Nonlinear Dyn (2018) 93:2089–2103
https://doi.org/10.1007/s11071-018-4309-8

ORIGINAL PAPER

# Adaptive dynamic programming-based stabilization of nonlinear systems with unknown actuator saturation

**Bo Zhao** · **Lihao Jia** · **Hongbing Xia** · **Yuanchun Li**

Received: 8 April 2017 / Accepted: 24 April 2018 / Published online: 7 May 2018


**Abstract** This paper addresses the stabilizing control problem for nonlinear systems subject to unknown actuator saturation by using adaptive dynamic programming algorithm. The control strategy is composed of an online nominal optimal control and a neural network (NN)-based feed-forward saturation compensator. For nominal systems without actuator saturation, a critic NN is established to deal with the Hamilton–Jacobi–Bellman equation. Thus, the online approximate nominal optimal control policy can be obtained without action NN. Then, the unknown actuator saturation, which is considered as saturation nonlinearity by simple transformation, is compensated by employing a NN-based feed-forward control loop. The stability of the closed-loop nonlinear system is analyzed to be ultimately uniformly bounded via Lyapunov's direct method. Finally, the effectiveness of the presented control method is demonstrated by two simulation examples.

B. Zhao (✉)
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
e-mail: zhaobo@ia.ac.cn

L. Jia
Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

H. Xia · Y. Li
Department of Control Science and Engineering, Changchun University of Technology, Changchun 130012, China

**Keywords** Adaptive dynamic programming · Unknown actuator saturation · Continuous-time nonlinear systems · Stabilizing control · Neural networks

## 1 Introduction

Along with the increasing complexity of modern industries, the optimal control problem has been paid considerable attention to nonlinear systems for many decades in the control community. To achieve this objective, a specified cost function or control policy is required to be minimized by solving Hamilton–Jacobi–Bellman (HJB) equation, which is difficult to be handled by analytical approach since it is actually a nonlinear partial differential equation. Although dynamic programming (DP) [1] gives a great strategy to handle the above issue, the increasing dimension of nonlinear systems increases the computation burden, which is the so-called curse of dimensionality.

Fortunately, adaptive dynamic programming (ADP) algorithm, which can avoid the above difficulty, is developed [2,3] with the aid of NNs [4–6] to tackle optimal control problems for nonlinear systems forward in time. Some synonyms are utilized for ADP, such as adaptive DP [7], approximate DP [8], adaptive critic designs [9], neuro-DP [10], neural DP [11], and reinforcement learning (RL) [12,13]. Werbos [2] categorized ADP approaches into heuristic dynamic programming (HDP), dual HDP (DHP), action-dependent HDP

(ADHDP), and action-dependent DHP (ADDHP). Afterward, globalized DHP (GDHP) and ADGDHP were reported [9]. In the past few years, ADP algorithms were developed further to deal with control problems for continuous-time systems [14,15], discrete-time systems [16,17], systems with uncertainties [18] and external disturbances [19], desired trajectory tracking [20,21], time-delay [22], fault-tolerant [23,24], zero-sum games [25], event-triggered systems [26], etc.

Specially, actuator saturation often emerges in many practical systems, such as spacecrafts [27], launch vehicles [28], robot manipulators [29,30], helicopters [31], teleoperation systems [32], suspension systems [33], interception systems [34]. Its presence may cause control performance reduction or even the closed-loop system unstable. Great efforts have been made to nonlinear systems in this situation [35–37]. It is worth pointing out that some ADP-based methods have been presented in recent years. He et al. [38] constructed a certain strategic utility function, which was approximated by a critic NN. And then, a RL-based output feedback controller was designed to transmit an expected tracking performance for systems with magnitude constraints. Abu-Khalaf et al. [39] presented a two-player policy iteration-based $L_2$-gain optimal feedback strategy for nonlinear systems in the presence of saturation constraints. Heydari et al. [40] developed a finite-horizon single network adaptive critic-based fixed-final-time control-constrained optimal controllers. Zhang et al. [41] presented an iterative two-stage DHP method for switched nonlinear systems subject to actuators saturation. Dong et al. [42] developed an actor-critic framework-based near-optimal control scheme with event-triggered strategy to reduce the computational and transmission cost. Modares et al. [43] proposed an actor-critic-based online learning policy iteration algorithm to tackle the optimal control problem for unknown nonlinear systems with input constraints. By constructing a nonquadratic cost function, Xu et al. [44] addressed the near-optimal regulation problem via NNs to handle the time-varying HJB equation for uncertain and quantized nonlinear discrete-time systems with input constraints. Song et al. [45] developed a HDP method for nonlinear discrete time-delay systems subject to actuator saturation. Zhang et al. [46] introduced a nonquadratic cost function to avoid the affection of control constraints, and three NNs are employed to facilitate the implementation of the iterative algorithm. Liu et al. [47] presented a robust adaptive control algorithm based on RL for uncertain nonlinear systems subject to input constraints. Meanwhile, they developed a triple-NN approximation-based GDHP framework for unknown discrete-time nonlinear systems [48]. Yang et al. respectively, developed an online identifier-critic architecture [49] and system data-based integral RL algorithm [50] for unknown nonlinear systems by updating the value function and control policy simultaneously. For practical systems, Pomprapa et al. [51] proposed a model-free policy iteration algorithm-based state feedback configuration for controlling arterial oxygen saturation of an interconnected three-tank systems.

From the above literature, we can conclude that most of existing results were concerned with nonlinear systems subject to actuator saturation with available limit bounds, which are always necessary for designing cost functions in ADP-based control methods directly or indirectly. However, the outputs of actuators may be biased or suddenly abrupt in many practical systems. It implies that the saturation bounds of the actuators are uncertain or unknown, which makes existing methods incapable of action. Thus, the main challenge lies in that we have to design a novel ADP-based optimal control scheme for nonlinear systems with unknown actuator saturation since the necessary saturation bounds in existing methods cannot be provided before designing the optimal control. This motivates our research.

This paper focuses on an ADP-based stabilizing scheme for nonlinear systems subject to unknown actuator saturation. The developed control method consists of the online nominal optimal control for nominal system and a NN-based feed-forward compensation for the unknown actuator saturation. The convergence of the closed-loop system is guaranteed via Lyapunov stability theorem. Simulation examples verify the effectiveness of the proposed stabilizing method.

The main contributions and highlights of this work in contrast to the existing literature are summarized as follows:

1. Unlike many existing works [22,24,52], this paper proposes an online learning optimal nominal control method, which removes the necessary requirements of the initial stabilizing control and the persisting excitation condition.
2. Different from the related methods in [46–50] that were concerned with nonlinear systems with avail-

able actuator saturation, this paper presents the ADP-based control methods for nonlinear systems subject to unknown actuator saturation. Thus, the developed control method avoids any priori knowledge of actuator saturation.

3. The optimal control is derived depending only on critic NN, rather than dual- or triple-NN-based architecture. Thus, it reduces the computational burden of traditional adaptive critic designs [43, 49].

The structure of this paper is organized as follows: In Sect. 2, the problem statement is provided. In Sect. 3, the ADP-based online nominal optimal control is developed for nominal nonlinear systems. Then, a NN-based saturation compensator is developed for eliminating the negative affection of unknown actuator saturation. In the following, the stability analysis is presented. In Sect. 4, two numerical examples are employed to verify the effectiveness of the proposed method. Finally, the conclusion is drawn in Sect. 5.

## 2 Problem statement

The considered nominal continuous-time nonlinear systems can be described as

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

where $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$ are the system state and control input vectors, respectively. $f(\cdot)$ and $g(\cdot)$ are assumed to be locally Lipschitz and differentiable in their arguments such that the solution $x(t)$ to nonlinear system (1) is unique for any given initial state $x(0) = x_0$ with $f(0) = 0$. Nonlinear system (1) is stable in the sense that there exists a continuous control $u$ which stabilizes the system asymptotically.

In order to better adapt practical control requirements, we are concerned with the stabilization problems for continuous-time nonlinear systems subject to unknown actuator saturation as

$$\dot{x} = f(x) + g(x)\tau, \tag{2}$$

where $\tau = [\tau_1, \tau_2, \ldots, \tau_m]^{\mathsf{T}} \in \mathbb{R}^m$ is the saturated actuator output vector, which is the actual applied control input of (2). It slopes between its lower and upper limits, i.e.,

$$\tau_i = \mathrm{sat}(u_i) = \begin{cases} u_{i\,\max}, & u_i > u_{i\,\max}, \\ u_i, & u_{i\,\min} \le u_i \le u_{i\,\max}, \\ u_{i\,\min}, & u_i < u_{i\,\min}, \end{cases} \tag{3}$$

where $i = 1, 2, \ldots, m$, and $u_{i\,\max}$ and $u_{i\,\min}$ are the unknown upper and lower limit bounds, respectively. That is to say, actuator saturation occurs if the commanded input $u_i$ falls outside of the set $[u_{i\,\min}, u_{i\,\max}]$, and the control input cannot be implemented to the device totally.

The main purpose of this paper is to propose a NN compensation-based ADP stabilization scheme for nonlinear systems subject to unknown actuator saturation and ensure all the signals of the closed-loop nonlinear system (2) to be ultimately uniformly bounded (UUB).

## 3 Online approximate optimal controller design and stability analysis

This section is divided into three parts. The online learning nominal optimal control scheme is presented in the first part for nominal system (1). Then, in the second part, a feed-forward NN compensator is developed to tackle the unknown actuator saturation for nonlinear system (2). In the third part, the UUB stability of the closed-loop nonlinear system is analyzed.

### 3.1 Online nominal optimal control

For nominal nonlinear system (1), a feedback control $u_n(x) \in \Psi(\Omega)$ will be derived to tackle its control problem such that the closed-loop nonlinear system is stable. The objective of this optimal control problem is to find the stabilizing nominal control $u_n(x)$ to minimize the infinite-horizon cost function which is given by

$$V(x_0) = \int_0^\infty U(x(s), u_n(s))\mathrm{d}s, \tag{4}$$

where $U(x, u_n) = x^{\mathsf{T}} Q x + u_n^{\mathsf{T}} R u_n$ is the utility function, $U(x, u_n) \ge 0$ for all $x$ and $u_n$ with $U(0, 0) = 0$, and $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are positive definite matrices. If the associated infinite-horizon cost function (4) is continuously differentiable, the infinitesimal

version of (4) is the so-called nonlinear Lyapunov equation

$$0 = U(x, u_n) + \nabla V^{\mathsf{T}}(x)\dot{x}.$$

In light of the nominal control policy $u_n(x)$ and the cost function $V(x)$, define the Hamiltonian as

$$H(x, u_n, \nabla V(x)) = U(x, u_n) + (\nabla V(x))^{\mathsf{T}} (f(x) + g(x)u_n), \tag{5}$$

and the optimal cost function as

$$V^*(x) = \min_{u_n \in \psi(\Omega)} \int_t^{\infty} U(x(s), u_n(s))\mathrm{d}s. \tag{6}$$

According to [1], the optimal cost function $V^*(x)$ of (6) can be derived from the solution of the HJB equation

$$0 = \min_{u_n(x) \in \Psi(\Omega)} H\left(x, u_n, \nabla V^*(x)\right) \tag{7}$$

with $V^*(0) = 0$, and the item $\nabla V^*(x)$ indicates the partial gradient of the cost function $V^*(x)$ in (6) with respect to $x$, i.e., $\nabla V^*(x) = \dfrac{\partial V^*(x)}{\partial x}$.

If the solution $V^*(x)$ of (7) exists, the closed-loop description for optimal control can be obtained as

$$u_n^*(x) = -\frac{1}{2} R^{-1} g^{\mathsf{T}}(x) \nabla V^*(x). \tag{8}$$

By simple transformation of (8), we get

$$\left(\nabla V^*(x)\right)^{\mathsf{T}} g(x) = -2\left(u_n^*(x)\right)^{\mathsf{T}} R. \tag{9}$$

As we know, NNs have strong capability of approximating any nonlinear functions. Since the differentiable cost function on the compact set $\Omega$ is usually highly nonlinear and nonanalytic, in the following, we approximate it with a critic NN with a single hidden layer as

$$V(x) = W_c^{\mathsf{T}} \sigma_c(x) + \varepsilon_c(x), \tag{10}$$

where $W_c \in \mathbb{R}^{l_1}$ and $\sigma_c(x) \in \mathrm{R}^{l_1}$ are, respectively, the ideal weight vector and the activation function of the critic NN, $l_1$ is the number of neurons in the hidden layer, and $\varepsilon_c(x)$ is the NN approximation error. Then, the partial gradient of $V(x)$ with respect to $x$ is

$$\nabla V(x) = (\nabla \sigma_c(x))^{\mathsf{T}} W_c + \nabla \varepsilon_c(x), \tag{11}$$

where $\nabla \sigma_c(x) = \dfrac{\partial \sigma_c(x)}{\partial x} \in \mathbb{R}^{l_1 \times n}$ and $\nabla \varepsilon_c(x)$ are the partial gradients of the activation function and the approximation error, respectively.

Thus, the Hamiltonian can be described as

$$H(x, u_n, W_c) = U(x, u_n) + \left(W_c^{\mathsf{T}} \nabla \sigma_c(x) + \nabla \varepsilon_c(x)\right)\dot{x}. \tag{12}$$

Combining (7) with (12), we have

$$U(x, u_n) + W_c^{\mathsf{T}} \nabla \sigma_c(x)\dot{x} = e_{cH}, \tag{13}$$

where $e_{cH} = -\nabla \varepsilon_c(x)\dot{x}$ indicates the NN approximation-caused residual error.

In virtue of the ideal weight vector $W_c$ is unavailable, the approximate critic NN can be expressed by

$$\hat{V}(x) = \hat{W}_c^{\mathsf{T}} \sigma_c(x), \tag{14}$$

where $\hat{W}_c$ is the estimation of the ideal weight vector $W_c$. Then, we have the partial gradient of $\hat{V}(x)$ with respect to $x$ as

$$\nabla \hat{V}(x) = (\nabla \sigma_c(x))^{\mathsf{T}} \hat{W}_c. \tag{15}$$

Thus, the Hamiltonian can be approximated as

$$H\left(x, u_n, \hat{W}_c\right) = U(x, u_n) + \hat{W}_c^{\mathsf{T}} \nabla \sigma_c(x)\dot{x} = e_c. \tag{16}$$

Define $\tilde{W}_c = W_c - \hat{W}_c$ as the weight approximation error. Comparing (12) with (16), we have

$$e_c = e_{cH} - \tilde{W}_c^{\mathsf{T}} \nabla \sigma_c(x)\dot{x}. \tag{17}$$

For adjusting the critic NN weight vector $\hat{W}_c$, the steepest descent algorithm is used to minimize the objective function $E_c = \dfrac{1}{2} e_c^{\mathsf{T}} e_c$. Thus, the weight vector approximation error can be updated adaptively by

$$\dot{\tilde{W}}_c = -\dot{\hat{W}}_c = l_c\left(e_{cH} - \tilde{W}_c^{\mathsf{T}}\theta\right)\theta, \tag{18}$$

where $\theta = \nabla\sigma_c(x)\dot{x}$.

Thus, $\hat{W}_c$ can be updated by

$$\dot{\hat{W}}_c = -l_c e_c \theta, \tag{19}$$

where $l_c > 0$ is the critic NN learning rate.

Therefore, the ideal nominal control policy can be expressed according to (8) and (10) as

$$u_n(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\left(\nabla\sigma_c^{\mathsf{T}}(x)W_c + \nabla\varepsilon_c(x)\right). \tag{20}$$

Thus, it can be approximated as

$$\hat{u}_n(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla\sigma_c^{\mathsf{T}}(x)\hat{W}_c. \tag{21}$$

From (20), we can observe that the nominal control policy can be obtained by the critic NN only and the training of the action NN is not required any more.

**Theorem 1** *For nonlinear system (1), the weight vector approximation error is guaranteed to be UUB if the critic NN weight vector is updated by (19).*

*Proof* Choose the Lyapunov function candidate as

$$L_1 = \frac{1}{2l_c}\tilde{W}_c^{\mathsf{T}}\tilde{W}_c. \tag{22}$$

The time derivative of (22) is

$$\begin{aligned}\dot{L}_1 &= \frac{1}{l_c}\tilde{W}_c^{\mathsf{T}}\dot{\tilde{W}}_c \\ &= \tilde{W}_c^{\mathsf{T}}\left(e_{cH} - \tilde{W}_c^{\mathsf{T}}\theta\right)\theta \\ &= \tilde{W}_c^{\mathsf{T}}e_{cH}\theta - \left\|\tilde{W}_c^{\mathsf{T}}\theta\right\|^2 \\ &\le \frac{1}{2}e_{cH}^2 - \frac{1}{2}\left\|\tilde{W}_c^{\mathsf{T}}\theta\right\|^2.\end{aligned} \tag{23}$$

Assume $\|\theta\| \le \theta_M$, where $\theta_M > 0$. Hence, $\dot{L}_1 < 0$ as long as $\tilde{W}_c$ lies outside of the following compact set

$$\Omega_{\tilde{W}_c} = \left\{\tilde{W}_c : \left\|\tilde{W}_c\right\| \le \left\|\frac{e_{cH}}{\theta_M}\right\|\right\}.$$

Therefore, according to Lyapunov's direct method, the approximation error of the weight vector is UUB. This completes the proof. □

### 3.2 Neural network-based unknown saturation compensation

In this subsection, a NN-based compensator is designed in detail as a feed-forward control loop, which is used for compensating the unknown nonlinear saturation.

In order to tackle the unknown actuator saturation, the vector $\delta(x) = u - \tau = [\delta_1, \delta_2, \ldots, \delta_m]^{\mathsf{T}} \in \mathbb{R}^m$, which is the so-called saturation nonlinearity, is introduced with the definition as

$$\begin{aligned}\delta_i(x) &= u_i - \tau_i \\ &= \begin{cases} u_i - u_{i\,\max}, & u_i > u_{i\,\max}, \\ 0, & u_{i\,\min} \le u_i \le u_{i\,\max}, \\ u_i - u_{i\,\min}, & u_i < u_{i\,\min}, \end{cases}\end{aligned} \tag{24}$$

where $i = 1, 2, \ldots, m$.

Noticing that in the case of no actuator saturation, $\delta(x)$ remains zero, and the control law becomes the same as the ideal nominal control law (20). However, $\delta(x)$ is nonzero in the presence of actuator saturation. Thus, the saturated nonlinear system (2) can be transformed into

$$\dot{x} = f(x) + g(x)(u - \delta). \tag{25}$$

Here, a backpropagation NN is introduced to approximate the unknown item $\delta(x)$ and it can be presented as

$$\delta(x) = W_\delta^{\mathsf{T}}\sigma_\delta(x) + \varepsilon_\delta(x), \tag{26}$$

where $W_\delta \in \mathbb{R}^{l_2}$ and $\sigma_\delta(x) \in \mathbb{R}^{l_2}$ are, respectively, the ideal weight vector and the activation function, $l_2$ is the number of neurons in the hidden layer, and $\varepsilon_\delta(x)$ is the NN approximation error.

To determine the unknown weight vector $W_\delta$, (26) is approximated by

$$\hat{\delta}(x) = \hat{W}_\delta^{\mathsf{T}}\sigma_\delta(x), \tag{27}$$

where $\hat{W}_\delta$ is the ideal weight vector estimation. It can be updated by

$$\begin{aligned}\dot{\hat{W}}_\delta &= -\dot{\tilde{W}}_\delta \\ &= \Gamma_\delta\sigma_\delta(x)\left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right) + k\Gamma_\delta\|x\|\hat{W}_\delta,\end{aligned} \tag{28}$$

where $\Gamma_\delta > 0$ and $k > 0$ are both NN learning rates.

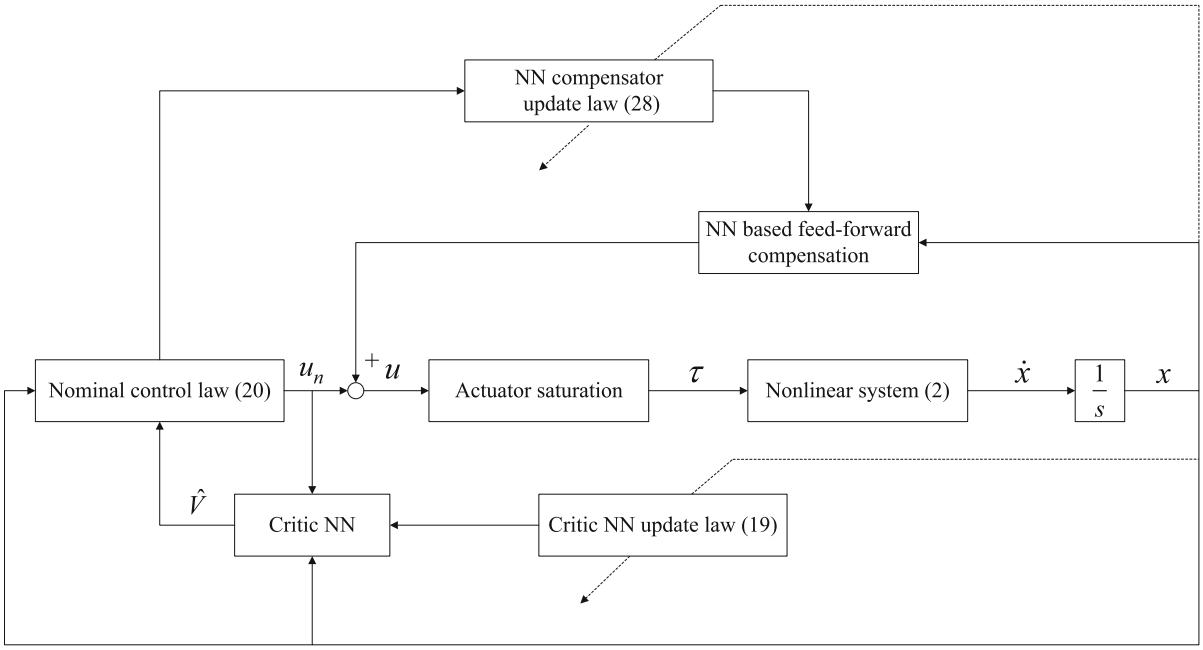


**Fig. 1** The presented NN compensation-based control architecture

Define $\tilde{\delta} = \delta - \hat{\delta}$ as the overall NN approximation error. We have

$$\begin{aligned}\tilde{\delta} &= W_\delta^{\mathsf{T}}\sigma_\delta(x) - \hat{W}_\delta^{\mathsf{T}}\sigma_\delta(x) + \varepsilon_\delta(x)\\ &= \tilde{W}_\delta^{\mathsf{T}}\sigma_\delta(x) + \varepsilon_\delta(x).\end{aligned} \tag{29}$$

Therefore, based on the approximated $\delta(x)$, the unknown actuator saturation problem can be tackled by designing a feed-forward compensation for the nominal optimal control (20). From this viewpoint, the overall control law for nonlinear system (2) is designed as

$$u = u_n + \hat{\delta}, \tag{30}$$

where the NN-based saturation compensator $\hat{\delta}$ is used to compensate for the saturation nonlinearity. In summary, the proposed NN compensation-based overall control architecture is illustrated in Fig. 1.

*Remark 1* Actually, external perturbations which include the exogenous signals and model uncertainties have to be considered to satisfy the requirements in real implementations. And indeed, some existing studies have focused on solving optimal control problems based on ADP for nonlinear systems with external perturbations by constructing improved cost function [26] or adding a compensation robust term [53]. To reduce the optimal controller design procedure, external perturbations are not handled in the ADP-based stabilization in this paper. We can design optimal controller for real implementations inspired from the literature such as [26,53].

### 3.3 Local online PI algorithm

The feed-forward NN compensation-based optimal stabilization algorithm (30) which consists of online optimal stabilization based on (20) and the NN compensator based on (27) is described as Algorithm 1.

From Algorithm 1, we can see that $V^{(0)}(x) = 0$ is required. It is required to prove the convergence of Algorithm 1, e.g., $V^{(p)}(x) \to V^*(x)$ and $u_n^{(p)}(x) \to u_n^*(x)$ as $p \to \infty$.

### 3.4 Stability analysis

Before the stability analysis, necessary assumptions should be listed as follows.

**Assumption 1** The NN approximation error of saturation nonlinearity is bounded, i.e., $\|\varepsilon_\delta(x)\| \le \varepsilon_{\delta M}$, where $\varepsilon_{\delta M} > 0$ is a constant.

**Algorithm 1** Feed-forward compensation-based online optimal stabilization algorithm

1: Select a set of small positive constants $\xi$, the maximum iteration time $M$, the maximum run step $N$, the initial values $\hat{W}_\delta^{(0)}$ and $\hat{W}_c^{(0)}$ of corresponding NN weight vectors. Let $p = 0$, $q = 0$ and $V^{(0)}(x) = 0$, and begin with a given nominal control policy $u_n^{(0)}(x)$.

2: **(Policy evaluation)** Let $p > 0$, solve the following nonlinear Lyapunov equation for the control policy $u^{(p)}(x)$:

$$0 = U(x, u_n^{(p)}) + \nabla V^{(p)\mathsf{T}}(x)\left(f(x) + g(x)u_n\right). \tag{31}$$

3: **(Policy improvement)** Update the control policy $u_n^{(p)}(x)$ by

$$u_n^{(p+1)}(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla V^{(p)}(x). \tag{32}$$

4: If $\left\|V^{(p+1)}(x) - V^{(p)}(x)\right\| \le \xi$, stop and obtain the approximated optimal control; else, let $p = p + 1$, if $p < M$, return to Step 2, otherwise go to Step 5.

5: **(Feed-forward compensation)** Update the weight vector $\hat{W}_\delta$ of NN by

$$\dot{\hat{W}}_\delta^{(q+1)} = \Gamma_\delta \sigma_\delta(x)\left(2u_n^{(p+1)\mathsf{T}}R - x^{\mathsf{T}}g(x)\right) + k\Gamma_\delta \|x\| \hat{W}_\delta^{(q)}. \tag{33}$$

And obtain the approximate unknown term $\hat{\delta}^{(q+1)}(x)$ as

$$\hat{\delta}^{(q+1)}(x) = \hat{W}_\delta^{(q)\mathsf{T}}\sigma_\delta(x). \tag{34}$$

6: **(Overall control policy)** Update the overall control policy $u^{(p)}(x)$ by

$$u^{(q+1)} = u_n^{(p+1)} + \hat{\delta}^{(q+1)}(x). \tag{35}$$

7: If $j < N$, return to Step 2; else, stop running.

**Assumption 2** There exist positive constants $\delta_M$ and $\delta_m$ such that $\|W_\delta\| \le \delta_M$ and $\left\|\tilde{W}_\delta\right\| \le \delta_m$, respectively.

**Theorem 2** *Consider nonlinear system subject to unknown actuator saturation (2), the transformed dynamics (25), as well as Assumptions 1 and 2, if the overall control law is designed as (30), which is composed of the online nominal optimal control (20) and NN-based feed-forward saturation compensation (27) via the update law (28), all the signals of the closed-loop nonlinear system can be guaranteed to be UUB.*

*Proof* Choose the Lyapunov function candidate as

$$L_2 = \frac{1}{2}x^{\mathsf{T}}x + V(x) + tr\left(\frac{1}{2}\tilde{W}_\delta^{\mathsf{T}}\Gamma_\delta^{-1}\tilde{W}_\delta\right), \tag{36}$$

where $tr(\cdot)$ indicates the trace of the matrix.

The time derivative of (36) is

$$\begin{aligned}\dot{L}_2 &= x^{\mathsf{T}}\dot{x} + \dot{V}(x) + tr\left(\tilde{W}_\delta^{\mathsf{T}}\Gamma_\delta^{-1}\dot{\tilde{W}}_\delta\right)\\ &= x^{\mathsf{T}}\left(f(x) + g(x)(u - \delta)\right) + \dot{V}(x)\\ &\quad + tr\left(\tilde{W}_\delta^{\mathsf{T}}\Gamma_\delta^{-1}\dot{\tilde{W}}_\delta\right).\end{aligned} \tag{37}$$

In the existence of saturation nonlinearity, for the second item of (37), we have

$$\begin{aligned}\dot{V}(x) &= \nabla V^{\mathsf{T}}(x)\dot{x}\\ &= \nabla V^{\mathsf{T}}(x)\left(f(x) + g(x)u\right) - \nabla V^{\mathsf{T}}(x)g(x)\delta.\end{aligned} \tag{38}$$

Define $\tilde{\delta} = \delta - \hat{\delta}$. Then, introducing the presented overall control law (30), (38) becomes

$$\dot{V}(x) = \nabla V^{\mathsf{T}}(x)\left(f(x) + g(x)u_n\right) - \nabla V^{\mathsf{T}}(x)g(x)\tilde{\delta}. \tag{39}$$

According to (7), (9), and (30), one has

$$\begin{aligned}\dot{L}_2 &= x^{\mathsf{T}}\left(f(x) + g(x)u_n\right) - x^{\mathsf{T}}Qx - u_n^{\mathsf{T}}Ru_n\\ &\quad + \left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right)\tilde{\delta} + tr\left(\tilde{W}_\delta^{\mathsf{T}}\Gamma_\delta^{-1}\dot{\tilde{W}}_\delta\right).\end{aligned} \tag{40}$$

Since $f(x)$ is locally Lipschitz, there exists a positive constant $D_f$ which satisfies $\|f(x)\| \le D_f\|x\|$. Suppose that $\|g(x)\| \le D_g$ [53]. Thus, (40) becomes

$$\begin{aligned}\dot{L}_2 \le & D_f\|x\|^2 + \frac{1}{2}\|x\|^2 + \frac{1}{2}D_g^2\|u_n\|^2 - \lambda_{\min}(Q)\|x\|^2\\ &- \lambda_{\min}(R)\|u_n\|^2 + \left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right)\tilde{\delta}\\ &+ tr\left(\tilde{W}_\delta^{\mathsf{T}}\Gamma_\delta^{-1}\dot{\tilde{W}}_\delta\right),\end{aligned} \tag{41}$$

where $\lambda_{\min}(\cdot)$ indicates the matrix minimum eigenvalue.

Combining (26), (27), and (28), one has

$$\begin{aligned}\dot{L}_2 \le & D_f\|x\|^2 + \frac{1}{2}\|x\|^2 + \frac{1}{2}D_g^2\|u_n\|^2 - \lambda_{\min}(Q)\|x\|^2\\ &- \lambda_{\min}(R)\|u_n\|^2 + \left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right)\tilde{\delta}\end{aligned}$$

$$
\begin{aligned}
&- tr\left(\tilde{W}_\delta^{\mathsf{T}}(\sigma_\delta(x)\left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right) + k\,\|x\|\,\hat{W}_\delta)\right)\\
=& -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}\right)\|x\|^2\\
&-\left(\lambda_{\min}(R) - \frac{1}{2}D_g^2\right)\|u_n\|^2\\
&+\left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right)\left(\tilde{W}_\delta^{\mathsf{T}}\sigma_\delta(x) + \varepsilon_\delta(x)\right)\\
&- tr\left(\tilde{W}_\delta^{\mathsf{T}}(\sigma_\delta(x)\left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right) + k\,\|x\|\,\hat{W}_\delta)\right)\\
=& -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}\right)\|x\|^2\\
&-\left(\lambda_{\min}(R) - \frac{1}{2}D_g^2\right)\|u_n\|^2\\
&+\left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right)\varepsilon_\delta(x) - k\,\|x\|\,tr\left(\tilde{W}_\delta^{\mathsf{T}}\hat{W}_\delta\right)\\
=& -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}\right)\|x\|^2\\
&-\left(\lambda_{\min}(R) - \frac{1}{2}D_g^2\right)\|u_n\|^2\\
&+\left(2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right)\varepsilon_\delta(x)\\
&- k\,\|x\|\,tr\left(\tilde{W}_\delta^{\mathsf{T}}\left(W_\delta - \tilde{W}_\delta\right)\right). \qquad (42)
\end{aligned}
$$

According to Assumptions 1 and 2, and supposing that $\left\|2u_n^{\mathsf{T}}R - x^{\mathsf{T}}g(x)\right\| \le \upsilon$, (42) becomes

$$
\begin{aligned}
\dot{L}_2 \le& -\left(\lambda_{\min}(Q) - D_f - \frac{1}{2}\right)\|x\|^2\\
&-\left(\lambda_{\min}(R) - \frac{1}{2}D_g^2\right)\|u_n\|^2\\
&+\varepsilon_{\delta M}\upsilon - k\,\|x\|\left(\delta_M\delta_m - \delta_m^2\right). \qquad (43)
\end{aligned}
$$

Let $A = \lambda_{\min}(Q) - D_f - \frac{1}{2}$, $B = k\left(\delta_M\delta_m - \delta_m^2\right)$. From (43), we can conclude that $\dot{L}_2 \le 0$ when the state $x$ lies outside of the compact set

$$
\Omega_x = \left\{x : \|x\| \le \frac{-B + \sqrt{B^2 + 4A\varepsilon_{\delta M}\upsilon}}{2A}\right\}
$$

with the following conditions:

$$
\begin{cases}
\lambda_{\min}(Q) > D_f + \frac{1}{2},\\
\lambda_{\min}(R) \ge \frac{1}{2}D_g^2.
\end{cases}
$$

It implies that all the signals of the closed-loop nonlinear system with unknown actuator saturation can be guaranteed to be UUB. This completes the proof. □

*Remark 2* Many existing works have tackled control problems for nonlinear systems [35–37], but they never considered their control performance in an optimal manner. On the other hand, some ADP-based optimal control approaches have been developed for nonlinear systems with available actuator saturation [46–50], rather than unknown actuator saturation, which is more common in practice since the outputs of actuators may be biased or suddenly abrupt in practice. To tackle this problem, a NN compensation-based ADP stabilizing algorithm is proposed in this paper. It guarantees not only nonlinear systems with unknown actuator saturation to be stable, but also in an optimal manner. In other words, the major improvement and advantage of this paper lie in that the proposed method not only deals with unknown actuator saturation, but also guarantees the control performance of nonlinear systems to be optimum.

*Remark 3* The model uncertainties and exogenous signals are important issues to be considered in practice. Actually, many existing ADP-based control schemes have been proposed for nonlinear systems with uncertainties and external disturbances [26,47,54–56]. From them, we can conclude that the control strategy to solve this problem is to add a robustifying term in cost functions, which are used for designing an ADP-based control approaches in this situation. The aim of this paper lies in that an ADP-based optimal control scheme is presented for nonlinear systems with unknown actuator saturation. Thus, optimal controllers which consider model uncertainties and exogenous signals can be designed by referring to these existing strategies, and it will be focused on in our future work.

## 4 Simulation studies

In the simulation section, two numerical examples are given to show the effectiveness of the developed ADP-based control scheme.

### 4.1 Example 1

Consider a torsional pendulum system which is described as [57]

$$\begin{cases} \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \omega, \\ J\dfrac{\mathrm{d}\omega}{\mathrm{d}t} = \tau - Mgl\sin\theta - f_d\dfrac{\mathrm{d}\theta}{\mathrm{d}t}, \end{cases}$$

where $M = \frac{1}{3}$ kg and $l = \frac{2}{3}$ m denote the mass and length of the pendulum bar, respectively. The angle $\theta$ and the angular velocity $\omega$ are the system states. $J = \frac{4}{3}Ml^2$ and $f_\mathrm{d} = 0.2$ are the rotary inertia and frictional factor, respectively. $g = 9.8\,\mathrm{m/s}^2$ is the gravitational acceleration. $\tau \in \mathbb{R}$ is the actual applied control input with unknown actuator saturation; in this simulation, it is chosen, respectively, for two cases as

*Case 1*

$$\tau = \mathrm{sat}(u) = \begin{cases} 0.1, & u > 0.1, \\ u, & -0.1 \le u \le 0.1, \\ -0.1, & u < -0.1. \end{cases}$$

*Case 2*

$$\tau = \mathrm{sat}(u) = \begin{cases} 0.2, & u > 0.2, \\ u, & -0.2 \le u \le 0.2, \\ -0.2, & u < -0.2. \end{cases}$$

Define $x = [x_1, x_2]^\mathsf{T} = [\theta, \omega]^\mathsf{T} \in \mathbb{R}^2$ as the state vector of the torsional pendulum system, whose initial state is $x_0 = [1, -1]^\mathsf{T}$. In this simulation, the cost function (4) is approximated by the critic NN, whose weight vector is indicated as $\hat{W}_c = [\hat{W}_{c1}, \hat{W}_{c2}, \hat{W}_{c3}]^\mathsf{T}$, and its initial value is selected as $\hat{W}_{c0} = [0.6, 0.2, 0.9]^\mathsf{T}$. The activation function of the critic NN is set as $\sigma_c(x) = \left[x_1^2,\ x_1x_2,\ x_2^2\right]$. Let $Q = 10I_2$ and $R = 10I$, where $I_n$ denotes identity matrix with $n$ dimensions, the critic NN learning rate be $l_c = 0.0002$, the learning rates of NN saturation compensator be $\Gamma_\delta = 0.01$ and $k = 1$, respectively.

In case 1, the simulation results are illustrated in Figs. 2, 3, 4, 5, and 6. As it is displayed in Fig. 2, one can observe that the critic NN weights converge to $[0.6108, 0.1764, 0.9576]^\mathsf{T}$, which indicates Theorem 1. From Fig. 3, the NN-based saturation compensator (27) is employed to overcome the negative affection from the unknown actuator saturation. We can see that the actual control input illustrated in Fig. 4 reaches the actuator bounds we predefined at the beginning of the operation process. After some settling time, the control input signal varies within the bounded values. Fortunately, with the proposed ADP-based stabilizing control scheme (30), the system states are still convergent as shown in Fig. 5. For the same simulation within the time interval $[0, 2000s]$, the amplified system states are shown in Fig. 6, from where we can see that all the signals of system states converge within a compact set $4.0 \times 10^{-6}$. This indicates the system states can be ensured to be UUB as concluded in Theorem 2. Alternatively in case 2, Figs. 7 and 8 provide the actual saturated control input and system states under the same controller as case 1, respectively. We can see that the developed stabilization scheme is effective even though the torsional pendulum system is driven by actuators with different unknown saturation bounds.

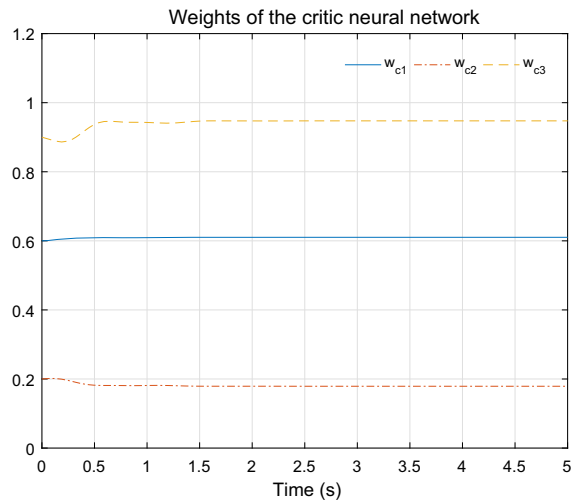


**Fig. 2** Critic NN weights of torsional pendulum system

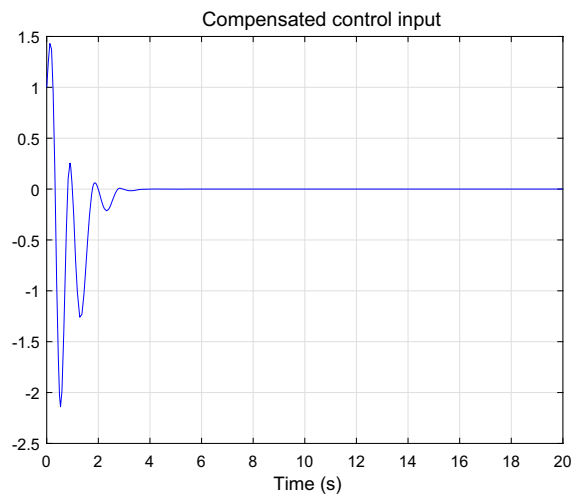


**Fig. 3** Compensated control input of torsional pendulum system

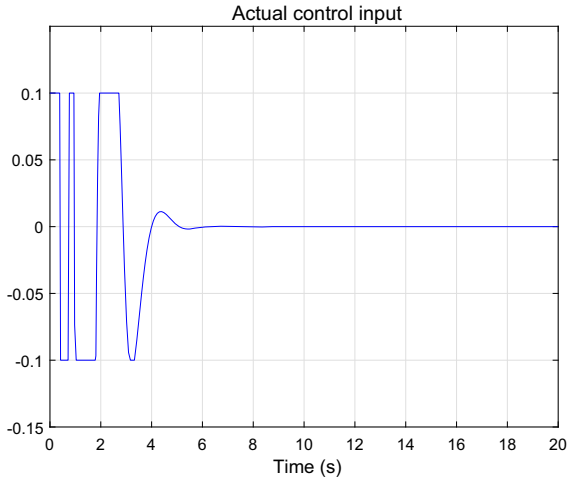


**Fig. 4** Actual control input of torsional pendulum system

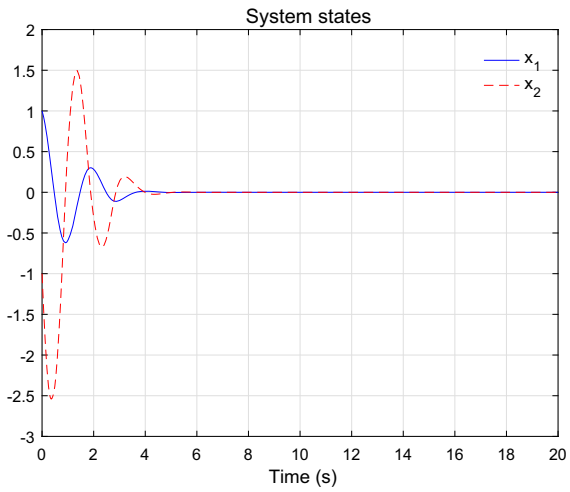


**Fig. 5** System states of torsional pendulum system

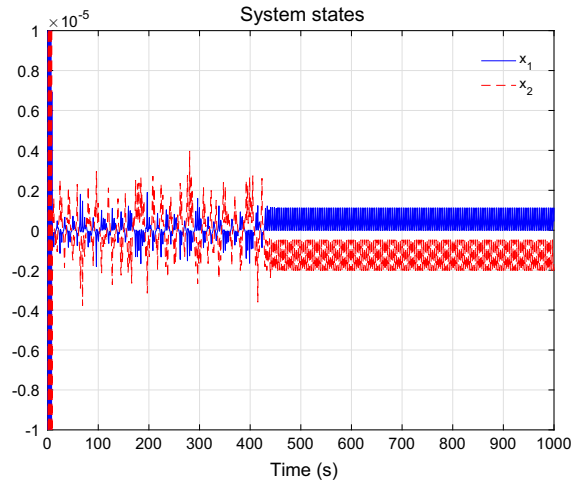


**Fig. 6** Amplified system states of torsional pendulum system

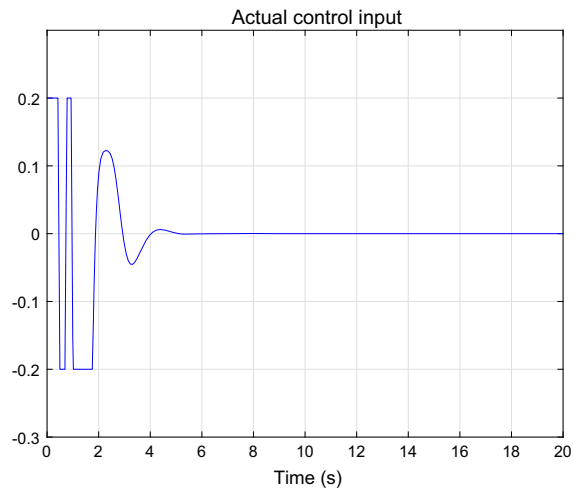


**Fig. 7** Actual control input of torsional pendulum system in case 2

## 4.2 Example 2

Consider the following nonlinear system [58]:

$$\dot{x} = \begin{bmatrix} x_2 - x_1 \\ -0.5x_1 - 0.5x_2 + 0.5x_2\,(\cos(2x_1) + 2)^2 \\ x_4 - x_3 \\ -x_3 - 0.5x_4 + 0.5x_4x_3^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \cos(2x_1) + 2 & 0 \\ 0 & 0 \\ 0 & x_3 \end{bmatrix} \tau, \tag{44}$$

where $x = [x_1, x_2, x_3, x_4]^{\mathsf{T}} \in \mathbb{R}^4$ is the system state vector and $\tau = [\tau_1, \tau_2]^{\mathsf{T}} \in \mathbb{R}^2$ is the actual control input vector. For the purpose of simulation, we, respectively, define the unknown actuator saturation of two cases as

*Case 1*

$$\tau_i = \mathrm{sat}(u_i) = \begin{cases} 3, & u_i > 3, \\ u_i, & -3 \le u_i \le 3, \\ -3, & u_i < -3, \end{cases}$$

where $i = 1, 2$.

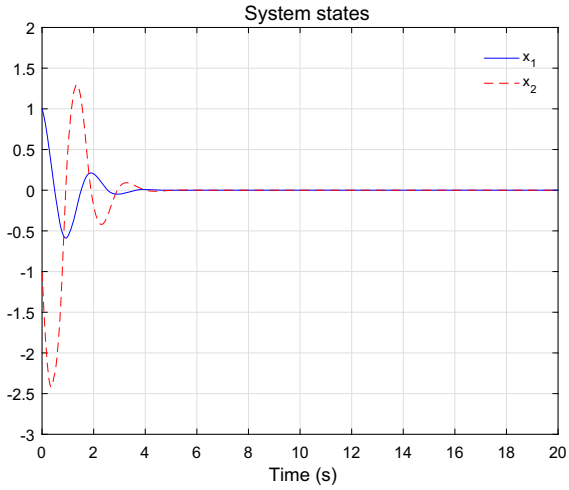


**Fig. 8** System states of torsional pendulum system in case 2

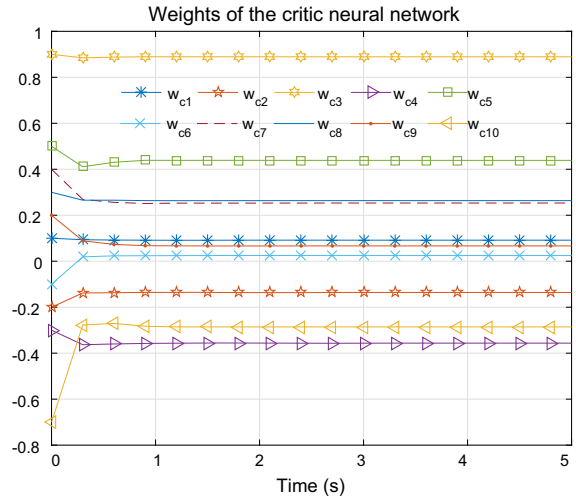


**Fig. 9** Critic NN weights of nonlinear system in Example 2

*Case 2*

$$\tau_1 = \mathrm{sat}(u_1) = \begin{cases} 5, & u_1 > 5, \\ u_1, & -5 \le u_1 \le 5, \\ -5, & u_1 < -5, \end{cases}$$

and

$$\tau_2 = \mathrm{sat}(u_2) = \begin{cases} 6, & u_2 > 6, \\ u_2, & -6 \le u_2 \le 6, \\ -6, & u_2 < -6. \end{cases}$$

Let the initial state vector be $x_0 = [1, -1, 2, -2]^{\mathsf{T}}$. The critic NN weight vector is indicated as $\hat{W}_c = [\hat{W}_{c1}, \hat{W}_{c2}, \ldots, \hat{W}_{c10}]^{\mathsf{T}}$, whose initial value is $\hat{W}_{c0} = [0.1, -0.2, 0.9, -0.3, 0.5, -0.1, 0.4, 0.3, 0.2, -0.7]^{\mathsf{T}}$. The critic NN activation function is selected as $\sigma_c(x) = [x_1^2, x_1x_2, x_1x_3, x_1x_4, x_2^2, x_2x_3, x_2x_4, x_3^2, x_3x_4, x_4^2]$. Let $Q = I_4$, $R = 0.1I_2$, the critic NN learning rate be $l_c = 0.0002$, the learning rates of the NN for saturation compensator be $\Gamma_\delta = 0.0001$ and $k = 1$, respectively.

The simulation results in case 1 are displayed in Figs. 9, 10, 11, 12, and 13. As shown in Fig. 9, it illustrates that the critic NN weights converge to $[0.0916, -0.1359, 0.8892, -0.3563, 0.4380, 0.0247, 0.2536, 0.2636, 0.0666, -0.2856]^{\mathsf{T}}$. Figure 10 shows the feed-forward compensation of the unknown saturation nonlinearity via the NN. By using the proposed control scheme, the actual control inputs illustrated in Fig. 11 drive the states of the nonlinear system (44) to convergence, which are shown in Fig. 12. In detail, the

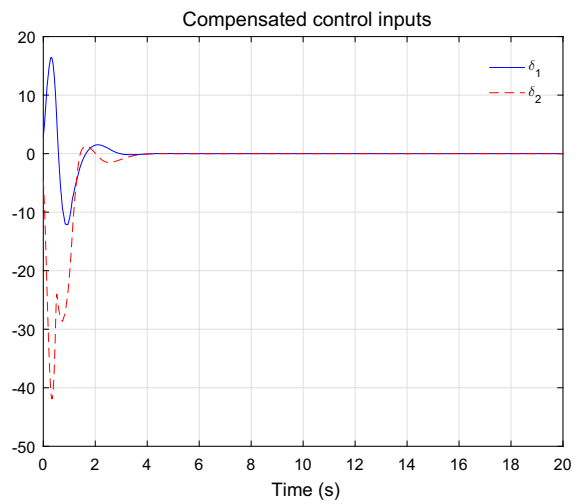


**Fig. 10** Compensated control inputs of nonlinear system in Example 2

amplified system states are illustrated in Fig. 13, which describes all the signals of system states converge to a compact set $1.0 \times 10^{-6}$ when the time sequence runs sufficiently long. This accords with the conclusion of Theorem 2. From these figures, we notice that the system states converge to equilibrium point, though the unknown actuator saturation exists. For case 2, Fig. 14 shows the converged system states under the actual control input as shown in Fig. 15 which is obtained from the same control algorithm as case 1. Thus, we can declare that the proposed ADP-based stabilizing con-

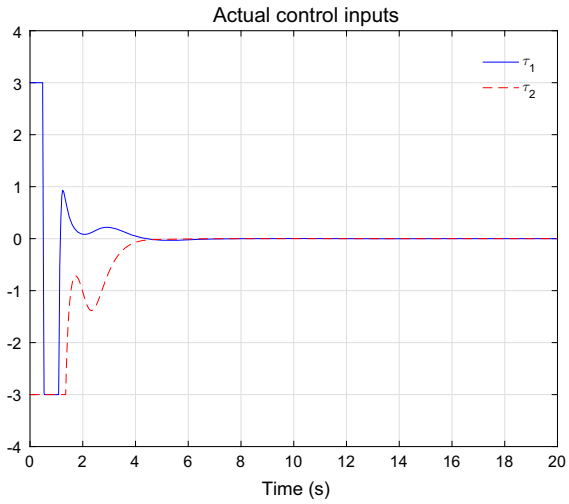


**Fig. 11** Actual control inputs of nonlinear system in Example 2

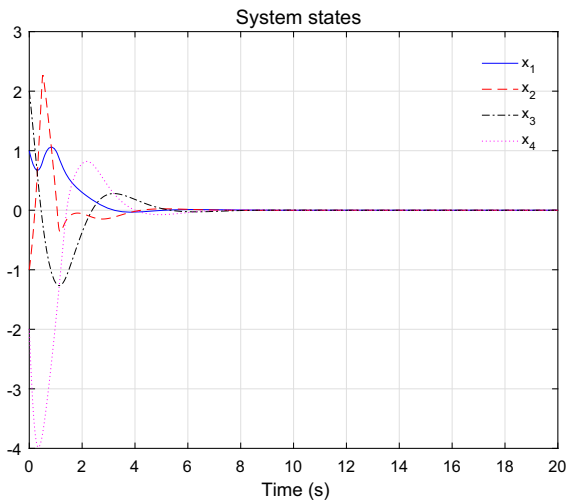


**Fig. 12** System states of nonlinear system in Example 2

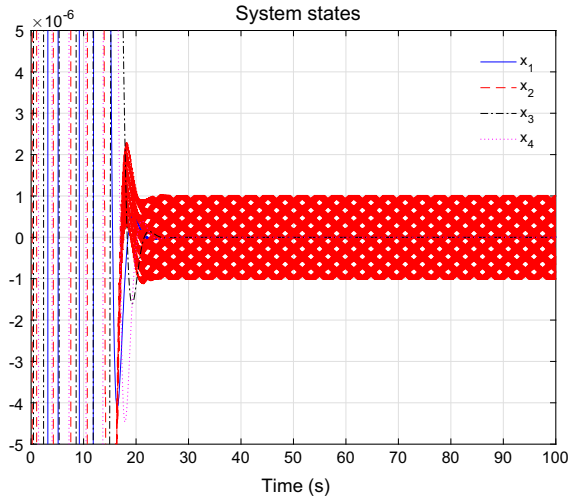


**Fig. 13** Amplified system states of nonlinear system in Example 2

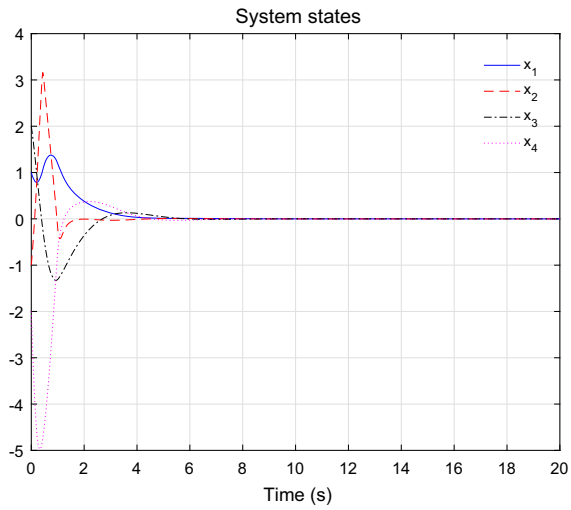


**Fig. 14** System states of nonlinear system in Example 2 of case 2

trol scheme is effective for nonlinear systems subject to unknown actuator saturation.

*Remark 4* The assumptions are feasible in these two examples. Actually, all signals in practical systems are norm-bounded. Taking Example 1 as an example, the compensated NN is used to approximate the unknown saturation nonlinearity (24), which is bounded since the driving ability of actuator is limited. Furthermore, the NN weight and its estimation are bounded such that they and their estimation errors have upper bound.

*Remark 5* In the existing methods [46–50], the authors have addressed the ADP-based optimal control problems for nonlinear systems with available actuator saturation. In contrast to them, this paper considers nonlinear systems in the presence of unknown actuator saturation, which is tackled by a feed-forward NN compensation-based ADP stabilization scheme. In other words, this paper extends the ADP method to a new controlled plant; this is the key contribution of this work.

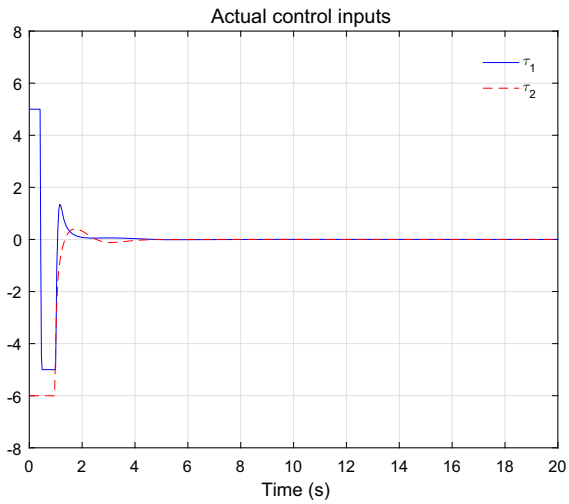


**Fig. 15** Actual control inputs of nonlinear system in Example 2 of case 2

## 5 Conclusion

In this paper, the stabilizing control problem of nonlinear systems subject to unknown actuator saturation is tackled by using NN compensation-based ADP algorithm. The online updated critic NN is adopted to derive the cost function approximately. As well, the nominal optimal control can be obtained thereby. By constructing a NN-based feed-forward saturation compensator, the overall ADP-based stabilizing control is implemented to reduce the influence of the unknown actuator saturation. Simulation results demonstrate that the proposed control scheme is effective. This strategy is utilized to deal with the stabilizing problem without any a priori knowledge of the limit bounds of saturated actuators, as well as the initial stabilizing control and the persisting of excitation condition, which are always required in traditional ADP methods. It is worth mentioning that the developed stabilization method is concerned with nonlinear systems with available dynamics. In the future, the challenges on unknown nonlinear systems and trajectory tracking problems will be enhanced for similar optimal control problems to better adapt the requirements of practical systems more. Meanwhile, general computational intelligence, such as fuzzy logic systems, evolutionary computation, can be applied to solve optimal control problems in the smart framework.

**Acknowledgements** This work was supported in part by the National Natural Science Foundation of China under Grants 61603387, 61533017, 61374051 and 61502494, and in part by the Early Career Development Award of SKLMCCS under Grant 20180201.

## References

1. Bellman, R.E.: Dynamic Programming. Princeton University Press, New Jersey (1957)
2. Werbos, P.J.: Approximate dynamic programming for real time control and neural modeling. In: White, D.A., Sofge, D.A. (eds.) Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches. Van Nostrand Reinhold, New York (1992)
3. Liu, D., Wei, Q., Wang, D., Yang, X., Li, H.: Adaptive dynamic programming with applications in optimal control. Springer, Cham (2017)
4. Ge, S., Hang, C., Zhang, T.: Adaptive neural network control of nonlinear systems by state and output feedback. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) **29**(6), 818–828 (1999)
5. Zhao, B., Li, Y.: Local joint information based active fault tolerant control for reconfigurable manipulator. Nonlinear Dyn. **77**(3), 859–876 (2014)
6. Zhao, B., Li, Y., Liu, D.: Self-tuned local feedback gain based decentralized fault tolerant control for a class of large-scale nonlinear systems. Neurocomputing **235**, 147–156 (2017)
7. Wang, F.Y., Zhang, H., Liu, D.: Adaptive dynamic programming: an introduction. IEEE Comput. Intell. Mag. **4**(2), 39–47 (2009)
8. Al-Tamimi, A., Lewis, F.L., Abu-Khalaf, M.: Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) **38**(4), 943–949 (2008)
9. Prokhorov, D.V., Wunsch, D.C.: Adaptive critic designs. IEEE Trans. Neural Netw. **8**(5), 997–1007 (1997)
10. Bertsekas, D.P., Tsitsiklis, J.N.: Neuro-dynamic programming: an overview. In: Proceedings of the 34th IEEE Conference on Decision and Control, pp. 560–564 (1995)
11. Si, J., Wang, Y.T.: On-line learning control by association and reinforcement. IEEE Trans. Neural Netw. **12**, 264–276 (2001)
12. Li, D.J., Tang, L., Liu, Y.J.: Adaptive intelligence learning for nonlinear chaotic systems. Nonlinear Dyn. **73**(4), 2103–2109 (2013)
13. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: a survey. J. Artif. Intell. Res. **4**, 237–285 (1996)
14. Liu, D., Wang, D., Li, H.: Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach. IEEE Trans. Neural Netw. Learn. Syst. **25**(2), 418–428 (2014)
15. Wang, D., He, H., Zhao, B., Liu, D.: Adaptive near-optimal controllers for nonlinear decentralised feedback stabilisation problems. IET Control Theory Appl. **11**(6), 799–806 (2017)

16. Mehraeen, S., Jagannathan, S.: Decentralized optimal control of a class of interconnected nonlinear discrete-time systems by using online Hamilton–Jacobi–Bellman formulation. IEEE Trans. Neural Netw. **22**(11), 1757–1769 (2011)
17. Lin, Q., Wei, Q., Zhao, B.: Optimal control for discrete-time systems with actuator saturation. Optim. Contr. Appl. Met. **38**, 1071–1080 (2017)
18. Gao, W., Jiang, Y., Jiang, Z.P., Chai, T.: Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming. Automatica **72**, 37–45 (2016)
19. Fan, Q., Yang, G.: Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances. IEEE Trans. Neural Netw. Learn. Syst. **27**(1), 165–177 (2016)
20. Zhao, B., Liu, D., Yang, X., Li, Y.: Observer-critic structure-based adaptive dynamic programming for decentralised tracking control of unknown large-scale nonlinear systems. Int. J. Syst. Sci. **48**(9), 1978–1989 (2017)
21. Zhao, B., Li, Y.: Model-free adaptive dynamic programming based near-optimal decentralized tracking Control of reconfigurable manipulators. Int. J. Control Autom. Syst. **16**(2), 478–490 (2018)
22. Zhang, H., Song, R., Wei, Q., Zhang, T.: Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming. IEEE Trans. Neural Netw. **22**(12), 1851–1862 (2011)
23. Xie, C., Yang, G.: Approximate guaranteed cost fault-tolerant control of unknown nonlinear systems with time-varying actuator faults. Nonlinear Dyn. **83**(1–2), 269–282 (2016)
24. Zhao, B., Liu, D., Li, Y.: Online fault compensation control based on policy iteration algorithm for a class of affine nonlinear systems with actuator failures. IET Control Theory Appl. **10**(15), 1816–1823 (2016)
25. Fu, Y., Fu, J., Chai, T.: Robust adaptive dynamic programming of two-player zero-sum games for continuous-time linear systems. IEEE Trans. Neural Netw. Learn. Syst. **26**(12), 3314–3319 (2015)
26. Wang, D., Mu, C., He, H., Liu, D.: Event-driven adaptive robust control of nonlinear systems with uncertainties through NDP strategy. IEEE Trans. Syst. Man Cybern. Syst. **47**(7), 1358–1370 (2017)
27. Sun, L., Huo, W., Jiao, Z.: Adaptive backstepping control of spacecraft rendezvous and proximity operations with input saturation and full-state constraint. IEEE Trans. Industr. Electron. **64**(1), 480–492 (2017)
28. Wang, F., Hua, C., Zong, Q.: Attitude control of reusable launch vehicle in reentry phase with input constraint via robust adaptive backstepping control. Int. J. Aadpt. Control. **29**(10), 1308–1327 (2015)
29. Mendoza, M., Zavala-Rło, A., Santibnez, V., Reyes, F.: Output-feedback proportional–integral-derivative-type control with simple tuning for the global regulation of robot manipulators with input constraints. IET Control Theory Appl. **9**(14), 2097–2106 (2015)
30. Chen, X., Jia, Y., Matsuno, F.: Tracking control for differential-drive mobile robots with diamond-shaped input constraints. IEEE Trans. Control Syst. Technol. **22**(5), 1999–2006 (2014)
31. Kiefer, T., Graichen, K., Kugi, A.: Trajectory tracking of a 3-DOF laboratory helicopter under input and state constraints. IEEE Trans. Control Syst. Technol. **18**(4), 944–952 (2010)
32. Zhai, D., Xia, Y.: Adaptive control for teleoperation system with varying time delays and input saturation constraints. IEEE Trans. Industr. Electron. **63**(11), 6921–6929 (2016)
33. Zhao, F., Ge, S., Tu, F., Qin, Y., Dong, M.: Adaptive neural network control for active suspension system with actuator saturation. IET Control Theory Appl. **10**(14), 1696–1705 (2016)
34. Sun, J., Liu, C.: Finite-horizon differential games for missile–target interception system using adaptive dynamic programming with input constraints. Int. J. Syst. Sci. **49**(2), 1–20 (2018)
35. Liu, Y., Tong, S., Li, D., Gao, Y.: Fuzzy adaptive control with state observer for a class of nonlinear discrete-time systems with input constraint. IEEE Trans. Fuzzy Syst. **24**(5), 1147–1158 (2016)
36. Chen, M., Ge, S., Ren, B.: Adaptive tracking control of uncertain MIMO nonlinear systems with input constraints. Automatica **47**(3), 452–465 (2011)
37. Song, G., Li, T., Hu, K., Zheng, B.: Observer-based quantized control of nonlinear systems with input saturation. Nonlinear Dyn. **86**(2), 1157–1169 (2016)
38. He, P., Jagannathan, S.: Reinforcement learning-based output feedback control of nonlinear systems with input constraints. IEEE Trans. Syst. Man Cybern. Part B (Cybernetics) **35**(1), 150–154 (2005)
39. Abu-Khalaf, M., Lewis, F.L., Huang, J.: Neurodynamic programming and zero-sum games for constrained control systems. IEEE Trans. Neural Netw. **19**(7), 1243–1252 (2008)
40. Heydari, A., Balakrishnan, S.N.: Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics. IEEE Trans. Neural Netw. Learn. Syst. **24**(1), 145–157 (2013)
41. Zhang, H., Qin, C., Luo, Y.: Neural-network-based constrained optimal control scheme for discrete-time switched nonlinear system using dual heuristic programming. IEEE Trans. Autom. Sci. Eng. **11**(3), 839–849 (2014)
42. Dong, L., Zhong, X., Sun, C., He, H.: Event-triggered adaptive dynamic programming for continuous-time systems with control constraints. IEEE Trans. Neural Netw. Learn. **28**(8), 1941–1952 (2017)
43. Modares, H., Lewis, F.L., Naghibi-Sistani, M.B.: Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks. IEEE Trans. Neural Netw. Learn. Syst. **24**(10), 1513–1525 (2013)
44. Xu, H., Zhao, Q., Jagannathan, S.: Finite-horizon near-optimal output feedback neural network control of quantized nonlinear discrete-time systems with input constraint. IEEE Trans. Neural Netw. Learn. Syst. **26**(8), 1776–1788 (2015)
45. Song, R., Zhang, H., Luo, Y., Wei, Q.: Optimal control laws for time-delay systems with saturating actuators based on heuristic dynamic programming. Neurocomputing **73**(16), 3020–3027 (2010)
46. Zhang, H., Luo, Y., Liu, D.: Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. IEEE Trans. Neural Netw. **20**(9), 1490–1503 (2009)
47. Liu, D., Yang, X., Wang, D., Wei, Q.: Reinforcement-learning-based robust controller design for continuous-time

uncertain nonlinear systems subject to input constraints. IEEE Trans. Cybern. **45**(7), 1372–1385 (2015)
48. Liu, D., Wang, D., Yang, X.: An iterative adaptive dynamic programming algorithm for optimal control of unknown discrete-time nonlinear systems with constrained inputs. Inform Sci. **220**, 331–342 (2013)
49. Yang, X., Liu, D., Ma, H., Xu, Y.: Online approximate solution of HJI equation for unknown constrained-input nonlinear continuous-time systems. Inform Sci. **328**, 435–454 (2016)
50. Yang, X., Liu, D., Luo, B., Li, C.: Data-based robust adaptive control for a class of unknown nonlinear constrained-input systems via integral reinforcement learning. Inform Sci. **369**, 731–747 (2016)
51. Pomprapa, A., Leonhardt, S., Misgeld, B.J.E.: Optimal learning control of oxygen saturation using a policy iteration algorithm and a proof-of-concept in an interconnecting three-tank system. Control Eng. Pract. **59**, 194–203 (2017)
52. Yang, X., Liu, D., Wang, D.: Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints. Int. J. Contr. **87**(3), 553–566 (2014)
53. Song, R., Lewis, F.L., Wei, Q., Zhang, H.: Off-policy actor-critic structure for optimal control of unknown systems with disturbances. IEEE Trans. Cybern. **46**(5), 1041–1050 (2016)
54. Wu, H.N., Li, M.M., Guo, L.: Finite-horizon approximate optimal guaranteed cost control of uncertain nonlinear systems with application to mars entry guidance. IEEE Trans. Neural Netw. Learn. Syst. **26**(7), 1456–1467 (2015)
55. Terra, M.H., Cerri, J.P., Ishihara, J.Y.: Optimal robust linear quadratic regulator for systems subject to uncertainties. IEEE Trans. Automat. Control **59**(9), 2586–2591 (2014)
56. Cui, X., Zhang, H., Luo, Y., Jiang, H.: Adaptive dynamic programming for $H_\infty$ tracking design of uncertain nonlinear systems with disturbances and input constraints. Int. J. Adapt. Control. **31**(11), 1567–1583 (2017)
57. Zhao, B., Liu, D., Li, Y.: Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems. Inform Sci. **384**, 21–33 (2017)
58. Liu, D., Wei, Q.: Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. IEEE Trans. Neural Netw. Learn. Syst. **25**(3), 621–634 (2014)