Soft Comput (2016) 20:707–716
DOI 10.1007/s00500-014-1534-z

METHODOLOGIES AND APPLICATION

A neural-network-based online optimal control approach for nonlinear robust decentralized stabilization

Ding Wang · Derong Liu · Hongliang Li · Hongwen Ma · Chao Li

Published online: 25 November 2014

Abstract In this paper, the robust decentralized stabilization of continuous-time uncertain nonlinear systems with multi control stations is developed using a neural network based online optimal control approach. The novelty lies in that the well-known adaptive dynamic programming method is extended to deal with the nonlinear feedback control problem under uncertain and large-scale environment. Through introducing an appropriate bounded function and defining a modified cost function, it can be observed that the decentralized optimal controller of the nominal system can achieve robust decentralized stabilization of original uncertain system. Then, a critic neural network is constructed for solving the modified Hamilton–Jacobi–Bellman equation corresponding to the nominal system in an online fashion. The weights of the critic network are tuned based on the standard steepest descent algorithm with an additional term provided to guarantee the boundedness of system states. The stability analysis of the closed-loop system is carried out via the Lyapunov approach. At last, two simulation examples are given to verify the effectiveness of the present control approach.

Keywords Adaptive dynamic programming · Approximate dynamic programming · Neural networks · Online optimal control · Robust decentralized stabilization · Uncertain nonlinear systems

Communicated by V. Loia.

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.

D. Wang · D. Liu (✉) · H. Li · H. Ma · C. Li
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
e-mail: derongliu@gmail.com; derong.liu@ia.ac.cn

D. Wang
e-mail: ding.wang@ia.ac.cn

H. Li
e-mail: hongliang.li@ia.ac.cn

H. Ma
e-mail: mahongwen2012@ia.ac.cn

C. Li
e-mail: lichao2012@ia.ac.cn

1 Introduction

How to construct truly brain-like systems has become one of the advanced research topics in the field of computational intelligence. Among them, adaptive or approximate dynamic programming (ADP) is a biologically inspired and computational method proposed by Werbos (1992) to solve optimization and optimal control problems efficiently. In general, it is implemented by solving the Hamilton–Jacobi–Bellman (HJB) equation based on function approximators, such as neural networks. In recent years, the research on ADP and related fields has gained much attention from scholars, see, e.g., Lewis et al. (2012), Liu et al. (2013c), Jiang and Jiang (2013) and the references therein. It is worth mentioning that the ADP method has been extensively used in feedback control applications, both for discrete-time systems (Al-Tamimi et al. 2008; Zhang et al. 2009; Wang et al. 2012a, b; Liu et al. 2012, 2013a, b; Heydari and Balakrishnan 2013; Ni et al. 2013; Ni and He 2013; Dierks and Jagannathan 2012; Zhang et al. 2014) and for continuous-time systems (Abu-Khalaf and Lewis 2005; Vamvoudakis and Lewis 2010; Bhasin et al. 2013; Wu and Luo 2012; Zhang et al. 2013; Yang et al. 2014; Dierks and Jagannathan 2010; Nodland et al. 2013; Zhao et

al. 2013; Adhyaru et al. 2011; Liu et al. 2014a, b). Thus, it gradually plays an important role in designing adaptive learning and intelligent control systems. In Wang et al. (2012b), the iterative globalized dual heuristic programming algorithm was developed to conduct the optimal control design of unknown nonaffine nonlinear discrete-time systems. In Ni et al. (2013), an adaptive learning approach for tracking control was given based on dual critic network design. In Wu and Luo (2012), a neural-network-based online simultaneous policy update algorithm was proposed for handling the nonlinear H_∞ control problem. In Liu et al. (2014b), an online synchronous approximate optimal learning algorithm was provided for multiplayer nonzero-sum games with unknown dynamics. It is not difficult to find that, however, most of the existing results are inapplicable when the controlled plant contains some kinds of uncertainties.

As is shown in Adhyaru et al. (2011), Haddad et al. (1998, 2000), Lin (2000), Wang et al. (2014) the unavoidable discrepancies between system models and real-world dynamics can sometimes result in the degradation of system performance. Hence, the feedback control should be designed to be robust with respect to system uncertainties. In Adhyaru et al. (2011), an optimal control algorithm was proposed to deal with the nonlinear robust control problem, but it was constructed using the least square method and performed in an offline fashion, not to mention the stability analysis of the closed-loop system was not conducted. Recently, Liu et al. (2014a) established an online learning optimal control approach to deal with the decentralized stabilization problem of nonlinear interconnected large-scale systems. It is a meaningful attempt for extending ADP approach to decentralized control of large-scale systems. However, the main algorithm is implemented based on an initial admissible control, which is not easy to acquire in some situations.

In this paper, we investigate the robust decentralized stabilization of continuous-time uncertain nonlinear systems using neural-network-based online solution of the HJB equation. The robust decentralized stabilization problem is transformed into an optimal control problem by introducing an appropriate cost function. It can be proved that the decentralized optimal controller of the nominal system is the robust decentralized controller of the uncertain system. Then, a critic network is constructed for facilitating the solution of the modified HJB equation. Moreover, inspired by the work of Dierks and Jagannathan (2010), Nodland et al. (2013), Zhao et al. (2013), an additional stabilizing term is introduced to verify the stability, which relaxes the need for an initial stabilizing control. It also can be regarded as the main idea of the reinforced training process of critic network. The uniform ultimate boundedness (UUB) of the closed-loop system is also proved using the well-known Lyapunov approach. Besides, the approximated control input can converge to the optimal control within a small bound. Significantly, the developed approach is applicable to design the robust decentralized control for a class of complex nonlinear systems under an uncertain and large-scale environment.

The rest of this paper is organized as follows: in Sect. 2, the problem statement of robust decentralized stabilization is provided. In Sect. 3, the studied problem is transformed into a decentralized optimal control problem with a modified cost function. In Sect. 4, a neural network is constructed to solve the modified HJB equation approximately in an online fashion. Then, the stability of the overall closed-loop system is proved. In Sect. 5, two numerical examples are given to demonstrate the effectiveness of the established approach. In Sect. 6, concluding remarks are presented to display the usability of the established method corresponding to nonlinear system with N control stations.

2 Problem statement

In this paper, we study the following continuous-time uncertain nonlinear systems with two control stations:

$$\begin{aligned}\dot{x} &= \bar{F}(x(t), u_1(t), u_2(t)) \\ &= f(x(t)) + g_1(x(t))u_1(t) + g_2(x(t))u_2(t) + \Delta f(x(t)),\end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector and $u_i(t) \in \mathbb{R}^{m_i}, i = 1, 2$, are the control inputs, $f(\cdot)$ and $g_i(\cdot)$, $i = 1, 2$, are differentiable in their arguments with $f(0) = 0$, and $\Delta f(x(t))$ is the nonlinear perturbation of the corresponding nominal system

$$\begin{aligned}\dot{x} &= F(x(t), u_1(t), u_2(t)) \\ &= f(x(t)) + g_1(x(t))u_1(t) + g_2(x(t))u_2(t).\end{aligned} \tag{2}$$

Here, we let $x(0) = x_0$ be the initial state. In addition, as in many other literature, we assume that $f + g_1u_1 + g_2u_2$ is Lipschitz continuous on a set Ω in $\mathbb{R}^n$ containing the origin and that the system (2) is controllable.

For the system uncertainty $\Delta f(x)$, we assume that it has the form

$$\Delta f(x) = G(x)d(\varphi(x)) \tag{3}$$

with

$$d^{\mathsf{T}}(\varphi(x))d(\varphi(x)) \le h^{\mathsf{T}}(\varphi(x))h(\varphi(x)), \tag{4}$$

where $G(\cdot) \in \mathbb{R}^{n\times r}$ and $\varphi(\cdot)$ satisfying $\varphi(0) = 0$ are fixed functions denoting the structure of the uncertainty, $d(\cdot) \in \mathbb{R}^r$ is an uncertain function with $d(0) = 0$, and $h(\cdot) \in \mathbb{R}^r$ is a given function with $h(0) = 0$.

Remark 1 Note that the formula described in (3) represents a general form of uncertainty, which includes the cases of matched uncertainty and unmatched one.

In this paper, we aim at finding two state feedback control functions $u_i(x)$, $i = 1, 2$, such that the control pair $(u_1(x), u_2(x))$ can stabilize system (1) for any possible uncertainties. In this sense, we say the control pair ensures robust decentralized stabilization of system (1).

3 Robust decentralized stabilization of the nonlinear systems via optimal control design

In this section, inspired by the work of Haddad et al. (1998, 2000), we derive the following theorem, which facilitates us to carry out the problem transformation between robust decentralized stabilization and decentralized optimal control:

Theorem 1 *If there exists a continuously differentiable cost function $V(x)$ satisfying $V(x) > 0$ for all $x \neq 0$ and $V(0) = 0$, a bounded function $\Gamma(x)$ satisfying $\Gamma(x) \geq 0$, and two feedback control functions $u_1(x)$ and $u_2(x)$ such that*

$$(\nabla V(x))^{\mathsf{T}} \Delta f(x) \leq \Gamma(x), \tag{5}$$

$$U(x, u_1, u_2) + \Gamma(x) + (\nabla V(x))^{\mathsf{T}} F(x, u_1, u_2) = 0, \tag{6}$$

where $\nabla V(x) = \partial V(x)/\partial x$ is the gradient of the cost function, $U(x, u_1, u_2) = Q(x) + u_1^{\mathsf{T}} R_1 u_1 + u_2^{\mathsf{T}} R_2 u_2$, $Q(x) \geq 0$, $Q(x) = 0$ if and only if $x = 0$, and $R_1 = R_1^{\mathsf{T}} > 0$ and $R_2 = R_2^{\mathsf{T}} > 0$ are constant matrices, then with the feedback control functions $u_1(x)$ and $u_2(x)$, there exists a neighborhood of the origin such that system (1) is asymptotically stable. Furthermore, if we define

$$J(x_0, u_1, u_2) = \int_0^\infty \big\{U(x(\tau), u_1(x(\tau)), u_2(x(\tau))) + \Gamma(x(\tau))\big\} d\tau \tag{7}$$

as the modified cost function of system (2), then we have $V(x_0) = J(x_0, u_1, u_2)$.

Proof First, we show the asymptotic stability of system (1) under the feedback control functions $u_1(x)$ and $u_2(x)$. Let

$$\dot{V}(x) \triangleq \frac{\mathrm{d}V(x)}{\mathrm{d}t} = (\nabla V(x))^{\mathsf{T}} \bar{F}(x, u_1, u_2). \tag{8}$$

Considering (5), (6), and (8), we can derive

$$\begin{aligned} \dot{V}(x(t)) &= (\nabla V(x))^{\mathsf{T}} F(x, u_1, u_2) + (\nabla V(x))^{\mathsf{T}} \Delta f(x) \\ &\leq (\nabla V(x))^{\mathsf{T}} F(x, u_1, u_2) + \Gamma(x) \\ &= -U(x, u_1, u_2) \\ &< 0 \end{aligned} \tag{9}$$

for any $x \neq 0$. This implies that $V(\cdot)$ is a Lyapunov function for system (1), which proves the asymptotic stability.

Next, note that (5) and (6) hold for any possible uncertainties. When $\Delta f(x) = 0$, we can easily derive that $\dot{V}(x) = (\nabla V(x))^{\mathsf{T}} F(x, u_1, u_2)$. According to (6), we obtain

$$\begin{aligned} U(x, u_1, u_2) + \Gamma(x) &= -\dot{V}(x) + U(x, u_1, u_2) + \Gamma(x) \\ &\quad + (\nabla V(x))^{\mathsf{T}} F(x, u_1, u_2) \\ &= -\dot{V}(x). \end{aligned} \tag{10}$$

By integrating over $[0, t)$, we have

$$\int_0^t \big\{U(x, u_1, u_2) + \Gamma(x)\big\} d\tau = -V(x(t)) + V(x_0). \tag{11}$$

Letting $t \to \infty$, we can easily find that $J(x_0, u_1, u_2) = V(x_0)$. This completes the proof. □

Lemma 1 *For any continuously differentiable function $V(x)$, if we define*

$$\Gamma(x) = h^{\mathsf{T}}(\varphi(x)) h(\varphi(x)) + \frac{1}{4} (\nabla V(x))^{\mathsf{T}} G(x) G^{\mathsf{T}}(x) \nabla V(x), \tag{12}$$

then, the relation $(\nabla V(x))^{\mathsf{T}} \Delta f(x) \leq \Gamma(x)$ holds.

Remark 2 This lemma can easily be proved by combining (3), (4), and (12) with the fact that $\xi^{\mathsf{T}}(x)\xi(x) \geq 0$, where

$$\xi(x) = d(\varphi(x)) - \frac{1}{2} G^{\mathsf{T}}(x) \nabla V(x). \tag{13}$$

Note that for system (1), with any continuously differentiable function $V(x)$, the bounded function $\Gamma(x)$ constructed as in (12) satisfies $(\nabla V(x))^{\mathsf{T}} \Delta f(x) \leq \Gamma(x)$. The importance of Lemma 1 lies in the fact that it presents a specific form of $\Gamma(x)$, which is significant in dealing with the dynamic uncertainty.

Remark 3 According to Theorem 1, the cost function $V(x)$, the bounded function $\Gamma(x)$, and feedback controls $u_1(x)$ and $u_2(x)$ satisfying (5) and (6) can guarantee the robust stabilization of system (1). It is important to notice that the optimal cost and optimal control of system (2) can provide specific forms of the cost function and feedback control. Hence, we should make great effort to solve the optimal control problem of system (2) with $V(x_0)$ considered as the cost function. In other words, we should minimize $J(x_0, u_1, u_2)$ with respect to u_1 and u_2.

Considering system (2), since

$$\begin{aligned} V(x_0) &= \int_0^\infty \big\{U(x, u_1, u_2) + \Gamma(x)\big\} d\tau \\ &= \int_0^T \big\{U(x, u_1, u_2) + \Gamma(x)\big\} d\tau + V(x(T)), \end{aligned} \tag{14}$$

we can find that

$$\lim_{T \to 0} \frac{1}{T} \bigg(V(x(T)) - V(x_0) + \int_0^T \big\{U(x, u_1, u_2) + \Gamma(x)\big\} d\tau \bigg) = 0, \tag{15}$$

which is equivalent to (6). Hence, (6) is an infinitesimal version of the modified cost function (14) and is the so-called nonlinear Lyapunov equation.

Now, for system (2) with modified cost function (14), we define the Hamiltonian function of the optimal control problem as

$$H(x,u_1,u_2,\nabla V(x)) = U(x,u_1,u_2)+\Gamma(x) +(\nabla V(x))^{\mathsf{T}}F(x,u_1,u_2). \tag{16}$$

Besides, the optimal cost function of system (2) can be defined as

$$J^*(x_0) = \min_{u_1,u_2\in\Psi(\Omega)}\int_0^\infty \big\{U(x(\tau),u_1(x(\tau)),u_2(x(\tau))) +\Gamma(x(\tau))\big\}\mathrm{d}\tau, \tag{17}$$

where $\Psi(\Omega)$ is the set of admissible controls on Ω. Note that $J^*(x)$ satisfies the modified HJB equation

$$0 = \min_{u_1,u_2\in\Psi(\Omega)} H(x,u_1,u_2,\nabla J^*(x)), \tag{18}$$

where $\nabla J^*(x) = \partial J^*(x)/\partial x$. Assume that the minimum on the right-hand side of (18) exists and is unique. Then, the optimal control of system (2) is

$$u_i^*(x) = -\frac{1}{2}R_i^{-1}g_i^{\mathsf{T}}(x)\nabla J^*(x),\quad i=1,2. \tag{19}$$

Hence, the modified HJB equation becomes

$$\begin{aligned}0 =\ & U(x,u_1^*,u_2^*)+(\nabla J^*(x))^{\mathsf{T}}F(x,u_1^*,u_2^*)\\ &+h^{\mathsf{T}}(\varphi(x))h(\varphi(x))+\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)\nabla J^*(x)\end{aligned} \tag{20}$$

with $J^*(0)=0$. Denoting $D_1 = g_1(x)R_1^{-1}g_1^{\mathsf{T}}(x)$ and $D_2 = g_2(x)R_2^{-1}g_2^{\mathsf{T}}(x)$ and substituting (19) into (20), we can obtain the formulation of the modified HJB equation in terms of $\nabla J^*(x)$ as follows:

$$\begin{aligned}0 =\ & Q(x)+(\nabla J^*(x))^{\mathsf{T}}f(x)+h^{\mathsf{T}}(\varphi(x))h(\varphi(x))\\ &-\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}(D_1+D_2)\nabla J^*(x)\\ &+\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)\nabla J^*(x)\end{aligned} \tag{21}$$

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

The following theorem proves that u_1^* and u_2^* can indeed realize the robust decentralized stabilization of system (1).

Theorem 2 *Let u_1^* and u_2^* given by (19) form the decentralized optimal control of system (2) with cost function (14). Then, the control pair (u_1^*,u_2^*) ensures robust decentralized stabilization of uncertain nonlinear system (1).*

Proof According to Lemma 1, we can find that

$$\begin{aligned}(\nabla J^*(x))^{\mathsf{T}}\Delta f(x) \le\ & h^{\mathsf{T}}(\varphi(x))h(\varphi(x))\\ &+\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)\nabla J^*(x).\end{aligned} \tag{22}$$

Then, we have

$$\begin{aligned}\dot J^*(x) =\ & (\nabla J^*(x))^{\mathsf{T}}\bar F(x,u_1^*,u_2^*)\\ =\ & (\nabla J^*(x))^{\mathsf{T}}F(x,u_1^*,u_2^*)+(\nabla J^*(x))^{\mathsf{T}}\Delta f(x)\\ \le\ & (\nabla J^*(x))^{\mathsf{T}}F(x,u_1^*,u_2^*)+h^{\mathsf{T}}(\varphi(x))h(\varphi(x))\\ &+\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)\nabla J^*(x).\end{aligned} \tag{23}$$

Since the HJB equation (20) implies that

$$\begin{aligned}&(\nabla J^*(x))^{\mathsf{T}}F(x,u_1^*,u_2^*)\\ &= -U(x,u_1^*,u_2^*)-h^{\mathsf{T}}(\varphi(x))h(\varphi(x))\\ &\quad -\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)\nabla J^*(x),\end{aligned} \tag{24}$$

we can further obtain that $\dot J^*(x)\le -U(x,u_1^*,u_2^*)<0$ holds for any $x\neq 0$. Therefore, the control pair (u_1^*,u_2^*) achieves robust decentralized stabilization of system (1). □

Next, let π_1 and π_2 be positive numbers. We can further obtain the following conclusion:

Theorem 3 *Let u_1^* and u_2^* given by (19) be the decentralized optimal control of system (2) with cost function (14). The control pair $(\pi_1u_1^*,\pi_2u_2^*)$ can ensure the robust decentralized stabilization of uncertain nonlinear system (1) provided that $\pi_1>1/2$ and $\pi_2>1/2$.*

Proof Based on (21), we have

$$\begin{aligned}(\nabla J^*(x))^{\mathsf{T}}f(x) =\ & -Q(x)-h^{\mathsf{T}}(\varphi(x))h(\varphi(x))\\ &+\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}(D_1+D_2)\nabla J^*(x)\\ &-\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)\nabla J^*(x).\end{aligned} \tag{25}$$

Hence, by combining (22) and (25), we can find that

$$\begin{aligned}\dot J^*(x) =\ & (\nabla J^*(x))^{\mathsf{T}}\bar F(x,\pi_1u_1^*,\pi_2u_2^*)\\ =\ & (\nabla J^*(x))^{\mathsf{T}}F(x,\pi_1u_1^*,\pi_2u_2^*)+(\nabla J^*(x))^{\mathsf{T}}\Delta f(x)\\ =\ & (\nabla J^*(x))^{\mathsf{T}}f(x)+(\nabla J^*(x))^{\mathsf{T}}\Delta f(x)\\ &+\pi_1(\nabla J^*(x))^{\mathsf{T}}g_1(x)u_1^*+\pi_2(\nabla J^*(x))^{\mathsf{T}}g_2(x)u_2^*\\ \le\ & -Q(x)+\frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}(D_1+D_2)\nabla J^*(x)\\ &+\pi_1(\nabla J^*(x))^{\mathsf{T}}g_1(x)u_1^*+\pi_2(\nabla J^*(x))^{\mathsf{T}}g_2(x)u_2^*.\end{aligned} \tag{26}$$

By observing (19), the equation (26) is in fact

$$\begin{aligned}\dot J^*(x) \le\ & -Q(x)-\frac{1}{2}\left(\pi_1-\frac{1}{2}\right)\left\|R_1^{-1/2}g_1^{\mathsf{T}}(x)\nabla J^*(x)\right\|^2\\ &-\frac{1}{2}\left(\pi_2-\frac{1}{2}\right)\left\|R_2^{-1/2}g_2^{\mathsf{T}}(x)\nabla J^*(x)\right\|^2,\end{aligned} \tag{27}$$

which implies that $\dot J^*(x)<0$ holds for any $x\neq 0$ and provided that $\pi_1>1/2$ and $\pi_2>1/2$. This completes the proof. □

Remark 4 According to Theorems 2 and 3, once the solution of the modified HJB equation (21) corresponding to system (2) is derived, we can establish the robust decentralized control scheme of system (1). Hence, we should put emphasis upon solving the modified HJB equation (21).

4 Online HJB solution of the decentralized optimal control problem based on neural networks

Note that for system (2), it is always difficult or even impossible to obtain the analytical solution of the modified HJB equation (21). Here, the ADP method is employed to help solving the problem by constructing a single neural network. Therefore, the idea of ADP is introduced to the framework of robust decentralized stabilization of nonlinear systems with multi-control-stations and dynamic uncertainties. Before proceeding, we recall the following assumption, which is typically used to facilitate designing the optimal control:

Assumption 1 (cf. Dierks and Jagannathan 2010; Nodland et al. 2013; Zhao et al. 2013) Consider system (2) with cost function (14) and optimal control (19). Let $J_s(x)$ be a continuously differentiable Lyapunov function candidate satisfying $\dot{J}_s(x) = (\nabla J_s(x))^{\mathsf{T}}(f(x) + g_1(x)u_1^* + g_2(x)u_2^*) < 0$, where $\nabla J_s(x) = \partial J_s(x)/\partial x$. Assume there exists a positive definite matrix $\Lambda(x)$ satisfying $\|\Lambda(x)\| = 0$ if and only if $\|x\| = 0$, such that $(\nabla J_s(x))^{\mathsf{T}}(f(x)+g_1(x)u_1^*+g_2(x)u_2^*) = -(\nabla J_s(x))^{\mathsf{T}}\Lambda(x)\nabla J_s(x)$ holds.

4.1 The critic network and online learning algorithm

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

$$V(x) = \omega_c^{\mathsf{T}}\sigma_c(x) + \varepsilon_c(x), \tag{28}$$

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 hidden layer, and $\varepsilon_c(x)$ is the approximation error of neural network. Then,

$$\nabla V(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x), \tag{29}$$

where $\nabla\sigma_c(x) = \partial\sigma_c(x)/\partial x$ and $\nabla\varepsilon_c(x) = \partial\varepsilon_c(x)/\partial x$. Here, we assume that the weight vector ω_c, the gradient $\nabla\sigma_c(x)$, and the approximation error $\varepsilon_c(x)$ and its derivative $\nabla\varepsilon_c(x)$ are all bounded on the compact set Ω (Vamvoudakis and Lewis 2010; Bhasin et al. 2013; Dierks and Jagannathan 2010).

In this paper, an artificial neural network, called critic network, is constructed to approximate the cost function as

$$\hat{V}(x) = \hat{\omega}_c^{\mathsf{T}}\sigma_c(x), \tag{30}$$

where $\hat{\omega}_c \in \mathbb{R}^l$ is the estimated weight of the ideal one and $\sigma_c(x)$ is selected such that $\hat{V}(x) > 0$ for any $x \neq 0$ and $\hat{V}(x) = 0$ when $x = 0$. Then, we have

$$\nabla\hat{V}(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c, \tag{31}$$

where $\nabla\hat{V}(x) = \partial\hat{V}(x)/\partial x$. Then, according to (19) and (29), we can derive the accurate expression of the optimal control function as

$$u_i(x) = -\frac{1}{2}R_i^{-1}g_i^{\mathsf{T}}(x)\big((\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x)\big), \quad i = 1, 2. \tag{32}$$

Besides, in light of (19) and (31), the corresponding approximate control function can be given as

$$\hat{u}_i(x) = -\frac{1}{2}R_i^{-1}g_i^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c, \quad i = 1, 2. \tag{33}$$

Applying the state feedback control function (33) to system (2), we can obtain the closed-loop system dynamics as

$$\dot{x} = f(x) - \frac{1}{2}(D_1 + D_2)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{34}$$

Based on (29), the Hamiltonian function (16) is in fact

$$\begin{aligned} H(x, \omega_c) = {} & Q(x) + \omega_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) \\ & -\frac{1}{4}\omega_c^{\mathsf{T}}\nabla\sigma_c(x)(D_1 + D_2)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c \\ & +h^{\mathsf{T}}(\varphi(x))h(\varphi(x)) \\ & +\frac{1}{4}\omega_c^{\mathsf{T}}\nabla\sigma_c(x)G(x)G^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + e_{cH} \\ = {} & 0, \end{aligned} \tag{35}$$

where

$$\begin{aligned} e_{cH} = {} & (\nabla\varepsilon_c(x))^{\mathsf{T}}f(x) \\ & -\frac{1}{2}(\nabla\varepsilon_c(x))^{\mathsf{T}}(D_1 + D_2)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c \\ & -\frac{1}{4}(\nabla\varepsilon_c(x))^{\mathsf{T}}(D_1 + D_2)\nabla\varepsilon_c(x) \\ & +\frac{1}{2}(\nabla\varepsilon_c(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c \\ & +\frac{1}{4}(\nabla\varepsilon_c(x))^{\mathsf{T}}G(x)G^{\mathsf{T}}(x)\nabla\varepsilon_c(x) \end{aligned} \tag{36}$$

is the residual error. Similarly, based on $\hat{\omega}_c$, we can derive the approximate Hamiltonian function as

$$\begin{aligned} \hat{H}(x, \hat{\omega}_c) = {} & Q(x) + \hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)f(x) \\ & -\frac{1}{4}\hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)(D_1 + D_2)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c \\ & +h^{\mathsf{T}}(\varphi(x))h(\varphi(x)) \\ & +\frac{1}{4}\hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)G(x)G^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c \\ \triangleq {} & e_c. \end{aligned} \tag{37}$$

Let the weight estimation error of the critic network be $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$. Then, by combining (35) with (37), we can obtain the expression of e_c with respect to $\tilde{\omega}_c$, which is useful to derive the dynamic information of $\tilde{\omega}_c$.

When training the critic network, it is desired to design $\hat{\omega}_c$ to minimize $E_c = (1/2)e_c^{\mathsf{T}}e_c$. The weights of the critic network are tuned based on the standard steepest descent algorithm with an additional term introduced to guarantee the boundedness of system state, i.e.,

$$\dot{\hat{\omega}}_c = -\alpha_c\left(\frac{\partial E_c}{\partial \hat{\omega}_c}\right) + \frac{1}{2}\alpha_s \Pi(x, \hat{u}_1, \hat{u}_2)\nabla\sigma_c(x)(D_1 + D_2)\nabla J_s(x), \quad (38)$$

where $\alpha_c > 0$ is the learning rate of the critic network, $\alpha_s > 0$ is the learning rate of the additional term, and $J_s(x)$ is the Lyapunov function candidate given in Assumption 1. Here, $\partial E_c/\partial\hat{\omega}_c = e_c(\partial e_c/\partial\hat{\omega}_c)$, where $\partial e_c/\partial\hat{\omega}_c$ can be derived from (37). The function $\Pi(x, \hat{u}_1, \hat{u}_2)$ denotes the additional stabilizing term defined based on the Lyapunov condition for stability, i.e.,

$$\Pi(x, \hat{u}_1, \hat{u}_2) = \begin{cases} 0, & \text{if } \dot{J}_s(x) = (\nabla J_s(x))^{\mathsf{T}} F(x, \hat{u}_1, \hat{u}_2) < 0, \\ 1, & \text{else.} \end{cases} \quad (39)$$

Note that the second term in (38) plays an important role of reinforcing the training process, which is conducted along the negative gradient direction of $(\nabla J_s(x))^{\mathsf{T}}F(x, \hat{u}_1, \hat{u}_2)$ with respect to $\hat{\omega}_c$, i.e.,

$$\begin{aligned} &-\frac{\partial\left((\nabla J_s(x))^{\mathsf{T}}F(x, \hat{u}_1, \hat{u}_2)\right)}{\partial\hat{\omega}_c} \\ &= -\left(\frac{\partial\hat{u}_1}{\partial\hat{\omega}_c}\right)^{\mathsf{T}}\frac{\partial\left((\nabla J_s(x))^{\mathsf{T}}F(x, \hat{u}_1, \hat{u}_2)\right)}{\partial\hat{u}_1} \\ &\quad -\left(\frac{\partial\hat{u}_2}{\partial\hat{\omega}_c}\right)^{\mathsf{T}}\frac{\partial\left((\nabla J_s(x))^{\mathsf{T}}F(x, \hat{u}_1, \hat{u}_2)\right)}{\partial\hat{u}_2} \\ &= \frac{1}{2}\nabla\sigma_c(x)(D_1 + D_2)\nabla J_s(x). \end{aligned} \quad (40)$$

The structural diagram of the online learning algorithm is illustrated in Fig. 1, noticing that the solid line and the dashed line represent the signal and the back-propagating path of the critic network, respectively.

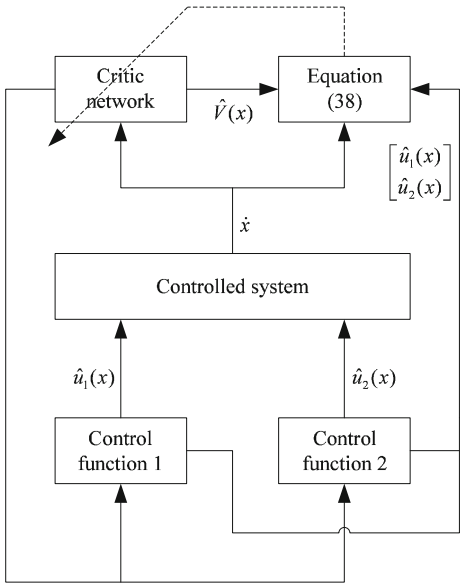

Fig. 1 Structural diagram of the online learning algorithm

4.2 Stability analysis of the closed-loop system

Theorem 4 *Consider the nonlinear system with two control stations given by* (2)*. Let the control inputs be provided by* (33) *and the weights of the critic network be tuned by* (38)*. Then, the state x and the weight estimation error $\tilde{\omega}_c$ are UUB. Moreover, the approximate control $\hat{u}_i$ given by* (33) *converges to optimal control u_i^* with bound ε_{iu}, $i = 1, 2$.*

Proof Here, we choose a Lyapunov function candidate as

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

where $J_s(x)$ is given in Assumption 1. Note that the dynamics of the weight estimation error can be obtained by considering the fact that $\dot{\tilde{\omega}}_c = -\dot{\hat{\omega}}_c$. Besides, the closed-loop system dynamics is presented in (34). Thus, by denoting that $A = \nabla\sigma_c(x)(D_1 + D_2)(\nabla\sigma_c(x))^{\mathsf{T}}$ and $B = \nabla\sigma_c(x)G(x)G^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}$, we can obtain the derivative of $L(t)$ with respect to time t, i.e.,

$$\begin{aligned} \dot{L}(t) =& \frac{1}{\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\dot{\tilde{\omega}}_c + \frac{\alpha_s}{\alpha_c}(\nabla J_s(x))^{\mathsf{T}}\dot{x} \\ =& -\left(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)\dot{x} - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}A\tilde{\omega}_c - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}B\tilde{\omega}_c \right. \\ &\left. + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}B\omega_c + e_{cH}\right) \\ &\times\left(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)\dot{x} - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}B\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}B\omega_c\right) \\ &-\frac{\alpha_s}{2\alpha_c}\Pi(x, \hat{u}_1, \hat{u}_2)\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)(D_1 + D_2)\nabla J_s(x) \\ &+\frac{\alpha_s}{\alpha_c}(\nabla J_s(x))^{\mathsf{T}}\dot{x}. \end{aligned} \quad (42)$$

Here, we assume that $\lambda_{1m} > 0$ and $\lambda_{1M} > 0$ are the lower and upper bounds of the norm of matrix A. Similarly, assume that $\lambda_{2m} > 0$ and $\lambda_{2M} > 0$ are the lower and upper bounds of the norm of matrix B. In addition, assume that $\|\nabla\sigma_c(x)\dot{x}\| \le \lambda_3$, $\|B\omega_c\| \le \lambda_4$, and $\|e_{cH}\| \le \lambda_5$, where λ_3, λ_4, and λ_5 are all positive constants. Then, we can find that

$$\begin{aligned} \dot{L}(t) \le& -\lambda_7\|\tilde{\omega}_c\|^4 + \lambda_8\|\tilde{\omega}_c\|^2 + \lambda_5^2 \\ &-\frac{\alpha_s}{2\alpha_c}\Pi(x, \hat{u}_1, \hat{u}_2)\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)(D_1 + D_2)\nabla J_s(x) \end{aligned}$$

$$+\frac{\alpha_s}{\alpha_c}(\nabla J_s(x))^{\mathsf{T}}\dot{x}, \tag{43}$$

where

$$\lambda_7 = \frac{1}{8}\lambda_{1m}\lambda_{2m} - \left(\frac{1}{8}\phi_1^2 + \frac{1}{16}\phi_2^2\right)\lambda_{1M}^2 + \frac{1}{8}\lambda_{2m}^2 - \left(\frac{3}{8}\phi_3^2 + \frac{3}{16}\phi_4^2\right)\lambda_{2M}^2, \tag{44}$$

$$\lambda_8 = \frac{1}{2}\lambda_{2M}\lambda_5 + \lambda_3\lambda_4 + \left(\frac{3}{4} + \frac{1}{8\phi_1^2} + \frac{3}{8\phi_3^2}\right)\lambda_3^2 + \left(\frac{1}{2} + \frac{1}{16\phi_2^2} + \frac{3}{16\phi_4^2}\right)\lambda_4^2 + \frac{1}{4}\lambda_5^2, \tag{45}$$

and ϕ_1, ϕ_2, ϕ_3, and ϕ_4 are constants chosen for the design purpose. In the following, the two cases, i.e., $\Pi(x, \hat{u}_1, \hat{u}_2) = 0$ and $\Pi(x, \hat{u}_1, \hat{u}_2) = 1$ will be considered, respectively.

Case 1 $\Pi(x, \hat{u}_1, \hat{u}_2) = 0$. Considering the fact that $(\nabla J_s(x))^{\mathsf{T}}\dot{x} < 0$, there exists a positive constant λ_6 such that $0 < \lambda_6\|\nabla J_s(x)\| \le -(\nabla J_s(x))^{\mathsf{T}}\dot{x}$, i.e., $(\nabla J_s(x))^{\mathsf{T}}\dot{x} \le -\lambda_6\|\nabla J_s(x)\|$. Then, the inequality (43) becomes

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

Hence, given the inequality

$$\|\tilde{\omega}_c\| \ge \sqrt{\frac{\lambda_8 + \sqrt{4\lambda_5^2\lambda_7 + \lambda_8^2}}{2\lambda_7}} \triangleq \mathscr{A}_1 \tag{47}$$

or

$$\|\nabla J_s(x)\| \ge \frac{\alpha_c(4\lambda_5^2\lambda_7 + \lambda_8^2)}{4\alpha_s\lambda_6\lambda_7} \triangleq \mathscr{B}_1 \tag{48}$$

holds, then $\dot{L}(t) < 0$.

Case 2 $\Pi(x, \hat{u}_1, \hat{u}_2) = 1$. Assume that $\|D_1 + D_2\| \le \lambda_9$ and $\|\nabla\varepsilon_c(x)\| \le \lambda_{10}$, where λ_9 and λ_{10} are also positive constants. Let λ_m be the minimum eigenvalue of the positive definite matrix $\Lambda(x)$; then we have $\lambda_m\|\nabla J_s(x)\|^2 \le (\nabla J_s(x))^{\mathsf{T}}\Lambda(x)\nabla J_s(x)$. By adding and subtracting the term $\alpha_s(\nabla J_s(x))^{\mathsf{T}}(D_1 + D_2)\nabla\varepsilon_c(x)/(2\alpha_c)$ to the right-hand side of (43) and according to Assumption 1, we have

$$\dot{L}(t) \le -\lambda_7\|\tilde{\omega}_c\|^4 + \lambda_8\|\tilde{\omega}_c\|^2 + \lambda_5^2 - \frac{\alpha_s}{\alpha_c}\lambda_m\|\nabla J_s(x)\|^2 + \frac{\alpha_s}{2\alpha_c}\lambda_9\lambda_{10}\|\nabla J_s(x)\|. \tag{49}$$

Hence, given the inequality

$$\|\tilde{\omega}_c\| \ge \sqrt{\frac{\lambda_8}{2\lambda_7} + \sqrt{\frac{\lambda_5^2}{\lambda_7} + \frac{\lambda_8^2}{4\lambda_7^2} + \frac{\alpha_s\lambda_9^2\lambda_{10}^2}{16\alpha_c\lambda_m\lambda_7}}} \triangleq \mathscr{A}_2 \tag{50}$$

or

$$\|\nabla J_s(x)\| \ge \frac{\lambda_9\lambda_{10}}{4\lambda_m} + \sqrt{\frac{\alpha_c(4\lambda_5^2\lambda_7 + \lambda_8^2)}{4\alpha_s\lambda_m\lambda_7} + \frac{\lambda_9^2\lambda_{10}^2}{16\lambda_m^2}} \triangleq \mathscr{B}_2 \tag{51}$$

holds, then $\dot{L}(t) < 0$.

Thus, if the inequality $\|\tilde{\omega}_c\| > \max(\mathscr{A}_1, \mathscr{A}_2) = \mathscr{A}$ or $\|\nabla J_s(x)\| > \max(\mathscr{B}_1, \mathscr{B}_2) = \mathscr{B}$ holds, then $\dot{L}(t) < 0$. According to the standard Lyapunov extension theorem (Lewis et al. 1999), we can derive that x and $\tilde{\omega}_c$ are UUB.

Next, according to (32) and (33), we have

$$u_i^* - \hat{u}_i = -\frac{1}{2}R_i^{-1}g_i^{\mathsf{T}}(x)\big((\nabla\sigma(x))^{\mathsf{T}}\tilde{\omega}_c + \nabla\varepsilon_c(x)\big), \tag{52}$$

where $i = 1, 2$. Assume that $\|R_i^{-1}\| \le R_{iM}^{-1}$, $\|g_i(x)\| \le g_{iM}$, and $\|\nabla\sigma(x)\| \le \sigma_{dM}$. Then, we can further obtain

$$\|u_i^* - \hat{u}_i\| \le \frac{1}{2}R_{iM}^{-1}g_{iM}(\sigma_{dM}\mathscr{A} + \lambda_{10}) \triangleq \varepsilon_{iu}, \tag{53}$$

where $i = 1, 2$. This completes the proof. □

5 Simulation examples

Example 1 Consider the continuous-time nonlinear system with two control stations and dynamic uncertainty

$$\dot{x} = \begin{bmatrix} -x_1 - 2x_2 \\ x_1 - 4x_2 - \cos x_1 \sin x_2^2 \end{bmatrix} + \begin{bmatrix} 1 \\ -3 \end{bmatrix} u_1 + \begin{bmatrix} 2 \\ 1 \end{bmatrix} u_2 + \Delta f(x), \tag{54}$$

where $x = [x_1, x_2]^{\mathsf{T}}$ and $\Delta f(x) = [px_1 \sin x_2, 0]^{\mathsf{T}}$ with $p \in [-0.5, 0.5]$. We choose $G(x) = [1, 0]^{\mathsf{T}}$ and $\varphi(x) = x$. Then, we have $d(\varphi(x)) = px_1 \sin x_2$. Hence, we can select $h(\varphi(x)) = 0.5x_1 \sin x_2$. Considering the nominal system with modified cost function, we let $Q(x) = x^{\mathsf{T}}x$, $R_1 = R_2 = I$, where I is an identity matrix with suitable dimension. For the purpose of solving the decentralized optimal control problem, a critic network is constructed as $\hat{V}(x) = \hat{\omega}_{c1}x_1^2 + \hat{\omega}_{c2}x_1x_2 + \hat{\omega}_{c3}x_2^2$.

In this example, we let the initial state of the controlled plant be $x_0 = [1, -1]^{\mathsf{T}}$. The Lyapunov function candidate $J_s(x)$ can be obtained by selecting a quadratic polynomial, such as $J_s(x) = (1/2)x^{\mathsf{T}}x$. Set the learning rates as $\alpha_c = 0.8$ and $\alpha_s = 0.5$. When conducting the online optimal control algorithm, an exploration noise described by $\mathscr{N}(t) = \sin^2(t)\cos(t) + \sin^2(2t)\cos(0.1t) + \sin^2(-1.2t)\cos(0.5t) + \sin^5(t)$ is introduced to satisfy the persistency of excitation condition. The evolution of system state is illustrated in Fig. 2. After a learning session, the weights of the critic network converge to $[0.2736, -0.1035, 0.1285]^{\mathsf{T}}$ as shown in Fig. 3, where the legends ω_{ac1}, ω_{ac2}, and ω_{ac3} represent the elements $\hat{\omega}_{c1}$, $\hat{\omega}_{c2}$, and $\hat{\omega}_{c3}$, respectively. It should be pointed out that the initial weights of the critic network are all set to zero, which implies that the initial stabilizing control is not needed in the developed approach.

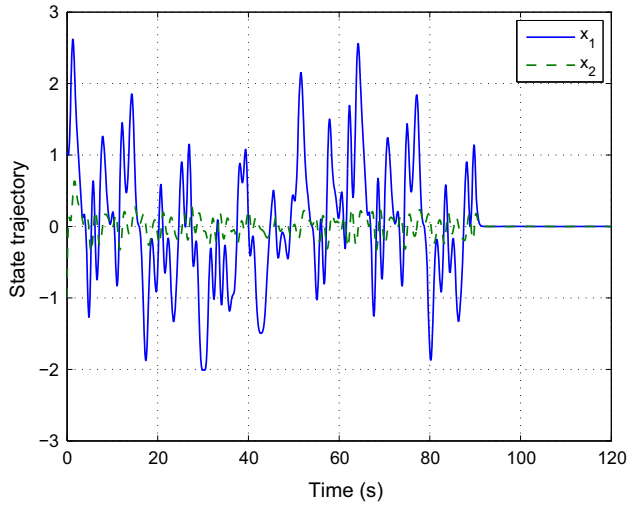

Fig. 2 Evolution of system state during the experiment

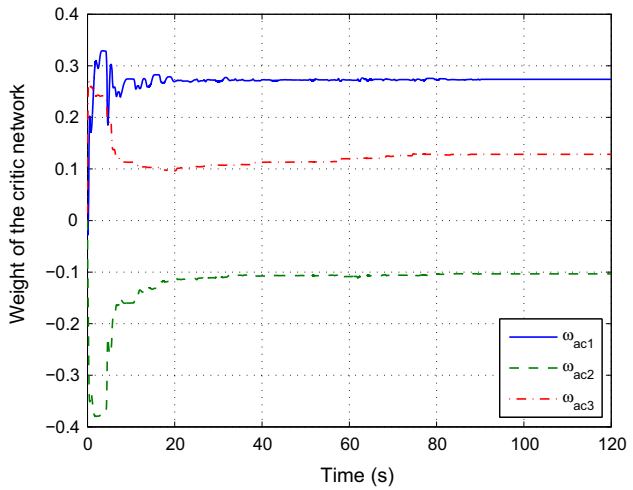

Fig. 3 Convergence of weight vector of the critic network

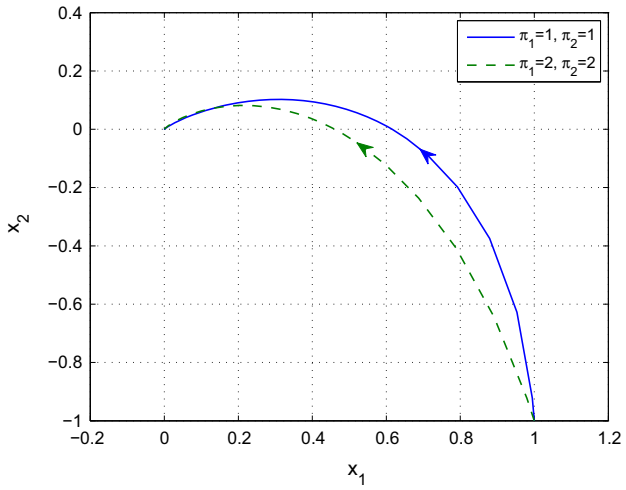

Fig. 4 The state trajectory ($p = -0.5$)

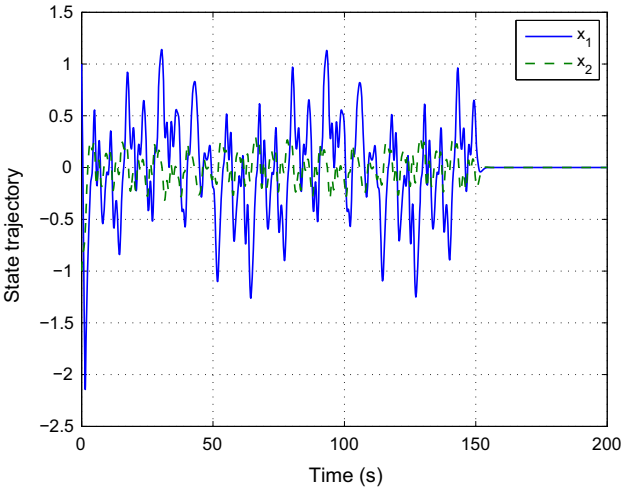

Fig. 5 Evolution of system state during the experiment

Next, based on the converged weight vector, the scalar parameter $p = -0.5$ is chosen for evaluating the control performance. The robust decentralized controllers when choosing $(\pi_1 = 1, \pi_2 = 1)$ and $(\pi_1 = 2, \pi_2 = 2)$ are applied to system (54), respectively. The system trajectories of the first 10 seconds are presented in Fig. 4, which verifies the conclusions of Theorems 2 and 3.

Example 2 Consider the continuous-time nonlinear system described by

$$\dot{x} = \begin{bmatrix} -0.5x_1 + x_2 + 0.5x_2^3 \\ -\sin x_1 \cos x_2 - 0.5x_2 - 0.5x_2^3 \end{bmatrix} + \begin{bmatrix} 0 \\ -0.5 \end{bmatrix} u_1 + \begin{bmatrix} -1 \\ 0 \end{bmatrix} u_2 + \Delta f(x), \tag{55}$$

where $x = [x_1, x_2]^{\mathsf{T}}$ is the state vector. The system uncertainty is $\Delta f(x) = [p_1 x_1 \sin x_2, p_2 x_1^3 \cos x_2]^{\mathsf{T}}$ with $p_1 \in [-0.5, 0.5]$ and $p_2 \in [-0.2, 0.2]$. In this example, we also choose $\varphi(x) = x$. However, according to the form of dynamic uncertainty, the $G(x)$ and $d(\varphi(x))$ are chosen as

$$G(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, d(\varphi(x)) = \begin{bmatrix} p_1 x_1 \sin x_2 \\ p_2 x_1^3 \cos x_2 \end{bmatrix}. \tag{56}$$

Then, we can select $h(\varphi(x)) = [0.5x_1 \sin x_2, 0.2x_1^3 \cos x_2]^{\mathsf{T}}$. Other parameters are set the same as Example 1. After carrying out a sufficient learning session in an online fashion, the evolution of system state during the experiment is given in Fig. 5. In addition, the weights of the critic network converge to $[0.9496, 0.1291, 1.0650]^{\mathsf{T}}$, which can be seen clearly by observing Fig. 6.

Next, using the converged weight vector, the scalar parameters $p_1 = 0.5$ and $p_2 = -0.2$ are given for evaluating the control performance. Based on the decentralized optimal controller, i.e., the case $\pi_1 = 1$ and $\pi_2 = 1$, the system trajectory of system (55) during the first 20 s is illustrated

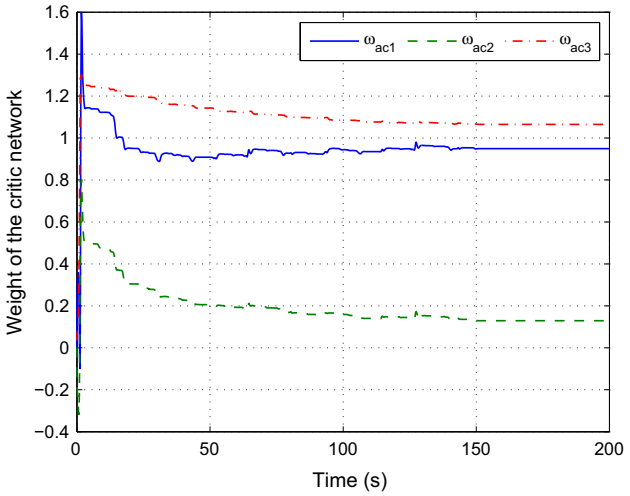

Fig. 6 Convergence of weight vector of the critic network

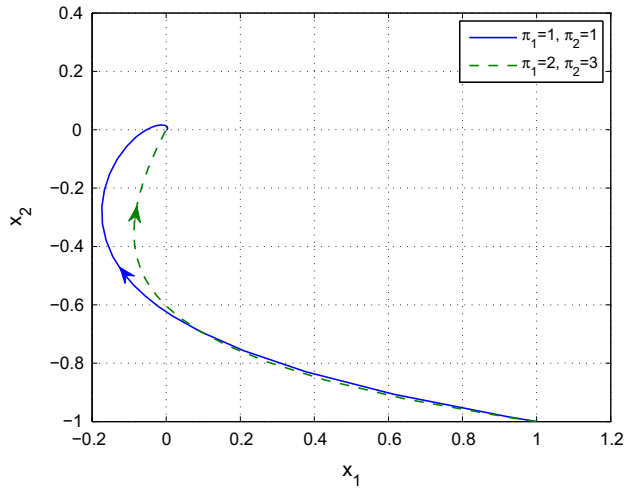

Fig. 7 The state trajectory ($p_1 = 0.5$, $p_2 = -0.2$)

in Fig. 7. Remarkably, as the statement of Theorem 3, the robust decentralized stabilization of system (55) also can be achieved when choosing $\pi_1 = 2$ and $\pi_2 = 3$ (see Fig. 7). To summarize, these results verify the effectiveness of the present control approach.

6 Conclusions

A novel strategy for robust decentralized stabilization of uncertain nonlinear systems is established. This is accomplished by properly modifying the cost function to account for system uncertainty so that the solution of the transformed decentralized optimal control problem is the robust decentralized controller of the uncertain nonlinear system. A critic network is constructed to solve the modified HJB equation online. Two numerical examples are provided to reinforce the theoretical results.

Actually, the extension of the developed method to the case that the original system contains N control stations can also be obtained. Consider the nonlinear system with N control stations

$$\begin{aligned}\dot{x}(t) &= \bar{F}(x(t), u_1(t), u_2(t), \ldots, u_N(t)) \\ &= f(x(t)) + \sum_{i=1}^{N} g_i(x(t))u_i(t) + \Delta f(x(t)).\end{aligned} \tag{57}$$

The corresponding nominal system is

$$\begin{aligned}\dot{x}(t) &= F(x(t), u_1(t), u_2(t), \ldots, u_N(t)) \\ &= f(x(t)) + \sum_{i=1}^{N} g_i(x(t))u_i(t).\end{aligned} \tag{58}$$

Under such circumstances, based on the established online optimal control strategy, we can derive N state feedback control functions $u_i(x)$, $i = 1, 2, \ldots, N$, such that the control pair $(u_1(x), u_2(x), \ldots, u_N(x))$ achieves robust decentralized stabilization of system (57).

References

Abu-Khalaf M, Lewis FL (2005) Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. Automatica 41(5):779–791

Adhyaru DM, Kar IN, Gopal M (2011) Bounded robust control of nonlinear systems using neural network-based HJB solution. Neural Comput Appl 20(1):91–103

Al-Tamimi A, Lewis FL, Abu-Khalaf M (2008) Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. IEEE Trans Syst Man Cybern Part B Cybern 38(4):943–949

Bhasin S, Kamalapurkar R, Johnson M, Vamvoudakis KG, Lewis FL, Dixon WE (2013) A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. Automatica 49(1):82–92

Dierks T, Jagannathan S (2012) 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(7):1118–1129

Dierks T, Jagannathan S (2010) Optimal control of affine nonlinear continuous-time systems. In: Proceedings of the American control conference, Baltimore, MD, USA, June 2010, pp 1568–1573

Haddad WM, Chellaboina V, Fausz JL (1998) Robust nonlinear feedback control for uncertain linear systems with nonquadratic performance criteria. Syst Control Lett 33(5):327–338

Haddad WM, Chellaboina V, Fausz JL (2000) Optimal non-linear robust control for non-linear uncertain systems. Int J Control 73(4):329–342

Heydari A, Balakrishnan SN (2013) Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics. IEEE Trans Neural Netw Learn Syst 24(1):145–157

Jiang ZP, Jiang Y (2013) Robust adaptive dynamic programming for linear and nonlinear systems: an overview. Eur J Control 19(5):417–425

Lewis FL, Jagannathan S, Yesildirek A (1999) Neural network control of robot manipulators and nonlinear systems. Taylor & Francis, London

Lewis FL, Vrabie D, Vamvoudakis KG (2012) Reinforcement learning and feedback control: using natural decision methods to design optimal adaptive controllers. IEEE Control Syst Mag 32(6):76–105

Lin F (2000) An optimal control approach to robust control design. Int J Control 73(3):177–186

Liu D, Wang D, Zhao D, Wei Q, Jin N (2012) 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(3):628–634

Liu D, Wang D, Yang X (2013a) An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs. Inf Sci 220:331–342

Liu D, Li H, Wang D (2013b) Neural-network-based zero-sum game for discrete-time nonlinear systems via iterative adaptive dynamic programming algorithm. Neurocomputing 110:92–100

Liu D, Li H, Wang D (2013c) Data-based self-learning optimal control: research progress and prospects. Acta Automatica Sinica 39(11):1858–1870

Liu D, Wang D, Li H (2014a) Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach. IEEE Trans Neural Netw Learn Syst 25(2):418–428

Liu D, Li H, Wang D (2014b) Online synchronous approximate optimal learning algorithm for multiplayer nonzero-sum games with unknown dynamics. IEEE Trans Syst Man Cybern Syst 44(8):1015–1027

Ni Z, He H, Wen J (2013) Adaptive learning in tracking control based on the dual critic network design. IEEE Trans Neural Netw Learn Syst 24(6):913–928

Ni Z, He H (2013) Heuristic dynamic programming with internal goal representation. Soft Comput 17(11):2101–2108

Nodland D, Zargarzadeh H, Jagannathan S (2013) Neural network-based optimal adaptive output feedback control of a helicopter UAV. IEEE Trans Neural Netw Learn Syst 24(7):1061–1073

Vamvoudakis KG, Lewis FL (2010) Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. Automatica 46(5):878–888

Wang D, Liu D, Wei Q (2012a) Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach. Neurocomputing 78(1):14–22

Wang D, Liu D, Wei Q, Zhao D, Jin N (2012b) Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming. Automatica 48(8):1825–1832

Wang D, Liu D, Li H (2014) Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems. IEEE Trans Autom Sci Eng 11(2):627–632

Werbos PJ (1992) Approximate dynamic programming for real-time control and neural modeling. In: White DA, Sofge DA (eds) Proceedings of handbook of intelligent control: neural, fuzzy, and adaptive approaches, ch 13, Van Nostrand Reinhold, New York

Wu HN, Luo B (2012) Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control. IEEE Trans Neural Netw Learn Syst 23(12):1884–1895

Yang X, Liu D, Wang D (2014) Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints. Int J Control 87(3):553–566

Zhang H, Luo Y, Liu D (2009) Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. IEEE Trans Neural Netw 20(9):1490–1503

Zhang H, Cui L, Luo Y (2013) Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP. IEEE Trans Cybern 43(1):206–216

Zhang D, Liu D, Wang D (2014) Approximate optimal solution of the DTHJB equation for a class of nonlinear affine systems with unknown dead-zone constraints. Soft Comput 18(2):349–357

Zhao Q, Xu H, Dierks T, Jagannathan S (2013) Finite-horizon neural network-based optimal control design for affine nonlinear continuous-time systems. In: Proceedings of the international joint conference on neural networks, Dallas, TX, USA, Aug 2013, pp 1–6