Data-Driven Robust Approximate Optimal Tracking Control for Unknown General Nonlinear Systems Using Adaptive Dynamic Programming Method

Huaguang Zhang, *Senior Member, IEEE*, Lili Cui, Xin Zhang, and Yanhong Luo, *Member, IEEE*

***Abstract*— In this paper, a novel data-driven robust approximate optimal tracking control scheme is proposed for unknown general nonlinear systems by using the adaptive dynamic programming (ADP) method. In the design of the controller, only available input–output data is required instead of known system dynamics. A data-driven model is established by a recurrent neural network (NN) to reconstruct the unknown system dynamics using available input–output data. By adding a novel adjustable term related to the modeling error, the resultant modeling error is first guaranteed to converge to zero. Then, based on the obtained data-driven model, the ADP method is utilized to design the approximate optimal tracking controller, which consists of the steady-state controller and the optimal feedback controller. Further, a robustifying term is developed to compensate for the NN approximation errors introduced by implementing the ADP method. Based on Lyapunov approach, stability analysis of the closed-loop system is performed to show that the proposed controller guarantees the system state asymptotically tracking the desired trajectory. Additionally, the obtained control input is proven to be close to the optimal control input within a small bound. Finally, two numerical examples are used to demonstrate the effectiveness of the proposed control scheme.**

***Index Terms*— Adaptive dynamic programming, data-driven model, neural networks, optimal tracking control, robust control.**

I. Introduction

DURING the last decades, the adaptive dynamic programming (ADP) as an effective intelligent control method has played an important role in seeking solutions for the optimal control [1]–[11]. Some recent surveys in [12]–[14] on ADP techniques present excellent overview of the state-of-the-art developments. However, it is worth mentioning that most of the existing results based on ADP technique require a knowledge of known nonlinear dynamics. Hence, some studies have attempted to solve the optimal control solution based on ADP technique without an *a priori* system model [15]–[18]. For linear discrete-time systems, Q-learning was introduced to relax some of the exact model-matching restrictions in [15] and [16], which allows model-free tuning of the action and critic networks. For linear continuous-time systems, Vrabie *et al.* proposed a new formulation of the proportional algorithm which converges to the optimal control solution without using internal dynamics of the system in [17]. Then this idea was extended to nonlinear continuous-time systems in [18], where the knowledge of the input-to-state dynamics was still required. Nevertheless, the requirement of system dynamics is hard to be satisfied, either fully or even partially known. Hence, for the unknown general nonlinear systems, ADP methods mentioned above cannot be applied directly.

Fortunately, we can access input–output data of the unknown general nonlinear systems in many practical control processes. So it is desirable to use available input–output data in the design of the controller. Such techniques belong to the field of data-driven control techniques [19]–[23]. The historical input–output data could be incorporated indirectly in the form of a data-driven model. The data-driven model could extract useful information contained in input–output data and capture input–output mapping. Markov models, neural network (NN) models, well-structured filters, wavelet models, and other function approximation models can be regarded as data-driven models [24]–[31]. In this paper, we proposed a data-driven model based on a recurrent neural network (RNN) to reconstruct the unknown system dynamics by using available input–output data. At first, a novel adjustable term related to the modeling error is added to the data-driven model, which guarantees the modeling error to converge to zero. Once the obtained data-driven model is established, it can be used to design controller. Consequently, it paves the way of applying the ADP method to deal with the optimal control problem of the unknown general nonlinear systems.

It is noted that most works based on ADP mentioned above deal with stabilization issues. However, few results consider the optimal tracking control problem based on ADP except in [32]–[37]. In fact, optimal tracking problem is at least as significant as the stabilization problem in the control field. To the best of our knowledge, to date there have been no attempts to solve the optimal tracking problem of unknown general nonlinear systems based on the ADP method. In this paper, for the first time, we propose a robust approximate

Manuscript received January 15, 2011; revised July 13, 2011; accepted September 4, 2011. Date of publication October 13, 2011; date of current version December 13, 2011. This work was supported in part by the National Natural Science Foundation of China under Grant 50977008, Grant 60821063, and Grant 61034005, and the National Basic Research Program of China under Grant 2009CB320601.

H. Zhang is with the School of Information Science and Engineering, Northeastern University, Shenyang 110004, China. He is also with the State Key Laboratory of Synthetic Automation for Process Industries, Northeastern University, Shenyang 110004, China (e-mail: hgzhang@ieee.org).

L. Cui, X. Zhang, and Y. Luo are with the School of Information Science and Engineering, Northeastern University, Shenyang 110004, China (e-mail: cuilili8396@163.com; jackie_zx@yahoo.com.cn; neuluo@gmail.com).

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/TNN.2011.2168538

optimal tracking controller via constructing a novel data-driven model for the unknown general nonlinear systems. The derived controller consists of the steady-state controller, the optimal feedback controller, and a robustifying term. The steady-state controller is designed to achieve the desired tracking performance at the steady-state stage. The optimal feedback controller is developed based on ADP to stabilize the state tracking error dynamics at the transient stage in an optimal way. In most existing literature [2]–[5], the critic NN and the action NN are updated sequentially. By contrast, both these are updated simultaneously in this paper. Because of the utilization of NNs, there exist NN approximation errors inevitably. Few of these papers mentioned above considered the NN approximation errors (except [6], [10], [35]). However, the results obtained in [6], [10], and [35] are uniformly ultimately bounded (UUB). In this paper, a robustifying term is developed to compensate for the NN approximation errors, and hence the asymptotical tracking results can be obtained. Moreover, the obtained control input is ensured to be close to the optimal control input within a small bound.

The main contributions of this paper include the following.

1) It is the first time that the optimal tracking problem of the unknown general nonlinear systems based on ADP method is investigated.
2) A novel data-driven model is established based on an RNN which guarantees the modeling error to asymptotically converge to zero.
3) A robust approximate optimal tracking controller is developed to ensure that the tracking error converges to zero asymptotically. Moreover, the proposed controller can ensure that the obtained control input is close to the optimal control input within a small bound.
4) In the design of the optimal feedback controller based on ADP, the critic NN and the action NN are updated simultaneously.

The rest of this paper is organized as follows. In Section II, the problem formulation is given. The establishment of data-driven model is presented in Section III. Then the ADP-based approximate optimal tracking control scheme is proposed with stability proof in Section IV. The ADP-based robust approximate optimal tracking control scheme is developed with stability proof in Section V. Two simulation examples are presented to show the satisfactory performance of the proposed scheme in Section VI. Finally, the conclusions are drawn in Section VII.

II. Problem Formulation

Consider the following general nonlinear continuous-time systems:

$$\dot{x}(t) = h(x(t), u(t)) \tag{1}$$

where $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T \in \mathbb{R}^n$ is the state vector, $u(t) = (u_1(t), u_2(t), \ldots, u_m(t))^T \in \mathbb{R}^m$ is the input vector, and $h(\cdot, \cdot)$ is an unknown continuous nonlinear smooth function with respect to $x(t)$ and $u(t)$.

For optimal tracking control problem, the control objective is to design an optimal controller for (1) which ensures that the state vector $x(t)$ tracks the specified trajectory $x_d(t)$ and minimizes the infinite horizon performance index function as follows:

$$V(e(t)) = \int_t^\infty r(e(\tau), u(e(\tau)))d\tau \tag{2}$$

where $e(t) = x(t) - x_d(t)$ denotes the state tracking error, $r(e(t), u(t)) = e^T(t)Qe(t) + u^T(t)Ru(t)$ is the utility function, and Q and R are symmetric positive definite matrices with appropriate dimensions.

Since the system dynamics is completely unknown, we cannot apply existing ADP methods to (1) directly. Therefore, it is desirable to devise a novel control scheme that does not need the exact system dynamics but only the input–output data which can be obtained during the operation of the system. In this paper, we propose a data-driven robust optimal tracking control scheme using the ADP method for unknown general nonlinear continuous-time systems. Specifically, the design of proposed controller is divided into two steps: 1) establishing a data-driven model based on an RNN by using available input–output data to reconstruct the unknown system dynamics, and 2) designing the robust approximate optimal tracking controller based on the obtained data-driven model. In the following, the establishment of data-driven model and the controller design are discussed in detail.

III. Establishment of Data-Driven Model

In this section, a data-driven model is established based on an RNN to reconstruct the unknown system dynamics by using available input–output data. By adding a novel adjustable term related to the modeling error, the resultant modeling error is guaranteed to converge to zero.

To begin with the development, the system dynamics (1) is rewritten in the form of an RNN as follows [38]:

$$\dot{x}(t) = C^{*T}x(t) + A^{*T}f(x(t)) + C_u^{*T}u(t) + A_u^{*T} + \varepsilon(t) \tag{3}$$

where $\varepsilon(t)$ is assumed to be bounded, and C^{*T}, A^{*T}, C_u^{*T}, and A_u^{*T} are the unknown ideal weight matrices. The activation function $f(\cdot)$ is selected as a monotonically increasing function satisfying

$$0 \le f(x) - f(y) \le k(x - y) \tag{4}$$

for any $x, y \in \mathbb{R}$ and $x \ge y, k > 0$, such as $f(x) = \tanh(x)$.

Based on (3), the data-driven model is then constructed as

$$\dot{\hat{x}}(t) = \hat{C}^T(t)\hat{x}(t) + \hat{A}^T(t)f(\hat{x}(t)) + \hat{C}_u^T(t)u(t) + \hat{A}_u^T(t) - \upsilon(t) \tag{5}$$

where $\hat{x}(t)$ is the estimated system state vector, $\hat{C}(t)$, $\hat{A}(t)$, $\hat{C}_u(t)$, and $\hat{A}_u(t)$ are the estimates of the ideal weight matrices C^*, A^*, C_u^*, and A_u^*, respectively, and $\upsilon(t)$ is defined as

$$\upsilon(t) = Se_m(t) + \frac{\hat{\lambda}(t)e_m(t)}{e_m^T(t)e_m(t) + \eta} \tag{6}$$

where $e_m(t) = x(t) - \hat{x}(t)$ is the system modeling error, $S \in \mathbb{R}^{n\times n}$ is a designed matrix, $\hat{\lambda}(t) \in \mathbb{R}$ is an additional tunable parameter, and $\eta > 1$ is a constant.

Assumption 1 [39]: The term $\varepsilon(t)$ is assumed to be upper bounded by a function of modeling error such that

$$\varepsilon^T(t)\varepsilon(t) \le \varepsilon_M(t) = \lambda^* e_m^T(t)e_m(t) \tag{7}$$

where λ^* is the bounded constant target value.

The modeling error dynamics is written as

$$\begin{aligned}\dot{e}_m(t) &= C^{*T}e_m(t) + \tilde{C}^T(t)\hat{x}(t) + A^{*T}\tilde{f}(e_m(t)) \\ &\quad + \tilde{A}^T(t)f(\hat{x}(t)) + \tilde{C}_u^T(t)u(t) + \tilde{A}_u^T(t) + \varepsilon(t) \\ &\quad + Se_m(t) - \frac{\tilde{\lambda}(t)e_m(t)}{e_m^T(t)e_m(t)+\eta} + \frac{\lambda^* e_m(t)}{e_m^T(t)e_m(t)+\eta} \quad (8)\end{aligned}$$

where $\tilde{C}(t) = C^* - \hat{C}(t)$, $\tilde{A}(t) = A^* - \hat{A}(t)$, $\tilde{C}_u(t) = C_u^* - \hat{C}_u(t)$, $\tilde{A}_u(t) = A_u^* - \hat{A}_u(t)$, $\tilde{f}(e_m(t)) = f(x(t)) - f(\hat{x}(t))$, and $\tilde{\lambda}(t) = \lambda^* - \hat{\lambda}(t)$.

Theorem 1: The modeling error $e_m(t)$ will asymptotically converge to zero as $t \to \infty$ if the weight matrices and the tunable parameter of the data-driven model (5) are updated through the following equations:

$$\begin{aligned}\dot{\hat{C}}(t) &= \Gamma_1 \hat{x}(t)e_m^T(t) \\ \dot{\hat{A}}(t) &= \Gamma_2 f(\hat{x}(t))e_m^T(t) \\ \dot{\hat{C}}_u(t) &= \Gamma_3 u(t)e_m^T(t) \\ \dot{\hat{A}}_u(t) &= \Gamma_4 e_m^T(t) \\ \dot{\hat{\lambda}}(t) &= -\Gamma_5 \frac{e_m^T(t)e_m(t)}{e_m^T(t)e_m(t)+\eta} \quad (9)\end{aligned}$$

where Γ_i is a positive definite matrix such that $\Gamma_i = \Gamma_i^T > 0$, $i = 1, 2, \ldots, 5$.

Proof: Choose the following Lyapunov function candidate:

$$J(t) = J_1(t) + J_2(t) \quad (10)$$

where

$$\begin{aligned}J_1(t) &= \frac{1}{2}e_m^T(t)e_m(t) \\ J_2(t) &= \frac{1}{2}tr\left\{\tilde{C}^T(t)\Gamma_1^{-1}\tilde{C}(t) + \tilde{A}^T(t)\Gamma_2^{-1}\tilde{A}(t) \right. \\ &\quad \left. + \tilde{C}_u^T(t)\Gamma_3^{-1}\tilde{C}_u(t) + \tilde{A}_u^T(t)\Gamma_4^{-1}\tilde{A}_u(t)\right\} \\ &\quad + \frac{1}{2}\tilde{\lambda}^T(t)\Gamma_5^{-1}\tilde{\lambda}(t).\end{aligned}$$

Then the time derivative of the Lyapunov function candidate (10) along the trajectories of the error system (8) is computed as follows:

$$\begin{aligned}\dot{J}_1(t) &= e_m^T(t)C^{*T}e_m(t) + e_m^T(t)\tilde{C}^T(t)\hat{x}(t) \\ &\quad + e_m^T(t)A^{*T}\tilde{f}(e_m(t)) + e_m^T(t)\tilde{A}^T(t)f(\hat{x}(t)) \\ &\quad + e_m^T(t)\tilde{C}_u^T(t)u(t) + e_m^T(t)\tilde{A}_u^T(t) + e_m^T(t)\varepsilon(t) \\ &\quad + e_m^T(t)Se_m(t) - \frac{e_m^T(t)\tilde{\lambda}(t)e_m(t)}{e_m^T(t)e_m(t)+\eta} \\ &\quad + \frac{e_m^T(t)\lambda^* e_m(t)}{e_m^T(t)e_m(t)+\eta}. \quad (11)\end{aligned}$$

From (4), we can obtain

$$\begin{aligned}e_m^T(t)A^{*T}\tilde{f}(e_m(t)) &\le \frac{1}{2}e_m^T(t)A^{*T}A^*e_m(t) \\ &\quad + \frac{1}{2}k^2 e_m^T(t)e_m(t). \quad (12)\end{aligned}$$

According to Assumption 1, we have

$$\begin{aligned}e_m^T(t)\varepsilon(t) &\le \frac{1}{2}e_m^T(t)e_m(t) + \frac{1}{2}\varepsilon^T(t)\varepsilon(t) \\ &\le \frac{1}{2}e_m^T(t)e_m(t) + \frac{1}{2}\lambda^* e_m^T(t)e_m(t). \quad (13)\end{aligned}$$

Therefore (11) can be rewritten as

$$\begin{aligned}\dot{J}_1(t) &\le e_m^T(t)C^{*T}e_m(t) + e_m^T(t)\tilde{C}^T(t)\hat{x}(t) \\ &\quad + \frac{1}{2}e_m^T(t)A^{*T}A^*e_m(t) \\ &\quad + \left(\frac{1}{2} + \frac{1}{2}\lambda^* + \frac{1}{2}k^2\right)e_m^T(t)e_m(t) \\ &\quad + e_m^T(t)\tilde{A}^T(t)f(\hat{x}(t)) \\ &\quad + e_m^T(t)\tilde{C}_u^T(t)u(t) + e_m^T(t)\tilde{A}_u^T(t) \\ &\quad + e_m^T(t)Se_m(t) - \frac{e_m^T(t)\tilde{\lambda}(t)e_m(t)}{e_m^T(t)e_m(t)+\eta} \\ &\quad + \frac{e_m^T(t)\lambda^* e_m(t)}{e_m^T(t)e_m(t)+\eta}. \quad (14)\end{aligned}$$

Computing the time derivative of $J_2(t)$ yields

$$\begin{aligned}\dot{J}_2(t) &= tr\left\{\tilde{C}^T(t)\Gamma_1^{-1}\dot{\tilde{C}}(t) + \tilde{A}^T(t)\Gamma_2^{-1}\dot{\tilde{A}}(t) \right. \\ &\quad \left. + \tilde{C}_u^T(t)\Gamma_3^{-1}\dot{\tilde{C}}_u(t) + \tilde{A}_u^T(t)\Gamma_4^{-1}\dot{\tilde{A}}_u(t)\right\} \\ &\quad + \tilde{\lambda}^T(t)\Gamma_5^{-1}\dot{\tilde{\lambda}}(t). \quad (15)\end{aligned}$$

Combining (14) with (15), we have

$$\begin{aligned}\dot{J}(t) &\le e_m^T(t)C^{*T}e_m(t) + \frac{1}{2}e_m^T(t)A^{*T}A^*e_m(t) \\ &\quad + e_m^T(t)\left(\left(\frac{1}{2} + \frac{1}{2}\lambda^* + \frac{1}{2}k^2\right)I_n + S\right)e_m(t) \\ &\quad + \frac{e_m^T(t)\lambda^* e_m(t)}{e_m^T(t)e_m(t)+\eta} \\ &\le e_m^T(t)\Xi e_m(t) \quad (16)\end{aligned}$$

where I_n denotes a $n \times n$ identity matrix and

$$\Xi = C^{*T} + \frac{1}{2}A^{*T}A^* + \left(\frac{1}{2} + \frac{3}{2}\lambda^* + \frac{1}{2}k^2\right)I_n + S.$$

S is selected to make $\Xi < 0$. Therefore, it can be concluded that $\dot{J}(t) < 0$. Since $J(t) > 0$, it follows from [40] that $e_m(t) \to 0$ as $t \to \infty$.

This completes the proof. ■

Remark 1: According to the results of Theorem 1, since $e_m(t) \to 0$ as $t \to \infty$, the term $\upsilon(t) \to 0$ as $t \to \infty$. In addition, $\dot{\hat{C}}(t) \to 0$, $\dot{\hat{A}}(t) \to 0$, $\dot{\hat{C}}_u(t) \to 0$, and $\dot{\hat{A}}_u(t) \to 0$ as $e_m(t) \to 0$. It means that $\hat{C}(t)$, $\hat{A}(t)$, $\hat{C}_u(t)$, and $\hat{A}_u(t)$ all tend to be constant matrices, which are denoted as C, A, C_u, and A_u, respectively.

Consequently, the nonlinear system (1) can be rewritten as

$$\dot{x}(t) = C^T x(t) + A^T f(x(t)) + C_u^T u(t) + A_u^T. \quad (17)$$

The original optimal tracking problem of (1) is transformed into the optimal tracking problem of (17). In the following sections, the controller design based on (17) will be given in detail.

IV. ADP-Based Approximate Optimal Tracking Controller Design

Assume that the desired trajectory $x_d(t)$ has the following form:

$$\dot{x}_d(t) = C^T x_d(t) + A^T f(x_d(t)) + C_u^T u_d(t) + A_u^T \quad (18)$$

where $u_d(t)$ is the control input of the desired system.

By using (17) and (18), the error system can be formulated as

$$\dot{e}(t) = C^T e(t) + A^T f_e(t) + C_u^T u_e(t) \quad (19)$$

where $f_e(t) = f(x(t)) - f(x_d(t))$ and $u_e(t) = u(t) - u_d(t)$. It is noted that the controller $u(t)$ is composed of two parts, i.e., the steady-state controller $u_d(t)$, and the feedback controller $u_e(t)$.

The steady-state controller $u_d(t)$ can be obtained from (18) as

$$u_d(t) = C_u^{-T}\left(\dot{x}_d(t) - C^T x_d(t) - A^T f(x_d(t)) - A_u^T\right) \quad (20)$$

where C_u^{-1} stands for the pseudo-inverse of C_u, which is used to maintain the tracking error close to zero at the steady-state stage. From (20), $u_d(t)$ is related to C_u, C, A, A_u, $x_d(t)$, and $\dot{x}_d(t)$. And it can be computed directly.

Next, the feedback controller $u_e(t)$ is designed to stabilize the state tracking error dynamics at transient stage in an optimal manner. In the following, for brevity, the notations $e(t)$, $u_d(t)$, $u_e(t)$, $u(t)$, and $V(e(t))$ are rewritten as e, u_d, u_e, u, and $V(e)$.

The infinite horizon performance index function (2) is transformed into

$$V(e) = \int_t^\infty r(e(\tau), u_e(e(\tau)))d\tau \quad (21)$$

where $r(e, u_e) = e^T Qe + u_e^T Ru_e$ is the utility function, and Q and R are symmetric positive definite matrices with appropriate dimensions.

It is desirable to find the optimal feedback control u_e^* that stabilizes (19) and minimizes the performance index (21). This kind of control is called admissible control.

Definition 1 [2]: A control policy $\mu(e)$ is defined as admissible with respect to (21) on Ω, denoted by $\mu(e) \in \psi(\Omega)$, if $\mu(e)$ is continuous on Ω, $\mu(0) = 0$, $\mu(e)$ stabilizes (19) on Ω, and $V(e)$ is finite, $\forall e \in \Omega$.

Define the Hamiltonian function as

$$H(e, u_e, V_e) = V_e^T\left(C^T e + A^T f_e + C_u^T u_e\right) + e^T Qe + u_e^T Ru_e \quad (22)$$

where $V_e = \partial V(e)/\partial e$.

The optimal cost function $V^*(e)$ is defined as

$$V^*(e) = \min_{u_e \in \psi(\Omega)}\left(\int_t^\infty r(e(\tau), u_e(e(\tau)))d\tau\right) \quad (23)$$

and satisfies

$$0 = \min_{u_e \in \psi(\Omega)}[H(e, u_e, V_e^*)]. \quad (24)$$

Further, we can obtain the optimal control u_e^* by solving $\partial H(e, u_e, V_e)/\partial u_e = 0$ as

$$u_e^* = -\frac{1}{2}R^{-1}C_u V_e^* \quad (25)$$

where $V_e^* = \partial V^*(e)/\partial e$. Then the overall optimal control input can be rewritten as $u^* = u_d + u_e^*$.

In the following, we will focus on the optimal feedback controller design using the ADP method, which is implemented by employing the critic NN and action NN.

A. Critic NN Design

A NN is utilized to approximate $V(e)$ as follows:

$$V(e) = W_1^T \phi_1(e) + \varepsilon_1 \quad (26)$$

where W_1 is the unknown ideal constant weights and $\phi_1(e): \mathbb{R}^n \to \mathbb{R}^{N_1}$ is called the critic NN activation function vector, N_1 is the number of neurons in the hidden layer, and ε_1 is the critic NN approximation error.

The derivative of the cost function $V(e)$ with respect to e is

$$V_e = \nabla\phi_1^T W_1 + \nabla\varepsilon_1 \quad (27)$$

where $\nabla\phi_1 \triangleq \partial\phi_1(e)/\partial e$ and $\nabla\varepsilon_1 \triangleq \partial\varepsilon_1/\partial e$.

Let $\hat{W}_1$ be an estimate of W_1, then we have the estimate of $V(e)$ as

$$\hat{V}(e) = \hat{W}_1^T \phi_1(e). \quad (28)$$

Then, the approximate Hamiltonian function can be derived as follows:

$$\begin{aligned} &H\left(e, u_e, \hat{W}_1\right) \\ &= \hat{W}_1^T \nabla\phi_1\left(C^T e + A^T f_e + C_u^T u_e\right) + e^T Qe + u_e^T Ru_e \\ &= e_1. \end{aligned} \quad (29)$$

Given any admissible control policy u_e, it is desired to select $\hat{W}_1$ to minimize the squared residual error $E_1(\hat{W}_1)$ as

$$E_1(\hat{W}_1) = \frac{1}{2}e_1^T e_1. \quad (30)$$

The weight update law for the critic NN is a gradient descent algorithm, which is given by

$$\dot{\hat{W}}_1 = -a_1\sigma_1\left(\sigma_1^T \hat{W}_1 + e^T Qe + u_e^T Ru_e\right) \quad (31)$$

where $a_1 > 0$ is the adaptive gain of the critic NN, $\sigma_1 = \sigma/(\sigma^T\sigma + 1)$, $\sigma = \nabla\phi_1(C^T e + A^T f_e + C_u^T u_e)$. According to the definition of σ_1, there exists a positive constant $\sigma_{1M} > 1$ such that $||\sigma_1|| \le \sigma_{1M}$. Define the weight estimation error of critic NN to be $\tilde{W}_1 = \hat{W}_1 - W_1$, and note that, for a fixed control policy u_e, the Hamiltonian function (22) becomes

$$\begin{aligned} &H(e, u_e, W_1) \\ &= W_1^T \nabla\phi_1\left(C^T e + A^T f_e + C_u^T u_e\right) + e^T Qe + u_e^T Ru_e \\ &= \varepsilon_{HJB} \end{aligned} \quad (32)$$

where the residual error due to the NN approximation is $\varepsilon_{HJB} = -\nabla\varepsilon_1(C^T e + A^T f_e + C_u^T u_e)$.

Rewriting (31) by using (32), we have

$$\dot{\tilde{W}}_1 = -a_1\sigma_1\left(\sigma_1^T \tilde{W}_1 + \varepsilon_{HJB}\right). \quad (33)$$

B. Action NN Design

To begin the development of the feedback control policy, u_e is approximated by the action NN as

$$u_e = W_2^T \phi_2(e) + \varepsilon_2 \tag{34}$$

where W_2 is the matrix of unknown ideal constant weights and $\phi_2(e): \mathbb{R}^n \to \mathbb{R}^{N_2}$ is called the action NN activation function vector, N_2 is the number of neurons in the hidden layer, and ε_2 is the action NN approximation error.

Let $\hat{W}_2$ be an estimate of W_2, the actual output can be expressed as

$$\hat{u}_e = \hat{W}_2^T \phi_2(e). \tag{35}$$

The feedback error signal used for tuning action NN is defined to be the difference between the feedback control input applied to the error system (19) and the control input minimizing (28) as

$$e_2 = \hat{W}_2^T \phi_2 + \frac{1}{2} R^{-1} C_u \nabla \phi_1^T \hat{W}_1. \tag{36}$$

The objective function to be minimized by the action NN is defined as

$$E_2(\hat{W}_2) = \frac{1}{2} e_2^T e_2. \tag{37}$$

The weight update law for the action NN is a gradient descent algorithm, which is given by

$$\dot{\hat{W}}_2 = -a_2 \phi_2 \left(\hat{W}_2^T \phi_2 + \frac{1}{2} R^{-1} C_u \nabla \phi_1^T \hat{W}_1 \right)^T \tag{38}$$

where $a_2 > 0$ is the adaptive gain of the action NN.

Define the weight estimation error of action NN to be $\tilde{W}_2 = \hat{W}_2 - W_2$. Since the control policy in (34) minimizes the infinite horizon performance index function (26), from (25) we have

$$\varepsilon_2 + W_2^T \phi_2 + \frac{1}{2} R^{-1} C_u \nabla \phi_1^T W_1 + \frac{1}{2} R^{-1} C_u \nabla \varepsilon_1 = 0. \tag{39}$$

Combining (38) with (39), we have

$$\dot{\tilde{W}}_2 = -a_2 \phi_2 \left(\tilde{W}_2^T \phi_2 + \frac{1}{2} R^{-1} C_u \nabla \phi_1^T \tilde{W}_1 + \varepsilon_{12} \right)^T \tag{40}$$

where $\varepsilon_{12} = -(\varepsilon_2 + R^{-1} C_u \nabla \varepsilon_1 / 2)$.

Remark 2: It is important to note that the tracking error must be persistently excited sufficiently for tuning critic NN and action NN. In order to satisfy the persistent excitation condition, probing noise is added to the control input [6]. Further, the persistent excitation condition ensures $\|\sigma_1\| \geq \sigma_{1m}$ and $\|\phi_2\| \geq \phi_{2m}$, with σ_{1m} and ϕ_{2m} being positive constants.

C. Stability Analysis

Based on the above analysis, the optimal tracking controller is composed of the steady-state controller u_d and the optimal feedback controller u_e. As a result, the control input is written as

$$u = u_d + \hat{u}_e. \tag{41}$$

According to (35) and the error system (19), we have

$$\dot{e} = C^T e + A^T f_e + C_u^T \hat{W}_2^T \phi_2. \tag{42}$$

Subtracting and adding $C_u^T W_2 \phi_2$ to (42), and recalling (34), (42) is rewritten as

$$\dot{e} = C^T e + A^T f_e + C_u^T \tilde{W}_2^T \phi_2 + C_u^T u_e - C_u^T \varepsilon_2. \tag{43}$$

In the following, the stability analysis will be performed. First, the following assumption is made, which can reasonably be satisfied under the current problem settings.

Assumption 2:

1) The unknown ideal constant weights for the critic NN and the action NN, i.e., W_1 and W_2, are upper bounded so that $\|W_1\| \leq W_{1M}$, $\|W_2\| \leq W_{2M}$, respectively.
2) The NN approximation errors ε_1 and ε_2 are upper bounded so that $\|\varepsilon_1\| \leq \varepsilon_{1M}$, $\|\varepsilon_2\| \leq \varepsilon_{2M}$, respectively.
3) The vectors of the activation functions of the critic NN and the action NN, i.e., ϕ_1 and ϕ_2, are upper bounded so that $\|\phi_1(\cdot)\| \leq \phi_{1M}$, $\|\phi_2(\cdot)\| \leq \phi_{2M}$, respectively.
4) The gradients of the critic NN approximation error and the activation function vector are upper bounded so that $\|\nabla \varepsilon_1\| \leq \varepsilon'_{1M}$, $\|\nabla \phi_1\| \leq \phi_{dM}$. And the residual error is upper bounded so that $\|\varepsilon_{HJB}\| \leq \varepsilon_{HJBM}$.

Now we are ready to prove the following theorem.

Theorem 2: Consider the system given by (17) and the desired trajectory given by (18). Let the control input be provided by (41). The weight updating laws of the critic NN and the action NN are given by (31) and (38), respectively. And let the initial action NN weights be chosen to generate an initial admissible control. Then the tracking error e and the weight estimate errors $\tilde{W}_1$ and $\tilde{W}_2$ are UUB with the bounds specifically given by (53)–(55). Moreover, the obtained control input u is close to the optimal control input u^* within a small bound ε_u, i.e., $\|u - u^*\| \leq \varepsilon_u$ as $t \to \infty$ for a small positive constant ε_u.

Proof: Choose the following Lyapunov function candidate:

$$L(t) = L_1(t) + L_2(t) + L_3(t) \tag{44}$$

where $L_1(t) = \mathrm{tr}\{\tilde{W}_1^T \tilde{W}_1\}/2a_1$ $L_2(t) = a_1 \mathrm{tr}\{\tilde{W}_2^T \tilde{W}_2\}/2a_2$ and $L_3(t) = a_1 a_2 (e^T e + \Gamma V(e))$ with $\Gamma > 0$.

According to Assumptions 1 and 2 and using (21), (33), and (40), the time derivative of the Lyapunov function candidate (44) along the trajectories of the error system (43) is computed as

$$\dot{L}(t) = \dot{L}_1(t) + \dot{L}_2(t) + \dot{L}_3(t) \tag{45}$$

where

$$\begin{aligned} \dot{L}_1(t) &= \frac{1}{a_1} \mathrm{tr} \left\{ \tilde{W}_1^T \dot{\tilde{W}}_1 \right\} \\ &= \frac{1}{a_1} \mathrm{tr} \left\{ \tilde{W}_1^T [-a_1 \sigma_1 (\sigma_1^T \tilde{W}_1 + \varepsilon_{HJB})] \right\} \\ &\leq -\left(\sigma_{1m}^2 - \frac{a_1}{2} \sigma_{1M}^2 \right) \|\tilde{W}_1\|^2 + \frac{1}{2a_1} \varepsilon_{HJB}^2 \end{aligned} \tag{46}$$

$$\begin{aligned} \dot{L}_2(t) &= \frac{a_1}{a_2} tr\{\tilde{W}_2^T \dot{\tilde{W}}_2\} \\ &= \frac{a_1}{a_2} tr \left\{ \tilde{W}_2^T [-a_2 \phi_2 (\tilde{W}_2^T \phi_2 \right. \end{aligned}$$

$$+\frac{1}{2}R^{-1}C_u\nabla\phi_1^T\tilde{W}_1+\varepsilon_{12})^T]\Big\}$$
$$\leq -\left(a_1\phi_{2m}^2-\frac{3}{4}a_1a_2\phi_{2M}^2\right)\|\tilde{W}_2\|^2 + \frac{a_1}{4a_2}\|R^{-1}\|^2\|C_u\|^2\phi_{dM}^2\|\tilde{W}_1\|^2+\frac{a_1}{2a_2}\varepsilon_{12}^T\varepsilon_{12} \quad (47)$$

$$\dot{L}_3(t) = 2a_1a_2e^T\dot{e}+a_1a_2\Gamma\left(-e^TQe-u_e^TRu_e\right)$$
$$= 2a_1a_2e^T\left(C^Te+A^Tf_e+C_u^T\tilde{W}_2^T\phi_2+C_u^Tu_e - C_u^T\varepsilon_2\right)+a_1a_2\Gamma\left(-e^TQe-u_e^TRu_e\right)$$
$$\leq a_1a_2(2\|C\|+3+\|A\|^2+k^2-\Gamma\lambda_{\min}(Q))\|e\|^2 + a_1a_2\phi_{2M}^2\|C_u\|^2\|\tilde{W}_2\|^2+a_1a_2\|C_u\|^2\varepsilon_2^T\varepsilon_2 + a_1a_2(\|C_u\|^2-\Gamma\lambda_{\min}(R))\|u_e\|^2.$$

Then

$$\dot{L}(t) \leq -\left(\sigma_{1m}^2-\frac{a_1}{2}\sigma_{1M}^2-\frac{a_1}{4a_2}\|R^{-1}\|^2\|C_u\|^2\phi_{dM}^2\right)\|\tilde{W}_1\|^2 - \left(a_1\phi_{2m}^2-\frac{3}{4}a_1a_2\phi_{2M}^2-a_1a_2\phi_{2M}^2\|C_u\|^2\right)\|\tilde{W}_2\|^2 - a_1a_2\left(-\|C_u\|^2+\Gamma\lambda_{\min}(R)\right)\|u_e\|^2 - a_1a_2\left(-2\|C\|-3-\|A\|^2-k^2+\Gamma\lambda_{\min}(Q)\right)\|e\|^2 + \frac{1}{2a_1}\varepsilon_{HJB}^2+\frac{a_1}{2a_2}\varepsilon_{12}^T\varepsilon_{12}+a_1a_2\|C_u\|^2\varepsilon_2^T\varepsilon_2. \quad (48)$$

By using Assumption 2, we have $\|\varepsilon_{12}\|\leq\varepsilon_{12M}$, where $\varepsilon_{12M}=\varepsilon_{2M}+R^{-1}C_u\varepsilon'_{1M}/2$. Then

$$\frac{1}{2a_1}\varepsilon_{HJB}^2+\frac{a_1}{2a_2}\varepsilon_{12}^T\varepsilon_{12}+a_1a_2\|C_u\|^2\varepsilon_{12}^T\varepsilon_{12}\leq D_M \quad (49)$$

where

$$D_M=\frac{\varepsilon_{HJBM}^2}{2a_1}+\frac{a_1\varepsilon_{12M}^2}{2a_2}+a_1a_2\|C_u\|^2\varepsilon_{12M}^2.$$

If Γ, a_1, and a_2 are selected to satisfy

$$\Gamma > \max\left\{\frac{\|C_u\|^2}{\lambda_{\min}(R)},\frac{2\|C\|+3+\|A\|^2+k^2}{\lambda_{\min}(Q)}\right\} \quad (50)$$

$$a_1 < \frac{4a_2\sigma_{1m}^2}{2a_2\sigma_{1M}^2+\|R^{-1}\|^2\|C_u\|^2\phi_{dM}^2} \quad (51)$$

$$a_2 < \frac{4\phi_{2m}^2}{3\phi_{2M}^2+4\phi_{2M}^2\|C_u\|^2} \quad (52)$$

and given that the inequalities

$$\|e\| > \sqrt{\frac{D_M}{a_1a_2(-2\|C\|-3-\|A\|^2-k^2+\Gamma\lambda_{\min}(Q))}} \triangleq b_e \quad (53)$$

or

$$\|\tilde{W}_1\| > \sqrt{\frac{D_M}{\sigma_{1m}^2-\frac{a_1}{2}\sigma_{1M}^2-\frac{a_1}{4a_2}\|R^{-1}\|^2\|C_u\|^2\phi_{dM}^2}} \triangleq b_{\tilde{W}_1} \quad (54)$$

or

$$\|\tilde{W}_2\| > \sqrt{\frac{D_M}{a_1\phi_{2m}^2-\frac{3}{4}a_1a_2\phi_{2M}^2-a_1a_2\phi_{2M}^2\|C_u\|^2}} \triangleq b_{\tilde{W}_2} \quad (55)$$

hold, then $\dot{L}(t)<0$. Therefore, using Lyapunov theory [41], it can be concluded that the tracking error e and the NN weight estimation errors $\tilde{W}_1$ and $\tilde{W}_2$ are UUB.

Next we will prove $\|u-u^*\|\leq\varepsilon_u$ as $t\to\infty$. Recalling the expression of u^* together with (34) and (41), we have

$$u-u^*=\tilde{W}_2^T\phi_2+\varepsilon_2. \quad (56)$$

When $t\to\infty$, the upper bound of (56) is

$$\|u-u^*\|\leq\varepsilon_u \quad (57)$$

where $\varepsilon_u=b_{\tilde{W}_2}\phi_{2M}+\varepsilon_{2M}$.

This completes the proof. ■

Remark 3: From (31) and (38), it is noted that the weights of critic NN and action NN are updated simultaneously in contrast to some standard ADP methods in which the weights of critic NN and action NN are updated sequentially.

Remark 4: If the NN approximation errors ε_1 and ε_2 are considered to be negligible, then from (49) we have $D_M=0$, with $u\to u^*$. Otherwise, the obtained control input u is close to the optimal input u^* within a small bound ε_u.

Due to the presence of the NN approximation errors ε_1 and ε_2, the tracking error is UUB instead of asymptotically converging to zero. In the following section, for improving the tracking performance, an additional robustifying term is developed to attenuate the NN approximation errors such that tracking error converges to zero asymptotically.

V. ADP-Based Robust Approximate Optimal Tracking Controller Design

In this section, a robustifying term is developed to compensate for the NN approximation errors in order to obtain asymptotic tracking results, which can be constructed in the form

$$u_r=\frac{K_re}{e^Te+\zeta} \quad (58)$$

where $\zeta>0$ is a constant, and $K_r>K_{r\min}$ is a designed parameter. $K_{r\min}$ is selected to satisfy the following inequality:

$$K_{r\min}\geq\frac{D_M\left(e^Te+\zeta\right)}{2a_1a_2\|C_u\|e^Te}. \quad (59)$$

Then the overall control input is given as

$$u_{ad}=u-u_r \quad (60)$$

where u is the same as (41).

Applying (60) to the error system (17) and using (18), a new error system is obtained as

$$\dot{e}=C^Te+A^Tf_e+C_u^T\tilde{W}_2^T\phi_2+C_u^Tu_e-C_u^T\varepsilon_2-C_u^Tu_r. \quad (61)$$

Theorem 3: Consider the system given by (17) and the desired trajectory given by (18). Let the control input be provided by (60). The weight updating laws of the critic NN and the action NN are given by (31) and (38), respectively.

And let the initial action NN weights be chosen to generate an initial admissible control. Then the tracking error e and the weight estimation errors $\tilde{W}_1$ and $\tilde{W}_2$ will asymptotically converge to zero. Moreover, the obtained control input u_{ad} is close to the optimal control input u^* within a small bound δ_u, i.e., $\|u_{ad} - u^*\| \leq \delta_u$ as $t \to \infty$ for a small positive constant δ_u.

Proof: Choose the same Lyapunov function candidate as in Theorem 2. Differentiating the Lyapunov function candidate in (44) along the trajectories of the error system in (61), similar to the proof of Theorem 2, by using (58) and (59), we can obtain

$$\begin{aligned}\dot{L}(t) \leq & -\left(\sigma_{1m}^2 - \frac{a_1}{2}\sigma_{1M}^2 - \frac{a_1}{4a_2}\|R^{-1}\|^2\|C_u\|^2\phi_{dM}^2\right)\|\tilde{W}_1\|^2 \\ & -\left(a_1\phi_{2m}^2 - \frac{3}{4}a_1a_2\phi_{2M}^2 - a_1a_2\phi_{2M}^2\|C_u\|^2\right)\|\tilde{W}_2\|^2 \\ & - a_1a_2\left(-\|C_u\|^2 + \Gamma\lambda_{\min}(R)\right)\|u_e\|^2 \\ & - a_1a_2(-2\|C\| - 3 - \|A\|^2 - k^2 + \Gamma\lambda_{\min}(Q))\|e\|^2. \end{aligned} \tag{62}$$

Choosing Γ, a_1, and a_2 as in Theorem 2, we have $\dot{L}(t) \leq 0$. Equations (44) and (62) guarantee that the tracking error e and NN weight estimation errors $\tilde{W}_1$ and $\tilde{W}_2$ are bounded since L is nonincreasing. Because all the variables on the right-hand side of (61) are bounded, $\dot{e}$ is also bounded. From (62), we have

$$\dot{L}(t) \leq -B_e\|e\|^2 \tag{63}$$

where $B_e = a_1a_2(-2\|C\| - 3 - \|A\|^2 - k^2 + \Gamma\lambda_{\min}(Q))$.

Integrating both sides of (63) and after some manipulations, we have

$$\int_0^\infty \|e\|^2 dt \leq B_e^{-1}(L(0) - L(\infty)). \tag{64}$$

Since the right side of (61) is bounded, $\|e\| \in \mathcal{L}_2$. Using Barbalat's lemma [41], we have $\lim_{t\to\infty}\|e\| = 0$. Similarly, we can prove that $\lim_{t\to\infty}\|\tilde{W}_1\| = 0$ and $\lim_{t\to\infty}\|\tilde{W}_2\| = 0$.

Next we will prove $\|u_{ad} - u^*\| \leq \delta_u$ as $t \to \infty$. From (34) and (60), we have

$$u_{ad} - u^* = \tilde{W}_2^T\phi_2 + \varepsilon_2 + u_r. \tag{65}$$

Since $\|e\| \to 0$ as $t \to \infty$, the robustifying control input $\|u_r\| \to 0$ as $t \to \infty$. Then the upper bound of (65) is

$$\|u_{ad} - u^*\| \leq \delta_u \tag{66}$$

where $\delta_u = \varepsilon_{2M}$.

This completes the proof. ∎

Remark 5: From (57) and (66), it can be seen that δ_u is smaller than ε_u, which reveals the role of the robustifying term in making the obtained control input closer to the optimal control input.

Remark 6: It is noted that the controller design is based on the result of the data-driven model established in this paper. During the process of modeling, sufficient learning time is needed to guarantee that the modeling error converges to zero asymptotically, which may be the weakness of this paper. Next, our research efforts will be directed toward implementing the system modeling and controller design simultaneously with the purpose of suiting for fast changing dynamics.

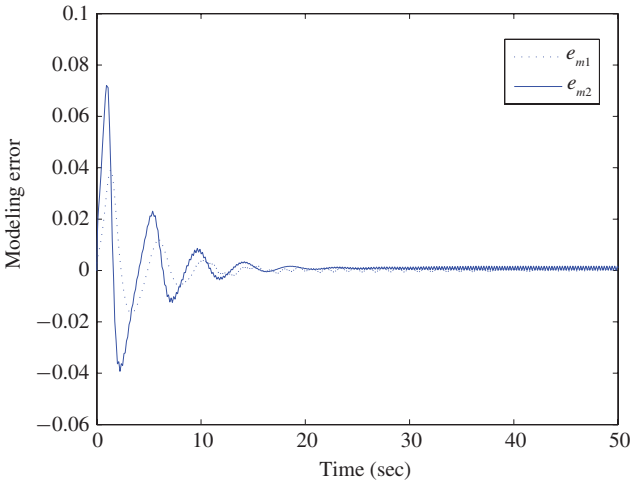

Fig. 1. Modeling error for the affine nonlinear system.

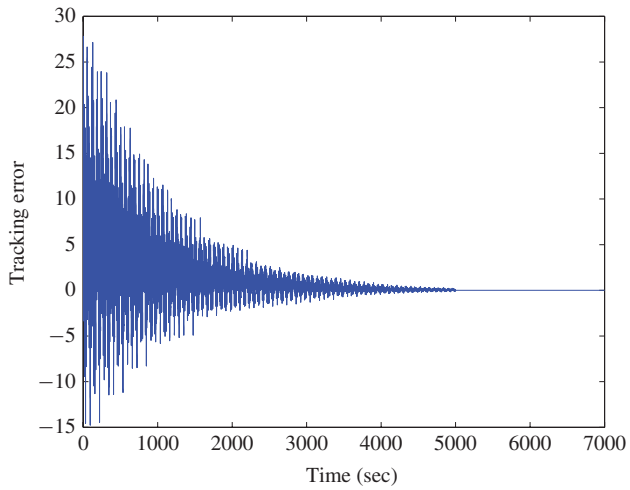

Fig. 2. Tracking error for the affine nonlinear system.

VI. Simulation

In this section, two examples are provided to demonstrate the effectiveness of the proposed approach.

Example 1: Consider the following affine nonlinear continuous-time system:

$$\begin{aligned}\dot{x}_1 &= -x_1 + x_2 \\ \dot{x}_2 &= -0.5x_1 - 0.5x_2\left(1 - (\cos(2x_1) + 2)^2\right) \\ &\quad + (\cos(2x_1) + 2)u. \end{aligned} \tag{67}$$

The performance index function is defined by (21), where the Q and R are chosen as identity matrices of appropriate dimensions. The control objective is to make x_1 follow the desired trajectory $x_{1d} = \sin(t)$. It is assumed that the system dynamics is unknown and input–output data are available.

At first, a data-driven model is established to estimate the nonlinear system dynamics. Let us select the RNN as (5) with $S = -30I_2$ and $\eta = 1.5$. The activation function $f(\hat{x})$ is selected as hyperbolic tangent function $\tanh(\hat{x})$. Select the

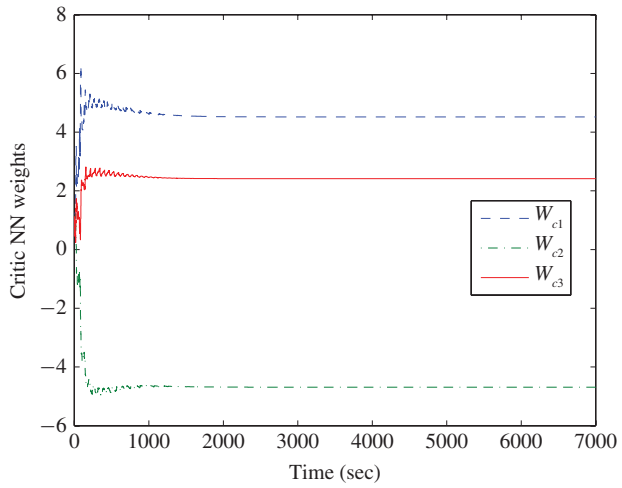

Fig. 3. Critic NN weights.

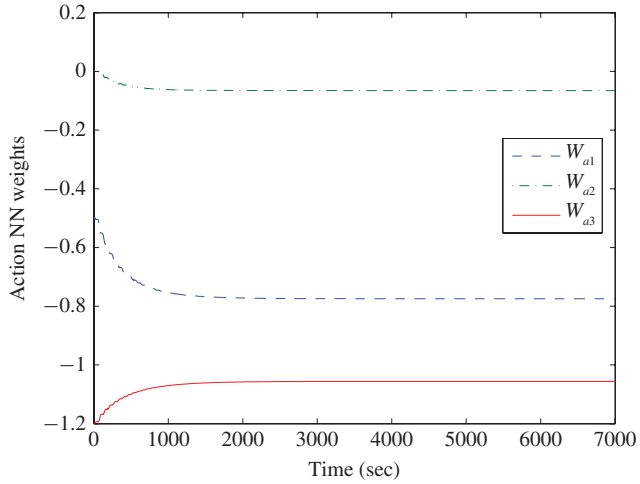

Fig. 4. Action NN weights.

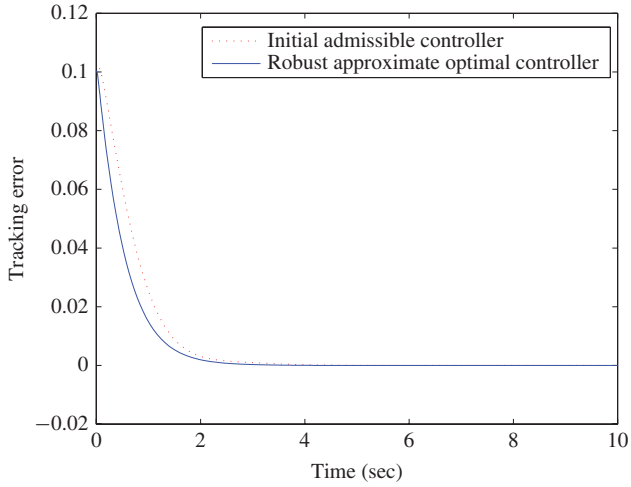

Fig. 5. Comparison result between initial admissible controller and robust approximate optimal controller.

design parameters in Theorem 1 as $\Gamma_1 = [1, 0.1; 0.1, 1]$, $\Gamma_2 = [1, 0.2; 0.2, 1]$, $\Gamma_3 = [1, 0.1; 0.1, 1]$, $\Gamma_4 = 0.2$, and $\Gamma_5 = 0.1$. Then, we can obtain the curves of the modeling error as shown in Fig. 1. It is observed that obtained data-driven model can reconstruct the nonlinear system dynamics successfully as Theorem 1 predicted.

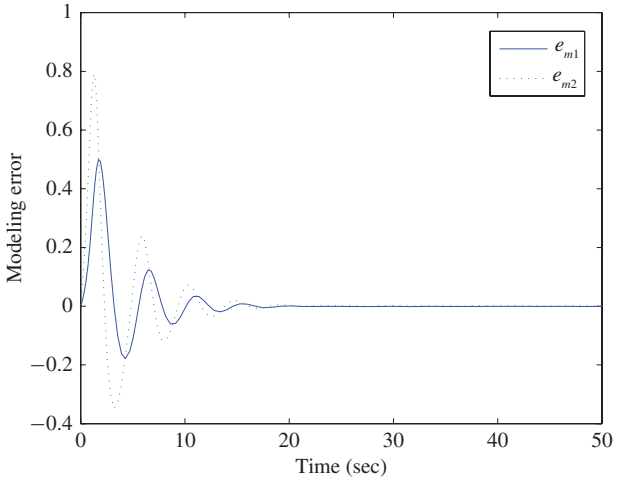

Fig. 6. Modeling error for the nonaffine nonlinear system.

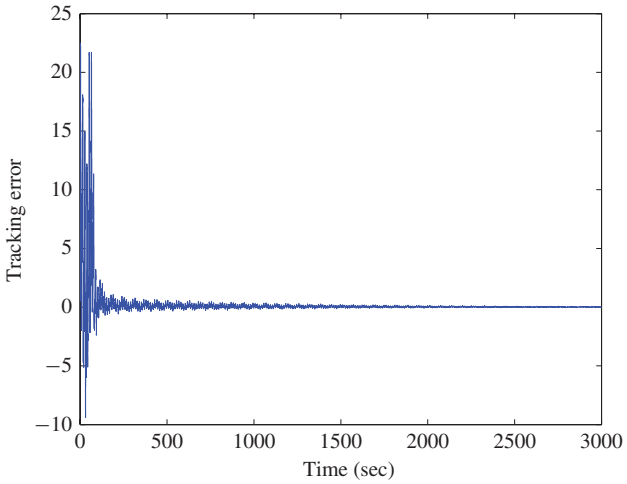

Fig. 7. Tracking error for the nonaffine nonlinear system.

Then, based on the obtained data-driven model, the approximate robust optimal controller is implemented for the unknown affine nonlinear continuous-time system (67). The activation function of critic NN is selected as $\phi_1 = [x_1^2 \; x_1x_2 \; x_2^2]^T$, and the critic NN weights are denoted as $\hat{W}_1 = [W_{c1} \; W_{c2} \; W_{c3}]^T$. The activation function of action NN ϕ_2 is chosen as the gradient of the critic NN, and the action NN weights are denoted as $\hat{W}_2 = [W_{a1} \; W_{a2} \; W_{a3}]^T$. The adaptive gains for the critic NN and action NN are selected as $a_1 = 0.8$ and $a_2 = 0.5$, and the design parameters of the robustifying term are selected as $K_r = [20, 20]$, and $\zeta = 1.2$. Additionally, the critic NN weights are set as $[1, 1, 1]^T$ at the beginning of the simulation with the initial weights of the action NN chosen to reflect the initial admissible control. To maintain the excitation condition, probing noise is added to the control input for the first 5000 s as in [6].

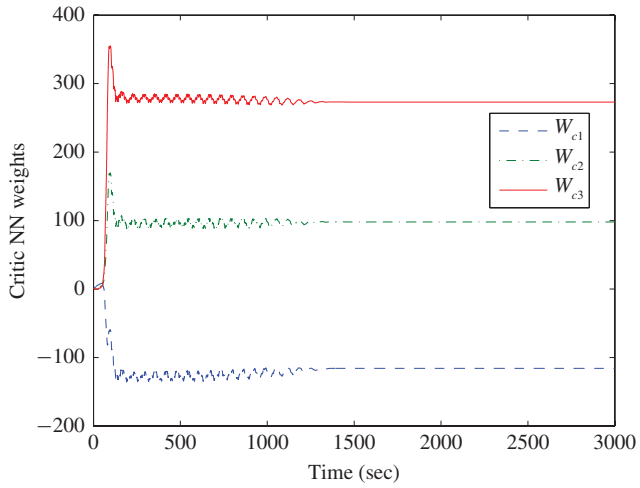

Fig. 8. Critic NN weights.

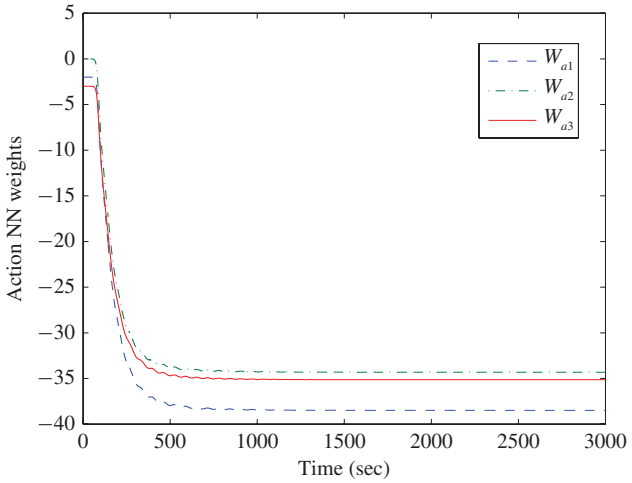

Fig. 9. Action NN weights.

After simulation, the curve of the tracking error is shown in Fig. 2. The convergence curves of the critic NN weights and action NN weights are shown in Figs. 3 and 4, respectively. For comparing the tracking performance, we apply the obtained robust optimal controller and initial admissible controller to (67) under the same initial state, and obtain the curves of tracking error as shown in Fig. 5. It can be seen from Fig. 5 that the proposed robust approximate optimal controller yields better tracking performance than the initial admissible controller.

Example 2: Consider the following nonaffine nonlinear continuous-time system:

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= x_1^2 + 0.15u^3 + 0.1(4 + x_2^2)u + \sin(0.1u). \end{aligned} \quad (68)$$

The performance index function is defined as in Example 1. The control objective is to make x_1 follow the desired trajectory $x_{1d} = \sin(t)$. It is assumed that the system dynamics is unknown and input–output data are available.

Using a similar method as in Example 1, we can obtain the curves of modeling error as shown in Fig. 6. It is observed

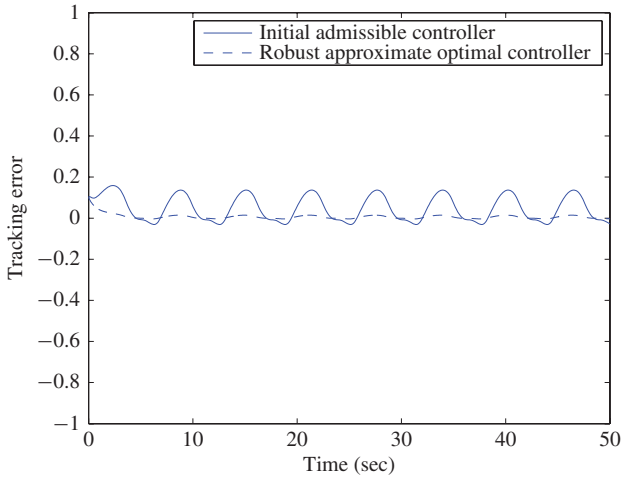

Fig. 10. Comparison result between initial admissible controller and robust approximate optimal controller.

that obtained data-driven model learns the nonlinear system dynamics successfully as Theorem 1 predicted. Then, based on the obtained data-driven model we design the robust approximate optimal controller, which is then applied to the unknown nonaffine nonlinear system (68). The activation functions of critic NN and action NN are the same as the ones in Example 1. The adaptive gains for the critic NN and action NN are selected as $a_1 = 0.5$ and $a_2 = 0.2$, and the parameters of the robustifying term are selected as $K_r = [10, 10], \zeta = 1.2$. Additionally, the critic NN weights are set as $[1, 1, 1]^T$ at the beginning of the simulation with the initial weights of the action NN chosen to reflect the initial admissible control. Similarly, to maintain the excitation condition, probing noise is added to the control input for the first 1500 s as in [6].

After simulation, the curve of the tracking error is shown in Fig. 7. The convergence curves of the critic NN weights and action NN weights are shown in Figs. 8 and 9, respectively. Similarly, for comparing the tracking performance, we apply obtained robust optimal controller and initial admissible controller to (68) under the same initial state, and obtain the curves of tracking error as shown in Fig. 10. It can be seen from Fig. 10 that the proposed robust approximate optimal controller yields better tracking performance than the initial admissible controller. The simulation results reveal that the proposed controller can be applied to nonaffine nonlinear systems and obtain satisfying tracking performance even for the unknown system dynamics.

VII. Conclusion

In this paper, we have proposed an effective scheme to design the data-driven robust optimal tracking controller for unknown general nonlinear systems. An RNN was employed to establish the data-driven model. Morever, based on the obtained model, the robust approximate optimal tracking controller was developed, which is composed of the steady-state controller for maintaining desired tracking at the steady-state stage, the feedback controller for stabilizing the state tracking

error dynamics at transient stage based on ADP method, and a robustifying term for attenuating the NN approximation errors. The simulation results have demonstrated the validity of the proposed data-driven robust approximate optimal tracking control scheme.

REFERENCES

[1] F. L. Lewis and V. Syrmos, *Optimal Control*. New York: Wiley, 1995.

[2] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, May 2005.

[3] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern., Part C: Appl. Rev.*, vol. 32, no. 2, pp. 140–153, May 2002.

[4] R. Beard, G. Saridis, and J. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, Dec. 1997.

[5] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 38, no. 4, pp. 943–949, Aug. 2008.

[6] 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.

[7] F.-Y. Wang, N. Jin, D. Liu, and Q. Wei, "Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ε-error bound," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 24–36, Jan. 2011.

[8] T. Hanselmann, L. Noakes, and A. Zaknich, "Continuous-time adaptive critics," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 631–647, May 2007.

[9] H. Zhang, Q. Wei, and D. Liu, "An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games," *Automatica*, vol. 47, no. 1, pp. 207–214, Jan. 2011.

[10] 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, nos. 5–6, pp. 851–860, Jul.–Aug. 2009.

[11] H. Zhang, Y. Luo, and D. Liu, "Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints," *IEEE Trans. Neural Netw.*, vol. 20, no. 9, pp. 1490–1503, Sep. 2009.

[12] J. Si, A. G. Barto, W. B. Powell, and D. Wunsch, *Handbook of Learning and Approximate Dynamic Programming*. New York: Wiley, 2004.

[13] F.-Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: An introduction," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 39–47, May 2009.

[14] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, Aug. 2009.

[15] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H_∞ control," *Automatica*, vol. 43, no. 3, pp. 473–481, Mar. 2007.

[16] Q. Wei, H. Zhang, and L. Cui, "Data-based optimal control for discrete-time zero-sum games of 2-D systems using adaptive critic designs," *ACTA Autom. Sinica*, vol. 35, no. 6, pp. 682–692, Jun. 2009.

[17] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, Feb. 2009.

[18] D. Vrabie and F. L. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Netw.*, vol. 22, no. 3, pp. 237–246, Apr. 2009.

[19] R. K. Lim and M. Q. Phan, "Identification of a multistep-ahead observer and its application to predictive control," *J. Guid. Control Dyn.*, vol. 20, no. 6, pp. 1200–1206, 1997.

[20] M. G. Safonov and T. C. Tsao, "The unfalsified control concept and learning," *IEEE Trans. Autom. Control*, vol. 42, no. 6, pp. 843–847, Jun. 1997.

[21] R. L. Toussaint, J. C. Boissy, M. L. Norg, M. Steinbuch, and O. H. Bosgra, "Suppressing non-periodically repeating disturbances in mechanical servo systems," in *Proc. IEEE Conf. Decis. Control*, Tampa, FL, Dec. 1998, pp. 2541–2542.

[22] J. M. Lee and J. H. Lee, "Approximate dynamic programming-based approaches for input-output data-driven control of nonlinear processes," *Automatica*, vol. 41, no. 7, pp. 1281–1288, Jul. 2005.

[23] A. Lecchini, M. C. Campi, and S. M. Savaresi, "Virtual reference feedback tuning for two degrees of freedom controllers," *Int. J. Adapt. Control Signal Process.*, vol. 16, no. 5, pp. 355–371, 2002.

[24] F. L. Lewis and K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 41, no. 1, pp. 14–25, Feb. 2011.

[25] Y. Zhang, L. Guo, H. Yu, and K. Zhao, "Fault tolerant control based on stochastic distributions via MLP neural networks," *Neurocomputing*, vol. 70, nos. 4–6, pp. 867–874, Jan. 2007.

[26] L. Guo and H. Wang, "Fault detection and diagnosis for general stochastic systems using B-spline expansions and nonlinear filters," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 8, pp. 1644–1652, Aug. 2005.

[27] X. Wu, J. Zhang, and Q. Zhu, "A generalized procedure in designing recurrent neural network identification and control of time-varying-delayed nonlinear dynamic systems," *Neurocomputing*, vol. 73, nos. 7–9, pp. 1376–1383, Mar. 2010.

[28] J. Xu and Y. Tan, "Nonlinear adaptive wavelet control using constructive wavelet networks," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 115–127, Jan. 2007.

[29] Z. Hou and J. Xu, "On data-driven control theory: The state of the art and perspective," *ACTA Autom. Sinica*, vol. 35, no. 6, pp. 650–667, 2009.

[30] R. E. Skelton and G. Shi, "Markov data-based LQG control," *J. Dyn. Syst., Meas., Control*, vol. 122, no. 3, pp. 551–559, 2000.

[31] W. Aangenent, D. Kostic, B. de Jager, R. de Molengraft, and M. Steinbuch, "Data-based optimal control," in *Proc. Amer. Control Conf.*, vol. 2. Portland, OR, Jun. 2005, pp. 1460–1465.

[32] Y. M. Park, M. S. Choi, and K. W. Lee, "An optimal tracking neuro-controller for nonlinear dynamic systems," *IEEE Trans. Neural Netw.*, vol. 7, no. 5, pp. 1099–1110, Sep. 1996.

[33] G. Tang, Y. Zhao, and B. Zhang, "Optimal output tracking control for nonlinear systems via successive approximation approach," *Nonlin. Anal.*, vol. 66, no. 6, pp. 1365–1377, Mar. 2007.

[34] H. Zhang, Q. Wei, and Y. Luo, "A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 38, no. 4, pp. 937–942, Aug. 2008.

[35] T. Dierks and S. Jagannathan, "Optimal tracking control of affine nonlinear discrete-time systems with unknown internal dynamics," in *Proc. 48th IEEE Conf. Decis. Control Conf. Chin. Control*, Shanghai, Dec. 2009, pp. 6750–6755.

[36] T. W. Mclain, C. A. Bailry, and R. W. Beard, "Synthesis and experimental testing of a nonlinear optimal tracking controller," in *Proc. Amer. Control Conf.*, vol. 4. San Diego, CA, Jun. 1999, pp. 2847–2851.

[37] V. Yadav, R. Padhi, and S. M. Balakrishnan, "Robust/optimal temperature profile control of a high-speed aerospace vehicle using neural networks," *IEEE Trans. Neural Netw.*, vol. 18, no. 4, pp. 1115–1128, Jul. 2007.

[38] J. D. J. Rubio and W. Yu, "Stability analysis of nonlinear system identification via delayed neural networks," *IEEE Trans. Circuits Syst. II*, vol. 54, no. 2, pp. 161–165, Feb. 2007.

[39] T. Hayakawa, W. M. Haddad, and N. Hovakimyan, "Neural network adaptive control for a class of nonlinear uncertain dynamical systems with asymptotic stability guarantees," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 80–89, Jan. 2008.

[40] V. Chellaboina and W. M. Haddad, "A unification between partial stability and stability theory for time-varying systems," *IEEE Control Syst. Mag.*, vol. 22, no. 6, pp. 66–75, Dec. 2002.

[41] H. K. Khalil, *Nonlinear System*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

Huaguag Zhang (SM'04) received the B.S. and M.S. degrees in control engineering from Northeast Dianli University, Jilin City, China, in 1982 and 1985, respectively, and the Ph.D. degree in thermal power engineering and automation from Southeast University, Nanjing, China, in 1991.

He was a Post-Doctoral Fellow with the Department of Automatic Control, Northeastern University, Shenyang, China, in 1992, for two years. Since 1994, he has been a Professor and Head of the Institute of Electric Automation, School of Information Science and Engineering, Northeastern University. He has authored or co-authored over 200 papers published in journals and conference proceedings and four monographs, and holds 20 patents. His current research interests include fuzzy controls, stochastic system controls, neural networks-based controls, nonlinear controls, and their applications.

Dr. Zhang is an Associate Editor of the *Automatica*, the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART-B: CYBERNETICS, and *Neurocomputing*. He was the recipient of the Outstanding Youth Science Foundation Award from the National Natural Science Foundation Committee of China in 2003. He was named the Cheung Kong Scholar by the Education Ministry of China in 2005.

Lili Cui received the B.S. degree in electrical engineering and automation from Dalian Jiaotong University, Dalian, China, in 2005. She is currently pursuing the Ph.D. degree in control theory and control engineering with Northeastern University, Shenyang, China.

Her current research interests include neural networks-based controls, adaptive optimal controls, approximate dynamic programming, and their industrial applications.

Xin Zhang received the B.S. degree in automation control from Northeastern University, Shenyang, China, in 2005. She is currently pursuing the Ph.D. degree in control theory and control engineering with the same university.

Her current research interests include approximate dynamic programming, neural networks-adaptive controls, game theory, and their industrial applications.

Yanhong Luo (M'09) received the B.S. degree in automation control and the M.S. and Ph.D. degrees in control theory and control engineering from Northeastern University, Shenyang, China, in 2003, 2006, and 2009, respectively.

She is currently an Associate Professor with Northeastern University. Her current research interests include fuzzy controls, neural networks-adaptive controls, approximate dynamic programming, and their industrial applications.