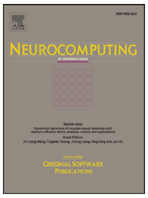
Neurocomputing 266 (2017) 353–360

Contents lists available at ScienceDirect

Neurocomputing

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



A novel neural optimal control framework with nonlinear dynamics: Closed-loop stability and simulation verification[☆]

Ding Wang[a,b], Chaoxu Mu[c,*]

[a] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[c] Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 20 February 2017
Revised 1 April 2017
Accepted 20 May 2017
Available online 25 May 2017

Communicated by Mou Chen

Keywords:
Adaptive dynamic programming
Adaptive system
Learning control
Neural network
Optimal regulator
Stability

ABSTRACT

In this paper, we focus on developing adaptive optimal regulators for a class of continuous-time nonlinear dynamical systems through an improved neural learning mechanism. The main objective lies in that establishing an additional stabilizing term to reinforce the traditional training process of the critic neural network, so that to reduce the requirement with respect to the initial stabilizing control, and therefore, bring in an obvious convenience to the adaptive-critic-based learning control implementation. It is exhibited that by employing the novel updating rule, the adaptive optimal control law can be obtained with an excellent approximation property. The closed-loop system is constructed and its stability issue is handled by considering the improved learning criterion. Experimental simulations are also conducted to verify the efficient performance of the present design method, especially the major role that the stabilizing term performed.

1. Introduction

As is known, linear optimal regulator design has been studied by control scientists and engineers for many years. For nonlinear systems, the optimal control problem always leads to cope with the nonlinear Hamilton–Jacobi–Bellman (HJB) equation, which is intractable to solve in general cases. Fortunately, a series of iterative methods have been established to tackle the optimal control problems approximately [1–3]. For adaptive/approximate dynamic programming (ADP) [3–9], the adaptive critic is taken as the basic structure and neural networks are often involved to serve as the function approximator. Generally speaking, employing the ADP method always results in approximate or adaptive optimal feedback controllers. Note that optimality and adaptivity are two important criteria of control theory and also possess grea t significance to control engineering, such as [10–16]. Hence, this kind of adaptive-critic-based optimal control design has great potentials in various control applications.

In the last decade, the methodology of ADP has been widely used for optimal control of discrete-time systems, such as [17–24] and continuous-time systems, like [25–32]. Heydari and Balakrishnan [18] investigated finite-horizon nonlinear optimal control with input constraints by adopting single network adaptive critic designs. Song et al. [19] proposed a novel ADP algorithm to solve the nearly optimal finite-horizon control problem for a class of deterministic nonaffine nonlinear time-delay systems. Mu et al. [21] studied the approximate optimal tracking control design for a class of discrete-time nonlinear systems based on the iterative globalized dual heuristic programming technique. Zhao et al. [22] gave a model-free optimal control method for optimal control of affine nonlinear systems without using the dynamics information. Qin et al. [23] studied the neural-network-based self-learning H_∞ control design for discrete-time input-affine nonlinear systems in light of ADP method. Zhong et al. [24] developed the theoretical basis of the new goal representation heuristic dynamic programming structure for general discrete-time nonlinear systems. Vamvoudakis and Lewis [25] proposed an important actor-critic algorithm to attain the continuous-time infinite horizon nonlinear optimal regulation design. Zhang et al. [26] studied the approximate optimal control for non-zero-sum differential games with continuous-time nonlinear dynamics based on single

[☆] This work was supported in part by National Natural Science Foundation of China under Grants U1501251, 61533017, and 61233001, in part by Beijing Natural Science Foundation under Grant 4162065, in part by Research Fund of Tianjin Key Laboratory of Process Measurement and Control under Grant TKLPMC-201612, and in part by the Early Career Development Award of SKLMCCS.

[*] Corresponding author.
E-mail addresses: ding.wang@ia.ac.cn (D. Wang), cxmu@tju.edu.cn (C. Mu).

http://dx.doi.org/10.1016/j.neucom.2017.05.051
0925-2312/© 2017 Elsevier B.V. All rights reserved.

network adaptive critics. Modares and Lewis [27] proposed a linear quadratic trajectory tracking control method for partially-unknown continuous-time systems based on the reinforcement learning technique. Na and Herrmann [28] proposed an online adaptive and approximate optimal trajectory tracking approach with a simplified dual approximation architecture for continuous-time unknown nonlinear controlled plants. Bian et al. [29] studied decentralized adaptive optimal control of a class of large-scale systems and its application toward the power systems. Jiang and Jiang [30] originally established the global ADP structure for continuous-time nonlinear systems. Luo et al. [31] provided the reinforcement learning solution for HJB equation with respect to the constrained optimal control problems. Gao and Jiang [32] applied ADP to design optimal output regulation of linear systems adaptively. This greatly promotes the development of the adaptive critic control designs of complex nonlinear systems. However, the traditional adaptive critic control design always depends on the choice of an initial stabilizing control, which is pretty difficult to find out in control practices. Actually, requiring an initial stabilizing control is a common property of [25,27], which weakens the application aspect of the adaptive-critic-based design to a certain extent, and correspondingly, motivates our research greatly. This paper focuses on developing nonlinear adaptive optimal regulators through an improved neural learning mechanism. The major contribution lies in that it constructs a simple reinforced structure to achieve the nonlinear optimal regulation design adaptively, without requiring the initial stabilizing controller. Moreover, the stability of the closed-loop system including the additional stabilizing term is presented with a simpler proof process. Finally, the important role that the stabilizing term plays is also verified by simulation study in detail. This can be regarded as an improvement to the traditional adaptive critic designs, like [25,27].

The rest of the current paper is organized as follows. The studied problem is described briefly in Section 2. The improved adaptive critic design technique of nonlinear adaptive optimal control is developed with closed-loop stability analysis in Section 3. The simulation studies and the concluding remarks are presented in Section 4 and Section 5, respectively. Incidentally, the main notations used in the paper are listed as follows. $\mathbb{R}$ stands for the set of all real numbers. $\mathbb{R}^n$ is the Euclidean space of all n-dimensional real vectors. $\mathbb{R}^{n\times m}$ is the space of all $n \times m$ real matrices. $\|\cdot\|$ denotes the vector norm of a vector in $\mathbb{R}^n$ or the matrix norm of a matrix in $\mathbb{R}^{n\times m}$. I_n represents the $n \times n$ identity matrix. $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ calculate the maximal and minimal eigenvalues of a matrix, respectively. Let Ω be a compact subset of $\mathbb{R}^n$ and $\mathscr{A}(\Omega)$ be the set of admissible control laws on Ω. The superscript "T" is taken for representing the transpose operation and $\nabla(\cdot)\triangleq\partial(\cdot)/\partial x$ is employed to denote the gradient operator.

2. Problem statement

In this paper, we study a class of continuous-time nonlinear systems with input-affine form given by

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

where $x(t) \in \Omega \subset \mathbb{R}^n$ is the state variable, $u(t) \in \Omega_u \subset \mathbb{R}^m$ is the control variable, and the system functions $f(\cdot) \in \mathbb{R}^n$ and $g(\cdot) \in \mathbb{R}^{n\times m}$ are known matrices and are differentiable in the arguments satisfying $f(0) = 0$. In this paper, we let the initial state at $t = 0$ be $x(0) = x_0$ and let $x = 0$ be the equilibrium point. In addition, we assume that $f(x)$ is Lipschitz continuous on a set Ω in $\mathbb{R}^n$ which contains the origin and the nonlinear plant (1) is controllable.

In order to design the optimal feedback control law $u(x)$, we let $Q(x) > 0$ when $x \neq 0$ and $Q(0) = 0$, set R as a positive definite matrix with appropriate dimension, take

$$U(x(\tau), u(\tau)) = Q(x(\tau)) + u^{\mathsf{T}}(\tau)Ru(\tau)$$

to stand for the utility function, and then define the infinite horizon cost function as

$$J(x(t), u) = \int_t^\infty U(x(\tau), u(\tau))\mathrm{d}\tau. \tag{2}$$

Notice here the cost $J(x(t), u)$ is often written as $J(x(t))$ or $J(x)$ for simplicity. For an admissible control law $u \in \mathscr{A}(\Omega)$, if the cost function (2) is continuously differentiable, then the related infinitesimal version is the nonlinear Lyapunov equation

$$0 = U(x, u) + (\nabla J(x))^{\mathsf{T}}[f(x) + g(x)u]$$

with $J(0) = 0$. Next, we define the Hamiltonian of system (1) as

$$H(x, u, \nabla J(x)) = U(x, u) + (\nabla J(x))^{\mathsf{T}}[f(x) + g(x)u].$$

According to Bellman's optimality principle, the optimal cost function $J^*(x)$

$$J^*(x) = \min_{u\in\mathscr{A}(\Omega)} \int_t^\infty U(x(\tau), u(\tau))\mathrm{d}\tau,$$

makes sure that the so-called HJB equation

$$\min_u H(x, u, \nabla J^*(x)) = 0$$

holds. Similar as [25,30], the optimal feedback control law is computed by

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla J^*(x). \tag{3}$$

Noticing the optimal control expression (3), the HJB equation is in fact

$$\begin{aligned} 0 &= U(x, u^*) + (\nabla J^*(x))^{\mathsf{T}}[f(x) + g(x)u^*] \\ &= Q(x) + (\nabla J^*(x))^{\mathsf{T}} f(x) \\ &\quad - \frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J^*(x), J^*(0) = 0. \end{aligned} \tag{4}$$

Eq. (4) is actually $H(x, u^*, \nabla J^*(x)) = 0$, which is difficult to get the solution theoretically. In other words, it is clearly not easy to obtain the optimal control law (3) for general nonlinear systems, which inspires us to effectively design a class of approximate optimal control schemes.

3. Approximate optimal control design and its stability

During the approximate control algorithm implementation, the idea of adaptive critic is adopted with neural network approximation. Using the universal approximation property, the optimal cost function $J^*(x)$ can be expressed by a neural network with a single hidden layer on a compact set Ω as

$$J^*(x) = \omega_c^{\mathsf{T}}\sigma_c(x) + \varepsilon_c(x), \tag{5}$$

where $\omega_c \in \mathbb{R}^{l_c}$ is the ideal weight vector that is upper bounded, l_c is the number of hidden neurons, $\sigma_c(x) \in \mathbb{R}^{l_c}$ is the activation function, and $\varepsilon_c(x) \in \mathbb{R}$ is the reconstruction error. Then, the gradient vector is

$$\nabla J^*(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x).$$

Noticing the ideal weight is unknown in advance, a critic network is developed to approximate the optimal cost function as

$$\hat{J}^*(x) = \hat{\omega}_c^{\mathsf{T}}\sigma_c(x), \tag{6}$$

where $\hat{\omega}_c \in \mathbb{R}^{l_c}$ denotes the estimated weight vector. Similarly, we derive the gradient vector as

$$\nabla\hat{J}^*(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c.$$

Considering the feedback formulation (3) and the neural network expression (5), the optimal control law can be rewritten as

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\big[(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x)\big]. \tag{7}$$

Using the critic neural network (6), the approximate optimal feedback control function is

$$\hat{u}^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{8}$$

Based on the neural network formulation, the approximate version of the Hamiltonian is expressed by

$$\hat{H}(x,\hat{u}^*(x),\nabla\hat{J}^*(x)) = U(x,\hat{u}^*(x)) + \hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)[f(x)+g(x)\hat{u}^*(x)]. \tag{9}$$

By considering the fact $H(x,u^*,\nabla J^*(x))=0,$ we have $e_c=\hat{H}(x,\hat{u}^*(x),\nabla\hat{J}^*(x))$ and then find that

$$\frac{\partial e_c}{\partial\hat{\omega}_c} = \nabla\sigma_c(x)[f(x)+g(x)\hat{u}^*(x)] \triangleq \phi, \tag{10}$$

where $\phi\in\mathbb{R}^{l_c}$ and the set containing the elements $\phi_1,\phi_2,\ldots,\phi_{l_c}$ is linearly independent.

Now, we show how to train the critic network and design the weight vector $\hat{\omega}_c$ to minimize the objective function $E_c=0.5e_c^2$. According to (9) and (10), we can employ the normalized steepest descent algorithm

$$\dot{\hat{\omega}}_c = -\alpha_c\frac{1}{(1+\phi^{\mathsf{T}}\phi)^2}\left(\frac{\partial E_c}{\partial\hat{\omega}_c}\right) = -\alpha_c\frac{\phi}{(1+\phi^{\mathsf{T}}\phi)^2}e_c$$

to adjust the weight vector, where $\alpha_c>0$ is the learning rate. Note that in this traditional design technique, we should choose a special weight vector to create the initial stabilizing control law and then start the training process. Otherwise, an unstable control may result in the instability of the closed-loop system.

Recently, a new near-optimal control algorithm was proposed in [33] and then applied for solving several control design problems [34,35]. Among that, an ADP-based guaranteed cost neural tracking control algorithm for a class of continuous-time uncertain nonlinear dynamics was developed in [35]. However, the stability proof of the above results is quite complicated. Inspired by Dierks and Jagannathan [33–35], we introduce an additional Lyapunov function to improve the critic learning mechanism and adopt it to facilitate updating the critic weight vector with a novel fashion. Similar as [34,35], we make the following assumption.

Assumption 1. Consider system (1) with the cost function (2) and its closed-loop form with the action of the optimal feedback control (7). Let $J_s(x)$ be a continuously differentiable Lyapunov function candidate that satisfies

$$\dot{J}_s(x) = (\nabla J_s(x))^{\mathsf{T}}[f(x)+g(x)u^*(x)] < 0.$$

Then, there exists a positive definite matrix $\Xi\in\mathbb{R}^{n\times n}$ such that

$$\begin{aligned}(\nabla J_s(x))^{\mathsf{T}}[f(x)+g(x)u^*(x)] &= -(\nabla J_s(x))^{\mathsf{T}}\Xi\nabla J_s(x)\\ &\le -\lambda_{\min}(\Xi)\|\nabla J_s(x)\|^2\end{aligned}$$

is true.

Remark 1. This is a common assumption which was used in the literature, for instance [26,33–35], in order to facilitate designing the control law and discussing the closed-loop stability. During the implementation, $J_s(x)$ can be obtained by suitably selecting a polynomial with respect to the state vector, such as $J_s(x)=0.5x^{\mathsf{T}}x$. It is an experimental choice incorporating engineering experience and intuition after considering a tradeoff between control accuracy and computation complexity.

When applying the approximate optimal control law (8) to the controlled plant and for the purpose of excluding the case that the closed-loop system is unstable, we can introduce an additional term to reinforce the training process by modulating the time derivative of $J_s(x)$ along the negative gradient direction with respect to the weight vector $\hat{\omega}_c$ as follows:

$$\begin{aligned}&-\frac{\partial\left[(\nabla J_s(x))^{\mathsf{T}}(f(x)+g(x)\hat{u}^*(x))\right]}{\partial\hat{\omega}_c}\\ &=-\left(\frac{\partial\hat{u}^*(x)}{\partial\hat{\omega}_c}\right)^{\mathsf{T}}\frac{\partial\left[(\nabla J_s(x))^{\mathsf{T}}(f(x)+g(x)\hat{u}^*(x))\right]}{\partial\hat{u}^*(x)}\\ &=\frac{1}{2}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x).\end{aligned}$$

Therefore, the novel critic learning rule developed in this paper is formulated as

$$\dot{\hat{\omega}}_c = -\alpha_c\frac{\phi}{(1+\phi^{\mathsf{T}}\phi)^2}e_c + \frac{1}{2}\alpha_s\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x), \tag{11}$$

where $\alpha_s>0$ is the designed learning constant.

In what follows, we focus on building the error dynamics with respect to the critic network and investigating its stability. We define the error vector between the ideal weight and the estimated value as $\tilde{\omega}_c=\omega_c-\hat{\omega}_c$ and then find that $\dot{\tilde{\omega}}_c=-\dot{\hat{\omega}}_c$. By using the tuning rule (11) and introducing two new variables

$$\phi_1=\frac{\phi}{(1+\phi^{\mathsf{T}}\phi)}\in\mathbb{R}^{l_c},\;\;\phi_2=1+\phi^{\mathsf{T}}\phi,$$

we derive that the critic error dynamics can be simply formulated as

$$\dot{\tilde{\omega}}_c = -\alpha_c\phi_1\phi_1^{\mathsf{T}}\tilde{\omega}_c + \alpha_c\frac{\phi_1}{\phi_2}e_{cH} - \frac{1}{2}\alpha_s\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x), \tag{12}$$

where the term

$$e_{cH} = -(\nabla\varepsilon_c(x))^{\mathsf{T}}[f(x)+g(x)\hat{u}^*(x)]$$

stands for the residual error arisen in the neural-network-based approximation process.

For the adaptive critic design, the persistence of excitation assumption is required since we want to identify the parameter of the critic network to approximate the optimal cost function. According to Vamvoudakis and Lewis [25], the persistence of excitation condition ensures that $\lambda_{\min}(\phi_1\phi_1^{\mathsf{T}})>0$, which is significant to conduct the closed-loop stability analysis in what follows.

Now, the closed-loop stability incorporating the novel learning mechanism is discussed. Before proceeding, the following assumption is needed, as usually proposed in literature as [26,35].

Assumption 2. The control function matrix $g(x)$ is upper bounded such that $\|g(x)\|\le\lambda_g$, where λ_g is a positive constant. On the compact set Ω, the terms $\nabla\sigma_c(x)$, $\nabla\varepsilon_c(x)$, and e_{cH} are all upper bounded such that $\|\nabla\sigma_c(x)\|\le\lambda_\sigma$, $\|\nabla\varepsilon_c(x)\|\le\lambda_\varepsilon$, and $|e_{cH}|\le\lambda_e$, where λ_σ, λ_ε, and λ_e are positive constants.

Theorem 1. *For the nonlinear system (1), we suppose that Assumption 2 holds. The approximate optimal control law is given by (8), where the constructed critic network is tuned by adopting the improved rule given as (11). Then, the closed-loop system state and the critic weight estimation error satisfy uniformly ultimately bounded stability.*

Proof. Let us choose a Lyapunov function candidate formulated as

$$L_c(t) = L_{c1}(t) + L_{c2}(t),$$

where

$$L_{c1}(t) = \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}(t)\tilde{\omega}_c(t), L_{c2}(t) = \alpha_sJ_s(x(t)).$$

Taking the time derivative to the above Lyapunov function and according to (12), we have

$$\dot{L}_{c1}(t) = -\alpha_c \tilde{\omega}_c^{\mathsf{T}} \phi_1 \phi_1^{\mathsf{T}} \tilde{\omega}_c + \alpha_c \frac{\tilde{\omega}_c^{\mathsf{T}} \phi_1}{\phi_2} e_{cH} - \frac{1}{2} \alpha_s \tilde{\omega}_c^{\mathsf{T}} \nabla \sigma_c(x) g(x) R^{-1} \times g^{\mathsf{T}}(x) \nabla J_s(x). \tag{13}$$

Besides, the derivative of $L_{c2}(t)$ is

$$\dot{L}_{c2}(t) = \alpha_s (\nabla J_s(x))^{\mathsf{T}} [f(x) + g(x) \hat{u}^*(x)]. \tag{14}$$

For $\dot{L}_{c1}(t)$, we apply the Young's inequality to the second term of (13), i.e.,

$$\alpha_c \frac{\tilde{\omega}_c^{\mathsf{T}} \phi_1}{\phi_2} e_{cH} \le \frac{1}{2} \left(\tilde{\omega}_c^{\mathsf{T}} \phi_1 \phi_1^{\mathsf{T}} \tilde{\omega}_c + \alpha_c^2 \frac{e_{cH}^2}{\phi_2^2} \right),$$

recall Assumption 2 and the fact $\phi_2 \ge 1$, and then derive that

$$\dot{L}_{c1}(t) \le -\left(\alpha_c - \frac{1}{2} \right) \lambda_{\min}(\phi_1 \phi_1^{\mathsf{T}}) \|\tilde{\omega}_c\|^2 + \frac{1}{2} \alpha_c^2 \lambda_e^2 - \frac{1}{2} \alpha_s \tilde{\omega}_c^{\mathsf{T}} \nabla \sigma_c(x) g(x) R^{-1} g^{\mathsf{T}}(x) \nabla J_s(x). \tag{15}$$

Substituting $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$ to the last term of (15), we have

$$\begin{aligned} \dot{L}_{c1}(t) \le & -\left(\alpha_c - \frac{1}{2} \right) \lambda_{\min}(\phi_1 \phi_1^{\mathsf{T}}) \|\tilde{\omega}_c\|^2 + \frac{1}{2} \alpha_c^2 \lambda_e^2 \\ & - \frac{1}{2} \alpha_s (\nabla J_s(x))^{\mathsf{T}} g(x) R^{-1} g^{\mathsf{T}}(x) (\nabla \sigma_c(x))^{\mathsf{T}} \omega_c \\ & - \alpha_s (\nabla J_s(x))^{\mathsf{T}} g(x) \hat{u}^*(x). \end{aligned} \tag{16}$$

By combining (14) and (16), we can obtain that the overall time derivative of $L_c(t)$ is

$$\begin{aligned} & \dot{L}_c(t) \\ & \le -\left[\left(\alpha_c - \frac{1}{2} \right) \lambda_{\min}(\phi_1 \phi_1^{\mathsf{T}}) - \lambda_g^2 \lambda_\sigma^2 \right] \|\tilde{\omega}_c\|^2 + \lambda_g^2 \lambda_\varepsilon^2 + \frac{1}{2} \alpha_c^2 \lambda_e^2 \\ & \quad + \alpha_s (\nabla J_s(x))^{\mathsf{T}} f(x) - \frac{1}{2} \alpha_s (\nabla J_s(x))^{\mathsf{T}} g(x) R^{-1} g^{\mathsf{T}}(x) (\nabla \sigma_c(x))^{\mathsf{T}} \omega_c. \end{aligned} \tag{17}$$

Recalling the optimal control law in (7), we find that (17) becomes

$$\begin{aligned} \dot{L}_c(t) \le & -\left[\left(\alpha_c - \frac{1}{2} \right) \lambda_{\min}(\phi_1 \phi_1^{\mathsf{T}}) - \lambda_g^2 \lambda_\sigma^2 \right] \|\tilde{\omega}_c\|^2 + \lambda_g^2 \lambda_\varepsilon^2 + \frac{1}{2} \alpha_c^2 \lambda_e^2 \\ & + \alpha_s (\nabla J_s(x))^{\mathsf{T}} [f(x) + g(x) u^*(x)] \\ & + \frac{1}{2} \alpha_s (\nabla J_s(x))^{\mathsf{T}} g(x) R^{-1} g^{\mathsf{T}}(x) \nabla \varepsilon_c(x). \end{aligned} \tag{18}$$

In light of Assumptions 1 and 2, it follows from (18) that

$$\begin{aligned} \dot{L}_c(t) \le & -\left[\left(\alpha_c - \frac{1}{2} \right) \lambda_{\min}(\phi_1 \phi_1^{\mathsf{T}}) - \lambda_g^2 \lambda_\sigma^2 \right] \|\tilde{\omega}_c\|^2 + \lambda_g^2 \lambda_\varepsilon^2 + \frac{1}{2} \alpha_c^2 \lambda_e^2 \\ & - \alpha_s \lambda_{\min}(\Xi) \|\nabla J_s(x)\|^2 + \frac{1}{2} \alpha_s \lambda_g^2 \lambda_\varepsilon \|R^{-1}\| \|\nabla J_s(x)\|. \end{aligned} \tag{19}$$

Performing some basic mathematical operations, (19) can be written as

$$\begin{aligned} \dot{L}_c(t) \le & -\left[\left(\alpha_c - \frac{1}{2} \right) \lambda_{\min}(\phi_1 \phi_1^{\mathsf{T}}) - \lambda_g^2 \lambda_\sigma^2 \right] \|\tilde{\omega}_c\|^2 \\ & + \lambda_\Sigma - \alpha_s \lambda_{\min}(\Xi) \left[\|\nabla J_s(x)\| - \frac{1}{4 \lambda_{\min}(\Xi)} \lambda_g^2 \lambda_\varepsilon \|R^{-1}\| \right]^2, \end{aligned}$$

where the constant term is denoted by

$$\lambda_\Sigma = \lambda_g^2 \lambda_\varepsilon^2 + \frac{1}{2} \alpha_c^2 \lambda_e^2 + \frac{1}{16 \lambda_{\min}(\Xi)} \alpha_s \lambda_g^4 \lambda_\varepsilon^2 \|R^{-1}\|^2.$$

This comes to a conclusion that, if the inequality

$$\|\tilde{\omega}_c\| > \sqrt{\frac{2 \lambda_\Sigma}{(2\alpha_c - 1) \lambda_{\min}(\phi_1 \phi_1^{\mathsf{T}}) - 2 \lambda_g^2 \lambda_\sigma^2}} \triangleq \mathscr{B}_{\tilde{\omega}_c}$$

or

$$\|\nabla J_s(x)\| > \frac{1}{4 \lambda_{\min}(\Xi)} \lambda_g^2 \lambda_\varepsilon \|R^{-1}\| + \sqrt{\frac{\lambda_\Sigma}{\alpha_s \lambda_{\min}(\Xi)}} \triangleq \mathscr{B}_{J_{sx}}$$

holds, we can accomplish the goal of $\dot{L}_c(t) < 0$. Note that $J_s(x)$ is selected as a polynomial and according to the standard Lyapunov extension theorem [36], we further come to the result that the system state x and the critic weight error $\tilde{\omega}_c$ are uniformly ultimately bounded. This is the end of the proof. □

Remark 2. According to Theorem 1, we observe that the critic weight error $\tilde{\omega}_c$ is upper bounded by a finite constant such as $\|\tilde{\omega}_c\| \le \mathscr{B}_{\tilde{\omega}_c}$. Then, according to (7) and (8), we can clearly find that

$$\begin{aligned} \|u^*(x) - \hat{u}^*(x)\| & = \frac{1}{2} \left\| R^{-1} g^{\mathsf{T}}(x) \left[(\nabla \sigma_c(x))^{\mathsf{T}} \tilde{\omega}_c + \nabla \varepsilon_c(x) \right] \right\| \\ & \le \frac{1}{2} \|R^{-1}\| \lambda_g (\lambda_\sigma \mathscr{B}_{\tilde{\omega}_c} + \lambda_\varepsilon) \triangleq \mathscr{B}_{u^*}, \end{aligned}$$

which implies that, the approximate optimal control $\hat{u}^*(x)$ converges to a neighborhood of its optimal value $u^*(x)$ with a finite bound $\mathscr{B}_{u^*}$, where $\mathscr{B}_{u^*}$ is a positive constant.

4. Simulation verification

In this section, some experimental simulations are conducted to display the effectiveness of the improved adaptive optimal control method. Consider a continuous-time nonlinear system with the following form:

$$\dot{x} = \begin{bmatrix} -x_1 + x_2 \\ -0.5x_1 - 0.5x_2 \left[1 - (\cos(2x_1) + 2)^2 \right] \end{bmatrix} + \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix} u, \tag{20}$$

where $x = [x_1, x_2]^{\mathsf{T}}$, we aim to derive a feedback control law $u(x)$ to minimize the infinite horizon cost function given by

$$J(x_0) = \int_0^\infty \left\{ Q(x) + u^{\mathsf{T}} R u \right\} \mathrm{d}\tau$$

with $Q(x) = x^{\mathsf{T}} x$ and $R = I$.

We adopt the improved adaptive control algorithm to cope with the optimal regulation problem, where a critic network should be built to approximate the optimal cost function. We denote the weight variable of the neural network as $\hat{\omega}_c = [\hat{\omega}_{c1}, \hat{\omega}_{c2}, \hat{\omega}_{c3}]^{\mathsf{T}}$ and choose the activation function as $\sigma_c(x) = [x_1^2, x_2^2, x_1 x_2]^{\mathsf{T}}$. Additionally, we set the basic learning rate of the neural network as $\alpha_c = 5$ and select the initial state vector of the controlled nonlinear plant be $x_0 = [1, -1]^{\mathsf{T}}$.

During the implement process of the improved neural learning algorithm, we bring in a probing noise to guarantee the persistence of excitation condition. The system state must be persistently excited long enough so as to guarantee the constructed critic network to learn the optimal cost and also to ensure us to obtain the optimal control law as accurately as possible. For keeping the stability property, we introduce the additional stabilizing term and update the weight vector according to the improved learning rule (11), where $J_s(x)$ is chosen as $J_s(x) = 0.5x^{\mathsf{T}} x$. When selecting $\alpha_s = 0.001$, the weight of the critic network converges to $[0.4975, 1.0013, 0.0014]^{\mathsf{T}}$ as shown in Fig. 1. Obviously, we see that the convergence of the weight elements has occurred at 550 s, so that the probing signal can be turned off after that. The evolution of the corresponding state trajectory is depicted in Fig. 2, which displays the adjustment trend in the neural network learning session.

Using the above converged weight and according to (6) and (8), the approximate optimal cost function and the approximate

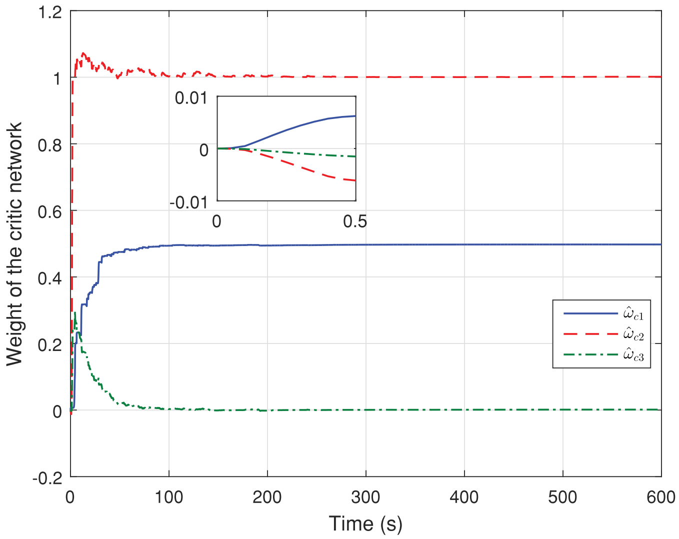


Fig. 1. Convergence of the weight vector when setting $\alpha_s = 0.001$.

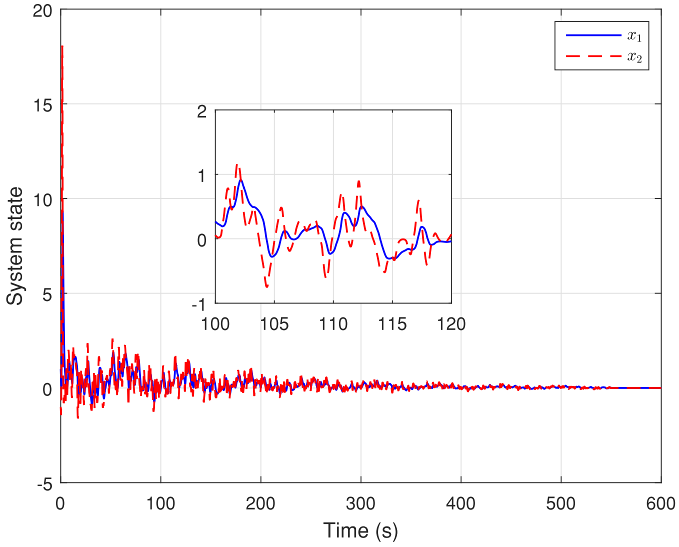


Fig. 2. State trajectories in the learning session.

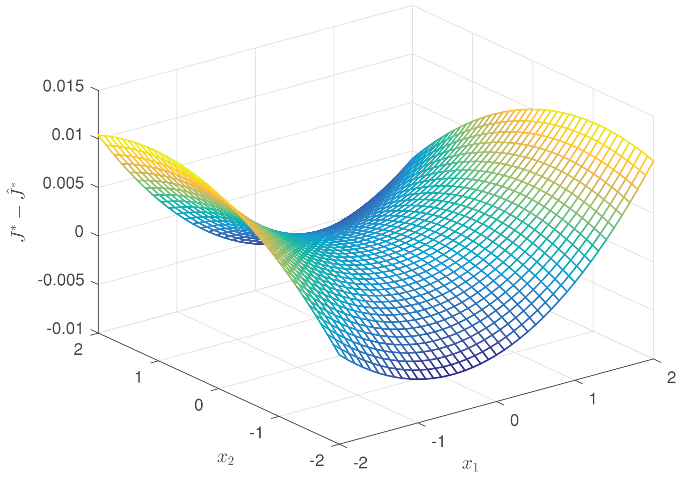


Fig. 3. 3D view of the approximation error of the cost function.

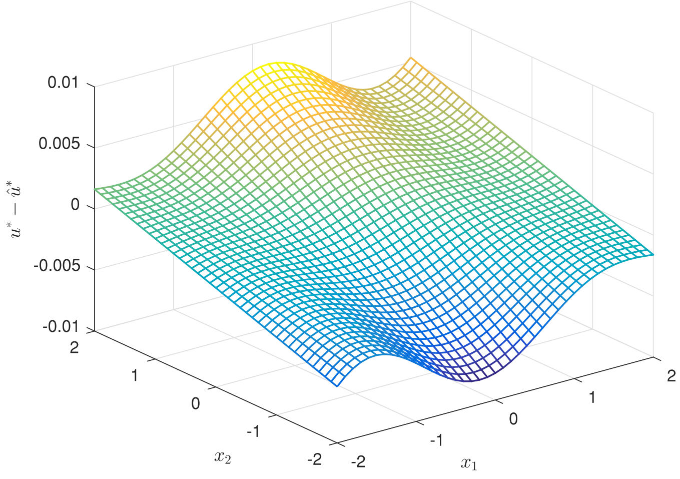


Fig. 4. 3D view of the approximation error of the control input.

optimal control law can be expressed by

$$\hat{J}^*(x) = \begin{bmatrix}0.4975\\1.0013\\0.0014\end{bmatrix}^{\mathsf{T}} \begin{bmatrix} x_1^2\\ x_2^2\\ x_1x_2\end{bmatrix}$$

and

$$\hat{u}^*(x) = -\frac{1}{2}R^{-1}\begin{bmatrix}0\\ \cos(2x_1)+2\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}2x_1 & 0\\ 0 & 2x_2\\ x_2 & x_1\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}0.4975\\1.0013\\0.0014\end{bmatrix},$$

respectively.

For the controlled nonlinear system with the given special form, using the similar strategy given in [25], the optimal cost function as well as the optimal control law are $J^*(x) = 0.5x_1^2 + x_2^2$ and $u^*(x) = -[\cos(2x_1)+2]x_2$, respectively. In this sense, the optimal weight vector should be $[0.5, 1, 0]^{\mathsf{T}}$. Hence, the converged weight $[0.4975, 1.0013, 0.0014]^{\mathsf{T}}$ possesses an excellent approximation ability. Moreover, we can plot the error illustration between the optimal cost and the approximate one as indicated in Fig. 3. Similarly, the error of the optimal control law compared with the approximate state feedback law is exhibited in Fig. 4. We can observe that the two approximation errors are pretty close to zero, which shows a satisfying approximate ability of the neural-network-based learning algorithm.

For further showing the action of the stabilizing term, we choose different parameters to observe the convergence process of the critic weight vector. When we set $\alpha_s = 0.01$ and still use $J_s(x) = 0.5x^{\mathsf{T}}x$, the weight of the critic network gradually converges to $[0.4763, 1.0133, 0.0130]^{\mathsf{T}}$ as shown in Fig. 5. If we continue to enlarge this parameter, the convergence ability becomes bad. For instance, when choosing $\alpha_s = 0.1$ and using $J_s(x) = 0.5x^{\mathsf{T}}x$, the weight vector of the critic network converges to $[0.2656, 1.1288, 0.1057]^{\mathsf{T}}$, which is exhibited in Fig. 6. Although the weight vectors of Figs. 1, 5, and 6 converge to different values, there exists a common property, i.e., the weights are all modulated from a zero vector. This illustrates the fact that the initial stabilizing control law is indeed not required under the improved adaptive critic control design.

The state curves of the first 20 s with respect to the above four cases are illustrated in Fig. 7. Therein, the four curves show the state trajectories obtained by the action of the four different control laws. The solid line represents the state curve by applying the approximate optimal control $\hat{u}^*$ derived by the converged weight $[0.4975, 1.0013, 0.0014]^{\mathsf{T}}$. The dash line stands for the state curve by employing the optimal control u^* with the ideal weight $[0.5, 1, 0]^{\mathsf{T}}$. The dash-dot line presents the state curve by using the approximate optimal control $\hat{u}^1$ derived from the weight vector

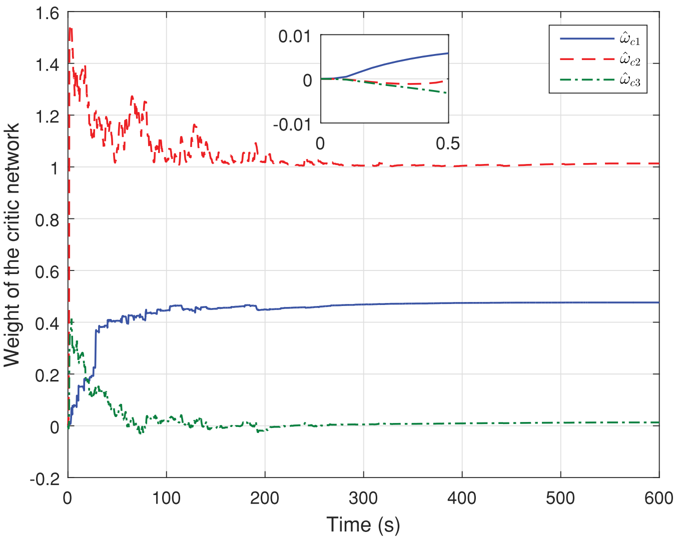


Fig. 5. Convergence of the weight vector when setting $\alpha_s = 0.01$.

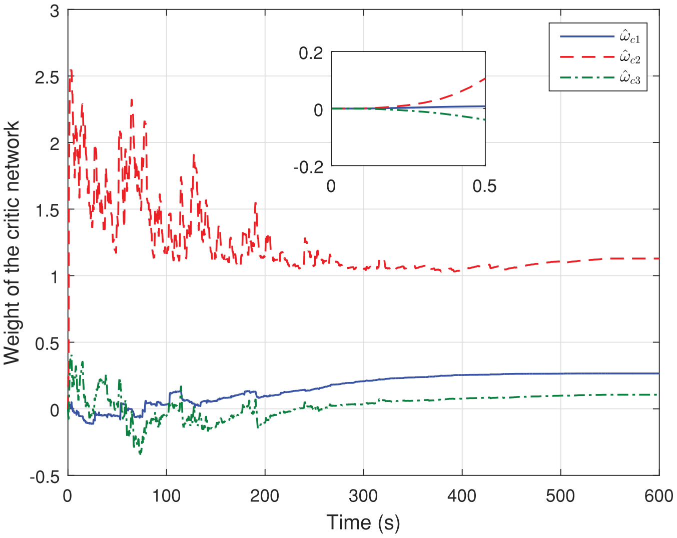


Fig. 6. Convergence of the weight vector when setting $\alpha_s = 0.1$.

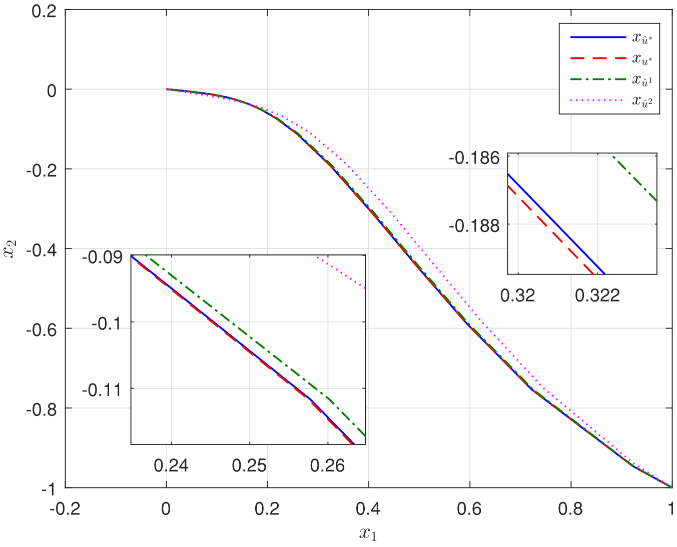


Fig. 7. State trajectories by adopting four difference control laws.

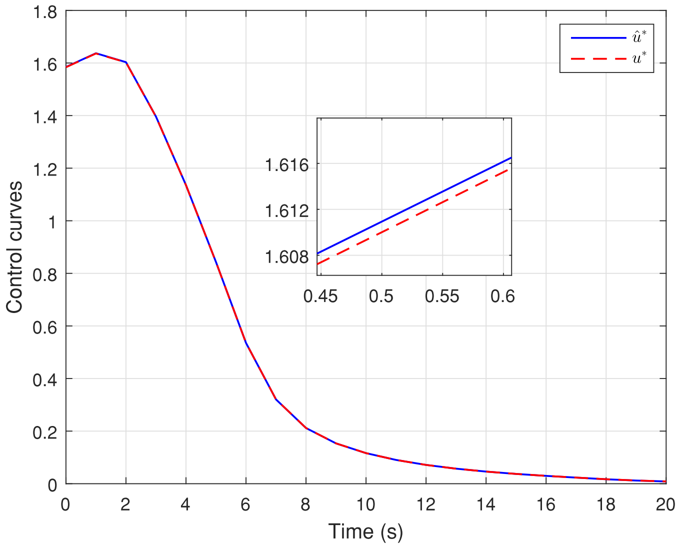


Fig. 8. Trajectories obtained from the optimal control law and approximate optimal control when using $\alpha_s = 0.001$.

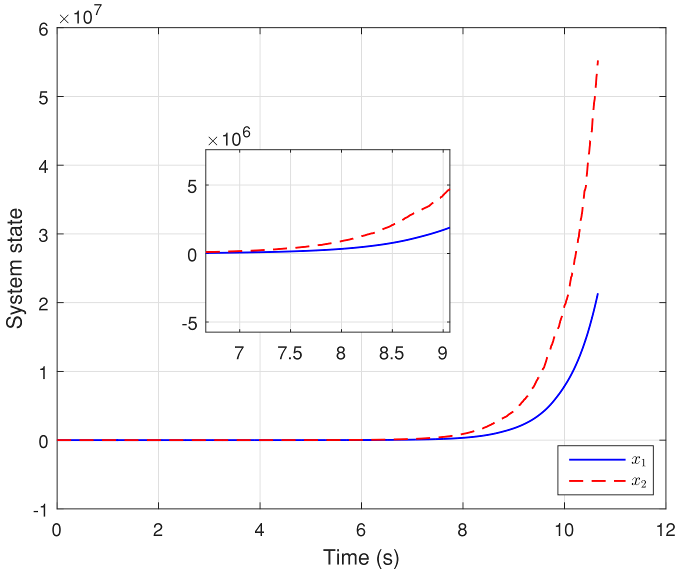


Fig. 9. State trajectories possessing divergent property without using the stabilizing term.

$[0.4763, 1.0133, 0.0130]^\mathsf{T}$. At last, the dot line shows the state curve by adopting the approximate optimal control $\hat{u}^2$ obtained with the weight vector $[0.2656, 1.1288, 0.1057]^\mathsf{T}$, which clearly, does not have satisfying performance.

From these comparison results, we prove that the weight vector obtained by using $\alpha_s = 0.001$ holds the best convergence trend among the three approximate values. The corresponding state trajectory has almost the same evolution as the curve derived by the optimal control law $u^*(x)$. Besides, the optimal control and approximate optimal control of the first 20 s when using $\alpha_s = 0.001$ are shown in Fig. 8. These two trajectories are also nearly the same with each other. Therefore, we come to a conclusion that $\alpha_s = 0.001$ is a very suitable choice of this experimental example.

Finally, we show the simulation result of removing the additional stabilizing term, i.e., setting $\alpha_s = 0$. The state trajectory possesses divergent property quickly as time goes on, which is displayed in Fig. 9. It means that, the approximate state feedback derived from the traditional learning algorithm is unable to control the plant expectedly, which firmly demonstrates the importance of the stabilizing term. However, we can conclude from the simulation process that the parameter related to the stabilizing term should not be chosen too large as well. Consequently, it is a parameter that must be selected properly during the adaptive control implementation. It should not be vanished completely and also is undesirable to set too large. The engineering experience is required and is also constructive to achieve a satisfying option.

All the aforementioned simulation results verify the effectiveness of the improved adaptive optimal feedback control strategy derived in this paper. Incidentally, the simulation plant (20) just represents a few nonlinear dynamical systems, where the optimal control law can be obtained only for the comparison purpose. Actually, the present method is particularly beneficial to design adaptive optimal control for nonlinear systems with more general form. In such situation, it is difficult to find optimal control laws in advance, hence it is considerably important to derive approximate (and adaptive) optimal regulators.

5. Conclusions

The adaptive optimal state feedback control design of nonlinear dynamical systems is studied with an improved adaptive critic structure. The approximate optimal control law is derived by training a critic network based on the new learning rule. The stability proof of the closed-loop system and the experimental verification of dynamical systems are carried out. The future work contains how to reduce the requirement with respect to the system dynamics, and then to develop more advanced adaptive optimal control techniques for general nonlinear systems (e.g., uncertain nonlinear systems) through the improved neural learning mechanism.

References

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

[2] K.G. Vamvoudakis, P.J. Antsaklis, W.E. Dixon, J.P. Hespanha, F.L. Lewis, H. Modares, B. Kiumarsi, Autonomy and machine intelligence in complex systems: a tutorial, in: Proceedings of the American Control Conference, Chicago, IL, USA, 2015, pp. 5062–5079.

[3] D. Wang, C. Mu, D. Liu, Data-driven nonlinear near-optimal regulation based on iterative neural dynamic programming, Acta Autom. Sin. 43 (2017) 366–375.

[4] X. Zhong, H. He, H. Zhang, Z. Wang, A neural network based online learning and control approach for Markov jump systems, Neurocomputing 149 (2015) 116–123.

[5] P.J. Werbos, Approximate dynamic programming for real-time control and neural modeling, in: D.A. White, D.A. Sofge (Eds.), Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches, Van Nostrand Reinhold, New York, NY, 1992. Ch. 13

[6] C. Mu, D. Wang, H. He, Novel iterative neural dynamic programming for data-based approximate optimal control design, Automatica 81 (2017) 240–252.

[7] C. Mu, Z. Ni, C. Sun, H. He, Air-breathing hypersonic vehicle tracking control based on adaptive dynamic programming, IEEE Trans. Neural Networks Learn. Syst. 28 (2017) 584–598.

[8] D. Wang, D. Liu, Q. Zhang, D. Zhao, in: Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics, IEEE Trans. Syst. Man Cybern. Syst. 46 (2016) 1544–1555.

[9] D. Wang, H. He, C. Mu, D. Liu, Intelligent critic control with disturbance attenuation for affine dynamics including an application to a micro-grid system, IEEE Trans. Ind. Electron. 64 (2017) 4935–4944.

[10] M. Chen, G. Tao, Adaptive fault-tolerant control of uncertain nonlinear large-scale systems with unknown dead zone, IEEE Trans. Cybern. 46 (2016) 1851–1862.

[11] Y. Wang, L. Cheng, Z.G. Hou, J. Yu, M. Tan, Optimal formation of multi-robot systems based on a recurrent neural network, IEEE Trans. Neural Netw. Learn. Syst. 27 (2016) 322–333.

[12] C. Li, J. Gao, J. Yi, G. Zhang, Analysis and design of functionally weighted single-input-rule-modules connected fuzzy inference systems, IEEE Trans. Fuzzy Syst. doi:10.1109/TFUZZ.2016.2637369

[13] W. He, Y. Dong, C. Sun, Adaptive neural impedance control of a robotic manipulator with input saturation, IEEE Trans. Syst. Man Cybern. Syst. 46 (2016) 334–344.

[14] T. Wang, S. Tong, Observer-based output-feedback asynchronous control for switched fuzzy systems, IEEE Trans. Cybern. doi:10.1109/TCYB.2016.2558821

[15] B. Xu, Robust adaptive neural control of flexible hypersonic flight vehicle with dead-zone input nonlinearity, Nonlinear Dyn. 80 (2015) 1509–1520.

[16] Y.J. Liu, S. Tong, C.L.P. Chen, D.J. Li, Neural controller design-based adaptive control for nonlinear MIMO systems with unknown hysteresis inputs, IEEE Trans. Cybern. 46 (2016) 9–19.

[17] T. Dierks, B.T. Thumati, S. Jagannathan, Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence, Neural Netw. 22 (2009) 851–860.

[18] A. Heydari, S.N. Balakrishnan, Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics, IEEE Trans. Neural Netw. Learn. Syst. 24 (2013) 145–157.

[19] R. Song, Q. Wei, Q. Sun, Nearly finite-horizon optimal control for a class of nonaffine time-delay nonlinear systems based on adaptive dynamic programming, Neurocomputing 156 (2015) 166–175.

[20] Q. Zhao, H. Xu, S. Jagannathan, Near optimal output feedback control of nonlinear discrete-time systems based on reinforcement neural network learning, IEEE/CAA J. Autom. Sin. 1 (2014) 372–384.

[21] C. Mu, C. Sun, A. Song, H. Yu, Iterative GDHP-based approximate optimal tracking control for a class of discrete-time nonlinear systems, Neurocomputing 214 (2016) 775–784.

[22] D. Zhao, Z. Xia, D. Wang, Model-free optimal control for affine nonlinear systems with convergence analysis, IEEE Trans. Autom. Sci. Eng. 12 (2015) 1461–1468.

[23] C. Qin, H. Zhang, Y. Wang, Y. Luo, Neural network-based online h_∞ control for discrete-time affine nonlinear system using adaptive dynamic programming, Neurocomputing 198 (2016) 91–99.

[24] X. Zhong, Z. Ni, H. He, A theoretical foundation of goal representation heuristic dynamic programming, IEEE Trans. Neural Netw. Learn. Syst. 27 (2016) 2513–2525.

[25] K.G. Vamvoudakis, F.L. Lewis, Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem, Automatica 46 (2010) 878–888.

[26] H. Zhang, L. Cui, Y. Luo, Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP, IEEE Trans. Cybern. 43 (2013) 206–216.

[27] H. Modares, F.L. Lewis, Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning, IEEE Trans. Autom. Control 59 (2014) 3051–3056.

[28] J. Na, G. Herrmann, Online adaptive approximate optimal tracking control with simplified dual approximation structure for continuous-time unknown nonlinear systems, IEEE/CAA J. Autom. Sin. 1 (2014) 412–422.

[29] T. Bian, Y. Jiang, Z.P. Jiang, Decentralized adaptive optimal control of large-scale systems with application to power systems, IEEE Trans. Ind. Electron. 62 (2015) 2439–2447.

[30] Y. Jiang, Z.P. Jiang, Global adaptive dynamic programming for continuous-time nonlinear systems, IEEE Trans. Autom. Control 60 (2015) 2917–2929.

[31] B. Luo, H.N. Wu, T. Huang, D. Liu, Reinforcement learning solution for HJB equation arising in constrained optimal control problem, Neural Netw. 71 (2015) 150–158.

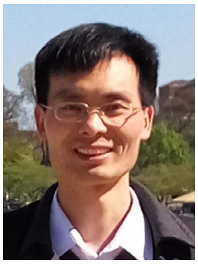
[32] W. Gao, Z.P. Jiang, Adaptive dynamic programming and adaptive optimal output regulation of linear systems, IEEE Trans. Autom. Control 61 (2016) 4164–4169.

[33] T. Dierks, S. Jagannathan, Optimal control of affine nonlinear continuous-time systems, in: Proceedings of the American Control Conference, Baltimore, MD, USA, 2010, pp. 1568–1573.

[34] D. Nodland, H. Zargarzadeh, S. Jagannathan, Neural network-based optimal adaptive output feedback control of a helicopter UAV, IEEE Trans. Neural Netw. Learn. Syst. 24 (2013) 1061–1073.

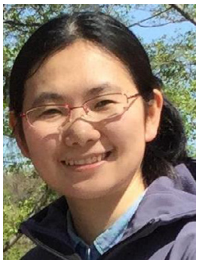
[35] X. Yang, D. Liu, Q. Wei, D. Wang, Guaranteed cost neural tracking control for a class of uncertain nonlinear systems using adaptive dynamic programming, Neurocomputing 198 (2016) 80–90.

[36] F.L. Lewis, S. Jagannathan, A. Yesildirek, Neural Network Control of Robot Manipulators and Nonlinear Systems, Taylor & Francis, London, 1999.



Ding Wang received the B.S. degree in mathematics from 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 Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007, 2009, and 2012, respectively. He was a Visiting Scholar with the Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI, USA, from December 2015 to January 2017. 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 research interests include adaptive and learning systems, computational intelligence, and intelligent control. He has published over 90 journal and conference papers, and coauthored two monographs. He is the

Publications Chair of the 24th International Conference on Neural Information Processing (ICONIP 2017). He was the Finance Chair of the 12th World Congress on Intelligent Control and Automation (WCICA 2016), the Secretariat of the 2014 IEEE World Congress on Computational Intelligence (IEEE WCCI 2014), and the Registration Chair of the 5th International Conference on Information Science and Technology (ICIST 2015) and the 4th International Conference on Intelligent Control and Information Processing (ICICIP 2013), and served as the program committee member of several international conferences. He was a recipient of the Excellent Doctoral Dissertation Award of Chinese Academy of Sciences in 2013, and a nomination of the Excellent Doctoral Dissertation Award of Chinese Association of Automation (CAA) in 2014. He serves as an Associate Editor of IEEE Transactions on Neural Networks and Learning Systems and Neurocomputing. He is a member of Institute of Electrical and Electronics Engineers (IEEE), Asia-Pacific Neural Network Society (APNNS), and CAA.



Chaoxu Mu received the Ph.D. degree in control science and engineering from Southeast University, Nanjing, China, in 2012. She was a Visiting Ph.D. Student with the Royal Melbourne Institute of Technology University, Melbourne, VIC, Australia, from 2010 to 2011. She has been a Post-Doctoral Fellow with the Department of Electrical, Computer and Biomedical Engineering, The University of Rhode Island, Kingston, RI, USA, from 2014 to 2016. She is currently an Associate Professor with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. She has authored over 50 journal and conference papers, and co-authored one monograph. Her current research interests include nonlinear system control and optimization, adaptive and learning systems, and machine learning. She is the IEEE member and a member of the Asia-Pacific Neural Network Society.