Adaptive-Critic-Based Robust Trajectory Tracking of Uncertain Dynamics and Its Application to a Spring–Mass–Damper System

Ding Wang, *Member, IEEE*, and Chaoxu Mu, *Member, IEEE*

***Abstract*—In this paper, the robust trajectory tracking design of uncertain nonlinear systems is investigated by virtue of a self-learning optimal control formulation. The primary novelty lies in that an effective learning based robust tracking control strategy is developed for nonlinear systems under a general uncertain environment. The augmented system construction is performed by combining the tracking error with the reference trajectory. Then, an improved adaptive critic technique, which does not depend on the initial stabilizing controller, is employed to solve the Hamilton–Jacobi–Bellman (HJB) equation with respect to the nominal augmented system. Using the obtained control law, the closed-loop form of the augmented system is built with stability proof. Moreover, the robust trajectory tracking performance is guaranteed via Lyapunov approach in theory and then through simulation demonstration, where an application to a practical spring–mass–damper system is included.**

***Index Terms*—Adaptive critic design, neural networks, optimal control, robust trajectory tracking, self-learning control, system uncertainty.**

I. Introduction

In recent research on information science, many efforts have been converged into the fields of system engineering and intelligent control, particularly on robust stabilization of uncertain dynamics. The robust control problem is traditionally addressed for dynamical systems with uncertainties [1], [2]. With the recent development of adaptive and learning approaches, many advanced techniques have been brought into the robust control community. Among them, the combination of optimal feedback control and robust stabilization has attracted special attention [3]. After that, by considering the idea of adaptive critic design [4], some related approaches were proposed gradually, including the system-transformation-based strategy [5]–[7], the sliding-mode-based scheme [8], and the robust adaptive dynamic programming method [9]–[11]. An evidently common property of these results [5]–[11] is the introduction of adaptive critic design, which is an intelligent optimization method involving the idea of reinforcement learning. When studying optimal control problems using adaptive critic formulation, the main idea lies in approximating the solution of the Hamilton–Jacobi–Bellman (HJB) equation, which is difficult to address directly [12]–[14]. The adaptive optimal control of nonlinear systems has been studied based on adaptive critic [12]–[14] with natural extension to nonzero-sum differential game design [15], [16]. Particularly, considering the dynamical uncertainties, the approximate HJB based solution can be applied to cope with the robust control problem [5], [6], [17]. Therein, the guaranteed cost control method was designed as an improvement to the basic robust stabilization [5], [6], since the upper bound of the cost function for the originally uncertain plant was considered [17]. It is clear to find that all of the above-mentioned results are obtained for basic regulation problems.

A large class of control design problems include the objective of following a reference signal rather than normally regulating the state at the origin, especially under noise and uncertainty environment. It is often of great significance to track a desired trajectory with specifically optimal performance, hence, it is one of the common problems of system and control communities. In particular, the trajectory tracking control problems have been studied under the adaptive critic framework [18]–[24]. A data-based near optimal tracking control scheme for unknown nonlinear systems was proposed in [18]. Then, the optimal trajectory tracking design for partially unknown nonlinear systems with input constraints was studied in [19]. After that, the H_∞ tracking control approach for input-affine nonlinear systems with completely unknown dynamics was developed in [20] and for real wheeled mobile robot without using the internal system dynamics was given in [21]. When considering the matched system uncertainty, the guaranteed cost tracking control method for a class of uncertain nonlinear systems was provided in [22].

Manuscript received December 17, 2016; revised May 6, 2017; accepted June 14, 2017. Date of publication June 30, 2017; date of current version November 16, 2017. This work was supported in part by Beijing Natural Science Foundation under Grant 4162065, in part by the National Natural Science Foundation of China under Grant U1501251, in part by Tianjin Natural Science Foundation under Grant 14JCQNJC05400, in part by the Open Project Program of The State Key Laboratory of Management and Control for Complex Systems under Grant 20170105, in part by China Postdoctoral Science Foundation under Grant 2014M561559, and in part by the Early Career Development Award of SKLMCCS. *(Corresponding author: Chaoxu Mu.)*

D. Wang is with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: ding.wang@ia.ac.cn).

C. Mu is with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China, and also with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: cxmu@tju.edu.cn).

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

Digital Object Identifier 10.1109/TIE.2017.2722424

A data-based adaptive tracking control scheme for disturbed continuous-time nonlinear systems via the new goal representation heuristic dynamic programming architecture was displayed in [23]. Recently, the novel event-triggered trajectory tracking design of nonlinear systems was proposed in [24], for the purpose of saving the communication resources. Note that these results are derived for trajectory tracking of normal nonlinear dynamics with input constraints or matched uncertainties.

As one of the basic algorithms of reinforcement learning, policy iteration has been utilized widely in various adaptive-critic-based optimization designs. However, when performing this method, an obvious difficulty is the choice of initial admissible control laws [5], [6], [9], [12]. Besides, the adaptive-critic-based robust control approach is often applicable to some special systems, where the uncertain term often matches with the control matrix [6], [8], [22]. Moreover, the tracking problem is often considered for normal optimal control design or matched robust control design [19], [20], [22]. For overcoming these drawbacks, in this paper, a novel self-learning-based robust trajectory tracking control method is developed for nonlinear systems with more general uncertainties, where the initial stabilizing controller is not needed.

The rest contents are arranged as follows. The robust trajectory tracking control problem description and the augmented system construction are stated in Section II. Then, the primary design method including the neural network control implementation and uniformly ultimately bounded (UUB) stability are analyzed in Section III. After that, the simulation verification and the concluding remark are given in Sections IV and V, respectively. In summary, the major contributions of this paper are listed as follows. First, the robust trajectory tracking design of nonlinear systems with more general dynamical uncertainties is originally investigated. Second, an improved learning criterion is established to reduce the requirement of the initial stabilizing controller for adaptive-critic-based optimal and robust control designs. Third, the application scope of the adaptive-critic-based tracking control method is enlarged to general uncertain nonlinear systems.

At the end of this section, the main notations used in the paper are listed as follows. $\mathbb{R}$ stands for the set of all real numbers. $\mathbb{R}^n$ is the Euclidean space of all n-dimensional real vectors. $\mathbb{R}^{n\times m}$ is the space of all $n \times m$ real matrices. $\|\cdot\|$ denotes the vector norm of a vector in $\mathbb{R}^n$ or the matrix norm of a matrix in $\mathbb{R}^{n\times m}$. I_n represents the $n \times n$ identity matrix and $0_{n\times m}$ stands for the $n \times m$ zero matrix. $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ calculate the maximal and minimal eigenvalues of a matrix, respectively. $\mathrm{diag}\{a_1, a_2, \ldots, a_n\}$ denotes the diagonal matrix composed of $a_1, a_2, \ldots, a_n$. Let Ω be a compact subset of $\mathbb{R}^n$ and $\mathscr{A}(\Omega)$ be the set of admissible control laws on Ω. The superscript "T" is taken to represent the transpose operation, and $\nabla(\cdot) \triangleq \partial(\cdot)/\partial x$ is used to denote the gradient operator.

II. Robust Tracking Problem Description

For describing the controlled plant, a class of continuous-time nonlinear systems given by

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

is considered, where $x(t) \in \Omega \subset \mathbb{R}^n$ is the state variable and $u(t) \in \mathbb{R}^m$ is the control vector, $f(\cdot)$ and $g(\cdot)$ are differentiable in their arguments satisfying $f(0) = 0$, and $\Delta f(x)$ is the unknown perturbation with $\Delta f(0) = 0$. Here, let $x(0) = x_0$ be the initial state and assume that the uncertain term $\Delta f(x)$ is bounded by a known function $\lambda_f(x)$, i.e., $\|\Delta f(x)\| \le \lambda_f(x)$ with $\lambda_f(0) = 0$.

In this paper, for achieving the purpose of trajectory tracking, one can introduce a reference system generated as follows:

$$\dot{r}(t) = \varphi(r(t)) \tag{2}$$

where $r(t) \in \mathbb{R}^n$ stands for the bounded desired trajectory with $r(0) = r_0$ and $\varphi(r(t))$ is a Lipschitz continuous function satisfying $\varphi(0) = 0$. Let the trajectory tracking error be

$$z(t) = x(t) - r(t) \tag{3}$$

and the initial error vector be $z(0) = z_0 = x_0 - r_0$. Then, combining (1)–(3), one can obtain the dynamics of the tracking error vector as follows:

$$\dot{z}(t) = f(x(t)) - \varphi(r(t)) + g(x(t))u(t) + \Delta f(x(t)). \tag{4}$$

Noticing $x(t) = z(t) + r(t)$, the system (4) can be written as

$$\begin{aligned}\dot{z}(t) = \; & f(z(t) + r(t)) + g(z(t) + r(t))u(t) \\ & - \varphi(r(t)) + \Delta f(z(t) + r(t)).\end{aligned} \tag{5}$$

Define an augmented state vector as $\xi(t) = [z^{\mathsf{T}}(t), r^{\mathsf{T}}(t)]^{\mathsf{T}} \in \mathbb{R}^{2n}$ with $\xi(0) = \xi_0 = [z_0^{\mathsf{T}}, r_0^{\mathsf{T}}]^{\mathsf{T}}$ being its initial condition, then the augmented dynamics based on (2) and (5) can be formulated as a concise form

$$\dot{\xi}(t) = \mathcal{F}(\xi(t)) + \mathcal{G}(\xi(t))u(t) + \Delta\mathcal{F}(\xi(t)) \tag{6}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{G}(\cdot)$ are new system matrices whereas $\Delta\mathcal{F}(\xi)$ can be still seen as the new uncertain term. Specifically, they are written as

$$\mathcal{F}(\xi(t)) = \begin{bmatrix} f(z(t) + r(t)) - \varphi(r(t)) \\ \varphi(r(t)) \end{bmatrix} \tag{7a}$$

$$\mathcal{G}(\xi(t)) = \begin{bmatrix} g(z(t) + r(t)) \\ 0_{n\times m} \end{bmatrix} \tag{7b}$$

$$\Delta\mathcal{F}(\xi(t)) = \begin{bmatrix} \Delta f(z(t) + r(t)) \\ 0_{n\times 1} \end{bmatrix}. \tag{7c}$$

Clearly, the new uncertain term is still upper bounded since

$$\begin{aligned}\|\Delta\mathcal{F}(\xi)\| &= \|\Delta f(z + r)\| = \|\Delta f(x)\| \\ &\le \lambda_f(x) = \lambda_f(z + r) \triangleq \lambda_f(\xi).\end{aligned} \tag{8}$$

In order to accomplish the robust tracking of system (1) to the reference trajectory (2), one can construct the augmented dynamics (6) and intend to find a feedback control law $u(\xi)$, under which the closed-loop system is asymptotically stable for the uncertainty $\Delta\mathcal{F}(\xi)$. In what follows, one shows that it can be transformed into designing the optimal controller of its nominal system by considering an appropriate cost function.

Now, one can pay important attention to the nominal part of the augmented system (6), that is

$$\dot{\xi}(t) = \mathcal{F}(\xi(t)) + \mathcal{G}(\xi(t))u(t). \tag{9}$$

One shall focus on the optimal feedback control design and want to find the control law $u(\xi)$ to minimize the cost function

$$J(\xi(t)) = \int_t^{\infty} \{\mathcal{Q}(\xi(\tau)) + U(\xi(\tau), u(\tau))\} \mathrm{d}\tau \tag{10}$$

where $\mathcal{Q}(\xi) \geq 0$, $U(\xi, u)$ is the basic part of the utility function, $U(0,0) = 0$, and $U(\xi, u) \geq 0$ for all ξ and u. Here, the basic utility function is chosen as the quadratic form $U(\xi, u) = \xi^{\mathsf{T}} \bar{Q} \xi + u^{\mathsf{T}} R u$, where $\bar{Q} = \mathrm{diag}\{Q, 0_{n \times n}\}$, Q and R are positive definite matrices with $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$. Note that the proposed cost function (10) reflects the uncertainty, regulation, and control terms at the same time, where the additional term $\mathcal{Q}(\xi)$ is closely connected with the dynamical uncertainty.

Addressing optimal control problem requires the designed feedback controller to be admissible [6], [12]. For any admissible control law $u \in \mathscr{A}(\Omega)$, if the associated cost function $J(\xi)$ is continuously differentiable, then, its infinitesimal version is called the nonlinear Lyapunov equation

$$0 = \mathcal{Q}(\xi) + U(\xi, u(\xi)) + (\nabla J(\xi))^{\mathsf{T}} [\mathcal{F}(\xi) + \mathcal{G}(\xi) u(\xi)] \tag{11}$$

with $J(0) = 0$. Define the Hamiltonian of the optimization problem as

$$\begin{aligned} H(\xi, u(\xi), \nabla J(\xi)) = \ & \mathcal{Q}(\xi) + U(\xi, u(\xi)) \\ & + (\nabla J(\xi))^{\mathsf{T}} [\mathcal{F}(\xi) + \mathcal{G}(\xi) u(\xi)]. \end{aligned} \tag{12}$$

The optimal cost function defined by

$$J^*(\xi(t)) = \min_{u \in \mathscr{A}(\Omega)} \int_t^{\infty} \{\mathcal{Q}(\xi(\tau)) + U(\xi(\tau), u(\tau))\} \mathrm{d}\tau \tag{13}$$

satisfies the HJB equation of the form

$$0 = \min_{u \in \mathscr{A}(\Omega)} H(\xi, u(\xi), \nabla J^*(\xi)). \tag{14}$$

The optimal feedback control law is derived by

$$u^*(\xi) = -\frac{1}{2} R^{-1} \mathcal{G}^{\mathsf{T}}(\xi) \nabla J^*(\xi). \tag{15}$$

Taking the optimal control law (15) into (11), one can rewrite the HJB equation as

$$\begin{aligned} 0 = \ & H(\xi, u^*(\xi), \nabla J^*(\xi)) \\ = \ & \mathcal{Q}(\xi) + U(\xi, u^*(\xi)) \\ & + (\nabla J^*(\xi))^{\mathsf{T}} [\mathcal{F}(\xi) + \mathcal{G}(\xi) u^*(\xi)] \end{aligned} \tag{16}$$

with $J^*(0) = 0$. In what follows, one shall show that solving the HJB equation (16) and deriving the optimal control law (15) for the nominal system (9) can let us accomplish robust stabilization for the uncertain system (6). However, it is usually a difficult task to solve the nonlinear HJB equation directly. This motivates us to find its approximate solution via adaptive critic design. To this end, one can specify the term $\mathcal{Q}(\xi)$ as

$$\mathcal{Q}(\xi) = \frac{1}{4} (\nabla J(\xi))^{\mathsf{T}} \nabla J(\xi) + \lambda_f^2(\xi) \tag{17}$$

and find that $\mathcal{Q}(\xi) \geq 0$ holds. By virtue of this term, the whole utility function and the cost function (10) can be well defined.

III. Robust Trajectory Tracking Design Scheme

The robust trajectory tracking problem can be handled through three steps. First, one can conduct the approximate tracking control design via adaptive critic technique and neural network implementation. Second, one shall prove the closed-loop stability of the nominal augmented system to obtain a final bound between the approximate and optimal tracking controllers. Third, one can check the robust trajectory tracking performance by considering the uncertain dynamics and the proposed approximate controller.

A. Neural Network Tracking Control Implementation

Recalling the classical universal approximation property, one can reconstruct $J^*(\xi)$ by a single hidden layer neural network on a compact set Ω as

$$J^*(\xi) = \omega_c^{\mathsf{T}} \sigma_c(\xi) + \varepsilon_c(\xi) \tag{18}$$

where $\omega_c \in \mathbb{R}^l$ is the ideal weight, $\sigma_c(\xi) \in \mathbb{R}^l$ is the activation function, l is the number of hidden neurons, and $\varepsilon_c(\xi)$ is the unknown approximation error. The related gradient vector is

$$\nabla J^*(\xi) = (\nabla \sigma_c(\xi))^{\mathsf{T}} \omega_c + \nabla \varepsilon_c(\xi) \tag{19}$$

which, clearly, is a $2n$-dimensional vector and different with the regulation problem [6], [12]. For performing adaptive critic design, a critic network is often built in terms of the estimated weight vector $\hat{\omega}_c$ to approximate the optimal cost function as

$$\hat{J}^*(\xi) = \hat{\omega}_c^{\mathsf{T}} \sigma_c(\xi). \tag{20}$$

Then, one can also get the gradient vector as follows:

$$\nabla \hat{J}^*(\xi) = (\nabla \sigma_c(\xi))^{\mathsf{T}} \hat{\omega}_c. \tag{21}$$

Noticing (15) and (19), one can write the optimal control law as

$$u^*(\xi) = -\frac{1}{2} R^{-1} \mathcal{G}^{\mathsf{T}}(\xi) \left[(\nabla \sigma_c(\xi))^{\mathsf{T}} \omega_c + \nabla \varepsilon_c(\xi) \right]. \tag{22}$$

Considering (21), the approximate optimal control law is

$$\hat{u}^*(\xi) = -\frac{1}{2} R^{-1} \mathcal{G}^{\mathsf{T}}(\xi) (\nabla \sigma_c(\xi))^{\mathsf{T}} \hat{\omega}_c. \tag{23}$$

Taking the applicable approximate optimal control law (23) into the nominal augmented plant (9), the closed-loop system dynamics can be formulated as

$$\dot{\xi} = \mathcal{F}(\xi) - \frac{1}{2} \mathcal{G}(\xi) R^{-1} \mathcal{G}^{\mathsf{T}}(\xi) (\nabla \sigma_c(\xi))^{\mathsf{T}} \hat{\omega}_c. \tag{24}$$

In what follows, for convenience of analysis, one shall introduce two nonnegative matrices as follows:

$$\mathcal{A}(\xi) = \nabla \sigma_c(\xi) \mathcal{G}(\xi) R^{-1} \mathcal{G}^{\mathsf{T}}(\xi) (\nabla \sigma_c(\xi))^{\mathsf{T}} \tag{25a}$$

$$\mathcal{B}(\xi) = \nabla \sigma_c(\xi) (\nabla \sigma_c(\xi))^{\mathsf{T}}. \tag{25b}$$

Note that both the cost function and control law can be expressed as functions of the weight vector with neural network formulation. Hence, the Hamiltonian can be written as a new equation including ξ and ω_c, i.e.,

$$\begin{aligned} H(\xi,\omega_c) &= \xi^{\mathsf{T}}\bar{Q}\xi + \omega_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{4}\omega_c^{\mathsf{T}}\mathcal{A}(\xi)\omega_c \\ &\quad + \lambda_f^2(\xi) + \frac{1}{4}\omega_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c + e_{cH} \\ &= 0 \end{aligned} \tag{26}$$

where the last term

$$\begin{aligned} e_{cH} &= (\nabla\varepsilon_c(\xi))^{\mathsf{T}}\mathcal{F}(\xi) \\ &\quad - \frac{1}{2}(\nabla\varepsilon_c(\xi))^{\mathsf{T}}\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)(\nabla\sigma_c(\xi))^{\mathsf{T}}\omega_c \\ &\quad - \frac{1}{4}(\nabla\varepsilon_c(\xi))^{\mathsf{T}}\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla\varepsilon_c(\xi) \\ &\quad + \frac{1}{2}(\nabla\varepsilon_c(\xi))^{\mathsf{T}}(\nabla\sigma_c(\xi))^{\mathsf{T}}\omega_c \\ &\quad + \frac{1}{4}(\nabla\varepsilon_c(\xi))^{\mathsf{T}}\nabla\varepsilon_c(\xi) \end{aligned} \tag{27}$$

denotes the residual error of neural network expression. By using the estimated weight, the approximate Hamiltonian is

$$\begin{aligned} \hat{H}(\xi,\hat{\omega}_c) &= \xi^{\mathsf{T}}\bar{Q}\xi + \hat{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{4}\hat{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\hat{\omega}_c \\ &\quad + \lambda_f^2(\xi) + \frac{1}{4}\hat{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\hat{\omega}_c. \end{aligned} \tag{28}$$

Denoting $e_c = \hat{H}(\xi,\hat{\omega}_c) - H(\xi,\omega_c)$ and considering (26), one easily observes that $e_c = \hat{H}(\xi,\hat{\omega}_c)$. Define the weight estimation error of the critic network as $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$. Combining (26) with (28), the formulation of e_c with $\tilde{\omega}_c$ is

$$\begin{aligned} e_c &= \hat{H}(\xi,\hat{\omega}_c) - H(\xi,\omega_c) \\ &= -\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\omega_c \\ &\quad + \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\tilde{\omega}_c - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c - e_{cH}. \end{aligned} \tag{29}$$

Now, one begins to train the critic neural network and designs the weight vector $\hat{\omega}_c$ to minimize the objective function $E_c = (1/2)e_c^2$. In the traditional adaptive critic design [6], [12], only the information $\partial E_c/\partial\hat{\omega}_c$ is considered to perform the neural network learning stage. Note that the initial admissible control law is required to start the learning control design. For overcoming the difficulty of searching the initial admissible controller, one can make the following assumption and develop an additional term to reinforce the learning process.

Assumption 1: Consider the nominal augmented system (9) with the cost function (10) and its closed-loop form with the optimal feedback control law (15). Let $J_s(\xi)$ be a continuously differentiable Lyapunov function candidate satisfying

$$\dot{J}_s(\xi) = (\nabla J_s(\xi))^{\mathsf{T}}(\mathcal{F}(\xi) + \mathcal{G}(\xi)u^*(\xi)) < 0. \tag{30}$$

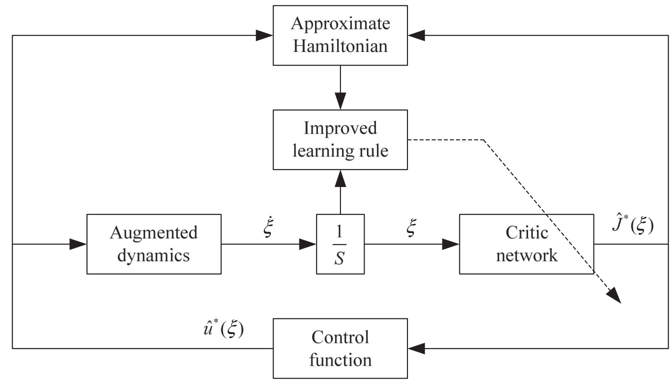


Fig. 1. Control structure.

Then, there exists a positive definite matrix $\mathcal{K} \in \mathbb{R}^{2n\times 2n}$ ensuring that

$$\begin{aligned} &(\nabla J_s(\xi))^{\mathsf{T}}[\mathcal{F}(\xi) + \mathcal{G}(\xi)u^*(\xi)] \\ &\quad = -(\nabla J_s(\xi))^{\mathsf{T}}\mathcal{K}\nabla J_s(\xi) \le -\lambda_{\min}(\mathcal{K})\|\nabla J_s(\xi)\|^2 \end{aligned} \tag{31}$$

is true. Note that during the implementation process, $J_s(\xi)$ can be obtained by suitably selecting a polynomial with respect to the augmented state vector, such as the form $J_s(\xi) = 0.5\xi^{\mathsf{T}}\xi$.

In this paper, the improved learning rule is developed as

$$\dot{\hat{\omega}}_c = -\alpha_c\left(\frac{\partial E_c}{\partial\hat{\omega}_c}\right) + \frac{1}{2}\alpha_s\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla J_s(\xi) \tag{32}$$

where $\alpha_c > 0$ denotes the basic learning rate of critic network and $\alpha_s > 0$ represents the adjusting rate for the additional stabilizing term. Compared with the traditional updating rule $\dot{\hat{\omega}}_c^o = -\alpha_c(\partial E_c/\partial\hat{\omega}_c)$ used in [5], [6], (32) provides a reinforced structure with two adjustable learning rates α_c and α_s. Hence, the designers can conduct more practical control tasks in light of their engineering experience and intuition. The structure of the proposed control method is shown in Fig. 1.

B. Closed-Loop Stability of the Nominal Augmented System

In this part, one can deduce the dynamics of the weight estimation error. Recalling (28), it is derived that

$$\frac{\partial e_c}{\partial\hat{\omega}_c} = \nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{2}\mathcal{A}(\xi)\hat{\omega}_c + \frac{1}{2}\mathcal{B}(\xi)\hat{\omega}_c. \tag{33}$$

Noticing (32), the dynamics of the weight estimation error, i.e., $\dot{\tilde{\omega}}_c = -\dot{\hat{\omega}}_c$, is written as

$$\dot{\tilde{\omega}}_c = \alpha_c\left(\frac{\partial E_c}{\partial\hat{\omega}_c}\right) - \frac{1}{2}\alpha_s\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla J_s(\xi). \tag{34}$$

Considering (29) and (33), it follows from (34) that

$$\begin{aligned}\dot{\tilde{\omega}}_c = \ & \alpha_c\Big(-\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\omega_c \\ & + \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\tilde{\omega}_c - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c - e_{cH}\Big) \\ & \times\Big(\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{2}\mathcal{A}(\xi)\omega_c + \frac{1}{2}\mathcal{A}(\xi)\tilde{\omega}_c \\ & + \frac{1}{2}\mathcal{B}(\xi)\omega_c - \frac{1}{2}\mathcal{B}(\xi)\tilde{\omega}_c\Big) \\ & - \frac{1}{2}\alpha_s\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla J_s(\xi) \end{aligned} \tag{35}$$

which is the weight error dynamics of the critic neural network. In the sequel, the stability of the critic error dynamics and the closed-loop augmented state including the approximate optimal controller are studied. The following common assumption is required [6], [8], [12], [16], [17].

Assumption 2: The control function matrix $g(x)$ is bounded as $\|g(x)\| \le \lambda_g$, where λ_g is a positive constant, and hence

$$\|\mathcal{G}(\xi)\| = \|g(z+r)\| = \|g(x)\| \le \lambda_g. \tag{36}$$

On the compact set Ω, the terms $\nabla\sigma_c(\xi)$, $\nabla\varepsilon_c(\xi)$, and e_{cH} are bounded as $\|\nabla\sigma_c(\xi)\| \le \lambda_\sigma$, $\|\nabla\varepsilon_c(\xi)\| \le \lambda_\varepsilon$, and $|e_{cH}| \le \lambda_e$, where λ_σ, λ_ε, and λ_e are positive constants.

Theorem 1: Considering the nominal augmented system (9), let the feedback control law be derived by (23) and the weight vector of the critic network be trained by (32). Then, based on this feedback control law, the closed-loop system state and the critic weight estimation error are both UUB.

Proof: A Lyapunov function candidate containing two terms is chosen as the following form:

$$L(t) = \frac{1}{2\alpha_c}\tilde{\omega}_c^{\mathsf{T}}(t)\tilde{\omega}_c(t) + \frac{\alpha_s}{\alpha_c}J_s(\xi(t)). \tag{37}$$

The time derivative along the two dynamics related with the augmented state (24) and the weight vector (35) is

$$\begin{aligned}\dot{L}(t) = \ & \frac{1}{\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\dot{\tilde{\omega}}_c + \frac{\alpha_s}{\alpha_c}(\nabla J_s(\xi))^{\mathsf{T}}\dot{\xi} \\ = \ & \tilde{\omega}_c^{\mathsf{T}}\Big(-\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\omega_c \\ & + \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\tilde{\omega}_c - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c - e_{cH}\Big) \\ & \times\Big(\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{2}\mathcal{A}(\xi)\omega_c + \frac{1}{2}\mathcal{A}(\xi)\tilde{\omega}_c \\ & + \frac{1}{2}\mathcal{B}(\xi)\omega_c - \frac{1}{2}\mathcal{B}(\xi)\tilde{\omega}_c\Big) \\ & - \frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla J_s(\xi) \\ & + \frac{\alpha_s}{\alpha_c}(\nabla J_s(\xi))^{\mathsf{T}}\dot{\xi}. \end{aligned} \tag{38}$$

Based on $\tilde{\omega}_c = \omega_c - \hat{\omega}_c$ and noticing (24) and (25a), one can find an additional formula

$$\nabla\sigma_c(\xi)\mathcal{F}(\xi) - \frac{1}{2}\mathcal{A}(\xi)\omega_c + \frac{1}{2}\mathcal{A}(\xi)\tilde{\omega}_c = \nabla\sigma_c(\xi)\dot{\xi} \tag{39}$$

which can be used to further derive that

$$\begin{aligned}\dot{L}(t) = \ & -\Big(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\dot{\xi} - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\tilde{\omega}_c - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\tilde{\omega}_c \\ & + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c + e_{cH}\Big) \\ & \times\Big(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\dot{\xi} - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c\Big) \\ & - \frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla J_s(\xi) \\ & + \frac{\alpha_s}{\alpha_c}(\nabla J_s(\xi))^{\mathsf{T}}\dot{\xi}. \end{aligned} \tag{40}$$

By observing the closed-loop augmented system $\dot{\xi}^* = \mathcal{F}(\xi) + \mathcal{G}(\xi)u^*(\xi)$ with the optimal controller $u^*(\xi)$, one can obtain

$$\begin{aligned}\dot{L}(t) = \ & -\Big(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\dot{\xi}^* + \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\tilde{\omega}_c \\ & + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla\varepsilon_c(\xi) \\ & - \frac{1}{4}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c + e_{cH}\Big) \\ & \times\Big(\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\dot{\xi}^* + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{A}(\xi)\tilde{\omega}_c \\ & + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla\varepsilon_c(\xi) \\ & - \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\tilde{\omega}_c + \frac{1}{2}\tilde{\omega}_c^{\mathsf{T}}\mathcal{B}(\xi)\omega_c\Big) \\ & - \frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla J_s(\xi) \\ & + \frac{\alpha_s}{\alpha_c}(\nabla J_s(\xi))^{\mathsf{T}}\dot{\xi}. \end{aligned} \tag{41}$$

Note that the term $\nabla\sigma_c(\xi)\dot{\xi}^*$ in (41) is upper bounded due to the boundedness of the element $\nabla\sigma_c(\xi)$ and the optimal state derivative $\dot{\xi}^*$. One can expand (41), perform mathematical operations, notice the Assumption 2, and then derive that

$$\begin{aligned}\dot{L}(t) \le \ & -\lambda_1\|\tilde{\omega}_c\|^4 + \lambda_2\|\tilde{\omega}_c\|^2 + \lambda_3^2 \\ & - \frac{\alpha_s}{2\alpha_c}\tilde{\omega}_c^{\mathsf{T}}\nabla\sigma_c(\xi)\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla J_s(\xi) \\ & + \frac{\alpha_s}{\alpha_c}(\nabla J_s(\xi))^{\mathsf{T}}\dot{\xi} \end{aligned} \tag{42}$$

where λ_1, λ_2, and λ_3 are all positive constants that can be guaranteed in theory. One can consider (42) as well as Assumptions

1 and 2 and further obtain

$$\begin{aligned}\dot{L}(t) \leq & -\lambda_1\|\tilde{\omega}_c\|^4 + \lambda_2\|\tilde{\omega}_c\|^2 + \lambda_3^2 \\ & + \frac{\alpha_s}{\alpha_c}(\nabla J_s(\xi))^{\mathsf{T}}(\mathcal{F}(\xi) + \mathcal{G}(\xi)u^*(\xi)) \\ & + \frac{\alpha_s}{2\alpha_c}(\nabla J_s(\xi))^{\mathsf{T}}\mathcal{G}(\xi)R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\nabla\varepsilon_c(\xi) \\ \leq & -\lambda_1\left(\|\tilde{\omega}_c\|^2 - \frac{\lambda_2}{2\lambda_1}\right)^2 + \lambda_\Sigma \\ & - \frac{\alpha_s}{\alpha_c}\lambda_{\min}(\mathcal{K})\left(\|\nabla J_s(\xi)\| - \frac{\lambda_g^2\lambda_\varepsilon\|R^{-1}\|}{4\lambda_{\min}(\mathcal{K})}\right)^2 \end{aligned} \tag{43}$$

where the constant term

$$\lambda_\Sigma = \frac{\lambda_2^2 + 4\lambda_1\lambda_3^2}{4\lambda_1} + \frac{\alpha_s\lambda_g^4\lambda_\varepsilon^2\|R^{-1}\|^2}{16\alpha_c\lambda_{\min}(\mathcal{K})}. \tag{44}$$

Therefore, if one of the following inequalities

$$\|\tilde{\omega}_c\| \geq \sqrt{\frac{\lambda_2}{2\lambda_1} + \sqrt{\frac{\lambda_\Sigma}{\lambda_1}}} \triangleq \lambda_{\tilde{\omega}_c} \tag{45}$$

or

$$\|\nabla J_s(\xi)\| \geq \frac{\lambda_g^2\lambda_\varepsilon\|R^{-1}\|}{4\lambda_{\min}(\mathcal{K})} + \sqrt{\frac{\alpha_c\lambda_\Sigma}{\alpha_s\lambda_{\min}(\mathcal{K})}} \triangleq \lambda_{J_s} \tag{46}$$

holds, one can derive that $\dot{L}(t) < 0$. Since $J_s(\xi)$ is often chosen as a polynomial and based on the standard Lyapunov extension theorem, one can conclude that the closed-loop state dynamics and the weight error dynamics are both UUB. ■

With the conclusion of Theorem 1, one can easily get the convergence property of the approximate control law function.

Corollary 1: The approximate control law $\hat{u}^*(\xi)$ derived by (23) converges to a neighborhood of the optimal feedback control law $u^*(\xi)$ with a finite bound.

Proof: By combining the control law functions given in (22) and (23), one shall obtain

$$u^*(\xi) - \hat{u}^*(\xi) = -\frac{1}{2}R^{-1}\mathcal{G}^{\mathsf{T}}(\xi)\big[(\nabla\sigma(\xi))^{\mathsf{T}}\tilde{\omega}_c + \nabla\varepsilon_c(\xi)\big]. \tag{47}$$

According to Theorem 1, one has derived that $\|\tilde{\omega}_c\| < \lambda_{\tilde{\omega}_c}$. Then, by considering Assumption 2, one can further determine a finite bound λ_u such that

$$\|u^*(\xi) - \hat{u}^*(\xi)\| \leq \frac{1}{2}\|R^{-1}\|\lambda_g(\lambda_\sigma\lambda_{\tilde{\omega}_c} + \lambda_\varepsilon) \triangleq \lambda_u \tag{48}$$

which completes the proof. ■

C. Robust Trajectory Tracking Property

Now, one shall check the robust trajectory tracking performance of the original uncertain nonlinear system (1) to the reference signal (2) in theory. To this end, it is required to prove that the tracking error of the uncertain augmented system is stable.

Theorem 2: For the nominal augmented system (9) and the cost function (10), the approximate optimal control law obtained by (23) ensures that the tracking error dynamics in the closed-loop form of system (6) is UUB.

Proof: Let the optimal cost function $J^*(\xi)$ be the Lyapunov function candidate. Note that the formula (15) implies

$$(\nabla J^*(\xi))^{\mathsf{T}}\mathcal{G}(\xi) = -2u^{*\mathsf{T}}(\xi)R. \tag{49}$$

Combining (16) with (49), one can easily find that

$$(\nabla J^*(\xi))^{\mathsf{T}}\mathcal{F}(\xi) = -\mathcal{Q}(\xi) - \xi^{\mathsf{T}}\bar{Q}\xi + u^{*\mathsf{T}}(\xi)Ru^*(\xi). \tag{50}$$

If one adopts the approximate optimal control law (23) and notices (50), then the time derivative along the uncertain augmented plant (6) can be obtained as

$$\begin{aligned}\dot{J}^*(\xi) = & \ (\nabla J^*(\xi))^{\mathsf{T}}[\mathcal{F}(\xi) + \mathcal{G}(\xi)\hat{u}^*(\xi) + \Delta\mathcal{F}(\xi)] \\ = & -\mathcal{Q}(\xi) - \xi^{\mathsf{T}}\bar{Q}\xi + u^{*\mathsf{T}}(\xi)Ru^*(\xi) \\ & - 2u^{*\mathsf{T}}(\xi)R\hat{u}^*(\xi) + (\nabla J^*(\xi))^{\mathsf{T}}\Delta\mathcal{F}(\xi). \end{aligned} \tag{51}$$

By considering the specified form of $\mathcal{Q}(\xi)$ as in (17) and introducing $(\Delta\mathcal{F}(\xi))^{\mathsf{T}}\Delta\mathcal{F}(\xi)$, one can further turn (51) to be

$$\begin{aligned}\dot{J}^*(\xi) = & -\xi^{\mathsf{T}}\bar{Q}\xi + u^{*\mathsf{T}}(\xi)Ru^*(\xi) - 2u^{*\mathsf{T}}(\xi)R\hat{u}^*(\xi) \\ & - \left[\frac{1}{2}\nabla J^*(\xi) - \Delta\mathcal{F}(\xi)\right]^{\mathsf{T}}\left[\frac{1}{2}\nabla J^*(\xi) - \Delta\mathcal{F}(\xi)\right] \\ & - \left[\lambda_f^2(\xi) - (\Delta\mathcal{F}(\xi))^{\mathsf{T}}\Delta\mathcal{F}(\xi)\right]. \end{aligned} \tag{52}$$

Note that the first term of (52) is equal to

$$-\xi^{\mathsf{T}}\bar{Q}\xi = -\left[z^{\mathsf{T}}, r^{\mathsf{T}}\right]\left[\begin{array}{c:c} Q & 0_{n\times n} \\ \hdashline 0_{n\times n} & 0_{n\times n}\end{array}\right]\begin{bmatrix} z \\ r\end{bmatrix} = -z^{\mathsf{T}}Qz. \tag{53}$$

Moving on, one can perform some mathematical operations to the second and third terms of (52) as follows:

$$\begin{aligned} & u^{*\mathsf{T}}(\xi)Ru^*(\xi) - 2u^{*\mathsf{T}}(\xi)R\hat{u}^*(\xi) \\ & \quad = [u^*(\xi) - \hat{u}^*(\xi)]^{\mathsf{T}}R[u^*(\xi) - \hat{u}^*(\xi)] - \hat{u}^{*\mathsf{T}}(\xi)R\hat{u}^*(\xi) \\ & \quad \leq \lambda_{\max}(R)\|u^*(\xi) - \hat{u}^*(\xi)\|^2. \end{aligned} \tag{54}$$

By noticing (8), (48), (53), and (54), it follows from (52) that

$$\begin{aligned}\dot{J}^*(\xi) & \leq -\xi^{\mathsf{T}}\bar{Q}\xi + \lambda_{\max}(R)\lambda_u^2 \\ & = -z^{\mathsf{T}}Qz + \lambda_{\max}(R)\lambda_u^2 \\ & \leq -\lambda_{\min}(Q)\|z\|^2 + \lambda_{\max}(R)\lambda_u^2. \end{aligned} \tag{55}$$

As a result, one can derive that $\dot{J}^*(\xi) < 0$, in case that $z(t)$ lies outside the compact set

$$\Omega_z = \left\{z : \|z\| \leq \sqrt{\frac{\lambda_{\max}(R)\lambda_u^2}{\lambda_{\min}(Q)}} \triangleq \lambda_z\right\} \tag{56}$$

where λ_z is a positive constant. In this sense, one can claim that with the approximate optimal control law (23), the tracking error of the closed-loop uncertain augmented system is UUB. In other words, this result shows that the state of the original uncertain nonlinear system (1) can follow the desired trajectory (2), which ends the proof. ■

Remark 1: The adaptive-critic-based robust trajectory tracking control of affine nonlinear systems is designed, where the involved uncertain term is more general than the matched case. It is important to note that this method is widely applicable and is more convenient to implement than those of [18]–[24]. This is because of the adoption of the improved learning rule and the new handling manner for the uncertain term. Besides, the robust trajectory tracking also achieves a meaningful generalization when compared with the traditional robust stabilization design such as in [5]–[11].

IV. Simulation Verification

In this part, one can verify the effectiveness of the present robust trajectory tracking control scheme through a typical nonlinear dynamics simulation and a practical application.

Example 1 (Nonlinear Dynamics Simulation). Consider a continuous-time uncertain nonlinear system

$$\dot{x} = \begin{bmatrix} -x_1 + x_2^2 \\ -x_1^3 - 2x_2 \end{bmatrix} + \begin{bmatrix} -0.5 \\ 1 \end{bmatrix} u + \begin{bmatrix} \theta_1 x_2 \cos(x_1) \\ \theta_2 x_1 \sin(x_2^2) \end{bmatrix} \tag{57}$$

where $x = [x_1, x_2]^{\mathsf{T}} \in \mathbb{R}^2$ is the state variable, $u \in \mathbb{R}$ is the control variable, and the uncertain parameters $\theta_1, \theta_2 \in [-1,1]$. The last term of system (57) is the uncertainty that is bounded by $\lambda_f(x) = \sqrt{x_1^2 + x_2^2}$. Let the initial system state vector be $x_0 = [-0.5, 1.5]^{\mathsf{T}}$. Here, the reference trajectory $r(t)$ is generated by the following system:

$$\dot{r} = \begin{bmatrix} -r_1 + \sin(r_2) \\ -2\sin^3(r_1) - 0.5r_2 \end{bmatrix} \tag{58}$$

where $r = [r_1, r_2]^{\mathsf{T}} \in \mathbb{R}^2$ is the reference state with the initial condition being $r_0 = [0.5, 0.5]^{\mathsf{T}}$.

One can define the tracking error as $z = x - r$ so that $\dot{z} = \dot{x} - \dot{r}$, let the augmented state vector be $\xi = [z^{\mathsf{T}}, r^{\mathsf{T}}]^{\mathsf{T}}$, and then combine (57) with (58) to obtain the augmented system dynamics as follows:

$$\begin{aligned} \dot{\xi} = &\begin{bmatrix} -(\xi_1 + \xi_3) + (\xi_2 + \xi_4)^2 + \xi_3 - \sin(\xi_4) \\ -(\xi_1 + \xi_3)^3 - 2(\xi_2 + \xi_4) + 2\sin^3(\xi_3) + 0.5\xi_4 \\ -\xi_3 + \sin(\xi_4) \\ -2\sin^3(\xi_3) - 0.5\xi_4 \end{bmatrix} \\ &+ [-0.5\ 1\ 0\ 0]^{\mathsf{T}} u + \Delta\mathcal{F}(\xi) \end{aligned} \tag{59}$$

where $\xi = [\xi_1, \xi_2, \xi_3, \xi_4]^{\mathsf{T}} \in \mathbb{R}^4$ with $\xi_1 = z_1$, $\xi_2 = z_2$, $\xi_3 = r_1$, $\xi_4 = r_2$, and $\Delta\mathcal{F}(\xi)$ is the uncertain term of the augmented system. According to (8), the upper bound is

$$\begin{aligned} \lambda_f(\xi) &= \sqrt{(z_1 + r_1)^2 + (z_2 + r_2)^2} \\ &= \sqrt{(\xi_1 + \xi_3)^2 + (\xi_2 + \xi_4)^2}. \end{aligned} \tag{60}$$

It is easy to compute that the initial error vector is $z_0 = x_0 - r_0 = [-1, 1]^{\mathsf{T}}$, so that the initial state of the augmented system is $\xi_0 = [-1, 1, 0.5, 0.5]^{\mathsf{T}}$.

For dealing with the approximate optimal control for the nominal augmented part of (59) with $\bar{Q} = \operatorname{diag}\{I_2, 0_{2\times 2}\}$ and $R = I$, one can construct a critic neural network as the form

$$\begin{aligned} \hat{J}^*(\xi) = &\ \hat{\omega}_{c1}\xi_1^2 + \hat{\omega}_{c2}\xi_1\xi_2 + \hat{\omega}_{c3}\xi_1\xi_3 \\ &+ \hat{\omega}_{c4}\xi_1\xi_4 + \hat{\omega}_{c5}\xi_2^2 + \hat{\omega}_{c6}\xi_2\xi_3 \\ &+ \hat{\omega}_{c7}\xi_2\xi_4 + \hat{\omega}_{c8}\xi_3^2 + \hat{\omega}_{c9}\xi_3\xi_4 + \hat{\omega}_{c10}\xi_4^2. \end{aligned} \tag{61}$$

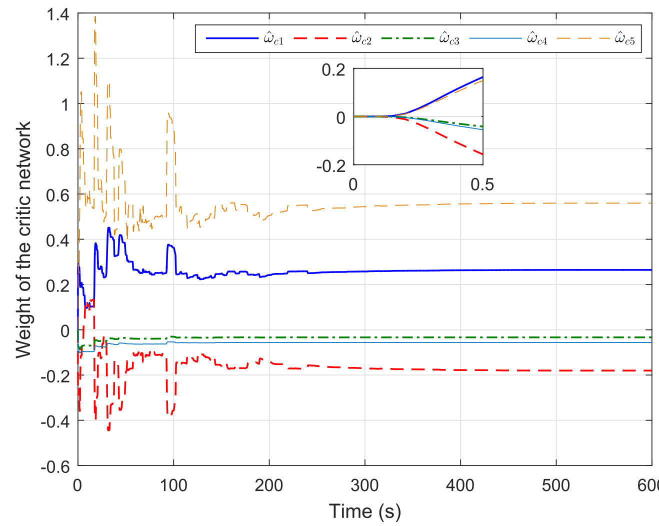


Fig. 2. Convergence process of the weight vector (first five elements).

This is always an experimental choice by considering a tradeoff between control accuracy and computational complexity. During the simulation process, one experimentally sets $\alpha_c = 1.2$ and $\alpha_s = 0.01$ as well as brings in a probing noise to satisfy the persistence of excitation condition. Through a learning stage within $t = 600$ s, the weight vector converges to $[0.2644, -0.1803, -0.0334, -0.0561, 0.5596, 0.1283, 0.1083, 0.0041, 0.0142, 0.0255]^{\mathsf{T}}$. Here, the first five elements of the weight vector are depicted in Fig. 2, whereas the last five elements are omitted. Clearly, one can find that the weight vector is initialized as a zero vector, which leads to an evident convenience when performing the adaptive critic control design.

Next, one shall check the robust trajectory tracking performance by adopting the obtained approximate optimal control law. One can select $\theta_1 = 1$ and $\theta_2 = -1$, and then apply the developed controller to the augmented uncertain plant (59) for $t = 16$ s. The tracking error and control input are shown in Figs. 3 and 4, where the tracking error gradually becomes zero. For conducting comparison, one can choose $\theta_1 = -1$ and $\theta_2 = 1$, and then perform the simulation and verification again. The tracking error and control input are depicted in Figs. 5 and 6, which still demonstrates the desired trajectory tracking performance.

The further discussion of the proposed strategy when compared with the existing work is exhibited in Table I, where the application scope and control achievement of different methods are included. It is clear to find that the previous approaches cannot be used to solve the robust trajectory control problem of this paper. Hence, the developed approach is indeed more general and applicable than the previous work.

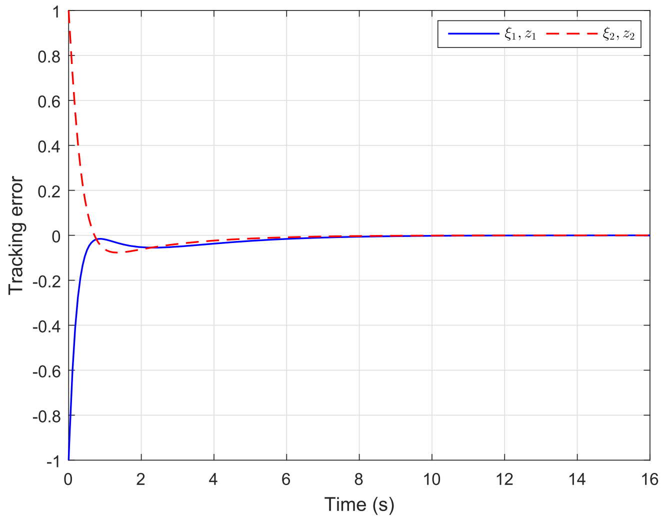


Fig. 3. Tracking error trajectories when $\theta_1 = 1$ and $\theta_2 = -1$.

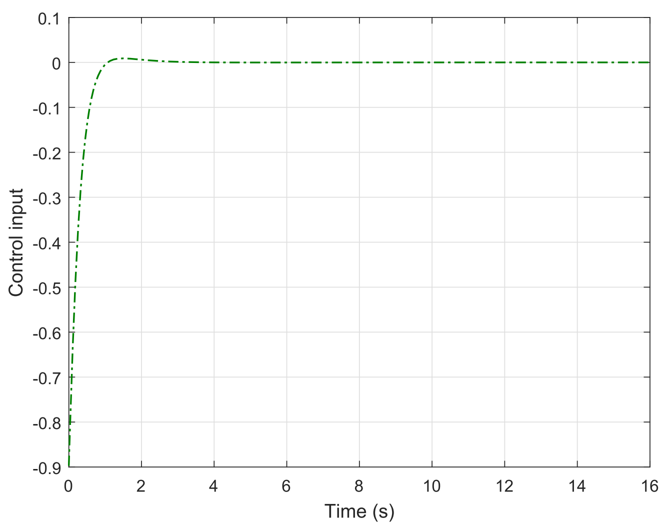


Fig. 4. Tracking control trajectory when $\theta_1 = 1$ and $\theta_2 = -1$.

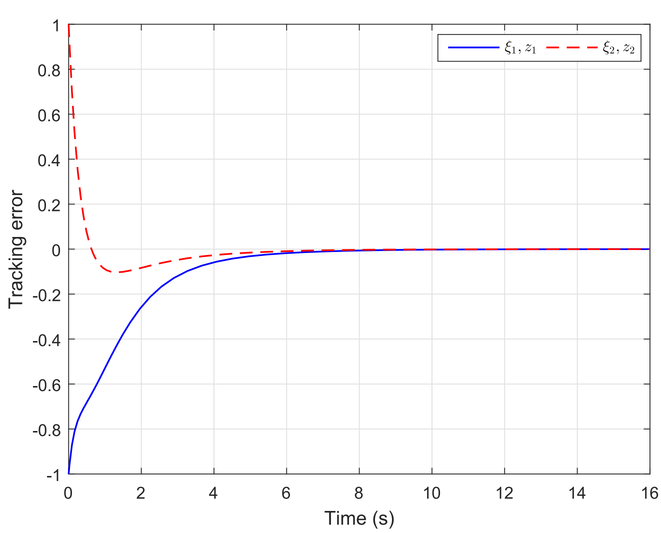


Fig. 5. Tracking error trajectories when $\theta_1 = -1$ and $\theta_2 = 1$.

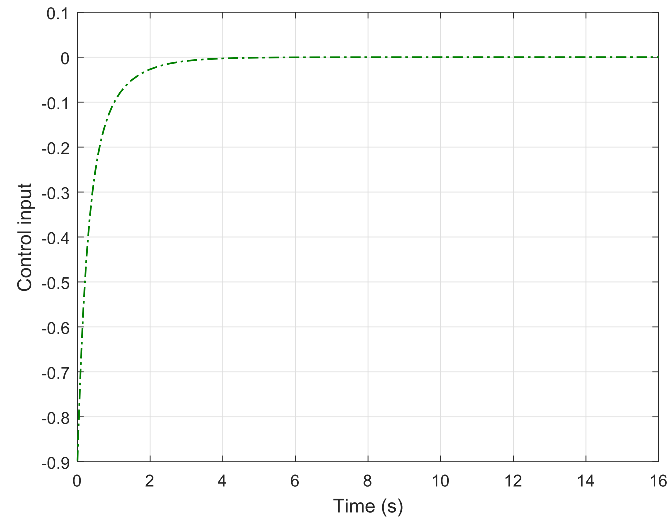


Fig. 6. Tracking control trajectory when $\theta_1 = -1$ and $\theta_2 = 1$.

TABLE I
COMPARISON DISCUSSION

Methods	Control plant	Control achievement
[5]–[11]	Specific uncertain system	Traditional robust stabilization
[12]–[17]	Specific nonlinear system	Near optimal regulation
[18]–[24]	Specific nonlinear system	Traditional tracking control
This paper	General uncertain system	Robust trajectory tracking

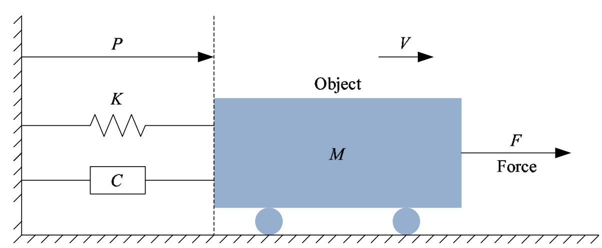


Fig. 7. Simple diagram of the mass–spring–damper system.

Example 2 (Application to a Spring–Mass–Damper System). Consider the spring–mass–damper system given in [19] and depicted in Fig. 7, where M denotes the mass of the object, K represents the stiffness constant of the spring, and C stands for the damping.

Letting P, V, and F be the position, the velocity, and the force applied to the object, respectively, one can obtain the system dynamics as follows:

$$\dot{P} = V \tag{62a}$$

$$M\dot{V} = -KP - CV + F. \tag{62b}$$

Defining $x_1 = P$, $x_2 = V$, and $u = F$, one can write the state-space description of the spring–mass–damper system as

$$\dot{x}_1 = x_2, \tag{63a}$$

$$\dot{x}_2 = -\frac{K}{M}x_1 - \frac{C}{M}x_2 + \frac{1}{M}u. \tag{63b}$$

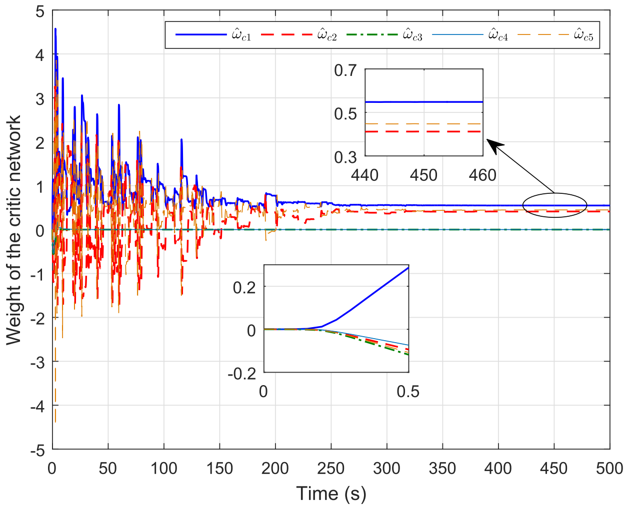


Fig. 8. Convergence process of the weight vector (first five elements).

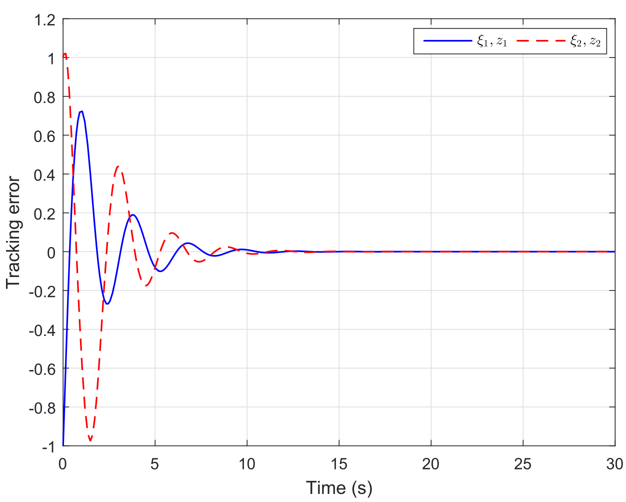


Fig. 9. Tracking error trajectories when $\theta_1 = 0.5$ and $\theta_2 = -0.5$.

Assuming the system (63) is disturbed by an uncertain term as Example 1 and setting practical parameters as $M = 1$ kg, $K = 3$ N/m, $C = 0.5$ Ns/m, one can construct the overall uncertain plant. In this example, the reference signal $r(t)$ is derived by

$$\dot{r} = \begin{bmatrix} -0.5r_1 - r_2 \cos(r_1) \\ \sin(r_1) - r_2 \end{bmatrix}. \tag{64}$$

Similar to Example 1, the augmented system dynamics can be built as follows:

$$\dot{\xi} = \begin{bmatrix} \xi_2 + \xi_4 + 0.5\xi_3 + \xi_4 \cos(\xi_3) \\ -3(\xi_1 + \xi_3) - 0.5(\xi_2 + \xi_4) - \sin(\xi_3) + \xi_4 \\ -0.5\xi_3 - \xi_4 \cos(\xi_3) \\ \sin(\xi_3) - \xi_4 \end{bmatrix} + [0 \;\; 1 \;\; 0 \;\; 0]^{\mathsf{T}} u + \Delta\mathcal{F}(\xi). \tag{65}$$

Note that the initial conditions and critic network are set and constructed both the same as Example 1. After a sufficient learning session, the weight of the critic network converges to $[0.5481, 0.4121, 0, 0, 0.4476, 0, 0, 0.2030, 0.1506, 0.1005]^{\mathsf{T}}$ and the first five elements are shown in Fig. 8. Finally, one can choose $\theta_1 = 0.5$ and $\theta_2 = -0.5$, apply the robust trajectory tracking control law to the uncertain system (65) for 30 s, and find that the tracking error gradually converges to zero. The tracking error and control trajectories are displayed in Figs. 9 and 10. This verifies the good tracking performance of the original uncertain dynamics to the desired trajectory.

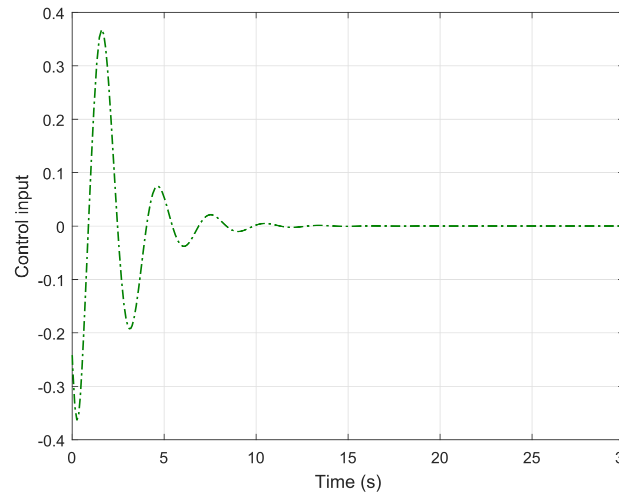


Fig. 10. Tracking control trajectory when $\theta_1 = 0.5$ and $\theta_2 = -0.5$.

V. Conclusion

In this paper, the self-learning robust trajectory tracking control of nonlinear systems with dynamical uncertainties was investigated. Remarkably, the involved uncertain term was more general than the matched case. Via system transformation and adaptive critic design, the approximate optimal control law of the nominal augmented plant can be applied to achieve the robust trajectory tracking of the original uncertain dynamics to the reference signal. The stability issues of the closed-loop augmented systems were analyzed, both for the nominal and uncertain plants. Note that all of the aforementioned results were derived for continuous-time general nonlinear systems. When regarding with the discrete-time case, there were also many results of adaptive-critic-based optimal tracking control design (see, e.g., [25]–[28] and the references therein), but most of them were conducted for normal systems without uncertain dynamics. Though discrete-time nonlinear systems were not the main topic of this paper, their robust trajectory tracking issue should be paid attention as well in the future study. In addition, considering the real constraints in optimization, the robust trajectory tracking problems with control constraints will also be investigated in the future work.

References

[1] H. Yan, F. Qian, H. Zhang, F. Yang, and G. Guo, "H_∞ fault detection for networked mechanical spring-mass systems with incomplete information," *IEEE Trans. Ind. Electron.*, vol. 63, no. 9, pp. 5622–5631, Sep. 2016.

[2] H. Zhang, Q. Hong, H. Yan, F. Yang, and G. Guo, "Event-based distributed H_∞ filtering networks of 2DOF quarter-car suspension systems," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 312–321, Feb. 2017.

[3] F. Lin, *Robust Control Design: An Optimal Control Approach*. New York, NY, USA: Wiley, 2007.

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

[5] X. Zhong, H. He, and D. V. Prokhorov, "Robust controller design of continuous-time nonlinear system using neural network," in *Proc. Int. Joint Conf. Neural Netw.*, Dallas, TX, USA, Aug. 2013, pp. 1–8.

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

[7] J. Sun, C. Liu, and Q. Ye, "Robust differential game guidance laws design for uncertain interceptor-target engagement via adaptive dynamic programming," *Int. J. Control*, vol. 90, no. 5, pp. 990–1004, 2017.

[8] Q. Y. Fan and G. H. Yang, "Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 165–177, Jan. 2016.

[9] Y. Jiang and Z. P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 882–893, May 2014.

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

[11] W. Gao, Y. Jiang, Z. P. Jiang, and T. Chai, "Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming," *Automatica*, vol. 72, pp. 37–45, 2016.

[12] 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.

[13] T. Dierks and S. Jagannathan, "Optimal control of affine nonlinear continuous-time systems," in *Proc. Amer. Control Conf.*, Baltimore, MD, USA, Jun. 2010, pp. 1568–1573.

[14] Y. Lv, J. Na, Q. Yang, X. Wu, and Y. Guo, "Online adaptive optimal control for continuous-time nonlinear systems with completely unknown dynamics," *Int. J. Control*, vol. 89, no. 1, pp. 99–112, 2016.

[15] D. Zhao, Q. Zhang, D. Wang, and Y. Zhu, "Experience replay for optimal control of nonzero-sum game systems with unknown dynamics," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 854–865, Mar. 2016.

[16] H. Zhang, L. Cui, and Y. Luo, "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 206–216, Feb. 2013.

[17] D. Liu, D. Wang, F. Y. Wang, H. Li, and X. Yang, "Neural-network-based online HJB solution for optimal robust guaranteed cost control of continuous-time uncertain nonlinear systems," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2834–2847, Dec. 2014.

[18] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2226–2236, Dec. 2011.

[19] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica*, vol. 50, pp. 1780–1792, 2014.

[20] H. Modares, F. L. Lewis, and Z. P. Jiang, "H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2550–2562, Oct. 2015.

[21] N. T. Luy, "Robust adaptive dynamic programming based online tracking control algorithm for real wheeled mobile robot with omni-directional vision system," *Trans. Inst. Meas. Control*, vol. 39, no. 6, pp. 832–847, 2017.

[22] X. Yang, D. Liu, Q. Wei, and D. Wang, "Guaranteed cost neural tracking control for a class of uncertain nonlinear systems using adaptive dynamic programming," *Neurocomput.*, vol. 198, pp. 80–90, 2016.

[23] C. Mu, Z. Ni, C. Sun, and H. He, "Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1460–1470, Jun. 2017.

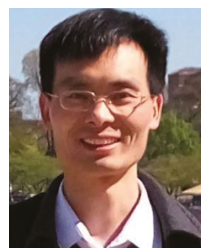
[24] K. G. Vamvoudakis, A. Mojoodi, and H. Ferraz, "Event-triggered optimal tracking control of nonlinear systems," *Int. J. Robust Nonlinear Control*, vol. 27, no. 4, pp. 598–619, 2017.

[25] C. Mu, C. Sun, A. Song, and H. Yu, "Iterative GDHP-based approximate optimal tracking control for a class of discrete-time nonlinear systems," *Neurocomput.*, vol. 214, pp. 775–784, 2016.

[26] A. Heydari and S. N. Balakrishnan, "Fixed-final-time optimal tracking control of input-affine nonlinear systems," *Neurocomput.*, vol. 129, pp. 528–539, 2014.

[27] B. Luo, D. Liu, T. Huang, and D. Wang, "Model-free optimal tracking control via critic-only Q-learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 10, pp. 2134–2144, Oct. 2016.

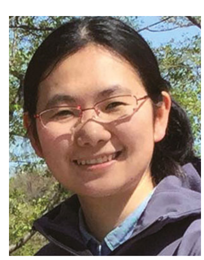
[28] H. Zhang, R. Song, Q. Wei, and T. Zhang, "Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 1851–1862, Dec. 2011.



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

From December 2015 to January 2017, he was a Visiting Scholar in the Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, Kingston, RI, USA. He is currently an Associate Professor with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published more than 100 journal and conference papers and coauthored two monographs. His research interests include adaptive and learning systems, computational intelligence, and intelligent control.

Dr. Wang received the Excellent Doctoral Dissertation Award of the Chinese Academy of Sciences in 2013, as well as a nomination of the Excellent Doctoral Dissertation Award of the Chinese Association of Automation in 2014. He was selected for the Young Elite Scientists Sponsorship Program by the China Association for Science and Technology in 2017. He serves as an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS AND NEUROCOMPUTING. He is the Publications Chair of the 24th International Conference on Neural Information Processing (ICONIP 2017).



Chaoxu Mu (M'15) received the Ph.D. degree in control science and engineering from Southeast University, Nanjing, China, in 2012.

From October 2010 to November 2011, she was a Visiting Ph.D. Student in the Royal Melbourne Institute of Technology University, Melbourne, VIC, Australia. From December 2014 to August 2016, she was a Postdoctoral Fellow in the Department of Electrical, Computer, and Biomedical Engineering, The University of Rhode Island, Kingston, RI, USA. She is currently an Associate Professor in the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. She is the author of more than 70 journal and conference papers and has coauthored one monograph. Her current research interests include nonlinear system control and optimization, adaptive and learning systems.

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