CrossMark

Data-based robust optimal control of continuous-time affine nonlinear systems with matched uncertainties ☆

Ding Wang [a,c,*], Chao Li [a], Derong Liu [b], Chaoxu Mu [c]

[a] *The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[b] *School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*
[c] *Tianjin Key Laboratory of Process Measurement and Control, School of Electrical Engineering and Automation, Tianjin University, Tianjin, 300072, China*

ABSTRACT

In this paper, the robust optimal control of continuous-time affine nonlinear systems with matched uncertainties is investigated by using a data-based integral policy iteration approach. It is a natural extension of the traditional optimal control design, under the framework of adaptive dynamic programming (ADP) method, to robust optimal control of nonlinear systems with matched uncertainties. In theoretical aspect, by increasing a feedback gain to the optimal controller of the nominal system, the robust controller of the matched uncertain system is obtained, which also achieves optimality with a newly well-defined cost function. When regarding the implementation, the data-based integral policy iteration algorithm is used to solve the Hamilton–Jacobi–Bellman equation corresponding to the nominal system with completely unknown dynamics information. Then, the actor-critic technique based on neural networks and least squares implementation method are employed to facilitate deriving the optimal control law iteratively, so that the closed-form expression of the robust optimal controller is available. Additionally, two simulation examples with application backgrounds are presented to illustrate the effectiveness of the established robust optimal control scheme. In summary, it is important to note that the result developed in this paper broadens the application scope of ADP-based optimal control approach to more general nonlinear systems possessing dynamical uncertainties.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The phenomenon of dynamical uncertainties is common in practical control systems. From the literature of modern nonlinear control, it is known that the presence of dynamical uncertainties makes the feedback control problem extremely challenging in the context of nonlinear systems. As a result, the problem of designing adaptive and robust controller for nonlinear systems with uncertainties has attained considerable attention [5,10,13,16,24,33–35,38,45,51–53,57]. Among them, Mu et al. [35] proposed a general design scheme of finite-time switching mode manifolds and corresponding nonsingular

controllers. Yang et al. [57] proposed a networked-predictive-control scheme to compensate for the network-induced delay, so that the problem of output feedback controller design for networked control systems with mixed communication delays can be addressed. These results are beneficial to the development of the modern control theory. Additionally, Lin et al. [24] showed that the robust control problem can be solved by means of studying the optimal control problem of the related nominal system, but the detailed procedure was not presented in that paper. Wang et al. [45] constructed a novel strategy to achieve robust stabilization for a class of uncertain nonlinear systems based on the online policy iteration algorithm. However, the optimality of robust controller with respective to a newly defined cost function was not taken into consideration. Moreover, the complete system dynamics, which were usually difficult to obtain for practical systems, were required during the algorithm implementation. To the best of our knowledge, there are no results on designing robust optimal control for uncertain nonlinear systems by using a data-based policy iteration approach. This is the motivation of our research. Actually, it is the first time to establish the robust optimal control method for a class of nonlinear systems possessing uncertainties via a data-based integral policy iteration learning technique with completely unknown dynamics.

Nowadays, the data-based control design has become a hot topic in the field of control theory and control engineering [12,29]. In this paper, the starting point of the obtained strategy is the data-based optimal control design. Note that studying the nonlinear optimal control problem always requires to solve the Hamilton–Jacobi–Bellman (HJB) equation. Though dynamic programming has been a classical method in solving optimization and optimal control problems, it often encounters the phenomenon of "curse of dimensionality" [2]. For avoiding this difficulty, adaptive/approximate dynamic programming (ADP) was introduced by Werbos [50] and Prokhorov and Wunsch [39] as an effective method to solve the optimal control problem forward-in-time, based on function approximation structures [7,8,21,25,32,42,46], such as neural networks, support vector machine, fuzzy logic, etc. Reinforcement learning is another computational method which can interactively find an optimal policy from the learning process between the agent and the environment. Remarkably, Lewis and Liu [20], and Lewis and Vrabie [21] have given some opinions that the idea of ADP is very closely related to the framework of reinforcement learning. Recently, the researches on ADP and reinforcement learning have gained much attention from scholars of numerous fields [3,6,11,23,25–28,30,32,36,37,41,42,44,46–49,54–56,58–62]. Among the various results, robust ADP was developed for the design of robust optimal controllers for linear and nonlinear systems subject to both parametric and dynamic uncertainties by Jiang and Jiang [16], in order to broaden the application scope of ADP theory in the presence of dynamic uncertainties. In addition, Jiang and Jiang [13] also extended the robust ADP approach to decentralized optimal control of a class of large-scale systems with uncertainties. Note that in [16], the control signal of the nonlinear system was only one-dimensional and the optimization issue with regard to the original uncertain nonlinear system was not presented, while in [13], the robust decentralized control approach was only suitable for a class of linear systems. These inevitably restrict the effect of the proposed methods to some extent.

In the existing literature of ADP-based optimal control, either policy iteration or value iteration is employed to solve the Bellman equation or the HJB equation. The information of control matrix is necessary when employing the traditional policy iteration algorithms. However, in many situations, it is difficult to acquire the accurate model of controlled plant. The ADP and reinforcement learning schemes, which have the learning and optimization capabilities, can relax the requirement for a complete and accurate model of the controlled plant, by virtue of considering compact parameterized function representations whose parameters can be adjusted through learning and adaption. Jiang and Jiang [15] presented a novel policy iteration approach for continuous-time linear systems with completely unknown dynamics. Vrabie and Lewis [43] derived an integral reinforcement learning method to obtain direct adaptive optimal control for nonlinear input-affine continuous-time systems with partially unknown dynamics. Lee et al. [18,19] presented an integral reinforcement learning algorithm for continuous-time systems without the exact knowledge of the system dynamics. Liu et al. [25] developed a neural-network-based decentralized control strategy of a class of continuous-time nonlinear interconnected systems without requirement of dynamical information. Bian et al. [4] proposed a novel optimal control design approach for continuous-time nonaffine nonlinear systems with unknown dynamics by the idea of ADP. However, the system uncertainties were not considered in the above results.

With this background, how to further extend the application scope of ADP approach to more general nonlinear systems with dynamic uncertainties arouses our wide concern. In this paper, we investigate the data-based robust optimal control of continuous-time nonlinear systems with matched uncertainties. To begin with, the problem statement and some preliminaries are provided. It is proved that the improvement of the optimal control law is nothing but the robust controller of the original uncertain system, which also attains the property of optimality with a newly defined cost function. This serves as the main theoretical result of the paper. Then, the optimal controller of the nominal system is obtained by the data-based integral policy iteration algorithm and the neural network technique with completely unknown system dynamics, which is regarded as the primary implementation procedure. At last, two simulation examples are given to show the good response performance of the present robust optimal control scheme.

2. Problem statement and preliminaries

In this paper, we study a class of continuous-time nonlinear systems with input-affine structure and matched uncertainties described as

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

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

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

we assume that $f(x) + g(x)u$ is Lipschitz continuous on a set $\Omega \subset \mathbb{R}^n$ containing the origin and that the system (2) is controllable.

For the purpose of designing the robust control of system (1), we should find a feedback control function $\bar{u}(x)$, such that the closed-loop system attains globally asymptotical stability for all uncertainties $\bar{d}(x)$. In the following, the relationship between the robust control of system (1) and optimal control of its nominal system will be investigated to facilitate solving the robust stabilization problem.

For convenience of analysis, we denote $d(x) = R^{1/2}\bar{d}(x)$, where $R \in \mathbb{R}^{m \times m}$ is a symmetric positive definite matrix and $d(x) \in \mathbb{R}^m$ is bounded by a known function $d_M(x)$ (i.e., $\|d(x)\| \le d_M(x)$ holds with $d_M(0) = 0$). When regarding system (2), in order to conduct the infinite horizon optimal control design, we should find the feedback control law $u(x)$ to minimize the cost function given by

$$\begin{aligned} J(x_0) &= \int_0^\infty \left\{ d_M^2(x(\tau)) + u^\mathsf{T}(x(\tau))Ru(x(\tau)) \right\} \mathrm{d}\tau \\ &= \int_0^\infty r(x(\tau), u(x(\tau))) \mathrm{d}\tau, \end{aligned} \tag{3}$$

where $r(x(t), u(x(t)))$ is seen as the utility function.

Recalling the optimal control theory [22,31], the designed feedback control law must not only stabilize the system on Ω, but also make sure that the cost function $J(x_0)$ is finite. That is to say, the control law must be admissible as defined in [42]. Denote $\Psi(\Omega)$ be the set of admissible control laws on Ω. For any admissible control $u \in \Psi(\Omega)$, if the related cost function (3) is continuously differentiable, then its infinitesimal version is the nonlinear Lyapunov equation that can be written as the form

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

with $J(0) = 0$, where $\nabla(\cdot) \triangleq \partial(\cdot)/\partial x$ denotes the gradient operation. Here, for instance, $\nabla J(x) = \partial J(x)/\partial x$.

Define the Hamiltonian function and the optimal cost function of system (2) as

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

and

$$J^*(x_0) = \min_{u \in \Psi(\Omega)} \int_0^\infty r(x(\tau), u(x(\tau))) \mathrm{d}\tau, \tag{6}$$

respectively, with $J^*(0) = 0$. Recalling the optimal control theory [22,31], we know that the optimal cost function $J^*(x)$ satisfies the HJB equation

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

The optimal feedback control law can be obtained by the formula

$$\begin{aligned} u^*(x) &= \arg \min_{u \in \Psi(\Omega)} H(x, u, \nabla J^*(x)) \\ &= -\frac{1}{2}R^{-1}g^\mathsf{T}(x)\nabla J^*(x). \end{aligned} \tag{8}$$

Based on (5) and (8), the HJB Eq. (7) takes the following form:

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

with $J^*(0) = 0$. Generally speaking, it is difficult to obtain $J^*(x)$ by solving (9) directly, hence the optimal control $u^*(x)$ is also not easy to derive. Fortunately, some successive approximation methods have been proposed to solve the problem iteratively.

3. Robust optimal control design of uncertain nonlinear systems: the theoretical result

In this part, we aim at establishing the relationship between the robust controller of uncertain nonlinear system (1) and optimal controller of its nominal system. By modifying the optimal control law (8) via a feedback gain as the form

$$\bar{u}(x) = \pi u^*(x) = -\frac{1}{2}\pi R^{-1}g^\mathsf{T}(x)\nabla J^*(x), \tag{10}$$

we can develop the robust optimal control strategy of system (1). This result has been given in [47,49] with theoretical proof. Here, for consistency, we recall the result by presenting the following two lemmas with brief analysis.

Lemma 1. *For the system (2), the feedback control law given by (10) ensures that the closed-loop system is asymptotically stable for all $\pi \geq 1/2$. Furthermore, for the system (1), there exists a positive number $\pi_1^* \geq 1$, such that for any feedback gain $\pi > \pi_1^*$, the control law developed by (10) ensures that the closed-loop system is asymptotically stable.*

Proof. The proof is divided into two parts as follows.

(1) Show the asymptotic stability of system (2) under the action of feedback control law (10). We show that the optimal cost function $J^*(x(t))$ is a Lyapunov function. From (6), it can be observed that $J^*(x(t))$ is a positive definite function for $x \neq 0$. In addition, based on (9) and (10), we find that the derivative of $J^*(x(t))$ along the trajectory of the closed-loop system (2) can be expressed as

$$\dot{J}^*(x(t)) = (\nabla J^*(x))^\mathsf{T}(f(x) + g(x)\bar{u}(x))$$
$$= -d_M^2(x) - \frac{1}{2}\left(\pi - \frac{1}{2}\right)\left\|R^{-1/2}g^\mathsf{T}(x)\nabla J^*(x)\right\|^2.$$

It can be found that $\dot{J}^*(x(t)) < 0$ whenever $\pi \geq 1/2$ and $x \neq 0$. Hence, the conditions for Lyapunov local stability theory are satisfied and the closed-loop system is asymptotically stable.

(2) Show the asymptotic stability of system (1) under the action of feedback control law (10). We select $L(t) = J^*(x(t))$ as the Lyapunov function candidate. By taking the time derivative of the Lyapunov function $L(t)$ along the trajectory of the closed-loop system (1) and letting $\xi = \left[d_M(x), \left\|R^{-1/2}g^\mathsf{T}(x)\nabla J^*(x)\right\|\right]^\mathsf{T}$, we can derive the following formula

$$\dot{L}(t) = (\nabla J^*(x))^\mathsf{T}\left(f(x) + g(x)(\bar{u}(x) + \bar{d}(x))\right)$$
$$\leq -\left\{d_M^2(x) + \frac{1}{2}\left(\pi - \frac{1}{2}\right)\left\|R^{-1/2}g^\mathsf{T}(x)\nabla J^*(x)\right\|^2 - \left\|R^{-1/2}g^\mathsf{T}(x)\nabla J^*(x)\right\|d_M(x)\right\}$$
$$= -\xi^\mathsf{T}\left[\begin{array}{c|c} 1 & -\dfrac{1}{2} \\ \hline -\dfrac{1}{2} & \dfrac{1}{2}\left(\pi - \dfrac{1}{2}\right) \end{array}\right]\xi. \tag{11}$$

By observing (11), there exists a positive number $\pi_1^* \geq 1$ such that any $\pi > \pi_1^*$ guarantees $\dot{L}(t) < 0$. This indicates the asymptotical stability of the closed-loop system. $\square$

From Lemma 1, it is found that $\bar{u}(x)$ is the robust stabilizing control law of system (1) for any gain $\pi > \pi_1^*$. In the following, we show that it also holds the property of optimality with appropriate feedback gain. For the system (1), we define the following cost function:

$$\bar{J}(x_0) = \int_0^\infty \left\{Q(x(\tau)) + \frac{1}{\pi}\bar{u}^\mathsf{T}(x(\tau))R\bar{u}(x(\tau))\right\}d\tau, \tag{12}$$

where

$$Q(x) = d_M^2(x) - (\nabla J^*(x))^\mathsf{T}g(x)\bar{d}(x) + \frac{1}{4}(\pi - 1)(\nabla J^*(x))^\mathsf{T}g(x)R^{-1}g^\mathsf{T}(x)\nabla J^*(x). \tag{13}$$

By adding and subtracting $(1/(\pi - 1))d^\mathsf{T}(x)d(x)$ to (13) and considering the bounded condition $\|d(x)\| \leq d_M(x)$, we can obtain the inequality $Q(x) \geq ((\pi - 2)/(\pi - 1))d_M^2(x)$. It can be observed that there exists a positive number $\pi_2^* \geq 2$ such that for all $\pi > \pi_2^*$, the function $Q(x)$ is positive definite. The cost function (12) for the uncertain system (1) is well defined in this sense. The following lemma presents the optimality of robust control law of the uncertain nonlinear system (1).

Lemma 2. *Consider the system (1) with the newly defined cost function (12). There exists a positive number π^* such that for any feedback gain $\pi > \pi^*$, the feedback control law computed by (10) can achieve robust optimal control of the original uncertain nonlinear system.*

Proof. Let the Hamiltonian function of the system (1) with the newly defined cost function (12) be

$$\bar{H}(\nabla\bar{J}(x)) = Q(x) + \frac{1}{\pi}\bar{u}^\mathsf{T}(x)R\bar{u}(x) + (\nabla\bar{J}(x))^\mathsf{T}\left(f(x) + g(x)(\bar{u}(x) + \bar{d}(x))\right),$$

where $\pi > \pi_2^* \geq 2$. Replacing $\bar{J}(x)$ with $J^*(x)$, using (9) and (10), and observing (13), we can acquire that $\bar{H}(\nabla J^*(x)) = 0$. This means that $J^*(x)$ is a solution of the HJB equation of the system (1). Then, we say that the control law (10) attains optimality with cost function (12). Considering Lemma 1, there exists a positive number $\pi^* \triangleq \max\{\pi_1^*, \pi_2^*\}$ such that for any $\pi > \pi^*$, the control law (10) achieves robust optimal control of the original uncertain nonlinear system. This completes the robust optimal control design. $\square$

According to Lemma 2, we should aim at solving the optimal control problem of the nominal system and then attain the robust optimal controller of the original system. Due to the powerfulness of ADP approach with respective to nonlinear optimal control problem, we will employ a data-based control method to design robust optimal control using actor-critic structure and neural network technique. The next section presents the implementation procedure in detail.

4. Data-based control method via integral policy iteration: the implementation procedure

In this section, we investigate a data-based approach to solve the optimal control problem for the nominal system (2). First, we introduce the model-free online integral policy iteration algorithm with completely unknown system dynamics. Then, we display the data-based implementation of the established model-free algorithm through neural network.

The feedback control developed in (8) denotes a closed-form solution, which avoids getting the optimal control law via optimization process. However, the existence of $J^*(x)$ satisfying (9) is the necessary and sufficient condition, which is difficult to derive analytically. Hence, instead of directly dealing with (9) to obtain the solution $J^*(x)$, we can successively solve the nonlinear Lyapunov Eq. (4) and then update the control law based on (8). This idea of successive approximation is known as the policy iteration algorithm [1] (see the following Algorithm 1).

Algorithm 1. Model-based policy iteration algorithm

Step 1. Give a small positive real number ϵ. Let $i = 0$ and start with an initial admissible control law $u_0(x)$.
Step 2. Policy evaluation: Based on the control law $u_i(x)$, solve $J_i(x)$ via the nonlinear Lyapunov equation

$$0 = r(x, u_i(x)) + (\nabla J_i(x))^\mathsf{T}(f(x) + g(x)u_i(x)).$$

Step 3. Policy improvement: Update the control law by

$$u_{i+1}(x) = -\frac{1}{2}R^{-1}g^\mathsf{T}(x)\nabla J_i(x). \tag{14}$$

Step 4. If $\|u_{i+1}(x) - u_i(x)\| \le \epsilon$, stop and obtain the approximate optimal control law for the nominal system; else, set $i = i + 1$ and go to Step 2.

In [40], it was shown that on the domain Ω, the cost function $J_i(x)$ uniformly converges to $J^*(x)$ with monotonicity $J_{i+1}(x) < J_i(x)$, and the control law $u_i(x)$ is admissible and converges to $u^*(x)$ as $i \to \infty$.

To deal with the optimal control problem with completely unknown system dynamics, we next develop a data-based online integral policy iteration algorithm. We consider the following nonlinear system explored by a known bounded piecewise continuous probing signal $e(t)$

$$\dot{x}(t) = f(x(t)) + g(x(t))(u(x(t)) + e(t)). \tag{15}$$

Considering the Lyapunov function (4), the derivative of the cost function (3) with respect to time along the trajectory of the explored nominal system (15) can be calculated as

$$\dot{J}(x) = (\nabla J(x))^\mathsf{T}(f(x) + g(x)(u(x) + e))$$
$$= -r(x, u(x)) + (\nabla J(x))^\mathsf{T}g(x)e. \tag{16}$$

Lemma 3. *Under the admissible control policy $u(x)$, if the state x is generated by the system (15), solving $J(x)$ from the following integral equation*

$$J(x(t+T)) - J(x(t)) = \int_t^{t+T}(\nabla J(x))^\mathsf{T}g(x)e\mathrm{d}\tau - \int_t^{t+T}r(x, u(x))\mathrm{d}\tau \tag{17}$$

is equivalent to finding the solution of (16).

Proof. Integrating (16) from t to $t + T$ along the trajectory generated by the explored nominal system (15), we obtain the integral equation (17) where the integral is well-defined since $J(x)$ and the interval $[t, t + T]$ are finite. This means that $J(x)$ as the unique solution of (16), also satisfies (17). To complete this proof, we show that (17) has a unique solution by contradiction.

We assume that there exists another cost function $V(x)$ which satisfies (17) with the condition $V(0) = 0$. As a result, the cost function satisfies

$$\dot{V}(x) = -r(x, u(x)) + (\nabla V(x))^\mathsf{T}g(x)e.$$

Subtracting this from (16), we obtain

$$0 = \dot{J}(x) - \dot{V}(x) - (\nabla J(x) - \nabla V(x))^\mathsf{T}g(x)e$$
$$= (\nabla J(x) - \nabla V(x))^\mathsf{T}(\dot{x} - g(x)e)$$
$$= \left(\frac{\mathrm{d}(J(x) - V(x))^\mathsf{T}}{\mathrm{d}x}\right)(f(x) + g(x)u(x)), \tag{18}$$

which holds for any state $x(t)$ on the system trajectories generated by the admissible control law $u(x)$. Considering (18), we get the formula $J(x) = V(x) + c$. Note that it must hold for $x(t) = 0$, so we have $J(0) = V(0) + c$, which implies that $c = 0$. Thus, $J(x) = V(x)$, i.e., (17) has a unique solution which is equal to the solution of (16). The proof is completed. $\square$

Using the symbols $J_i(x)$ and $u_i(x)$, and considering the policy improvement (14), the formulation (17) can be rewritten as

$$J_i(x(t+T)) - J_i(x(t)) = -2\int_t^{t+T} u_{i+1}^\mathsf{T}(x)Re\,\mathrm{d}\tau - \int_t^{t+T} r(x, u_i(x))\mathrm{d}\tau. \tag{19}$$

Since the terms $f(x)$ and $g(x)$ do not appear in the integral Eq. (19), it is significant to find that the policy iteration can be conducted without using the system dynamics. Thus, we can obtain the online model-free integral policy iteration algorithm as follows (see Algorithm 2).

Algorithm 2. Model-free integral policy iteration algorithm

Step 1. Give a small positive real number ϵ. Let $i = 0$ and start with an initial admissible control law $u_0(x)$.
Step 2. Policy evaluation and improvement:
      Based on the control policy $u_i(x)$, solve $J_i(x)$ and $u_{i+1}(x)$ from the integral Eq. (19).
Step 3. If $\|u_{i+1}(x) - u_i(x)\| \le \epsilon$, stop and obtain the approximate optimal control law for the nominal system; else, set $i = i + 1$ and go to Step 2.

The convergence of the model-free integral policy iteration algorithm is presented in the following main theorem.

Theorem 1. *Give an initial admissible control law $u_0(x)$ for the nominal system (2). Using the integral policy iteration algorithm established in Algorithm 2, the cost function and the control law converge to the optimal ones as $i \to \infty$, i.e., $J_i(x) \to J^*(x)$ and $u_i(x) \to u^*(x)$.*

Proof. If the initial control law $u_0(x)$ is admissible, according to (14) and (16), all the subsequent control laws will be admissible [43] and the iteration process will converge to the solution of the HJB equation as well. Considering (19) and Lemma 3, we can conclude that the developed integral policy iteration algorithm will converge to the solution of the optimal control of (2) without using the knowledge of system dynamics. The proof is completed. $\square$

Next, we discuss the data-based implementation method of the established model-free policy iteration algorithm using the neural network technique. A critic neural network and an actor neural network are introduced to approximate the cost function and the control law of the nominal system, respectively. We assume that for the nominal system, $J_i(x)$ and $u_{i+1}(x)$ are represented on a compact set Ω by single-layer neural networks as

$$J_i(x) = \sum_{j=1}^{N_c} \omega_{ij}\phi_j(x) + \varepsilon_c(x),$$

$$u_{i+1}(x) = \sum_{j=1}^{N_a} v_{ij}\psi_j(x) + \varepsilon_a(x),$$

where $\omega_{ij} \in \mathbb{R}$ and $v_{ij} \in \mathbb{R}^m$ are unknown bounded ideal weight parameters, $\phi_j(x) \in \mathbb{R}$ and $\psi_j(x) \in \mathbb{R}$, $\{\phi_j\}_{j=1}^{N_c}$ and $\{\psi_j\}_{j=1}^{N_a}$ are the sequences of real-valued activation functions that are linearly independent and complete, and $\varepsilon_c(x) \in \mathbb{R}$ and $\varepsilon_a(x) \in \mathbb{R}^m$ are the bounded neural network approximation errors. Since the ideal weights are unknown, the outputs of the critic network and the actor network are denoted as

$$\hat{J}_i(x) = \sum_{j=1}^{N_c} \hat{\omega}_{ij}\phi_j(x) = \hat{\omega}_i^\mathsf{T}\phi(x), \tag{20}$$

$$\hat{u}_{i+1}(x) = \sum_{j=1}^{N_a} \hat{v}_{ij}\psi_j(x) = \hat{v}_i^\mathsf{T}\psi(x), \tag{21}$$

where $\hat{\omega}_i$ and $\hat{v}_i$ are the current estimated weights, and

$$\phi(x) = [\phi_1(x), \phi_2(x), \dots, \phi_{N_c}(x)]^\mathsf{T} \in \mathbb{R}^{N_c},$$
$$\psi(x) = [\psi_1(x), \psi_2(x), \dots, \psi_{N_a}(x)]^\mathsf{T} \in \mathbb{R}^{N_a},$$
$$\hat{\omega}_i = [\hat{\omega}_{i1}, \hat{\omega}_{i2}, \dots, \hat{\omega}_{iN_c}]^\mathsf{T} \in \mathbb{R}^{N_c},$$
$$\hat{v}_i = [\hat{v}_{i1}, \hat{v}_{i2}, \dots, \hat{v}_{iN_a}]^\mathsf{T} \in \mathbb{R}^{N_a \times m}.$$

Define $\mathrm{col}\{\hat{v}_i^\mathsf{T}\} = [\hat{v}_{i1}^\mathsf{T}, \hat{v}_{i2}^\mathsf{T}, \dots, \hat{v}_{iN_a}^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{mN_a}$, then

$$\hat{u}_{i+1}^\mathsf{T}(x)Re = \left(\hat{v}_i^\mathsf{T}\psi(x)\right)^\mathsf{T}Re$$

$$= (\psi(x) \otimes (Re))^{\mathsf{T}} \mathrm{col}\{\hat{v}_i^{\mathsf{T}}\},$$

where $\otimes$ represents the Kronecker product. Substituting (20) and (21) into the integral Eq. (19), we obtain the following expression

$$\lambda_k^{\mathsf{T}} \begin{bmatrix} \hat{\omega}_i \\ \mathrm{col}\{\hat{v}_i^{\mathsf{T}}\} \end{bmatrix} = \theta_k \tag{22}$$

with

$$\theta_k = \int_{t+(k-1)T}^{t+kT} r(x, \hat{u}_i(x)) \mathrm{d}\tau,$$

$$\lambda_k = \left[\left(\phi(x(t + (k-1)T)) - \phi(x(t + kT)) \right)^{\mathsf{T}}, -2 \int_{t+(k-1)T}^{t+kT} (\psi(x) \otimes (Re))^{\mathsf{T}} \mathrm{d}\tau \right]^{\mathsf{T}},$$

where the measurement time is considered from $t + (k-1)T$ to $t + kT$. Note that (22) is only a 1-dimensional equation, we cannot ensure the uniqueness of the solution. Inspired by [18], we introduce the least squares method to solve the parameter vector over the compact set Ω. For any positive integer K, we denote $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_K]$ and $\Theta = [\theta_1, \theta_2, \dots, \theta_K]^{\mathsf{T}}$. Then, we have the following K-dimensional equation

$$\Lambda^{\mathsf{T}} \begin{bmatrix} \hat{\omega}_i \\ \mathrm{col}\{\hat{v}_i^{\mathsf{T}}\} \end{bmatrix} = \Theta.$$

If Λ^{T} has full column rank, the parameters can be solved by the following operation

$$\begin{bmatrix} \hat{\omega}_i \\ \mathrm{col}\{\hat{v}_i^{\mathsf{T}}\} \end{bmatrix} = (\Lambda \Lambda^{\mathsf{T}})^{-1} \Lambda \Theta. \tag{23}$$

Here, the number of collected points K should be set satifying $K \geq \mathrm{rank}(\Lambda) = N_c + mN_a$ to guarantee the existance of $(\Lambda \Lambda^{\mathsf{T}})^{-1}$. The least squares problem in (23) can be solved in real time by collecting enough data points generated by the explored nominal system (15).

Remark 1. Based on the integral policy iteration algorithm and neural network technique, we solve the optimal control problem iteratively and hence the approximate optimal control law $\hat{u}^*(x)$ can be obtained. According to (10), we can derive the robust control law $\bar{u}(x) = \pi \hat{u}^*(x)$. Therefore, the closed-form expression of the robust optimal controller of the uncertain nonlinear system is available. This completes the data-based robust optimal control design of continuous-time affine nonlinear system with matched uncertainties in theory and implementation.

5. Simulation studies

In this section, two simulation examples with application backgrounds are presented to illustrate the effectiveness of the established robust optimal control scheme.

Example 1. Consider an input-affine continuous-time nonlinear system described as [47]

$$\dot{x} = \begin{bmatrix} -0.5x_1 + x_2(1 + 0.5x_2^2) \\ -0.8(x_1 + x_2) + 0.5x_2(1 - 0.3x_2^2) \end{bmatrix} + \begin{bmatrix} 0 \\ -0.6 \end{bmatrix} (\bar{u}(x) + \bar{d}(x)), \tag{24}$$

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 matched uncertainty of system (24) is $\bar{d}(x) = \delta_1 x_2 \cos(\delta_2 x_1 + \delta_3 x_2)$, 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 $R = I$ (I is an identity matrix with appropriate dimension) so $d(x) = \bar{d}(x)$ and choose $d_M(x) = \|x\|$ as the bound of the uncertain term $d(x)$.

Based on the theoretical results of this paper, we should solve the optimal control problem of the nominal system

$$\dot{x} = \begin{bmatrix} -0.5x_1 + x_2(1 + 0.5x_2^2) \\ -0.8(x_1 + x_2) + 0.5x_2(1 - 0.3x_2^2) \end{bmatrix} + \begin{bmatrix} 0 \\ -0.6 \end{bmatrix} u \tag{25}$$

with a newly defined cost function of the form

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

In the following, two case studies are provided with comparison remarks between the data-based integral policy iteration algorithm and the traditional model-based policy iteration algorithm.

Case 1: Assume that the exact knowledge of dynamical system (25) is fully unknown. We adopt the data-based integral policy iteration algorithm to tackle the optimal control problem. In this example, the activation functions are chosen as

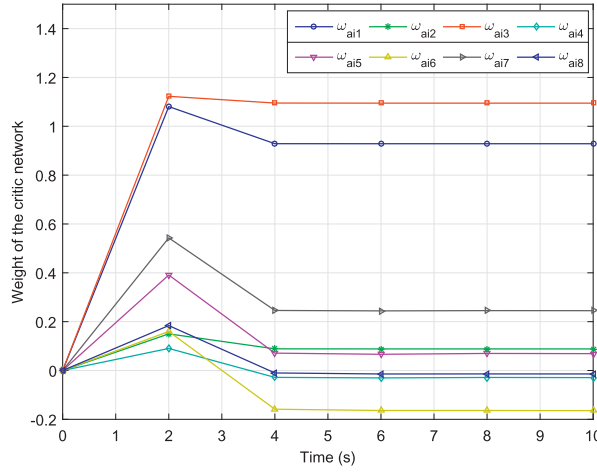$$\phi(x) = [x_1^2, x_1 x_2, x_2^2, x_1^4, x_1^3 x_2, x_1^2 x_2^2, x_1 x_2^3, x_2^4]^{\mathsf{T}},$$

Fig. 1. Evolution of the weights of the critic network (ω_{aij} represents $\hat{\omega}_{ij}$, $j = 1, 2, \ldots, 8$).
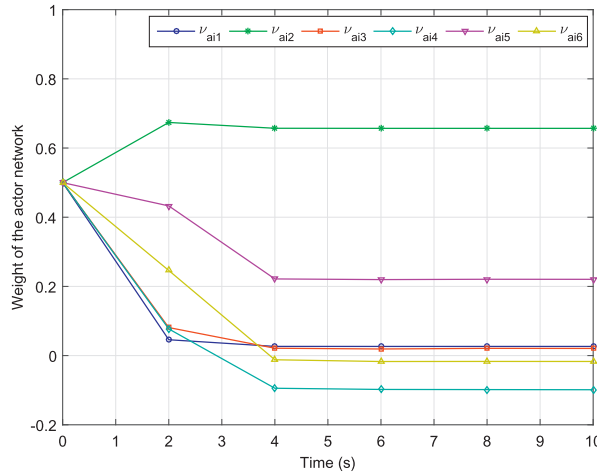

Fig. 2. Evolution of the weights of the actor network (ν_{aij} represents $\hat{v}_{ij}$, $j = 1, 2, \ldots, 6$).

$$\psi(x) = [x_1, x_2, x_1^3, x_1^2 x_2, x_1 x_2^2, x_2^3]^\mathsf{T}.$$

Observing the fact that $N_c = 8$ and $N_a = 6$, we can conduct the iteration algorithm with $K = 20$. During the simulation process, the initial weights of the critic network and the actor network are chosen as

$$\hat{\omega}_0 = [0, 0, 0, 0, 0, 0, 0, 0]^\mathsf{T},$$
$$\hat{v}_0 = [0.5, 0.5, 0.5, 0.5, 0.5, 0.5]^\mathsf{T}.$$

Let the initial state be $x_0 = [0.5, -0.5]^\mathsf{T}$. The time interval $T = 0.1$s and the probing signal $e(t) = 0.1 \sin(2\pi t) + 0.1 \cos(2\pi t)$ are chosen in the learning process. The least squares problem is solved after 20 samples are acquired. Hence, the weights of the neural networks are updated every 2s. During simulation, Figs. 1 and 2 illustrate the evolutions of the weights of the critic network and the actor network, respectively. It is clear that the weights are convergent after five iterations. At $t = 10$s, we have

$$\hat{\omega}_5 = [0.9286, 0.0884, 1.0948, -0.0291, 0.0690, -0.1644, 0.2449, -0.0141]^\mathsf{T},$$
$$\hat{v}_5 = [0.0265, 0.6569, 0.0207, -0.0987, 0.2205, -0.0169]^\mathsf{T}.$$

At last, a set of scalar parameters: $\pi = 3$, $\delta_1 = 0.8$, $\delta_2 = -5$, and $\delta_3 = 3$, is selected in order to evaluate the performance of robust controller. Then, the state response of system (24) combined with the robust controller during the first 20s is given in Fig. 3. According to the conclusion of Lemma 2, it also holds the property of optimality with cost function (12). These results verify the effectiveness of the data-based robust optimal control strategy given in this paper.

Case 2: To demonstrate the effectiveness of the developed algorithm without system dynamics, we use Algorithm 1 to solve the optimal control problem of (25) with the knowledge of system dynamics. Using the implementation method de-
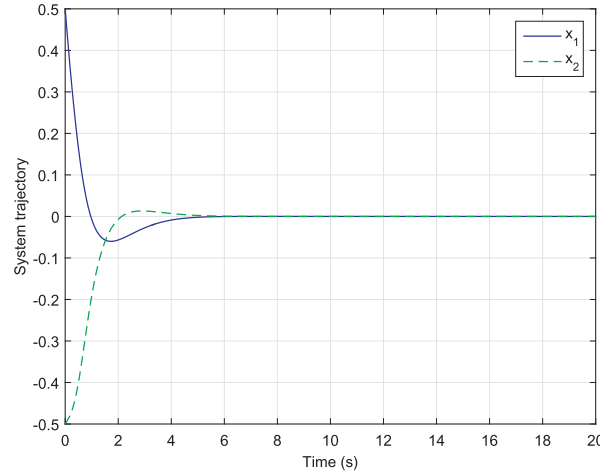
Fig. 3. The state trajectory of the original nonlinear system when setting $\pi = 3$, $\delta_1 = 0.8$, $\delta_2 = -5$, and $\delta_3 = 3$ (Case 1).
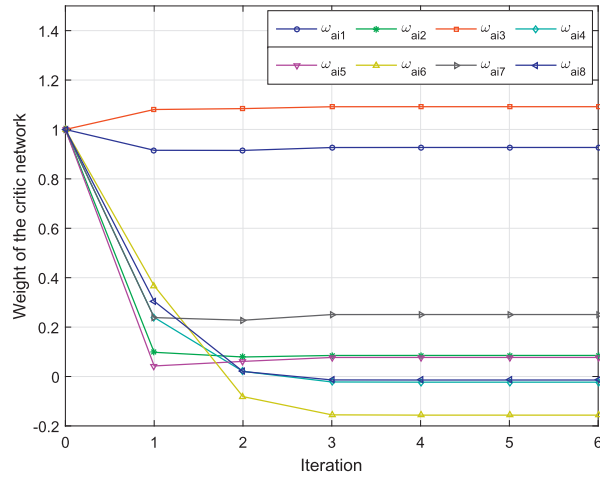

Fig. 4. Evolution of the weights of the critic network (ω_{aij} represents $\hat{\omega}_{ij}$, $j = 1, 2, \ldots, 8$).

scribed in [1], the cost function can be approximated by the activation function $\phi(x) = [x_1^2, x_1 x_2, x_2^2, x_1^4, x_1^3 x_2, x_1^2 x_2^2, x_1 x_2^3, x_2^4]^T$. Using the information of system dynamics, the weights of the critic network can be updated iteratively by solving the generalized-HJB equation. In the simulation process, Fig. 4 illustrates the evolution of the weights of the critic network. It is clear to find that the weights are convergent after six iterations, that is $\hat{\omega}_6 = [0.9271, 0.0853, 1.0921, -0.0229, 0.0772, -0.1563, 0.2509, -0.0142]^T$. Setting the parameters the same as in Case 1, the performance of the robust control strategy is displayed in Fig. 5, which is difficult to observe difference compared with Fig. 3.

Remark 2. In this example, we use the model-based policy iteration algorithm and the integral policy iteration algorithm, which is regarded as a model-free algorithm, to solve the optimal control problem of the nominal system (25), respectively. The simulation results testify the equivalence of the two algorithms. Compared with Algorithm 1, the integral policy iteration algorithm can be implemented with completely unknown system dynamics in an online manner. Moreover, using the relationship between the robust optimal control of the uncertain system and optimal control of its nominal system, a model-free robust optimal control approach can be developed. In this sense, we establish the data-based robust optimal control strategy of continuous-time affine nonlinear systems under uncertain environment.

Example 2. In this example, we consider the classical multi-machine power system with governor controllers [9,13,14,17]

$$\dot{\delta}_i(t) = \omega_i(t),$$

$$\dot{\omega}_i(t) = -\frac{D_i}{2H_i}\omega_i(t) + \frac{\omega_0}{2H_i}(P_{mi}(t) - P_{ei}(t)),$$

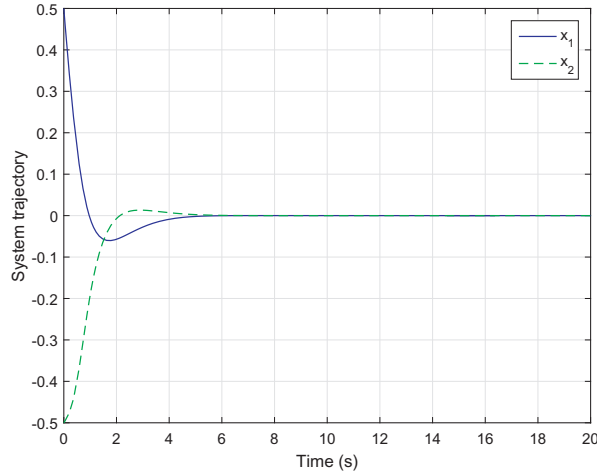$$\dot{P}_{mi}(t) = \frac{1}{T_i}(-P_{mi}(t) + u_{gi}(t)),$$

Fig. 5. The state trajectory of the original nonlinear system when setting $\pi = 3$, $\delta_1 = 0.8$, $\delta_2 = -5$, and $\delta_3 = 3$ (Case 2).

Table 1
Summary of parameters used in the multi-machine power system.

$\delta_i(t)$	Angle of the ith generator	ω_0	Steady state frequency
$\omega_i(t)$	Relative rotor speed	E'_{qi}	Transient electromotive force constant
$P_{mi}(t)$	Mechanical power	B_{ij}	Imaginary part of the admittance matrix
$P_{ei}(t)$	Electrical power	G_{ij}	Real part of the admittance matrix
D_i	Damping constant	$u_{gi}(t)$	Speed governor control signal for the ith generator
H_i	Inertia constant	$\delta_{ij}(t)$	The angular difference between the ith and jth generators
T_i	Governor time constant	N	Number of the generators

$$P_{ei}(t) = E'_{qi} \sum_{j=1}^{N} E'_{qj}(B_{ij} \sin \delta_{ij}(t) + G_{ij} \cos \delta_{ij}(t)),$$

where $1 \leq i, j \leq N$. Table 1 shows the summary of parameters used in the multi-machine power system. The values of these parameters are set the same as that in [13].

We consider the third generator of the power system in this numerical simulation. Similarly, as in [13], we rewrite the third generator as the following form

$$\dot{x} = \begin{bmatrix} x_2 \\ -\dfrac{D}{2H}x_2 + \dfrac{\omega_0}{2H}x_3 \\ -\dfrac{1}{T}x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \dfrac{1}{T} \end{bmatrix} (\bar{u}(x) + \bar{d}(x)), \tag{26}$$

where the state vector x is denoted as $x = [x_1, x_2, x_3]^\mathsf{T} \in \mathbb{R}^3$. Here, the components of state x are defined as $x_1 = \Delta\delta(t) = \delta(t) - \delta_0$, $x_2 = \Delta\omega(t) = \omega(t) - \omega_0$, $x_3 = \Delta P_m(t) = P_m(t) - P_e(t)$, and the system control is defined as $\bar{u}(x(t)) = u_g(t) - P_e(t)$. The term $\bar{d}(t) = -E'_q(\delta_1 \cos(x_1 - \delta_3) - \delta_2 \sin(x_1 - \delta_3))(x_2 - \delta_4)$ reflects the uncertainty caused by the other generators of the multi-machine power system, with unknown parameters $\delta_1 \in [0, 0.9]$, $\delta_2 \in [-0.45, 0.45]$, $\delta_3 \in [-60, 60]$, and $\delta_4 \in [-2, 2]$ included. We set $R = I$ and select $d_M(x) = 10\sqrt{10}\|x\|$ as the bound of the uncertain function $d(x)$. Using the obtained theoretical results, the cost function can be represented as

$$J(x_0) = \int_0^{\infty} \left\{ 1000\|x(\tau)\|^2 + u^\mathsf{T}(x(\tau))Ru(x(\tau)) \right\} \mathrm{d}\tau.$$

Assume that the exact knowledge of the dynamics (26) is fully unknown. We adopt the data-based integral policy iteration algorithm to tackle the optimal control problem of the nominal system (which is omitted here). In this simulation study, the activation functions are chosen as

$$\phi(x) = [x_1^2, x_1 x_2, x_1 x_3, x_2^2, x_2 x_3, x_3^2]^\mathsf{T},$$
$$\psi(x) = [x_1, x_2, x_3]^\mathsf{T}.$$

Clearly, we find that $N_c = 6$ and $N_a = 3$ and then we can conduct the simulation with $K = 10$. During the simulation process, the initial weights of the two networks are chosen as
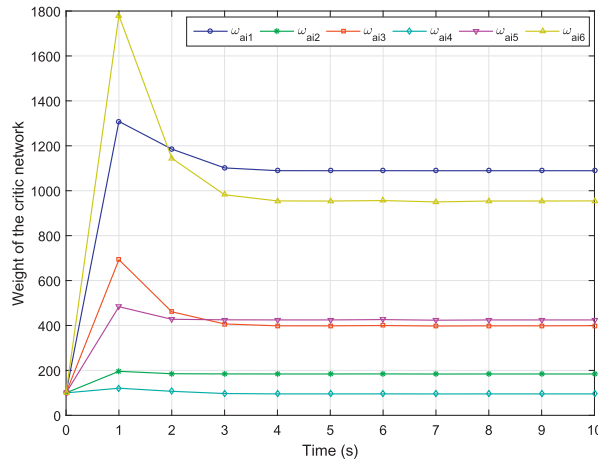
Fig. 6. Evolution of the weights of the critic network (ω_{aij} represents $\hat{\omega}_{ij}$, $j = 1, 2, \ldots, 6$).
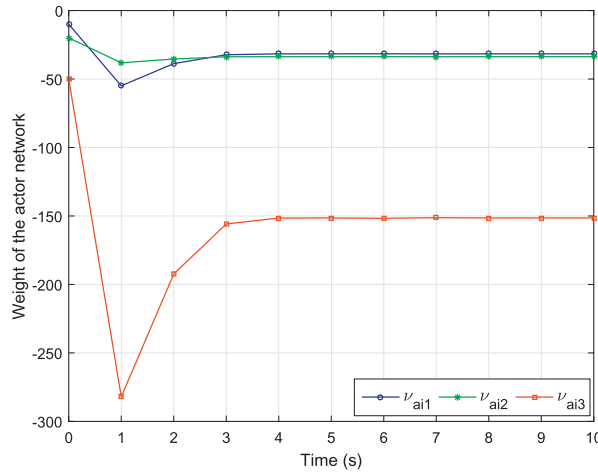

Fig. 7. Evolution of the weights of the actor network (ν_{aij} represents $\hat{\nu}_{ij}$, $j = 1, 2, 3$).

$$\hat{\omega}_0 = [100, 100, 100, 100, 100, 100]^{\mathsf{T}},$$

$$\hat{\nu}_0 = -[10, 20, 50]^{\mathsf{T}}.$$

Let the initial state be $x_0 = [1, 1, 1]^{\mathsf{T}}$. The time interval $T = 0.1$s and the probing signal $e(t) = 0.01 \sin(2\pi t) + 0.01 \cos(2\pi t)$ are also introduced to the learning process. The least squares problem is solved after 10 samples are acquired. Thus, the weights of the neural networks are updated every 1s. In this simulation, Figs. 6 and 7 illustrate the evolutions of the weights of the critic network and the actor network, respectively. We can observe that the weights of two networks are convergent after 10 iterations. At $t = 10$s, we have

$$\hat{\omega}_{10} = [1089.0219, 184.1026, 399.1418, 95.3567, 424.7228, 954.6765]^{\mathsf{T}},$$

$$\hat{\nu}_{10} = [-31.61284, -33.7111, -151.4410]^{\mathsf{T}}.$$

Finally, the scalar parameters $\pi = 3$, $\delta_1 = 0.5$, $\delta_2 = 0.3$, $\delta_3 = 50$, and $\delta_4 = 2$, are chosen to display the performance of the robust control method. When employing the derived robust control strategy, the state response of system (26) during the first 20s is displayed in Fig. 8, which verifies the availability of the present robust optimal control scheme.

6. Conclusion

A novel integral policy iteration approach for robust optimal control of input-affine nonlinear systems with matched uncertainties is developed in this paper, based on the novel idea of data-based ADP. It is proved that the robust controller of the original uncertain system can attain optimality with a newly specified cost function. Then, the problem of designing the robust optimal control is transformed into an optimal control problem. Using model-free integral policy iteration algorithm,
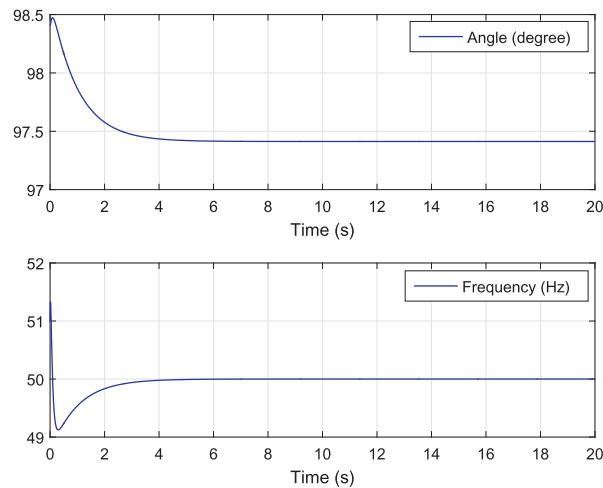
Fig. 8. The angle and frequency trajectories of the controlled generator when setting $\pi = 3$, $\delta_1 = 0.5$, $\delta_2 = 0.3$, $\delta_3 = 50$, and $\delta_4 = 2$.

the optimal controller of the nominal system can be developed without relying on the knowledge of system dynamics. The obtained results are a natural extension of the traditional ADP-based optimal control design to robust optimal control of nonlinear systems under uncertain environment. The two simulation examples verify the good control performance.

References

[1] R.W. Beard, G.N. Saridis, J.T. Wen, Galerkin approximations of the generalized Hamilton–Jacobi–Bellman equation, Automatica 33 (1997) 33–1997.
[2] R.E. Bellman, Dynamic Programming, Princeton University Press, Princeton, NJ, 1957.
[3] 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) 49–2013.
[4] T. Bian, Y. Jiang, Z.P. Jiang, Adaptive dynamic programming and optimal control of nonlinear nonaffine systems, Automatica 50 (2014) 50–2014.
[5] L. Cheng, Z.G. Hou, M. Tan, Y. Lin, W. Zhang, Neural-network-based adaptive leader-following control for multi-agent systems with uncertainties, IEEE Trans. Neural Netw. 21 (2010) 21–2010.
[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) 23–2012.
[7] B. Gu, V.S. Sheng, K.Y. Tay, W. Romano, S. Li, Incremental support vector learning for ordinal regression, IEEE Trans. Neural Netw. Learn. Syst. 26 (2015) 26–2015.
[8] B. Gu, V.S. Sheng, Z. Wang, D. Ho, S. Osman, S. Li, Incremental learning for ν-support vector regression, Neural Netw. 67 (2015) 140–150.
[9] G. Guo, Y. Wang, D.J. Hill, Nonlinear output stabilization control for multimachine power systems, IEEE Trans. Circuits Syst.–I 47 (2000) 47–2000.
[10] W. He, S. Zhang, S.S. Ge, Adaptive control of a flexible crane system with the boundary output constraint, IEEE Trans. Ind. Electron. 61 (2014) 61–2014.
[11] 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.
[12] Z.S. Hou, Z. Wang, From model-based control to data-driven control: Survey, classification and perspective, Inf. Sci. 235 (2013) 3–35.
[13] Y. Jiang, Z.P. Jiang, Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems, IEEE Trans. Circuits Syst.–II 59 (2012) 59–2012.
[14] Y. Jiang, Z.P. Jiang, Robust Adaptive Dynamic Programming for Large-scale Systems with an Application to Multimachine Power Systems, New York University, Brooklyn, NY, 2012 Technical report. [Online] Available: http://files.nyu.edu/yj348/public/papers/2012/tcas12tr.pdf.
[15] Y. Jiang, Z.P. Jiang, Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics, Automatica 48 (2012) 48–2012.
[16] Y. Jiang, Z.P. Jiang, Robust adaptive dynamic programming and feedback stabilization of nonlinear systems, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 25–2014.
[17] P. Kundur, N.J. Balu, M.G. Lauby, Power System Stability and Control, McGraw-Hill, New York, NY, 1994.
[18] J.Y. Lee, J.B. Park, Y.H. Choi, Integral q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems, Automatica 48 (2012) 2850–2859.
[19] J.Y. Lee, J.B. Park, Y.H. Choi, Integral reinforcement learning for continuous-time input-affine nonlinear systems with simultaneous invariant explorations, IEEE Trans. Neural Netw. Learn. Syst. 26 (2015) 916–932.
[20] F.L. Lewis, D. Liu, Reinforcement Learning and Approximate Dynamic Programming for Feedback Control, Wiley, Hoboken, NJ, 2013.
[21] F.L. Lewis, D. Vrabie, Reinforcement learning and adaptive dynamic programming for feedback control, IEEE Circuits Syst. Mag. 9 (2009) 9–2009.
[22] F.L. Lewis, D. Vrabie, V. Syrmos, Optimal Control, Wiley, Hoboken, NJ, 2012.
[23] 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) 3–2012.
[24] F. Lin, R.D. Brand, J. Sun, Robust control of nonlinear systems: Compensating for uncertainty, Int. J. Control 56 (1992) 1453–1459.
[25] D. Liu, C. Li, H. Li, D. Wang, H. Ma, Neural-network-based decentralized control of continuous-time nonlinear interconnected systems with unknown dynamics, Neurocomputing 165 (2015) 165–2015.
[26] 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) 25–2014.
[27] D. Liu, D. Wang, X. Yang, An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs, Inf. Sci. 220 (2013) 331–342.
[28] 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.

[29] B. Luo, H.N. Wu, T. Huang, D. Liu, Data-based approximate policy iteration for affine nonlinear continuous-time optimal control design, Automatica 50 (2014) 3281–3290.

[30] 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) 52–2013.

[31] D.S. Naidu, Optimal Control Systems, CRC Press, Boca Raton, FL, 2003.

[32] C. Mu, Z. Ni, C. Sun, H. He, Air-breathing hypersonic vehicle tracking control based on adaptive dynamic programming, IEEE Trans. Neural Netw. Learn. Syst. (2016), doi:10.1109/TNNLS.2016.2516948. (In press).

[33] C. Mu, C. Sun, A new finite time convergence condition for super-twisting observer based on Lyapunov analysis, Asian J. Control 17 (2015) 17–2015.

[34] C. Mu, C. Sun, W. Xu, Fast sliding mode control on air-breathing hypersonic vehicles with transient response analysis, proceedings of the institution of mechanical engineers, Part I: J. Syst. Control Eng. 230 (2016a) 230–2016.

[35] C. Mu, W. Xu, C. Sun, On switching manifold design for terminal sliding mode control, J. Frankl. Inst. 353 (2016b) 1553–1572.

[36] 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) 24–2013.

[37] Z. Ni, H. He, D. Zhao, X. Xu, D.V. Prokhorov, GrDHP: a general utility function representation for dual heuristic dynamic programming, IEEE Trans. Neural Netw. Learn. Syst. 26 (2015) 26–2015.

[38] A.E. Ougli, B. Tidhaf, Optimal type-2 fuzzy adaptive control for a class of uncertain nonlinear systems using an LMI approach, Int. J. Innov. Comput. Inf. Control 11 (2015) 11–2015.

[39] D.V. Prokhorov, D.C. Wunsch, Adaptive critic designs, IEEE Trans. Neural Netw. 8 (1997) 8–1997.

[40] G.N. Saridis, C.G. Lee, An approximation theory of optimal control for trainable manipulators, IEEE Trans. Syst. Man Cybern. – Part B 9 (1979) 9–1979.

[41] Y. Tang, H. He, Z. Ni, D. Zhao, X. Xu, Fuzzy-based goal representation adaptive dynamic programming, IEEE Trans. Fuzzy Syst. (2016), doi:10.1109/TFUZZ.2015.2505327. (in press).

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

[43] D. Vrabie, F.L. Lewis, Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems, Neural Netw. 22 (2009) 22–2009.

[44] D. Wang, D. Liu, Neuro-optimal control for a class of unknown nonlinear dynamic systems using SN-DHP technique, Neurocomputing 121 (2013) 218–225.

[45] 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.

[46] D. Wang, D. Liu, H. Li, B. Luo, H. Ma, An approximate optimal control approach for robust stabilization of a class of discrete-time nonlinear systems with uncertainties, IEEE Trans. Syst. Man Cybern. 46 (2016) 713–717.

[47] D. Wang, D. Liu, H. Li, H. Ma, Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming, Inf. Sci. 282 (2014) 167–179.

[48] 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) 48–2012.

[49] D. Wang, D. Liu, Q. Zhang, D. Zhao, Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics, IEEE Trans. Syst. Man Cybern. (2016), doi:10.1109/TSMC.2015.2492941. (In press).

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

[51] L. Wu, R. Yang, P. Shi, X. Su, Stability analysis and stabilization of 2-d switched systems under arbitrary and restricted switchings, Automatica 59 (2015) 59–2015.

[52] T.S. Wu, M. Karkoub, H.S. Chen, W.S. Yu, M.G. Her, Robust tracking observer-based adaptive fuzzy control design for uncertain nonlinear MIMO systems with time delayed states, Inf. Sci. 290 (2015) 86–105.

[53] H. Xing, X. Zhong, J. Li, Linear extended state observer based back-stepping control for uncertain SISO nonlinear systems, Int. J. Innov. Comput. Inf. Control 11 (2015) 11–2015.

[54] B. Xu, C. Yang, Z. Shi, Reinforcement learning output feedback NN control using deterministic learning technique, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 25–2014.

[55] 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.

[56] X. Xu, L. Zuo, Z. Huang, Reinforcement learning algorithms with function approximation: recent advances and applications, Inf. Sci. 261 (2014) 261–2014.

[57] R. Yang, G.P. Liu, P. Shi, C. Thomas, M.V. Basin, Predictive output feedback control for networked control systems, IEEE Trans. Ind. Electron. 61 (2014) 61–2014.

[58] X. Yang, D. Liu, H. Ma, Y. Xu, Online approximate solution of HJI equation for unknown constrained-input nonlinear continuous-time systems, Inf. Sci. 328 (2016) 328–2016.

[59] H. Zhang, D. Liu, Y. Luo, D. Wang, Adaptive Dynamic Programming for Control: Algorithms and Stability, Springer, London, UK, 2013.

[60] H. Zhang, C. Qin, Y. Luo, Neural-network-based constrained optimal control scheme for discrete-time switched nonlinear system using dual heuristic programming, IEEE Trans. Autom. Sci. Eng. 11 (2014) 11–2014.

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

[62] X. Zhong, Z. Ni, H. He, A theoretical foundation of goal representation heuristic dynamic programming, IEEE Trans. Neural Netw. Learn. Syst. (2016), doi:10.1109/TNNLS.2015.2490698. (in press).