Neurocomputing 165 (2015) 90–98

Contents lists available at ScienceDirect

Neurocomputing

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

Neural-network-based decentralized control of continuous-time nonlinear interconnected systems with unknown dynamics[☆]

Derong Liu[*], Chao Li, Hongliang Li, Ding Wang, Hongwen Ma

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

ARTICLE INFO

Article history:
Received 28 February 2014
Received in revised form
24 June 2014
Accepted 5 July 2014
Available online 17 April 2015

Keywords:
Adaptive dynamic programming
Decentralized control
Optimal control
Policy iteration
Neural networks

ABSTRACT

In this paper, we establish a neural-network-based decentralized control law to stabilize a class of continuous-time nonlinear interconnected large-scale systems using an online model-free integral policy iteration (PI) algorithm. The model-free PI approach can solve the decentralized control problem for the interconnected system which has unknown dynamics. The stabilizing decentralized control law is derived based on the optimal control policies of the isolated subsystems. The online model-free integral PI algorithm is developed to solve the optimal control problems for the isolated subsystems with unknown system dynamics. We use the actor-critic technique based on the neural network and the least squares implementation method to obtain the optimal control policies. Two simulation examples are given to verify the applicability of the decentralized control law.

1. Introduction

Decentralized control method using local information of each subsystem is an efficient and effective way in the control of interconnected systems. This overcomes the limitations of the traditional control method that requires sufficient information between subsystems. Unlike a centralized controller, a decentralized controller can be designed independently for local subsystems and make full use of the local available signals for feedback. Therefore, the decentralized controllers have simpler architecture, and are more practical than the traditional centralized controllers. Various decentralized controllers have been established for large-scale interconnected systems in the presence of uncertainties and information structure constraints [1–7]. Generally speaking, a decentralized control law is composed of some noninteracting local controllers corresponding to the isolated subsystems, not the overall system. In many situations, the design of the isolated subsystems is very important. In [8], the decentralized controller was derived for the large-scale system using the optimal control policies of the isolated subsystems. Therefore, the optimal control method can be applied to facilitate the design process of the decentralized control law.

The optimal control problem of nonlinear systems has been widely studied in the past few decades. The optimal control policy can be obtained by solving Hamilton–Jacobi–Bellman (HJB) equation which is a partial differential equation. Because of the curse of dimensionality [9], this is a difficult task even in the case of completely known dynamics. Among the methods of solving the HJB equation, adaptive dynamic programming (ADP) has received increasing attention owing to its learning and optimal capacities [10–20]. Reinforcement learning (RL) is another computational method and it can interactively find an optimal policy [21–24]. Al-Tamimi et al. [25] proposed a greedy iterative ADP to solve the optimal control problem for nonlinear discrete-time systems. Park et al. [26] used multilayer neural networks (NNs) to design a finite-horizon optimal tracking neuro-controller for discrete-time nonlinear systems with quadratic cost function. Abu-Khalaf and Lewis [27] established an offline optimal control law for nonlinear systems with saturating actuators. Vamvoudakis and Lewis [28] derived a synchronous policy iteration (PI) algorithm to learn online continuous-time optimal control with known dynamics. Vrabie and Lewis [29] derived an integral RL method to obtain direct adaptive optimal control for nonlinear input-affine continuous-time systems with partially unknown dynamics. Jiang and Jiang [30] presented a novel PI approach for continuous-time linear systems with complete unknown dynamics. Liu et al. [31] extended the PI algorithm to nonlinear optimal control problem with unknown dynamics and discounted cost function. Lee et al. [32,33] presented an integral Q-learning algorithm for continuous-time systems without the exact knowledge of the system dynamics.

It is difficult to obtain the exact knowledge of the system dynamics for large-scale systems, such as transportation systems and power systems. The novelty of this paper is that we relax the assumptions of exact knowledge of the system dynamics required in the optimal

[☆] This work was supported in part by the National Natural Science Foundation of China under Grants 61034002, 61233001, 61273140, 61304086, and 61374105.

[*] Corresponding author.
E-mail addresses: derong.liu@ia.ac.cn (D. Liu), lichao2012@ia.ac.cn (C. Li), hongliang.li@ia.ac.cn (H. Li), ding.wang@ia.ac.cn (D. Wang), mahongwen2012@ia.ac.cn (H. Ma).

http://dx.doi.org/10.1016/j.neucom.2014.07.082
0925-2312/© 2015 Elsevier B.V. All rights reserved.

controller design presented in [8]. In this paper, we use an online model-free integral PI to solve the decentralized control of a class of continuous-time nonlinear interconnected systems. We establish the stabilizing decentralized control law by adding feedback gains to the local optimal polices of the isolated subsystems. The optimal control problems for the isolated subsystems with unknown dynamics are related to develop the decentralized control law. To implement this algorithm, a critic NN and an action NN are used to approximate the value function and control policy of the isolated subsystem, respectively. The effectiveness of the decentralized control law established in this paper is demonstrated by two simulation examples.

The rest of this paper is organized as follows. In Section 2, we present the decentralized control problem of the continuous-time nonlinear large-scale interconnected system. Section 3 presents the decentralized stabilization control law for the continuous-time interconnected system by adding appropriate feedback gains to the local optimal polices of the isolated subsystems. In Section 4, we derive a model-free PI algorithm using NN implementation to obtain the decentralized control law. Two simulation examples are provided in Section 5 to illustrate the effectiveness of the derived decentralized control law. In Section 6, we conclude the paper with a few remarks.

2. Problem formulation

We consider a continuous-time nonlinear large-scale system Σ composed of N interconnected subsystems described by

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

where $x_i(t) \in \mathbb{R}^{n_i}$ is the state, $u_i(x_i(t)) \in \mathbb{R}^{m_i}$ is the control input vector of the ith subsystem. The overall state of the large-scale system Σ is denoted by $x = [x_1^{\mathsf{T}}\ x_2^{\mathsf{T}}\ \ldots\ x_N^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^n$, where $n = \sum_{i=1}^N n_i$. The local states are represented by $x_1, x_2, \ldots, x_N$, whereas $u_1(x_1)$, $u_2(x_2), \ldots, u_N(x_N)$ are local controls. For the ith subsystem, f_i is a continuous nonlinear internal dynamics function from $\mathbb{R}^{n_i}$ into $\mathbb{R}^{n_i}$ such that $f_i(0) = 0$. $g_i(x_i)$ is the input gain function from $\mathbb{R}^{n_i}$ into $\mathbb{R}^{n_i \times m_i}$. $Z_i(x(t))$ is the interconnected term for the ith subsystem.

The ith isolated subsystem Σ_i is given by

$$\Sigma_i : \dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))u_i(x_i(t)). \tag{2}$$

For the ith isolated subsystem, we assume that the subsystem is controllable, $f_i + g_i u_i$ is Lipschitz continuous on a set Ω_i in $\mathbb{R}^{n_i}$, and there exists a continuous control policy that asymptotically stabilizes the subsystem. Additionally, we let the following assumptions hold through this paper.

Assumption 1. The state vector $x_i = 0$ is the equilibrium of the ith subsystem, $i = 1, 2, \ldots, N$.

Assumption 2. The functions $f_i(\cdot)$ and $g_i(\cdot)$ are differentiable in their arguments with $f_i(0) = 0$, where $i = 1, 2, \ldots, N$.

Assumption 3. The feedback control vector $u_i(x_i) = 0$ when $x_i = 0$, where $i = 1, 2, \ldots, N$.

In this paper, we aim at finding N feedback control policies $u_1(x_1)$, $u_2(x_2), \ldots, u_N(x_N)$ as the decentralized control law to stabilize the large-scale system (1) when dealing with the decentralized control problem. In the control pair $(u_1(x_1), u_2(x_2), \ldots, u_N(x_N))$, the ith control policy $u_i(x_i)$ is only a function of the corresponding local state, namely x_i. As shown in [8], the decentralized control law of the interconnected system is related to the optimal controllers of the isolated subsystems. To deal with the optimal control problem, we need to find the optimal control policy $u_i^*(x_i)$ of the ith isolated subsystem. The optimal control policy minimizes the following infinite horizon cost function:

$$J_i(x_i(t)) = \int_t^\infty r_i(x_i(\tau), u_i(\tau))\, \mathrm{d}\tau \tag{3}$$

where $x_i(\tau)$ denotes the solution of the ith subsystem (2) for the initial condition $x_i(t) \in \Omega_i$ and the input $\{u_i(\tau); \tau > t\}$. $r_i(x_i, u_i) = Q_i(x_i) + u_i^{\mathsf{T}}(x_i) R_i u_i(x_i)$, where $Q_i(x_i)$ is a positive definite function, i.e., $\forall x_i \neq 0$, $Q_i(x_i) > 0$ and $x_i = 0 \Rightarrow Q_i(x_i) = 0$, and $R_i \in \mathbb{R}^{m_i \times m_i}$ is a positive definite matrix.

3. Decentralized control law

In this section, we present the decentralized controller design. The optimal control problem of the isolated subsystems is described under the framework of HJB equations. The decentralized control law is derived by adding some local feedback gains to the isolated optimal control policies.

3.1. Optimal control

In this paper, to design the decentralized control law, we need to solve the optimal control problems for the N isolated subsystems. According to the optimal control theory, we know that the designed feedback control policy must not only stabilize the subsystem on Ω_i, but also guarantee that the cost function (3) is finite. That is to say, the control policy must be admissible.

Definition 1. Consider the ith isolated subsystem, a control policy $\mu_i(x_i)$ is defined as admissible with respect to (3) on Ω_i, denoted by $\mu_i(x_i) \in \Psi_i(\Omega_i)$, if $\mu_i(x_i)$ is continuous on Ω_i, $\mu_i(0) = 0$, $\mu_i(x_i)$ stabilizes the ith isolated subsystem (2) on Ω_i, and $J_i(x_i(t))$ is finite $\forall x_{i0} \in \Omega_i$.

We consider the ith isolated subsystem Σ_i in (2). For any admissible control policy $\mu_i(x_i) \in \Psi_i(\Omega_i)$, we assume that the associated value function

$$V_i(x_i(t)) = \int_t^\infty r_i(x_i(\tau), \mu_i(\tau))\, \mathrm{d}\tau$$

is continuously differentiable. The infinitesimal version of this value function is the nonlinear Lyapunov equation

$$r_i(x_i, \mu_i) + (\nabla V_i(x_i))^{\mathsf{T}}(f_i(x_i) + g_i(x_i)\mu_i(x_i)) = 0 \tag{4}$$

with $V_i(0) = 0$. In (4), the term $\nabla V_i(x_i) = \partial V_i(x_i)/\partial x_i$ denotes the partial derivative of the local value function $V_i(x_i)$ with respect to the local state x_i.

The optimal value function of the ith isolated subsystem can be formulated as

$$V_i^*(x_i(t)) = \min_{\mu_i \in \Psi_i(\Omega_i)} \int_t^\infty r_i(x_i(\tau), \mu_i(\tau))\, \mathrm{d}\tau, \tag{5}$$

and it satisfies the so-called HJB equation

$$0 = \min_{\mu_i \in \Psi_i(\Omega_i)} H_i(x_i, \mu_i, \nabla V_i^*(x_i))$$

where $\nabla V_i^*(x_i) = \partial V_i^*(x_i)/\partial x_i$. The Hamiltonian function of the ith isolated subsystem is defined by

$$H_i(x_i, \mu_i, \nabla V_i(x_i))$$
$$= r_i(x_i, \mu_i) + (\nabla V_i(x_i))^{\mathsf{T}}(f_i(x_i) + g_i(x_i)\mu_i(x_i)). \tag{6}$$

By minimizing the Hamiltonian function (6), the optimal control policy for the ith isolated subsystem can be obtained as

$$u_i^*(x_i) = \arg \min_{\mu_i \in \Psi_i(\Omega_i)} H_i(x_i, \mu_i, \nabla V_i^*(x_i))$$
$$= -\tfrac{1}{2} R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla V_i^*(x_i). \tag{7}$$

Substituting the optimal control policy (7) into the nonlinear Lyapunov equation (4), we can obtain the formulation of the HJB equation in terms of $\nabla V_i^*(x_i)$ as follows:

$$0 = Q_i(x_i) + (\nabla V_i^*(x_i))^{\mathsf{T}} f_i(x_i) - \tfrac{1}{4}(\nabla V_i^*(x_i))^{\mathsf{T}} g_i(x_i) R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla V_i^*(x_i) \tag{8}$$

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

3.2. Stabilizing decentralized control law

According to [2], we modify the local optimal control laws $u_1^*(x_1)$, $u_2^*(x_2)$, …, $u_N^*(x_N)$ by proportionally adding some local feedback gains to obtain a stabilizing decentralized control law for the interconnected large-scale system (1). Now, we give the following theorem to indicate how to add the feedback gains and how to guarantee the asymptotic stability of the subsystems.

Theorem 1. *Considering the ith isolated subsystem Σ_i (2), the feedback control*

$$u_i(x_i) = \pi_i u_i^*(x_i) = -\tfrac{1}{2}\pi_i R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla V_i^*(x_i) \tag{9}$$

can ensure that the ith closed-loop isolated subsystem is asymptotically stable $\forall \pi_i \geq 1/2$.

Proof. The theorem can be proved by showing $V_i^*(x_i)$ is a Lyapunov function. Considering (5), we notice that $V_i^*(x_i) > 0$ for any $x_i \neq 0$ and $V_i^*(x_i) = 0$ when $x_i = 0$, which implies that $V_i^*(x_i)$ is a positive definite function. Then, the derivative of $V_i^*(x_i)$ along the corresponding trajectory of the closed-loop isolated subsystem is given by

$$\begin{aligned}\dot{V}_i^*(x_i) &= (\nabla V_i^*(x_i))^{\mathsf{T}} \dot{x}_i \\ &= (\nabla V_i^*(x_i))^{\mathsf{T}} (f_i(x_i) + g_i(x_i) u_i(x_i)).\end{aligned} \tag{10}$$

Adding and subtracting $(1/2)(\nabla V_i^*(x_i))^T g_i(x_i) u_i^*(x_i)$ to (10), and considering (7)–(9), we have

$$\begin{aligned}\dot{V}_i^*(x_i) &= (\nabla V_i^*(x_i))^{\mathsf{T}} f_i(x_i) - \tfrac{1}{2}(\pi_i - \tfrac{1}{2}) \\ &\quad \times (\nabla V_i^*(x_i))^{\mathsf{T}} g_i(x_i) R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla V_i^*(x_i) \\ &\quad - \tfrac{1}{4}(\nabla V_i^*(x_i))^{\mathsf{T}} g_i(x_i) R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla V_i^*(x_i) \\ &= -Q_i(x_i) - \tfrac{1}{2}(\pi_i - \tfrac{1}{2}) \| R_i^{-1/2} g_i^{\mathsf{T}}(x_i) \nabla V_i^*(x_i) \|^2.\end{aligned} \tag{11}$$

In light of (11), we can obtain that $\dot{V}_i^*(x_i) < 0$ for all $\pi_i \geq 1/2$ and $x_i \neq 0$. Therefore, the conditions for Lyapunov local stability theory are satisfied. The proof is completed.□

To demonstrate the theorem related to the stabilizing decentralized control law, we assume that the interconnected term $Z_i(x(t))$ is characterized by a bound on its magnitude as

$$\| \overline{Z}_i(x) \| \leq \sum_{j=1}^{N} \rho_{ij} h_{ij}(x_j), \quad i = 1, 2, \ldots, N \tag{12}$$

where $\overline{Z}_i(x) = R_i^{1/2} Z_i(x)$ and R_i is the positive definite matrix defined in (3). $h_{ij}(x_j)$ is a positive semi-definite function, and ρ_{ij} is a non-negative constant with $i, j = 1, 2, \ldots, N$. If we define $h_i(x_i) = \max\{h_{1i}(x_i), h_{2i}(x_i), \ldots, h_{Ni}(x_i)\}$, the condition (12) can be rewritten as

$$\| \overline{Z}_i(x) \| \leq \sum_{j=1}^{N} \lambda_{ij} h_j(x_j), \quad i = 1, 2, \ldots, N \tag{13}$$

where $\lambda_{ij} \geq \rho_{ij} h_{ij}(x_j)/h_j(x_j)$ is also a non-negative constant. We assume that $h_i(x_i)$ satisfies

$$h_i^2(x_i) \leq Q_i(x_i), \quad i = 1, 2, \ldots, N \tag{14}$$

where $Q_i(x_i)$ is the positive definite function in (3).

Next, we provide the modified theorem which can be used to establish the stabilizing decentralized control law for the large-scale system (1).

Theorem 2. *For interconnected system (1), there exist N positive numbers $\pi_i^* > 0$, $i = 1, 2, \ldots, N$, such that for any $\pi_i > \pi_i^*$, the feedback controls developed by (9) ensure that the closed-loop interconnected system is asymptotically stable. That is to say, the control pair $(u_1(x_1), u_2(x_2), \ldots, u_N(x_N))$ is the decentralized control law of the large-scale interconnected system (1).*

Proof. According to Theorem 1, we observe that $V_i^*(x_i)$ is Lyapunov function. Here, we select a composite Lyapunov function given by

$$L(x) = \sum_{i=1}^{N} \theta_i V_i^*(x_i) \tag{15}$$

where θ_i is an arbitrary positive constant. Taking the time derivative of $L(x)$ along the trajectories of the closed-loop interconnected system, we have

$$\begin{aligned}\dot{L}(x) &= \sum_{i=1}^{N} \theta_i \dot{V}_i^*(x_i) \\ &= \sum_{i=1}^{N} \theta_i \{(\nabla V_i^*(x_i))^{\mathsf{T}} (f_i(x_i) + g_i(x_i) u_i(x_i)) + (\nabla V_i^*(x_i))^{\mathsf{T}} g_i(x_i) Z_i(x)\}.\end{aligned} \tag{16}$$

Then, considering (11), (13) and (14), and after some basic manipulations, (16) can be turned into the following form:

$$\begin{aligned}\dot{L}(x) \leq -\sum_{i=1}^{N} \theta_i \Bigg\{ & Q_i(x_i) + \frac{1}{2}\left(\pi_i - \frac{1}{2}\right) \| (\nabla J_i^*(x_i))^{\mathsf{T}} g_i(x_i) R_i^{-1/2} \|^2 \\ & - \| (\nabla J_i^*(x_i))^{\mathsf{T}} g_i(x_i) R_i^{-1/2} \| \sum_{j=1}^{N} \lambda_{ij} Q_j^{1/2}(x_j) \Bigg\}.\end{aligned} \tag{17}$$

Like the result presented in [8], we can transform (17) to the following compact form:

$$\begin{aligned}\dot{L}(x) &\leq -\xi^T \begin{bmatrix} \Theta & -\frac{1}{2}\Lambda^T \Theta \\ -\frac{1}{2}\Theta\Lambda & \Theta\Pi \end{bmatrix} \xi \\ &\triangleq -\xi^T \mathcal{A} \xi\end{aligned} \tag{18}$$

where Θ, Λ, Π, and ξ are chosen as those denoted in [8]. In light of (18), we know that sufficiently large π_i can be chosen to guarantee that the matrix $\mathcal{A}$ is positive definite. That is to say, there exist π_i^* so that all $\pi_i \geq \pi_i^*$ are large enough to guarantee the positive definiteness of $\mathcal{A}$. Then, we have $\dot{L}(x) < 0$. Therefore, the conditions for Lyapunov stability theory are satisfied, and the closed-loop interconnected system is asymptotically stable under the action of control pair $(u_1(x_1), u_2(x_2), \ldots, u_N(x_N))$. The proof is completed.□

4. NN-based implementation using online model-free PI algorithm

In this section, we discuss the implementation of the decentralized control law presented in Section 3. We introduce the online PI algorithm in the first subsection. A model-free integral PI algorithm is derived to solve the optimal control problem with completely unknown dynamics in the second subsection. A NN-based implementation of the established model-free integral PI algorithm is discussed at last.

4.1. Online PI algorithm

The formulation developed in (7) displays an array of closed-form expression of the optimal control policy for the ith isolated subsystem, which obviates the need to search for the optimal control policy via optimization process. The existence of $V_i^*(x_i)$ satisfying (8) is the necessary and sufficient condition for optimality. However, it is generally difficult and impossible to obtain the solution $V_i^*(x_i)$ of the HJB equation.

We make effort to obtain the approximation solution of the HJB equation related to the optimal control problem. Instead of directly solving (8), the solution $V_i^*(x_i)$ can be obtained by successively solving the nonlinear Lyapunov equation (4) and updating the policy based on (7). This successive approximation is known as the PI algorithm, and it is described in Algorithm 1 as the fundamental for the model-free PI method. In [29], it was shown that for Algorithm 1 on the domain Ω_i, $V_i^{(p)}(x_i)$ uniformly converges to $V_i^*(x_i)$ with monotonicity $0 < V_i^{(p+1)}(x_i) < V_i^{(p)}(x_i)$, and $\mu_i^{(p)}(x_i)$ is admissible and converges to $u_i^*(x_i)$. The online PI algorithm consisting of policy evaluation and policy improvement can be demonstrated as follows.

Algorithm 1. Online PI.

```
1: Give a small positive real number $\epsilon$. Let $p=0$ and start with
   an initial admissible control policy $\mu_i^{(0)}(x_i)$.
2: Policy Evaluation: Based on the control policy $\mu_i^{(p)}(x_i)$,solve
   the following nonlinear Lyapunov equations for $V_i^{(p)}(x_i)$:
   $0 = Q_i(x_i) + (\mu_i^{(p)}(x_i))^{\mathsf{T}} R_i \mu_i^{(p)}(x_i)$
       $+ (\nabla V_i^{(p)}(x_i))^{\mathsf{T}} (f_i(x_i) + g_i(x_i)\mu_i^{(p)}(x_i)).$      (19)
3: Policy Improvement: Update the control policy by
   $\mu_i^{(p+1)}(x_i) = -\frac{1}{2} R_i^{-1} g_i^{\mathsf{T}}(x_i) \nabla V_i^{(p)}(x_i).$      (20)
4: If $\| V_i^{(p)}(x_i) - V_i^{(p-1)}(x_i) \| \le \epsilon$, stop and obtain the approximate
   optimal control law of the $i$th isolated subsystem; else, set
   $p = p+1$ and go to Step 2.
```

4.2. Model-free PI algorithm

We will develop an online model-free integral PI algorithm for optimal control problems with completely unknown system dynamics. To deal with exploration which relaxes the assumptions of exact knowledge on $f_i(x_i)$ and $g_i(x_i)$, we consider the following nonlinear subsystem explored by a known bounded piecewise continuous signal $e_i(t)$:

$$\dot{x}_i(t) = f_i(x_i(t)) + g_i(x_i(t))[u_i(x_i(t)) + e_i(t)]. \tag{21}$$

The derivative of the value function with respect to time along the trajectory of the subsystem (21) is calculated as

$$\begin{aligned}\dot{V}_i(x_i) &= \nabla V_i^{\mathsf{T}}(x_i)(f_i(x_i) + g_i(x_i)[\mu_i(x_i) + e_i]) \\ &= -r_i(x_i, \mu_i) + \nabla V_i^{\mathsf{T}}(x_i) g_i(x_i) e_i.\end{aligned} \tag{22}$$

We present a lemma which is essential to prove the convergence of the model-free PI algorithm for the isolated subsystems.

Lemma 1. *Solving for $V_i(x_i)$ in the following equation:*

$$\begin{aligned}V_i(x_i(t+T)) - V_i(x_i(t)) = &\int_t^{t+T} \nabla V_i^{\mathsf{T}}(x_i) g_i(x_i) e_i \, \mathrm{d}\tau \\ &- \int_t^{t+T} r_i(x_i, \mu_i(x_i)) \, \mathrm{d}\tau\end{aligned} \tag{23}$$

is equivalent to finding the solution of (22).

Proof. Since $\mu_i(x_i) \in \Psi_i(\Omega_i)$, the value function $V_i(x_i)$ is a Lyapunov function for the subsystem (21), and it satisfies (22) with $r_i(x_i, \mu_i) > 0$, $x_i \neq 0$. We integrate (22) over the interval $[t, t+T]$ to obtain (23). This means that the unique solution of (22), $V_i(x_i)$, also satisfies (23). To complete the proof, we show that (23) has a unique solution by contradiction.

Thus, we assume that there exists another value function $\overline{V}_i(x_i)$ which satisfies (23) with the end condition $\overline{V}_i = 0$. This value function also satisfies $\dot{\overline{V}}_i(x_i) = -r_i(x_i, \mu_i) + \nabla \overline{V}_i^{\mathsf{T}}(x_i) g_i(x_i) e_i$. Subtracting this from (22), we obtain

$$\begin{aligned}0 &= \left(\frac{\mathrm{d}[\overline{V}_i(x_i) - V_i(x_i)]^{\mathsf{T}}}{\mathrm{d}x_i} \right) \times (\dot{x}_i - g_i(x_i) e_i) \\ &= \left(\frac{\mathrm{d}[\overline{V}_i(x_i) - V_i(x_i)]^{\mathsf{T}}}{\mathrm{d}x_i} \right) \times (f_i(x_i) + g_i(x_i)\mu_i(x_i)),\end{aligned} \tag{24}$$

which must hold for any x_i on the system trajectories generated by the stabilizing policy $\mu_i(x_i)$. According to (24), we have $\overline{V}_i(x_i) = V_i(x_i) + c$. As this relation must hold for $x_i(t) = 0$, we know $\overline{V}_i(0) = V_i(0) + c \Rightarrow c = 0$. Thus, $\overline{V}_i(x_i) = V_i(x_i)$, i.e., (23) has a unique solution which is equal to the solution of (22). The proof is completed. □

Integrating (22) from t to $t+T$ with any time interval $T > 0$, and considering (19) and (20), we have

$$\begin{aligned}&V_i^{(p)}(x_i(t+T)) - V_i^{(p)}(x_i(t)) \\ &= -2 \int_t^{t+T} (\mu_i^{(p+1)}(x_i))^{\mathsf{T}} \\ &\quad \times R_i e_i \, \mathrm{d}\tau - \int_t^{t+T} \{Q_i(x_i) + (\mu_i^{(p)}(x_i))^{\mathsf{T}} R_i \mu_i^{(p)}(x_i)\} \, \mathrm{d}\tau.\end{aligned} \tag{25}$$

Eq. (25) which is derived by (20) and (23) plays an important role in relaxing the assumption of knowing the system dynamics, since $f_i(x_i)$ and $g_i(x_i)$ do not appear in the equation. It means that the iteration can be done without knowing the system dynamics. Thus, we obtain the online model-free integral PI algorithm.

Algorithm 2. Online Model-free Integral PI.

```
1: Give a small positive real number $\epsilon$. Let $p=0$ and start with
   an initial admissible control policy $\mu_i^{(0)}(x_i)$.
2: Policy Evaluation and Improvement: Based on the control
   policy $\mu_i^{(p)}(x_i)$, solve the following nonlinear Lyapunov
   equations for $V_i^{(p)}(x_i)$ and $\mu_i^{(p+1)}(x_i)$:
   $V_i^{(p)}(x_i(t)) = \int_t^{t+T} \{Q_i(x_i) + (\mu_i^{(p)}(x_i))^{\mathsf{T}} R_i \mu_i^{(p)}(x_i)\} \, \mathrm{d}\tau$
       $+ 2 \int_t^{t+T} (\mu_i^{(p+1)}(x_i))^{\mathsf{T}} R_i e_i \, \mathrm{d}\tau + V_i^{(p)}(x_i(t+T)).$      (26)
3: If $\| V_i^{(p)}(x_i) - V_i^{(p-1)}(x_i) \| \le \epsilon$, stop and obtain the approximate
   optimal control law of the $i$th isolated subsystem; else, set
   $p = p+1$ and go to Step 2.
```

Remark 1. In Algorithms 1 and 2, we let $V_i^{(p-1)}(x_i) = 0$, when $p=0$. Note that N initial admissible control policies are required in Algorithms 1 and 2.

Theorem 3. *Considering the isolated subsystem* (2), *we give N initial admissible control policies $\mu_1^{(0)}(x_1)$, $\mu_2^{(0)}(x_2), \ldots, \mu_N^{(0)}(x_N)$. Then, using the policy iteration algorithm established in* (26), *the value functions and control policies converge to the optimal ones as $p \to \infty$, i.e.,*

$$V_i^{(p)}(x_i) \to V_i^*(x_i), \quad \mu_i^{(p)}(x_i) \to u_i^*(x_i).$$

Proof. In [27], it was shown that during the iteration process in (20) and (22), if the initial policy $\mu_i^{(0)}(x_i)$ is admissible, all the subsequent control policies will be admissible. Moreover, the iteration result will converge to the solution of the HJB equation. Based on the formation process of (25) and the proven equivalence between (22) and (23), we can conclude that the proposed online model-free PI algorithm will converge to the solution of the optimal control problem for subsystem (21) without using the knowledge of system dynamics. The proof is completed.□

4.3. Online NN implementation

In this subsection, we discuss the NN-based implementation method of the established model-free PI algorithm. A critic NN and an action NN are used to approximate the value function and the control policy of the subsystem, respectively. We assume that for the ith subsystem, $V_i^{(p)}(x_i)$ and $\mu_i^{(p+1)}(x_i)$ are represented on a compact set Ω_i by single-layer NNs as

$$V_i^{(p)}(x_i) = (w_c^i)^{\mathsf T}\phi_c^i(x_i)+\varepsilon_c^i(x_i)$$
$$\mu_i^{(p+1)}(x_i) = (w_a^i)^{\mathsf T}\phi_a^i(x_i)+\varepsilon_a^i(x_i)$$

where $w_c^i \in \mathbb{R}^{N_c^i}$ and $w_a^i \in \mathbb{R}^{N_a^i}$ are unknown bounded ideal weight parameters which will be determined by the established model-free PI algorithm, $\phi_c^i(x_i) \in \mathbb{R}^{N_c^i}$ and $\phi_a^i(x_i) \in \mathbb{R}^{N_a^i}$ are the continuously differentiable nonlinear activation functions, and $\varepsilon_c^i(x_i) \in \mathbb{R}$ and $\varepsilon_a^i(x_i) \in \mathbb{R}$ are the bounded NN approximation errors. Here, the subscripts 'c' and 'a' denote the critic and the action, respectively. Since the ideal weights are unknown, the outputs of the critic NN and the action NN are

$$\hat{V}_i^{(p)}(x_i) = (\hat{w}_c^i)^{\mathsf T}\phi_c^i(x_i) \tag{27}$$

$$\hat{\mu}_i^{(p+1)}(x_i) = (\hat{w}_a^i)^{\mathsf T}\phi_a^i(x_i) \tag{28}$$

where $\hat{w}_c^i$ and $\hat{w}_a^i$ are the current estimated weights.

Using the expressions (27) and (28), (26) can be rewritten as a general form

$$[\psi_k^i]^{\mathsf T}\begin{bmatrix}\hat{w}_c^i\\ \hat{w}_a^i\end{bmatrix} = \theta_k^i \tag{29}$$

with

$$\theta_k^i = \int_{t+(k-1)T}^{t+kT}\{Q_i(x_i)+(\mu_i^{(p)}(x_i))^{\mathsf T}R_i\mu_i^{(p)}(x_i)\}\,\mathrm{d}\tau$$
$$\psi_k^i = \Big[(\phi_c^i(x_i(t+(k-1)T))-\phi_c^i(x_i(t+kT)))^{\mathsf T},$$
$$-2\int_{t+(k-1)T}^{t+kT} R_i e_i(\phi_a^i(x_i))^{\mathsf T}\mathrm{d}\tau\Big]^{\mathsf T}$$

where the measurement time is from $t+(k-1)T$ to $t+kT$. Since (29) is only a 1-dimensional equation, we cannot guarantee the uniqueness of the solution. Similar to [32], we use the least squares sense method to solve the parameter vector over a compact set Ω_i. For any positive integral K_i, we denote $\Phi_i = [\psi_1^i, \psi_2^i, \dots, \psi_{K_i}^i]$ and $\Theta_i = [\theta_1^i, \theta_2^i, \dots, \theta_{K_i}^i]^{\mathsf T}$. Then, we have the following K_i-dimensional equation:

$$\Phi_i^{\mathsf T}\begin{bmatrix}\hat{w}_c^i\\ \hat{w}_a^i\end{bmatrix} = \Theta_i.$$

If $\Phi_i^{\mathsf T}$ has full column rank, the parameters can be solved by

$$\begin{bmatrix}\hat{w}_c^i\\ \hat{w}_a^i\end{bmatrix} = (\Phi_i\Phi_i^{\mathsf T})^{-1}\Phi_i\Theta_i. \tag{30}$$

Therefore, we need to guarantee that the number of collected points K_i satisfies $K_i \ge \operatorname{rank}(\Phi_i) = N_c^i + N_a^i$, which will make $(\Phi_i\Phi_i^{\mathsf T})^{-1}$ exist. The least squares problem in (30) can be solved in real time by collecting enough data points generated from the system (21).

Clearly, the problem of designing the decentralized control law becomes to derive the optimal controllers for the N isolated subsystems. Based on the online model-free integral PI algorithm and NN techniques, we obtain the approximation solutions of HJB equations. We can conclude that the approximate optimal control policies $\hat{\mu}_i(x_i)$ can be obtained. As shown in [8], we have the decentralized control law

$$u_i(x_i) = \pi_i\hat{\mu}_i(x_i). \tag{31}$$

Therefore, the stabilizing decentralized control law of the interconnected large-scale system is derived.

5. Numerical simulations

Two simulation examples are provided in this section to demonstrate the effectiveness of the decentralized control law established in this paper.

5.1. Simulation Example 1

We consider the following nonlinear interconnected system consisting of two subsystems:

$$\begin{aligned}\dot{x}_1 &= \begin{bmatrix}-x_{11}+x_{12}\\ -0.5x_{11}-0.5x_{12}-0.5x_{12}(\cos\,(2x_{11})+2)^2\end{bmatrix}\\ &\quad+\begin{bmatrix}0\\ \cos\,(2x_{11})+2\end{bmatrix}(u_1(x_1)+(x_{11}+x_{12})\sin\,x_{12}^2\cos\,(0.5x_{21}))\\ \dot{x}_2 &= \begin{bmatrix}x_{22}\\ -x_{21}-0.5x_{22}+0.5x_{21}^2x_{22}\end{bmatrix}\\ &\quad+\begin{bmatrix}0\\ x_{21}\end{bmatrix}(u_2(x_2)+0.5(x_{12}+x_{22})\cos\,(e^{x_{21}^2}))\end{aligned} \tag{32}$$

where $x_1 = [x_{11}\ x_{12}]^{\mathsf T} \in \mathbb{R}^2$ and $u_1(x_1) \in \mathbb{R}$ are the state and control variables of subsystem 1, and $x_2 = [x_{21}\ x_{22}]^{\mathsf T} \in \mathbb{R}^2$ and $u_2(x_2) \in \mathbb{R}$ are the state and control variables of subsystem 2. We deal with the optimal control problem of this two isolated subsystems. According to [8], the cost functions of the optimal control problem are

$$J_1(x_{10}) = \int_0^\infty \{x_{11}^2+x_{12}^2+u_1^{\mathsf T}u_1\}\,\mathrm{d}\tau$$
$$J_2(x_{20}) = \int_0^\infty \{x_{22}^2+u_2^{\mathsf T}u_2\}\,\mathrm{d}\tau.$$

Assume that the exact knowledge of the dynamics (32) is fully unknown. We adopt the online model-free PI algorithm to tackle the optimal control problem.

For the isolated subsystem 1

$$\dot{x}_1 = \begin{bmatrix}-x_{11}+x_{12}\\ -0.5x_{11}-0.5x_{12}-0.5x_{12}(\cos\,(2x_{11})+2)^2\end{bmatrix}+\begin{bmatrix}0\\ \cos\,(2x_{11})+2\end{bmatrix}u_1(x_1),$$

we denote the weight vectors of the critic and action networks as

$$\hat{w}_c^1 = [\hat{w}_{c1}^1\ \hat{w}_{c2}^1\ \hat{w}_{c3}^1]^{\mathsf T}$$
$$\hat{w}_a^1 = [\hat{w}_{a1}^1\ \hat{w}_{a2}^1]^{\mathsf T}.$$

The activation functions are chosen as

$$\phi_c^1(x_1) = [x_{11}^2\ x_{11}x_{12}\ x_{12}^2]^{\mathsf T}$$
$$\phi_a^1(x_1) = [x_{11}(2+\cos\,(2x_{11}))\ x_{12}(2+\cos\,(2x_{11}))]^{\mathsf T}.$$

From these parameters, we know $N_c^1 = 3$ and $N_a^1 = 2$, so we conduct the simulation with $K_1 = 10$. We set the initial state and the initial weights as $x_{10} = [1\ -1]^{\mathsf T}$, $\hat{w}_c^1 = [0\ 0\ 0]^{\mathsf T}$ and $\hat{w}_a^1 = [-0.3\ -0.9]^{\mathsf T}$. The period time $T = 0.1$ s and the exploration $e_1(t) = 0.5\sin\,(2\pi t)$ are used in the learning process. The least squares problem is solved after 10 samples are acquired, and thus the weights of the NNs are updated every 1 s. According to [28], the optimal cost function and control policy of the isolated subsystem 1 are $J_1^*(x_1)=0.5x_{11}^2+x_{12}^2$ and $u_1^*(x_1)=-(\cos\ (2x_{11})+2)x_{12}$, respectively. The optimal weights are $w_c^{1*}=[0.5\ 0\ 1]^{\mathsf{T}}$ and $w_a^{1*}=[0\ \ -1]^{\mathsf{T}}$. Figs. 1 and 2 illustrate the evolutions of the weights of the critic network and the action network, respectively. It is clear that the weights approximately converge to the optimal ones. At $t=7$ s, $\hat{w}_c^1=[0.5012\ 0.0003\ 1.0000]^{\mathsf{T}}$ and $\hat{w}_a^1=[-0.0002\ \ -1.0000]^{\mathsf{T}}$.

Similarly, for the isolated subsystem 2, the activation functions are chosen as

$$\phi_c^2(x_2)=[x_{21}^2\ x_{21}x_{22}\ x_{22}^2]^{\mathsf{T}}$$
$$\phi_a^2(x_2)=[x_{21}^2\ x_{21}x_{22}]^{\mathsf{T}}.$$

As $N_c^2=3$ and $N_a^2=2$, we conduct the simulation with $K_2=10$. We set the initial state and the initial weights as $x_{20}=[1\ \ -1]^{\mathsf{T}}$, $\hat{w}_c^2=[0\ 0\ 0]^{\mathsf{T}}$ and $\hat{w}_a^2=[0\ 0]^{\mathsf{T}}$. The period time $T=0.1$ s and the exploration $e_2(t)=0.5\sin\ (2\pi t)$ are used in the learning process. The optimal cost function and control policy of the isolated subsystem 2 are $J_2^*(x_2)=x_{21}^2+x_{22}^2$ and $u_2^*(x_2)=-x_{21}x_{22}$. The optimal weights are $w_c^{2*}=[1\ 0\ 1]^{\mathsf{T}}$ and $w_a^{2*}=[0\ \ -1]^{\mathsf{T}}$. Figs. 3 and 4 illustrate the evolutions of the weights of the critic network and the action network, respectively. It is clear that the weights approximately converge to the optimal ones. At $t=9$ s, $\hat{w}_c^2=[1.0000\ \ -0.0000\ 1.0000]^{\mathsf{T}}$ and $\hat{w}_a^2=[-0.0000\ \ -1.0000]^{\mathsf{T}}$.

According to (31), we choose $\pi_1=\pi_2=2$ to obtain the decentralized control law $(\pi_1\hat{\mu}_1(x_1),\pi_2\hat{\mu}_2(x_2))$ of the interconnected system (32). By applying the decentralized control law to control the interconnected system for 60 s, we obtain the evolution process of the state trajectories shown in Figs. 5 and 6. Obviously, the applicability of the decentralized control law developed in this paper has been testified by these simulation results.

5.2. Simulation Example 2

Consider the classical multimachine power system with governor controllers [16]

$$\dot{\delta}_i(t)=\omega_i(t)$$
$$\dot{\omega}_i(t)=-\frac{D_i}{2H_i}\omega_i(t)+\frac{\omega_0}{2H_i}[P_{mi}(t)-P_{ei}(t)]$$
$$\dot{P}_{mi}(t)=\frac{1}{T_i}[-P_{mi}(t)+u_{gi}(t)]$$
$$P_{ei}(t)=E'_{qi}\sum_{j=1}^{N}E'_{qj}[B_{ij}\sin\ \delta_{ij}(t)+G_{ij}\cos\ \delta_{ij}(t)]$$

where for $1\le i$ and $j\le N$, $\delta_i(t)$ represents the angle of the ith generator; $\delta_{ij}(t)=\delta_i(t)-\delta_j(t)$ is the angular difference between the ith and jth generators; $\omega_i(t)$ is the relative rotor speed; $P_{mi}(t)$ and $P_{ei}(t)$ are the mechanical power and the electrical power, respectively; E'_{qi} is the transient electromotive force in quadrature axis and is assumed to be constant under high-gain SCR controllers; D_i, H_i, and T_i are the damping constant, the inertia constant, and the governor time constant, respectively; B_{ij} and G_{ij} are the imaginary and real parts of the admittance matrix, respectively; and $u_{gi}(t)$ is the speed governor control signal for the ith generator; ω_0 is the steady state frequency.

A three-machine power system is considered in our numerical simulation. The parameters of the system are the same as those in [16]. The weighting matrices are set to be $Q_i(x_i)=x_i^{\mathsf{T}}\times 1000I_3\times x_i$ and R_i=1, for $i=1,2,3$. Similarly, as in [16], the multimachine power system can be rewritten as the following form:

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

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

$$\Delta\dot{P}_{mi}(t)=\frac{1}{T_i}[-\Delta P_{mi}(t)+u_i(t)-d_i(t)].$$

We define the state $x_i=[\Delta\delta_i(t)\ \Delta\omega_i(t)\ \Delta P_{mi}(t)]^{\mathsf{T}}=[x_{i1}\ x_{i2}\ x_{i3}]^{\mathsf{T}}$, where $\Delta\delta_i(t)=\delta_i(t)-\delta_{i0}$, $\Delta\omega_i(t)=\omega_i(t)-\omega_{i0}$,

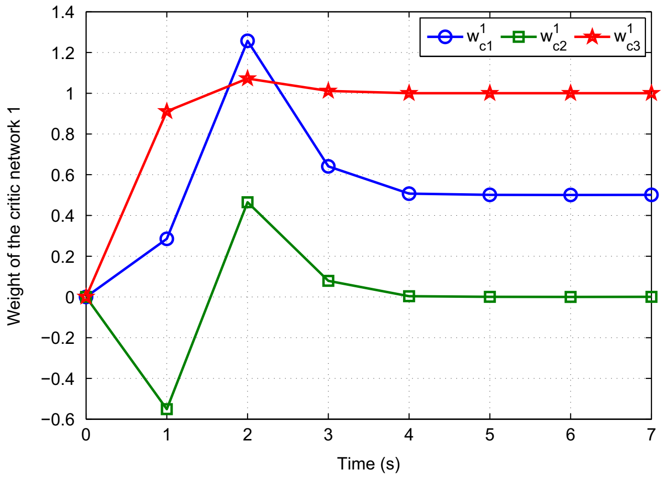


Fig. 1. Evolutions of the weight of the critic network 1.

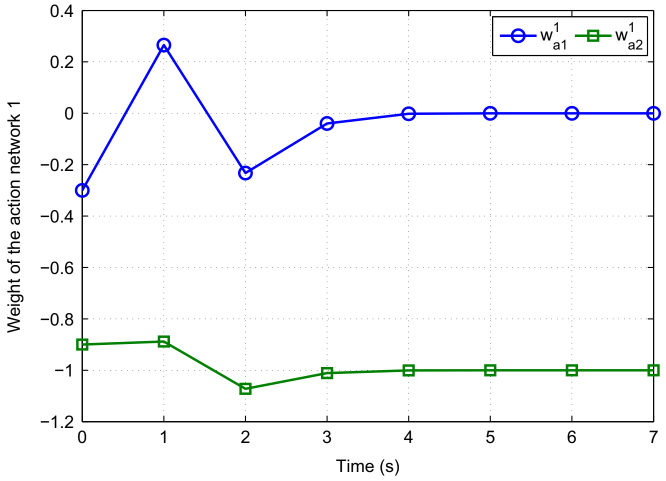


Fig. 2. Evolutions of the weight of the action network 1.

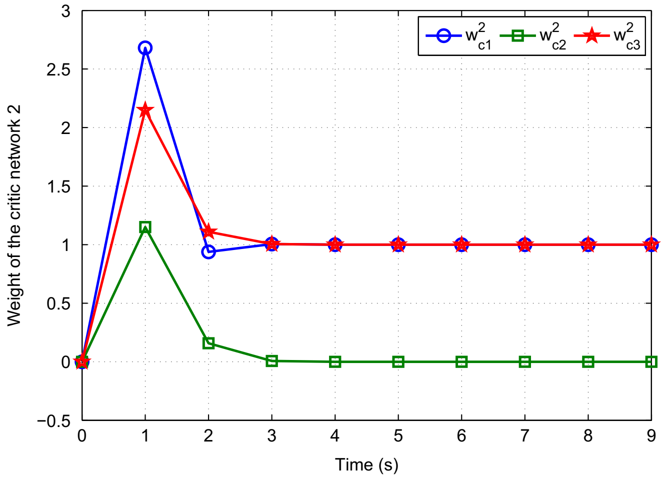


Fig. 3. Evolutions of the weight of the critic network 2.

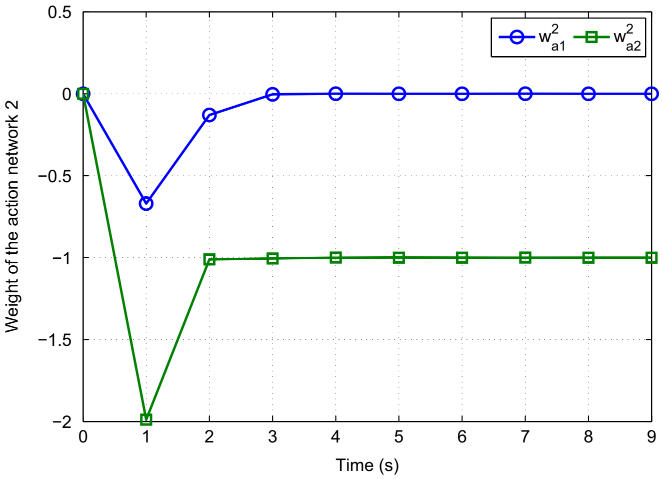


Fig. 4. Evolutions of the weight of the action network 2.

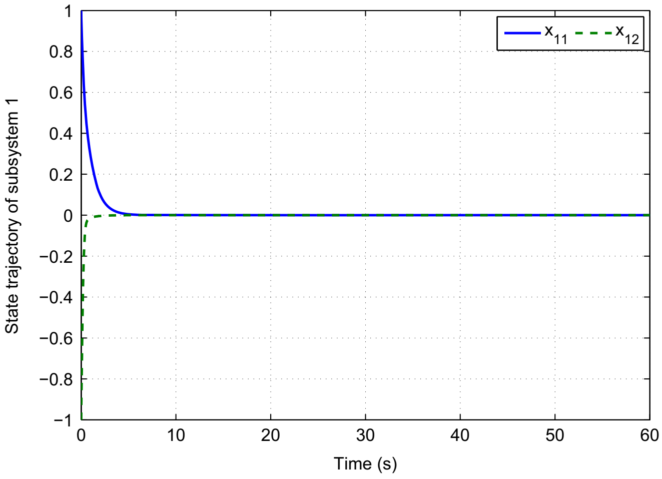


Fig. 5. State trajectory of subsystem 1 under the action of the decentralized control law.

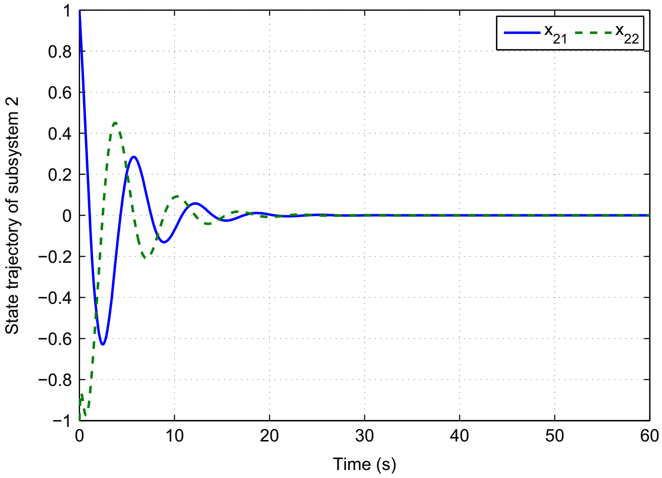


Fig. 6. State trajectory of subsystem 2 under the action of the decentralized control law.

$\Delta P_{mi}(t) = P_{mi}(t) - P_{ei}(t)$, $u_i(t) = u_{gi}(t) - P_{ei}(t)$, and

$$d_i(t) = E'_{qi} \sum_{j=1, j \neq i}^{N} \{E'_{qj}[B_{ij}\cos\delta_{ij}(t) - G_{ij}\sin\ \delta_{ij}(t)] \times [\Delta\omega_i(t) - \Delta\omega_j(t)]\}.$$

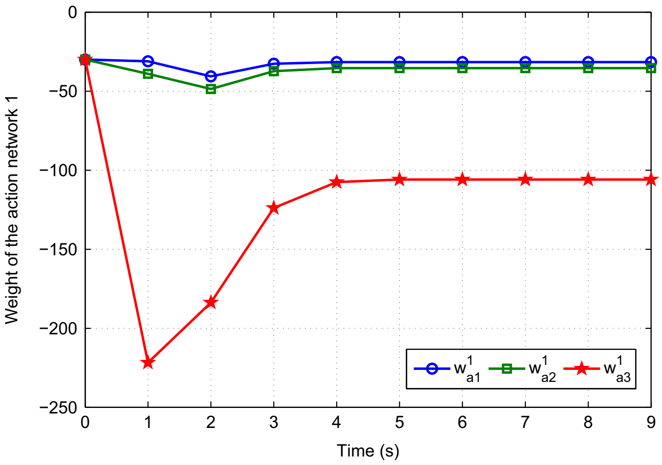


Fig. 7. Evolutions of the weight of the action network 1.

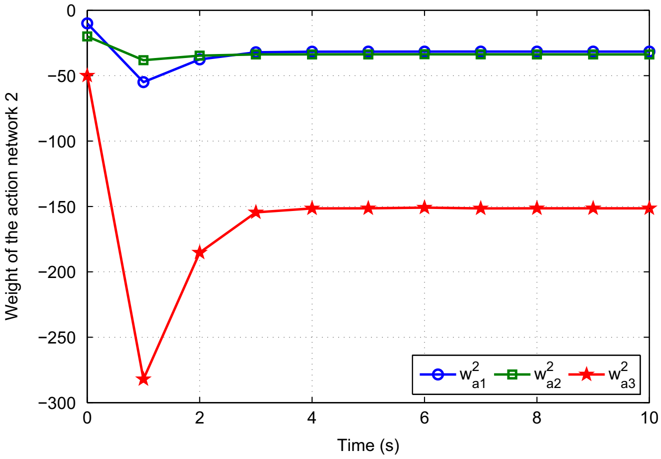


Fig. 8. Evolutions of the weight of the action network 2.

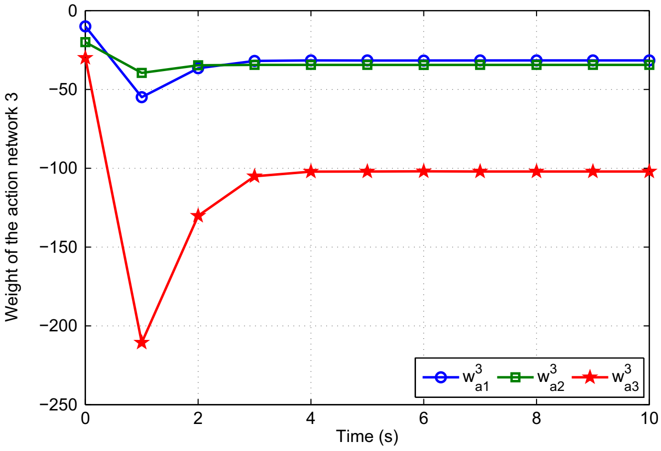


Fig. 9. Evolutions of the weight of the action network 3.

For each isolated subsystem, we denote the weight vectors of the critic and action networks as

$$\hat{w}^i_c = [\hat{w}^i_{c1}\ \hat{w}^i_{c2}\ \hat{w}^i_{c3}\ \hat{w}^i_{c4}\ \hat{w}^i_{c5}\ \hat{w}^i_{c6}]^{\mathsf{T}}$$

$$\hat{w}^i_a = [\hat{w}^i_{a1}\ \hat{w}^i_{a2}\ \hat{w}^i_{a3}]^{\mathsf{T}}.$$

The activation functions are chosen as

$$\phi^i_c(x_i) = [x^2_{i1}\ x_{i1}x_{i2}\ x_{i1}x_{i3}\ x^2_{i2}\ x_{i2}x_{i3}\ x^2_{i3}]^{\mathsf{T}}$$

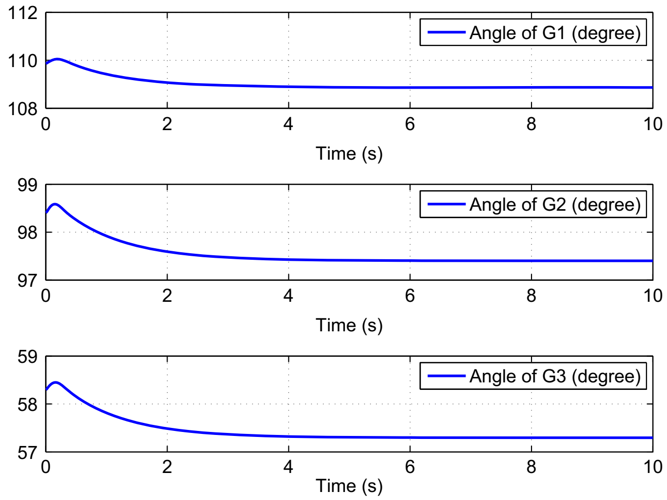


Fig. 10. Angle of the generators under the action of the decentralized control law.

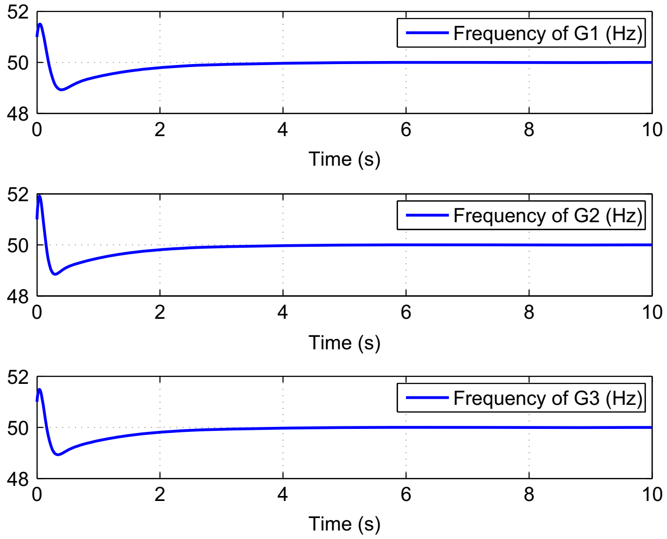


Fig. 11. Frequency of the generators under the action of the decentralized control law.

$\phi_a^i(x_i) = [x_{i1}\ x_{i2}\ x_{i3}]^{\mathrm{T}}$.

From these parameters, we know $N_c^i = 6$ and $N_a^i = 3$, so we conduct the simulation with $K_i = 10$. We set the initial state and the initial weights of the critic networks as $x_{i0} = [1\ 1\ 1]^{\mathrm{T}}$, $\hat{w}_c^i = 100\times[1\ 1\ 1\ 1\ 1\ 1]^{\mathrm{T}}$, for $i = 1, 2, 3$. The initial weights of the action networks are chosen as $\hat{w}_a^1 = -[30\ 30\ 30]^{\mathrm{T}}$, $\hat{w}_a^2 = -[10\ 20\ 50]^{\mathrm{T}}$ and $\hat{w}_a^3 = \ -[10\ 20\ 30]^{\mathrm{T}}$. The period time $T = 0.1$ s and the exploration $e_i(t) = 0.01(\sin(2\pi t) + \cos(2\pi t))$ are used in the learning process. The least squares problem is solved after 10 samples are acquired, and thus the weights of the NNs are updated every 1 s. Figs. 7, 8, and 9 illustrate the evolutions of the weights of the action network for the isolated subsystem 1, 2 and 3, respectively. It is clear that the weights approximately converge after some update steps.

According to (31), we choose $\pi_1 = \pi_2 = \pi_3 = 1$ to obtain the control pair $(\pi_1\hat{\mu}_1(x_1), \pi_2\hat{\mu}_2(x_2), \pi_3\hat{\mu}_3(x_3))$ as the stabilizing decentralized control law of the interconnected system. By applying the decentralized control law to control the interconnected power system for 10 s, we obtain the evolution process of the power angle deviations and frequencies of the generators shown in Figs. 10 and 11, respectively. Obviously, the applicability of the decentralized control law developed in this paper has been testified by these simulation results.

6. Conclusion

In this paper, a stabilizing decentralized control law for a class of nonlinear large-scale systems with unknown dynamics is established using a NN-based online model-free integral PI algorithm. The decentralized control law is derived by the optimal controllers of the isolated subsystems. We use an online model-free integral PI algorithm with an exploration to solve the HJB equations related to the optimal control problem of the isolated subsystems. To implement the constructed algorithm, we use the actor-critic technique and the least squares implementation method. We demonstrate the effectiveness of the developed decentralized control law by two simulation examples.

References

[1] L. Bakule, Decentralized control: an overview, Ann. Rev. Control 32 (2008) 87–98.
[2] A. Saberi, On optimality of decentralized control for a class of nonlinear interconnected systems, Automatica 24 (1988) 101–104.
[3] P. Ioannou, Decentralized adaptive control of interconnected systems, IEEE Trans. Autom. Control 31 (1986) 291–298.
[4] J.T. Spooner, K.M. Passino, Decentralized adaptive control of nonlinear systems using radial basis neural networks, IEEE Trans Autom. Control 44 (1999) 2050–2057.
[5] K. Kalsi, J. Lian, S.H. Zak, Decentralized dynamic output feedback control of nonlinear interconnected systems, IEEE Trans. Autom. Control 55 (2010) 1964–1970.
[6] J. Lavaei, Decentralized implementation of centralized controllers for interconnected systems, IEEE Trans. Autom. Control 57 (2012) 1860–1865.
[7] T. Li, R. Li, J. Li, Decentralized adaptive neural control of nonlinear interconnected large-scale systems with unknown time delays and input saturation, Neurocomputing 74 (2011) 2277–2283.
[8] D. Liu, D. Wang, H. Li, Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 418–428.
[9] R. Bellman, Dynamic Programming, Princeton University Press, Princeton, NJ, 1957.
[10] F.-Y. Wang, H. Zhang, D. Liu, Adaptive dynamic programming: an introduction, IEEE Comput. Intell. Mag. 4 (2009) 39–47.
[11] H. Zhang, Q. Wei, D. Liu, An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games, Automatica 47 (2011) 207–214.
[12] D. Wang, D. Liu, Q. Wei, Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach, Neurocomputing 78 (2012) 14–22.
[13] D. Liu, Q. Wei, Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems, IEEE Trans. Cybern. 43 (2013) 779–789.
[14] D. Liu, Y. Zhang, H. Zhang, A self-learning call admission control scheme for CDMA cellular networks, IEEE Trans. Neural Netw. 16 (2005) 1219–1228.
[15] D. Liu, H. Javaherian, O. Kovalenko, T. Huang, Adaptive critic learning techniques for engine torque and air–fuel ratio control, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 38 (2008) 988–993.
[16] 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: Express Briefs 59 (2012) 693–697.
[17] 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. Learning Syst. 23 (2012) 1118–1129.
[18] T. Huang, D. Liu, A self-learning scheme for residential energy system control and management, Neural Comput. Appl. 22 (2013) 259–269.
[19] D. Zhao, Z. Zhang, Y. Dai, Self-teaching adaptive dynamic programming for gomoku, Neurocomputing 78 (2012) 23–29.
[20] Y. Jiang, Z.-P. Jiang, Robust adaptive dynamic programming with an application to power systems, IEEE Trans. Neural Netw. Learn. Syst. 24 (2013) 1150–1156.
[21] F.L. Lewis, D. Vrabie, Reinforcement learning and adaptive dynamic programming for feedback control, IEEE Circuits Syst. Mag. 9 (2009) 32–50.
[22] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, vol. 1, Cambridge University Press, Cambridge, MA, 1998.
[23] J. Si, A.G. Barto, W.B. Powell, D.C. Wunsch, et al., Handbook of Learning and Approximate Dynamic Programming, IEEE Press, Los Alamitos, 2004.
[24] S.J. Bradtke, B.E. Ydstie, A.G. Barto, Adaptive linear quadratic control using policy iteration, in: American Control Conference, 1994, vol. 3, IEEE, Baltimore, MD, 1994, pp. 3475–3479.
[25] A. Al-Tamimi, F.L. Lewis, M. Abu-Khalaf, Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 38 (2008) 943–949.
[26] Y.-M. Park, M.-S. Choi, K.Y. Lee, An optimal tracking neurocontroller for nonlinear dynamic systems, IEEE Trans. Neural Netw. 7 (1996) 1099–1110.

[27] M. Abu-Khalaf, F.L. Lewis, Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach, Automatica 41 (2005) 779–791.

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

[29] D. Vrabie, F. Lewis, Neural network approach to continuous time direct adaptive optimal control for partially unknown nonlinear systems, Neural Netw. 22 (2009) 237–246.

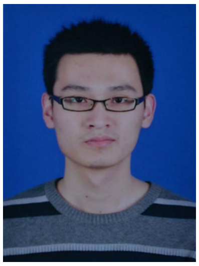
[30] Y. Jiang, Z.-P. Jiang, Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics, Automatica 48 (2012) 2699–2704.

[31] D. Liu, X. Yang, H. Li, Adaptive optimal control for a class of continuous-time affine nonlinear systems with unknown internal dynamics, Neural Comput. Appl. (2012) 1843–1850.

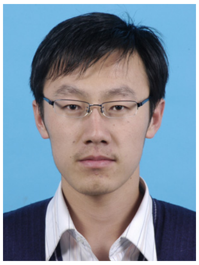
[32] J.Y. Lee, J.B. Park, Y.H. Choi, Integral Q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems, Automatica 48 (2012) 2850–2859.

[33] J.Y. Lee, J.B. Park, Y.H. Choi, Integral reinforcement learning with explorations for continuous-time nonlinear systems, in: The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE, Brisbane, Australia, 2012, pp. 1042–1047.

Derong Liu received the B.S. degree in Mechanical Engineering from the East China Institute of Technology (now Nanjing University of Science and Technology), Nanjing, China, in 1982, the M.S. degree in Automatic Control Theory and Applications from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1987, and the Ph.D. degree in Electrical Engineering from the University of Notre Dame, Indiana, USA, in 1994. Dr. Liu was a Product Design Engineer with China North Industries Corporation, Jilin, China, from 1982 to 1984. He was an Instructor with the Graduate School of the Chinese Academy of Sciences, Beijing, from 1987 to 1990. He was a Staff Fellow with General Motors Research and Development Center, from 1993 to 1995. He was an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, from 1995 to 1999. He joined the University of Illinois at Chicago in 1999, and became a Full Professor of Electrical and Computer Engineering and of Computer Science in 2006. He was selected for the "100 Talents Program" by the Chinese Academy of Sciences in 2008, and now he serves as the Associate Director of The State Key Laboratory of Management and Control for Complex Systems at the Institute of Automation. He has published 15 books (six research monographs and nine edited volumes). Dr. Liu was an Associate Editor of Automatica from 2006 to 2009. Currently, he is an elected AdCom member of the IEEE Computational Intelligence Society and he is the Editor-in-Chief of the IEEE Transactions on Neural Networks and Learning Systems. He also serves as an Associate Editor of IEEE Transactions on Control Systems Technology, IEEE Transactions on Systems, Man, and Cybernetics: Systems, IEEE Transactions on Intelligent Transportation Systems, Soft Computing, Neurocomputing, Neural Computing and Applications, and Science in China Series F: Information Sciences. He was an Associate Editor of the IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications from 1997 to 1999, the IEEE Transactions on Signal Processing from 2001 to 2003, the IEEE Transactions on Neural Networks from 2004 to 2009, the IEEE Computational Intelligence Magazine from 2006 to 2009, and the IEEE Circuits and Systems Magazine from 2008 to 2009, and the Letters Editor of the IEEE Transactions on Neural Networks from 2006 to 2008. He received the Faculty Early Career Development Award from the National Science Foundation in1999, the University Scholar Award from University of Illinois from 2006 to 2009, and the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008. He is a Fellow of the IEEE and a Fellow of the International Neural Network Society.



Chao Li received the B.S. degree in Mechatronics from the Nanjing University of Science and Technology in 2012. He is currently working toward the Ph.D. degree in the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is also with the University of Chinese Academy of Sciences, Beijing. His research interests include neural networks, reinforcement learning and adaptive dynamic programming.



Hongliang Li received the B.S. degree in Mechanical Engineering and Automation from Beijing University of Posts and Telecommunications in 2010. He is currently working toward the Ph.D. degree in the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is also with the University of Chinese Academy of Sciences, Beijing. His research interests include neural networks, reinforcement learning, adaptive dynamic programming, game theory and multi-agent systems.

Ding Wang received the B.S. degree in mathematics from the Zhengzhou University of Light Industry, Zhengzhou, China, the M.S. degree in operational research and cybernetics from Northeastern University, Shenyang, China, and the Ph.D. degree in Control Theory and Control Engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007, 2009, and 2012, respectively. He is currently an assistant professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interests include adaptive dynamic programming, neural networks, and intelligent control.

Hongwen Ma received the B.S. degree in Electric Engineering and Automation from the Nanjing University of Science and Technology in 2012. He is currently working toward the Ph.D. degree in the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is also with the University of Chinese Academy of Sciences, Beijing. His research interests include neural networks, networked control systems, and multi-agent systems.