IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 18, 2019, accepted October 3, 2019, date of publication October 7, 2019, date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945970

Optimal Control of Nonlinear Systems With Time Delays: An Online ADP Perspective

JING ZHU [1,2], (Member, IEEE), YIJING HOU [1], (Student Member, IEEE), AND TAO LI [1,2]
[1]College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[2]Key Laboratory of Navigation, Control and Health-Management Technologies of Advanced Aerocraft, Ministry of Industry and Information Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Corresponding author: Jing Zhu (drzhujing@nuaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61603179, Grant 61873123, and Grant 61873127, and in part by the China Postdoctoral Science Foundation under Grant 2019T120427.

ABSTRACT Drawing upon Lyapunov stability theories and online adaptive dynamic programming (ADP) technique, we propose a novel optimal control scheme for the nonlinear time delay system. Our contribution is twofold. First, we investigate the asymptotical stability problem and obtain a generalized stability condition in terms of linear matrix inequalities (LMIs). An explicit, easy-computing delay bound is presented by virtue of *Gronwall's inequality*. Second, we propose the neural network (NN)-based optimal control strategy by utilizing two approximate NNs. The NN-based optimal control law converges to the real optimal control law since that the estimation errors of NNs weights converge to zero. Numerical examples are presented to illustrate our results.

INDEX TERMS Adaptive dynamic programming (ADP), delay systems, neural network (NN), nonlinear systems, optimal control.

I. INTRODUCTION

For various control systems and communication networks, time delays are widely existed and may cause bad performance, poor robustness, and even task failure. Therefore, it is of fundamental significance to analyze the stability and control problems on systems subject to time delays. In general, time delay systems are analyzed either by time-domain or frequency-domain approach (see [1]–[3], and the references therein). For linear systems with constant time delay, it is readily available to the necessary and sufficient conditions on stability in terms of *linear matrix inequalities* (LMIs) from time-domain theories [4] and *small gain theorem* from frequency-domain techniques [5], to name a few. On the other hand, it is relatively far more complicated when dealing with the stability of nonlinear time delay systems. Over the current literature, the stability on nonlinear systems subject to constant time delays is mainly investigated based on *sliding mode control* theories [6], H_∞ *optimization* method [7], LMI techniques [8], neural networkds (NNs) and *adaptive control* maneuver [9], etc. The past work, nevertheless, sheds few lights on the development of nonlinear time delay system theories, and leaves it an open problem yet up to now.

At the same time, optimal control theory has gained lots of attention and progress during the last decade. There are numerous results applicable to optimal control of linear systems by using classical methods, such as *maximum principle*, *convex optimization*, *dynamic programming* [10], while the counterparts for nonlinear systems are much fewer, especially for delayed nonlinear systems. The difficulties in optimal control of nonlinear systems largely arises from solving the Hamilton-Jacobi-Bellman (HJB) equations, which for the case usually consist of nonlinear partial differential equations. To solve nonlinear HJB equations, an *adaptive dynamic programming* (ADP) algorithm was first proposed in [11]. In recent studies, an iterative ADP (also termed as online ADP) algorithm was advocated to set up the infinite horizon oriented optimal control concerning nonlinear systems [12], while in [13], [14] and [15], the online ADP algorithm was adopted to discrete-time nonlinear systems with internal dynamics, zero-sum nonlinear differential games and continuous-time chaotic systems, respectively. The optimal control policy of nonlinear system thus exhibits a somehow flourishing development (see [16], [17] and the references therein).

Despite the considerable advances on the investigation of nonlinear plants, the optimal control of nonlinear systems subject to time delays poses a great challenge to researchers

The associate editor coordinating the review of this manuscript and approving it for publication was Bohui Wang.

as the lack of useful methodologies and tools. The difficulties by nature arise from the coupled delay states in the nominal plant. In this vein, the current state is determined upon the past states in terms of delays and the obtained control law, which is unknown until the the current time. Available results concerning the optimal control of nonlinear time-delay systems remain minor.

In this paper, we consider a class of nonlinear plants against uncertain delays. Our contribution is twofold. First, we seek to find out the largest admissible time delay of the objective plant and sufficient asymptotical stability conditions are obtained in virtue of Lyapunov-Krasvovskii functionals and mathematic inequalities theories. Moreover, an explicit, easy-computing delay bound is presented by virtue of *Gronwall's inequality*. On the other hand, drawing upon online ADP technique, the NN-based optimal control law is proposed by using two NNs to approximate the performance index function and optimal input, respectively. It is proved the weight estimation errors converge to zero. Therefore, the NN-based approximate optimal control law converges to the real optimal control law. Numerical examples are presented to show the effectiveness of our results.

The remainder of this paper is organized as follows. In Section II, we present the mathematic formulation of the nonlinear time delay system and introduce some preliminary lemmas. The stabiliyt conditions and NN-based optimal control solution are obtained in Section III. Illustrative examples are demonstrated in Section IV and conclusions are subsequently followed in Section V.

II. PROBLEM FORMULATION

In the beginning of this section, we introduce our notations used in this paper. Let $\mathbb{R}$ refer to the space of real numbers, $\mathbb{R}^n$ and $\mathbb{R}^n_+$ refer to the space of real vectors and positive real vectors with n-dimension, and $\mathbb{R}^{l\times p}$ refers to the space of real vectors with $l \times p$-dimension. For any real matrix A and function $x(t)$, A^T and $x^T(t)$ denote the transpose of A and $x(t)$, respectively. For a Hermitian matrix A, its largest eigenvalue is denoted as $\lambda(A)$. The expression $A \geq 0$ means A is nonnegative definite, and $A > 0$ means it is positive definite, while $A < 0$ means it is negative definite. Let $\|x(t)\|$ denote the Euclidean norm of $x(t)$. Then it is defined as

$$\|x(t)\| = \sqrt{x^T(t)x(t)}.$$

The symbol ∇_x denotes the partial differential operator $\frac{\partial\cdot}{\partial x}$.

Consider the following nonlinear system

$$\dot{x}(t) = Ax(t-\tau) + f(t, x(t), x(t-\sigma))u(t), \tag{1}$$

where $x(t), u(t) \in \mathbb{R}^n$ are the system state and input signal, $\tau, \sigma > 0$ are unknown constant, referring the delay parameter in the linear and nonlinear plant, respectively. In NN community, τ is also known as leakage delay, and σ is termed as transmission or processing delay, see [18], [19], etc. By considering two independent time delay constants, we deal with the generalized circumstance of nonlinear systems with leakage delays and/or transmission delays. Assume that A is a constant real matrix with appropriate dimension, and $f(t, x(t), x(t-\sigma))$ is nonlinear and Lipschitz continuous with $f(t, 0, 0) = 0$. By employing the so-called *model transformation* [2], the original system (1) can be expressed as

$$\frac{d}{dt}\left(x(t)+A\int_{t-\tau}^{t} x(u)du\right)=Ax(t)+f(t, x(t), x(t-\sigma))u(t). \tag{2}$$

It is worth noting that by taking model transformation additional dynamics will be introduced into the system. As such, the stability of the system (2) implies that of the system (1) but not vice-versa [20].

Our purpose in this paper is to design the optimal control law in terms of input $u(t)$ to optimize the performance index function below

$$J(x(t), u(t)) = \int_t^\infty L(x(s), u(s))ds, \tag{3}$$

where

$$L(x(s), u(s)) = x^T(s)Qx(s) + u^T(s)Ru(s), \tag{4}$$

with matrix Q being nonnegative definite, and $R = R^T$ being symmetric and positive definite. As a consequence, the HJB function is

$$H(x, u, t) = L(x(t), u(t)) + J_x^T(Ax + fu), \tag{5}$$

with $J_x = \partial J/\partial x$ referring to the partial derivative of $J(x(t), u(t))$. Obviously, equation (5) satisfies

$$\min\{H(x, u, t)\} = 0. \tag{6}$$

Hence, the optimal performance index function is derived as

$$J^*(x(t), u(t)) = \min \int_t^\infty L(x(s), u(s))ds,$$

and the corresponding optimal input is

$$u^* = -\frac{1}{2}R^{-1}f^T J_x^*.$$

Consequently, the nominal system (2) is equivalent to the following if the optimal control is achieved

$$\frac{d}{dt}\left(x(t)+A\int_{t-\tau}^{t} x(u)du\right) = Ax(t) - \frac{1}{2}f(t, x(t), x(t-\sigma))R^{-1}f^T J_x^*,$$

which can be further expressed as

$$\frac{d}{dt}\left(x(t)+A\int_{t-\tau}^{t} x(u)du\right)=Ax(t)+\mathcal{F}(t, x(t), x(t-\sigma)), \tag{7}$$

with $\mathcal{F}(\cdot)$ being a nonlinear function

$$\mathcal{F}(t, x(t), x(t-\sigma)) = -\frac{1}{2}f\left(t, x(t), x(t-\sigma)\right)R^{-1}f^T J_x^*.$$

As such, we are led to examine the stability on system (7) instead of the original system (1). To facilitate control system design, the following assumption is presented.

Assumption 1: The nonlinear function $\mathcal{F}(t,x,y)$ satisfies

$$\|\mathcal{F}(t,x,y)-\mathcal{F}(t,x_1,y_1)\| \le \alpha\|x-x_1\|^2+\beta\|y-y_1\|^2, \quad (8)$$

where $t,x,y,x_1,y_1 \in \mathbb{R}^n$, and α, β are some positive scalar.

Under Assumption 1, one can conclude that the original system (1) has unique equilibrium if

$$\sqrt{\alpha+\beta}\|A^{-1}\| \le 1, \quad (9)$$

according to the famous *contraction mapping* theorem [21]. The detailed proof can be found in [22], thus is omitted.

III. MAIN RESULTS

A. STABILITY ANALYSIS ON NONLINEAR TIME-DELAY SYSTEMS

In this section, we aim to obtain stability condition for nonlinear systems with time delays, including a sufficient asymptotically stable condition in terms of LMIs and an explicit delay bound independent of the nonlinear delay parameter. The following preliminary lemma, adopted from [23], concerned with mathematic inequalities on matrices and plays a vital role in our subsequent stability results.

Lemma 1: Given real matrices Ξ_1, $\Xi_2 \in \mathbb{R}^{l\times p}$. The following inequality holds

$$\Xi_1^T\Xi_2+\Xi_2^T\Xi_1 \le \eta\Xi_1^T\Xi_3\Xi_1+\eta^{-1}\Xi_2^T\Xi_3^{-1}\Xi_2, \quad (10)$$

for some real matrix $\Xi_3 \in \mathbb{R}^{p\times p}$ and real scalar η with $\Xi_3=\Xi_3^T>0$, $\eta>0$.

Upon Lemma 1 and Lyapunov stability theories, we now investigate the asymptotical stability of the original system (1). This gives rise to the following theorem.

Theorem 1: The system (1) is asymptotically stable if for some positive real scalar α, β, η_1, and η_2, there exists a real matrix $P=P^T>0$ such that the following LMI holds

$$\Phi<0, \quad (11)$$

where

$$\begin{aligned}\Phi &= Q_1+\tau Q_2+Q_3,\\ Q_1 &= PA+A^TP+\eta_1^{-1}P^2+\alpha(\eta_1+\tau\eta_2)I+\tau A^TPA,\\ Q_2 &= A^TPA+\eta_2^{-1}A^TP^2A,\ Q_3=\beta(\eta_1+\tau\eta_2)I,\end{aligned}$$

and I is an identity matrix with appropriate dimension.

Proof: We begin by constructing the following Lyapunov-Krasovskii functional $V(t)$, which consists of $V_1(t)$ and $V_2(t)$. Let

$$V_1(t)=\left(x(t)+A\int_{t-\tau}^{t}x(u)du\right)^T P\left(x(t)+A\int_{t-\tau}^{t}x(u)du\right),$$

and

$$V_2(t)=\int_{t-\tau}^{t}\int_{s}^{t}x^T(u)Q_2x(u)duds+\int_{t-\sigma}^{t}x^T(u)Q_3\,x(u)du.$$

In light of model transformation and equation (7), we are led to

$$\dot V_1(t)=2\left(x(t)+A\int_{t-\tau}^{t}x(u)du\right)^T P\,(Ax(t)+\mathcal{F}(t,x(t),x(t-\sigma))).$$

It is worth noting that above function has an upper bound

$$\dot V_1(t) \le x^T(x)Q_1x(t)+\int_{t-\tau}^{t}x^T(u)Q_2x(u)du + x^T(t-\sigma)Q_3x(t-\sigma),$$

according to Lemma 1.

In the similar manner, the time derivative of $V_2(t)$ yields to

$$\dot V_2(t)=x^T(t)(\tau Q_2+Q_3)x(t)-\int_{t-\tau}^{t}x^T(u)Q_2x(u)du - x^T(t-\sigma)Q_3x(t-\sigma).$$

Consequently, its upper bound can be computed as

$$\dot V(t) \le x^T(t)\Phi x(t),$$

which indicates that $\dot V(t)$ is negative if Φ is negative definite. The proof is thus completed. □

Theorem 1 provides the generalized sufficient LMI stability criterion for the nonlinear system (1) subject to time delays with the aid of matrix inequality and model transformation techniques. In what follows, we shall introduce another matrix inequality tool, the *Gronwall's inequality*, to generate an alternative, easy computing stability condition [24]. The obtained condition might be conservative to some systems since that it is independent of the nonlinear delay parameter. However, there do exist important instances where the condition (14) tend to be nonconservative, which will be subsequently demonstrated in Section IV.

Lemma 2: Assume that $f(t) \in \mathbb{R}$ and $\eta(t) \in \mathbb{R}$ with $\eta(t)>0$ are continuous functions. If $\zeta(t)$ is a non-decreasing function and satisfies

$$f(t) \le \zeta(t)+\int_{a}^{t}\eta(s)f(s)ds, \quad (12)$$

then the following inequality holds

$$f(t) \le \zeta(t)e^{\int_a^t \eta(s)ds}. \quad (13)$$

According Lemma 1 and 2, we obtain the following corollary in which explicit delay bound on τ are derived.

Corollary 1: Consider the nonlinear system (1) subject to time delays. The system (1) is asymptotically stable if A is stable and

$$\tau < \|A\|^{-1}. \quad (14)$$

Proof: Let the Lyapunov-Krasovskii functional $V(t)$ be the same as in Theorem 1. Then for $t=t_0$, we have

$$\begin{aligned}V(t_0)=&\left(x(t_0)+A\int_{t_0-\tau}^{t_0}x(u)du\right)^T P\left(x(t_0)+A\int_{t_0-\tau}^{t_0}x(u)du\right)\\ &+\int_{t_0-\tau}^{t_0}\int_{s}^{t_0}x^T(u)Q_2x(u)duds\\ &+\int_{t_0-\sigma}^{t_0}x^T(u)Q_3x(u)du,\end{aligned}$$

where Q_2, Q_3 are as defined in Theorem 1. Therefore,

$$\begin{aligned} V(t) \leq\ & \lambda(P)\left\|x(t_0)+A\int_{t_0-\tau}^{t_0}x(u)du\right\|^2 \\ &+\lambda(Q_2)\int_{t_0-\tau}^{t_0}\int_{s}^{t_0}\|x(u)\|^2duds \\ &+\lambda(Q_3)\int_{t_0-\sigma}^{t_0}\|x(u)\|^2du \\ \leq\ & \left((1+\tau\|A\|)^2\lambda(P)+\frac{1}{2}\tau^2\lambda\left(Q_2\right)+\sigma\lambda(Q_3)\right)\|x(t_0)\|^2. \end{aligned}$$

Assume that $\|x(t_0)\|^2 = \Gamma$, and M is some positive scalar. As proved in Theorem 1, it yields to

$$V(t) < V(t_0) < M\Gamma < \infty, \quad \text{for } t > t_0.$$

Evidently for $t > t_0$,

$$\lambda(P)\left\|x(t)+A\int_{t-\tau}^{t}x(u)du\right\|^2 \leq M\Gamma,$$

which leads to

$$\|x(t)\| \leq \|A\|\int_{t-\tau}^{t}\|x(u)\|du+\sqrt{\frac{M\Gamma}{\lambda(P)}}.$$

Based on Lemma 2, we are led to

$$\|x(t)\| \leq \sqrt{\frac{M\Gamma}{\lambda(P)}}e^{\tau\|A\|}, \quad t > t_0. \tag{15}$$

Hence the system is asymptotically stable if

$$\tau\|A\| < 1.$$

This completes the proof. □

Note that the stability of A implies that the system (1) is stable when the linear plant is delay free and the nonlinear delay parameter goes infinity, i.e., $\tau = 0$, $\sigma \to \infty$. Hence the stability condition obtained in Corollary 1 is independent of the nonlinear delay parameter σ. In other words, σ can be arbitrarily large as long as the linear delay parameter τ satisfies condition (14). As such, it has large freedom in the design of optimal controller; nevertheless, the stability criterion herein is somehow conservative. One may develop refined delay-dependent stability conditions with less conservatism by means of advanced mathematic inequalities techniques.

B. NN-BASED ONLINE ADP ALGORITHM

In the following paragraph, we shall develop an online ADP algorithm to achieve the optimal control policy with the use of NNs approximations. We utilize a two-layer NN, consists of critic NN and actor NN, to estimate the performance index function and optimal input respectively. Under this structure, the weights for both NN are iteratively tuned in real-time.

Let $W_c \in \mathbb{R}^{l\times p}$ refer to the ideal weight matrix in the critic NN, $\phi_c(x)$ refer to the activation function and $\epsilon_c(x)$ refer to the approximation error. As for the actor NN, let $W_a \in \mathbb{R}^{l\times p}$ refer to the real weights, $\phi_a(x)$ refer to the activation function, and $\epsilon_a(x)$ refer to the approximation error. It is useful to make the following assumption.

Assumption 2: a) *The approximation errors of critic NN and actor NN are positively bounded by $\|\epsilon_c\| \leq \epsilon_{cM}$ and $\|\epsilon_a\| \leq \epsilon_{aM}$.*

b) *The residual error ϵ_H is positively bounded by $\|\epsilon_H\| \leq \epsilon_{HM}$.*

c) *The activation function of the actor NN is positively bounded by $\phi_m \leq \|\phi_a\| \leq \phi_M$.*

Upon Assumption 2, the performance index function $J(x(t), u(t))$ of critic NN thus turns to be

$$J(x) = W_c^T\phi_c(x) + \epsilon_c(x). \tag{16}$$

Based on the *universal approximation property* [25], there always exists a NN such that $\epsilon_c(x)$ is bounded. The HJB function is

$$\begin{aligned} H(x(t), u(t), W_c) = W_c^T\nabla_x\phi_c\left(\dot{x}\right)+x^TQx+u^TRu \\ +\nabla_x\epsilon_c^T\left(\dot{x}\right)-\epsilon_H, \end{aligned} \tag{17}$$

thus leading to

$$\epsilon_H = W_c^T\nabla_x\phi_c\left(\dot{x}\right)+x^TQx+u^TRu+\nabla_x\epsilon_c\left(\dot{x}\right),$$

since that

$$H(x(t), u(t), W_c) = 0.$$

Let $\hat{W}_c \in \mathbb{R}^{l\times p}$ refer to the real weight matrix of critic NN. Then the estimation of equation (16) is

$$\hat{J}(x) = \hat{W}_c^T\phi_c(x). \tag{18}$$

The corresponding HJB function can be further expressed as

$$H(x(t), u(t), \hat{W}_c) = \hat{W}_c^T\nabla_x\phi_c\left(\dot{x}\right)+x^TQx+u^TRu. \tag{19}$$

Let $\tilde{W}_c$ refer to the error of weight estimation with

$$\tilde{W}_c = W_c - \hat{W}_c.$$

The estimation error of HJB equation e_c thus is computed as

$$e_c = H(x(t), u(t), W_c) - H(x(t), u(t), \hat{W}_c),$$

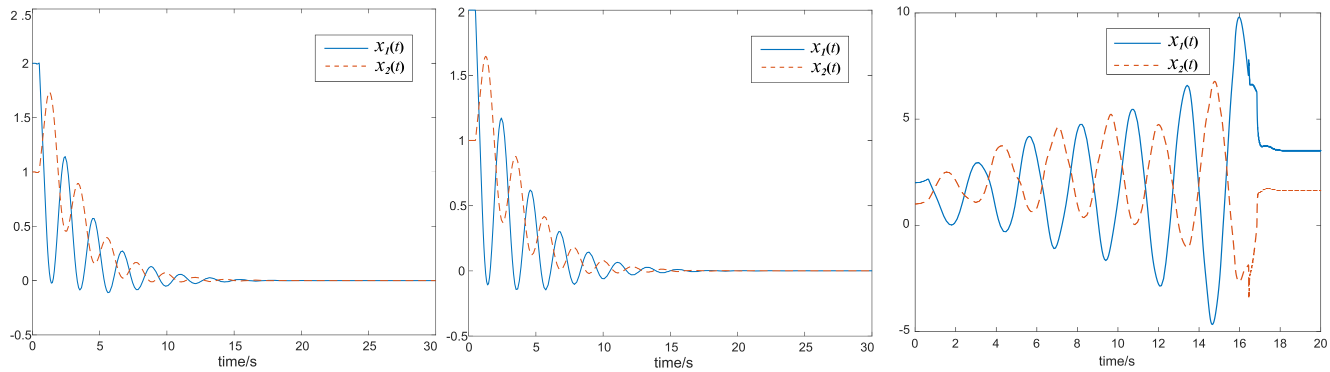
or equivalently,

$$e_c = H(x(t), u(t), \tilde{W}_c) - \epsilon_H.$$

Hence, the squared estimation error is $E_c = \frac{1}{2}e_c^Te_c$. We are then led to find the optimal weight update law such that E_c is minimized.

In light of *Levenberg-Marquardt algorithm* (LMA) [26], the tuning rule of critic NN weight is derived as

$$\begin{aligned} \dot{\hat{W}}_c &= -\theta_c\frac{\partial E_1}{\partial\hat{W}_c} \\ &= -\frac{\theta_c\varsigma\left(\varsigma^T\hat{W}_c+x^TQx+u^TRu\right)}{\left(\varsigma^T\varsigma+1\right)^2}, \end{aligned} \tag{20}$$



(a) $\tau = 0.5,\ \sigma = 0.2$ (b) $\tau = 0.5,\ \sigma = 1$ (c) $\tau = 0.8,\ \sigma = 1$

FIGURE 1. **State response of the system (28).**

with the learning rate $\theta_c > 0$ and $\varsigma = \nabla_x \phi(\dot{x}(t))$. Moreover, the estimation error of critic NN weight is dynamically updated according to

$$\dot{\tilde{W}}_c = -\frac{\theta_c \varsigma \left(\varsigma^T \tilde{W}_c + \epsilon_H\right)}{\left(\varsigma^T \varsigma + 1\right)^2}. \tag{21}$$

On the other hand, we construct the performance index function of actor NN as

$$u(t) = W_a^T \phi_a(x) + \epsilon_a(x). \tag{22}$$

Let $\hat{u}(t)$ denote the estimated input with $\hat{u}(t) = \hat{W}_a^T \phi_a(x)$, where $\hat{W}_a$ denotes the current weight of actor NN. Hence, the estimation error ϵ_u is equal to

$$\epsilon_u = \hat{W}_a \phi_a + \frac{1}{2} R^{-1} f^T \nabla_x \phi_a^T(x) \hat{W}_c. \tag{23}$$

As a result, the tuning rule of actor NN weight is

$$\dot{\hat{W}}_a = -\theta_a \phi_a \left(\hat{W}_a \phi_a + \frac{1}{2} R^{-1} f^T \nabla_x \phi_a^T(x) \hat{W}_c\right) \tag{24}$$

with the learning rate $\theta_a > 0$. Denote the estimation error of actor NN weight by $\tilde{W}_a$, then

$$\tilde{W}_a = W_a - \hat{W}_a.$$

Consequently, the estimation error of actor NN weight is dynamically updated along with

$$\dot{\tilde{W}}_a = -\theta_a \phi_a \times \left(\tilde{W}_a \phi_a + \frac{1}{2} R^{-1} f^T \nabla\left(\phi_a^T(x)\right) \hat{W}_c + \epsilon_a + \frac{1}{2} R^{-1} f^T \nabla_x \epsilon_c\right) \tag{25}$$

We shall prove that the estimation errors of critic and actor NNs weights are bounded in the following part.

Theorem 2: Let $\tilde{W}_c$, $\tilde{W}_a$ denote the weights estimation errors of critic and actor NNs, and θ_c, θ_a denote the learning rate. Then $\tilde{W}_c$ and $\tilde{W}_a$ are convergent to zero if their weight update policies are as stated in (21), (25), and satisfying

$$0 < \theta_c < \min\left\{\frac{4\theta_a \varsigma_m^2}{2\varsigma_m^2 + \left\|R^{-1} f \nabla_x \phi\right\|^2}, \frac{2\varsigma_m^2}{\varsigma_M^2}\right\}, \tag{26}$$

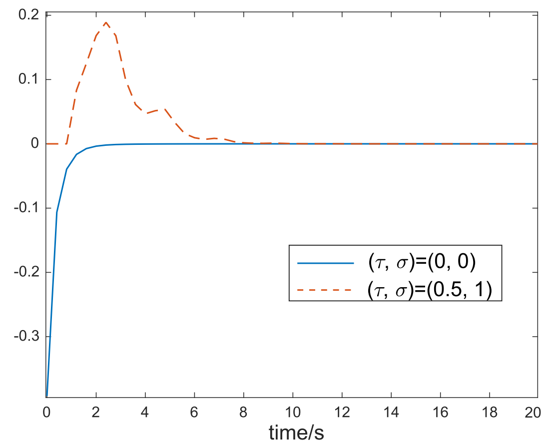


FIGURE 2. **The control trajectory of the system (28).**

$$0 < \theta_a < \frac{4\phi_M^2}{3\phi_m^2}, \tag{27}$$

where ς_m is the lower bound of $\left\|\frac{\varsigma}{(\varsigma^T\varsigma+1)^2}\right\|$, ς_M is the corresponding upper bound, ϕ_m and ϕ_M are as defined in Assumption 2.

Proof: Let the Lyapunov function $\tilde{V}$ be $\tilde{V} = \tilde{V}_1 + \tilde{V}_2$, where

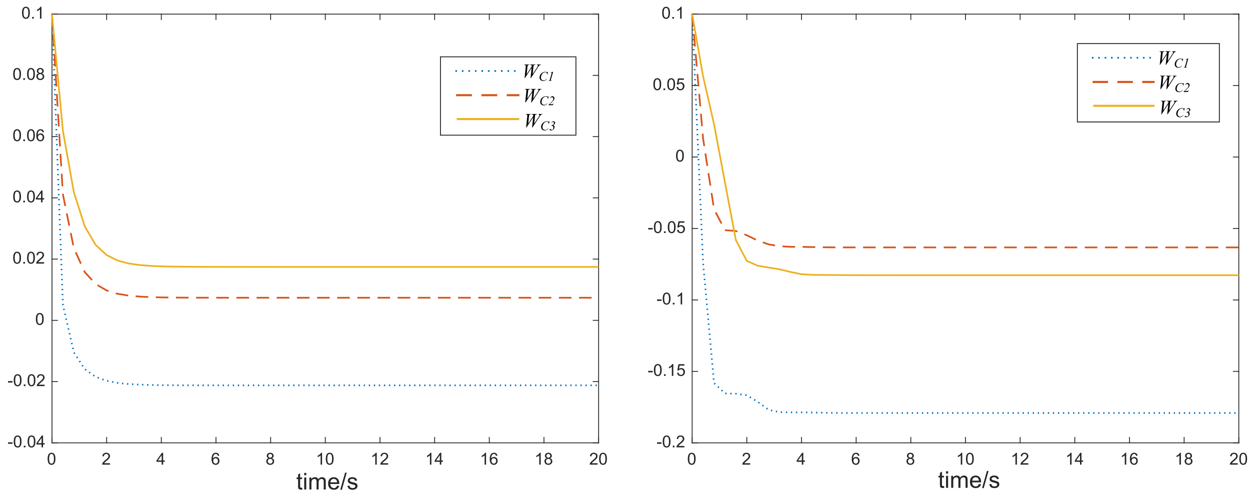
$$\tilde{V}_1 = \frac{1}{2} tr\left(\tilde{W}_c^T \theta_c^{-1} \tilde{W}_c\right),$$

and

$$\tilde{V}_2 = tr\left(\tilde{W}_a^T \theta_c \theta_a^{-1} \tilde{W}_a\right).$$

Consequently, the upper bounds of $\dot{\tilde{V}}_1$ and $\dot{\tilde{V}}_2$ are computed as

$$\begin{aligned}\dot{\tilde{V}}_1 &= \frac{1}{\theta_c} \tilde{W}_c^T \dot{\tilde{W}}_c \\ &= -\frac{\tilde{W}_c^T \theta_c \varsigma \left(\varsigma^T \tilde{W}_c + \epsilon_H\right)}{\left(\varsigma^T \varsigma + 1\right)^2} \\ &\le -\left(\varsigma_m^2 - \frac{1}{2}\theta_c \varsigma_M^2\right) \left\|\tilde{W}_c^2\right\| + \frac{\epsilon_H^2}{2\theta_c},\end{aligned}$$



(a) $\tau = 0,\ \sigma = 0$

(b) $\tau = 0.5,\ \sigma = 1$

FIGURE 3. **The simulation curve of critic NN weights.**

and

$$\begin{aligned}\dot{\tilde{V}}_2 &= \frac{2\theta_c}{\theta_a}\tilde{W}_a^T\tilde{W}_a \\ &\leq -\left(\theta_c - \frac{3\theta_a\theta_c}{4}\right)\left\|\tilde{W}_a\right\|^2\|\phi_m\|^2 \\ &\quad + \frac{\theta_c\left\|R^{-1}f\nabla_x\phi\right\|^2\left\|\tilde{W}_c\right\|^2}{4\theta_a} \\ &\quad + \frac{\theta_c\left(\epsilon_a + 1/2R^{-1}f^T\nabla_x\epsilon_c\right)^2}{2\theta_a}.\end{aligned}$$

The rest proof follows directly from Lyapunov stability theories, thus is omitted. □

In fact, we can achieve only a nearly optimal control by using NN based approach with the existence of NN reconstruction errors in general. Theorem 2, nevertheless, indicates the proposed NN-based optimal control law converges to the real optimal control policy as the convergence of the estimation errors of NNs weights.

IV. EXAMPLES

Metzler systems are widely used in economics, population systems, and biochemical system systems for growth behavior modelling. A nonlinear Metzler delay system is in the form of (1) with $A = [a_{ij}]$ being a Metzlerian matrix satisfying $a_{ii} < 0$ and $a_{ij} \geq 0$ for $i \neq j$ [27]. In this section, several numerical examples with regard to nonlinear Metzler delay systems are presented to examine our algorithm.

Example 1: Consider the following Metzler delay system

$$\dot{x}(t) = \begin{bmatrix} -2 & 1 \\ 1 & -1 \end{bmatrix} x(t-\tau) + \begin{bmatrix} \sin x_2(t-\sigma) \\ \sin x_1(t-\sigma) \end{bmatrix} u(t), \quad (28)$$

with initial state $x(0) = [2\ 1]^T$.

Since that the Metzlerian matrix A is stable, Corollary 1 is applicable to the system (28). As suggested by Condition (14), an easy computing stability condition is obtained by calculating $\|A\|$. Hence, the system (28) is asymptotically stable if $\tau < 0.5$ for all possible σ. However, above stability condition is somehow conservative. Indeed, Fig. 1 (a)-(b) indicate that the nominal plant is stable when $(\tau,\ \sigma) = (0.5, 0.2)$ and $(\tau,\ \sigma) = (0.5, 1)$. However, the nominal system turns to be unstable when $(\tau,\ \sigma) = (0.8, 1)$, as shown in Fig. 1 (c).

More specifically, we consider the Metzler system (28) with $(\tau,\ \sigma)$ being $(0, 0)$ and $(0.5, 1)$, respectively. In other words, we consider a nonlinear delay-free Metzler system and a nonlinear Metzler system with tow distinct delay parameters. Assume the initial weights are

$$W_c = [0.1\ \ 0.1\ \ 0.1]^T,\ W_a = [0\ \ 0\ \ 0]^T.$$

Let $Q = R = 1$, the activation function of critic and actor NNs to be

$$\phi_c(x) = \phi_a(x) = \left[x_1^2\ \ x_1x_2\ \ x_2^2\right]^T.$$

By selecting the learning rates of critic and actor NNs as

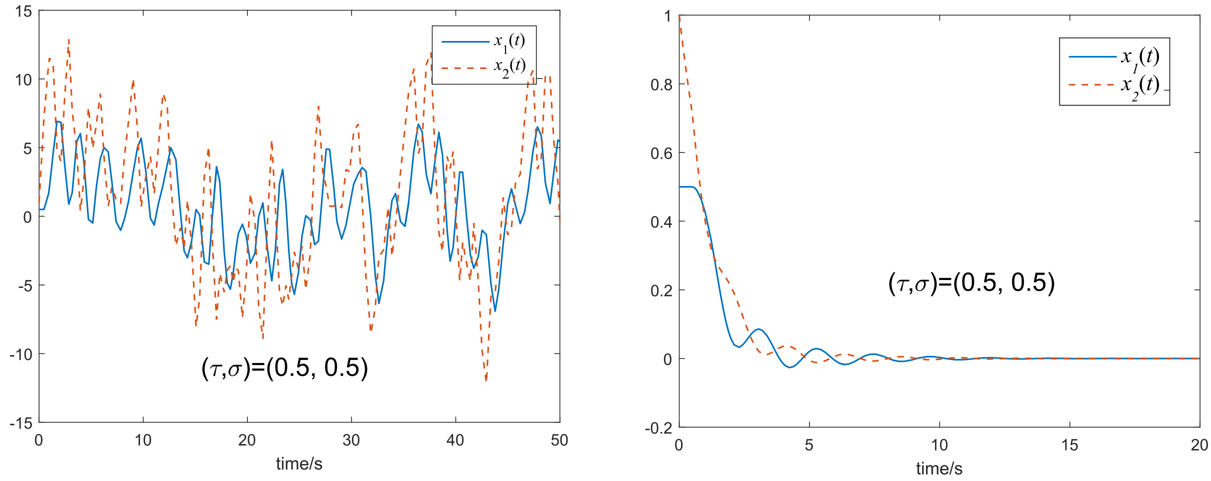
$$\theta_c = \theta_a = 0.01,$$

the control trajectories of above mentioned two systems are compared in Fig. 2. From Fig. 2 we can see that it merely takes 2 seconds to achieve the optimal control when $(\tau, \sigma) = (0, 0)$, while it takes slightly longer, about 10 seconds, to reach the control objective for system (28) with $(\tau, \sigma) = (0.5, 1)$. The convergence of the corresponding critic NN weights is shown in Fig. 3, which indicates that the NN-based optimal control approach is convergent to the real optimal control policy. Consequently, we can conclude that the NN-based optimal control is efficiently achieved under both cases.

Example 2: Consider another Metzler delay system

$$\dot{x}(t) = \begin{bmatrix} -2 & 1 \\ 1 & -1 \end{bmatrix} x(t-\tau) + \begin{bmatrix} 0 \\ \cos x_1(t-\sigma) \end{bmatrix} u(t), \quad (29)$$

with initial state $x(0) = [0.5\ 1]^T$.



(a) Without optimal control (b) With optimal control

FIGURE 4. State response of the system (29).

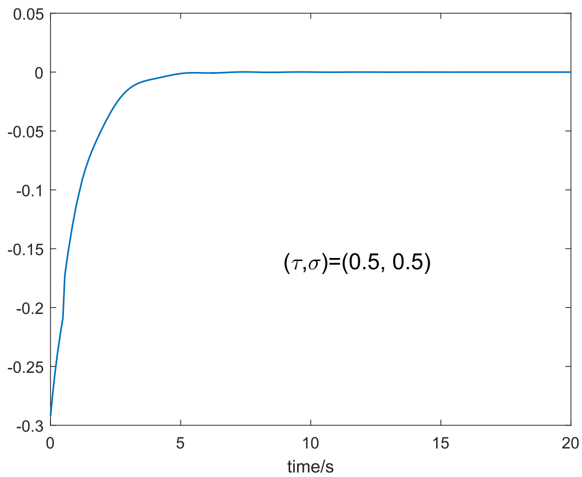


FIGURE 5. The control trajectory of the system (29).

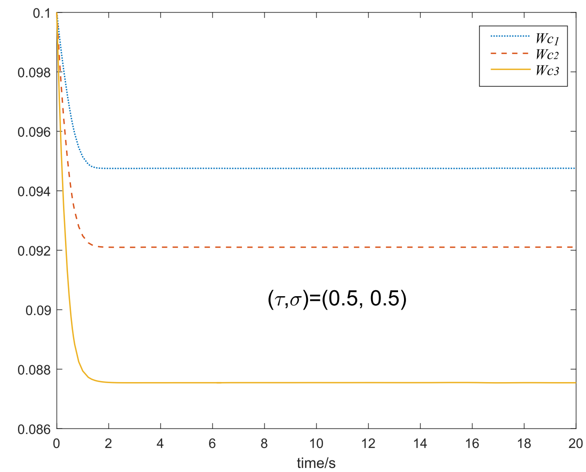


FIGURE 6. The simulation curve of critic NN weights.

Note that the linear time-delay plant of the system (28) and system (29) are the same. Consequently, we are led to the stability condition as in Example 1 that the system (28) is asymptotically stable as long as $\tau < 0.5$, based on Condition (14) in Corollary 1. From Fig. 4(a) we can see that the nominal system (29) is indeed unstable when $\tau = 0.5$, which means that sufficient stability condition herein is tight.

Meanwhile, we set the initial weights to be

$$W_c = [0.1 \;\; 0.1 \;\; 0.1]^T, \qquad W_a = [0 \;\; 0 \;\; 0]^T,$$

and the learning rate of critic and actor NNs to be

$$\theta_c = \theta_a = 0.01.$$

As a result, the NN-based optimal control scheme is eventually achieved with the unstable system being stabilized, as shown in Fig. 4(b) and the control trajectory converging to zero, as shown in Fig. 5. Fig. 6 demonstrates the dynamic of critic NN weights, from which we can conclude that the NN-based optimal control scheme converges to the real optimal control scheme.

V. CONCLUSION

In this paper we propose a NN-based optimal control policy for nonlinear systems subject to time delays. Sufficient asymptotically stability criteria in terms of LMIs and Gronwall's inequality are developed. The NN-based algorithm is derived in virtue of online ADP strategy by using critic NN and actor NN to approximate the optimal cost function and optimal control input. The convergency of the weights estimation errors is proved, thus indicating the optimal control is actually achieved. Our work can be extended in a straightforward manner to nonlinear systems with time-varying delays.

REFERENCES

[1] E. Boukas and Z. K. Liu, *Deterministic and Stochastic Time-Delay Systems*. Boston, MA, USA: Birkhäuser, 2002.

[2] K. Gu, V. Kharitonov, and J. Chen, *Stability of Time-Delay Systems*. Boston, MA, USA: Bitkhäuser, 2003.

[3] W. Michiels and S.-I. Niculescu, *Stability and Stabilization of Time-Delay Systems: An Eigenvalue-Based Approach*. Philadelphia, PA, USA: SIAM, 2007.
[4] P.-L. Liu, "Exponential stability for linear time-delay systems with delay dependence," *J. Franklin Inst.*, vol. 340, nos. 6–7, pp. 481–488, Sep./Nov. 2003.
[5] J. Zhu, T. Qi, D. Ma, and J. Chen, *Limits of Stability and Stabilization of Time-Delay Systems: A Small-Gain Approach*. Cham, Switzerland: Springer, 2018.
[6] Y. Niu, D. W. C. Ho, and J. Lam, "Robust integral sliding mode control for uncertain stochastic systems with time-varying delay," *Automatica*, vol. 41, no. 5, pp. 873–880, 2005.
[7] E. Fridman and U. Shaked, "Stability and guaranteed cost control of uncertain discrete delay systems," *Int. J. Control*, vol. 78, no. 4, pp. 235–246, Aug. 2005.
[8] A. J. Koshkouei and K. J. Burnham, "Discontinuous observers for nonlinear time-delay systems," *Int. J. Syst. Sci.*, vol. 40, no. 4, pp. 383–392, 2009.
[9] S. S. Ge, F. Hong, and T. H. Lee, "Adaptive neural network control of nonlinear systems with unknown time delays," *IEEE Trans. Autom. Control*, vol. 48, no. 11, pp. 2004–2010, Nov. 2003.
[10] K. Ito and K. Kunisch, *Optimal Control*. Hoboken, NJ, USA: Wiley, 1999.
[11] P. J. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *General Syst. Yearbook*, vol. 22, pp. 25–38, 1977.
[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, 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, 2009.
[14] 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.
[15] R. Song, W. Xiao, H. Zhang, and C. Sun, "Adaptive dynamic programming for a class of complex-valued nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 9, pp. 1733–1739, Sep. 2014.
[16] 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.
[17] D. Liu, Q. Wei, X. Yang, H. Li, and D. Wang, *Adaptive Dynamic Programming with Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.
[18] K. Gopalsamy, *Stability and Oscillations in Delay Differential Equations of Population Dynamics*. Norwell, MA, USA: Kluwer, 1992.
[19] B. Kosok, *Neural Networks and Fuzzy Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 1992.
[20] J. Zhu and J. Chen, "Stability of systems with time-varying delays: An L_1 small-gain perspective," *Automatica*, vol. 52, pp. 260–265, 2015.
[21] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, vol. 53, nos. 5–6, pp. 475–504, 2004.
[22] C. Li and T. Huang, "On the stability of nonlinear systems with leakage delay," *J. Franklin Inst.*, vol. 346, no. 4, pp. 366–377, May 2009.
[23] E. N. Sanchez and J. P. Perez, "Input-to-state stability (ISS) analysis for dynamic neural networks," *IEEE Trans. Circuits Syst. I. Fundam. Theory Appl.*, vol. 46, no. 11, pp. 1395–1398, Nov. 1999.
[24] R. Bellman, "The stability of solutions of linear differential equations," *Duke Math. J.*, vol. 10, no. 4, pp. 643–647, 1943.
[25] S. Jagannathan, *Neural Network Control of Nonlinear Discrete-Time Systems*. Boca Raton, FL, USA: CRC Press, 2006.
[26] J. J. Moré, "*The Levenberg-Marquardt Algorithm: Implementation and Theory* (Lecture Notes in Mathematics), vol. 630. Cham, Switzerland: Springer, 1978, pp. 105–116.
[27] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia, PA, USA: SIAM, 1994.

JING ZHU (M'19) was born in Linfen, Shanxi, China, in 1988. She received the B.Eng. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2010, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, in 2015.

She was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of California at Riverside, Riverside, CA, USA, from December 2013 to June 2014. She is currently a Lecturer with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. Her research interests include time-delay systems, robust control, adaptive control, and optimal control.

YIJING HOU (S'19) received the B.S. degree in automation control from Suzhou University, Anhui, China, in 2016. She is currently pursuing the M.S. degree in control engineering with the Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research interests include optimal control, neural networks, adaptive control, nonlinear systems, and time-delay systems.

TAO LI received the Ph.D. degree in engineering from Southeast University, China, in 2008, where he was a Postdoctoral Research Fellow with the School of Instrument Science and Engineering, from 2008 to 2011. He was a Visiting Scholar with the Control System Center, The Manchester University, U.K., from 2016 to 2017. He is currently an Associate Professor with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, China. His current research interests include neural networks, time-delay systems, and networked control systems.