International Journal of Applied Mathematics in Control Engineering 3 (2020) 206-212

Contents lists available at YXpublications

International Journal of Applied Mathematics in Control Engineering

Journal homepage: http://www.ijamce.com

Asymptotically-stable Optimal Control for Chaotic Systems with Input Saturation Using ADP Algorithm

Fule Wang [a], Qiuxia Qu [a,*] Jinyu Liu [a], Liangliang Sun [a], Juan Wang [a], Qiong Xia [a]

[a] *School of Information and Control Engineering, Shenyang Jianzhu University, Shenyang, 110168 China*

ARTICLE INFO

Article history:
Received September 5 2020
Accepted December 10 2020
Available online December 15 2020

Keywords:
adaptive dynamic programming
chaotic systems
input saturation
asymptotically-stable optimal control

ABSTRACT

Based on the adaptive dynamic programming (ADP) approach, a novel approximate optimal control is proposed in this papers for continuous-time chaotic systems with input saturation. This online adaptive algorithm is implemented as an actor/critic structure which involves a critic neural network and an actor neural network to obtain in real-time the approximations of the optimal control cost and optimal policy, respectively. Further, a robustifying term is developed to eliminate the effect caused by the neural network approximation errors, leading to asymptotical stability of the closed loop chaotic system. Lyapunov techniques are used to prove that the proposed optimal constrained controller guarantees the chaotic system states asymptotically stable. The feasibility of the proposed method is confirmed by simulating the control of a Lorenz system.

1. Introduction

Chaotic systems as a nonlinear dynamic behavior have been widely applied in many practical systems and procedures such as in chemical reactions, biological systems, information processing, and so on[1–3]. As an interesting nonlinear phenomenon, chaotic behaviour has been intensively studied during recent years. Many different approaches have been developed to achieve chaos control successfully, such as adaptive control, optimal control, adaptive impulsive control, feedback linearization, and variable structure control[4–8]. Most studies are based on the assumption that the actuator is not saturated during the control process[9–12]. However, actuator will saturate due to its physical limitations in practice. Besides, the saturation of control input may cause the chaotic systems unpredictable results due to its high sensitivity to system parameters[13-16]. Hence, the derivation of controller with input saturation is a challenging problem.

Considering this fact, the controller design for systems with saturation constraints has attracted extensive attention, and some interesting methods have been developed[17-20]. There is a difficulty in adaptive control design of constrained-input nonlinear systems, that is, the stability analysis of closed-loop systems is relatively difficult. At the same time, it is noted that optimal control is a very important aspect in the control field. The accurate constrained optimal control under the minimized performance index is not easy to be obtained, because of the tackle problem of directly obtaining the optimal solution of the Hamilton-Jacobi-Bellman (HJB) equation[17].

In recent years, ADP approach and the related research have received much attention as an effective intelligent control method from researchers [21–34]. The basic principle of ADP is to approximate the optimal performance index function and optimal control function by using function approximation structure or neural network by using reinforcement learning mechanism to meet the optimality principle of dynamic programming. An online adaptive algorithm has been employed to solve optimal tracking control problem for chaotic systems and the result was very effective[23]. But without input saturation taken into consideration, and the iterative cost functions converge to a finite neighborhood of the optimal value.

The need for adaptive algorithm to learn optimal constrained control for chaotic systems, while still guaranteeing asymptotic stability motivates our research. In this paper, based on an online ADP algorithm, the approximate optimal control is designed for chaotic systems with input saturation for the first time. To deal with input saturation in chaotic systems, a suitable nonquadratic functional is used to encode the constraints into the optimization formulation. Then, a policy iteration (PI) algorithm based on an actor/critic NNs structure is developed to solve the associated HJB

* Corresponding author.
E-mail addresses: quqiuxia@sjzu.edu.cn (Q. Qu)

equation online for chaotic Systems. That is, the optimal control policy and the optimal value function are approximated as the output of two NNs, namely an actor NN and a critic NN. The problem of solving the HJB equation is then converted to a problem of simultaneously adjusting the weights of two NNs.

The paper is organized as follows. Section 2 provides the formulation of the optimal control problem for chaotic systems. Section 3 introduces the online ADP algorithm for the actor and critic networks. Results for convergence and stability are given in Section 4. Section 5 presents a simulation example that show the effectiveness of the online PI algorithm. Finally, section 6 provides the conclusions.

2. Problem statement

Consider the following nonlinear continuous-time chaotic systems given by

$$\dot{x}(t) = f\big(x(t)\big) + g\big(x(t)\big)u(t);\ x(0) = x_0, t \geq 0 \qquad (1)$$

where $x \in \mathbb{R}^n$ is the state vector, the control input $u \in U, U = \{u = (u_1, \dots, u_m)\} \subset \mathbb{R}^m, |u_i| \leq \lambda, \forall i \in 1, \dots, m$,where $\lambda \in \mathbb{R}$ is the saturating bound for the actuators. $f(x(t)) \in \mathbb{R}^n$, and $g \in \mathbb{R}^{n\times m}$are known locally Lipschitz continuous functions so that the solution $x(t)$ exists and is unique for any initial condition $x_0$and piece wise continuous control $u(t) \in U$.

The optimal control problem discussed in this paper is to find an admissible policy $u^*(t)$ which satisfies the constraints mentioned above and minimizes the following infinite horizon performance index associated with the system (1)

$$V(x(t)) = \int_t^\infty r\,(x(\tau), u(\tau))d\tau \qquad (2)$$

where $r(x,u) = Q(x) + U(u), Q(x)$ is a positive definite monotonically increasing function and $U(u)$ is a non-negative function.

In fact, many chaotic systems can be described by Eq. (1), such as the Chen system, Lorenz system, Rossler system, Lu system, several variants of Chua's circuits, and Duffing oscillator[7,8,13] .

To guarantee that the control signals satisfy the input constraints, the following nonquadratic functional is proposed in [27,29]. The well-known hyperbolic tangent is used as

$$U(u) = 2\int_0^u (\lambda\psi^{-1}(v/\lambda))^T R dv = 2\int_0^u (\lambda tanh^{-1}(v/\lambda))^T R dv \qquad (3)$$

Using (2)-(3), the performance index becomes

$$V(x(t)) = \int_t^\infty (\,Q(x(\tau) + 2\int_0^u (\lambda\, tanh^{-1}(v/\lambda))^T R dv)d\tau \qquad (4)$$

Differentiating $V(x(t))$ along the system trajectories, the following nonlinear Lyapunov equation(LE) is obtained

$$Q(x) + U(u) + \nabla V^T(x)\big(f(x) + g(x)u(x)\big) = 0, V(0) = 0 \qquad (5)$$

where $\nabla V(x) = \partial V/\partial x$, which denotes the gradient of the value function $V(x)$. Eq. (5) is an infinitesimal version of (4)

Let $V^*(x)$be the optimal cost function defined as

$$V^*(x(t)) = \min_{u(t)\in U} \int_t^\infty r(x(\tau), u(x(\tau))d\tau \qquad (6)$$

Then $V^*(x)$satisfies the following HJB equation

$$\min_{u(t)\in U}[\,Q(x) + U(u) + \nabla V^{*T}(x)(f(x) + g(x)u(x))] = 0 \qquad (7)$$

The optimal control input for the given problem is obtained by differentiating (7) with respect to u. The result is

$$u^*(x) = -\lambda \tanh(\,1/(2\lambda)R^{-1}g^T(x)\nabla V^*(x)) = -\lambda \tanh(\,D^*) \qquad (8)$$

Putting (8) in (3) results in

$$U(u^*) = \lambda\nabla V^{*T}(x)g(x)\tanh(D^*) + \lambda^2 R\ln(\,1 - tanh^2(D^*)) \qquad (9)$$

Where $D^* = 1/(2\lambda)R^{-1}g^T(x)\nabla V^*(x)$.

Substituting $u^*(x)$ and $U(u^*)$ into (5), the HJB equation becomes

$$Q(x) + U(u^*) + \nabla V^{*T}(x)\big(f(x) + g(x)u^*(x)\big) = Q(x) + \nabla V^{*T}(x)f(x) + \lambda^2 R\ln(\,1 - tanh^2(D^*)) = 0, \quad V(0) = 0 \qquad (10)$$

The HJB equation (10) is a nonlinear partial differential equation which is extremely difficult to solve. This is the motivation of introducing an synchronous online Policy Iteration Algorithm (PI) for approximating the HJB solution.

The following section provides approximate techniques to converge to the solution of the HJB equation (10).

3. Online ADP algorithm for chaotic systems with input saturation

The learning structure in this paper uses value function approximation (Werbos, 1992) with two neural networks(NNs), namely an actor NN and a critic NN. The critic NN is trained to become an approximation of the value function solution at the policy evaluation step, while the actor NN is trained to approximate an optimal policy at the policy improving step. Both actor and critic NNs are updated simultaneously in real time. We call this synchronous online PI. In the following two subsections, an online PI algorithm is now given to learn the optimal control solution for chaotic systems with constrained-input.

3.1 Value function approximation using critic NN

In the NN, if the number of hidden layer neurons is L, the weight matrix between the input layer and hidden layer isW_1, the weight matrix between the hidden layer and output layer isW_2, and the input vector of the neural network is X, then the output of a three-layer neural network is expressed as

$$F(X, Y_1, Y_2) = Y_2^T\sigma'\,(Y_1^T X) \qquad (11)$$

where $\sigma'\,(Y_1^T X)$is the activation function. For convenience of analysis, only the output weight Y_2 is updated during the training, while the hidden weight Y_1 is kept unchanged. Hence, in the following part, the neural network function can be simplified into

$$F(X, Y_2) = Y_2^T\sigma(X) \qquad (12)$$

Assuming the value function solution to the HJB equation (10) is a smooth function, then according to the Weierstrass high-order approximation theorem, there exists a NN such that the solution $V^*(x)$ and its gradients $\nabla V^*(x)$ with respect to X can be uniformly approximated as

$$V^*(x) = W_1^T\phi_1(x) + \varepsilon_1(x) \qquad (13)$$

$$\nabla V^*(x) = \nabla\phi_1^T W_1 + \nabla\varepsilon_1 \qquad (14)$$

where $W_1 \in \mathbb{R}^L$is the ideal constant weights, $\phi_1(x)$: $\mathbb{R}^n \to \mathbb{R}^L$is the suitable basis activation function of critic network, and $\varepsilon_1(x)$ is the approximation error. Besides, the gradients of$\phi_1$and $\varepsilon_1(x)$ is $\nabla\phi_1 = \partial\phi_1/\partial x, \nabla\varepsilon_1 = \partial\varepsilon_1/\partial x$, respectively.

Using the NN value function approximation, considering a fixed control policy $u(t)$, the nonlinear LE (5) becomes

$$H(x, u, W_1) = Q + U(u) + W_1^T\nabla\phi_1(f + gu) = e_L \qquad (15)$$

where the residual error due to the function approximation error is

$$e_L = -\nabla\varepsilon_1^T(f + gu) \qquad (16)$$

The following standard assumptions are considered for the critic

NN approximators in this paper.

Assumption 1. The critic NN activation functions ϕ_1, reconstruction error ε_1, their gradients, and the Hamiltonian residual error e_L are bounded over the compact set. In the sense, there exit finite constants ϕ_{1M}, ε_{1M}, ϕ_{1dM}, ε_{1dM}, $e_{LM} \in \mathbb{R}^+$ such that $\|\phi_1\| \le \phi_{1M}$, $\|\varepsilon_1\| \le \varepsilon_{1M}$, $\|\nabla\phi_1\| \le \phi_{1dM}$, $\|\nabla\varepsilon_1\| \le \varepsilon_{1dM}$, $\|e_L\| \le e_{LM}$.

Consider a fixed control policy u and assume that its corresponding value function is approximated by the critic NN (13). Since the ideal weights W_1 of the critic NN in (13) are unknown and must be approximated in real time. Hence, consider the critic weight estimates $\widehat{W}_1 \in \mathbb{R}^L$, then the approximate value function and the approximate nonlinear LE can be written as

$$\hat{V}(x) = \widehat{W}_1^T \phi_1(x), \forall x \tag{17}$$

$$H(x, u, \widehat{W}_1) = Q + U(u) + \widehat{W}_1^T \nabla\phi_1(f + gu) = e_H, \forall x, u \tag{18}$$

where residual error e_H is produced by the NN approximation error. Under the Lipschitz assumption on the dynamics, this residual error is bounded on a compact set. Therefore, it is desired to select tuning law for the weight estimates $\widehat{W}_1$ so that they converge to the ideal values W_1 and the following squared residual error is minimized.

$$E_1(\widehat{W}_1) = \frac{1}{2} e_H^T e_H \tag{19}$$

Lemma 1.[21] For any admissible policy $u(t)$, the least-squares solution to (15) exists and is unique for each N. Denote this solution as W_1 and define

$$V_1(x) = W_1^T \phi_1(x) \tag{20}$$

Then, as $N \to \infty$:

$a. sup \ \|e_L\| \le e_{LM}$,

$b. \|W_1 - W\| \to 0$,

$c. sup \ \|V_1 - V\| \to 0$,

$d. sup \ \|\nabla V_1 - \nabla V\| \to 0$.

This result shows that $V_1(x)$ converges uniformly in Sobolev norm to the exact solution $V(x)$ to (5) as $N \to \infty$, and the weights W_1 converge to the first N of the weights W, which exactly solve (5).

Considering lemma 1, define the critic weight estimation error

$$\widetilde{W} = W_1 - \widehat{W}_1 \tag{21}$$

Then, according to (18) and (5), the following equation can be obtained

$$\begin{aligned} e_H &= -\widetilde{W}^T \nabla\phi_1(f + gu) + e_L \\ &= -\widetilde{W}^T \nabla\phi_1(f + gu) - \nabla\varepsilon_1^T(f + gu) \end{aligned} \tag{22}$$

Therefore, from (20)-(22), we can get $\widehat{W}_1 \to W_1$, and $e_H \to e_L$. The tuning for the critic NN is obtained by a gradient descent rule as follows:

$$\dot{\widehat{W}}_1 = -a_1 \frac{\sigma_1}{\sigma^T\sigma + 1}(\sigma^T \widehat{W}_1 + Q + U) \tag{23}$$

where $a_1 > 0$ is the learning rate, $\sigma = \nabla\phi_1(f + gu)$, $\sigma_1 = \sigma/(\sigma^T\sigma + 1)$. From the definition of σ_1 we can get $\sigma_{1m} < \|\sigma_1\| \le \sigma_{1M}$, σ_{1m} and σ_{1M} are positive constants.

Remark 1. From (23), we can see that once the system states have converged to zero, the $\widehat{W}_1$ is no longer updated. This can be viewed as a persistency of excitation (PE) requirement for the inputs to $\widehat{W}_1$ wherein the system, states must be persistently existing long enough for the optimal cost function to be learned.

Then, the critic NN weights error dynamics can be written as

$$\dot{\widetilde{W}}_1 = -\dot{\widehat{W}}_1 = -a_1\sigma_1\sigma_1^T\widetilde{W}_1 + a_1\frac{\sigma_1}{\sigma^T\sigma + 1}e_L \tag{24}$$

Remark 2. The residual error e_L is produced by the NN reconstruction error and is near to zero since the number of hidden layer neurons L is enough.

3.2. Control policy approximation using actor NN

In order to flexibly adjust the critic NN and actor NN to establish convergence to the optimal solution and guarantee Lyapunov-based stability, we set independent weights for critic NN and actor NN. Hence, the optimal control policy can be approximated by an actor NN as follows,

$$u^*(x) = W_2^T \phi_2(x) + \varepsilon_2(x) \tag{25}$$

where $W_2 \in \mathbb{R}^{N_1 \times m}$ is the matrix of ideal unknown constant weights, $\phi_2 : \mathbb{R}^n \to \mathbb{R}^{N_1}$, is called the action NN activation vector, N_1 is the number of neurons in the hidden layer, and $\varepsilon_2(x)$ is the action NN approximation error. As before, with the following assumption satisfied, the NN activation functions must define a complete independent basis set so that u^* can be uniformly approximated on a compact set.

Assumption 2 The actor NN activation functions and actor reconstruction error are bounded over the compact set. In the sense, there exit finite constants ϕ_{2M}, $\varepsilon_{2M} \in \mathbb{R}^+$, such that $\|\phi_2\| \le \phi_{2M}$, $\|\varepsilon_2\| \le \varepsilon_{2M}$.

Since the ideal weights W_2 are unknown, we use estimate weights $\widehat{W}_2 \in \mathbb{R}^{N_1 \times m}$ to approximate the optimal control in (25) by the following equation:

$$\hat{u}(x) = \widehat{W}_2^T \phi_2(x), \forall x \tag{26}$$

Then, the error between the policy estimate (26) and the approximate control based on critic NN's estimate (17) is

$$e_u = \widehat{W}_2^T \phi_2 + \lambda \tanh(1/(2\lambda) R^{-1} g^T(x) \nabla\phi_1^T \widehat{W}_1) \tag{27}$$

In the following, our goal is to tune $\widehat{W}_2$ such that the error is minimized

$$E_2(\widehat{W}_2) = \frac{1}{2} e_u^T e_u \tag{28}$$

The tuning law for the actor NN is also obtained by a gradient descent rule as follows:

$$\begin{aligned} \dot{\widehat{W}}_2 &= -a_2 \frac{\partial E_2}{\partial \widehat{W}_2} \\ &= -a_2 \phi_2 e_u \\ &= -a_2 \phi_2 (\widehat{W}_2{}^T \phi_2 \\ &\quad + \lambda \tanh(1/(2\lambda) R^{-1} g^T(x) \nabla\phi_1^T \widehat{W}_1)) \end{aligned} \tag{29}$$

where $a_2 > 0$ is the learning rate. The actor NN weights error is defined as

$$\widetilde{W}_2 = W_2 - \widehat{W}_2 \tag{30}$$

the error dynamics can be written as

$$\begin{aligned} \dot{\widetilde{W}}_2 = -\dot{\widehat{W}}_2 &= -a_2\phi_2\phi_2^T\widetilde{W}_2 - a_2\phi_2\varepsilon_2 - \lambda a_2 \phi_2 \tanh(D^*) \\ &\quad + \lambda a_2 \phi_2 \tanh(1/(2\lambda) R^{-1} g^T(x) \nabla\phi_1^T \widehat{W}_1) \end{aligned} \tag{31}$$

Remark 3. Note that the fourth term of (31) is a function of $\widehat{W}_1$ but since it appears inside the saturation function tanh(·), this term is always bounded and will be treated appropriately in the stability analysis that follows.

4. Stability analysis

In this subsection, based on the above analysis, theorems are presented to indicate the closed-loop system state and network weights estimation errors are uniformly ultimately bounded. First, the regularity assumption is needed for the stability results presented below.

Assumption 3. The process input function g is uniformly

bounded on a set $\Omega \subset \mathbb{R}^n$, i.e. $\|g(x)\| \le g_M$, $\forall x \subset \mathbb{R}^n$, T is a positive constant.

Then, according to (26) and the system (1), we have

$$\begin{aligned} \dot{x} &= f(x) + g(x)\widehat{W}_2^T\phi_2(x) \\ &= f(x) + g(x)(u^* - \varepsilon_2 - \widetilde{W}_2^T\phi_2(x)) \end{aligned} \quad (32)$$

Now we are ready to prove the following theorem.

Theorem 1. Consider the system given by (1), let the control input be provided by (26). The weight updating laws of the critic NN and the action NN are given by (23) and (29), respectively. And let the initial action NN weights be chosen to generate an initial admissible control. Then the weight estimate errors $\widetilde{W}_1$ and $\widetilde{W}_2$ are UUB with the bounds specifically given by (41)-(43). Moreover, the obtained control input u is close to the optimal control input u^* within a small bound ε_u, i.e., $\|u - u^*\| \le \varepsilon_u$ as $t \to \infty$ for a small positive constant ε_u.

Proof: Choose the following Lyapunov function candidate:

$$L = L_1 + L_2 + L_3 \quad (33)$$

Where, $L_1 = \frac{1}{2}tr\{\widetilde{W}_1^T a_1^{-1}\widetilde{W}_1\}$, $L_2 = \frac{1}{2}tr\{\widetilde{W}_2^T a_2^{-1}\widetilde{W}_2\}$, $L_3 = V^*$.

According to Assumptions 1 and 2 and using (10), (24), and (31), the time derivative of the Lyapunov function candidate (33) along the trajectories of the error system (32) is computed as

$$\dot{L} = \dot{L}_1 + \dot{L}_2 + \dot{L}_3 \quad (34)$$

where

$$\begin{aligned} \dot{L}_1(t) &= \tfrac{1}{a_1}tr\left\{\widetilde{W}_1^T\dot{\widetilde{W}}_1\right\} \\ &= \frac{1}{a_1}tr\left\{\widetilde{W}_1^T\left(-a_1\sigma_1\sigma_1^T\widetilde{W}_1 + a_1\frac{\sigma_1}{\sigma^T\sigma+1}e_L\right)\right\} \\ &\le -(\sigma_{1m}^2 - \tfrac{a_1}{2}\sigma_{1M}^2)\|\widetilde{W}_1\|^2 + \tfrac{1}{2a_1}e_{LM}^2 \end{aligned} \quad (35)$$

$$\begin{aligned} \dot{L}_2(t) &= \frac{1}{a_2}tr\left\{\widetilde{W}_2^T\dot{\widetilde{W}}_2\right\} \\ &= \frac{1}{a_2}tr\{\widetilde{W}_2[-a_2\phi_2\phi_2^T\widetilde{W}_2 - \lambda a_2\phi_2\tanh(D^*) \\ &\quad +\lambda a_2\phi_2\tanh(1/(2\lambda)R^{-1}g^T(x)\nabla\phi_1^T\widehat{W}_1) - a_2\phi_2\varepsilon_2]\} \\ &\le -(\phi_{2M}^2 - 1)\|\widetilde{W}_2\|^2 + \frac{1}{2}\lambda^2\phi_{2M}^2 \\ &\quad +\tfrac{1}{2}(\lambda\phi_{2M} + \varepsilon_{2M}\phi_{2M})^2 \end{aligned} \quad (36)$$

$$\begin{aligned} \dot{L}_3(t) &= \nabla V^{*T}(f + g\hat{u}) \\ &= \nabla V^{*T}f + \nabla V^{*T}g(u^* - \varepsilon_2 - \widetilde{W}_2^T\phi_2) \\ &= -Q(x) - U(u^*) - \nabla V^{*T}(x)g(\varepsilon_2 + \widetilde{W}_2^T\phi_2) \\ &\le -Q(x) - U(u^*) \\ &\quad +g(W_{1M}\phi_{1dM} + \varepsilon_{1dM})(\varepsilon_{2M} + \|\widetilde{W}_2\|\phi_{2M}) \\ &\le -\lambda_{\min}(Q)\|x\|^2 - U(u^*) + \frac{g_M}{2}(\phi_{2M} + 1)(W_{1M}\phi_{1dM} \\ &\quad +\varepsilon_{1dM})^2 + \tfrac{g_M}{2}\varepsilon_{2M}^2 + \tfrac{g_M}{2}\phi_{2M}\|\widetilde{W}_2\|^2 \end{aligned} \quad (37)$$

Then

$$\begin{aligned} \dot{L}(t) \le &-\left(\sigma_{1m}^2 - \frac{a_1}{2}\sigma_{1M}^2\right)\|\widetilde{W}_1\|^2 - \left(\phi_{2M}^2 - \frac{g_M}{2}\phi_{2M} - 1\right)\|\widetilde{W}_2\|^2 \\ &-\lambda_{\min}(Q)\|x\|^2 - U(u^*) + D_M \end{aligned} \quad (38)$$

where $D_M = \frac{1}{2a_1}e_{LM}^2 + \frac{1}{2}\lambda^2\phi_{2M}^2 + \frac{1}{2}(\lambda\phi_{2M} + \varepsilon_{2M}\phi_{2M})^2 + \frac{g_M}{2}(\phi_{2M} + 1)(W_{1M}\phi_{1dM} + \varepsilon_{1dM})^2 + \frac{g_M}{2}\varepsilon_{2M}^2$.

If σ_{1m}, σ_{1M}, and ϕ_{2M} are selected to satisfy

$$a_1 < \frac{2\sigma_{1m}^2}{\sigma_{1M}^2} \quad (39)$$

$$\phi_{2M} > \frac{g_M + \sqrt{g_M^2 + 16}}{4} \quad (40)$$

and given the following inequalities

$$\|x\| > \sqrt{\frac{D_M}{\lambda_{\min}(Q)}} \triangleq l_x \quad (41)$$

$$\|\widetilde{W}_1\| > \sqrt{\frac{D_M}{\sigma_{1m}^2 - \frac{1}{2}\sigma_{1M}^2}} \triangleq l_{\widetilde{W}_1} \quad (42)$$

$$\|\widetilde{W}_2\| > \sqrt{\frac{D_M}{\phi_{2M}^2 - \frac{g_M}{2}\phi_{2M} - 1}} \triangleq l_{\widetilde{W}_2} \quad (43)$$

all hold, then $\dot{L} < 0$.Therefore, using Lyapunov theory[31], it can be concluded that the system state x and the NN weight estimation errors are UUB.

Next we will prove $\|\hat{u} - u^*\| \le \varepsilon_u$ as $t \to \infty$. Recalling the expression of u^*, we have

$$\hat{u} - u^* = \widetilde{W}_2^T\phi_2 + \varepsilon_2 \quad (44)$$

When $t \to \infty$, the upper bound of (44) is

$$\|\hat{u} - u^*\| \le \varepsilon_u \quad (45)$$

Where $\varepsilon_u = l_{\widetilde{W}_2}\phi_{2M} + \varepsilon_{2M}$. This completes the proof.

To remove the effect of the NN approximation errors ε_1, ε_2 (and their partial derivatives) and obtain a closed-loop system with an asymptotically stable equilibrium point, one needs to add a robustifying term to the control law (26) as

$$u_{ad} = \hat{u} + \zeta = \widehat{W}_2^T\phi_2 + \zeta, \forall x \quad (46)$$

where

$$\zeta = -K_{ad}\|x\|^2\tfrac{1_m}{h + x^Tx}, \forall x \quad (47)$$

with h a positive constant, K_{ad} satisfies

$$K_{ad}\|x\|^2 \ge \tfrac{D_M(h + x^Tx)}{g_M(W_{1M}\phi_{1dM} + \varepsilon_{1dM})}, \forall x \quad (48)$$

The following theorem is the main result of the paper and proves asymptotic stability of the learning scheme of resulting closed-loop dynamics:

$$\dot{x} = f(x) + g(x)((W_2 - \widetilde{W}_2)\phi_2(x) + \zeta) \quad (49)$$

Theorem 2: Consider the system given by (1), let the control input be provided by (46). The weight updating laws of the critic NN and the action NN are given by (23) and (29), respectively. And let the initial action NN weights be chosen to generate an initial admissible control. Then the system state x and the weight estimate errors $\widetilde{W}_1$ and $\widetilde{W}_2$ will asymptotically converge to zero. Moreover, the obtained control input u is close to the optimal control input u^* within a small bound δ_u, i.e., $\|u - u^*\| \le \delta_u$ as $t \to \infty$ for a small positive constant δ_u.

Proof: Choose the same Lyapunov function candidate as in Theorem 1. Differentiating the Lyapunov function candidate in (33) along the trajectories of the system in (49), similar to the proof of Theorem 1, by using (47) and (48), we can obtain.

$$\begin{aligned} \dot{L}(t) \le &-(\sigma_{1m}^2 - \tfrac{a_1}{2}\sigma_{1M}^2)\|\widetilde{W}_1\|^2 \\ &-(\phi_{2M}^2 - \frac{g_M}{2}\phi_{2M} - 1)\|\widetilde{W}_2\|^2 \\ &-\lambda_{min}(Q)\|x\|^2 - U(u^*) \end{aligned} \quad (50)$$

Choosing a_1, and ϕ_{2M} as in Theorem 1, we have $\dot{L}(t) \leq 0, t \geq 0$. Using Barbalat's lemma [17], we have $\|x\| \to 0$ as $t \to \infty$. Similarly, we can prove that $\|\widetilde{W}_1\| \to 0$ and $\|\widetilde{W}_2\| \to 0$ as $t \to \infty$. Next we will prove $\|u - u^*\| \leq \delta_u$ as $t \to \infty$. From (46), we have

$$u_{ad} - u^* = \widetilde{W}_2^T \phi_2(x) + \varepsilon_2 + \zeta, \forall x \tag{51}$$

Since$\|x\| \to 0$ as $t \to \infty$, the robustifying control input $\|\zeta\| \to 0$ as $t \to \infty$, then the upper bound of (51) is

$$\|u_{ad} - u^*\| \leq \delta_u \tag{52}$$

Where $\delta_u = \varepsilon_{2M}$. This completes the proof.

Remark 4. For the inequality (39) to hold, one needs to pick the appropriate activation function for the critic NN. Regarding (40), since ϕ_{2M} is simply an upper bound that appears in Assumptions 2, one can have it as large as needed. However, one must keep in mind that a large value for ϕ_{2M}, requires an appropriate value for the function K_{ad} in the robustness term in (48).

Remark 5. From (45) and (52), it can be seen that δ_u is smaller than ε_u ,which reveals the role of the robustifying term in making the obtained control input closer to the optimal control input.

5. Simulation

The well-known Lorenz system is given by

$$\begin{aligned} \dot{x}_1 &= \sigma(-x_1 + x_2) \\ \dot{x}_2 &= \gamma x_1 - x_2 - x_1 x_3 + u \\ \dot{x}_3 &= x_1 x_2 - b x_3 \end{aligned} \tag{53}$$

where the parameters $\sigma = 10$, $\gamma = 28$, and $b = 8/3$. The phase plane trajectory of the Lorenz system is shown in Fig. 1.

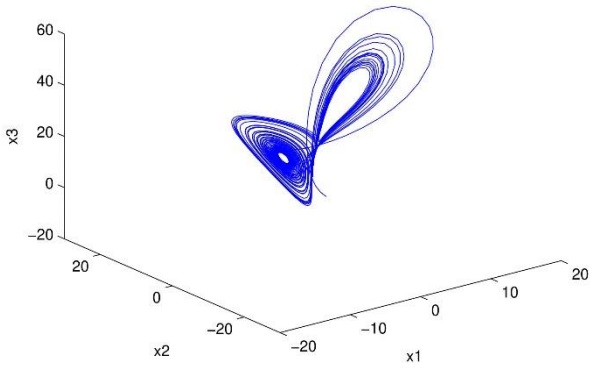

Fig. 1. The phase plane trajectory of Lorenz system

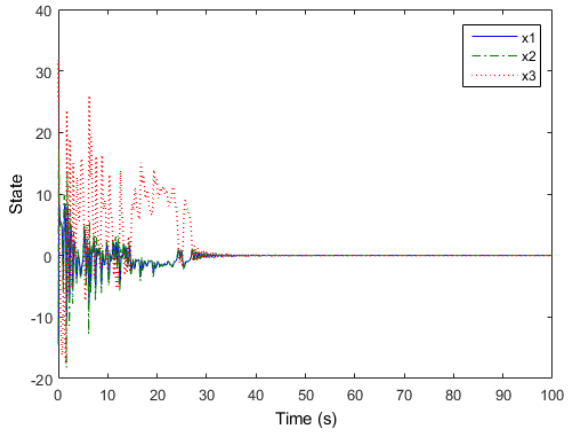

Fig. 2. The convergence of states.

We now consider the optimal control of Lorenz system (53) with the input saturated $\|u\| \leq 10$ and the cost defined by (2), (3), and (4) with Q and R in the utility function are identity matrices of appropriate dimensions, for which the optimal feedback law is not known in closed form.

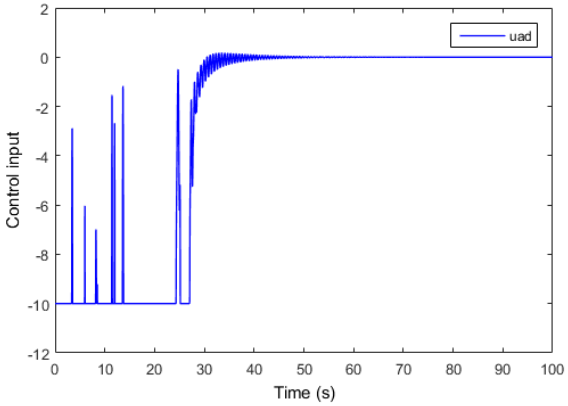

Fig. 3. The convergence of control policy.

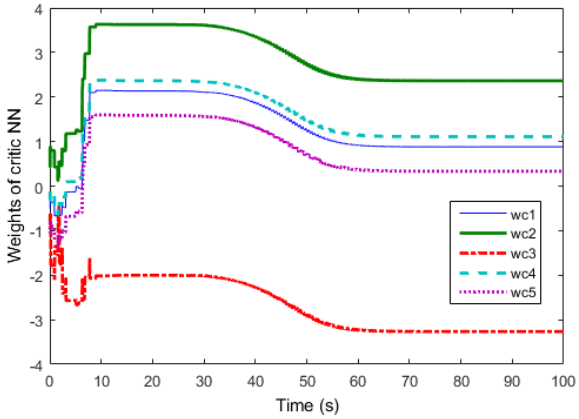

Fig. 4. The convergence of critic NN weights.

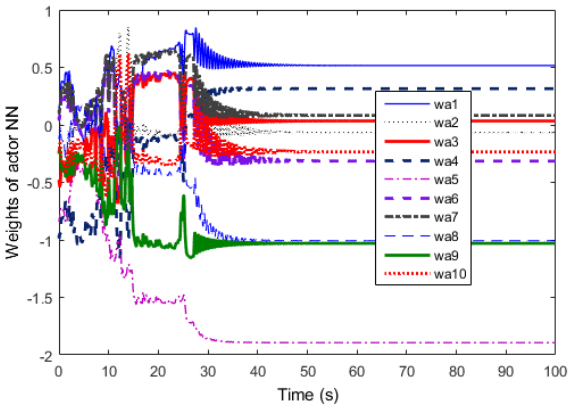

Fig. 5. The convergence of actor NN weights.

To find the optimal solution using the proposed method, the structures of action network and critic network are chosen 3-10-1 and 3-5-1, respectively. As no verifiable method exists to ensure PE in nonlinear systems, a small exploratory signal consisting of sinusoids of varying frequencies, i.e. $100exp(-0.1t)((sin(t)^2 cos(t) + sin(2t)^2 cos(0.1t)) + 1.2sin(-1.2t)^2 cos(0.5t) + cos(2.4t)sin(2.4t)^3)$, is added to the control input before the first 25s to excite the system states and

ensure the PE holds. The initial weights of the two networks are selected from (−1.5, 0.5). Adaptive parameters of the critic and action networks are $a_1 = a_2 = 0.5$. The parameters of robustifying term are selected as $K_{ad} = 165,\ h = 70$. The initial state of the chaotic system (54) is [-14.5; -14.5; 25.8]. The activation functions in the critic network and the action network are hyperbolic tangent functions. After 100 time steps, we obtain that $\widehat{W}_c$ converges to [0.8803, 2.3644, − 3.2656, 1.1088, 0.3327], and $\widehat{W}_a$ converges to [0.5160, − 0.0660, 0.0344, 0.3152, − 1.8937, − 0.3147, 0.0822, − 1.0074, − 1.0300, − 0.2350].

Fig. 2 presents the state trajectory of the system (53), from which we can see that the closed-loop system state converges to zero as the time step increases. Fig. 3 shows the optimal control input, which is saturated when it reaches the maximum and minimum saturation limits. Fig. 4 and 5 show the convergence of the critic NN weights and actor NN weights. Thus, the proposed optimal control method for chaotic systems in this paper is very effective.

6. Conclusions

This paper has developed an asymptotically-stable optimal control scheme based on an online ADP algorithm for continuous-time chaotic systems with input saturation. The ADP algorithm is used to obtain the approximate optimal control input which minimizes the values of the specified performance index. The critic network and action network are used to approximate performance index function and control input, respectively. A robustifying term is designed to eliminate the effect caused by the neural network approximation errors, leading to asymptotical stability of the closed-loop chaotic system. The effectiveness of the proposed approach has been demonstrated by a simulation of a Lorenz system.

Acknowledgements

Thank you for your great help and support in the completion of this paper, Wang Fule and Liu Jinyu would like to acknowledge by Qu Qiuxia and Sun Liangliang.

This work was supported in part by the National Natural Science Foundation of China under Grants 61873174 and 61703288, in part by Liaoning Natural Science Foundation Guidance Program under Grant 2019-ZD-0679, in part by Liaoning Natural Science Foundation under Grant 2019-BS-195, and in part by the Open project of State Key Laboratory of Robotics under Grant 2019-O20

References

[1] Gao SG., Dong HR., Sun XB., Ning B., Neural adaptive chaotic control with constrained input using state and output feedback[J]. Chinese Physics B ,2015.24(1): p. 170-176.

[2] Fuh CC., Optimal control of chaotic systems with input saturation using an input-state linearization scheme[J]. Commun Nonlinear Sci Numer Simulat. 2009.14(8): p. 3424-3431.

[3] Wang Y., Wong KW., Liao XF., Chen GR., A new chaos based fast image encryption algorithm[J]. Applied Soft Computing, 2011. 11(1): p. 514-522.

[4] Liu GG., Zhao Y., Adaptive control on a class of uncertain chaotic systems[J]. Chinese Physics Letters, 2005. 22(5): p. 1069-1071.

[5] Wang J., Gao JF., Ma XX., Liang ZH., A general response system control method based on backstepping design for synchronization of continuous scalar chaotic signal[J]. Chinese Physics Letters, 2006. 2(8): p. 2027-2029.

[6] Wang XY., He YJ., Projective synchronization of fractional order chaotic system based on linear separation[J]. Phys Lett A, 2008. 372(4): p. 435-441.

[7] Wang XY., Nian FZ., Guo G., High precision fast projective synchronization in chaotic (hyperchaotic) systems[J]. Phys Lett A, 2009. 373(20): p. 1754-1761.

[8] Sinha SC., Henrichs JT., Ravindra B., A general approach in the design of active controllers for nonlinear systems exhibiting chaos[J]. International Journal of Bifurcation Chaos, 2000. 10(1): p. 165-178.

[9] Zhang H., Wang ZL., Liu DR., Chaotifying fuzzy hyperbolic model using adaptive inverse optimal control approach[J]. International Journal of Bifurcation Chaos, 2004. 14(10): p. 3505-3517.

[10] Yassen MT., Chaos control of chaotic dynamical systems using backstepping design[J]. Chaos, Solitons Fractals, 2006. 27(2): p. 537-548.

[11] Chen GR., Dong XN., On feedback control of chaotic continuous-time systems[J]. IEEE Transactions on Circults Systems, 1993. 40(9): p. 591-601.

[12] Hua CC., Guan XP., Adaptive control for chaotic systems[J]. Chaos, Solitons Fractals, 2004. p. 22(1): 55-60.

[13] Liu S., Chen LQ., Chaos synchronization of a chain network based on a sliding mode control[J]. Chinese Physics B, 2013. 22(10): p. 176-181.

[14] Werbos PJ., Approximate dynamic programming for real time control and neural modeling[M]. Multiscience Press, 1992.

[15] Murray JJ., Cox CJ., Lendaris GG, Saeks R. Adaptive dynamic programming[J]. IEEE Transactions Systems Man Cybernetics Part C, 2002. 32(2): p. 140-153.

[16] Song RZ., Xiao WD., Wei QL., A new approach of optimal control for a class of continuous-time chaotic systems by an online ADP algorithm[J]. Chin Phys B, 2014. 23(5): p. 401-407.

[17] Khalil H. Nonlinear systems[M]. NJ: Prentice-Hall. 2002.

[18] Luo, X., et al., Time Delay Estimation-based Adaptive Sliding-Mode Control for Nonholonomic Mobile Robots[J]. International Journal of Applied Mathematics in Control Engineering, 2018. 1(1): p. 1-8.

[19] Wu, WJ., Parametric Lyapunov Equation Approach to Stabilization and Set Invariance Conditions of Linear Systems with Input Saturation[J]. International Journal of Applied Mathematics in Control Engineering, 2018. 1(1): p.79-84.

[20] Vamvoudakis KG., Lewis FL., Online actor-critic algorithm to solve the continuous infinite-time horizon optimal control problem[J]. Automatica, 2010. 46(5): p. 878-888.

[21] Dierks T., Jaganathan S., Online optimal control of nonlinear discrete-time systems using approximate dynamic programming[J]. Control Theory Applied, 2011. 9(3): p. 361-369.

[22] Zhang HG., Wei QL., Luo YH., A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear system based on greedy HDP iteration algorithm[J]. IEEE Transactions on Systems Man & Cybernetics Part B, 2008. 38(4): p. 937-942.

[23] Mir A. S., Bhasin S., Senroyn N., Decentralized nonlinear adaptive optimal control scheme for enhancement of power system stability[J]. IEEE Transaction on Power Systems, 2020. 35(2): p. 1400-1410.

[24] He S., Fang H., Zhang M., Liu F., Ding Z., Adaptive optimal control for a class of nonlinear systems: the online policy iteration approach[J]. IEEE Transaction on neural network and learning Systems, 2020. 31(2): p. 549-558.

[25] Wei Q., Liu D., Lin H., Value iteration adaptive dynamic programming for optimal control of discrete-time nonlinear systems[J]. Transaction on neural network and learning Systems, 2016. 46(3): p. 840-853.

[26] Shi Z., Wang Z., Optimal control for a class of complex singular system based on adaptive dynamic programming[J]. IEEE/CAA Journal of Automatica Sinica, 2019. 6(1): p. 191-200.

[27] Luo B., Yang Y., Liu D., Wu HN., Event-triggered optimal control with performance guarantees using adaptive dynamic programming[J]. IEEE Transactions on neural networks and learning systems, 2020. 31(1): p. 76-88.

[28] Vamvoudakis K. G., Ferraz H., Model-free event-triggered control algorithm for continuous-time linear systems with optimal performance[J]. Automatica, 2018. 87(1): p. 412-420.

[29] Chen C., Modares H., Xie K., Lewis F. L., Wan Y., Xie S., Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics[J]. IEEE transactions Automatic Control, 2019. 64(1): p. 4423-4438.

[30] Wei Q., Liu D., Lin Q., Discrete-time local value iteration adaptive dynamic programming: admissibility and termination analysis[J]. IEEE Transaction on neural network and learning Systems, 2017. 28(11): p.2490-2502.

[31] Abu-Khalaf M., Lewis FL., Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach[J]. Automatica, 2005. 41(5): p. 779-791.
[32] Modares H., Lewis F., Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforce-ment learning[J]. Automatica, 2014. 50(7): p. 1780-1792.
[33] Modares H., Sistani M. N., Lewis F., A policy iteration approach to online optimal control of continuous-time constrained-input systems[J]. ISA transactions, 2013. 52(5): p. 611–621.
[34] Gao, X., et al., Adaptive High Order Neural Network Identification and Control for Hysteresis Nonlinear System with Bouc-Wen Model[J]. International Journal of Applied Mathematics in Control Engineering, 2019. 2(1): p. 17-24.

Fule Wang obtained his bachelor's degree in building electrical and intelligent from Shenyang Institute of urban construction in 2015, and began to study for a master's degree in control science and engineering from Shenyang Jianzhu University in 2019. His current research interests include adaptive dynamic programming, reinforcement learning, nonlinear systems and neural networks.

Qiuxia Qu received the B.S. degree in Electrical Engineering and Automation from Qufu normal university, China, in 2008, and the M.S. and Ph.D. degrees both in Control Theory and Control Engineering from Northeastern University, China, in 2010 and 2018, respectively. She is currently a Lecturer with Shenyang Jianzhu University. Her current research interests include adaptive dynamic programming, reinforcement learning, nonlinear systems, and neural network.

Jinyu Liu obtained a bachelor's degree in building electrical and intelligent from Shenyang Jianzhu University in 2015, and has been studying for a master's degree in control engineering from Shenyang Jianzhu University since 2019. Her current research interests include adaptive dynamic programming, reinforcement learning, nonlinear systems and neural networks

Liangliang Sun received his Ph.D degree in Control Theory and Control Engineering from Northeastern University, China, in 2015. He is currently an Associate Professor of Information and Control Engineering at Shenyang Jianzhu University. His current research interests include Intelligent Manufacturing Systems, planning, scheduling, and coordination of design, manufacturing, and service activities intelligence field.

Juan Wang received the B.S. degree in Mathematics and Applied Mathematics from Bohai University, China, in 2007, and M.S. degree in Operational Research and Cybernetics from Northeastern University, China, in 2009. She completed her Ph.D. in Control Theory and Applications in 2016 at Northeastern University, China. She is an associate professor for Faculty of Information and Control Engineering of Shenyang Jianzhu University, China. Her research interests include switched systems, positive system and actuator saturation.

Qiong Xia received the B.S. and M.S. degrees in Automation and System Engineering from Harbin Institute of Technology and Northeastern University, China, in 2005 and 2010, respectively. She completed her Ph.D. degree in System Engineering in 2018 at Northeastern University, China. She is currently a Lecturer with Shenyang Jianzhu University. Her research interests include ferrous metallurgy, modeling and optimization for complex process of industry.