International Journal of Control, 2013
Vol. 86, No. 9, 1554–1566, http://dx.doi.org/10.1080/00207179.2013.790562

Taylor & Francis
Taylor & Francis Group

Neural-network-observer-based optimal control for unknown nonlinear systems using adaptive dynamic programming

Derong Liu*, Yuzhu Huang, Ding Wang and Qinglai Wei

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

(*Received 13 October 2012; final version received 25 March 2013*)

In this paper, an observer-based optimal control scheme is developed for unknown nonlinear systems using adaptive dynamic programming (ADP) algorithm. First, a neural-network (NN) observer is designed to estimate system states. Then, based on the observed states, a neuro-controller is constructed via ADP method to obtain the optimal control. In this design, two NN structures are used: a three-layer NN is used to construct the observer which can be applied to systems with higher degrees of nonlinearity and without a priori knowledge of system dynamics, and a critic NN is employed to approximate the value function. The optimal control law is computed using the critic NN and the observer NN. Uniform ultimate boundedness of the closed-loop system is guaranteed. The actor, critic, and observer structures are all implemented in real-time, continuously and simultaneously. Finally, simulation results are presented to demonstrate the effectiveness of the proposed control scheme.

Keywords: nonlinear observer; adaptive dynamic programming; neural network; uniformly ultimately bounded; nonlinear system

1. Introduction

As is well known, various control schemes have been developed in the literature for optimal control based on full state measurements. However, in most real cases, the state variables are unavailable for direct online measurements, and merely input and output of the system are measurable. Therefore, estimating the state variables by observers plays an important role in the control of processes to achieve better performances. During the past several decades, many nonlinear observers have been developed to obtain the estimated states. However, these conventional nonlinear observers, such as high-gain observers, and sliding mode observers (Farza, Sboui, Cherrier, & M'Saad, 2010; Jo & Seo, 2002; Jung, Huh, & Lee, 2008; Nicosia, Tomei, & Tornambe, 1989; Slotine & Li, 1991) are only applicable to systems with specific model structures. Furthermore, most of them rely on completely knowing the system nonlinearities a priori. Note that, for most practical processes, obtaining an exact model is a difficult task or is not possible at all.

Moreover, in recent years, neural-network (NN) techniques have shown a good promise as competitive methods for nonlinear control, signal processing, and other applications. The capability of NN for identification, observation, and control of nonlinear systems has been investigated in online and offline environments (Chen & Khalil, 1995; Michael & Harley, 1995; Narendra & Parthasarathy, 1990; Park, Huh, Kim, & Seo, 2005; Yu, 2009). In fact, due to the properties of nonlinearity, adaptivity, self-learning, fault tolerance, and advanced input–output mapping (Igelnik & Pao, 1995; Jagannathan, 2006; Lewis, Jagannathan, & Yesildirek, 1999), NNs show powerful potentials in solving the nonlinear state observation problems without a priori knowledge of system dynamics. In Ahmed and Riyaz (2002), a general multiple-input-multiple-output (MIMO) nonlinear system was linearised and an extended Kalman filter was used to estimate the system states. The gain of the proposed observer was computed by a multilayer feedforward NN. In Selmic and Lewis (2001), multimodel identification and failure detection using radial basis function were presented, where one tuneable layer NN was considered and the persistency of excitation condition was developed to guarantee the convergence of the parameters of the identifier to the ideal parameters. In Abdollahi, Talebi, and Patel (2006), an NN-based observer for nonlinear systems was proposed by using a backpropagation algorithm with a modification term.

In this paper, inspired by Abdollahi et al. (2006), a multilayer feedforward NN observer for unknown nonlinear systems is developed, where the observer NN is used to parameterise the nonlinearities of the system and trained using the error backpropagation algorithm. In the following, after obtaining the observed states, it is necessary to derive the optimal control of the nonlinear system based on

*Corresponding author. Email: derong.liu@ia.ac.cn

the observed states. In the optimal control field, dynamic programming (DP) has been a useful computational technique in solving optimal control problems for many years. However, due to the backward numerical process required for its solutions, i.e., the well-known 'curse of dimensionality' (Bellman, 1957; Lewis & Syrmos, 1995; Wang, Zhang, & Liu, 2009), it is often computationally untenable to run DP to obtain the optimal solution.

By means of constructing a module called 'critic' to approximate the cost function in DP, adaptive dynamic programming (ADP) successfully avoids the 'curse of dimensionality'. Therefore, in recent years, ADP has attracted much attention from researchers (Bertsekas & Tsitsiklis, 1996; Dierks, Thumati, & Jagannathan, 2009; He & Jagannathan, 2007; Lewis & Vrabie, 2009; Liu, Xiong, & Zhang, 2001; Liu, Zhang, & Zhang, 2005; Murray, Cox, Lendaris, & Saeks, 2002; Si, Barto, Powell, & Wunsch, 2004; Wang, Liu, & Wei, 2012; Zhang, Luo, & Liu, 2009; Zhang, Wei, & Luo, 2008). ADP was proposed in Werbos (1977) and Werbos (1992), as a way to solve optimal control problems forward in time. In Prokhorov and Wunsch (1997), ADP approaches were classified into several main schemes including heuristic dynamic programming (HDP), action-dependent heuristic dynamic programming (ADHDP), dual heuristic dynamic programming (DHP), ADDHP, globalised DHP (GDHP), and ADGDHP. In Al-Tamimi, Lewis, and Abu-Khalaf (2008), a greedy iterative HDP was proposed to solve the optimal control problem for nonlinear discrete-time systems. Vrabie and Lewis (2009) studied the continuous-time optimal control problem using ADP. Wang, Jin, Liu, and Wei (2011) developed an ε-ADP algorithm for studying finite-horizon optimal control of discrete-time nonlinear systems.

Taking account of practical application conditions, a novel control scheme is developed for unknown nonlinear systems based on the ADP algorithm and NN observer in this paper. First, an NN observer is designed to estimate system states. Then, based on the observed states, a feedforward neuro-controller is constructed using ADP algorithm to obtain the optimal control. Moreover, uniformly ultimately bounded (UUB) stability of the closed-loop system is guaranteed. The actor, critic, and observer structures are implemented in real-time, continuously and simultaneously.

The rest of this paper is organised as follows. In Section 2, the problem formulation is presented. In Section 3, by using a multilayer feedforward NN, an observer is designed for the unknown nonlinear system. Moreover, the Lyapunov approach is used to show that state estimation errors and weight estimation errors are all bounded. In Section 4, a feedforward neuro-controller is constructed by using ADP algorithm based on the observed states. Meanwhile, the boundedness of all signals in the closed-loop observer and controller is shown. In Section 5, simulation results are presented to demonstrate the effectiveness of the proposed optimal control scheme. Several conclusions are drawn in Section 6.

2. Problem formulation

Consider the nonlinear continuous-time system described by

$$\begin{aligned} \dot{x}(t) &= F(x(t), u(t)), \\ y(t) &= Cx(t), \end{aligned} \tag{1}$$

where $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^{\mathsf{T}} \in \mathbb{R}^n$ is the state vector, $u(t) = [u_1(t), u_2(t), \ldots, u_m(t)]^{\mathsf{T}} \in \mathbb{R}^m$ is the control input vector, $y(t) = [y_1(t), y_2(t), \ldots, y_l(t)]^{\mathsf{T}} \in \mathbb{R}^l$ is the output vector, and $F(x, u)$ is an unknown continuous nonlinear smooth function with respect to $x(t)$ and $u(t)$. Moreover, it is assumed that system (1) is observable and system states are bounded in L_∞ (Abdollahi et al., 2006). This is a common assumption in identification schemes.

For optimal output regulator problems, the control objective is to design an optimal controller for system (1) which minimises the generalised infinite-horizon cost function

$$V(x(t), t) = \int_t^\infty (y^{\mathsf{T}}(\tau)Qy(\tau) + u^{\mathsf{T}}(\tau)Ru(\tau))\mathrm{d}\tau, \tag{2}$$

where t is the initial time, Q and R are symmetric positive definite matrices with appropriate dimensions. Noticing that $y(t) = Cx(t)$, (2) can be rewritten as

$$V(x(t), t) = \int_t^\infty r(x(\tau), u(\tau))\mathrm{d}\tau, \tag{3}$$

where $r(x(\tau), u(\tau)) = x^{\mathsf{T}}(\tau)Q_c x(\tau) + u^{\mathsf{T}}(\tau)Ru(\tau)$ with $Q_c = C^{\mathsf{T}}QC$, and Q_c is symmetric semi-definite due to the observability of system (1). Meanwhile, for optimal control problems, it should be noted that the designed feedback control $u(x)$ must not only stabilise system (1) but also guarantee that (3) is finite, i.e., the control must be admissible (Abu-Khalaf & Lewis, 2005).

Definition 2.1: A control law $u(x)$ is defined to be admissible with respect to (3) on a compact set Ω, denoted by $u \in \Psi(\Omega)$, if $u(x)$ is continuous on Ω, $u(0) = 0$, u stabilises system (1) on Ω, and $V(x(t))$ in (3) is finite.

Since the knowledge of system dynamics is completely unknown and system states are not available, we cannot apply existing ADP methods to system (1) directly. Therefore, it is desirable to design a novel control scheme that does not need the exact knowledge of system dynamics but only the input and output data measured during the operation of the system. In this paper, we develop an NN-observer-based optimal control scheme for unknown nonlinear continuous-time systems using ADP algorithms. In detail, the design of

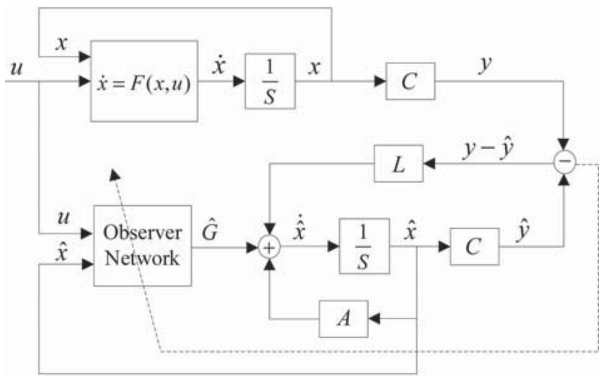

Figure 1. The structure diagram of the NN observer.

proposed controller is divided into two steps: (1) establish a multilayer feedforward NN observer for unknown nonlinear systems by using the measured input and output data of the system and (2) based on observed states, we design an optimal neuro-controller using ADP algorithms.

3. Neural-network-observer design

In this section, a multilayer feedforward NN observer is designed to obtain estimated states. Specially, the feedforward NN is used to parameterise the nonlinearities of the system and trained using the error backpropagation algorithm. Moreover, the observer error dynamics are described for analysing the stability of the NN observer.

Considering system (1), by adding and subtracting Ax, we have

$$\begin{aligned} \dot{x}(t) &= Ax + G(x,u), \\ y(t) &= Cx(t), \end{aligned} \tag{4}$$

where A is a Hurwitz matrix, the pair (C, A) is observable, and $G(x, u) = F(x, u) - Ax$. Now, the state observer for system (1) can be described by

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + \hat{G}(\hat{x},u) + L(y - C\hat{x}), \\ \hat{y}(t) &= C\hat{x}(t), \end{aligned} \tag{5}$$

where $\hat{x}$ and $\hat{y}$ denote the state and output of the observer, respectively, and the observer gain $L \in \mathbb{R}^{n\times l}$ is selected such that $A - LC$ is a Hurwitz matrix. Since A is selected such that (C, A) is observable, it ensures that L exists.

The key to designing an NN observer is to employ an NN to identify the nonlinearity and a conventional observer to estimate the states (Abdollahi et al., 2006; Ahmed & Riyaz, 2000; Selmic & Lewis, 2001). The structure of the designed NN observer is shown in Figure 1.

As is well known, a three-layer NN with a single-hidden layer is sufficient to approximate nonlinear systems with any degree of nonlinearity. Here, the function approximation capability of NNs is used. In fact, it has been shown by many researchers that for restricted to a compact set Ω of $x \in \mathbb{R}^n$ and for some sufficiently large number of hidden layer neurons, there exist weights and thresholds such that any continuous function has an NN representation on the compact set Ω (Igelnik & Pao, 1995; Jagannathan, 2006; Lewis et al., 1999; Yu, 2009). Thus, according to the universal approximation property of NNs, $G(x, u)$ can be represented as

$$G(x,u) = W\sigma(V\bar{x}) + \varepsilon(x), \tag{6}$$

where $W \in \mathbb{R}^{n\times k}$ and $V \in \mathbb{R}^{k\times(n+m)}$ are the ideal weight matrices of the output and hidden layers, k is the number of hidden layer neurons, $\bar{x} = [x^{\mathsf{T}}, u^{\mathsf{T}}]^{\mathsf{T}}$ is the NN input, and $\varepsilon(x)$ is the bounded NN functional approximation error, i.e., $\|\varepsilon(x)\| \le \varepsilon_M$, $\sigma(\cdot)$ is the NN activation function and selected to be a hyperbolic tangent function. Besides, NN activation functions are also bounded such that $\|\sigma(\cdot)\| \le \sigma_M$ for a positive constant σ_M.

It is assumed that the upper bounds of the fixed ideal weights W and V exist such that

$$\|W\|_F \le W_M, \ \|V\|_F \le V_M. \tag{7}$$

Then, $G(x, u)$ can be approximated by

$$\hat{G}(\hat{x},u) = \hat{W}\sigma(\hat{V}\hat{\bar{x}}), \tag{8}$$

where $\hat{x}$ is the estimated state vector, $\hat{\bar{x}} = [\hat{x}^{\mathsf{T}}, u^{\mathsf{T}}]^{\mathsf{T}}$, $\hat{W}$ and $\hat{V}$ are the corresponding estimates of the ideal weight matrices.

Therefore, the dynamics of the NN observer are given by

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x} + \hat{W}\sigma(\hat{V}\hat{\bar{x}}) + L(y - C\hat{x}), \\ \hat{y}(t) &= C\hat{x}(t). \end{aligned} \tag{9}$$

Let the state and output estimation errors be defined as $\tilde{x} = x - \hat{x}$ and $\tilde{y} = y - \hat{y}$, respectively. Then, considering (6) and subtracting (9) from (4), the error dynamics can be expressed as

$$\begin{aligned} \dot{\tilde{x}}(t) &= Ax + W\sigma(V\bar{x}) - A\hat{x} - \hat{W}\sigma(\hat{V}\hat{\bar{x}}), \\ &\quad - L(Cx - C\hat{x}) + \varepsilon(x), \\ \tilde{y}(t) &= C\tilde{x}(t). \end{aligned} \tag{10}$$

Considering (10), by adding and subtracting $W\sigma(\hat{V}\hat{\bar{x}})$, we obtain

$$\begin{aligned} \dot{\tilde{x}}(t) &= A_c\tilde{x} + \tilde{W}\sigma(\hat{V}\hat{\bar{x}}) + \zeta(t), \\ \tilde{y}(t) &= C\tilde{x}(t), \end{aligned} \tag{11}$$

where $\tilde{W} = W - \hat{W}$ and $A_c = A - LC$, and $\zeta(t) = W[\sigma(V\bar{x}) - \sigma(\hat{V}\hat{\bar{x}})] + \varepsilon(x)$ is a bounded disturbance term, i.e., $\|\zeta(t)\| \leq \zeta_M$ for some positive constant, due to the boundedness of the hyperbolic tangent function, the NN approximation error $\varepsilon(x)$, and ideal NN weights (W, V).

In order to guarantee the stability of the NN observer, a suitable tuning algorithm should be provided for the NN weights in the design. In this paper, inspired by Abdollahi et al. (2006), we design the weight tuning algorithm based on the error backpropagation algorithm plus some modification terms to guarantee the stability of the state observer and the NN weight errors, as detailed in the following theorem.

Theorem 3.1: *Consider system (1) and the observer dynamics (9). If the modified NN weight tuning algorithm with modification terms is provided as*

$$\begin{aligned}\dot{\hat{W}} &= -\eta_1(\tilde{y}^{\mathsf{T}} C A_c^{-1})^{\mathsf{T}} \sigma^{\mathsf{T}}(\hat{V}\hat{\bar{x}}) - \theta_1\|\tilde{y}\|\hat{W},\\ \dot{\hat{V}} &= -\eta_2(\tilde{y}^{\mathsf{T}} C A_c^{-1}\hat{W}(I - \Gamma(\hat{V}\hat{\bar{x}})))^{\mathsf{T}}\mathrm{sgn}^{\mathsf{T}}(\hat{\bar{x}}) - \theta_2\|\tilde{y}\|\hat{V},\end{aligned} \tag{12}$$

where $\Gamma(\hat{V}\hat{\bar{x}}) = \mathrm{diag}\{\sigma_i^2(\hat{V}\hat{\bar{x}})\}$, *$i = 1, 2, \dots, m$, and* $\mathrm{sgn}(\hat{\bar{x}}) = [\mathrm{sgn}(\hat{\bar{x}}_1), \mathrm{sgn}(\hat{\bar{x}}_2), \dots, \mathrm{sgn}(\hat{\bar{x}}_{n+m})]^{\mathsf{T}}$ *with*

$$\mathrm{sgn}(\hat{\bar{x}}_j) = \begin{cases} 1, & \text{for} \hat{\bar{x}}_j > 0\\ 0, & \text{for} \hat{\bar{x}}_j = 0,\\ -1, & \text{for} \hat{\bar{x}}_j < 0\end{cases} \tag{13}$$

$j = 1, 2, \dots, n + m$, and η_1, $\eta_2 > 0$ are the learning rates, θ_1,θ_2 are the designed positive numbers, then the state estimation error $\tilde{x}$ and weight estimation errors $\tilde{W} = W - \hat{W}$ and $\tilde{V} = V - \hat{V}$ are UUB.

Proof: Consider Lyapunov function candidate

$$J_o = \frac{1}{2}\tilde{x}^{\mathsf{T}} P\tilde{x} + \frac{1}{2}\mathrm{tr}\{\tilde{W}^{\mathsf{T}}\tilde{W}\} + \frac{1}{2}\mathrm{tr}\{\tilde{V}^{\mathsf{T}}\tilde{V}\}, \tag{14}$$

where P is a positive definite matrix that satisfies

$$A_c^{\mathsf{T}} P + PA_c = -\Lambda \tag{15}$$

for the Hurwitz matrix A_c and some positive definite matrix Λ. The time derivative of the Lyapunov function candidate is given by

$$\dot{J}_o = \frac{1}{2}\dot{\tilde{x}}^{\mathsf{T}} P\tilde{x} + \frac{1}{2}\tilde{x}^{\mathsf{T}} P\dot{\tilde{x}} + \mathrm{tr}(\tilde{W}^{\mathsf{T}}\dot{\tilde{W}}) + \mathrm{tr}(\tilde{V}^{\mathsf{T}}\dot{\tilde{V}}). \tag{16}$$

Using Equation (12), we obtain

$$\begin{aligned}\dot{\tilde{W}} &= \eta_1(\tilde{y}^{\mathsf{T}} C A_c^{-1})^{\mathsf{T}} \sigma^{\mathsf{T}}(\hat{V}\hat{\bar{x}}) + \theta_1\|\tilde{y}\|\hat{W},\\ \dot{\tilde{V}} &= \eta_2(\tilde{y}^{\mathsf{T}} C A_c^{-1}\hat{W}(I - \Gamma(\hat{V}\hat{\bar{x}})))^{\mathsf{T}}\mathrm{sgn}^{\mathsf{T}}(\hat{\bar{x}}) + \theta_2\|\tilde{y}\|\hat{V}.\end{aligned} \tag{17}$$

Substituting (11), (15), and (17) into (16), we have

$$\begin{aligned}\dot{J}_o = &-\frac{1}{2}\tilde{x}^{\mathsf{T}}\Lambda\tilde{x} + \tilde{x}P(\tilde{W}\sigma(\hat{V}\hat{\bar{x}}) + \zeta)\\ &+ \mathrm{tr}(\tilde{W}^{\mathsf{T}}\eta_1(\tilde{y}^{\mathsf{T}} C A_c^{-1})^{\mathsf{T}}\sigma^{\mathsf{T}}(\hat{V}\hat{\bar{x}}) + \tilde{W}^{\mathsf{T}}\theta_1\|\tilde{y}\|\hat{W})\\ &+ \mathrm{tr}(\tilde{V}^{\mathsf{T}}\eta_2(\tilde{y}^{\mathsf{T}} C A_c^{-1}\hat{W}(I - \Gamma(\hat{V}\hat{\bar{x}})))^{\mathsf{T}}\mathrm{sgn}^{\mathsf{T}}(\hat{\bar{x}})\\ &+ \tilde{V}^{\mathsf{T}}\theta_2\|\tilde{y}\|\hat{V}).\end{aligned} \tag{18}$$

By using some polynomial adjustments and (11), Equation (18) can be rewritten as

$$\begin{aligned}\dot{J}_o = &-\frac{1}{2}\tilde{x}^{\mathsf{T}}\Lambda\tilde{x} + \tilde{x}P(\tilde{W}\sigma(\hat{V}\hat{\bar{x}}) + \zeta)\\ &+ \mathrm{tr}(\tilde{W}^{\mathsf{T}} l_1\tilde{x}\sigma^{\mathsf{T}}(\hat{V}\hat{\bar{x}}) + \tilde{W}^{\mathsf{T}}\theta_1\|C\tilde{x}\|(W - \tilde{W}))\\ &+ \mathrm{tr}(\tilde{V}^{\mathsf{T}}(I - \Gamma(\hat{V}\hat{\bar{x}}))^{\mathsf{T}}\hat{W} l_2\tilde{x}\mathrm{sgn}^{\mathsf{T}}(\hat{\bar{x}})\\ &+ \tilde{V}^{\mathsf{T}}\theta_2\|C\tilde{x}\|(V - \tilde{V})),\end{aligned} \tag{19}$$

where $l_1 = \eta_1 A_c^{-\mathsf{T}} C^{\mathsf{T}} C$ and $l_2 = \eta_2 A_c^{-\mathsf{T}} C^{\mathsf{T}} C$. Before proceeding, we provide the following inequalities:

$$\begin{aligned}&\mathrm{tr}(\tilde{W}^{\mathsf{T}}(W - \tilde{W})) \leq W_M\|\tilde{W}\| - \|\tilde{W}\|^2,\\ &\mathrm{tr}(\tilde{V}^{\mathsf{T}}(V - \tilde{V})) \leq V_M\|\tilde{V}\| - \|\tilde{V}\|^2,\\ &\mathrm{tr}(\tilde{W}^{\mathsf{T}} l_1\tilde{x}\sigma^{\mathsf{T}}(\hat{V}\hat{\bar{x}})) \leq \sigma_M\|\tilde{W}\|\|l_1\|\|\tilde{x}\|.\end{aligned} \tag{20}$$

Note that the last inequality in (20) is obtained based on the fact that, for two matrices A and B, the following relationship holds:

$$\mathrm{tr}(A^{\mathsf{T}} B) = \mathrm{tr}(B^{\mathsf{T}} A). \tag{21}$$

On the other hand, by $\|\hat{W}\| = \|W - \tilde{W}\| \leq W_M + \|\tilde{W}\|$, $1 - \sigma_M^2 \leq 1$, and (21), we obtain

$$\begin{aligned}&\mathrm{tr}(\tilde{V}^{\mathsf{T}}(I - \Gamma(\hat{V}\hat{\bar{x}}))^{\mathsf{T}}\hat{W} l_2\tilde{x}\mathrm{sgn}^{\mathsf{T}}(\hat{\bar{x}}))\\ &\leq \|\tilde{V}\|(W_M + \|\tilde{W}\|)\|l_2\|\|\tilde{x}\|.\end{aligned} \tag{22}$$

Then, from (20) and (22), we have

$$\begin{aligned}\dot{J}_o \leq &-\frac{1}{2}\lambda_{\min}(\Lambda)\|\tilde{x}\|^2 + \|\tilde{x}\|\|P\|(\|\tilde{W}\|\sigma_M + \zeta_M)\\ &+ \sigma_M\|\tilde{W}\|\|l_1\|\|\tilde{x}\| + (W_M\|\tilde{W}\| - \|\tilde{W}\|^2)\theta_1\|C\|\|\tilde{x}\|\\ &+ \|\tilde{V}\|\|l_2\|(W_M + \|\tilde{W}\|)\|\tilde{x}\|\\ &+ \theta_2\|C\|\|\tilde{x}\|(V_M\|\tilde{V}\| - \|\tilde{V}\|^2),\end{aligned} \tag{23}$$

where $\lambda_{\min}(\Lambda)$ is the minimum eigenvalue of Λ.

Next, let $K_1 = \|l_2\|/2$, then, by adding and subtracting $K_1^2\|\tilde{W}\|^2\|\tilde{x}\|$ and $\|\tilde{V}\|^2\|\tilde{x}\|$ to the right-hand side of (23), we obtain

$$\begin{aligned}\dot{J}_o \leq &-\frac{1}{2}\lambda_{\min}(\Lambda)\|\tilde{x}\|^2 + \big[\|P\|\zeta_M - (\theta_1\|C\| - K_1^2)\|\tilde{W}\|^2\\ &+ (\|P\|\sigma_M + \sigma_M\|l_1\| + \theta_1\|C\|W_M)\|\tilde{W}\|\end{aligned}$$

$$+(\theta_2\|C\|V_M+\|l_2\|W_M)\|\tilde{V}\|$$
$$-(\theta_2\|C\|-1)\|\tilde{V}\|^2-(K_1\|\tilde{W}\|-\|\tilde{V}\|)^2\big]\|\tilde{x}\|. \quad (24)$$

Denote K_2 and K_3 as

$$K_2=\frac{\|P\|\sigma_M+\sigma_M\|l_1\|+\theta_1\|C\|W_M}{2(\theta_1\|C\|-K_1^2)}$$
$$K_3=\frac{\theta_2\|C\|V_M+\|l_2\|W_M}{2(\theta_2\|C\|-1)}. \quad (25)$$

Then, in order to complete the squares for the terms $\|\tilde{W}\|$ and $\|\tilde{V}\|$, the terms $K_2^2\|\tilde{x}\|$ and $K_3^2\|\tilde{x}\|$ are added to and subtracted from (24), and we have

$$\begin{aligned}\dot{J}_o\le&-\frac{1}{2}\lambda_{\min}(\Lambda)\|\tilde{x}\|^2+\big[\|P\|\bar{\omega}+(\theta_1\|C\|-K_1^2)K_2^2\\&+(\theta_2\|C\|-1)K_3^2-(\theta_1\|C\|-K_1^2)(K_2-\|\tilde{W}\|)^2\\&-(\theta_2\|C\|-1)(K_3-\|\tilde{V}\|)^2-(K_1\|\tilde{W}\|-\|\tilde{V}\|)^2\big]\|\tilde{x}\|.\end{aligned} \quad (26)$$

Select $\theta_1\ge K_1^2/\|C\|$ and $\theta_2\ge 1/\|C\|$. Then, (26) becomes

$$\begin{aligned}\dot{J}_o\le&-\frac{1}{2}\lambda_{\min}(\Lambda)\|\tilde{x}\|^2+\|\tilde{x}\|\big(\|P\|\bar{\omega}+(\theta_1\|C\|-K_1^2)K_2^2\\&+(\theta_2\|C\|-1)K_3^2\big).\end{aligned} \quad (27)$$

Therefore, for guaranteeing the negativeness of $\dot{J}_o$, the following condition on $\|\tilde{x}\|$ should hold, i.e.,

$$\begin{aligned}\|\tilde{x}\|>&\frac{2\left(\|P\|\bar{\omega}+\left(\theta_1\|C\|-K_1^2\right)K_2^2+\left(\theta_2\|C\|-1\right)K_3^2\right)}{\lambda_{\min}(\Lambda)}\\=&\,d.\end{aligned} \quad (28)$$

Furthermore, according to the standard Lyapunov theorem (Lewis & Syrmos, 1995), as long as (28) is satisfied, we can demonstrate that the observation error $\tilde{x}$ and the weight estimation errors $\tilde{W}$ and $\tilde{V}$ are UUB. □

Remark 1: $\dot{J}_o$ is negative definite outside the ball with radius d described as $X=\{\tilde{x}\,|\,\|\tilde{x}\|>d\}$, and $\tilde{x}$ is UUB. The size of the estimation error bound d can be kept arbitrarily small by proper selection of the design parameters θ_1, θ_2, and the learning rates η_1, η_2 such that the higher accuracy can be achieved.

Remark 2: The explanation about selecting an NN observer rather than system identification technique is given here. In control engineering, a common approach is to start from measurements of the behaviour of the system and the external influences (inputs to the system) and try to determine a mathematical relation between them without going into the details of what is actually happening inside the system (Goodwin & Payne, 1977; Walter & Pronzato, 1997). This approach is called system identification. Therefore, we can conclude that based on system identification, we are generally able to obtain a 'black box' model of the nonlinear system (Jin, Sain, Pham, Spencer, & Ramallo, 2001), but do not obtain any in-depth knowledge about system states because they are the internal properties of the system. In most real cases, the state variables are unavailable for direct online measurements, and merely input and output of the system are measurable. Therefore, estimating the state variables by observers plays an important role in the control of processes to achieve better performances. Once obtaining the estimated states, we can design a state feedback controller to achieve the optimisation of system performance directly (Theocharis & Petridis, 1994). In conclusion, we choose an NN observer rather than system identification technique in this paper.

4. Optimal neuro-controller design based on ADP

In this section, based on observed states, a neuro-controller is developed for obtaining optimal control using ADP algorithms. Moreover, all signals in the closed-loop observer and controller are proved to be UUB.

By (1) and (3), the Hamiltonian function can be defined as

$$H(x,u,V_x)=r(x(t),u(t))+V_x^{\mathsf{T}}F(x(t),u(t)), \quad (29)$$

where $V_x=\partial V(x)/\partial x$. The optimal cost function $V^*(x)$ is defined as

$$V^*(x)=\min_{u\in\Psi(\Omega)}\left(\int_t^\infty r(x(\tau),u(\tau))\mathrm{d}\tau\right) \quad (30)$$

and satisfies the HJB equation

$$\min_{u\in\Psi(\Omega)}\big[H(x,u,V_x^*)\big]=0, \quad (31)$$

where $V_x^*=\partial V^*(x)/\partial x$. Assume that the minimum on the right-hand side of (31) exists and is unique. Then, by solving $\partial H(x,u,V_x)/\partial u=0$, the optimal control can be obtained as

$$u^*=-\frac{1}{2}R^{-1}\left(\frac{\partial F(x,u)}{\partial u}\right)^{\mathsf{T}}V_x^*. \quad (32)$$

Substituting (32) into (31), we obtain

$$\begin{aligned}0=&\,x^{\mathsf{T}}Q_cx+\frac{1}{4}V_x^{*\mathsf{T}}\frac{\partial F(x,u)}{\partial u}R^{-1}\left(\frac{\partial F(x,u)}{\partial u}\right)^{\mathsf{T}}V_x^*\\&+V_x^{*\mathsf{T}}F\left(x,-\frac{1}{2}R^{-1}\left(\frac{\partial F(x,u)}{\partial u}\right)^{\mathsf{T}}V_x^*\right).\end{aligned} \quad (33)$$

Note that, in order to find the optimal control solution of the problem, we only need to solve (33) for the cost

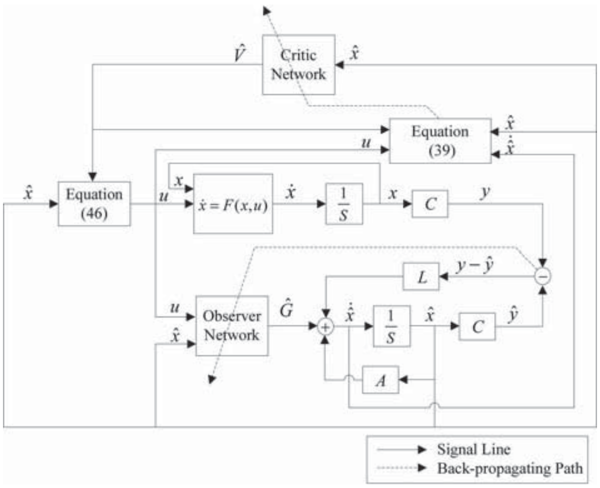

Figure 2. The structure diagram of the NN-observer-based controller.

function and then substitute the solution in (32) to obtain the optimal control. However, due to the nonlinear nature of the HJB equation, finding its solution is generally difficult or impossible.

Therefore, based on the designed observer, a neuro-controller is developed by using ADP methods. The structure diagram of the NN-observer-based controller is shown in Figure 2.

In the following, we focus on the optimal feedback controller design by using the ADP method, which is implemented by employing a critic NN to approximate the cost function. According to the universal approximation property of NNs, the cost function $V(\hat{x})$ can be represented by the critic NN as

$$V(\hat{x}) = W_c^{\mathsf{T}} \sigma_c \left(V_c^{\mathsf{T}} \hat{x} \right) + \varepsilon_c(\hat{x}), \tag{34}$$

where $W_c \in \mathbb{R}^{k_c \times 1}$ and $V_c \in \mathbb{R}^{n \times k_c}$ are the ideal weight matrices of the output and hidden layer, k_c is the number of hidden layer neurons, and ε_c is the bounded NN functional approximation error. In our design, based on Igelnik and Pao (1995), for both simplicity of learning and efficiency of approximation, the output layer weight matrix W_c is adapted online, whereas the input layer weight matrix V_c is selected initially at random and held fixed during the entire learning process. It is demonstrated in Igelnik and Pao (1995) that if the number of hidden layer neurons k_c is sufficiently large, the NN approximation error ε_c can be made arbitrarily small.

For the critic NN, its output can be expressed as

$$\hat{V}(\hat{x}) = \hat{W}_c^{\mathsf{T}} \sigma_c \left(V_c^{\mathsf{T}} \hat{x} \right) = \hat{W}_c^{\mathsf{T}} \sigma_c(z), \tag{35}$$

where $\hat{W}_c$ is the estimate of the ideal weights W_c. Since the hidden layer weight matrix V_c is fixed, the activation function vector $\sigma_c(V_c^{\mathsf{T}} \hat{x})$ is denoted as $\sigma_c(z) : \mathbb{R}^n \to \mathbb{R}^{k_c}$ with $z = V_c^{\mathsf{T}} \hat{x}$.

The derivative of the cost function $V(\hat{x})$ with respect to $\hat{x}$ is

$$V_{\hat{x}} = \nabla \sigma_c^{\mathsf{T}} W_c + \nabla \varepsilon_c, \tag{36}$$

where $\nabla \sigma_c^{\mathsf{T}} = V_c (\partial \sigma_c^{\mathsf{T}}(z) / \partial z)$ and $\nabla \varepsilon_c = \partial \varepsilon_c / \partial \hat{x}$. Note that the gradient of the reconstruction error $\nabla \varepsilon_c$ is also bounded. In addition, the derivative of $\hat{V}(\hat{x})$ with respect to $\hat{x}$ is derived as

$$\hat{V}_{\hat{x}} = \nabla \sigma_c^{\mathsf{T}} \hat{W}_c. \tag{37}$$

Then, the approximate Hamiltonian function can be derived as

$$H(\hat{x}, u, \hat{W}_c) = \hat{W}_c^{\mathsf{T}} \nabla \sigma_c F(\hat{x}, u) + r(\hat{x}, u) = e_c. \tag{38}$$

In addition, it is worth pointing out that, in the expression of the error e_c, the knowledge of the system dynamics is required. To overcome this limitation, the NN observer $\dot{\hat{x}}$, developed in (9), is used to replace the system dynamics $F(\hat{x}, u)$ in (38) to yield a modified expression of e_c as

$$e_c = \hat{W}_c^{\mathsf{T}} \nabla \sigma_c \dot{\hat{x}} + r(\hat{x}, u). \tag{39}$$

Given any admissible control policy u, it is desired to select $\hat{W}_c$ to minimise the squared residual error $E_c(\hat{W}_c)$ as

$$E_c(\hat{W}_c) = \frac{1}{2} e_c^{\mathsf{T}} e_c. \tag{40}$$

The weight update law for the critic NN is selected as the normalised gradient descent algorithm

$$\dot{\hat{W}}_c = -\alpha \frac{\psi}{(\psi^{\mathsf{T}} \psi + 1)^2} (\psi^{\mathsf{T}} \hat{W}_c + r(\hat{x}, u)), \tag{41}$$

where $\alpha > 0$ is the learning rate and $\psi = \nabla \sigma_c^{\mathsf{T}} \dot{\hat{x}}$. This is a modified Levenberg–Marquardt algorithm where $(\psi^{\mathsf{T}} \psi + 1)^2$ is used for normalisation instead of $(\psi^{\mathsf{T}} \psi + 1)$. This is required in the proofs, where one needs both appearances of $\psi / (\psi^{\mathsf{T}} \psi + 1)$ in (41) to be bounded (Ioannou & Fidan, 2006). Let the weight estimation error of critic NN be $\tilde{W}_c = W_c - \hat{W}_c$.

Before proceeding, we present an assumption as follows.

Assumption 1:

(1) *The NN approximate error and its gradient are bounded on a compact set containing* Ω *so that* $\|\varepsilon_c\| < \varepsilon_{cM}$ *and* $\|\nabla \varepsilon_c\| < \varepsilon_{dM}$.

(2) The NN activation function and its gradient are bounded such that
$\|\sigma_c\| < \sigma_{cM}$ *and* $\|\nabla\sigma_c\| < \sigma_{dM}$.

Based on Vamvoudakis and Lewis (2010)*, these assumptions are standard. Assumption* 1(2) *is satisfied, e.g., by sigmoids, tanh, and other standard NN activation functions.*

By (29) and (34), we obtain

$$0 = r(\hat{x}, u) + W_c^{\mathsf{T}} \nabla\sigma_c \dot{\hat{x}} - \vartheta, \tag{42}$$

where $\vartheta = -\nabla\varepsilon_c \dot{\hat{x}}$ is the residual error due to the NN approximation.

Substituting (42) into (41) and using the notation

$$\psi_1 = \psi/(\psi^{\mathsf{T}}\psi + 1), \quad \psi_2 = \psi^{\mathsf{T}}\psi + 1, \tag{43}$$

we can obtain the dynamics of the critic NN weight estimation error as

$$\dot{\tilde{W}}_c = -\alpha\psi_1\psi_1^{\mathsf{T}}\tilde{W}_c + \alpha\psi_1\frac{\vartheta}{\psi_2}. \tag{44}$$

From the form of ψ_1, there exists a positive constant $\psi_{1M} > 1$ such that $\|\psi_1\| < \psi_{1M}$. In addition, it is important to note that the persistence excitation condition is required for tuning critic NN. In order to satisfy the persistent excitation condition, probing noise is added to the control input (Vamvoudakis & Lewis 2010). Furthermore, the persistent excitation condition ensures $\|\psi_1\| \geq \psi_{1m}$, with ψ_{1m} being a positive constant.

Next, by using (32) and (36), the corresponding feedback control u is given by

$$u = -\frac{1}{2}R^{-1}\left(\frac{\partial F(x, u)}{\partial u}\right)^{\mathsf{T}}(\nabla\sigma_c^{\mathsf{T}} W_c + \nabla\varepsilon_c). \tag{45}$$

The approximate expression of u can be developed as

$$\hat{u} = -\frac{1}{2}R^{-1}\left(\frac{\partial \hat{F}(\hat{x}, u)}{\partial u}\right)^{\mathsf{T}}\nabla\sigma_c^{\mathsf{T}}\hat{W}_c. \tag{46}$$

Additionally, by (45), it is important to note that the term $\partial F(x, u)/\partial u$ is required for computing the control u. However, for unknown nonlinear systems, this term cannot be obtained directly. In this paper, using the observer NN, its estimates can be obtained by

$$\frac{\partial \hat{F}(\hat{x}, u)}{\partial u} = \frac{\partial \hat{G}(\hat{x}, u)}{\partial u} = \frac{\partial \hat{W}\sigma(\hat{V}\hat{\bar{x}})}{\partial u} = \hat{W}\frac{\partial\sigma(\hat{V}\hat{\bar{x}})}{\partial \hat{V}\hat{\bar{x}}}\hat{V}\frac{\partial\hat{\bar{x}}}{\partial u}. \tag{47}$$

Thus, it is believed that $\partial\hat{F}(\hat{x}, u)/\partial u$ can be obtained by the backpropagation from the outputs of the observer NN to its input u.

In the following, the stability analysis will be performed. For the design of the NN-observer-based control system, it seems natural to take a Lyapunov function candidate that consists of a combination of the Lyapunov functions for the NN observer and the controller. The following theorem shows the stability of the whole system.

Theorem 4.1: *Consider the NN observer system (9) and the feedback control provided by (45). Let weight tuning laws for the observer and the controller be provided by*

$$\begin{aligned}\dot{\hat{W}} &= -\eta_1(\tilde{y}^{\mathsf{T}}CA_c^{-1})^{\mathsf{T}}\sigma^{\mathsf{T}}(\hat{V}\hat{\bar{x}}) - \theta_1\|\tilde{y}\|\hat{W}\\ \dot{\hat{V}} &= -\eta_2(\tilde{y}^{\mathsf{T}}CA_c^{-1}\hat{W}(I - \Gamma(\hat{V}\hat{\bar{x}})))^{\mathsf{T}}\mathrm{sgn}^{\mathsf{T}}(\hat{\bar{x}}) - \theta_2\|\tilde{y}\|\hat{V}\end{aligned} \tag{48}$$

and

$$\dot{\hat{W}}_c = -\alpha\frac{\psi_1}{\psi^{\mathsf{T}}\psi + 1}(\psi^{\mathsf{T}}\hat{W} + r(\hat{x}, u)), \tag{49}$$

then all the signals x, $\tilde{x}$, $\tilde{W}$, $\tilde{V}$, and $\tilde{W}_c$ in the NN-observer-based control system are UUB.

Proof: Choose the following Lyapunov function candidate:

$$J(t) = J_o(t) + J_c(t), \tag{50}$$

where J_o is defined as in (14) and J_c is given by

$$J_c = \frac{1}{2\alpha}\mathrm{tr}\{\tilde{W}_c^{\mathsf{T}}\tilde{W}_c\} + \alpha(x^{\mathsf{T}}x + \gamma V(x)) \tag{51}$$

with $\gamma > 0$. The time derivative of J_c is derived as

$$\dot{J}_c = \dot{J}_{c1} + \dot{J}_{c2}, \tag{52}$$

where

$$\begin{aligned}\dot{J}_{c1} &= \frac{1}{\alpha}\mathrm{tr}\{\tilde{W}_c^{\mathsf{T}}\dot{\tilde{W}}_c\} = \frac{1}{\alpha}\mathrm{tr}\left\{\tilde{W}_c^{\mathsf{T}}\left(-\alpha\psi_1\psi_1^{\mathsf{T}}\tilde{W}_c + \alpha\psi_1\frac{\vartheta}{\psi_2}\right)\right\}\\ &= \frac{1}{\alpha}\mathrm{tr}\left\{-\alpha\tilde{W}_c^{\mathsf{T}}\psi_1\psi_1^{\mathsf{T}}\tilde{W}_c + \frac{1}{2}\left(2\alpha\tilde{W}_c^{\mathsf{T}}\psi_1\frac{\vartheta}{\psi_2}\right)\right\}\\ &\leq -\psi_1^2\|\tilde{W}_c\|^2 + \frac{1}{2\alpha}(\alpha^2\psi_1^2\|\tilde{W}_c\|^2 + \vartheta^2)\\ &\leq -\left(\psi_{1m} - \frac{\alpha}{2}\psi_{1M}\right)\|\tilde{W}_c\|^2 + \frac{1}{2\alpha}\vartheta^2,\end{aligned} \tag{53}$$

$$\begin{aligned}\dot{J}_{c2} &= 2\alpha x^{\mathsf{T}}\dot{x} + \alpha\gamma(-x^{\mathsf{T}}Q_c x - u^{\mathsf{T}}Ru)\\ &= 2\alpha x^{\mathsf{T}}(Ax + W\sigma(V\bar{x}) + \varepsilon) + \alpha\gamma(-x^{\mathsf{T}}Q_c x - u^{\mathsf{T}}Ru)\\ &\leq \alpha(2\|A\| + 2)\|x\|^2 + \alpha\|W\sigma(V\bar{x})\|^2 + \|\varepsilon\|^2\\ &\quad - \alpha\gamma\lambda_{\min}(Q_c)\|x\|^2 - \alpha\gamma\lambda_{\min}(R)\|u\|^2\end{aligned}$$

$$\leq \alpha\big(2\|A\| + 2 - \gamma\lambda_{\min}(Q_c)\big)\|x\|^2 - \alpha\gamma\lambda_{\min}(R)\|u\|^2 + \alpha W_M^2\sigma_M^2 + \varepsilon_M^2. \quad (54)$$

Thus, we have

$$\begin{aligned}\dot{J}_c \leq &-\Big(\psi_{1m} - \frac{\alpha}{2}\psi_{1M}\Big)\|\tilde{W}_c\|^2 \\ &- \alpha\big(-2\|A\| - 2 + \gamma\lambda_{\min}(Q_c)\big)\|x\|^2 \\ &- \alpha\gamma\lambda_{\min}(R)\|u\|^2 + \frac{1}{2\alpha}\vartheta^2 + \alpha W_M^2\sigma_M^2 + \varepsilon_M^2. \quad (55)\end{aligned}$$

Note that, according to Assumption 1, it is assumed that the gradients of the critic NN approximation error and the activation function vector are upper bounded, i.e., $\nabla\varepsilon_c \leq \varepsilon_{dM}$, $\nabla\sigma_c \leq \sigma_{dM}$, and the residual error is upper bounded, i.e., $\vartheta \leq \vartheta_M$. Hence, we have

$$\begin{aligned}\dot{J}_c \leq &-\Big(\psi_{1m} - \frac{\alpha}{2}\psi_{1M}\Big)\|\tilde{W}_c\|^2 \\ &- \alpha\big(-2\|A\| - 2 + \gamma\lambda_{\min}(Q_c)\big)\|x\|^2 \\ &- \alpha\gamma\lambda_{\min}(R)\|u\|^2 + D_M, \quad (56)\end{aligned}$$

where

$$D_M = \frac{1}{2\alpha}\vartheta_M^2 + \alpha W_M^2\sigma_M^2 + \varepsilon_M^2. \quad (57)$$

Then, based on (27) and (55), combining $J_o(t)$ and $J_c(t)$, $\dot{J}(t)$ becomes

$$\begin{aligned}\dot{J}(t) \leq &-\frac{1}{2}\lambda_{\min}(\Lambda)\|\tilde{x}\|^2 + \|\tilde{x}\|\big(\|P\|\bar{\omega} + \big(\theta_1\|C\| - K_1^2\big)K_2^2 \\ &+ (\theta_2\|C\| - 1)K_3^2\big) - \Big(\psi_{1m} - \frac{\alpha}{2}\psi_{1M}\Big)\|\tilde{W}_c\|^2 \\ &- \alpha\big(-2\|A\| - 2 + \gamma\lambda_{\min}(Q_c)\big)\|x\|^2 \\ &- \alpha\gamma\lambda_{\min}(R)\|u\|^2 + D_M. \quad (58)\end{aligned}$$

Therefore, if θ_1, θ_2, γ, and α are selected to satisfy

$$\begin{aligned}&\theta_1 \geq K_1^2/\|C\|, \quad \theta_2 \geq 1/\|C\|, \\ &\gamma > \frac{2\|A\| + 2}{\lambda_{\min}(Q_c)}, \quad \alpha < \frac{2\psi_{1m}}{\psi_{1M}},\end{aligned} \quad (59)$$

and given that the inequalities

$$\begin{aligned}\|\tilde{x}\| &> \frac{2\left(\|P\|\bar{\omega} + \big(\theta_1\|C\| - K_1^2\big)K_2^2 + (\theta_2\|C\| - 1)K_3^2\right)}{\lambda_{\min}(\Lambda)} \\ \|\tilde{W}_c\| &> \sqrt{\frac{D_M}{\psi_{1m} - \frac{\alpha}{2}\psi_{1M}}} \quad (60)\end{aligned}$$

hold, then $\dot{J}(t) < 0$. Hence, using Lyapunov theory (Lewis et al., 1999), it can be concluded that the observer error $\tilde{x}$, the state x, and the NN weight estimation errors $\tilde{W}$, $\tilde{V}$, and $\tilde{W}_c$ are UUB in the NN-observer-based control system. □

Remark 3: It should be noted that in (59) and (60), the constraints for θ_1, θ_2, and $\tilde{x}$ are set the same as the NN observer designed earlier. In fact, a nonlinear separation principle is not valid. However, for the proof of the NN-observer-based control system, the closed-loop dynamics incorporates the observer dynamics, then we can develop simultaneous weight tuning algorithms for the NN observer and the neuro-controller.

5. Simulation study

In this section, two examples are provided to demonstrate the effectiveness of the NN-observer-based optimal control scheme developed in this paper.

5.1. Example 1

Consider the affine nonlinear continuous-time system described by

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= -0.5x_2 - 0.5x_3(1 - (\cos(2x_1) + 2)^2) + \cos(2x_1)u \\ &\quad + 2u \\ y &= x_1 + x_3, \end{aligned} \quad (61)$$

with initial conditions $x_1(0) = 1$, $x_2(0) = -1$, and $x_3(0) = 1$. The performance index function is defined by (2), where Q and R are chosen as identity matrices of appropriate dimensions. It is assumed that the system dynamics are unknown, the system states are not available for measurements, and only the input and output of the system are measurable.

During the design process, the following statements are need. In Bernard (1970), the square matrix A is called Hurwitz matrix if every eigenvalue of A has strictly negative real part, i.e., $\mathrm{Re}[A] < 0$ for each eigenvalue. With regard to observable (Dorf, 1991; Singh, 1975), a system is completely observable if and only if there exists a finite time T such that the initial state $x(0)$ can be determined from the observation history $y(t)$ given the control $u(t)$. Here, the system is observable when the determinant of the observability matrix P_o is nonzero, where $P_o = [CCA \ldots CA^{n-1}]^{\mathsf{T}}$ which is an $n \times n$ matrix; that is, if the row rank of the observability matrix P_o is equal to n, then the system is observable (Dorf, 1991).

At first, an NN observer is established to estimate system states. For ensuring that A is Hurwitz matrix, the pair (C, A) is observable and $A - LC$ is Hurwitz matrix, we select

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -12 & -16 & -7 \end{bmatrix}, \quad L = \begin{bmatrix} 28 \\ -30 \\ 15 \end{bmatrix}. \tag{62}$$

The observer NN is a three-layer NN with one hidden layer containing eight neurons. The input layer involves four neurons and the output layer contains three neurons. The activation function $\sigma(\cdot)$ is selected as hyperbolic tangent function $\tanh(\cdot)$. Let the learning rates be $\eta_1 = \eta_2 = 100$ and the design parameters be $\theta_1 = \theta_2 = 1.5$. Additionally, the initial weights of W and V are all set to be random within [0, 0.2]. Then, according to Figure 1, we can complete the design of the NN observer for system (61).

Then, based on the observed states, a feedforward neuro-controller is constructed via the ADP method to obtain the optimal control of the system. The basic idea of ADP is to obtain the nearly optimal control by constructing a critic NN to approximate the cost function. In the design, for both simplicity of learning and efficiency of approximation, based on Igelnik and Pao (1995), the activation functions of the critic NN are chosen from the fourth-order series expansion of the value function. Only polynomial terms of even order are considered, therefore,

$$\begin{aligned}\sigma_c = \big[&x_1^2, x_1x_2, x_1x_3, x_2^2, x_2x_3, x_3^2, x_1^4, x_2^4, x_3^4, x_1^3x_2, x_1^3x_3,\\ &x_2^3x_1, x_2^3x_3, x_3^3x_1, x_3^3x_2, x_1^2x_2^2, x_2^2x_3^2, x_1^2x_3^2, x_1^2x_2x_3,\\ &x_1x_2^2x_3, x_1x_2x_3^2\big].\end{aligned}$$

Then, the critic NN weights are denoted as $\hat{W}_c = [\hat{W}_{c1}, \hat{W}_{c2}, \ldots, \hat{W}_{c21}]^{\mathsf{T}}$. The learning rate for the critic NN is selected as $\alpha = 0.5$. Additionally, in the beginning, the initial weights of $\hat{W}_c$ are set as $[0.7, 0.7, \ldots, 0.7]^{\mathsf{T}}$. Moreover, based on the critic NN and the observer NN, the control is updated by calculating (46). In order to maintain the excitation condition, probing noise is added to the control input for the first 10 s as in Vamvoudakis and Lewis (2010). Note that for initialisation of network weights, the ideal initial values for weights, i.e., those weights will maximise the effectiveness and speed with which an NN learns. However, the ideal initial weights cannot yet be determined theoretically (Tamura & Tateishi, 1997). Here, the best possible NN parameters containing the initial weight are ascertained by repeating experiment. Furthermore, for different initial weight (Haykin, 1999), there exist some differences on the effectiveness and speed with which an NN learns. Moreover, when the initialisation of weights is irrational, the convergence results of NNs are probably bad. The structure diagram in Figure 2 illustrates the design of the NN-observer-based controller using the ADP method.

Upon completion of simulation, the observed-state trajectories are shown in Figures 3–5, where the corresponding real-state trajectories are also plotted for assessing the performance of the NN observer. Moreover, the errors between the observed and real states are shown in Figure 6. From Figure 6, it is clear that the observed states x_{o1}, x_{o2}, x_{o3}, i.e., $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$, quickly approach the real states. The convergence curves of norms of the observer NN weights and critic NN weights are shown in Figure 7. Figures 8

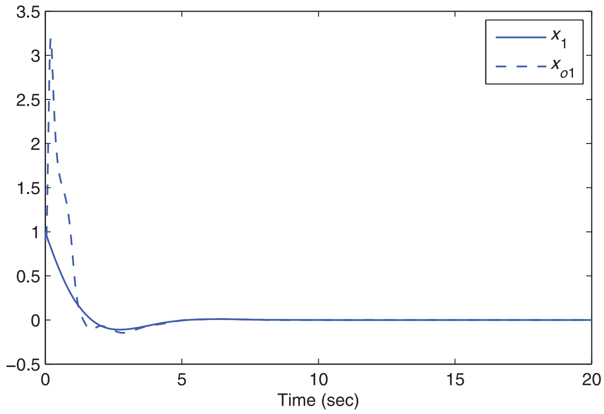

Figure 3. The trajectories of real state x_1 and observed state x_{o1}.

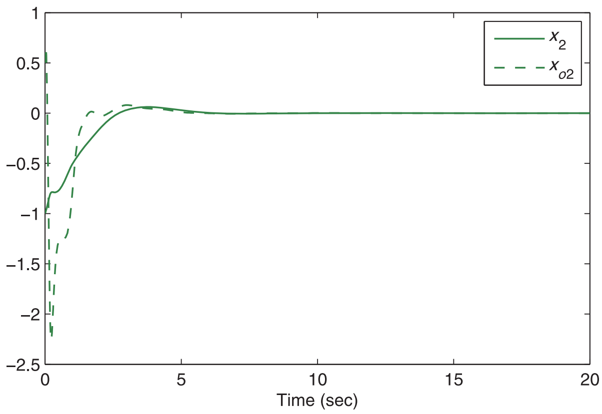

Figure 4. The trajectories of real state x_2 and observed state x_{o2}.

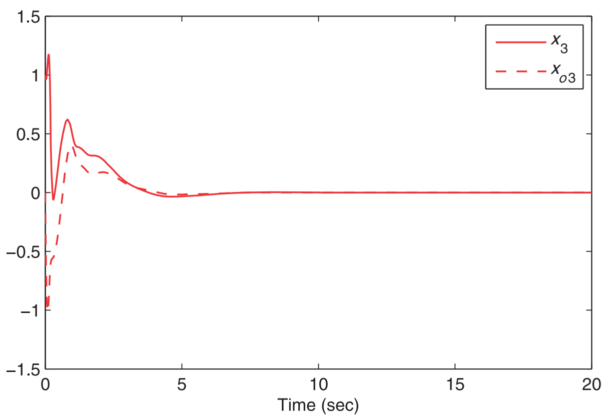

Figure 5. The trajectories of real state x_3 and observed state x_{o3}.

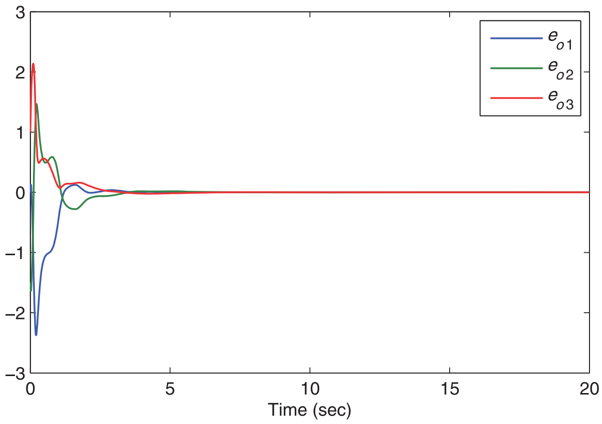

Figure 6. The NN observer errors e_{o1}, e_{o2}, and e_{o3}.

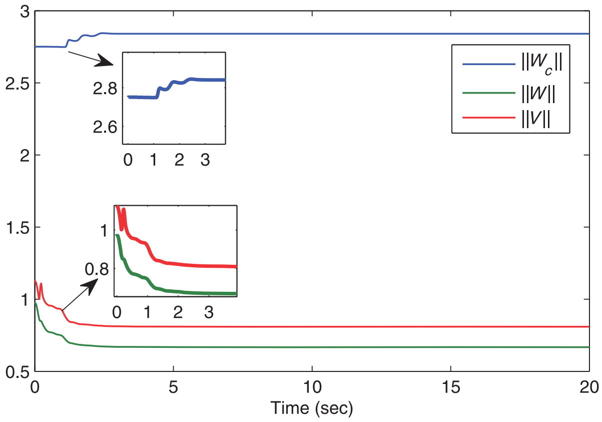

Figure 7. The norms of observer NN and critic NN weights $\|W\|$, $\|V\|$, $\|W_c\|$.

and 9 depict the system output trajectory y and the nearly optimal control signal u, respectively. It can be seen from Figures 8 and 9 that proposed NN-observer-based optimal controller yields very good control effect.

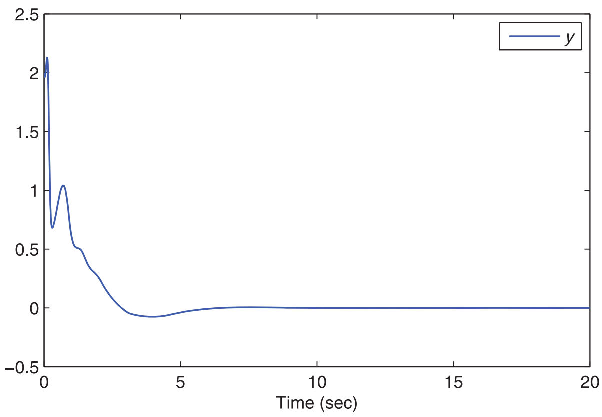

Figure 8. The system output y.

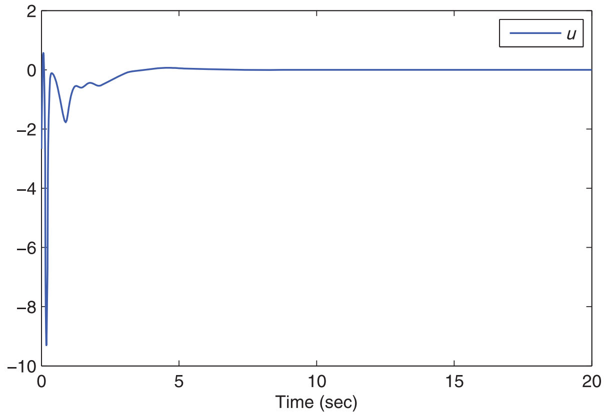

Figure 9. The control input u.

Additionally, it is significant to state that based on Dorf (1991), only the real parts of all eigenvalues of $(A - LC)$ are negative, $\tilde{x}$ can be attenuated to zero, and $\hat{x}$ approximates the actual state x. Moreover, the rate of state approximation depends mainly on the choice of L and the eigenvalue assignment of $(A - LC)$. In simulation, we select A and L by repetitious experiments for yielding better performance of the whole system which contains the NN observer and the neuro-controller. From Figures 3–5, we can find that the observed states x_{o1}, x_{o2}, and x_{o3} approach the real states at different rates. This is mainly due to the difference of the negative real parts of eigenvalues of $(A - LC)$.

5.2. *Example 2*

Consider the nonaffine nonlinear continuous-time system

$$\begin{aligned} \dot{x_1} &= -x_1 + x_2, \\ \dot{x_2} &= -x_1 - (1 - \sin^2(x_1))x_2 + \sin(x_1)u + 0.1u^2, \\ y &= x_1, \end{aligned} \tag{63}$$

with initial conditions $x_1(0) = 1$ and $x_2(0) = -0.5$. The performance index function is also defined by (2), where Q and R are chosen as identity matrices of appropriate dimensions. It is assumed that the system dynamics are unknown, the system states are not available for measurements, and only the input and output of the system are measurable.

In order to estimate the system states, an NN observer is set up and the corresponding parameters are chosen as

$$A = \begin{bmatrix} 0 & 1 \\ -6 & -5 \end{bmatrix}, \quad L = \begin{bmatrix} 10 \\ -2 \end{bmatrix}. \tag{64}$$

The observer NN is a three-layer NN with one hidden layer containing eight neurons. The input layer involves three neurons and the output layer contains two neurons. The activation function $\sigma(\cdot)$, the learning rates η_1, η_2, and the design parameters θ_1, θ_2 are set the same as Example 1.

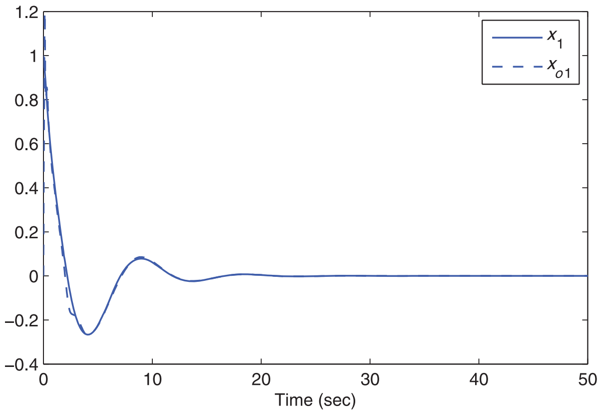

Figure 10. The trajectories of real state x_1 and observed state x_{o1}.

The initial weights of W and V are all set to be random within [0.5, 1]. From Figures 10 and 11, it is clear that the observed states x_{o1}, x_{o2}, i.e., $\hat{x}_1$, $\hat{x}_2$, quickly approach the real states.

Then, based on the observed states, similar to Example 1, a critic NN is constructed to obtain the nearly optimal control. The activation functions of the critic NN are chosen from the sixth-order series expansion of the value function. Only polynomial terms of even order are considered, therefore,

$$\sigma_c = \left[x_1^2, x_1x_2, x_2^2, x_1^4, x_1^3x_2, x_1^2x_2^2, x_1x_2^3, x_2^4, x_1^6, x_1^5x_2, x_1^4x_2^2, x_1^3x_2^3, x_1^2x_2^4, x_1x_2^5, x_2^6\right].$$

The corresponding parameters are set the same as Example 1. Additionally, the initial weights of $\hat{W}_c$ are set as $[1, 1, \ldots, 1]^\mathsf{T}$. Moreover, in order to maintain the excitation condition, probing noise is added to the control input for the first 10 s as in Vamvoudakis and Lewis (2010).

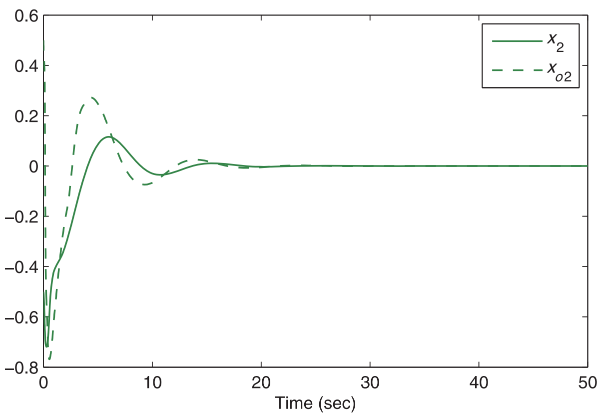

Figure 11. The trajectories of real state x_2 and observed state x_{o2}.

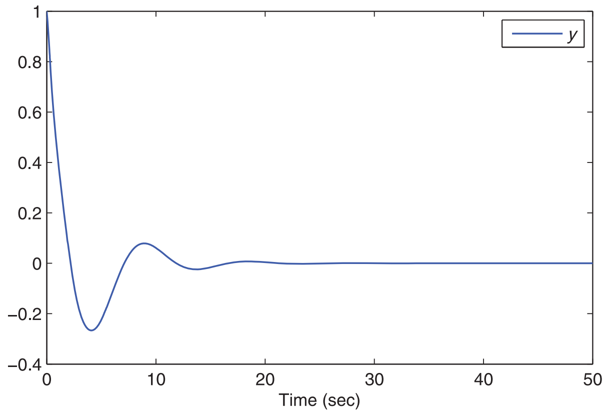

Figure 12. The system output y.

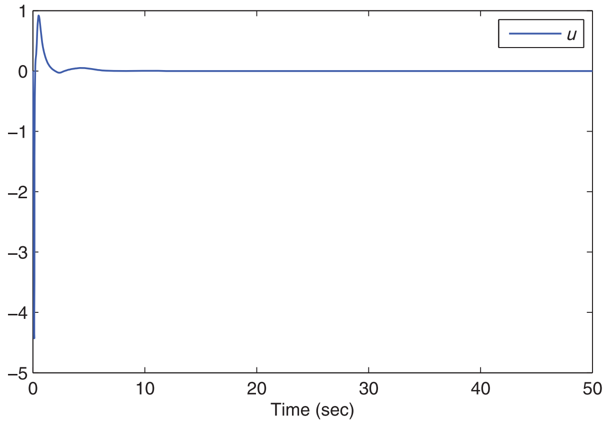

Figure 13. The control input u.

After simulation, the observed-state trajectories are shown in Figures 10 and 11, where the corresponding real-state trajectories are also plotted for assessing the performance of the NN observer. Figures 12 and 13 depict the system output trajectory y and the nearly optimal control signal u, respectively. It can be seen from Figures 12 and 13 that the proposed NN-observer-based optimal controller is valid.

6. Conclusion

In this paper, we develop an observer-based optimal control scheme for unknown nonlinear continuous-time systems. An NN observer is designed to estimate the system states. Then, based on the observed states, the feedforward neuro-controller is developed based on the ADP method. In the implementation of the scheme, two NN structures are used: a three-layer feedforward NN is used to construct the NN observer which can be applied to systems with higher degrees of nonlinearity and without a priori knowledge of the system dynamics, and a critic NN is employed to approximate the value function. Moreover, the UUB stability of the NN-observer-based control system is proved. The simulation results have confirmed the validity of the proposed observer-based optimal control scheme based on ADP.

Acknowledgements

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

References

Abdollahi, F., Talebi, H.A., & Patel, R.V. (2006). A stable neural network-based observer with application to flexible joint manipulators. *IEEE Transactions on Neural Networks*, *17*, 118–129.

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

Ahmed, M.S., & Riyaz, S.H. (2000). Dynamic observer – a neural net approach. *Journal of Intelligent & Fuzzy Systems*, *9*, 113–127.

Al-Tamimi, A., Lewis, F.L., & Abu-Khalaf, M. (2008). Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof. *IEEE Transactions on Systems, Man, Cybernetics, Part B, Cybernetics*, *38*, 943–949.

Bellman, R.E. (1957). *Dynamic programming*. New Jersey: Princeton University Press.

Bernard, A.A. (1970). On the total negativity of the Hurwitz matrix. *SIAM Journal on Applied Mathematics*, *18*, 407–414.

Bertsekas, D.P., & Tsitsiklis, J.N. (1996). *Neuro-dynamic programming*. Belmont: Athena Scientific.

Chen, F.C., & Khalil, H.K. (1995). Adaptive control of a class of nonlinear discrete-time systems using neural networks. *IEEE Transactions on Automatic Control*, *40*, 791–801.

Dierks, T., Thumati, B.T., & Jagannathan, S. (2009). Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence. *Neural Networks*, *22*, 851–860.

Dorf, R.C. (1991). *Modern control systems*. Boston: Addison-Wesley Longman Publishing Co., Inc.

Farza, M., Sboui, A., Cherrier, E., & M'Saad, M. (2010). High-gain observer for a class of time-delay nonlinear systems. *International Journal of Control*, *83*, 273–280.

Goodwin, G.G., & Payne, R.L. (1977). *Dynamic system identification: Experiment design and data analysis*. New York: Academic Press.

Haykin, S.S. (1999). *Neural networks: A comprehensive foundation*. New Jersey: Prentice Hall.

He, P., & Jagannathan, S. (2007). Reinforcement learning neural-network-based controller for nonlinear discrete-time systems with input constraints. *IEEE Transactions on Systems, Man, Cybernetics, Part B, Cybernetics*, *37*, 425–436.

Igelnik, B., & Pao, Y.H. (1995). Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Transactions on Neural Networks*, *6*, 1320–1329.

Ioannou, P.A., & Fidan, B. (2006). *Adaptive control tutorial*. Philadelphia: Society for Industrial and Applied Mathematics.

Jagannathan, S. (2006). *Neural network control of nonlinear discrete-time systems*. Boca Raton: CRC Press.

Jin, G., Sain, M.K., Pham, K.D., Spencer, B.F., & Ramallo, J.C. (2001). Modeling MR-dampers: A nonlinear blackbox approach. In *Proceedings of the American control conference* (pp. 429–434). Arlington, VA.

Jo, N.H., & Seo, J.H. (2002). Observer design for non-linear systems that are not uniformly observable. *International Journal of Control*, *75*, 369–380.

Jung, J.C., Huh, K., & Lee, T.H. (2008). Observer design methodology for stochastic and deterministic robustness. *International Journal of Control*, *81*, 1172–1182.

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

Lewis, F.L., & Syrmos, V.L. (1995). *Optimal control*. New York: Wiley.

Lewis, F.L., & Vrabie, D. (2009). Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits and Systems Magazine*, *9*, 32–50.

Liu, D., Xiong, X., & Zhang, Y. (2001). Action-dependent adaptive critic designs. In *Proceedings of international joint conference on neural networks* (pp. 990–995). Washington, DC.

Liu, D., Zhang, Y., & Zhang, H. (2005). A self-learning call admission control scheme for CDMA cellular networks. *IEEE Transactions on Neural Networks*, *16*, 1219–1228.

Michael, T., & Harley, R. (1995). Identification and control of induction machines using artificial neural networks. *IEEE Transactions on Industry Applications*, *31*, 612–619.

Murray, J.J., Cox, C.J., Lendaris, G.G., & Saeks, R. (2002). Adaptive dynamic programming. *IEEE Transactions on Systems, Man, Cybernetics, Part C, Applications and Reviews*, *32*, 140–153.

Narendra, K., & Parthasarathy, K. (1990). Identification and control of dynamic system using neural networks. *IEEE Transactions on Neural Networks*, *1*, 4–27.

Nicosia, S., Tomei P., & Tornambe, A. (1989). An approximate observer for a class of nonlinear systems. *Systems & Control Letters*, *12*, 43–51.

Park, J.H., Huh, S.H., Kim, S.H., Seo, S.J., & Park, G.T. (2005). Direct adaptive controller for nonaffine nonlinear systems using self-structuring neural networks. *IEEE Transactions on Neural Networks*, *16*, 414–422.

Prokhorov, D.V., & Wunsch, D.C. (1997). Adaptive critic designs. *IEEE Transactions on Neural Networks*, *8*, 997–1007.

Selmic, R., & Lewis, F.L. (2001). Multimodel neural networks identification and failure detection of nonlinear systems. In *Proceeding of the 40th IEEE conference on decision and control* (pp. 3128–3133). Orlando, FL.

Si, J., Barto, A.G., Powell, W.B., & Wunsch, D.C. (2004). *Handbook of learning and approximate dynamic programming*. New York: Wiley.

Singh, S.N. (1975). Observability in nonlinear systems with immeasurable inputs. *International Journal of Systems Science*, *6*, 723–732.

Slotine, J.J.E., & Li, W. (1991). *Applied nonlinear control*. New Jersey: Prentice Hall.

Tamura, S., & Tateishi, M. (1997). Capabilities of a four-layered feedforward neural network: Four layers versus three. *IEEE Transactions on Neural Networks*, *8*, 251–255.

Theocharis, J., & Petridis, V. (1994). Neural network observer for induction motor control. *IEEE Control Systems*, *14*, 26–37.

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

Vrabie, D., & Lewis, F.L. (2009). Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Networks*, *22*, 237–246.

Walter, E., & Pronzato, L. (1997). *Identification of parametric models from experimental data, communications and control engineering series*. New York: Springer.

Wang, F.Y., Jin, N., Liu, D., & Wei, Q. (2011). Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ε-error bound. *IEEE Transactions on Neural Networks*, *22*, 24–36.

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

Wang, F.Y., Zhang, H., & Liu, D. (2009). Adaptive dynamic programming: An introduction. *IEEE Computational Intelligence Magazine*, *4*, 39–47.

Werbos, P.J. (1977). Advanced forecasting methods for global crisis warning and models of intelligence. *General Systems Yearbook*, *22*, 25–38.

Yu, W. (2009). *Recent advances in intelligent control systems*. London: Springer.

Zhang, H., Luo, Y., & Liu, D. (2009). Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. *IEEE Transactions on Neural Networks*, *20*, 1490–1503.

Zhang, H., Wei, Q., & Luo, Y. (2008). A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear system via the greedy HDP iteration algorithm. *IEEE Transactions on Systems, Man, Cybernetics, Part B, Cybernetics*, *38*, 937–942.