Intelligent Critic Control With Disturbance Attenuation for Affine Dynamics Including an Application to a Microgrid System

Ding Wang, *Member, IEEE*, Haibo He, *Senior Member, IEEE*, Chaoxu Mu, *Member, IEEE*, and Derong Liu, *Fellow, IEEE*

***Abstract*—In this paper, a computationally efficient framework for intelligent critic control design and application of continuous-time input-affine systems is established with the purpose of disturbance attenuation. The described problem is formulated as a two-player zero-sum differential game and the adaptive critic mechanism with intelligent component is employed to solve the minimax optimization problem. First, a neural identifier is developed to reconstruct the unknown dynamical information incorporating stability analysis. Next, the optimal control law and the worst-case disturbance law are designed by introducing and tuning a critic neural network. Moreover, the closed-loop system is proved to possess the uniform ultimate boundedness. At last, the present method is applied to a smart microgrid and then is further adopted to control a general nonlinear system via simulation, thereby substantiating the performance of disturbance attenuation.**

***Index Terms*—Adaptive/approximate dynamic programming, adaptive critic control, disturbance attenuation, intelligent control, neural identification, smart microgrid.**

Manuscript received September 21, 2016; revised December 8, 2016; accepted January 30, 2017. Date of publication February 24, 2017; date of current version May 10, 2017. This work was supported in part by the National Natural Science Foundation of China under Grants U1501251, 61533017, 61233001, 61304018, 51529701, and 61520106009, in part by the Beijing Natural Science Foundation under Grant 4162065, in part by the Tianjin Natural Science Foundation under Grant 14JCQNJC05400, in part by the U.S. National Science Foundation under Grant ECCS 1053717, in part by the Early Career Development Award of SKLMCCS, and in part by the Research Fund of the Tianjin Key Laboratory of Process Measurement and Control under Grant TKLPMC-201612. *(Corresponding author: Haibo He.)*

D. Wang is with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: ding.wang@ia.ac.cn). This research was conducted when he was a Visiting Scholar at the Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881 USA.

H. He is with the Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881 USA (e-mail: he@ele.uri.edu).

C. Mu is with the Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: cxmu@tju.edu.cn).

D. Liu is with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: derong@ustb.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2017.2674633

I. Introduction

IN CONTROL theory and engineering, robustness is an important criterion to evaluate the performance of the designed controller with respect to uncertain parameters or disturbances of the dynamical plant [1]–[5], where stability analysis is the basic issue as studied in [6]–[9]. For example, there are several excellent robust control algorithms developed for microgrids to improve their performance with respect to disturbances and uncertainties [10]–[13]. In particular, the H_∞ method generally focuses on constructing the worst-case control law for specified plants including additive disturbances or dynamical uncertainties [14], [15]. In order to obtain a controller that minimizes the cost function in the worst-case disturbance, the H_∞ design requires to find the Nash equilibrium solution by considering the Hamilton–Jacobi–Isaacs equation. However, it is intractable to acquire the analytic solution for general nonlinear systems. Hence, the adaptive/approximate dynamic programming strategy was developed [16], as an effective method to solve optimal control problems using a design manner of forward-in-time. Therein, function approximation structures, such as artificial neural networks, were always included [17], [18]. Remarkably, the core of the adaptive/approximate dynamic programming approach lies in the adaptive critic mechanism with intelligent component [16]–[18]. In other words, it can be regarded as an intelligent control implementation of the traditional optimization design, especially for complex systems with nonlinearities and uncertainties.

When mentioning the research of adaptive/approximate dynamic programming, in the last decade, it has gained much progress in term of optimal control design for discrete-time systems [19]–[21], continuous-time systems [22]–[26], and potential applications [27]–[31], particularly for power system design and control [27]–[30]. In addition, the nonlinear H_∞ control [32]–[36] and multiagent differential game design [37] also have been revisited and studied under this framework incorporating adaptivity and learning ability. However, existing works are mostly conducted for optimal regulation problem or H_∞ control design with known dynamics, which lacks an extension to zero-sum differential game problems for unknown nonlinear plant. Additionally, building an architecture composed of actor, critic, and disturbance elements, is often complicated since it relies on a number of neural networks and occupies large computational resource. This, of course, motivates our research on developing an effectively intelligent H_∞ control method with simple identification structure and adaptive critic learning module.

Compared with the traditional mathematical programming methods, the adaptive/approximate dynamic programming is appropriate for solving sequential optimization and control problems under uncertain environment, which are common in real-world applications [31]. Hence, this paper focuses on designing the intelligent critic control with unknown nonlinear dynamics for the purpose of achieving disturbance attenuation. The main contribution lies in that the neural network identification framework is combined with the adaptive critic learning technique, in order to study the nonlinear H_∞ feedback control and application with unknown dynamical information.

Notations: Throughout this paper, $\mathbb{R}$ represents the set of all real numbers. $\mathbb{R}^n$ is the Euclidean space of all n-dimensional real vectors. $\mathbb{R}^{n\times m}$ is the space of all $n \times m$ real matrices. $\|\cdot\|$ denotes the vector norm of a vector in $\mathbb{R}^n$ or the matrix norm of a matrix in $\mathbb{R}^{n\times m}$. I_n represents the $n \times n$ identity matrix. $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ stand for the maximal and minimal eigenvalues of a matrix, respectively. Let Ω be a compact subset of $\mathbb{R}^n$, Ω_u be a compact subset of $\mathbb{R}^m$, and $\mathscr{A}(\Omega)$ be the set of admissible controls on Ω. $\mathcal{L}_2[0,\infty)$ denotes a space of functions where the Lebesgue integral of the element is finite. ρ is the $\mathcal{L}_2$-gain performance level. "T" is used for representing the transpose operation, $\mathrm{tr}(\cdot)$ is adopted to conduct the trace operation, and $\nabla(\cdot) \triangleq \partial(\cdot)/\partial x$ is employed to denote the gradient operator.

II. Problem Statement and Preliminaries

Consider a class of continuous-time input-affine systems with external perturbations described by

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

where $x(t) \in \Omega \subset \mathbb{R}^n$ is the state vector, $u(t) \in \Omega_u \subset \mathbb{R}^m$ is the control vector, $v(t) \in \mathbb{R}^q$ is the perturbation vector with $v(t) \in \mathcal{L}_2[0,\infty)$, $z(t) = Q(x(t)) \in \mathbb{R}^p$ is the objective output, and $f(\cdot)$, $g(\cdot)$, $h(\cdot)$ are differentiable in their arguments with $f(0) = 0$. We let the initial state at $t = 0$ be $x(0) = x_0$ and $x = 0$ be the equilibrium point of the controlled plant. The system (1) is assumed to be controllable.

Assumption 1: The control and disturbance matrices $g(x)$ and $h(x)$ are upper bounded such that $\|g(x)\| \le \lambda_g$ and $\|h(x)\| \le \lambda_h$, where λ_g and λ_h are positive constants.

The nonlinear H_∞ design needs to derive a feedback control law $u(x)$ such that the closed-loop dynamics is asymptotically stable and has $\mathcal{L}_2$-gain no larger than ϱ, that is,

$$\int_0^\infty \left(\|Q(x(\tau))\|^2 + u^{\mathsf{T}}(\tau)u(\tau)\right)\mathrm{d}\tau \le \varrho^2 \int_0^\infty \|v(\tau)\|^2 \mathrm{d}\tau \tag{2}$$

where $\|Q(x)\|^2 = x^{\mathsf{T}}(t)Qx(t)$ and $Q \in \mathbb{R}^{n\times n}$ is a positive definite matrix.

In light of [32]–[36], designing the H_∞ control can be regarded as a two-player zero-sum differential game. The solution of H_∞ control problem is the saddle point of zero-sum game theory, denoted as a pair of laws (u^*, v^*), where u^* and v^* are called the optimal control and the worst-case disturbance, respectively. Let

$$\begin{aligned} U(x(\tau), u(\tau), v(\tau)) = \ & x^{\mathsf{T}}(\tau)Qx(\tau) + u^{\mathsf{T}}(\tau)u(\tau) \\ & - \varrho^2 v^{\mathsf{T}}(\tau)v(\tau) \end{aligned} \tag{3}$$

represent the utility function and define the infinite horizon cost function as

$$J(x(t), u, v) = \int_t^\infty U(x(\tau), u(\tau), v(\tau))\mathrm{d}\tau. \tag{4}$$

For simplicity, the cost $J(x(t), u, v)$ is often written as $J(x(t))$ or $J(x)$ in the sequel. What we always concern is the cost function starting from $t = 0$, which is denoted as $J(x(0)) = J(x_0)$. Here, our goal is to find the feedback saddle point solution (u^*, v^*), such that the Nash condition

$$J^*(x_0) = \min_u \max_v J(x_0, u, v) = \max_v \min_u J(x_0, u, v) \tag{5}$$

holds. For an admissible control $u \in \mathscr{A}(\Omega)$, if the related cost function (4) is continuously differentiable, then its infinitesimal version is the nonlinear Lyapunov equation

$$0 = U(x, u, v) + (\nabla J(x))^{\mathsf{T}}(f(x) + g(x)u + h(x)v) \tag{6}$$

with $J(0) = 0$. Define the Hamiltonian of system (1) as

$$H(x, u, v, \nabla J(x)) = U(x, u, v) + (\nabla J(x))^{\mathsf{T}}(f + gu + hv). \tag{7}$$

Employing Bellman's optimality principle, the optimal cost function $J^*(x)$ makes sure that the so-called Hamilton–Jacobi–Isaacs equation $\min_u \max_v H(x, u, v, \nabla J^*(x)) = 0$ holds. The saddle point solution (u^*, v^*) satisfies the stationary condition [34], which can be used to derive the optimal control law and the worst-case disturbance law by [32]–[36]

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

$$v^*(x) = \frac{1}{2\varrho^2}h^{\mathsf{T}}(x)\nabla J^*(x). \tag{8b}$$

Considering the two formulas in (8), the Hamilton–Jacobi–Isaacs equation turns to the form

$$\begin{aligned} 0 = \ & x^{\mathsf{T}}Qx + (\nabla J^*(x))^{\mathsf{T}}f(x) \\ & - \frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}g(x)g^{\mathsf{T}}(x)\nabla J^*(x) \\ & + \frac{1}{4\varrho^2}(\nabla J^*(x))^{\mathsf{T}}h(x)h^{\mathsf{T}}(x)\nabla J^*(x), J^*(0) = 0. \end{aligned} \tag{9}$$

The formula (9) is called the Hamilton–Jacobi–Isaacs equation and is difficult to solve in theory. This inspires us to find an alternate avenue to overcome the difficulty by adopting the adaptive critic mechanism.

III. Intelligent Critic Control Design With Disturbance Attenuation

A. Identification of the Controlled Plant With Stability

Here, we introduce a three-layer neural network identifier to reconstruct the dynamics (1) by using the input/output data. Let the number of neurons in the hidden layer be denoted by l_m and then the system (1) can be approximated by

$$\dot{x} = Ax + \omega_m^{\mathsf{T}} \sigma_m(\bar{z}) + \varepsilon_m \tag{10}$$

where $A \in \mathbb{R}^{n\times n}$ is a stable design matrix, $\omega_m \in \mathbb{R}^{l_m\times n}$ is the ideal weight matrix between the hidden layer and the output layer, $\sigma_m(\cdot) \in \mathbb{R}^{l_m}$ is a differentiable and monotonically increasing activation function such as $\sigma_m(\cdot) = \tanh(\cdot)$, $\bar{z} = \nu_m^{\mathsf{T}} z$ with $\bar{z} \in \mathbb{R}^{l_m}$, $\nu_m \in \mathbb{R}^{(n+m+q)\times l_m}$ is the ideal weight matrix between the input layer and the hidden layer, $z = [x^{\mathsf{T}}, u^{\mathsf{T}}, v^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^{n+m+q}$ is the augmented input vector, and $\varepsilon_m \in \mathbb{R}^n$ is the reconstruction error. With the differentiable activation function $\sigma_m(\cdot)$, for any $a, b \in \mathbb{R}$ $(a \geq b)$, there exists a constant λ_0 $(\lambda_0 > 0)$, such that the relationship

$$\sigma_m(a) - \sigma_m(b) \leq \lambda_0(a - b) \tag{11}$$

holds [23], [26]. Note that when performed for a vector, (11) is applied to each element of the vector. Hence, under the condition (11), we can further derive the following inequality

$$\|\sigma_m(\xi_a) - \sigma_m(\xi_b)\| \leq \lambda_0 \|\xi_a - \xi_b\| \tag{12}$$

for any two vectors ξ_a and ξ_b with the same dimensions.

For simplicity, we let the input-hidden weight matrix ν_m be constant and only tune the hidden-output weight matrix. Actually, we can initialize the input-hidden matrix randomly and keep it unchanged during the identification. Then, the output of neural network identifier is

$$\dot{\hat{x}} = A\hat{x} + \hat{\omega}_m^{\mathsf{T}}(t)\sigma_m(\hat{z}) \tag{13}$$

where $\hat{\omega}_m(t)$ is the currently estimated weight matrix of the ideal value ω_m at time t , $\hat{x}$ is the estimated system state, and $\hat{z} = \nu_m^{\mathsf{T}}[\hat{x}^{\mathsf{T}}, u^{\mathsf{T}}, v^{\mathsf{T}}]^{\mathsf{T}}$.

Let $\tilde{\omega}_m = \hat{\omega}_m - \omega_m$ be the weight estimation error of the neural identifier and $\tilde{x} = \hat{x} - x$ be the identification error. Then, according to (10) and (13), the dynamical equation with respect to the identification error can be derived as

$$\dot{\tilde{x}} = A\tilde{x} + \tilde{\omega}_m^{\mathsf{T}}(t)\sigma_m(\hat{z}) + \omega_m^{\mathsf{T}}(\sigma_m(\hat{z}) - \sigma_m(\bar{z})) - \varepsilon_m. \tag{14}$$

Observing the identifier weight matrices and the reconstruction error, we present two common assumptions often used in the community, such as [23] and [26], which are helpful to analyze the stability of the identification error dynamics. Note that the reconstruction error ε_m can be arbitrarily small, as long as the number of the hidden layer node l_m is large enough. In the neural identification field, the reconstruction error ε_m is often considered to be bounded by a known constant. However, observing (14), we know that ε_m is closely linked with $\tilde{x}$. From a mathematical perspective, the assumption that ε_m is bounded by a function of $\tilde{x}$ is regarded to be more general.

Assumption 2: The ideal weight matrices are bounded such as $\|\omega_m\| \leq \lambda_{\omega_m}$ and $\|\nu_m\| \leq \lambda_{\nu_m}$, where λ_{ω_m} and λ_{ν_m} are positive constants.

Assumption 3: The neural reconstruction error ε_m is upper bounded by a function of the identification error, such that $\varepsilon_m^{\mathsf{T}}\varepsilon_m \leq \lambda_{\varepsilon_m}\tilde{x}^{\mathsf{T}}\tilde{x}$, where λ_{ε_m} is a positive constant.

Theorem 1: Using the neural identifier (13) with a suitable stable matrix A, if the network weight is tuned by $\dot{\hat{\omega}}_m = -\alpha_m\sigma_m(\hat{z})\tilde{x}^{\mathsf{T}}$, where $\alpha_m > 0$ is the learning rate, then the state estimation error $\tilde{x}$ is asymptotically stable.

Proof: Choose a Lyapunov function candidate as the form $L_1(t) = L_{11}(t) + L_{12}(t)$, where

$$L_{11}(t) = \tilde{x}^{\mathsf{T}}(t)\tilde{x}(t), L_{12}(t) = \frac{1}{\alpha_m}\text{tr}\{\tilde{\omega}_m^{\mathsf{T}}(t)\tilde{\omega}_m(t)\}. \tag{15}$$

We take the derivative of $L_{11}(t)$ along the trajectory of the error system (14) and obtain

$$\begin{aligned}\dot{L}_{11}(t) = \; & 2\tilde{x}^{\mathsf{T}}[A\tilde{x} + \omega_m^{\mathsf{T}}(\sigma_m(\hat{z}) - \sigma_m(\bar{z})) - \varepsilon_m] \\ & + 2\tilde{x}^{\mathsf{T}}\tilde{\omega}_m^{\mathsf{T}}(t)\sigma_m(\hat{z}).\end{aligned} \tag{16}$$

Using the adjusting criterion $\dot{\hat{\omega}}_m = -\alpha_m\sigma_m(\hat{z})\tilde{x}^{\mathsf{T}}$, the fact that $\dot{\tilde{\omega}}_m = \dot{\hat{\omega}}_m$, and the property of trace operation, we find that

$$\dot{L}_{12}(t) = \frac{2}{\alpha_m}\text{tr}(\tilde{\omega}_m^{\mathsf{T}}\dot{\tilde{\omega}}_m) = -2\tilde{x}^{\mathsf{T}}\tilde{\omega}_m^{\mathsf{T}}\sigma_m(\hat{z}). \tag{17}$$

According to (16) and (17), we can obtain

$$\dot{L}_1(t) = 2\tilde{x}^{\mathsf{T}}A\tilde{x} + 2\tilde{x}^{\mathsf{T}}\omega_m^{\mathsf{T}}(\sigma_m(\hat{z}) - \sigma_m(\bar{z})) - 2\tilde{x}^{\mathsf{T}}\varepsilon_m. \tag{18}$$

Adopting (12) and observing Assumption 2, we have $\|\hat{z} - \bar{z}\| \leq \|\nu_m\|\|\hat{x} - x\| \leq \lambda_{\nu_m}\|\tilde{x}\|$, such that

$$\begin{aligned}& 2\tilde{x}^{\mathsf{T}}\omega_m^{\mathsf{T}}(\sigma_m(\hat{z}) - \sigma_m(\bar{z})) \\ & \leq \tilde{x}^{\mathsf{T}}\omega_m^{\mathsf{T}}\omega_m\tilde{x} + (\sigma_m(\hat{z}) - \sigma_m(\bar{z}))^{\mathsf{T}}(\sigma_m(\hat{z}) - \sigma_m(\bar{z})) \\ & \leq \tilde{x}^{\mathsf{T}}\omega_m^{\mathsf{T}}\omega_m\tilde{x} + \lambda_0^2\lambda_{\nu_m}^2\tilde{x}^{\mathsf{T}}\tilde{x}.\end{aligned} \tag{19}$$

Recalling Assumption 3, we derive $-2\tilde{x}^{\mathsf{T}}\varepsilon_m \leq (1 + \lambda_{\varepsilon_m})\tilde{x}^{\mathsf{T}}\tilde{x}$, which is combined with (19) to further obtain the reduction of (18) to

$$\begin{aligned}\dot{L}_1(t) \leq \; & \tilde{x}^{\mathsf{T}}[2A + \omega_m^{\mathsf{T}}\omega_m + (1 + \lambda_{\varepsilon_m} + \lambda_0^2\lambda_{\nu_m}^2)I_n]\tilde{x} \\ \triangleq \; & -\tilde{x}^{\mathsf{T}}\Xi\tilde{x}\end{aligned} \tag{20}$$

where the square matrix

$$\Xi = -2A - \omega_m^{\mathsf{T}}\omega_m - (1 + \lambda_{\varepsilon_m} + \lambda_0^2\lambda_{\nu_m}^2)I_n. \tag{21}$$

If A is selected to ensure that $\Xi > 0$, then the time derivative of the Lyapunov function is $\dot{L}_1(t) < 0$ for any $\tilde{x} \neq 0$. Thus, we find that the identification error can approach zero as time goes to infinity (i.e., $\tilde{x}(t) \to 0$ as $t \to \infty$), which completes the proof. ■

According to Theorem 1, we observe that the model neural network is actually an asymptotically stable identifier. Hence, after a sufficient learning stage, we can obtain an available neural identifier with finally converged weights as follows:

$$\dot{x} = f(x) + g(x)u + h(x)v = Ax + \omega_m^{\mathsf{T}}\sigma_m(\bar{z}) \tag{22}$$

which, in fact, represents the information of the state derivative of the controlled plant. In addition, we respectively take the partial derivative of (22) with regard to the control u and the disturbance v and derive that

$$g(x) = \omega_m^{\mathsf{T}} \left(\frac{\partial \sigma(\bar{z})}{\partial \bar{z}} \right) \nu_m^{\mathsf{T}} \begin{bmatrix} 0_{n\times m} \\ I_m \\ 0_{q\times m} \end{bmatrix} \tag{23a}$$

$$h(x) = \omega_m^{\mathsf{T}} \left(\frac{\partial \sigma(\bar{z})}{\partial \bar{z}} \right) \nu_m^{\mathsf{T}} \begin{bmatrix} 0_{n\times q} \\ 0_{m\times q} \\ I_q \end{bmatrix} \tag{23b}$$

where the term $\partial\sigma(\bar{z})/\partial\bar{z}$ is in fact a l_m-dimensional square matrix. The two formulas in (23) reconstruct the information of the control matrix and the disturbance matrix. Remarkably, the obtained neural dynamics reflects the data-based learning of the controlled plant and thus is helpful for the intelligent H_∞ control design in the sequel.

Remark 1: Strictly speaking, the state derivative $\dot{x}$ in (22), the control matrix $g(x)$ in (23a), and the disturbance matrix $h(x)$ in (23b) should be denoted by $\hat{\dot{x}}$, $\hat{g}(x)$, and $\hat{h}(x)$, respectively, as approximated values. However, this may cause the complication of symbols as well as the confusion of control design description. For convenience of analysis, from the next part, we keep on using the notations $\dot{x}$, $g(x)$, and $h(x)$, without stating that they are actually the converged variables after the sufficient learning session.

B. Intelligent Critic Control Design With Stability

For performing the neural control implementation, we denote l_c as the number of neurons in the hidden layer. According to the universal approximation property [38], the cost function $J^*(x)$ can be reconstructed by a neural network with a single hidden layer on a compact set Ω as $J^*(x) = \omega_c^{\mathsf{T}}\sigma_c(x) + \varepsilon_c(x)$, where $\omega_c \in \mathbb{R}^{l_c}$ is the ideal weight vector, $\sigma_c(x) \in \mathbb{R}^{l_c}$ is the activation function, and $\varepsilon_c(x) \in \mathbb{R}$ is the reconstruction error. Then, the gradient vector is

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

Since the ideal weight is unknown, a critic neural network is introduced and used for approximating the cost function as $\hat{J}(x) = \hat{\omega}_c^{\mathsf{T}}\sigma_c(x)$, where $\hat{\omega}_c \in \mathbb{R}^{l_c}$ denotes the estimated weight vector. Similarly, we have the gradient vector

$$\nabla\hat{J}(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{25}$$

Adopting the neural network expression (24), the optimal control law (8a) and the worst-case disturbance law (8b) are written as

$$u^*(x) = -\frac{1}{2}g^{\mathsf{T}}(x)((\nabla\sigma_c(x)^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x)) \tag{26a}$$

$$v^*(x) = \frac{1}{2\varrho^2}h^{\mathsf{T}}(x)((\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x)). \tag{26b}$$

Incorporating the critic neural network, the approximate expressions of the above two laws are

$$\hat{u}(x) = -\frac{1}{2}g^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c \tag{27a}$$

$$\hat{v}(x) = \frac{1}{2\varrho^2}h^{\mathsf{T}}(x)(\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c. \tag{27b}$$

For the control u and the disturbance v, we apply the neural network expression to the Hamiltonian and derive that

$$\begin{aligned} H(x, u(x), v(x), \omega_c) &= U(x, u(x), v(x)) + \omega_c^{\mathsf{T}}\nabla\sigma_c(x) \\ &\times (f(x) + g(x)u(x) + h(x)v(x)) \triangleq e_{cH} \end{aligned} \tag{28}$$

where the term

$$e_{cH} = -(\nabla\varepsilon_c(x))^{\mathsf{T}}(f(x) + g(x)u(x) + h(x)v(x)) \tag{29}$$

represents the residual error arisen in the approximate operation. Meanwhile, the approximate Hamiltonian is

$$\begin{aligned} \hat{H}(x, u(x), v(x), \hat{\omega}_c) &= U(x, u(x), v(x)) + \hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x) \\ &\times (f(x) + g(x)u(x) + h(x)v(x)) \triangleq e_c. \end{aligned} \tag{30}$$

Let us define the error vector between the ideal weight and the estimated value as $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$. Then, we combine (28) with (30) and yield

$$e_c = -\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(x)(f(x) + g(x)u(x) + h(x)v(x)) + e_{cH} \tag{31}$$

which comprises the relationship of the above two versions of the Hamiltonian.

Next, we turn to train the critic neural network as the main learning component and need to design the weight vector $\hat{\omega}_c$ to minimize the objective function $E_c = (1/2)e_c^2$. In the learning stage, the approximated control and disturbance laws are used. We employ the normalized steepest descent algorithm to adjust the weight as

$$\begin{aligned} \dot{\hat{\omega}}_c &= -\alpha_c \frac{1}{(1+\phi^{\mathsf{T}}\phi)^2}\left(\frac{\partial E_c}{\partial\hat{\omega}_c}\right) \\ &= -\alpha_c \frac{\phi}{(1+\phi^{\mathsf{T}}\phi)^2}(U(x, \hat{u}(x), \hat{v}(x)) + \phi^{\mathsf{T}}\hat{\omega}_c) \end{aligned} \tag{32}$$

where $\alpha_c > 0.5$ represents the learning rate to be determined

$$\phi = \nabla\sigma_c(x)\,(f(x) + g(x)\hat{u}(x) + h(x)\hat{v}(x)) \tag{33}$$

is a l_c-dimensional column vector, and the term $(1+\phi^{\mathsf{T}}\phi)^2$ is utilized for normalization.

In what follows, we construct the error dynamics of the critic network and focus on its stability. By recalling $\dot{\tilde{\omega}}_c = -\dot{\hat{\omega}}_c$ and introducing

$$\phi_1 = \frac{\phi}{(1+\phi^{\mathsf{T}}\phi)}, \phi_2 = 1 + \phi^{\mathsf{T}}\phi \geq 1 \tag{34}$$

we further derive that the critic error dynamics is written as

$$\dot{\tilde{\omega}}_c = -\alpha_c\phi_1\phi_1^{\mathsf{T}}\tilde{\omega}_c + \alpha_c\frac{\phi_1}{\phi_2}e_{cH}. \tag{35}$$

When designing an adaptive control system, the persistence of excitation assumption is necessary to perform system identification [39]. In the adaptive critic control community, it is also required because we need to identify the parameter of the critic network to approximate the optimal cost function.

Assumption 4 (cf. [22]): The signal ϕ_1 is persistently exciting within the interval $[t, t+T]$, $T>0$, i.e., there exist two constants $\varsigma_1>0$, $\varsigma_2>0$ such that

$$\varsigma_1 I_{l_c} \le \int_t^{t+T} \phi_1(\tau)\phi_1^{\mathsf{T}}(\tau)\mathrm{d}\tau \le \varsigma_2 I_{l_c} \tag{36}$$

holds for all t.

According to Assumption 4, the persistence of excitation condition guarantees $\lambda_{\min}(\phi_1\phi_1^{\mathsf{T}})>0$, which is important to perform the stability analysis. In the sequel, the uniformly ultimately bounded stability [40] of the closed-loop system is analyzed. Before proceeding, the following assumption is required, as usually stated and used in the literature [23], [26].

Assumption 5: On the given compact set Ω, the terms ω_c, $\nabla\sigma_c(x)$, $\nabla\varepsilon_c(x)$, and e_{cH} are upper bounded such that $\|\omega_c\| \le \lambda_{\omega_c}$, $\|\nabla\sigma_c(x)\| \le \lambda_{d\sigma_c}$, $\|\nabla\varepsilon_c(x)\| \le \lambda_{d\varepsilon_c}$, and $|e_{cH}| \le \lambda_{e_c}$, where λ_{ω_c}, $\lambda_{d\sigma_c}$, $\lambda_{d\varepsilon_c}$, and λ_{e_c} are positive constants.

Theorem 2: For the nonlinear system (1), we suppose that Assumptions 1 and 5 hold. The neural identifier is constructed by (13) with $\tilde{x}=\hat{x}-x$ being the identification error. The approximate optimal control law and worst-case disturbance law are given by (27a) and (27b), respectively, where the constructed critic network is tuned by adopting (32). Then, the closed-loop system state x, the system identification error $\tilde{x}$, and the critic weight error $\tilde{\omega}_c$ are uniformly ultimately bounded, respectively, by

$$\sqrt{\frac{\lambda_1}{2\varrho^2\lambda_{\min}(Q)}} \triangleq \mathcal{B}_x, \sqrt{\frac{\lambda_1}{2\varrho^2\lambda_{\min}(\Xi)}} \triangleq \mathcal{B}_{\tilde{x}} \tag{37a}$$

$$\sqrt{\frac{\lambda_1}{\varrho^2(2\alpha_c-1)\lambda_{\min}(\phi_1\phi_1^{\mathsf{T}})-\lambda_{d\sigma_c}^2(\varrho^2\lambda_g^2+\lambda_h^2)}} \triangleq \mathcal{B}_{\tilde{\omega}_c} \tag{37b}$$

where $\lambda_1=\varrho^2(\lambda_g^2\lambda_{d\varepsilon_c}^2+\alpha_c^2\lambda_{e_c}^2)+\lambda_h^2\lambda_{d\sigma_c}^2\lambda_{\omega_c}^2$ is a constant.

Proof: Choose a Lyapunov function candidate composed of three terms as $L_2(t)=L_{21}(t)+L_{22}(t)+L_{23}(t)$, where

$$L_{21}(t)=J^*(x(t)), L_{22}(t)=L_1(t), L_{23}(t)=\frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}(t)\tilde{\omega}_c(t). \tag{38}$$

We compute the time derivative of the Lyapunov function $L_2(t)$ along the dynamics (1), (14), and (35) and obtain

$$\dot{L}_{21}(t)=(\nabla J^*(x))^{\mathsf{T}}(f(x)+g(x)\hat{u}(x)+h(x)\hat{v}(x)) \tag{39a}$$

$$\dot{L}_{22}(t) \le -\tilde{x}^{\mathsf{T}}\Xi\tilde{x} \tag{39b}$$

$$\dot{L}_{23}(t)=-\alpha_c\tilde{\omega}_c^{\mathsf{T}}\phi_1\phi_1^{\mathsf{T}}\tilde{\omega}_c+\alpha_c\frac{\tilde{\omega}_c^{\mathsf{T}}\phi_1}{\phi_2}e_{cH}. \tag{39c}$$

Note that the formula (8) implies that

$$(\nabla J^*(x))^{\mathsf{T}}g(x)=-2u^{*\mathsf{T}}(x) \tag{40a}$$

$$(\nabla J^*(x))^{\mathsf{T}}h(x)=2\varrho^2v^{*\mathsf{T}}(x). \tag{40b}$$

In addition, (9) reveals

$$(\nabla J^*(x))^{\mathsf{T}}f(x)=-x^{\mathsf{T}}Qx+u^{*\mathsf{T}}(x)u^*-\varrho^2v^{*\mathsf{T}}(x)v^*(x). \tag{41}$$

Considering (39a) and based on (40) and (41), we derive

$$\begin{aligned}\dot{L}_{21}(t)=&-x^{\mathsf{T}}Qx+u^{*\mathsf{T}}(x)u^*(x)-\varrho^2v^{*\mathsf{T}}(x)v^*(x)\\&-2u^{*\mathsf{T}}(x)\hat{u}(x)+2\varrho^2v^{*\mathsf{T}}(x)\hat{v}(x)\\ \le&-x^{\mathsf{T}}Qx+\|u^*(x)-\hat{u}(x)\|^2+\varrho^2\|\hat{v}(x)\|^2.\end{aligned} \tag{42}$$

Recalling the neural-network-related formulas of $u^*(x)$ and $\hat{u}(x)$, i.e., (26a) and (27a), it follows from the fact $\omega_c=\hat{\omega}_c+\tilde{\omega}_c$ and Assumptions 1 and 5 that

$$\begin{aligned}&\|u^*(x)-\hat{u}(x)\|^2\\&\quad=\frac{1}{4}\|g^{\mathsf{T}}(x)((\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c+\nabla\varepsilon_c(x))\|^2\\&\quad\le\frac{1}{2}\lambda_g^2(\lambda_{d\sigma_c}^2\|\tilde{\omega}_c\|^2+\lambda_{d\varepsilon_c}^2).\end{aligned} \tag{43}$$

Then, it follows from (42) that

$$\begin{aligned}\dot{L}_{21}(t)\le&-\lambda_{\min}(Q)\|x\|^2+\frac{1}{2\varrho^2}\lambda_{d\sigma_c}^2(\varrho^2\lambda_g^2+\lambda_h^2)\|\tilde{\omega}_c\|^2\\&+\frac{1}{2\varrho^2}(\varrho^2\lambda_g^2\lambda_{d\varepsilon_c}^2+\lambda_h^2\lambda_{d\sigma_c}^2\lambda_{\omega_c}^2).\end{aligned} \tag{44}$$

When considering (39c) and Assumption 5, we employ the Young's inequality and derive that

$$\dot{L}_{23}(t)\le-\left(\alpha_c-\frac{1}{2}\right)\lambda_{\min}(\phi_1\phi_1^{\mathsf{T}})\|\tilde{\omega}_c\|^2+\frac{1}{2}\alpha_c^2\lambda_{e_c}^2. \tag{45}$$

By combining (44) and (45), it follows from (39) that

$$\begin{aligned}\dot{L}_2(t)\le&-\lambda_{\min}(Q)\|x\|^2-\lambda_{\min}(\Xi)\|\tilde{x}\|^2+\frac{1}{2\varrho^2}\lambda_1\\&-\left[\left(\alpha_c-\frac{1}{2}\right)\lambda_{\min}(\phi_1\phi_1^{\mathsf{T}})\right.\\&\left.-\frac{1}{2\varrho^2}\lambda_{d\sigma_c}^2(\varrho^2\lambda_g^2+\lambda_h^2)\right]\|\tilde{\omega}_c\|^2.\end{aligned} \tag{46}$$

Hence, if one of the following inequalities holds:

$$\|x\|>\mathcal{B}_x, \|\tilde{x}\|>\mathcal{B}_{\tilde{x}}, \|\tilde{\omega}_c\|>\mathcal{B}_{\tilde{\omega}_c} \tag{47}$$

then $\dot{L}_2(t)<0$. Thus, based on the standard Lyapunov extension theorem, we obtain the conclusion that the state vector x, the neural identification error $\tilde{x}$, and the critic weight error $\tilde{\omega}_c$ are all uniformly ultimately bounded. Clearly, their upper bounds are, respectively, written as $\mathcal{B}_x$, $\mathcal{B}_{\tilde{x}}$, and $\mathcal{B}_{\tilde{\omega}_c}$, which completes the proof. ■

Corollary 1: The approximate control law $\hat{u}$ in (27a) and disturbance law $\hat{v}$ designed in (27b) converge to the neighborhood

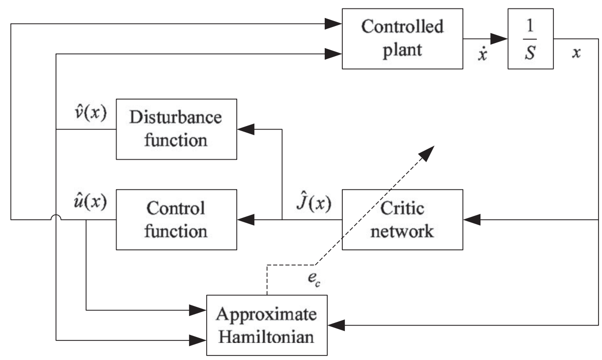

Fig. 1. Simple control structure (the solid line represents the signal and the dashed line represents the back-propagating path).

of their optimal values, u^* and v^*, with finite bounds

$$\frac{1}{2}\lambda_g(\lambda_{d\sigma_c}\mathcal{B}_{\tilde{\omega}_c}+\lambda_{d\varepsilon_c}) \triangleq \mathcal{B}_u \quad (48a)$$

$$\frac{1}{2\varrho^2}\lambda_h(\lambda_{d\sigma_c}\mathcal{B}_{\tilde{\omega}_c}+\lambda_{d\varepsilon_c}) \triangleq \mathcal{B}_v \quad (48b)$$

respectively.

Proof: According to Theorem 2, we derive that $\|\tilde{\omega}_c\| < \mathcal{B}_{\tilde{\omega}_c}$. Based on (26a) and (27a), we find that

$$\|u^*-\hat{u}\| = \frac{1}{2}\|g^{\mathsf{T}}(x)((\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c+\nabla\varepsilon_c(x))\| \le \mathcal{B}_u \quad (49)$$

where $\mathcal{B}_u$ stands for the finite bound with respect to the control signal. Using a similar mechanism, we can obtain from (26b) and (27b) that

$$\|v^*-\hat{v}\| = \frac{1}{2\varrho^2}\|h^{\mathsf{T}}(x)((\nabla\sigma_c(x))^{\mathsf{T}}\tilde{\omega}_c+\nabla\varepsilon_c(x))\| \le \mathcal{B}_v \quad (50)$$

where $\mathcal{B}_v$ denotes the finite bound with respect to the disturbance signal. This actually completes the proof. ■

Remark 2: According to Theorem 2, we can find that the bounds of the x, $\tilde{x}$, and $\tilde{\omega}_c$ can be adjusted to be arbitrarily small, if we enlarge the related parameters such as $\lambda_{\min}(Q)$, $\lambda_{\min}(\Xi)$, and α_c. Clearly, in light of Corollary 1, the bounds of approximate control with respect to the optimal value and approximate disturbance law with respect to the optimal one also can be modulated based on the initial parameter settings. This kind of stability is weaker than the asymptotic stability.

At the end of the section, we give a simple diagram of the present intelligent critic control scheme as shown in Fig. 1.

IV. Experimental Verification

In this section, we first apply the present control approach to a microgrid system with linear dynamics and then turn to the simulation verification of a more general nonlinear plant.

A. Application to a Microgrid System

Smart grids including various load changes and multiple renewable generations have received intensive attention in recent years. In modern power systems, many kinds of distributed and renewable energies have been frequently integrated into microgrids. However, the involvement of the intermittent power may bring in some unforeseeable causes, which will inevitably affect the stability of microgrids. In particular, the imbalance between load consumptions and power generations is a common phenomenon, which may result in the frequency deviation, especially for microgrids [41], [42]. Hence, the frequency stability of microgrids has been a significant topic to the development of modern power systems [43]. The load frequency control is seemed as an essential control design strategy to guarantee the reliable operation in the field of power systems [44]. It also requires a robust controller to ensure the balance between all power generations and load consumptions under uncertain and disturbed environment.

We consider a benchmark power system constructed in Fig. 2, which is composed of regular generations (microturbines), renewable energy generation sources (photovoltaic arrays), and a set of demand sides (smart homes and loads). The benchmark power system can be regarded as a microgrid, which is affected and controlled by the local smart microgrid management center [45]. The active power is produced by microturbines and photovoltaic arrays to balance all local loads. The states of the microgrid system incorporate the frequency deviation, the turbine power, and the governor position. All the variables can be measured by distributed sensors and then transmitted to the microgrid management center via a communication medium, where the collected data is also processed. Then, the generated control signals are sent back to each participating unit in the local system, so as to guarantee the frequency stability.

Now, we formulate the load frequency control problem inspired by the excellent work in [46] and [47]. The primary design objective is to guarantee that the load frequency of the microgrid system can maintain the command frequency level even if there exist load disturbances and energy uncertainties. Some mathematical notations are provided in Table I to facilitate describing the benchmark system. The dynamics of this system is given as follows:

$$\Delta\dot{\xi}_f = -\frac{1}{T_p}\Delta\xi_f+\frac{k_p}{T_p}\Delta\xi_t+\frac{k_p}{T_p}v \quad (51a)$$

$$\Delta\dot{\xi}_t = -\frac{1}{T_t}\Delta\xi_t+\frac{1}{T_t}\Delta\xi_p, \quad (51b)$$

$$\Delta\dot{\xi}_p = -\frac{1}{s_pT_g}\Delta\xi_f-\frac{1}{T_g}\Delta\xi_p+\frac{1}{T_g}u \quad (51c)$$

where $u \in \mathbb{R}$ and $v \in \mathbb{R}$ are, respectively, regarded as the control signal to be designed and the perturbation signal caused by photovoltaic power and load demand change.

Under this circumstance, we define $x = [\Delta\xi_f, \Delta\xi_t, \Delta\xi_p]^{\mathsf{T}} \in \mathbb{R}^3$ as the state vector, where $x_1 = \Delta\xi_f$, $x_2 = \Delta\xi_t$, and $x_3 = \Delta\xi_p$. Then, the state-space description of the proposed power system dynamics (51) can be written as

$$\dot{x} = \begin{bmatrix} -\frac{1}{T_p} & \frac{k_p}{T_p} & 0 \\ 0 & -\frac{1}{T_t} & \frac{1}{T_t} \\ -\frac{1}{s_pT_g} & 0 & -\frac{1}{T_g} \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{T_g} \end{bmatrix} u + \begin{bmatrix} \frac{k_p}{T_p} \\ 0 \\ 0 \end{bmatrix} v. \quad (52)$$

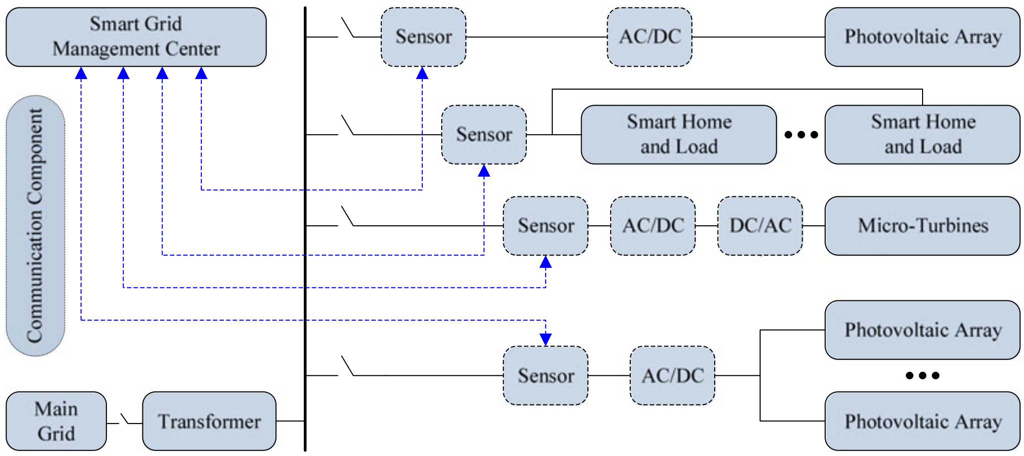

Fig. 2. Simple diagram of the proposed microgrid system, which is considered as a microgrid. The module "ac/dc" represents the power conversion between alternating current and direct current. The dashed blue line denotes the signal transmission via a communication channel component.

TABLE I
PARAMETERS OF THE MICROGRID SYSTEM

Parameters	Meaning
$\Delta\xi_f$	The frequency deviation
$\Delta\xi_t$	The turbine power
$\Delta\xi_p$	The governor position value
T_t	Time constant of the turbine
T_g	Time constant of the governor
T_p	Time constant of the power system
k_p	Gain of the power system
s_p	The speed regulation coefficient

TABLE II
VALUES OF THE MICROGRID SYSTEM

Parameters	T_t	T_g	T_p	k_p	s_p
Values	5	0.2	2	0.5	0.5

Here, the control matrix and the disturbance matrix, i.e., $g = [0, 0, 1/T_g]^\mathsf{T}$ and $h = [k_p/T_p, 0, 0]^\mathsf{T}$ are both constant, which obviously, satisfies the bounded assumption. Note that (52) is an input-affine form (1) with linear dynamics.

In the sequel, we design the intelligent critic controller and then evaluate the H_∞ control performance of the dynamical plant (52). For the simulation purpose, we select the values of the related parameters as shown in Table II. Let the initial state of the controlled plant be $x_0 = [0.1, -0.2, 0.2]^\mathsf{T}$ and choose $Q = I_3$ and $\varrho = 5$. During the simulation process, we select six hidden neurons, i.e., $l_m = 6$, and perform the neural identification algorithm for 100 s with $\sigma_m(\cdot) = \tanh(\cdot)$, $\lambda_0 = 1$, and $\alpha_m = 0.3$, to obtain the converged weight matrices of input-to-hidden (5×6) and hidden-to-output (6×3) given as

$$\begin{bmatrix} 0.2356 & -0.0832 & -0.1201 & 0.1075 & 0.3984 & -0.9423 \\ -0.8418 & 0.3918 & -0.6774 & 0.2900 & 0.6064 & 0.7372 \\ -0.3848 & 0.5043 & 0.8364 & 0.1357 & -0.9555 & 0.7439 \\ 0.2989 & -0.1097 & 0.5611 & 0.0228 & -0.3531 & -0.4407 \\ -0.2576 & -0.6032 & -0.4062 & 0.0979 & -0.8700 & -0.6095 \end{bmatrix}$$

$$\begin{bmatrix} -0.0439 & 0.0886 & -0.1490 & 1.0762 & -0.0325 & -0.1926 \\ -0.4653 & 0.4896 & 0.1111 & 1.4269 & 0.2398 & -0.4250 \\ 4.5871 & -2.3624 & 4.5912 & 2.8955 & 1.6888 & -4.3937 \end{bmatrix}^\mathsf{T}.$$

Note that the above input-hidden matrix is initialized randomly and kept unchanged. Besides, the approximated values of g and h can be derived after the identification stage. Then, the critic network is constructed as follows:

$$\begin{aligned} \hat{J}(x) = \hat{\omega}_c^\mathsf{T}\sigma_c(x) = {} & \hat{\omega}_{c1}x_1^2 + \hat{\omega}_{c2}x_2^2 + \hat{\omega}_{c3}x_3^2 \\ & + \hat{\omega}_{c4}x_1x_2 + \hat{\omega}_{c5}x_1x_3 + \hat{\omega}_{c6}x_2x_3 \end{aligned} \quad (53)$$

where the weight $\hat{\omega}_c = [\hat{\omega}_{c1}, \hat{\omega}_{c2}, \hat{\omega}_{c3}, \hat{\omega}_{c4}, \hat{\omega}_{c5}, \hat{\omega}_{c6}]^\mathsf{T}$ and the activation function $\sigma_c(x) = [x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3]^\mathsf{T}$ with $l_c = 6$. In what follows, we will make effort to derive an applicable weight vector. It is worth mentioning that the number of hidden layer neurons is often determined by computer experiment. Actually, the choice of the activation function is more of an art than science, involving a tradeoff between control accuracy and computational complexity.

For adjusting the critic network, we set the learning rate of the critic network as $\alpha_c = 2$. For performing effective learning and approximation, we add a probing noise to guarantee the persistence of excitation condition within the first 550 s. The simulation result of the neural network learning stage in illustrated in Fig. 3(a). Therein, we find that the weight vector of the critic network gradually converges to $[0.1600, 0.2775, 0.0771, 0.2218, -0.1657, 0.1221]^\mathsf{T}$, which reflects the learning ability of the intelligent critic controller.

Next, we evaluate the H_∞ control performance by applying the obtained intelligent critic controller to the plant (52) for 60s and introducing an external perturbation $v(t) = e^{-0.2t}\cos(-0.6t), t > 0$. The simulation results of the H_∞ control implementation stage are exhibited in Fig. 3(b) and (c). Among them, the 3-D (three-dimensional) view of the system state trajectory is depicted in Fig. 3(b). Besides, in order to

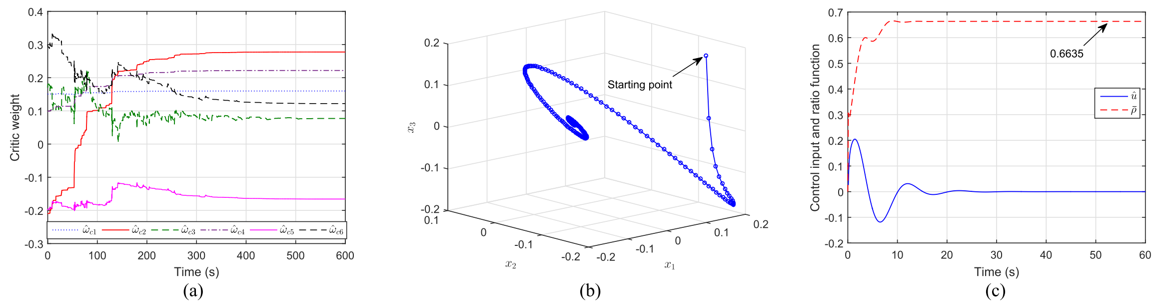

Fig. 3. (a) Convergence of the weight vector. (b) 3-D view of the system state curve. (c) Control input and ratio function.

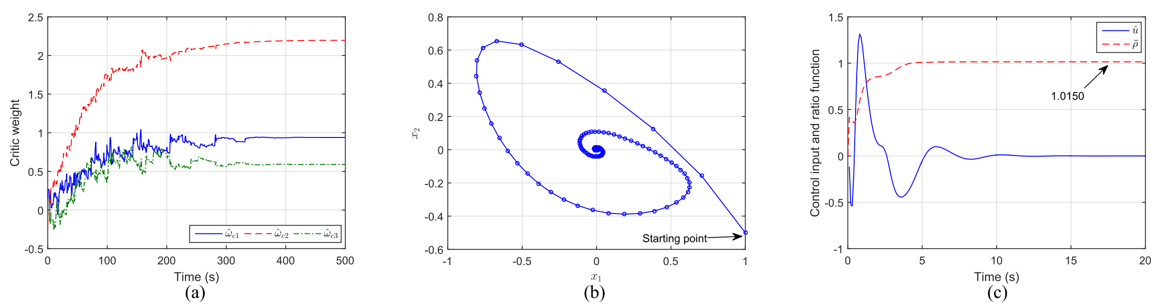

Fig. 4. (a) Convergence of the weight vector. (b) System state trajectory. (c) Control input and ratio function.

reflect the disturbance attenuation of the H_∞ control problem, we define a ratio function $\bar{\varrho}(t)$ as the form

$$\bar{\varrho}(t) = \sqrt{\int_0^t \left(x^{\mathsf{T}}(\tau)Qx(\tau) + u^{\mathsf{T}}(\tau)u(\tau)\right)\mathrm{d}\tau \Big/ \int_0^t \|v(\tau)\|^2 \mathrm{d}\tau}. \tag{54}$$

As time goes on, the approximate control law $\hat{u}(x(t))$ and the ratio $\bar{\varrho}(t)$ converge to 0 and 0.6635, respectively, which are both shown in Fig. 3(c). Since $\bar{\varrho}(t) \to 0.6635 < \varrho = 5$, we successfully observe a required $\mathcal{L}_2$-gain performance level for the closed-loop system. Consequently, the designed intelligent critic control law possesses an excellent ability of disturbance attenuation.

Remark 3: Note that during the simulation, the "Time (s)" marked in the figures is in fact the time steps. The system state should be persistently excited long enough so as to let the critic network acquire the optimal cost as accurately as possible. Hence, it requires sufficient time steps to perform the learning task. However, it is certainly not the actually elapsed time of the CPU. In this example, using the computer with the processor Intel Core i7-4790, the actual elapsed time of the learning and control processes is 14.9562 and 0.9934 s, respectively, rather than the time steps marked in Fig. 3(a) and (c). Of course, the elapsed time is related to the computer configuration. This fact is also true for the next example.

B. Simulation of a Nonlinear Dynamical Plant

Consider a continuous-time nonlinear system with input-affine structure and external disturbance given as follows:

$$\dot{x} = \begin{bmatrix} -x_1^3 - 2x_2 \\ x_1 + 0.5\cos x_1^2 \sin x_2^3 \end{bmatrix} + \begin{bmatrix} 1 \\ \sin x_1 \end{bmatrix} u + \begin{bmatrix} -1 \\ \cos x_2 \end{bmatrix} v \tag{55}$$

where $x = [x_1, x_2]^{\mathsf{T}} \in \mathbb{R}^2$, $u \in \mathbb{R}$, and $v \in \mathbb{R}$ are the state, control, and perturbation variables, respectively. Clearly, the bounded condition of the control and disturbance matrices is true due to the fact that $\sqrt{1+\sin^2 x_1} \le \sqrt{2}$ and $\sqrt{1+\cos^2 x_2} \le \sqrt{2}$. We consider the H_∞ control problem with an initial state $x_0 = [1, -0.5]^{\mathsf{T}}$ and choose $Q = I_2$ and $\varrho = 3$. In this simulation, we select six hidden neurons ($l_m = 6$) and set the learning rate as $\alpha_m = 0.8$. Then, the critic network is introduced with the structure

$$\hat{J}(x) = \hat{\omega}_c^{\mathsf{T}}\sigma_c(x) = \hat{\omega}_{c1}x_1^2 + \hat{\omega}_{c2}x_2^2 + \hat{\omega}_{c3}x_1x_2 \tag{56}$$

where $\hat{\omega}_c = [\hat{\omega}_{c1}, \hat{\omega}_{c2}, \hat{\omega}_{c3}]^{\mathsf{T}}$ and $\sigma_c(x) = [x_1^2, x_2^2, x_1x_2]^{\mathsf{T}}$ with $l_c = 3$. Note that the probing noise should also be brought into the implementation process to satisfy the persistence of excitation condition with the learning rate $\alpha_c = 3.5$. We can observe that the convergence of the weight vector occurred after 450 s. In addition, the convergence process of the weight vector to $[0.9390, 2.1959, 0.5902]^{\mathsf{T}}$ is depicted in Fig. 4(a).

At last, we apply the approximated control law to the controlled plant (55) for 20 s with an external perturbation $v(t) = 3e^{-t}\cos t, t > 0$ being employed and then obtain the system state trajectory is shown in Fig. 4(b). In addition, the adjustments of the control input and the ratio function are illustrated in Fig. 4(c), which reveals that the designed H_∞ feedback controller attains a prespecified $\mathcal{L}_2$-gain performance level for the closed-loop system (i.e., $\bar{\varrho}(t) \to 1.0150 < \varrho = 3$). The above simulation results substantiate the effectiveness of the intelligent critic control strategy with respect to the external disturbance.

V. Conclusion

The intelligent H_∞ control of continuous-time affine dynamic systems was investigated with adaptive critic framework. The approximate optimal control and worst-case disturbance laws were derived with the help of regulating an identifier and training a critic network with stability proof. The application to a power system and the simulation for a nonlinear system was presented as experimental verification. The general discussion on discrete-time systems is worth further studying in the future.

References

[1] M. Ruderman and M. Iwasaki, "Observer of nonlinear friction dynamics for motion control," *IEEE Trans. Ind. Electron.*, vol. 62, no. 9, pp. 5941–5949, Sep. 2015.

[2] W. Sun, Z. Zhao, and H. Gao, "Saturated adaptive robust control for active suspension systems," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3889–3896, Sep. 2013.

[3] W. He, S. Zhang and S. S. Ge, "Adaptive control of a flexible crane system with the boundary output constraint," *IEEE Trans. Ind. Electron.*, vol. 61, no. 8, pp. 4126–4133, Aug. 2014.

[4] L. Cheng, W. Liu, Z. G. Hou, J. Yu, and M. Tan, "Neural-network-based nonlinear model predictive control for piezoelectric actuators," *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7717–7727, Dec. 2015.

[5] B. Xu, "Robust adaptive neural control of flexible hypersonic flight vehicle with dead-zone input nonlinearity," *Nonlinear Dyn.*, vol. 80, no. 3, pp. 1509–1520, 2015.

[6] S. Jafarzadeh and M. S. Fadali, "On the stability and control of continuous-time TSK fuzzy systems," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 1073–1087, Jun. 2013.

[7] R. E. Precup, M. B. Radac, M. L. Tomescu, E. M. Petriu, and S. Preitl, "Stable and convergent iterative feedback tuning of fuzzy controllers for discrete-time SISO systems," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 188–199, 2013.

[8] A. Kahrobaeian and Y. A. R. I. Mohamed, "Analysis and mitigation of low-frequency instabilities in autonomous medium-voltage converter-based microgrids with dynamic loads," *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 1643–1658, Apr. 2014.

[9] A. Polyakov and L. Fridman, "Stability notions and Lyapunovfunctions for sliding mode control systems," *J. Franklin Inst.*, vol. 351, no. 4, pp. 1831–1865, 2014.

[10] M. Cucuzzella, G. P. Incremona, and A. Ferrara, "Design of robust higher order sliding mode control for microgrids," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 5, no. 3, pp. 393–401, Sep. 2015.

[11] A. Kahrobaeian and Y. A. R. I. Mohamed, "Suppression of interaction dynamics in DG converter-based microgrids via robust system-oriented control approach," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1800–1811, Dec. 2012.

[12] M. A. Mahmud, M. J. Hossain, H. R. Pota, and A. M. T. Oo, "Robust nonlinear distributed controller design for active and reactive power sharing in islanded microgrids," *IEEE Trans. Energy Convers.*, vol. 29, no. 4, pp. 893–903, Dec. 2014.

[13] Y. A. R. I. Mohamed, H. H. Zeineldin, M. M. A. Salama, and R. Seethapathy, "Seamless formation and robust control of distributed generation microgrids via direct voltage control and optimized dynamic power sharing," *IEEE Trans. Power Electron.*, vol. 27, no. 3, pp. 1283–1294, Mar. 2012.

[14] T. Basar and P. Bernhard, *H_∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, 2nd ed. Boston, MA, USA: Birkhauser, 2008.

[15] H. Gao, J. Wu, and P. Shi, "Robust sampled-data H_∞ control with stochastic sampling," *Automatica*, vol. 45, pp. 1729–1736, 2009.

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

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

[18] H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive Dynamic Programming for Control: Algorithms and Stability*. London, U.K.: Springer, 2013.

[19] T. Dierks, B. T. Thumati, and S. Jagannathan, "Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence," *Neural Netw.*, vol. 22, no. 5–6, pp. 851–860, 2009.

[20] A. Heydari and S. N. Balakrishnan, "Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 1, pp. 145–157, Jan. 2013.

[21] Q. Zhao, H. Xu, and S. Jagannathan, "Near optimal output feedback control of nonlinear discrete-time systems based on reinforcement neural network learning," *IEEE/CAA J. Automat. Sinica*, vol. 1, no. 4, pp. 372–384, Oct. 2014.

[22] K. G. Vamvoudakis and F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, May 2010.

[23] X. Yang, D. Liu, and D. Wang, "Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints," *Int. J. Control*, vol. 87, no. 3, pp. 553–566, 2014.

[24] Y. Jiang and Z. P. Jiang, "Global adaptive dynamic programming for continuous-time nonlinear systems," *IEEE Trans. Autom. Control*, vol. 60, no. 11, pp. 2917–2929, Nov. 2015.

[25] W. Gao and Z. P. Jiang, "Adaptive dynamic programming and adaptive optimal output regulation of linear systems," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 4164–4169, Dec. 2016.

[26] D. Wang, D. Liu, Q. Zhang, and D. Zhao, "Data-based adaptive critic designs for nonlinear robust optimal control with uncertain dynamics," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 11, pp. 1544–1555, Nov. 2016.

[27] J. Liang, G. K. Venayagamoorthy, and 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*, vol. 3, no. 1, pp. 59–69, Mar. 2012.

[28] C. Wang, D. Liu, Q. Wei, D. Zhao, and Z. Xia, "Iterative adaptive dynamic programming approach to power optimal control for smart grid with energy storage devices," *Acta Autom. Sinica*, vol. 40, no. 9, pp. 1984–1990, 2014.

[29] T. Bian, Y. Jiang, and Z. P. Jiang, "Decentralized adaptive optimal control of large-scale systems with application to power systems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2439–2447, Apr. 2015.

[30] Y. Tang, H. He, J. Wen, and J. Liu, "Power system stability control for a wind farm based on adaptive dynamic programming," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 166–177, Jan. 2015.

[31] J. Wang, X. Xu, D. Liu, Z. Sun, and Q. Chen, "Self-learning cruise control using Kernel-based least squares policy iteration," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 3, pp. 1078–1087, May 2014.

[32] M. Abu-Khalaf, F. L. Lewis, and J. Huang, "Policy iterations on the Hamilton-Jacobi-Isaacs equation for H_∞ state feedback control with input saturation," *IEEE Trans. Autom. Control*, vol. 51, no. 12, pp. 1989–1995, Dec. 2006.

[33] K. G. Vamvoudakis and F. L. Lewis, "Online solution of nonlinear two-player zero-sum games using synchronous policy iteration," *Int. J. Robust Nonlinear Control*, vol. 22, pp. 1460–1483, 2012.

[34] H. Modares, F. L. Lewis, and M. B. N. Sistani, "Online solution of non-quadratic two-player zero-sum games arising in the H_∞ control of constrained input systems," *Int. J. Adapt. Control Signal Process.*, vol. 28, pp. 232–254, 2014.

[35] B. Luo, H. N. Wu, and T. Huang, "Off-policy reinforcement learning for H_∞ control design," *IEEE Trans. Cybern.*, vol. 45, no. 1, pp. 65–76, Jan. 2015.

[36] C. Qin, H. Zhang, Y. Wang, and Y. Luo, "Neural network-based online H_∞ control for discrete-time affine nonlinear system using adaptive dynamic programming," *Neurocomputing*, vol. 198, pp. 91–99, 2016.

[37] H. Zhang, J. Zhang, G. H. Yang, and Y. Luo, "Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 1, pp. 152–163, Feb. 2015.

[38] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[39] M. Krstic, I. Kanellakopoulos, and P. Kokotovic, *Nonlinear and Adaptive Control Design*. New York, NY, USA: Wiley, 1995.

[40] M. J. Corless and G. Leitmann, "Continuous state feedback guaranteeing uniform ultimate boundedness for uncertain dynamic systems," *IEEE Trans. Autom. Control*, vol. AC-26, no. 5, pp. 1139–1144, Oct. 1981.

[41] E. Romero-Cadaval, G. Spagnuolo, L. G. Franquelo, C. A. Ramos-Paja, T. Suntio, and W. M. Xiao, "Grid-connected photovoltaic generation plants: Components and operation," *IEEE Ind. Electron. Mag.*, vol. 7, no. 3, pp. 6–20, Sep. 2013.

[42] S. K. Pandey, S. R. Mohanty, and N. Kishor, "A literature survey on load-frequency control for conventional and distribution generation power systems," *Renew. Sustain. Energy Rev.*, vol. 25, pp. 318–334, 2013.

[43] K. P. S. Parmar, S. Majhi, and D. P. Kothari, "Load frequency control of a realistic power system with multi-source power generation," *Int. J. Elect. Power Energy Syst.*, vol. 42, pp. 426–433, 2012.

[44] R. Francis and I. A. Chidambaram, "Optimized PI+ load-frequency controller using BWNN approach for an interconnected reheat power system with RFB and hydrogen electrolyser units," *Int. J. Elect. Power Energy Syst.*, vol. 67, pp. 381–392, 2015.

[45] I. Kamwa, R. Grondin, and Y. Hebert, "Wide-area measurement based stabilizing control of large power systems–A decentralized/hierarchical approach," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 136–153, Feb. 2001.

[46] C. Mu, Y. Tang, and H. He, "Observer-based sliding mode frequency control for micro-grid with photovoltaic energy integration," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Boston, MA, USA, Jul. 2016, pp. 1–5.

[47] S. Sonmez, S. Ayasun, and C. O. Nwankpa, "An exact method for computing delay margin for stability of load frequency control systems with constant communication delays," *IEEE Trans. Power Syst.*, vol. 31, no. 1, 370–377, Jan. 2016.

Ding Wang (M'15) received the B.S. degree in mathematics from Zhengzhou University of Light Industry, Zhengzhou, China, the M.S. degree in operations research and cybernetics from Northeastern University, Shenyang, China, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007, 2009, and 2012, respectively.

He was a Visiting Scholar in the Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI, USA, from December 2015 to January 2017. He is currently an Associate Professor in the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interests include adaptive and learning systems, computational intelligence, and intelligent control. He has published more than 80 journal and conference papers, and coauthored two monographs.

Dr. Wang is the Publications Chair of the 24th International Conference on Neural Information Processing (ICONIP 2017). He received the Excellent Doctoral Dissertation Award of the Chinese Academy of Sciences in 2013, and a nomination for the Excellent Doctoral Dissertation Award of the Chinese Association of Automation (CAA) in 2014. He serves as an Associate Editor for the IEEE Transactions on Neural Networks and Learning Systems and *Neurocomputing*. He is a Member of the Asia-Pacific Neural Network Society (APNNS) and CAA.

Haibo He (SM'11) received the B.S. and M.S. degrees from Huazhong University of Science and Technology, China, in 1999 and 2002, respectively, and the Ph.D. degree from Ohio University, Athens, OH, USA, in 2006, all in electrical engineering.

From 2006 to 2009, he was an Assistant Professor in the Department of Electrical and Computer Engineering, Stevens Institute of Technology. He is currently the Robert Haas Endowed Chair Professor in the Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI, USA. His research interests include adaptive dynamic programming, computational intelligence, machine learning and data mining, and various applications. He has published 1 sole-author research book (Wiley), edited 1 book (Wiley-IEEE) and 6 conference proceedings (Springer), and authored or coauthored more than 200 peer-reviewed journal and conference papers.

Prof. He served as the General Chair of the IEEE Symposium Series on Computational Intelligence (SSCI 2014). He received the IEEE International Conference on Communications Best Paper Award (2014), IEEE Computational Intelligence Society (CIS) Outstanding Early Career Award (2014), National Science Foundation (NSF) CAREER Award (2011), and Providence Business News (PBN) "Rising Star Innovator Award" (2011). He is currently the Editor-in-Chief of the IEEE Transactions on Neural Networks and Learning Systems.

Chaoxu Mu (M'15) received the B.S. degree from Harbin Institute of Technology, Harbin, China, in 2006, the M.S. degree from Hohai University, Nanjing, China, in 2009, and the Ph.D. degree from Southeast University, Nanjing, China, in 2012, all in control science and engineering.

She was a Visiting Ph.D. Student with the Royal Melbourne Institute of Technology University, Melbourne, Australia, from October 2010 to November 2011. She was a Post-Doctoral Fellow in the Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI, USA, from December 2014 to August 2016. She is currently an Associate Professor in the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. She has authored more than 50 journal and conference papers, and coauthored one monograph. Her current research interests include nonlinear system control and optimization, and adaptive and learning systems.

Dr. Mu is a member of the Asia-Pacific Neural Network Society.

Derong Liu (S'91–M'94–SM'96–F'05) received the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1994.

He was a Staff Fellow with the General Motors Research and Development Center from 1993 to 1995. He was an Assistant Professor in 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 he served as the Associate Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, from 2010 to 2015. He is currently a Full Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. He has published 18 books.

Dr. Liu is the Editor-in-Chief of *Artificial Intelligence Review*. He was the Editor-in-Chief of the IEEE Transactions on Neural Networks and Learning Systems from 2010 to 2015. He received the Faculty Early Career Development Award from the National Science Foundation in 1999, the University Scholar Award from the University of Illinois from 2006 to 2009, the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008, and the Outstanding Achievement Award from the Asia Pacific Neural Network Assembly in 2014. He is a Fellow of the International Neural Network Society and of the International Association of Pattern Recognition.