Information Sciences 282 (2014) 167–179

Contents lists available at ScienceDirect

Information Sciences

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

Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming ☆

Ding Wang, Derong Liu *, Hongliang Li, Hongwen Ma

The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 24 December 2013
Received in revised form 13 May 2014
Accepted 21 May 2014
Available online 11 June 2014

Keywords:
Adaptive dynamic programming
Approximate dynamic programming
Decentralized optimal control
Neural network
Robust optimal control

ABSTRACT

In this paper, the neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming approach is investigated. First, the robust controller of the original uncertain system is derived by adding a feedback gain to the optimal controller of the nominal system. It is also shown that this robust controller can achieve optimality under a specified cost function, which serves as the basic idea of the robust optimal control design. Then, a critic network is constructed to solve the Hamilton–Jacobi–Bellman equation corresponding to the nominal system, where an additional stabilizing term is introduced to verify the stability. The uniform ultimate boundedness of the closed-loop system is also proved by using the Lyapunov approach. Moreover, the obtained results are extended to solve decentralized optimal control problem of continuous-time nonlinear interconnected large-scale systems. Finally, two simulation examples are presented to illustrate the effectiveness of the established control scheme.

1. Introduction

In practical control systems, model uncertainties arise frequently and can severely degrade the closed-loop system performance. Hence, the problem of designing robust controller for nonlinear systems with uncertainties has drawn considerable attention in recent literature [43,15,31]. Lin et al. [15] showed that the robust control problem can be solved by studying the optimal control problem of the corresponding nominal system, but the detailed procedure was not presented. In [31], the authors developed an iterative algorithm for online design of robust control for a class of continuous-time nonlinear systems. However, the optimality of the robust controller with respect to a specified cost function was not discussed. In [43], the authors addressed the problem of designing robust tracking controls for a class of uncertain nonholonomic systems actuated by brushed direct current motors, while the research was not related with the optimality.

The starting point of the obtained strategy of this paper is optimal control. The nonlinear optimal control problem always requires to solve the Hamilton–Jacobi–Bellman (HJB) equation. Though dynamic programming has been a conventional method in solving optimization and optimal control problems, it often suffers from the curse of dimensionality, which was primarily due to the backward-in-time approach. To avoid the difficulty, based on function approximators, such as neural networks, adaptive/approximate dynamic programming (ADP) was proposed by Werbos [35] as a method to solve

☆ This work was supported in part by the National Natural Science Foundation of China under Grants 61034002, 61233001, 61273140, 61304086, and 61374105, in part by Beijing Natural Science Foundation under Grant 4132078, and in part by the Early Career Development Award of SKLMCCS.
* Corresponding author. Tel.: +86 10 82544761; fax: +86 10 82544799.
E-mail addresses: ding.wang@ia.ac.cn (D. Wang), derong.liu@ia.ac.cn (D. Liu), hongliang.li@ia.ac.cn (H. Li), mahongwen2012@ia.ac.cn (H. Ma).

http://dx.doi.org/10.1016/j.ins.2014.05.050
0020-0255/

optimal control problems forward-in-time. Recently, the study on ADP and related fields have gained much attention from various scholars [1–10,12–14,16–25,28–30,32–34,36–38,40–42,44–46]. Lewis and Vrabie [13] stated that the ADP technique is closely related to the field of reinforcement learning. As is known to all, policy iteration is one of the basic algorithms of reinforcement learning. In addition, the initial admissible control is necessary when employing the policy iteration algorithm. However, in many situations, it is difficult to find the initial admissible control.

To the best of our knowledge, there are few results on robust optimal control of uncertain nonlinear systems based on ADP, not to mention the decentralized optimal control of large-scale systems. This is the motivation of our research. Actually, it is the first time that the robust optimal control scheme for a class of uncertain nonlinear systems via ADP technique and without using an initial admissible control is established. To begin with, the optimal controller of the nominal system is designed. It can be proved that the modification of optimal control law is in fact the robust controller of the original uncertain system, which also achieves optimality under the definition of a cost function. Then, a critic network is constructed for solving the HJB equation corresponding to the nominal system. In addition, inspired by the work of [5,24], an additional stabilizing term is introduced to verify the stability, which relaxes the need for an initial stabilizing control. The uniform ultimate boundedness (UUB) of the closed-loop system is also proved via the Lyapunov approach. Furthermore, the aforementioned results are extended to deal with the decentralized optimal control for a class of continuous-time nonlinear interconnected systems. At last, two simulation examples are given to show the effectiveness of the robust optimal control scheme.

2. Problem statement and preliminaries

In this paper, we study the continuous-time uncertain nonlinear systems given by

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

where $x(t) \in \mathbb{R}^n$ is the state vector and $\bar{u}(t) \in \mathbb{R}^m$ is the control vector, $f(\cdot)$ and $g(\cdot)$ are differentiable in their arguments with $f(0) = 0$, and $\bar{d}(x)$ is the unknown nonlinear perturbation. Let $x(0) = x_0$ be the initial state. We assume that $\bar{d}(0) = 0$, so that $x = 0$ is an equilibrium of system (1). As in many other literature, for the nominal system

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

we also assume that $f + gu$ is Lipschitz continuous on a set Ω in $\mathbb{R}^n$ containing the origin and that system (2) is controllable.

For system (1), in order to deal with the robust control problem, we should find a feedback control policy $\bar{u}(x)$, such that the closed-loop system is globally asymptotically stable for all uncertainties $\bar{d}(x)$. In this paper, we will show that this problem can be converted into designing an optimal controller for the corresponding nominal system with appropriate cost function.

Let $R \in \mathbb{R}^{m\times m}$ be a symmetric positive definite matrix. Then, we denote $d(x) = R^{1/2}\bar{d}(x)$, where $d(x) \in \mathbb{R}^m$ is bounded by a known function $d_M(x)$, i.e., $\|d(x)\| \leqslant d_M(x)$ with $d_M(0) = 0$. For system (2), in order to deal with the infinite horizon optimal control problem, we have to find the control policy $u(x)$, which minimizes the cost function given by

$$J(x_0) = \int_0^\infty \left\{ d_M^2(x(\tau)) + u^{\mathsf{T}}(x(\tau))Ru(x(\tau)) \right\} \mathrm{d}\tau. \tag{3}$$

Based on optimal control theory, the designed feedback control must not only stabilize the system on Ω, but also guarantee that the cost function (3) is finite. In other words, the control policy must be admissible [1,28]. Let $\Psi(\Omega)$ be the set of admissible controls on Ω. For any admissible control policy $u \in \Psi(\Omega)$, if the associated cost function (3) is continuously differentiable, then its infinitesimal version is the nonlinear Lyapunov equation given by

$$0 = d_M^2(x) + u^{\mathsf{T}}(x)Ru(x) + (\nabla J(x))^{\mathsf{T}}(f(x) + g(x)u(x)), \tag{4}$$

with $J(0) = 0$. In Eq. (4), the symbol $\nabla(\cdot) \triangleq \partial(\cdot)/\partial x$ is the notation of gradient operator, for example, $\nabla J(x) = \partial J(x)/\partial x$.

Define the Hamiltonian function of system (2) as follows:

$$H(x, u, \nabla J(x)) = d_M^2(x) + u^{\mathsf{T}}(x)Ru(x) + (\nabla J(x))^{\mathsf{T}}(f(x) + g(x)u(x)). \tag{5}$$

The optimal cost function of system (2) can be formulated as

$$J^*(x_0) = \min_{u\in\Psi(\Omega)} \int_0^\infty \left\{ d_M^2(x(\tau)) + u^{\mathsf{T}}(x(\tau))Ru(x(\tau)) \right\} \mathrm{d}\tau. \tag{6}$$

According to optimal control theory, the optimal cost function $J^*(x)$ satisfies the HJB equation

$$0 = \min_{u\in\Psi(\Omega)} H(x, u, \nabla J^*(x)). \tag{7}$$

Assume that the minimum on the right hand side of (7) exists and is unique. Then, the optimal control policy is

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

Based on (5) and (8), the HJB Eq. (7) becomes

$$0 = d_M^2(x) + (\nabla J^*(x))^{\mathsf T} f(x) - \frac{1}{4}(\nabla J^*(x))^{\mathsf T} g(x) R^{-1} g^{\mathsf T}(x) \nabla J^*(x), \tag{9}$$

with $J^*(0) = 0$.

Consider system (2) with cost function (3) and the optimal feedback control (8). The following assumption is presented for the robust optimal control design.

Assumption 1. Let $J_s(x)$ be a continuously differentiable Lyapunov function candidate satisfying

$$\dot J_s(x) = (\nabla J_s(x))^{\mathsf T} \dot x = (\nabla J_s(x))^{\mathsf T} (f(x) + g(x)u^*) < 0. \tag{10}$$

Assume there exists a positive definite matrix $\Lambda(x)$ such that the following relation holds:

$$(\nabla J_s(x))^{\mathsf T} (f(x) + g(x)u^*) = -(\nabla J_s(x))^{\mathsf T} \Lambda(x) \nabla J_s(x). \tag{11}$$

Remark 1. This is a common assumption that has been used in some literature, for instance [5,24], to facilitate discussing the stability issue of closed-loop system. According to [5], we assume that the closed-loop dynamics with optimal control can be bounded by a function of system state. Without loss of generality, we assume that $\|f(x) + g(x)u^*\| \leqslant \eta \|\nabla J_s(x)\|$ with $\eta > 0$. Hence, we can further obtain $\|(\nabla J_s(x))^{\mathsf T}(f(x) + g(x)u^*)\| \leqslant \eta \|\nabla J_s(x)\|^2$. Let λ_m and λ_M be the minimum and maximum eigenvalues of matrix $\Lambda(x)$. Considering $(\nabla J_s(x))^{\mathsf T}(f(x) + g(x)u^*) < 0$ and the fact that $\lambda_m \|\nabla J_s(x)\|^2 \leqslant (\nabla J_s(x))^{\mathsf T} \Lambda(x) \nabla J_s(x) \leqslant \lambda_M \|\nabla J_s(x)\|^2$, we can conclude that Assumption 1 is reasonable. Specifically, in this paper, $J_s(x)$ can be obtained by properly selecting a quadratic polynomial.

3. Robust optimal control design of uncertain nonlinear systems

In this section, for establishing the robust stabilizing control strategy of system (1), we modify the optimal control law (8) of system (2) by proportionally increasing a feedback gain, i.e.,

$$\bar u(x) = \zeta u^*(x) = -\frac{1}{2}\zeta R^{-1} g^{\mathsf T}(x) \nabla J^*(x). \tag{12}$$

Now, we present the following lemma to indicate that the optimal control has infinite gain margin.

Lemma 1. *For system (2), the feedback control given by (12) ensures that the closed-loop system is asymptotically stable for all* $\zeta \geqslant 1/2$.

Proof. We show that $J^*(x)$ is a Lyapunov function. In light of (6), we can easily find that $J^*(x)$ is positive definite. Considering (9) and (12), the derivative of $J^*(x)$ along the trajectory of the closed-loop system is

$$\begin{aligned} \dot J^*(x) &= (\nabla J^*(x))^{\mathsf T} (f(x) + g(x)\bar u(x)) \\ &= -d_M^2(x) - \frac{1}{2}\left(\zeta - \frac{1}{2}\right) \left\| R^{-1/2} g^{\mathsf T}(x) \nabla J^*(x) \right\|^2. \end{aligned} \tag{13}$$

Hence, $\dot J^*(x) < 0$ whenever $\zeta \geqslant 1/2$ and $x \neq 0$. Then, the conditions for Lyapunov local stability theory are satisfied. □

Theorem 1. *For system (1), there exists a positive number* $\zeta_1^* \geqslant 1$*, such that for any* $\zeta > \zeta_1^*$*, the feedback control developed by (12) ensures that the closed-loop system is asymptotically stable.*

Proof. We select $\bar L(x) = J^*(x)$ as the Lyapunov function candidate. Taking the time derivative of $\bar L(x)$ along the trajectory of the closed-loop system, we obtain

$$\dot{\bar L}(x) = (\nabla J^*(x))^{\mathsf T} (f(x) + g(x)(\bar u(x) + \bar d(x))). \tag{14}$$

Based on (13), we find that

$$\dot{\bar L}(x) \leqslant -\left\{ d_M^2(x) + \frac{1}{2}\left(\zeta - \frac{1}{2}\right) \left\| (\nabla J^*(x))^{\mathsf T} g(x) R^{-1/2} \right\|^2 - \left\| (\nabla J^*(x))^{\mathsf T} g(x) R^{-1/2} \right\| d_M(x) \right\}. \tag{15}$$

Let $\xi = \left[d_M(x), \left\| (\nabla J^*(x))^{\mathsf T} g(x) R^{-1/2} \right\| \right]^{\mathsf T}$. Then, we have $\dot{\bar L}(x) \leqslant -\xi^{\mathsf T} \Theta \xi$, where

$$\Theta = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2}\left(\zeta - \frac{1}{2}\right) \end{bmatrix}. \tag{16}$$

From (16), we observe that there exists a positive number $\zeta_1^* \geqslant 1$ such that any $\zeta > \zeta_1^*$ can guarantee the positive definiteness of Θ. Then, we have $\dot{L}(x) < 0$, which implies that the closed-loop system is asymptotically stable. □

According to Theorem 1, $\bar{u}(x)$ is the robust control strategy of the original system (1). Next, we will show that it also possesses the property of optimality.

For system (1), we define the following cost function

$$\bar{J}(x_0) = \int_0^\infty \left\{ Q(x(\tau)) + \frac{1}{\zeta}\bar{u}^\mathsf{T}(x(\tau))R\bar{u}(x(\tau)) \right\} \mathrm{d}\tau, \tag{17}$$

where

$$Q(x) = d_M^2(x) - (\nabla J^*(x))^\mathsf{T} g(x)\bar{d}(x) + \frac{1}{4}(\zeta - 1)(\nabla J^*(x))^\mathsf{T} g(x)R^{-1}g^\mathsf{T}(x)\nabla J^*(x). \tag{18}$$

Lemma 2. *There exists a positive number $\zeta_2^* \geqslant 2$ such that for all $\zeta > \zeta_2^*$, the function $Q(x)$ is positive definite.*

Proof. Adding and subtracting $(1/(\zeta - 1))d^\mathsf{T}(x)d(x)$ to (18), we find that

$$Q(x) = d_M^2(x) + \frac{1}{4(\zeta - 1)}\Big((\zeta - 1)(\nabla J^*(x))^\mathsf{T} g(x)R^{-1/2} - 2d^\mathsf{T}(x)\Big)\Big((\zeta - 1)(\nabla J^*(x))^\mathsf{T} g(x)R^{-1/2} - 2d^\mathsf{T}(x)\Big)^\mathsf{T} - \frac{1}{\zeta - 1}d^\mathsf{T}(x)d(x). \tag{19}$$

Recalling $\|d(x)\| \leqslant d_M(x)$and $\zeta > 2$, we can obtain

$$Q(x) \geqslant d_M^2(x) - \frac{1}{\zeta - 1}d^\mathsf{T}(x)d(x) \geqslant \frac{\zeta - 2}{\zeta - 1}d_M^2(x). \tag{20}$$

This proves that $Q(x)$ is a positive definite function. □

Theorem 2. *Consider system (1) with cost function (17). There exists a positive number ζ^* such that for any $\zeta > \zeta^*$, the feedback control law obtained by (12) is an asymptotically stabilizing solution of the optimal control problem.*

Proof. The Hamiltonian function of system (1) with cost function (17) is

$$\overline{H}(\nabla\bar{J}(x)) = Q(x) + \frac{1}{\zeta}\bar{u}^\mathsf{T}(x)R\bar{u}(x) + (\nabla\bar{J}(x))^\mathsf{T}(f(x) + g(x)(\bar{u}(x) + \bar{d}(x))), \tag{21}$$

where $\zeta > \zeta_2^* \geqslant 2$. Replacing $\bar{J}(x)$ with $J^*(x)$ and observing (18), the Eq. (21) becomes

$$\overline{H}(\nabla J^*(x)) = d_M^2(x) + (\nabla J^*(x))^\mathsf{T} f(x) + \frac{1}{4}(\zeta - 1)(\nabla J^*(x))^\mathsf{T} g(x)R^{-1}g^\mathsf{T}(x)\nabla J^*(x) + \frac{1}{\zeta}\bar{u}^\mathsf{T}(x)R\bar{u}(x) + (\nabla J^*(x))^\mathsf{T} g(x)\bar{u}(x). \tag{22}$$

Using (9) and (12), we can further obtain that $\overline{H}(\nabla J^*(x)) = 0$, which shows that $J^*(x)$ is a solution of the HJB equation of system (1). Then, we say that the control law (12) achieves optimality with cost function (17). Furthermore, there exists a positive number $\zeta^* \triangleq \max\{\zeta_1^*, \zeta_2^*\}$ such that for any $\zeta > \zeta^*$, the control law (12) is an asymptotically stabilizing solution of the corresponding optimal control problem. □

Remark 2. Based on Theorems 1 and 2, there exists $\zeta > \zeta^*$ such that the control law (12) cannot only stabilize system (1), but also achieve optimality with the defined cost function. That is to say, for a fixed $\zeta > \zeta^*$, the derived control law is the robust optimal control of the original uncertain nonlinear system.

Remark 3. According to Theorems 1 and 2, in order to complete the robust optimal control design, we should put emphasis upon solving the optimal control problem of the nominal system. As we see in the introduction, the ADP approach is effective in nonlinear optimal control design. Hence, in next section, we will present the design method based on neural network and the corresponding stability proof of the closed-loop system.

4. Optimal control design via ADP approach and the stability proof

According to the universal approximation property of neural networks, $J^*(x)$ can be reconstructed by a single-layer neural network on a compact set Ω as

$$J^*(x) = \omega_c^{\mathsf{T}}\sigma_c(x) + \varepsilon_c(x), \tag{23}$$

where $\omega_c \in \mathbb{R}^l$ is the ideal weight, $\sigma_c(x) \in \mathbb{R}^l$ is the activation function, l is the number of neurons in the hidden layer, and $\varepsilon_c(x)$ is the approximation error. Then, we have

$$\nabla J^*(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x). \tag{24}$$

Based on (24), the Lyapunov Eq. (4) becomes

$$0 = d_M^2(x) + u^{\mathsf{T}}(x)Ru(x) + \left(\omega_c^{\mathsf{T}}\nabla\sigma_c(x) + (\nabla\varepsilon_c(x))^{\mathsf{T}}\right)(f(x) + g(x)u(x)). \tag{25}$$

In light of [28,4,5], in this paper, we also assume that $\omega_c, \nabla\sigma_c(x)$, and $\varepsilon_c(x)$ and its derivative $\nabla\varepsilon_c(x)$ are all bounded on a compact set Ω.

Since the ideal weights are unknown, a critic neural network is built as

$$\hat{J}(x) = \hat{\omega}_c^{\mathsf{T}}\sigma_c(x), \tag{26}$$

to approximate the optimal cost function. Similarly, we have

$$\nabla\hat{J}(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{27}$$

According to (8) and (24), we have

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\left((\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x)\right). \tag{28}$$

In light of (8) and (27), the approximate control function can be given as

$$\hat{u}(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{29}$$

Applying (29) to system (2), the closed-loop system dynamics is expressed as

$$\dot{x} = f(x) - \frac{1}{2}g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{30}$$

Using the neural network expression (24), the Hamiltonian function becomes

$$H(x,\omega_c) = d_M^2(x) + \omega_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) + e_{cH} - \frac{1}{4}\omega_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c = 0, \tag{31}$$

where

$$e_{cH} = (\nabla\varepsilon_c(x))^{\mathsf{T}}f(x) - \frac{1}{2}(\nabla\varepsilon_c(x))^{\mathsf{T}}g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c - \frac{1}{4}(\nabla\varepsilon_c(x))^{\mathsf{T}}g(x)R^{-1}g^{\mathsf{T}}(x)\nabla\varepsilon_c(x), \tag{32}$$

denotes the residual error. Using the estimated weight vector, the approximate Hamiltonian function can be derived as

$$\hat{H}(x,\hat{\omega}_c) = d_M^2(x) + \hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) - \frac{1}{4}\hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c \triangleq e_c. \tag{33}$$

Let the weight estimation error of the critic network be $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$. Then, based on (31) and (33), we obtain

$$e_c = -\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c - e_{cH}. \tag{34}$$

In order to train the critic network, we aim at designing $\hat{\omega}_c$ to minimize the objective function

$$E_c = \frac{1}{2}e_c^{\mathsf{T}}e_c. \tag{35}$$

Here, the weights of the critic network are tuned based on the standard steepest descent algorithm with an additional term introduced to ensure the boundedness of system state, i.e.,

$$\dot{\hat{\omega}}_c = -\alpha_c\left(\frac{\partial E_c}{\partial\hat{\omega}_c}\right) + \frac{1}{2}\alpha_s\Pi(x,\hat{u})\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x), \tag{36}$$

where $\alpha_c > 0$ is the learning rate of the critic network, $\alpha_s > 0$ is the learning rate of the additional term, $J_s(x)$ is the Lyapunov function candidate given in Assumption 1, and $\Pi(x,\hat{u})$ is the additional stabilizing term defined as

$$\Pi(x,\hat{u}) = \begin{cases} 0, & \text{if } \dot{J}_s(x) = (\nabla J_s(x))^{\mathsf{T}}(f(x) + g(x)\hat{u}) < 0, \\ 1, & \text{else.} \end{cases} \tag{37}$$

Remark 4. It is important to note that the term $\Pi(x,\hat{u})$ is defined based on the Lyapunov condition for stability. The second term of (36) is removed when the nonlinear system exhibits stable behavior. However, in case of the controlled system exhibits signs of instability, the second term of (36) is activated for reinforcing the training process.

Next, we will find the dynamics of the weight estimation error $\tilde{\omega}_c$. According to (33), we have

$$\frac{\partial e_c}{\partial \hat{\omega}_c} = \nabla\sigma_c(x)f(x) - \frac{1}{2}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{38}$$

In light of (36), the dynamics of the weight estimation error is

$$\dot{\tilde{\omega}}_c = \alpha_c e_c\left(\frac{\partial e_c}{\partial \hat{\omega}_c}\right) - \frac{1}{2}\alpha_s\Pi(x,\hat{u})\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x). \tag{39}$$

Then, combining (34) and (38), the error dynamics (39) becomes

$$\begin{aligned}\dot{\tilde{\omega}}_c =\, & \alpha_c\Big(-\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c \\ & + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c - e_{cH}\Big) \\ & \times\Big(\nabla\sigma_c(x)f(x) - \frac{1}{2}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \frac{1}{2}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c\Big) \\ & - \frac{1}{2}\alpha_s\Pi(x,\hat{u})\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x).\end{aligned} \tag{40}$$

Theorem 3. *For system (2), let the control input be provided by (29) and the weight of the critic network be tuned by (36). Then, the state x of the closed-loop system and the weight estimation error $\tilde{\omega}_c$ of the critic network are UUB.*

Proof. Choose the Lyapunov function candidate as

$$L = \frac{1}{2\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\tilde{\omega}_c + \frac{\alpha_s}{2\alpha_c}J_s(x). \tag{41}$$

The derivative of (41) with respect to time along the dynamics (30) and (40) is

$$\dot{L} = \frac{1}{\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\dot{\tilde{\omega}}_c + \frac{\alpha_s}{2\alpha_c}(\nabla J_s(x))^{\mathsf{T}}\dot{x}. \tag{42}$$

Substituting (40) into (42), we have

$$\begin{aligned}\dot{L} =\, & \tilde{\omega}_c^{\mathsf{T}}\Big(-\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c \\ & + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c - e_{cH}\Big) \\ & \times\Big(\nabla\sigma_c(x)f(x) - \frac{1}{2}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \frac{1}{2}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c\Big) \\ & - \frac{\alpha_s}{2\alpha_c}\Pi(x,\hat{u})\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x) + \frac{\alpha_s}{2\alpha_c}(\nabla J_s(x))^{\mathsf{T}}\dot{x}.\end{aligned} \tag{43}$$

Denote $A = \nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}$. Then, (43) becomes

$$\begin{aligned}\dot{L} =\, & -\Big(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) + \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}A\tilde{\omega}_c - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}A\omega_c + e_{cH}\Big)\Big(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}A\tilde{\omega}_c - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}A\omega_c\Big) \\ & - \frac{\alpha_s}{2\alpha_c}\Pi(x,\hat{u})\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x) + \frac{\alpha_s}{2\alpha_c}(\nabla J_s(x))^{\mathsf{T}}\dot{x}.\end{aligned} \tag{44}$$

We assume $\lambda_{1m} > 0$ and $\lambda_{1M} > 0$ are the lower and upper bounds of the norm of matrix A, respectively. Additionally, assume that $\|\nabla\sigma_c(x)f(x)\| \leqslant \lambda_3$, $\|A\omega_c\| \leqslant \lambda_4$, and $\|e_{cH}\| \leqslant \lambda_5$, where λ_3, λ_4, and λ_5 are positive constants. Hence, the inequality (44) becomes

$$\begin{aligned}\dot{L} \leqslant\, & -\left(\frac{1}{8} - \frac{3}{8}\phi_1^2 - \frac{3}{16}\phi_2^2\right)\lambda_{1m}^2\|\tilde{\omega}_c\|^4 + \left\{\frac{1}{2}\lambda_{1M}\lambda_5 + \left(1 + \frac{3}{8\phi_1^2}\right)\lambda_3^2 + \left(\frac{3}{4} + \frac{3}{16\phi_2^2}\right)\lambda_4^2\right\}\|\tilde{\omega}_c\|^2 + \frac{3}{4}\lambda_5^2 \\ & - \frac{\alpha_s}{2\alpha_c}\Pi(x,\hat{u})\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x) + \frac{\alpha_s}{2\alpha_c}(\nabla J_s(x))^{\mathsf{T}}\dot{x},\end{aligned} \tag{45}$$

where ϕ_1 and ϕ_2 are constants chosen for the design purpose.

Case 1: $\Pi(x,\hat{u}) = 0$. Since $(\nabla J_s(x))^{\mathsf{T}}\dot{x} < 0$, there exists a positive constant λ_6 such that $0 < \lambda_6\|\nabla J_s(x)\| \leqslant -(\nabla J_s(x))^{\mathsf{T}}\dot{x}$ holds. Then, the inequality (45) becomes

$$\dot{L} \leqslant -\lambda_7\|\tilde{\omega}_c\|^4 + \lambda_8\|\tilde{\omega}_c\|^2 + \frac{3}{4}\lambda_5^2 - \frac{\alpha_s}{2\alpha_c}\lambda_6\|\nabla J_s(x)\|, \tag{46}$$

where

$$\lambda_7 = \left(\frac{1}{8} - \frac{3}{8}\phi_1^2 - \frac{3}{16}\phi_2^2\right)\lambda_{1m}^2, \tag{47}$$

$$\lambda_8 = \frac{1}{2}\lambda_{1M}\lambda_5 + \left(1 + \frac{3}{8\phi_1^2}\right)\lambda_3^2 + \left(\frac{3}{4} + \frac{3}{16\phi_2^2}\right)\lambda_4^2. \tag{48}$$

Therefore, whenever the inequality

$$\|\tilde{\omega}_c\| \geqslant \sqrt{\frac{\lambda_8 + \sqrt{3\lambda_5^2\lambda_7 + \lambda_8^2}}{2\lambda_7}} \triangleq \mathcal{A}_1 \tag{49}$$

or

$$\|\nabla J_s(x)\| \geqslant \frac{\alpha_c(3\lambda_5^2\lambda_7 + \lambda_8^2)}{2\alpha_s\lambda_6\lambda_7} \triangleq \mathcal{B}_1 \tag{50}$$

holds, we have $\dot{L} < 0$.

Case 2: $\Pi(x,\hat{u}) = 1$. According to (28) and (29), we have

$$u^* - \hat{u} = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\Big((\nabla\sigma(x))^{\mathsf{T}}\tilde{\omega}_c + \nabla\varepsilon_c(x)\Big). \tag{51}$$

In addition, we assume that $\|g(x)R^{-1}g^{\mathsf{T}}(x)\| \leqslant \lambda_9$, $\|\nabla\sigma_c(x)\| \leqslant \lambda_{10}$, and $\|\nabla\varepsilon_c(x)\| \leqslant \lambda_{11}$, where λ_9, λ_{10}, and λ_{11} are also positive constants. Then, considering (11) and (51), the inequality (45) becomes

$$\begin{aligned}\dot{L} &\leqslant -\left(\frac{1}{8} - \frac{3}{8}\phi_1^2 - \frac{3}{16}\phi_2^2\right)\lambda_{1m}^2\|\tilde{\omega}_c\|^4 + \left\{\frac{1}{2}\lambda_{1M}\lambda_5 + \left(1 + \frac{3}{8\phi_1^2}\right)\lambda_3^2 + \left(\frac{3}{4} + \frac{3}{16\phi_2^2}\right)\lambda_4^2\right\}\|\tilde{\omega}_c\|^2 + \frac{3}{4}\lambda_5^2 \\ &\quad + \frac{\alpha_s}{2\alpha_c}(\nabla J_s(x))^{\mathsf{T}}(f(x) + g(x)u^*) - \frac{\alpha_s}{4\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J_s(x) \\ &\quad + \frac{\alpha_s}{4\alpha_c}(\nabla J_s(x))^{\mathsf{T}}g(x)R^{-1}g^{\mathsf{T}}(x)\nabla\varepsilon_c(x) \\ &\leqslant -\lambda_7\|\tilde{\omega}_c\|^4 + \left(\lambda_8 + \frac{\alpha_s}{8\alpha_c}\right)\|\tilde{\omega}_c\|^2 + \frac{3}{4}\lambda_5^2 - \lambda_{12}\|\nabla J_s(x)\|^2 + \frac{\alpha_s}{4\alpha_c}\lambda_9\lambda_{11}\|\nabla J_s(x)\|,\end{aligned} \tag{52}$$

where

$$\lambda_{12} = \frac{\alpha_s}{2\alpha_c}\left(\lambda_m - \frac{1}{4}\lambda_9^2\lambda_{10}^2\right). \tag{53}$$

Therefore, whenever the inequality

$$\|\tilde{\omega}_c\| \geqslant \sqrt{\frac{8\alpha_c\lambda_8 + \alpha_s}{16\alpha_c\lambda_7} + \sqrt{\frac{(8\alpha_c\lambda_8 + \alpha_s)^2}{256\alpha_c^2\lambda_7^2} + \frac{3\lambda_5^2}{4\lambda_7} + \frac{\alpha_s^2\lambda_9^2\lambda_{11}^2}{64\alpha_c^2\lambda_7\lambda_{12}}}} \triangleq \mathcal{A}_2 \tag{54}$$

or

$$\|\nabla J_s(x)\| \geqslant \frac{\alpha_s\lambda_9\lambda_{11}}{8\alpha_c\lambda_{12}} + \sqrt{\frac{(8\alpha_c\lambda_8 + \alpha_s)^2}{256\alpha_c^2\lambda_7\lambda_{12}} + \frac{3\lambda_5^2}{4\lambda_{12}} + \frac{\alpha_s^2\lambda_9^2\lambda_{11}^2}{64\alpha_c^2\lambda_{12}^2}} \triangleq \mathcal{B}_2 \tag{55}$$

holds, we obtain $\dot{L} < 0$.

According to Cases 1 and 2, if the inequality $\|\tilde{\omega}_c\| > \max(\mathcal{A}_1, \mathcal{A}_2)$ or $\|\nabla J_s(x)\| > \max(\mathcal{B}_1, \mathcal{B}_2)$ holds, then $\dot{L} < 0$. Thus, by using the standard Lyapunov extension theorem [11], the state x and the error $\tilde{\omega}_c$ are UUB. □

5. Decentralized optimal control design of nonlinear interconnected systems

Large-scale systems are common in engineering area when doing research on complex dynamical systems that can be partitioned into a set of interconnected subsystems. The decentralized control is one of the effective design approaches and has attracted a great amount of interest due to its advantages in easier implementation and lower dimensionality

[17,10,26,27,39]. In this section, we generalize the aforementioned results to decentralized optimal control for a class of continuous-time nonlinear interconnected large-scale systems. This part is also an extension of the decentralized control strategy developed in [17].

Consider nonlinear large-scale systems composed of N interconnected subsystems which are described by

$$\dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))\big(\bar{u}_i(t) + \bar{d}_i(x_D(t))\big),\ i = 1,2,\ldots,N, \tag{56}$$

where $x_i(t) \in \mathbb{R}^{n_i}$ and $\bar{u}_i(t) \in \mathbb{R}^{m_i}$ are the state vector and the control vector of the ith subsystem, respectively, and $x_D = [x_1^{\mathsf{T}}, x_2^{\mathsf{T}}, \ldots, x_N^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^{n_D}$ is the overall state with $n_D = \sum_{i=1}^{N} n_i$. Note that for subsystem i, $f_i(x_i), g_i(x_i)$, and $g_i(x_i)\bar{d}_i(x_D)$ represent the nonlinear internal dynamics, the input gain matrix, and the interconnected term, respectively. Here, $x_1, x_2, \ldots, x_N$ are called local states while $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_N$ are local controls.

Let $x_i(0) = x_{i0}$ be the initial state of the ith subsystem, $i = 1,2,\ldots,N$. Let $R_i \in \mathbb{R}^{m_i \times m_i}$, $i = 1,2,\ldots,N$, be a set of symmetric positive definite matrices. Denote $d_i(x_D) = R_i^{1/2}\bar{d}_i(x_D), i = 1,2,\ldots,N$, which are bounded as

$$\|d_i(x_D)\| \leqslant \sum_{j=1}^{N} \rho_{ij} h_{ij}(x_j),\ i = 1,2,\ldots,N. \tag{57}$$

In (57), ρ_{ij} are nonnegative constants and $h_{ij}(x_j)$ are positive semi-definite functions with $i,j = 1,2,\ldots,N$. If we define $h_i(x_i) = \max\{h_{1i}(x_i), h_{2i}(x_i), \ldots, h_{Ni}(x_i)\}$, $i = 1,2,\ldots,N$, then the Eq. (57) becomes

$$\|d_i(x_D)\| \leqslant \sum_{j=1}^{N} \tau_{ij} h_j(x_j),\ i = 1,2,\ldots,N, \tag{58}$$

where $\tau_{ij} \geqslant \rho_{ij} h_{ij}(x_j)/h_j(x_j)$, $i,j = 1,2,\ldots,N$, are also nonnegative constants.

In the following, we focus on designing the decentralized optimal control law. First, we should find N control policies $\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N)$, such that the constituted control vector $(\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N))$ can stabilize system (56). As is shown in [17], the decentralized control can be obtained by solving the optimal control problem of the N isolated subsystems given by

$$\dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))u_i(t),\ i = 1,2,\ldots,N. \tag{59}$$

Let $h_i(x_i) \leqslant d_{iM}(x_i),\ i = 1,2,\ldots,N$. According to [17], we can find a set of optimal control policies $u_i^*(x_i), i = 1,2,\ldots,N$, which minimize the local cost functions

$$J_i(x_{i0}) = \int_0^{\infty} \left\{ d_{iM}^2(x_i(\tau)) + u_i^{\mathsf{T}}(x_i(\tau)) R_i u_i(x_i(\tau)) \right\} \mathrm{d}\tau,\ i = 1,2,\ldots,N. \tag{60}$$

Using the notation of optimal cost functions $J_i^*(x_i)$, $i = 1,2,\ldots,N$, the HJB equations of the isolated subsystems are

$$0 = d_{iM}^2(x_i) + (\nabla J_i^*(x_i))^{\mathsf{T}} f_i(x_i) - \frac{1}{4}(\nabla J_i^*(x_i))^{\mathsf{T}} g_i(x_i) R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla J_i^*(x_i), \tag{61}$$

with $J_i^*(0) = 0$. Then, there exist N positive numbers such that the feedback controls

$$\bar{u}_i(x_i) = \zeta_i u_i^*(x_i) = -\frac{1}{2}\zeta_i R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla J_i^*(x_i),\ i = 1,2,\ldots,N, \tag{62}$$

can form a control pair $(\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N))$, which is just the decentralized control strategy of system (56).

Next, we study the optimality of the decentralized control scheme with a specified overall cost function. Denote

$$f_D(x_D) = \begin{bmatrix} f_1(x_1) \\ f_2(x_2) \\ \vdots \\ f_N(x_N) \end{bmatrix},\ \bar{u}_D(x_D) = \begin{bmatrix} \bar{u}_1(x_1) \\ \bar{u}_2(x_2) \\ \vdots \\ \bar{u}_N(x_N) \end{bmatrix},\ \bar{d}_D(x_D) = \begin{bmatrix} \bar{d}_1(x_D) \\ \bar{d}_2(x_D) \\ \vdots \\ \bar{d}_N(x_D) \end{bmatrix},$$

$$R_D = \mathrm{diag}\left\{ \frac{1}{\zeta_1} R_1, \frac{1}{\zeta_2} R_2, \ldots, \frac{1}{\zeta_N} R_N \right\},\ g_D(x_D) = \mathrm{diag}\{g_1(x_1), g_2(x_2), \ldots, g_N(x_N)\}. \tag{63}$$

For system (56), we define the following cost function

$$\bar{J}_D(x_{D0}) = \int_0^{\infty} \{ Q_D(x_D(\tau)) + \bar{u}_D^{\mathsf{T}}(x_D(\tau)) R_D \bar{u}_D(x_D(\tau)) \} \mathrm{d}\tau, \tag{64}$$

where

$$Q_D(x_D) = \sum_{i=1}^{N} \left\{ d_{iM}^2(x_i) - (\nabla J_i^*(x_i))^{\mathsf{T}} g_i(x_i) \bar{d}_i(x_D) + \frac{1}{4}(\zeta_i - 1)(\nabla J_i^*(x_i))^{\mathsf{T}} g_i(x_i) R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla J_i^*(x_i) \right\}. \tag{65}$$

Then, we have the following theorem.

Theorem 4. *Consider system (56) with cost function (64). There exists a set of positive numbers such that the feedback control laws obtained by (62) constitute the decentralized optimal control of the interconnected large-scale system.*

Proof. Similar to Lemma 2 and by considering (58), we have

$$\begin{aligned} Q_D(x_D) &\geqslant \sum_{i=1}^{N}\left\{d_{iM}^2(x_i) - \frac{1}{\zeta_i - 1}d_i^{\mathsf{T}}(x_D)d_i(x_D)\right\} \\ &\geqslant \sum_{i=1}^{N}\left\{d_{iM}^2(x_i) - \frac{1}{\zeta_i - 1}\left(\sum_{j=1}^{N}\tau_{ij}d_{jM}(x_j)\right)^2\right\}. \end{aligned} \tag{66}$$

Then, we find that the positive definiteness of $Q_D(x_D)$ can be guaranteed if $\zeta_i, i = 1, 2, \ldots, N$, are sufficiently large. Moreover, based on (61) and (65), we can prove that $J_D^*(x_D) = \sum_{i=1}^{N} J_i^*(x_i)$ satisfies the equation

$$Q_D(x_D) + \bar{u}_D^{\mathsf{T}}(x_D)R_D\bar{u}_D(x_D) + (\nabla J_D^*(x_D))^{\mathsf{T}}\big(f_D(x_D) + g_D(x_D)(\bar{u}_D(x_D) + \bar{d}_D(x_D))\big) = 0. \tag{67}$$

Therefore, consider system (56) with cost function (64), there exists a set of positive numbers such that the control pair $(\bar{u}_1(x_1), \bar{u}_2(x_2), \ldots, \bar{u}_N(x_N))$ obtained by (62) is the decentralized optimal control law. □

Remark 5. It should be pointed out that, in this part, the ADP technique can also be employed to design the optimal controls of the isolated subsystems, where N critic networks will be constructed to facilitate the implementation procedure. Then, the decentralized control of the interconnected system can be obtained, which is, simultaneously, the optimal control with respect to an overall cost function. In this sense, we accomplish the decentralized optimal control design of the nonlinear interconnected large-scale system based on ADP approach.

6. Simulation studies

Two examples are provided in this section to demonstrate the effectiveness of the robust optimal control strategy.

Example 1. Consider the following continuous-time nonlinear system:

$$\dot{x} = \begin{bmatrix} -0.5x_1 + x_2(1 + 0.5x_2^2) \\ -0.8(x_1 + x_2) + 0.5x_2(1 - 0.3x_2^2) \end{bmatrix} + \begin{bmatrix} 0 \\ -0.6 \end{bmatrix}(\bar{u} + \bar{d}(x)), \tag{68}$$

where $x = [x_1, x_2]^{\mathsf{T}} \in \mathbb{R}^2$ and $\bar{u} \in \mathbb{R}$ are the state and control variables, respectively. The term $\bar{d}(x) = \delta_1 x_2 \cos(\delta_2 x_1 + \delta_3 x_2)$ reflects the uncertainty of the controlled plant, where δ_1, δ_2, and δ_3 are unknown parameters with $\delta_1 \in [-1, 1], \delta_2 \in [-5, 5]$, and $\delta_3 \in [-3, 3]$. We set $R = I$ and choose $d_M(x) = \|x\|$ as the bound of $d(x)$.

According to the results in this paper, in order to derive the optimal control of the nominal system

$$\dot{x} = \begin{bmatrix} -0.5x_1 + x_2(1 + 0.5x_2^2) \\ -0.8(x_1 + x_2) + 0.5x_2(1 - 0.3x_2^2) \end{bmatrix} + \begin{bmatrix} 0 \\ -0.6 \end{bmatrix}u \tag{69}$$

with cost function

$$J(x_0) = \int_0^\infty \left\{\|x(\tau)\|^2 + u^{\mathsf{T}}(x(\tau))Ru(x(\tau))\right\}\mathrm{d}\tau, \tag{70}$$

we have to construct a neural network based on the idea of ADP. In this example, the critic network is built as

$$\hat{J}(x) = \hat{\omega}_{c1}x_1^2 + \hat{\omega}_{c2}x_1x_2 + \hat{\omega}_{c3}x_2^2 + \hat{\omega}_{c4}x_1^4 + \hat{\omega}_{c5}x_1^3x_2 + \hat{\omega}_{c6}x_1^2x_2^2 + \hat{\omega}_{c7}x_1x_2^3 + \hat{\omega}_{c8}x_2^4. \tag{71}$$

During the simulation process, the probing noise is introduced to satisfy the persistency of excitation condition. Let the learning rates of the critic network and the additional term be $\alpha_c = 0.8$ and $\alpha_s = 0.5$, respectively. In addition, let the initial weight of the critic network be zero vector and the initial state of the controlled plant be $x_0 = [0.5, -0.5]^{\mathsf{T}}$. After simulation, we can observe that the convergence of the weights has occurred after 2500 s. Then, the probing signal is turned off. In fact, the weights of the critic network converge to $[0.8709, 0.1291, 1.0617, 0.0868, -0.1566, 0.2053, -0.0059, 0.0651]^{\mathsf{T}}$, which is displayed in Fig. 1.

Next, the scalar parameters are chosen as $\zeta = 3$, $\delta_1 = 0.8$, $\delta_2 = -5$, and $\delta_3 = 3$, respectively, so as to evaluate the robust control performance. Under the action of the robust control strategy, the state trajectory of system (68) during the first 20 s is shown in Fig. 2. In light of Theorem 2, it also achieves optimality with cost function defined as in (17). These results authenticate the validity of the robust optimal control scheme developed in this paper.

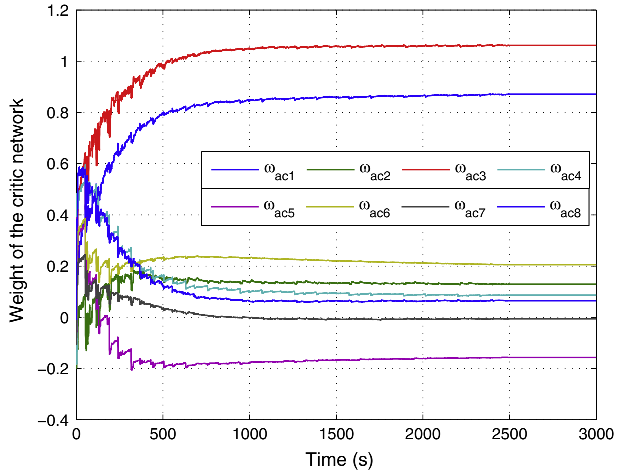

Fig. 1. Convergence of the weight vector of the critic network (ω_{aci} represents $\hat{\omega}_{ci},\ i = 1, 2, \ldots, 8$).

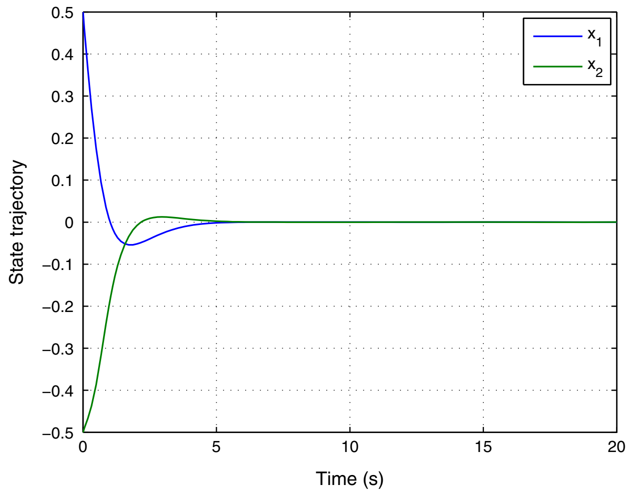

Fig. 2. The state trajectory.

Example 2. Consider the continuous-time nonlinear large-scale system [17]

$$\begin{aligned}\dot{x}_1 &= \begin{bmatrix} -x_{11} + x_{12} \\ -0.5x_{11} - 0.5x_{12} - 0.5x_{12}(\cos(2x_{11}) + 2)^2 \end{bmatrix} + \begin{bmatrix} 0 \\ \cos(2x_{11}) + 2 \end{bmatrix}(\bar{u}_1(x_1) + \bar{d}_1(x_D)),\\ \dot{x}_2 &= \begin{bmatrix} x_{22} \\ -x_{21} - 0.5x_{22} + 0.5x_{21}^2 x_{22} \end{bmatrix} + \begin{bmatrix} 0 \\ x_{21} \end{bmatrix}(\bar{u}_2(x_2) + \bar{d}_2(x_D)),\end{aligned} \tag{72}$$

where $x_1 = [x_{11}, x_{12}]^{\mathsf{T}} \in \mathbb{R}^2$ and $\bar{u}_1(x_1) \in \mathbb{R}$ are the state and control variables of subsystem 1, $x_2 = [x_{21}, x_{22}]^{\mathsf{T}} \in \mathbb{R}^2$ and $\bar{u}_2(x_2) \in \mathbb{R}$ are the state and control variables of subsystem 2, and $x_D = [x_1^{\mathsf{T}}, x_2^{\mathsf{T}}]^{\mathsf{T}}$ is the overall state. The interconnected terms are $\bar{d}_1(x_D) = (x_{11} + x_{22}) \sin x_{12}^2 \cos(0.5x_{21})$ and $\bar{d}_2(x_D) = 0.5(x_{12} + x_{22}) \cos\left(e^{x_{21}^2}\right)$. Let $R_1 = R_2 = I$, $h_1(x_1) = \|x_1\|$, and $h_2(x_2) = | x_{22} |$.

In order to design the decentralized optimal controller of interconnected system (72), we first aim at solving the optimal control problem of the two isolated subsystems

$$\begin{aligned}\dot{x}_1 &= \begin{bmatrix} -x_{11} + x_{12} \\ -0.5x_{11} - 0.5x_{12} - 0.5x_{12}(\cos(2x_{11}) + 2)^2 \end{bmatrix} + \begin{bmatrix} 0 \\ \cos(2x_{11}) + 2 \end{bmatrix} u_1(x_1),\\ \dot{x}_2 &= \begin{bmatrix} x_{22} \\ -x_{21} - 0.5x_{22} + 0.5x_{21}^2 x_{22} \end{bmatrix} + \begin{bmatrix} 0 \\ x_{21} \end{bmatrix} u_2(x_2).\end{aligned} \tag{73}$$

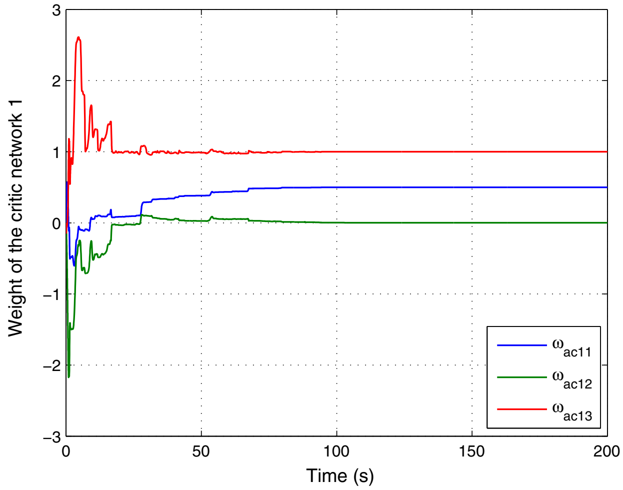

Fig. 3. Convergence of the weight vector of the critic network 1 ($\omega_{ac11}, \omega_{ac12}$, and ω_{ac13} represent $\hat{\omega}_{c11}, \hat{\omega}_{c12}$, and $\hat{\omega}_{c13}$, respectively).

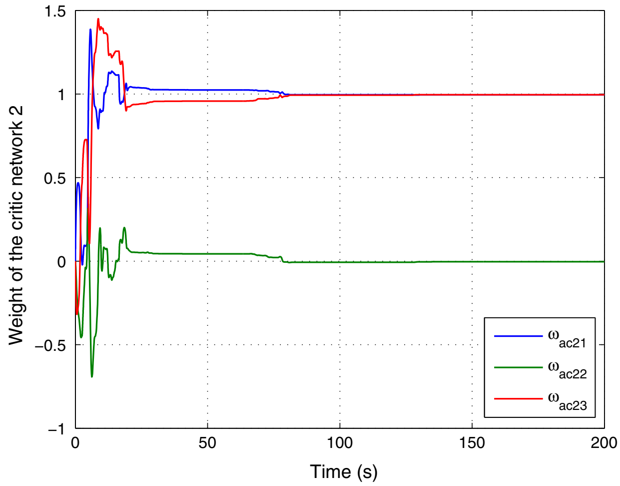

Fig. 4. Convergence of the weight vector of the critic network 2 ($\omega_{ac21}, \omega_{ac22}$, and ω_{ac23} represent $\hat{\omega}_{c21}, \hat{\omega}_{c22}$, and $\hat{\omega}_{c23}$, respectively).

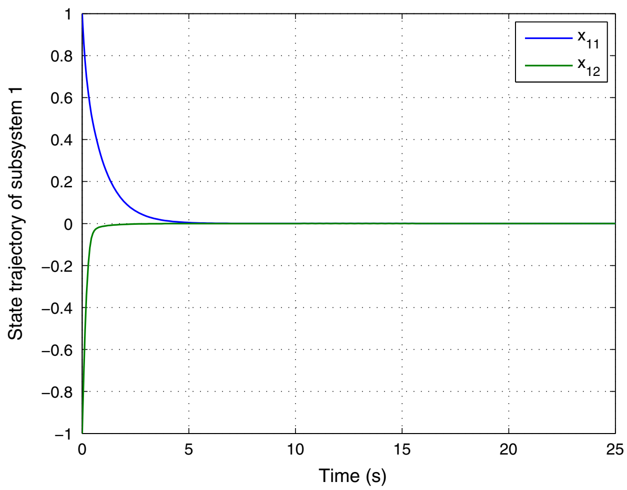

Fig. 5. The state trajectory of subsystem 1.

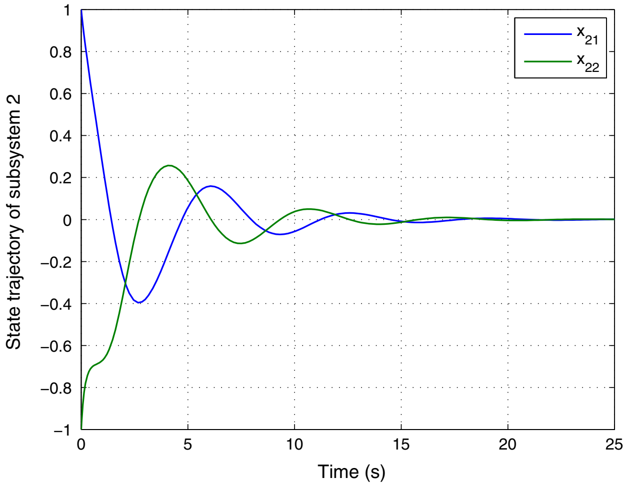

Fig. 6. The state trajectory of subsystem 2.

Here, we choose $d_{1M}(x_1) = \|x_1\|$ and $d_{2M}(x_2) = | x_{22} |$. Hence, the cost functions of the optimal control problem are

$$J_1(x_{10}) = \int_0^\infty \left\{ \|x_1(\tau)\|^2 + u_1^{\mathsf{T}}(x(\tau))R_1u_1(x(\tau)) \right\} \mathrm{d}\tau, \tag{74}$$

and

$$J_2(x_{20}) = \int_0^\infty \left\{ | x_{22}(\tau)|^2 + u_2^{\mathsf{T}}(x(\tau))R_2u_2(x(\tau)) \right\} \mathrm{d}\tau, \tag{75}$$

respectively.

Here, two critic networks are constructed with activation functions chosen as $[x_{11}^2, x_{11}x_{12}, x_{12}^2]^{\mathsf{T}}$ and $[x_{21}^2, x_{21}x_{22}, x_{22}^2]^{\mathsf{T}}$, respectively. Besides, let the learning rates be the same as in Example 1, the initial weights of the two critic networks be $[0, 0, 0]^{\mathsf{T}}$, and the initial states of the two isolated subsystems be $x_{10} = x_{20} = [1, -1]^{\mathsf{T}}$. During the simulation, we can find that after 180s, the weights of the critic networks converge to $[0.5000, 0.0001, 1.0000]^{\mathsf{T}}$ and $[0.9949, -0.0034, 0.9959]^{\mathsf{T}}$ (see Figs. 3 and 4). Next, we apply the decentralized control scheme to controlled plant (72) for 25s and obtain the evolution processes of the state trajectories illustrated in Figs. 5 and 6. These simulation results verify the validity of the decentralized optimal control scheme developed in this paper.

7. Conclusion

A novel robust optimal control scheme for a class of uncertain nonlinear systems via ADP approach is developed in this paper. It is proved that the robust controller of the original uncertain system achieves optimality under a specified cost function. During the implementation process, a critic network is constructed to solve the HJB equation of the nominal system and an additional stabilizing term is introduced to verify the stability. The obtained results are also extended to design the decentralized optimal control for a class of nonlinear interconnected large-scale systems. Simulation studies verify the good control performance. In the future, we will focus on studying the robust and decentralized optimal control for nonlinear systems with unknown dynamics. In this sense, the requirement of system dynamics will be further reduced, which reflects the superiority of ADP technique in dealing with the optimal control problem under nonlinear, uncertain, and complex environment.

References

[1] M. Abu-Khalaf, F.L. Lewis, Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach, Automatica 41 (2005) 779–791.
[2] A. Al-Tamimi, F.L. Lewis, M. Abu-Khalaf, Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof, IEEE Trans. Syst. Man Cybernet. – Part B: Cybernet. 38 (2008) 943–949.
[3] D.P. Bertsekas, M.L. Homer, D.A. Logan, S.D. Patek, N.R. Sandell, Missile defense and interceptor allocation by neuro-dynamic programming, IEEE Trans. Syst. Man Cybernet. – Part A: Syst. Humans 30 (2000) 42–51.
[4] S. Bhasin, R. Kamalapurkar, M. Johnson, K.G. Vamvoudakis, F.L. Lewis, W.E. Dixon, A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems, Automatica 49 (2013) 82–92.
[5] 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.
[6] T. Dierks, S. Jagannathan, Online optimal control of affine nonlinear discrete-time systems with unknown internal dynamics by using time-based policy update, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 1118–1129.
[7] J. Fu, H. He, X. Zhou, Adaptive learning and control for MIMO system based on adaptive dynamic programming, IEEE Trans. Neural Netw. 22 (2011) 1133–1148.

[8] 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.
[9] Y. Jiang, Z.P. Jiang, Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics, Automatica 48 (2012) 2699–2704.
[10] Z.P. Jiang, Y. Jiang, Robust adaptive dynamic programming for linear and nonlinear systems: an overview, Eur. J. Control 19 (2013) 417–425.
[11] F.L. Lewis, S. Jagannathan, A. Yesildirek, Neural Network Control of Robot Manipulators and Nonlinear Systems, Taylor & Francis, London, 1999.
[12] F.L. Lewis, D. Liu, Reinforcement Learning and Approximate Dynamic Programming for Feedback Control, Wiley, Hoboken, NJ, 2013.
[13] F.L. Lewis, D. Vrabie, Reinforcement learning and adaptive dynamic programming for feedback control, IEEE Circ. Syst. Mag. 9 (2009) 32–50.
[14] J. Liang, G.K. Venayagamoorthy, R.G. Harley, Wide-area measurement based dynamic stochastic optimal power flow control for smart grids with high variability and uncertainty, IEEE Trans. Smart Grid 3 (2012) 59–69.
[15] F. Lin, R.D. Brand, J. Sun, Robust control of nonlinear systems: compensating for uncertainty, Int. J. Control 56 (1992) 1453–1459.
[16] D. Liu, H. Li, D. Wang, Neural-network-based zero-sum game for discrete-time nonlinear systems via iterative adaptive dynamic programming algorithm, Neurocomputing 110 (2013) 92–100.
[17] D. Liu, D. Wang, H. Li, Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 418–428.
[18] D. Liu, D. Wang, X. Yang, An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs, Inform. Sci. 220 (2013) 331–342.
[19] D. Liu, D. Wang, D. Zhao, Q. Wei, N. Jin, Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming, IEEE Trans. Autom. Sci. Eng. 9 (2012) 628–634.
[20] B. Luo, H.N. Wu, Approximate optimal control design for nonlinear one-dimensional parabolic PDE systems using empirical eigenfunctions and neural network, IEEE Trans. Syst. Man Cybernet. – Part B: Cybernet. 42 (2012) 1538–1549.
[21] H. Modares, F.L. Lewis, M.B. Naghibi-Sistani, Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems, Automatica 50 (2014) 193–202.
[22] H. Modares, M.B. Naghibi-Sistani, F.L. Lewis, A policy iteration approach to online optimal control of continuous-time constrained-input systems, ISA Trans. 52 (2013) 611–621.
[23] Z. Ni, H. He, J. Wen, Adaptive learning in tracking control based on the dual critic network design, IEEE Trans. Neural Netw. Learn. Syst. 24 (2013) 913–928.
[24] 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.
[25] D.V. Prokhorov, D.C. Wunsch, Adaptive critic designs, IEEE Trans. Neural Netw. 8 (1997) 997–1007.
[26] A. Saberi, On optimality of decentralized control for a class of nonlinear interconnected systems, Automatica 24 (1988) 101–104.
[27] D.D. Siljak, Decentralized Control of Complex Systems, Academic Press, Boston, MA, 2012.
[28] 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.
[29] D. Wang, D. Liu, Neuro-optimal control for a class of unknown nonlinear dynamic systems using SN-DHP technique, Neurocomputing 121 (2013) 218–225.
[30] D. Wang, D. Liu, H. Li, Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems, IEEE Trans. Autom. Sci. Eng. 11 (2014) 627–632.
[31] D. Wang, D. Liu, H. Li, X. Yang, A learning optimal control scheme for robust stabilization of a class of uncertain nonlinear systems, in: Proceedings of the Chinese Control Conference, Xi'an, China, 2013, pp. 7834–7839.
[32] D. Wang, D. Liu, Q. Wei, Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach, Neurocomputing 78 (2012) 14–22.
[33] D. Wang, D. Liu, Q. Wei, D. Zhao, N. Jin, Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming, Automatica 48 (2012) 1825–1832.
[34] F.Y. Wang, H. Zhang, D. Liu, Adaptive dynamic programming: an introduction, IEEE Comput. Intell. Mag. 4 (2009) 39–47.
[35] 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, 1992.
[36] H.N. Wu, B. Luo, Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 1884–1895.
[37] H.N. Wu, B. Luo, Simultaneous policy update algorithms for learning the solution of linear continuous-time H_∞ state feedback control, Inform. Sci. 222 (2013) 472–485.
[38] B. Xu, C. Yang, Z. Shi, Reinforcement learning output feedback NN control using deterministic learning technique, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 635–641.
[39] D. Xu, V. Ugrinovskii, Decentralized measurement feedback stabilization of large-scale systems via control vector Lyapunov functions, Syst. Control Lett. 62 (2013) 1187–1195.
[40] X. Xu, Z. Hou, C. Lian, H. He, Online learning control using adaptive critic designs with sparse kernel machines, IEEE Trans. Neural Netw. Learn. Syst. 24 (2013) 762–775.
[41] X. Xu, L. Zuo, Z. Huang, Reinforcement learning algorithms with function approximation: recent advances and applications, Inform. Sci. 261 (2014) 1–31.
[42] X. Yang, D. Liu, D. Wang, Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints, Int. J. Control 87 (2014) 553–566.
[43] H.M. Yen, T.H.S. Li, Y.C. Chang, Design of a robust neural network-based tracking controller for a class of electrically driven nonholonomic mechanical systems, Inform. Sci. 222 (2013) 559–575.
[44] 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. Cybernet. 43 (2013) 206–216.
[45] H. Zhang, D. Liu, Y. Luo, D. Wang, Adaptive Dynamic Programming for Control: Algorithms and Stability, Springer, London, UK, 2013.
[46] H. Zhang, Y. Luo, D. Liu, Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints, IEEE Trans. Neural Netw. 20 (2009) 1490–1503.