big[\kappa\nabla\sigma g(x)\mathcal{F}(x)\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c + \Big(P_2 - P_1\varphi^{\mathsf{T}}\Big)\hat{W}_c\Big] \quad (37) \end{aligned}$$

where $\mathcal{F}(x) = \mathrm{sgn}(\Phi_2(x)) - \tanh(\Phi_2(x))$.

Traditionally, for utilizing RL approaches, a second NN called the action NN is introduced to approximate the control policy [23], [31]–[33], [39], [40]. However, in this paper, the action NN is not required. The reasons are as follows.

1) As pointed out in [23], the action NN is mainly employed to avoid the need for knowledge of the internal dynamics $f(x)$. Nevertheless, both the knowledge of $f(x)$ and $g(x)$ is assumed to be available in our case.
2) From (25) and (27), we can find that the value function shares the same weight $\hat{W}_c$ with the control policy. Therefore, if the value function can be approximated by the critic NN given in (25), then the control policy is obtained via (27). In other words, the action NN can be replaced with (27).

Based on the above analyzes, the schematic of the developed control algorithm is shown in Fig. 1.

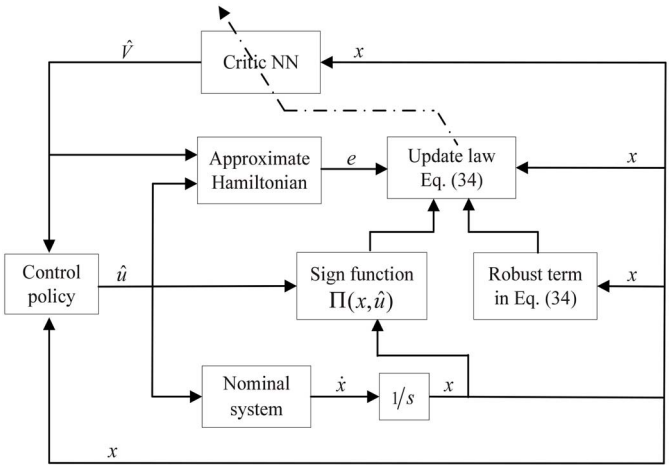


Fig. 1. Schematic of the developed control algorithm.

V. Stability Analysis

In this section, we present our main results via Lyapunov's direct method. Before proceeding, we provide the following two assumptions, which have been used in [30]–[34].

Assumption 5: The ideal NN weight W_c is bounded by a known constant $W_M > 0$, i.e., $\|W_c\| \le W_M$. There exist known constants $b_\varepsilon > 0$ and $b_{\varepsilon x} > 0$ such that $\|\varepsilon(x)\| < b_\varepsilon$, $\|\nabla\varepsilon(x)\| < b_{\varepsilon x}$, for every $x \in \Omega$. In addition, ε_{u^*} given in (24) is bounded by a known constant $b_{\varepsilon_{u^*}} > 0$ over Ω, i.e., $\|\varepsilon_{u^*}\| \le b_{\varepsilon_{u^*}}$, for every $x \in \Omega$.

Assumption 6: There exist known constants $b_\sigma > 0$ and $b_{\sigma x} > 0$ such that $\|\sigma(x)\| \le b_\sigma$, $\|\nabla\sigma(x)\| \le b_{\sigma x}$, for every $x \in \Omega$.

Let $G(\Phi_i) = \tanh(\Phi_i(x))$, $i = 1, 2$. By employing Taylor series, we have

$$\begin{aligned} G(\Phi_1) - G(\Phi_2) = {} & \frac{\partial G(\Phi_2)}{\partial\Phi_2}\big(\Phi_1(x) - \Phi_2(x)\big) \\ & + O\big((\Phi_1(x) - \Phi_2(x))^2\big) \\ = {} & \frac{1}{2\kappa}\big[I_m - \mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}\nabla\sigma^{\mathsf{T}}\tilde{W}_c \\ & + O\big((\Phi_1(x) - \Phi_2(x))^2\big) \quad (38) \end{aligned}$$

where $\mathscr{B}(\Phi_2(x)) = \mathrm{diag}\big\{\tanh^2(\Phi_{2i}(x))\big\}$, $i = 1,\ldots,m$ and $O\big((\Phi_1(x) - \Phi_2(x))^2\big)$ is the higher-order terms of

the Taylor series [51]. Then, by using (38), we derive

$$O\Big((\Phi_1(x)-\Phi_2(x))^2\Big) = G(\Phi_1)-G(\Phi_2)+\frac{1}{2\kappa}[\mathscr{B}(\Phi_2(x))-I_m]g^{\mathsf{T}}\nabla\sigma^{\mathsf{T}}\tilde{W}_c.$$

Observing that $\|\tanh(\Phi_i)\| \le 1$ $(i=1,2)$, we can obtain a fact as follows. It should be noted that a similar fact has been stated in [53].

Fact 1: For hyperbolic function tanh, the higher-order term in the Taylor series is bounded by

$$\left\|O\Big((\Phi_1(x)-\Phi_2(x))^2\Big)\right\| \le c_1+c_2\|\tilde{W}_c\|$$

where c_i $(i=1,2)$ are computable positive constants.

Theorem 2: Consider the nominal nonlinear CT system described by (2) with the associated HJB equation (10). Let Assumptions 2–6 hold and take the control input for system (2) as given in (27). Meanwhile, let the critic NN weight tuning law be described by (34). Then, the function J_x and the critic NN weight estimation error $\tilde{W}_c$ are guaranteed to be UUB.

Proof: Consider the Lyapunov function candidate

$$L(t)=L_1(x)+\frac{1}{2}\tilde{W}_c^{\mathsf{T}}\gamma^{-1}\tilde{W}_c \tag{39}$$

where $L_1(x)=J(x)$ with $J(x)$ given in Assumption 4.

Taking the time derivative of (39), we have

$$\begin{aligned}\dot{L}(t) &= J_x^{\mathsf{T}}\big(f(x)+g(x)\hat{u}\big)+\dot{\hat{W}}_c^{\mathsf{T}}\gamma^{-1}\tilde{W}_c\\ &= J_x^{\mathsf{T}}\big[f(x)-\kappa g(x)\tanh(\Phi_2(x))\big]\\ &\quad+\dot{\hat{W}}_c^{\mathsf{T}}\gamma^{-1}\tilde{W}_c.\end{aligned} \tag{40}$$

Using (37), the last term of (40) can be represented as

$$\begin{aligned}\dot{\hat{W}}_c^{\mathsf{T}}\gamma^{-1}\tilde{W}_c =& \Big[-\tilde{W}_c^{\mathsf{T}}\phi+\kappa\tilde{W}_c^{\mathsf{T}}\nabla\sigma g\mathcal{F}(x)+\rho(x)\Big]\frac{\varphi^{\mathsf{T}}}{m_s}\tilde{W}_c\\ &-\frac{1}{2}\Pi\big(x,\hat{u}\big)J_x^{\mathsf{T}}g(x)\big[I_m-\mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)\\ &\times\nabla\sigma^{\mathsf{T}}\tilde{W}_c+\kappa\tilde{W}_c^{\mathsf{T}}\nabla\sigma g(x)\mathcal{F}(x)\frac{\varphi^{\mathsf{T}}}{m_s}\hat{W}_c\\ &+\tilde{W}_c^{\mathsf{T}}\big(P_2\hat{W}_c-P_1\varphi^{\mathsf{T}}\hat{W}_c\big)\\ =&-\tilde{W}_c^{\mathsf{T}}\varphi\varphi^{\mathsf{T}}\tilde{W}_c+\alpha(x)\varphi^{\mathsf{T}}\tilde{W}_c+\tilde{W}_c^{\mathsf{T}}\beta(x)\\ &-\frac{1}{2}\Pi\big(x,\hat{u}\big)J_x^{\mathsf{T}}g(x)\big[I_m-\mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)\\ &\times\nabla\sigma^{\mathsf{T}}\tilde{W}_c+\tilde{W}_c^{\mathsf{T}}\big(P_2\hat{W}_c-P_1\varphi^{\mathsf{T}}\hat{W}_c\big)\end{aligned} \tag{41}$$

where $\alpha(x)=\rho(x)/m_s$ and $\beta(x)=\kappa\nabla\sigma g(x)\mathcal{F}(x)\ (\varphi^{\mathsf{T}}/m_s)W_c$.

By the definition of $\tilde{W}_c$ given in (26), we derive the last term in (41) as

$$\begin{aligned}\tilde{W}_c^{\mathsf{T}}\big(P_2\hat{W}_c-P_1\varphi^{\mathsf{T}}\hat{W}_c\big) =& \tilde{W}_c^{\mathsf{T}}P_2W_c-\tilde{W}_c^{\mathsf{T}}P_2\tilde{W}_c\\ &-\tilde{W}_c^{\mathsf{T}}P_1\varphi^{\mathsf{T}}W_c+\tilde{W}_c^{\mathsf{T}}P_1\varphi^{\mathsf{T}}\tilde{W}_c.\end{aligned}$$

Let $\mathcal{Y}^{\mathsf{T}}=\big[\tilde{W}_c^{\mathsf{T}}\varphi\quad\tilde{W}_c^{\mathsf{T}}\big]$. Then, (41) can be rewritten as

$$\begin{aligned}\dot{\hat{W}}_c^{\mathsf{T}}\gamma^{-1}\tilde{W}_c =& -\mathcal{Y}^{\mathsf{T}}M\mathcal{Y}+\mathcal{Y}^{\mathsf{T}}N-\frac{1}{2}\Pi(x,\hat{u})J_x^{\mathsf{T}}g(x)\\ &\times\big[I_m-\mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)\nabla\sigma^{\mathsf{T}}\tilde{W}_c\end{aligned} \tag{42}$$

where

$$M=\begin{bmatrix} I & -\frac{1}{2}P_1^{\mathsf{T}}\\ -\frac{1}{2}P_1 & P_2\end{bmatrix},\quad N=\begin{bmatrix}\alpha(x)\\ \beta(x)+P_2W_c-P_1\varphi^{\mathsf{T}}W_c\end{bmatrix}.$$

Substituting (42) into (40) and choosing P_i $(i=1,2)$ such that the matrix M is positive definite, we have

$$\begin{aligned}\dot{L}(t) \le& J_x^{\mathsf{T}}\big(f(x)+g(x)\hat{u}\big)-\lambda_{\min}(M)\|\mathcal{Y}\|^2\\ &-\frac{1}{2}\Pi(x,\hat{u})J_x^{\mathsf{T}}g(x)[I_m-\mathscr{B}(\Phi_2(x))]\\ &\times g^{\mathsf{T}}(x)\nabla\sigma^{\mathsf{T}}\tilde{W}_c+\zeta_N\|\mathcal{Y}\|\end{aligned} \tag{43}$$

where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of M, and ζ_N is an upper bound of $\|N\|$, i.e., $\|N\|\le\zeta_N$.

Based on $\Pi(x,\hat{u})$ given in (35), we divide (43) into the following two cases for discussion.

Case I ($\Pi(x,\hat{u})=0$): In this sense, the first term in (43) is negative. Since $\|x\|>0$ is guaranteed by adding the PE signal, one can obtain that there exists a positive constant μ such that $0<\mu\le\|\dot{x}\|$ implies $J_x^{\mathsf{T}}\dot{x}\le-\|J_x\|\mu<0$ based on Archimedean property of $\mathbb{R}$ [51]. Then, (43) is developed as

$$\begin{aligned}\dot{L}(t) \le& \ J_x\dot{x}-\lambda_{\min}(M)\|\mathcal{Y}\|^2+\zeta_N\|\mathcal{Y}\|\\ \le& -\|J_x\|\mu+\frac{1}{4}\zeta_N^2/\lambda_{\min}(M)\\ &-\lambda_{\min}(M)\Big(\|\mathcal{Y}\|-\frac{1}{2}\zeta_N/\lambda_{\min}(M)\Big)^2.\end{aligned} \tag{44}$$

Thus, (44) yields $\dot{L}(t)<0$ as long as one of the following conditions holds:

$$\|J_x\|>\frac{\zeta_N^2}{4\mu\lambda_{\min}(M)}\triangleq\mathfrak{D}_1,\quad\text{or}\quad\|\mathcal{Y}\|>\frac{\zeta_N}{\lambda_{\min}(M)}. \tag{45}$$

Noticing that $\|\mathcal{Y}\|\le\sqrt{1+\|\varphi\|^2}\|\tilde{W}_c\|$ and observing the fact that $\|\varphi\|\le(1/2)$, we can derive $\|\mathcal{Y}\|\le(\sqrt{5}/2)\|\tilde{W}_c\|$. Then, by using (45), we obtain

$$\|\tilde{W}_c\|>\frac{2\zeta_N}{\sqrt{5}\lambda_{\min}(M)}\triangleq\mathfrak{D}_2.$$

Case II ($\Pi(x,\hat{u})=1$): In this circumstance, the first term in (43) is nonnegative. It implies that the control (27) might not stabilize system (2). Then, by using (27), (43) becomes

$$\begin{aligned}\dot{L}(t) \le& J_x^{\mathsf{T}}f(x)-\kappa J_x^{\mathsf{T}}g(x)\Big[\tanh(\Phi_2(x))\\ &+\frac{1}{2\kappa}\big[I_m-\mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)\nabla\sigma^{\mathsf{T}}\tilde{W}_c\Big]\\ &-\lambda_{\min}(M)\|\mathcal{Y}\|^2+\zeta_N\|\mathcal{Y}\|.\end{aligned} \tag{46}$$

Utilizing (38), we get

$$\begin{aligned}&\tanh(\Phi_2(x))+\frac{1}{2\kappa}\big[I_m-\mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}(x)\nabla\sigma^{\mathsf{T}}\tilde{W}_c\\ &\quad=\tanh(\Phi_1(x))-O\big((\Phi_1(x)-\Phi_2(x))^2\big).\end{aligned} \tag{47}$$

Substituting (47) into (46) and using (24), we have

$$\begin{aligned}\dot{L}(t) \le& J_x^{\mathsf{T}}\big(f(x)+g(x)u^*\big)-J_x^{\mathsf{T}}g(x)\varepsilon_{u^*}\\ &-\kappa J_x^{\mathsf{T}}g(x)O\big((\Phi_1(x)-\Phi_2(x))^2\big)\\ &-\lambda_{\min}(M)\|\mathcal{Y}\|^2+\zeta_N\|\mathcal{Y}\|.\end{aligned} \tag{48}$$

Using Assumptions 3–5 and Fact 1, (48) is developed as

$$\begin{aligned}\dot{L}(t) \le\ & J_x^{\mathsf{T}}\big(f(x)+g(x)u^*\big) - J_x^{\mathsf{T}}g(x)\varepsilon_{u^*} \\ & - \kappa J_x^{\mathsf{T}}g(x)O\big((\Phi_1(x)-\Phi_2(x))^2\big) \\ & - \lambda_{\min}(M)\|\mathcal{Y}\|^2 + \zeta_N\|\mathcal{Y}\| \\ \le\ & -\lambda_{\min}(\mathcal{B}(x))\|J_x\|^2 + g_M(\kappa c_1 + b_{\varepsilon_{u^*}})\|J_x\| \\ & + g_M\kappa c_2\|J_x\|\|\tilde{W}_c\| - \lambda_{\min}(M)\|\mathcal{Y}\|^2 \\ & + \zeta_N\|\mathcal{Y}\|. \end{aligned} \tag{49}$$

For every $\ell_i \in (0,1)$, $i=1,2$, let $\ell_1+\ell_2=1$. Then, (49) can be represented as

$$\begin{aligned}\dot{L}(t) \le\ & -\ell_1\lambda_{\min}(\mathcal{B}(x))\|J_x\|^2 + g_M(\kappa c_1 + b_{\varepsilon_{u^*}})\|J_x\| \\ & - \ell_2\lambda_{\min}(\mathcal{B}(x))\left(\|J_x\| - \frac{g_M\kappa c_2\|\tilde{W}_c\|}{2\ell_2\lambda_{\min}(\mathcal{B}(x))}\right)^2 \\ & + \frac{(g_M\kappa c_2)^2}{4\ell_2\lambda_{\min}(\mathcal{B}(x))}\|\tilde{W}_c\|^2 - \lambda_{\min}(M)\|\mathcal{Y}\|^2 \\ & + \zeta_N\|\mathcal{Y}\|. \end{aligned} \tag{50}$$

Noticing that $\|\tilde{W}_c\|^2 \le \|\mathcal{Y}\|^2$, we develop (50) as

$$\begin{aligned}\dot{L}(t) \le\ & -\ell_1\lambda_{\min}(\mathcal{B}(x))\|J_x\|^2 + g_M(\kappa c_1 + b_{\varepsilon_{u^*}})\|J_x\| \\ & - \left(\lambda_{\min}(M) - \frac{(g_M\kappa c_2)^2}{4\ell_2\lambda_{\min}(\mathcal{B}(x))}\right)\|\mathcal{Y}\|^2 + \zeta_N\|\mathcal{Y}\| \\ =\ & -\ell_1\lambda_{\min}(\mathcal{B}(x))\left(\|J_x\| - \frac{g_M(\kappa c_1 + b_{\varepsilon_{u^*}})}{2\ell_1\lambda_{\min}(\mathcal{B}(x))}\right)^2 \\ & - \frac{\hbar}{4\ell_2\lambda_{\min}(\mathcal{B}(x))}\left(\|\mathcal{Y}\| - \frac{2\ell_2\lambda_{\min}(\mathcal{B}(x))\zeta_N}{\hbar}\right)^2 \\ & + \frac{g_M^2(\kappa c_1 + b_{\varepsilon_{u^*}})^2}{4\ell_1\lambda_{\min}(\mathcal{B}(x))} + \frac{\ell_2\lambda_{\min}(\mathcal{B}(x))\zeta_N^2}{\hbar} \end{aligned} \tag{51}$$

where $\hbar = 4\ell_2\lambda_{\min}(\mathcal{B}(x))\lambda_{\min}(M) - g_M^2\kappa^2c_2^2$. Observe that the value of $\hbar$ depends on the parameters ℓ_2, $\mathcal{B}(x)$, and P_i $(i=1,2)$. Hence, the value of $\hbar$ can be kept positive by properly selecting these parameters.

For convenience, we denote

$$\mathfrak{N} = \frac{g_M^2(\kappa c_1 + b_{\varepsilon_{u^*}})^2}{4\ell_1\lambda_{\min}(\mathcal{B}(x))} + \frac{\ell_2\lambda_{\min}(\mathcal{B}(x))\zeta_N^2}{\hbar}.$$

Then, (51) implies $\dot{L}(t) < 0$ as long as one of the following conditions holds:

$$\|J_x\| > \frac{g_M(\kappa c_1 + b_{\varepsilon_{u^*}})}{2\ell_1\lambda_{\min}(\mathcal{B}(x))} + \sqrt{\frac{\mathfrak{N}}{\ell_1\lambda_{\min}(\mathcal{B}(x))}} \triangleq \mathfrak{D}_1'$$

or

$$\|\mathcal{Y}\| > \frac{2\ell_2\lambda_{\min}(\mathcal{B}(x))\zeta_N}{\hbar} + \sqrt{\frac{4\ell_2\lambda_{\min}(\mathcal{B}(x))\mathfrak{N}}{\hbar}}. \tag{52}$$

Observe that $\|\mathcal{Y}\| \le (\sqrt{5}/2)\|\tilde{W}_c\|$. Then, by using (52), we have

$$\|\tilde{W}_c\| > \frac{4\ell_2\lambda_{\min}(\mathcal{B}(x))\zeta_N}{\sqrt{5}\hbar} + 4\sqrt{\frac{\ell_2\lambda_{\min}(\mathcal{B}(x))\mathfrak{N}}{5\hbar}} \triangleq \mathfrak{D}_2'.$$

Combining cases I and II and using the standard Lyapunov extension theorem [53], we can obtain that J_x is UUB with ultimate bound $\mathfrak{D}_1$ (or $\mathfrak{D}_1'$) and the weight estimation error $\tilde{W}_c$ is also UUB with ultimate bound $\mathfrak{D}_2$ (or $\mathfrak{D}_2'$). ■

Remark 5: Note that Δ_Φ given in (31) satisfies that $\Delta_\Phi \in (-m\ln 4, m\ln 4)$. By Assumptions 3, 5, and 6, we derive that $\rho(x)$ given in (31) is bounded. Meanwhile, by φ and m_s given in (34), we can obtain $\|\varphi\| \le 1/2$ and $1/m_s \le 1$. Therefore, N given in (42) is bounded.

Remark 6: Because $J(x)$ given in Assumption 4 is often obtained by selecting polynomials, one can derive that J_x is also a polynomial with respect to x. Since Theorem 2 has verified that J_x is UUB, one can easily obtain that the trajectory of the closed-loop system is also UUB.

Next, we develop a theorem to show rigorously that the closed-loop system is stable in the sense of uniform ultimate boundedness during NN learning process.

Theorem 3: Consider system (2) with associated HJB equation (9). Let Assumptions 3, 5, and 6 hold and let the control input be given in (27) for system (2). Meanwhile, the weight update law for the critic NN is given in (34). Then, the closed-loop system is guaranteed to be UUB with the ultimate bound $\tilde{\Omega}_x$ given in the subsequent (60).

Proof: Consider the Lyapunov function candidate $V^*(x)$ given in (6). Taking the derivative of $V^*(x)$ along the system trajectory $\dot{x} = f(x) + g(x)\hat{u}$, we get

$$\dot{V}^*(x) = V_x^{*\mathsf{T}}f(x) + V_x^{*\mathsf{T}}g(x)\hat{u}. \tag{53}$$

From (8) and (9), we have

$$V_x^{*\mathsf{T}}f(x) = -V_x^{*\mathsf{T}}g(x)u^* - d_M^2(x) - x^{\mathsf{T}}Qx - 2\kappa\int_0^{u^*}\tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon. \tag{54}$$

Substituting (54) into (53), we derive

$$\dot{V}^*(x) = V_x^{*\mathsf{T}}g(x)\big(\hat{u}-u^*\big) - x^{\mathsf{T}}Qx - d_M^2(x) - 2\kappa\int_0^{u^*}\tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon. \tag{55}$$

Combining (20) with (55) and observing the fact that $2\kappa\int_0^{u^*}\tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon$ is positive definite, we have

$$\dot{V}^*(x) \le -x^{\mathsf{T}}Qx + W_c^{\mathsf{T}}\nabla\sigma g(x)\big(\hat{u}-u^*\big) + \nabla\varepsilon^{\mathsf{T}}g(x)\big(\hat{u}-u^*\big). \tag{56}$$

On the other hand, by using (24), (27), and (38), we get

$$\begin{aligned}\hat{u}-u^* =\ & \kappa\big[\tanh(\Phi_1(x)) - \tanh(\Phi_2(x))\big] - \varepsilon_{u^*} \\ =\ & \frac{1}{2}\big[I_m - \mathscr{B}(\Phi_2(x))\big]g^{\mathsf{T}}\nabla\sigma^{\mathsf{T}}\tilde{W}_c \\ & + \kappa O\big((\Phi_1(x)-\Phi_2(x))^2\big) - \varepsilon_{u^*}. \end{aligned}$$

Then, employing Assumptions 5 and 6 and Fact 1, we derive

$$\|\hat{u}-u^*\| \le \big(g_M b_{\sigma x} + \kappa c_2\big)\|\tilde{W}_c\| + \kappa c_1 + b_{\varepsilon_{u^*}}. \tag{57}$$

By Theorem 2, we know that $\tilde{W}_c$ is UUB with ultimate bound $\mathfrak{D}_2$ (or $\mathfrak{D}_2'$). Let $\mathfrak{M} = \max\{\mathfrak{D}_2, \mathfrak{D}_2'\}$. From (57), we have

$$\|\hat{u}-u^*\| \le \big(g_M b_{\sigma x} + \kappa c_2\big)\mathfrak{M} + \kappa c_1 + b_{\varepsilon_{u^*}} \triangleq \mathfrak{T}_1. \tag{58}$$

Hence, by using (58) and Assumptions 3, 5, and 6, (56) becomes

$$\dot{V}^*(x) \le -\lambda_{\min}(Q)\|x\|^2 + g_M(W_M b_{\sigma x} + b_{\varepsilon x})\mathfrak{T}_1. \quad (59)$$

Therefore, (59) yields $\dot{V}^*(x) < 0$ as long as x is out of

$$\tilde{\Omega}_x = \left\{ x\colon \|x\| \le \sqrt{\frac{g_M(W_M b_{\sigma x} + b_{\varepsilon x})\mathfrak{T}_1}{\lambda_{\min}(Q)}} \right\}. \quad (60)$$

According to the standard Lyapunov extension theorem [53], this verifies that the trajectory of the closed-loop system is UUB. That is, the closed-loop system is stable in the sense of uniform ultimate boundedness during NN learning process. ■

Corollary 1: The control $\hat{u}$ given in (27) can approximate the optimal control u^* within a finite bound $\mathfrak{T}_1$ given in (58). Meanwhile, $\hat{V}(x)$ given in (25) will be close to the optimal value $V^*(x)$ within a finite bound $\mathfrak{T}_2$ given in (61).

Proof: The first part of Corollary 1 has been proved in Theorem 3. The second part of Corollary 1 is derived as follows. Using (20), (25), and Assumptions 5 and 6, we have

$$\|\hat{V} - V^*\| \le b_\sigma \mathfrak{M} + b_{\varepsilon_{u^*}} \triangleq \mathfrak{T}_2 \quad (61)$$

where $\mathfrak{M}$ is given in (58). ■

Remark 7: Noticing the expressions of $\mathfrak{D}_2$ and $\mathfrak{D}_2'$, we find that $\mathfrak{M}$ can be kept small by selecting proper parameters (e.g., $\lambda_{\min}(M)$ is large enough). In addition, as pointed out in [54] and [55], if the number of neurons N_0 goes to infinity, there exist $\varepsilon \to 0$ and $\nabla\varepsilon \to 0$. Hence, ε_{u^*} can be arbitrarily small when N_0 is large enough. That is, $b_{\varepsilon_{u^*}}$ can be kept sufficiently small. Therefore, $\mathfrak{T}_2$ given in (61) can be made very small.

VI. Simulation Results

In this section, two examples are provided to illustrate the effectiveness of the developed theoretical results.

A. Example 1

Consider the uncertain CT linear system given by

$$\dot{x} = Ax + B\left[u(x) + qx_1 \sin^5(x_2)\cos^2(x_3)\right] \quad (62)$$

where

$$A = \begin{bmatrix} -1.01887 & 0.90506 & -0.00215 \\ 0.82225 & -1.07741 & -0.17555 \\ 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

with the state $x = [x_1, x_2, x_3]^{\mathsf{T}} \in \mathbb{R}^3$, the control $u \in \mathfrak{A} = \{u \in \mathbb{R} : |u| \le 1\}$, and q is an unknown parameter. The term $d(x) = qx_1 \sin^5(x_2)\cos^2(x_3)$ reflects the uncertainty of system (62). For simplicity of discussion, we assume that $q \in [-2, 2]$ and $d_M(x) = 2\|x\|$.

The nominal system is $\dot{x} = Ax + Bu$ with A and B given in (62). The corresponding value function is

$$V(x) = \int_0^\infty \left(4\|x\|^2 + x^{\mathsf{T}}Qx + \mathcal{W}(u)\right)dt$$

where $Q = I_3$ and $\mathcal{W}(u) = 2\kappa \int_0^u \tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon$.

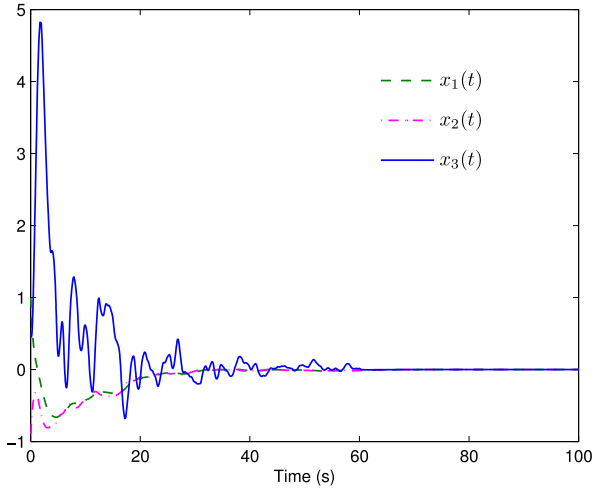


Fig. 2. Evolution of nominal system state $x(t)$ for NN learning process.

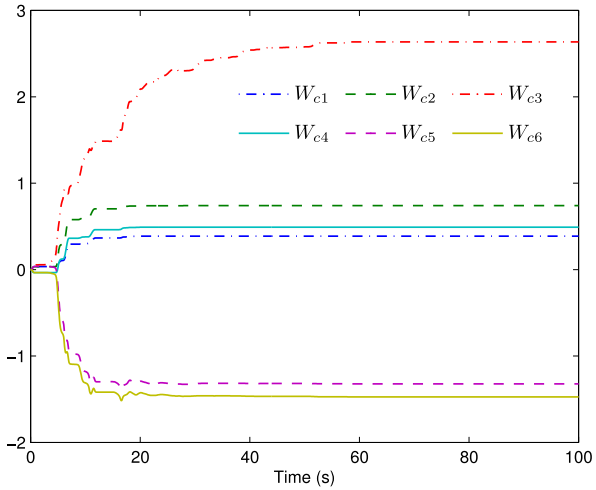


Fig. 3. Convergence of the critic NN weight $\hat{W}_c$.

The activation function for the critic NN is chosen with $N_0 = 6$ neurons as

$$\sigma(x) = \left[x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3\right]^{\mathsf{T}}$$

and $\hat{W}_c = [W_{c1}, W_{c2}, \ldots, W_{c6}]^{\mathsf{T}}$ is the critic NN weight. It should be emphasized that choosing the proper neurons for NNs is still an open question [56]. In this example, the number of neurons is obtained by computer simulations. We find that selecting six neurons in the hidden layer for the critic NN can lead to satisfactory simulation results.

The initial weight for the critic NN is chosen to be zero, and the initial state is set to be $x_0 = [1, -1, 0.5]^{\mathsf{T}}$. The parameters are designed as $\kappa = 1$, $\gamma = 0.95$. To guarantee the PE condition, a small exploratory signal $n(t) = \sin^2(t)\cos(t) + \sin^2(2t)\cos(0.1t) + \sin^2(1.2t)\cos(0.5t) + \sin^5(t) + \sin^2(1.12t) + \cos(2.4t)\sin^3(2.4t)$ is added to the control $u(t)$ for the first 60 s. The developed control algorithm is implemented by using (27) and (34).

The computer simulation results are shown in Figs. 2–5. Fig. 2 illustrates the evolution of the nominal system state x

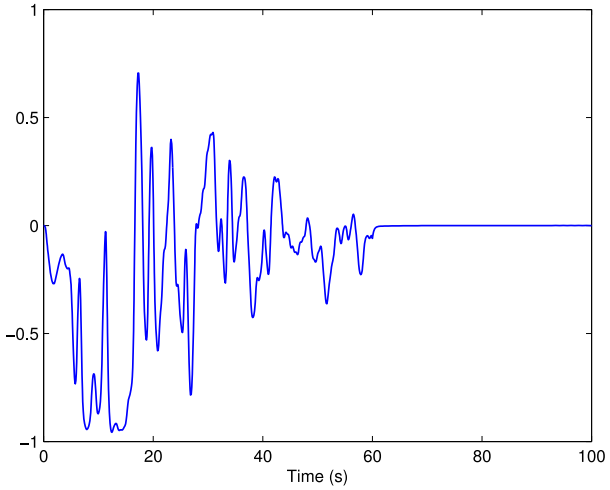


Fig. 4. Control input u.

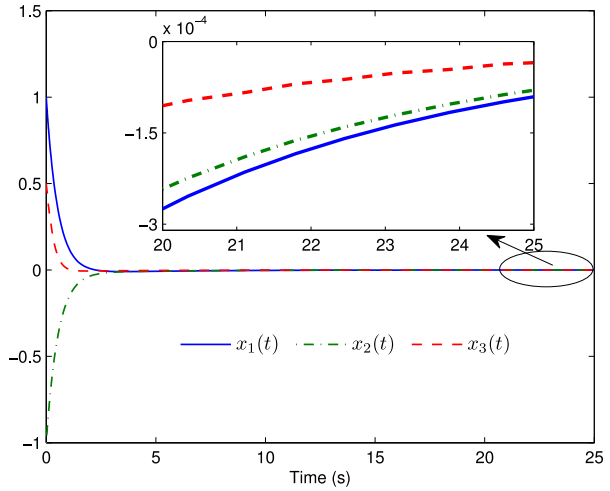


Fig. 5. Trajectories of system (62) under approximate optimal control.

during the critic NN learning process. Fig. 3 indicates the convergence of the critic NN weights. As shown in Fig. 3, the critic NN weights converge to $W_c = [0.3871, 0.7402, 2.6356, 0.4915, -1.3225, -1.4737]^{\mathsf{T}}$. Fig. 4 presents the control input. Fig. 5 shows the trajectories of system (62) under approximate optimal control. From Figs. 2–4, one can find that the developed adaptive control guarantees that all signals in the closed-loop optimal control system are UUB. Moreover, from Fig. 3, one can find the PE condition ensures the critic NN weight to be convergent in approximately 60 s. By Corollary 1, the critic is considered to be convergent to the approximate optimal value. Then, we apply the derived approximate optimal control to system (62). Fig. 5 shows the approximate optimal control can guarantee system (62) to be stable in the sense of uniform ultimate boundedness.

B. Example 2

Consider the uncertain CT nonlinear system given by [43]

$$\dot{x} = f(x) + g(x)\big(u(x) + qx_1 \sin x_2\big) \tag{63}$$

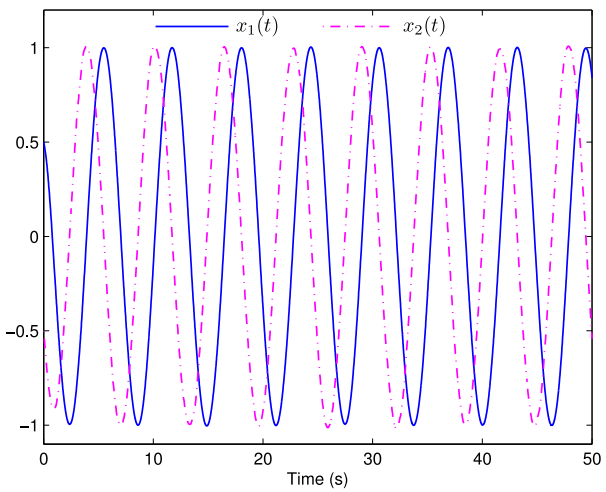


Fig. 6. Trajectories of system (63) with initial control $u = 0$.

where

$$f(x) = \begin{bmatrix} x_1 + x_2 - x_1\big(x_1^2 + x_2^2\big) \\ -x_1 + x_2 - x_2\big(x_1^2 + x_2^2\big) \end{bmatrix}, \quad g(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

with the state $x = [x_1, x_2]^{\mathsf{T}} \in \mathbb{R}^2$, the control $u \in \mathfrak{A} = \{u \in \mathbb{R} : |u| \le 1\}$, and q is an unknown parameter. The term $d(x) = qx_1 \sin x_2$ gives rise to the uncertainty of system (63). For simplicity of discussion, in this example, we assume that $q \in [-1.4, 1.4]$ and $d_M(x) = 1.4\|x\|$.

The nominal system is $\dot{x} = f(x) + g(x)u$ with $f(x)$ and $g(x)$ given in (63). The corresponding value function is

$$V(x) = \int_0^\infty \big(2\|x\|^2 + x^{\mathsf{T}}Qx + \mathcal{W}(u)\big)dt$$

where $Q = I_2$ and $\mathcal{W}(u) = 2\kappa \int_0^u \tanh^{-\mathsf{T}}(\upsilon/\kappa)d\upsilon$.

The activation function for the critic NN is chosen with $N_0 = 24$ neurons as

$$\begin{aligned}\sigma(x) = \Big[&x_1^2, x_2^2, x_1x_2, x_1^4, x_2^4, x_1^3x_2, x_1^2x_2^2, x_1x_2^3, x_1^6 \\ &x_2^6, x_1^5x_2, x_1^4x_2^2, x_1^3x_2^3, x_1^2x_2^4, x_1x_2^5, x_1^8, x_2^8 \\ &x_1^7x_2, x_1^6x_2^2, x_1^5x_2^3, x_1^4x_2^4, x_1^3x_2^5, x_1^2x_2^6, x_1x_2^7\Big]^{\mathsf{T}}\end{aligned}$$

and $\hat{W}_c = [W_{c1}, W_{c2}, \ldots, W_{c24}]^{\mathsf{T}}$ is the critic NN weight. Similar to Example 1, the number of neurons is also obtained by computer simulations.

The initial weight for the critic NN is selected to be zero (i.e., the initial control $u = 0$), and the initial system state is set to be $x_0 = [0.5, -0.5]^{\mathsf{T}}$. It is significant to point out that, in this circumstance, the initial control cannot stabilize system (63). To illustrate this fact, we present Fig. 6 (since the closed-loop system turns out to be periodic oscillation, we provide the trajectory of the system for the first 50 s). It also verifies that there is no requirement of the initial stabilizing control for implementing the new developed algorithm. The other parameters and the exploratory signal are chosen to be the same as in Example 1. The exploratory signal is added to $u(t)$ for the first 600 s.

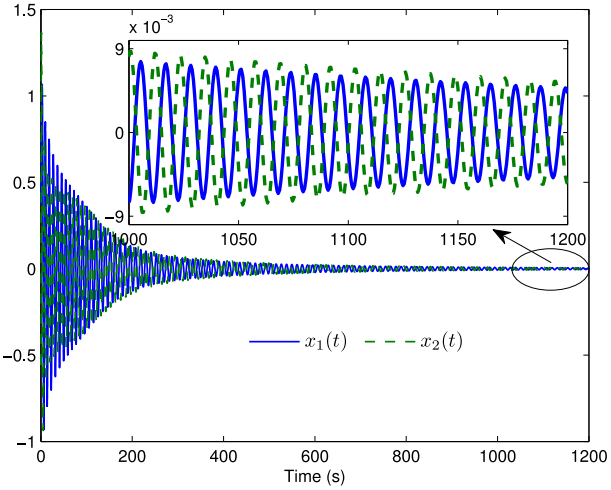


Fig. 7. Evolution of nominal system state $x(t)$ for NN learning process.

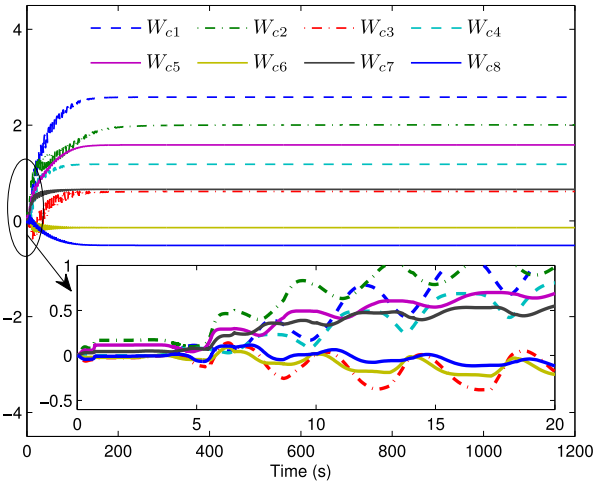


Fig. 8. Convergence of the critic NN weight $\hat{W}_c$.

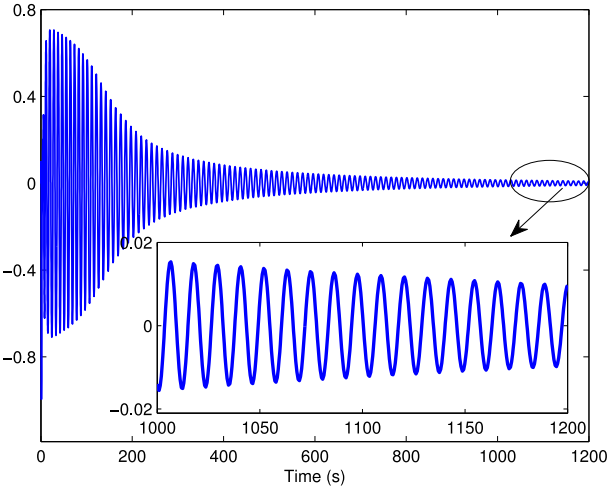


Fig. 9. Control input u.

The computer simulation results are illustrated in Figs. 7–10. Fig. 7 shows the evolution of the nominal system state x during the critic NN learning process. Fig. 8 presents the convergence curves of the first eight weights of the critic NN. In fact, the weights of the critic NN converge to

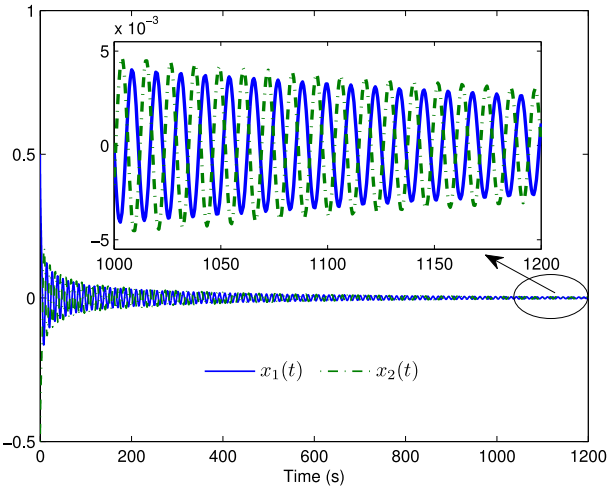


Fig. 10. Trajectories of system (63) under approximate optimal control.

W_c = [2.5849, 2.0037, 0.6158, 1.1825, 1.5860, −0.1390, 0.6583, −0.5108, 0.6364, 0.6695, −0.1333, 0.3175, −0.1374, 0.3578, −0.2267, 0.3878, 0.2951, −0.0874, 0.1738, −0.0517, 0.1243, −0.0437, 0.1497, −0.0854]^T. Fig. 9 shows the control input. Fig. 10 indicates the trajectories of system (63) under approximate optimal control. From Figs. 7–9, one can find that the developed adaptive control guarantees that all signals in the closed-loop optimal control system are UUB. In addition, from Fig. 8, one can find that the PE condition ensures the weight to be convergent in approximately 600 s. By Corollary 1, the critic is considered to be convergent to the approximate optimal value. Then, we apply the derived approximate optimal control to system (63). Fig. 10 shows that the derived optimal control can keep system (63) stable in the sense of uniform ultimate boundedness. Furthermore, it should be emphasized that, the present algorithm for deriving optimal control differs significantly from the algorithms proposed in [38], [39], [41], and [43]. In our case, no initial stabilizing control is required. This feature has been shown by Fig. 8, where the initial weight for the critic NN can be zeros. In this situation, the closed-loop system is unstable (see Fig. 6).

VII. Conclusion

In this paper, we have developed a novel RL-based robust control algorithm for constrained-input uncertain nonlinear CT systems by solving constrained optimal control problems. The present algorithm employs a single critic NN to obtain the approximate optimal control, which guarantees the uncertain nonlinear CT system to be stable in the sense of uniform ultimate boundedness. By using the developed algorithm, no initial stabilizing control is required. A limitation of the present method is that the prior knowledge of $f(x)$ and $g(x)$ is required to be available. In our future work, we will relax the restrictive condition. In addition, if system (1) is nonaffine, then it will be a rather challenging task to design a robust controller. Therefore, we shall also focus on developing RL-based robust control algorithms for unknown nonaffine nonlinear CT systems in the presence of input constraints.

REFERENCES

[1] T. Hu and Z. Lin, *Control Systems With Actuator Saturation: Analysis and Design*. Boston, MA, USA: Birkhauser, 2001.

[2] F. L. Lewis, J. Campos, and R. R. Selmic, *Neuro-Fuzzy Control of Industrial Systems With Actuator Nonlinearities*. Philadelphia, PA, USA: SIAM Press, 2002.

[3] W. Gao and R. R. Selmic, "Neural network control of a class of nonlinear systems with actuator saturation," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 147–156, Jan. 2006.

[4] H. Li and Y. Shi, "Robust distributed model predictive control of constrained continuous-time nonlinear systems: A robustness constraint approach," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1673–1678, Jun. 2014.

[5] A. E. Bryson and Y. C. Ho, *Applied Optimal Control: Optimization, Estimation and Control*. Boca Raton, FL, USA: CRC Press, 1975.

[6] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. Hoboken, NJ, USA: Wiley, 2012.

[7] R. E. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[8] P. J. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Dept. Appl. Math., Harvard University, Cambridge, MA, USA, 1974.

[9] H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive Dynamic Programming for Control: Algorithms and Stability*. London, U.K.: Springer, 2013.

[10] Q. Wei, F. Y. Wang, D. Liu, and X. Yang, "Finite-approximation-error-based discrete-time iterative adaptive dynamic programming," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2820–2833, Dec. 2014.

[11] D. Liu, D. Wang, and X. Yang, "An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs," *Inf. Sci.*, vol. 220, pp. 331–342, Jan. 2013.

[12] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 621–634, Mar. 2014.

[13] Z. Ni, H. He, J. Wen, and X. Xu, "Goal representation heuristic dynamic programming on maze navigation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2038–2050, Dec. 2013.

[14] Z. Ni, H. He, and J. Wen, "Adaptive learning in tracking control based on the dual critic network design," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 913–928, Jun. 2013.

[15] H. Zhang, L. Cui, and Y. Luo, "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 206–216, Feb. 2013.

[16] H. Zhang, C. Qin, and Y. Luo, "Neural-network-based constrained optimal control scheme for discrete-time switched nonlinear system using dual heuristic programming," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 3, pp. 839–849, Jul. 2014.

[17] Q. Wei, D. Liu, and X. Yang, "Infinite horizon self-learning optimal control of nonaffine discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 4, pp. 866–879, Apr. 2015.

[18] S. Mehraeen, T. Dierks, S. Jagannathan, and M. L. Crow, "Zero-sum two-player game theoretic formulation of affine nonlinear discrete-time systems using neural networks," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1641–1655, Dec. 2013.

[19] J. Nascimento and W. B. Powell, "An optimal approximate dynamic programming algorithm for concave, scalar storage problems with vector-valued controls," *IEEE Trans. Autom. Control*, vol. 58, no. 12, pp. 2995–3010, Dec. 2013.

[20] A. Heydari, "Revisiting approximate dynamic programming and its convergence," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2733–2743, Dec. 2014.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning—An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[22] M. Kusy and R. Zajdel, "Application of reinforcement learning algorithms for the adaptive computation of the smoothing parameter for probabilistic neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[23] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Syst. Mag.*, vol. 32, no. 6, pp. 76–105, Dec. 2012.

[24] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 14–25, Feb. 2011.

[25] X. Yang, D. Liu, and Q. Wei, "Near-optimal online control of uncertain nonlinear continuous-time systems based on concurrent learning," in *Proc. Int. Joint Conf. Neural Netw.*, Beijing, China, Jul. 2014, pp. 231–238.

[26] B. Luo, H. N. Wu, and T. Huang, "Off-policy reinforcement learning for H_∞ control design," *IEEE Trans. Cybern.*, vol. 45, no. 1, pp. 65–76, Jan. 2015.

[27] X. Zhong, H. He, H. Zhang, and Z. Wang, "Optimal control for unknown discrete-time nonlinear Markov jump systems using adaptive dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2141–2155, Dec. 2014.

[28] Y. Liu, L. Tang, S. Tong, C. L. P. Chen, and D. Li, "Reinforcement learning design-based adaptive tracking control with less learning parameters for nonlinear discrete-time MIMO systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 1, pp. 165–176, Jan. 2015.

[29] J. Y. Lee, J. B. Park, and Y. H. Choi, "Integral reinforcement learning for continuous-time input-affine nonlinear systems with simultaneous invariant explorations," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[30] D. Wang, D. Liu, and H. Li, "Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 627–632, Apr. 2014.

[31] 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, May 2010.

[32] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2226–2236, Dec. 2011.

[33] S. Bhasin *et al.*, "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," *Automatica*, vol. 49, no. 1, pp. 82–92, Jan. 2013.

[34] F. L. Lewis and D. Liu, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Hoboken, NJ, USA: Wiley, 2013.

[35] X. Yang, D. Liu, and Y. Huang, "Neural-network-based online optimal control for uncertain non-linear continuous-time systems with control constraints," *IET Control Theory Appl.*, vol. 7, no. 17, pp. 2037–2047, Nov. 2013.

[36] X. Yang, D. Liu, D. Wang, and Q. Wei, "Discrete-time online learning control for a class of unknown nonaffine nonlinear systems using reinforcement learning," *Neural Netw.*, vol. 55, pp. 30–41, Jul. 2014.

[37] M. Palanisamy, H. Modares, F. L. Lewis, and M. Aurangzeb, "Continuous-time Q-learning for infinite-horizon discounted cost linear quadratic regulator problems," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 165–176, Feb. 2015.

[38] 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, May 2005.

[39] H. Modares, F. L. Lewis, and M. Naghibi-Sistani, "Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems," *Automatica*, vol. 50, no. 1, pp. 193–202, Jan. 2014.

[40] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica*, vol. 50, no. 7, pp. 1780–1792, Jul. 2014.

[41] X. Yang, D. Liu, and D. Wang, "Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints," *Int. J. Control*, vol. 87, no. 3, pp. 553–566, 2014.

[42] Y. Jiang and Z. P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, Oct. 2012.

[43] D. M. Adhyaru, I. N. Kar, and M. Gopal, "Bounded robust control of nonlinear systems using neural network-based HJB solution," *Neural Comput. Appl.*, vol. 20, no. 1, pp. 91–103, Feb. 2011.

[44] Y. Jiang and Z. P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 882–893, May 2014.

[45] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," in *Proc. Amer. Control Conf.*, Baltimore, MD, USA, Jun./Jul. 2010, pp. 1568–1573.

[46] M. J. Corless and G. Leitmann, "Continuous state feedback guaranteeing uniform ultimate boundedness for uncertain dynamic systems," *IEEE Trans. Autom. Control*, vol. 26, no. 5, pp. 1139–1144, Oct. 1981.
[47] F. Lin and R. D. Brandt, "An optimal control approach to robust control of robot manipulators," *IEEE Trans. Robot. Autom.*, vol. 14, no. 1, pp. 69–77, Feb. 1998.
[48] H. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[49] W. Rudin, *Functional Analysis*, 2nd ed. Singapore: McGraw-Hill, 1991.
[50] R. Beard, G. Saridis, and J. Wen, "Galerkin approximations of the generalized Hamilton–Jacobi–Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, Dec. 1997.
[51] W. Rudin, *Principles of Mathematical Analysis*. 3rd ed. Singapore: McGraw-Hill, 1976.
[52] D. Nodland, H. Zargarzadeh, and S. Jagannathan, "Neural network-based optimal adaptive output feedback control of a helicopter UAV," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1061–1073, Jul. 2013.
[53] F. L. Lewis, S. Jagannathan, and A. Yesildirek, *Neural Network Control of Robot Manipulators and Nonlinear Systems*. London, U.K.: Taylor and Francis, 1999.
[54] K. Hornic and M. Stinchombe, "Multilayer feedforward neural networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
[55] B. A. Finlayson, *The Method of Weighted Residuals and Variational Principles*. New York, NY, USA: Academic Press, 1972.
[56] R. Padhi, N. Unnikrishnan, X. Wang, and S. N. Balakrishnan, "A single network adaptive critic (SNAC) architecture for optimal control synthesis for a class of nonlinear systems," *Neural Netw.*, vol. 19, no. 10, pp. 1648–1660, Dec. 2006.

Derong Liu (S'91–M'94–SM'96–F'05) received the B.S. degree in mechanical engineering from the East China Institute of Technology (now Nanjing University of Science and Technology), Nanjing, China, the M.S. degree in automatic control theory and applications from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, and the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1982, 1987, and 1994, respectively.

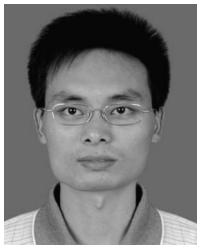
He was a Product Design Engineer with the China North Industries Corporation, Jilin, China, from 1982 to 1984. He was an Instructor with the Graduate School of the Chinese Academy of Sciences, from 1987 to 1990. He was a Staff Fellow with General Motors Research and Development Center, Warren, MI, USA, from 1993 to 1995. He was an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, from 1995 to 1999. He joined the University of Illinois at Chicago, Chicago, IL, USA, in 1999, and became a Full Professor of Electrical and Computer Engineering and of Computer Science in 2006. He was selected for the 100 Talents Program by the Chinese Academy of Sciences in 2008, and he currently serves as an Associate Director with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation. He has published 15 books (six research monographs and nine edited volumes).

Dr. Liu was a recipient of the Faculty Early Career Development Award from the National Science Foundation in 1999, the University Scholar Award from the University of Illinois from 2006 to 2009, and the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008. He was an Associate Editor of *Automatica* from 2006 to 2009. He is currently an elected AdCom member of the IEEE Computational Intelligence Society and he is an Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. He also serves as an Associate Editor for the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, *Soft Computing*, *Neurocomputing*, *Neural Computing and Applications*, and *Science in China Series F: Information Sciences*. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I: FUNDAMENTAL THEORY AND APPLICATIONS from 1997 to 1999, the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2001 to 2003, the IEEE TRANSACTIONS ON NEURAL NETWORKS from 2004 to 2009, the *IEEE Computational Intelligence Magazine* from 2006 to 2009, and the *IEEE Circuits and Systems Magazine* from 2008 to 2009. He was the Letters Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS from 2006 to 2008. He is a fellow of the International Neural Network Society.

Xiong Yang received the B.S. degree in mathematics and applied mathematics from Central China Normal University, Wuhan, China, the M.S. degree in pure mathematics from Shandong University, Jinan, China, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2008, 2011, and 2014, respectively.

He is currently an Assistant Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His current research interests include adaptive dynamic programming, reinforcement learning, adaptive control, and neural networks.

Dr. Yang was a recipient of the Presidential Award of Excellence from the Chinese Academy of Sciences in 2014.



Ding Wang (M'15) received the B.S. degree in mathematics from the Zhengzhou University of Light Industry, Zhengzhou, China, the M.S. degree in operations research and cybernetics from Northeastern University, Shenyang, China, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007, 2009, and 2012, respectively.

He is currently an Associate Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His current research interests include neural networks, learning systems, and complex systems and intelligent control.

Qinglai Wei (M'11) received the B.S. degree in automation, the M.S. degree in control theory and control engineering, and the Ph.D. degree in control theory and control engineering, all from the Northeastern University, Shenyang, China, in 2002, 2005, and 2008, respectively.

He was a Post-Doctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2009 to 2011, where he is currently an Associate Professor. His current research interests include neural network-based control, adaptive dynamic programming, optimal control, nonlinear systems, and their industrial applications.

Dr. Wei has been an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS since 2014. He is currently an Associate Editor of *Acta Automatica Sinica*.