Optimal Tracking Control for a Class of Unknown Discrete-time Systems with Actuator Saturation via Data-based ADP Algorithm

SONG Rui-Zhuo[1] XIAO Wen-Dong[1] SUN Chang-Yin[1]

Abstract A novel optimal tracking control method for a class of discrete-time systems with actuator saturation and unknown dynamics is proposed in this paper. The scheme is based on the iterative adaptive dynamic programming (ADP) algorithm. In order to implement the control scheme, a data-based identifier is first constructed for the unknown system dynamics. By introducing the M network, the explicit formula of the steady control is achieved. In order to eliminate the effect of the actuator saturation, a nonquadratic performance functional is presented, and then an iterative ADP algorithm is established to achieve the optimal tracking control solution with convergence analysis. For implementing the optimal control method, neural networks are used to establish the data-based identifier, compute the performance index functional, approximate the optimal control policy and solve the steady control, respectively. Simulation example is provided to verify the effectiveness of the presented optimal tracking control scheme.

Key words Adaptive dynamic programming (ADP), iterative algorithm, optimal tracking control, data-based, identifier

Citation Song Rui-Zhuo, Xiao Wen-Dong, Sun Chang-Yin. Optimal tracking control for a class of unknown discrete-time systems with actuator saturation via data-based ADP algorithm. *Acta Automatica Sinica*, 2013, **39**(9): 1413−1420

DOI 10.3724/SP.J.1004.2013.01413

In most industrial process control systems, the system dynamics are not accurately known, which makes it difficult to design the optimal controller. Fortunately, many industrial processes generate and store huge amounts of process data at every time instant of every day, containing all the valuable state information of process operations and equipment. Using online and offline data, to design controllers directly, predict and assess system states, evaluate performance, make decisions, or even diagnose faults, would be very significant, especially in the absence of accurate process models[1−2]. In recent years, some data-based control schemes are proposed by many scholars. Based only on available system data, [3] proposed a novel switching mechanism to deal with the unknown control direction. Using a set of input/output data, [4] proposed piecewise affine autoregressive exogenous models of the system dynamics. However, the most current studies focus on designing controller for general nonlinear systems, and only few works involve the optimal tracking control problem for unknown system dynamics with actuator saturation.

It is generally known that the actuator saturation is one of the most common nonlinearities in practical applications[5]. In recent decades, the control problem of nonlinear systems with saturating actuators is one of the hot issues in the field of systems control[6−8]. However, most of works do not take into account the optimal tracking control problem for unknown system dynamics. Therefore, this problem will be discussed through Hamilton-Jacobi-Bellman (HJB) framework[9] in this paper.

As it is well known to solve the optimal control problems effectively, dynamic programming is one of the most helpful method[10]. However, except for very special cases, the HJB equation is untenable computationally by dynamic programming, due to the reason of "curse of dimensionality". In consequence, adaptive/approximate dynamic programming (ADP) algorithms have gained a lot of attention from scholars[11−14], to get the approximate solution of HJB equation, in which the neural networks have been widely used[15], and it promotes the development of the neural networks[16−17]. By now, ADP has successfully solved the nonlinear zero-sum differential games problem[18], optimal control problem for MIMO systems[19], on line optimization problem[20], etc. Reference [21] considered the systems with state delays and control delays, and proposed a new optimal control scheme with the M network, which is the highlight of the paper. Reference [22] defined a novel performance index function for the optimal tracking problem. In addition, [5] proposed a new optimal control algorithm for systems with control constraints. But the research results are lacking of detailed analysis of the unknown systems. Aiming at the unknown system dynamics, [23] proposed a forward-in-time method to obtain the optimal control solution. However, the rigorous analyses of the model convergence are absent. Furthermore, [24] presented an optimal control method for unknown nonlinear systems. But the actuator saturation nonlinearities are not considered, which are very important in industrial systems.

It is undeniable that for ADP algorithm in the field of optimal control, the scholars have made significant progress. However, as far as we know, how to obtain the solution of the optimal tracking control problem for unknown system dynamics with actuator saturation via ADP algorithm is very intractable. Therefore, this paper will propose a method to get over this problem. First, the data-based identifier is established with convergence proof. Then, the M network is used to obtain the steady control. Next, the corresponding HJB equation of the optimal tracking problem is derived. A novel iterative ADP algorithm is established to obtain the solution of the HJB equation. And then, for implementing the proposed optimal tracking control method, the neural networks are adopted as the approximators. Briefly, the major highlights of the paper are summarized as follows.

1) The input/output data are used to establish the identifier with rigorous convergence proof.

2) A novel nonquadratic functional is proposed for the optimal tracking problem and the corresponding HJB equa-

Manuscript received December 13, 2012; accepted March 15, 2013
Supported by National Natural Science Foundation of China (61304079, 61125306, 61034002), the Open Research Project from SKLMCCS (20120106), the Fundamental Research Funds for the Central Universities (FRF-TP-13-018A), and the China Postdoctoral Science Foundation (2013M530527)
Recommended by Associate Editor LIU De-Rong
1. School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

tion of the optimal tracking problem is derived.

3) It is proven that the iterative performance index functional sequence converges to the minimum. It is also shown that the HJB equation is established for the optimal tracking control problem.

4) The M network is utilized to approximate the steady control, and one neural network is used to establish the data-based identifier.

The organization of the rest of the paper is as follows. Section 1 presents the motivations and preliminaries. Section 2 is about the ADP-based optimal tracking control method, and theorems of the convergence. Section 3 discusses the implementation of the optimal tracking control method. Section 4 gives an example to show the performance of the proposed method. Section 5 concludes the paper.

1 Motivations and preliminaries

In this paper, we consider the following unknown nonlinear systems

$$\boldsymbol{x}(t+1) = F(\boldsymbol{x}(t), \boldsymbol{u}(t)) \tag{1}$$

where the control $\boldsymbol{u}(t) \in \mathbf{R}^m$ and the state $\boldsymbol{x}(t) \in \mathbf{R}^n$. $F(\boldsymbol{x}(t), \boldsymbol{u}(t))$ is unknown continuous function. It is assumed that the state is completely controllable and bounded in $\Omega \subset \mathbf{R}^n$ and $F(\mathbf{0}, \mathbf{0}) = 0$. We denote $\boldsymbol{u}(t) = [\boldsymbol{u}_1(t) \quad \boldsymbol{u}_2(t) \quad \cdots \quad \boldsymbol{u}_m(t)]^{\mathrm{T}} \in \mathbf{R}^m$, $\boldsymbol{u}_i^{\min} \leq \boldsymbol{u}_i(t) \leq \boldsymbol{u}_i^{\max}$, $i = 1, \cdots, m$, where $\boldsymbol{u}_i^{\min}$ and $\boldsymbol{u}_i^{\max}$ are the saturating bounds for the ith actuator. In this paper, we assume that the unknown system (1) is controllable in Ω, i.e., there exists analytic control policy which makes the system track the specified signal.

Assumption 1. For the system state $\boldsymbol{x}(t)$ and the control input $\boldsymbol{u}(t)$, the Jacobian matrix of system (1) $\dfrac{\partial F(\boldsymbol{x}(t), \boldsymbol{u}(t))}{\partial \boldsymbol{u}(t)}$ is nonsingular.

The assumption is important to guarantee the existence of the control policy. So in the light of the theorem of implicit function, the control $\boldsymbol{u}(t)$ is existent and defined as follows:

$$\boldsymbol{u}(t) = G(\boldsymbol{x}(t), \boldsymbol{x}(t+1)) \tag{2}$$

The reference signal $\boldsymbol{\eta}(t)$ is defined as follows:

$$\boldsymbol{\eta}(t+1) = F_d(\boldsymbol{\eta}(t)) \tag{3}$$

Furthermore, the steady control is

$$\boldsymbol{u}_s(t) = G_d(\boldsymbol{\eta}(t), \boldsymbol{\eta}(t+1)) \tag{4}$$

where G_d satisfies $\boldsymbol{\eta}(t+1) = F(\boldsymbol{\eta}(t), G_d(\boldsymbol{\eta}(t), \boldsymbol{\eta}(t+1)))$.

Remark 1. If the system dynamics of (1) are known, then the steady control $\boldsymbol{u}_s(t)$ can be obtained by the method in [22]. However, in this paper, the system dynamics of (1) are unknown, so a method is proposed in Subsection 2.2 to obtain the steady control $G_d(\cdot)$. Furthermore, in order to ensure that the optimization problem of this paper is solvable, the saturating bounds of system (1) satisfy $\boldsymbol{u}_i^{\max} \geq \boldsymbol{u}_{si}$ and $\boldsymbol{u}_i^{\min} \leq \boldsymbol{u}_{si}$, where $\boldsymbol{u}_s = [\boldsymbol{u}_{s1}, \cdots, \boldsymbol{u}_{sm}]^{\mathrm{T}}$.

Based on the above preparative work, the state error is defined as follows:

$$\boldsymbol{e}(t) = \boldsymbol{x}(t) - \boldsymbol{\eta}(t) \tag{5}$$

and the control error is defined

$$\boldsymbol{v}(t) = \boldsymbol{u}(t) - \boldsymbol{u}_s(t) \tag{6}$$

Here we denote $\boldsymbol{x}(t) = \boldsymbol{e}(t) + \boldsymbol{\eta}(t)$ and $\boldsymbol{u}(t) = \boldsymbol{v}(t) + \boldsymbol{u}_s(t)$, then the following expression is obtained

$$\begin{cases} \boldsymbol{e}(t+1) = \boldsymbol{x}(t+1) - \boldsymbol{\eta}(t+1) = \\ \qquad F(\boldsymbol{x}(t), \boldsymbol{u}(t)) - F_d(\boldsymbol{\eta}(t)) = \\ \qquad F(\boldsymbol{e}(t) + \boldsymbol{\eta}(t), \boldsymbol{v}(t) + \boldsymbol{u}_s(t)) - F_d(\boldsymbol{\eta}(t)) \\ \boldsymbol{\eta}(t+1) = F_d(\boldsymbol{\eta}(t)) \end{cases} \tag{7}$$

With the above preparation, the performance index functional is presented as follows:

$$H(\boldsymbol{e}(t), \boldsymbol{v}(t)) = \sum_{i=t}^{\infty} \{M(\boldsymbol{e}(i)) + N(\boldsymbol{v}(i))\} \tag{8}$$

where $U(\boldsymbol{e}(i), \boldsymbol{v}(i)) = M(\boldsymbol{e}(i)) + N(\boldsymbol{v}(i))$ is positive definite.

For convenience, the optimal performance index functional is defined as $H^*(\boldsymbol{e}(t))$, which is expressed as

$$H^*(\boldsymbol{e}(t)) = \inf_{\boldsymbol{v}(t)} \{H(\boldsymbol{e}(t), \boldsymbol{v}(t)), \boldsymbol{v}(t) \in \Omega_v\} \tag{9}$$

where Ω_u is the set of admissible controls.

Definition 1. For the expression (9), the control policy $\boldsymbol{v}(t)$ is regarded as to be admissible, if $\boldsymbol{v}(t)$ is continuous, $H(\boldsymbol{e}(t), \boldsymbol{v}(t))$ is finite, for $\forall\, \boldsymbol{e}(t)$, and $\boldsymbol{v}(t) = \mathbf{0}$ as $\boldsymbol{e}(t) = \mathbf{0}$.

On the basis of Bellman′s optimality principle, for the optimal performance index functional $H^*(\boldsymbol{e}(t))$, the HJB equation should be satisfied

$$H^*(\boldsymbol{e}(t)) = \inf_{\boldsymbol{v}(t)} \{M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^*(\boldsymbol{e}(t+1))\} \tag{10}$$

which means that the optimal control policy $\boldsymbol{v}^*(t)$ is obtained by

$$\boldsymbol{v}^*(t) = \arg \inf_{\boldsymbol{v}(t)} \{M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^*(\boldsymbol{e}(t+1))\} \tag{11}$$

Therefore, it is clear that if the optimal performance index functional $H^*(\boldsymbol{e}(t))$ is acquired by (10), then the optimal tracking control problem of this paper is solved. However, there are three difficulties need to be overcome.

1) The dynamics of system (1) are unknown. And it leads to the dynamics of $\boldsymbol{e}(t+1)$ in (10) are also unknown. So the data-based identifier should be established to reconstruct $\boldsymbol{e}(t+1)$.

2) The expression of $\boldsymbol{u}_s(t)$ in (4) is undefined, which means that the expression of $\boldsymbol{v}(t)$ is unknown. So $\boldsymbol{v}(t)$ in (10) should be defined.

3) The input $\boldsymbol{u}(t)$ in system (1) is saturated, that means $\boldsymbol{v}(t)$ is also saturated, so $N(\boldsymbol{v}(t))$ in the HJB equation (10) should be defined reasonably.

In the following part, the difficulties will be solved one by one.

2 Efficient optimal solution based on ADP

This section contains four subsections. The unknown nonlinear system is identified by a neural network identification scheme with stability proof in the first subsection. The M network is given to obtain $\boldsymbol{u}_s(t)$ in the second subsection. In the third subsection, according to the iterative ADP algorithm, the optimal tracking control scheme is proposed. In the fourth subsection, the corresponding convergence theorem is developed.

2.1 Data-based identifier of unknown system dynamics

In this subsection, the data-based identifier is established with convergence proof. The three-layer neural networks are used as the identifier. In the neural networks, l is the hidden layer neurons number, the constant matrices w^{in} and w^{out} are the weight matrices relating the input layer and hidden layer, the hidden layer and output layer, respectively. Here w^{in} is assumed to be constant, so w^{out} is the only one to be tuned. Thus, we have the identification method as follows:

$$\hat{\boldsymbol{e}}(t+1) = (w^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) \tag{12}$$

where $\hat{\boldsymbol{e}}(t+1)$ is the estimate of $\boldsymbol{e}(t+1)$, $\boldsymbol{z}(t) = (w^{\text{in}})^{\text{T}}[\boldsymbol{e}^{\text{T}}(t) \ \ \boldsymbol{v}^{\text{T}}(t)]^{\text{T}}$. So we can obtain

$$\boldsymbol{e}(t+1) = (w^{\text{out}})^{*\text{T}}\psi(\boldsymbol{z}(t)) + \boldsymbol{\tau}(t) \tag{13}$$

where the neural networks function approximation error $\boldsymbol{\tau}(t)$ is bounded, $(w^{\text{out}})^{*\text{T}}$ is the ideal weight matrix, $\psi(\boldsymbol{z}(t)) \in \mathbf{R}^l$ is the bounded activation function, and $\|\psi(\boldsymbol{z})\| \le \bar{\psi}$.

The system identification error is denoted as follows:

$$\tilde{\boldsymbol{e}}(t+1) = \hat{\boldsymbol{e}}(t+1) - \boldsymbol{e}(t+1) \tag{14}$$

and the weight matrix error is denoted by

$$\tilde{w}^{\text{out}}(t) = w^{\text{out}}(t) - (w^{\text{out}})^* \tag{15}$$

Then we have

$$\tilde{\boldsymbol{e}}(t+1) = (\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) - \boldsymbol{\tau}(t) \tag{16}$$

Define

$$E(t+1) = \frac{1}{2}\tilde{\boldsymbol{e}}^{\text{T}}(t+1)\tilde{\boldsymbol{e}}(t+1) \tag{17}$$

Then we have

$$w^{\text{out}}(t+1) = w^{\text{out}}(t) - \alpha_m\psi(\boldsymbol{z}(t))\tilde{\boldsymbol{e}}^{\text{T}}(t+1) \tag{18}$$

where the learning rate α_m is a positive number, and $\alpha_m \le \dfrac{\alpha^2}{2\bar{\psi}^2}$, in which α is an adjustable parameter. The weight error is defined as follows:

$$\tilde{w}^{\text{out}}(t+1) = \tilde{w}^{\text{out}}(t) - \alpha_m\psi(\boldsymbol{z}(t))\tilde{\boldsymbol{e}}^{\text{T}}(t+1) \tag{19}$$

For the convergence proof of the identification error, the following assumption and lemma are necessary[25].

Assumption 2. It is assumed that the term $\boldsymbol{\tau}(t)$ satisfies

$$\boldsymbol{\tau}^{\text{T}}(t)\boldsymbol{\tau}(t) \le \lambda\tilde{\boldsymbol{e}}^{\text{T}}(t)\tilde{\boldsymbol{e}}(t) \tag{20}$$

where λ is constant.

Lemma 1[26]. The vector D_i, $i = 1, \cdots, n$, satisfies

$$(D_1 + \cdots + D_n)^{\text{T}}(D_1 + \cdots + D_n) \le n(D_1^{\text{T}}D_1 + \cdots + D_n^{\text{T}}D_n) \tag{21}$$

In the following part, the stability analysis will be given.

Theorem 1. If the weights update law is as in (18), the learning rate α_m satisfies $\alpha_m \le \dfrac{\alpha^2}{2\bar{\psi}^2}$, and α satisfies

$$\max\left\{-\sqrt{\frac{1-\lambda}{\lambda}}, -1\right\} \le \alpha \le \min\left\{\sqrt{\frac{1-\lambda}{\lambda}}, 1\right\} \tag{22}$$

where $0 < \lambda < 1$.

Then asymptotic stability stable of the identification error $\tilde{\boldsymbol{e}}(t+1)$ in (16) is guaranteed, and the boundedness of $\tilde{w}^{\text{out}}(t)$ in (15) is also obtained.

Proof. The Lyapunov functional is as follows:

$$V(t) = V_1(t) + V_2(t) \tag{23}$$

where

$$V_1(t) = \tilde{\boldsymbol{e}}^{\text{T}}(t)\tilde{\boldsymbol{e}}(t) \tag{24}$$

and

$$V_2(t) = \alpha_m^{-1}\text{tr}\{(\tilde{w}^{\text{out}}(t))^{\text{T}}\tilde{w}^{\text{out}}(t)\} \tag{25}$$

Then the first difference of the Lyapunov functional candidate is computed as follows:

$$\begin{aligned}\Delta V_1(t) =& \\ &\tilde{\boldsymbol{e}}^{\text{T}}(t+1)\tilde{\boldsymbol{e}}(t+1) - \tilde{\boldsymbol{e}}^{\text{T}}(t)\tilde{\boldsymbol{e}}(t) = \\ &((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)))^{\text{T}} \times \\ &(\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) - 2((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)))^{\text{T}}\boldsymbol{\tau}(t) + \\ &\boldsymbol{\tau}^{\text{T}}(t)\boldsymbol{\tau}(t) - \tilde{\boldsymbol{e}}^{\text{T}}(t)\tilde{\boldsymbol{e}}(t)\end{aligned} \tag{26}$$

and

$$\begin{aligned}\Delta V_2(t) =& \\ &\alpha_m^{-1}\text{tr}\{(w^{\text{out}}(t+1) - (w^{\text{out}})^*)^{\text{T}}(w^{\text{out}}(t+1) - (w^{\text{out}})^*)\} - \\ &\alpha_m^{-1}\text{tr}\{(\tilde{w}^{\text{out}}(t))^{\text{T}}\tilde{w}^{\text{out}}(t)\}\end{aligned} \tag{27}$$

By the derivation, we have

$$\begin{aligned}\Delta V_2(t) =& \\ &\alpha_m^{-1}\text{tr}\{(w^{\text{out}}(t) - \alpha_m\psi(z(t))\tilde{\boldsymbol{e}}^{\text{T}}(t+1) - (w^{\text{out}})^*)^{\text{T}} \times \\ &(w^{\text{out}}(t) - \alpha_m\psi(\boldsymbol{z}(t))\tilde{\boldsymbol{e}}^{\text{T}}(t+1) - (w^{\text{out}})^*)\} - \\ &\alpha_m^{-1}\text{tr}\{(\tilde{w}^{\text{out}}(t))^{\text{T}}\tilde{w}^{\text{out}}(t)\} = \\ &\alpha_m\psi^{\text{T}}(\boldsymbol{z}(t))\psi(\boldsymbol{z}(t))((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) - \boldsymbol{\tau}(t))^{\text{T}} \times \\ &((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) - \boldsymbol{\tau}(t)) - \\ &2\psi^{\text{T}}(\boldsymbol{z}(t))\tilde{w}^{\text{out}}(t)((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) - \boldsymbol{\tau}(t))\end{aligned} \tag{28}$$

According to Lemma 1, we have

$$\begin{aligned}\Delta V_2(t) \le& \\ &2\alpha_m\psi^{\text{T}}(\boldsymbol{z}(t))\psi(\boldsymbol{z}(t))(((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)))^{\text{T}} \times \\ &(\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) + \boldsymbol{\tau}^{\text{T}}(t)\boldsymbol{\tau}(t)) - \\ &2\psi^{\text{T}}(\boldsymbol{z}(t))\tilde{w}^{\text{out}}(t)((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) - \boldsymbol{\tau}(t))\end{aligned} \tag{29}$$

So the following expression

$$\begin{aligned}\Delta V(t) =& \\ &\Delta V_1(t) + \Delta V_2(t) \le \\ &-((\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)))^{\text{T}}(\tilde{w}^{\text{out}}(t))^{\text{T}}\psi(\boldsymbol{z}(t)) +\end{aligned}$$

$$\boldsymbol{\tau}^{\mathrm{T}}(t)\boldsymbol{\tau}(t) - \tilde{\boldsymbol{e}}^{\mathrm{T}}(t)\tilde{\boldsymbol{e}}(t) +$$
$$2\alpha_m\psi^{\mathrm{T}}(\boldsymbol{z}(t))\psi(\boldsymbol{z}(t))(((\tilde{w}^{\mathrm{out}}(t))^{\mathrm{T}}\psi(\boldsymbol{z}(t)))^{\mathrm{T}} \times$$
$$(\tilde{w}^{\mathrm{out}}(t))^{\mathrm{T}}\psi(\boldsymbol{z}(t)) + \boldsymbol{\tau}^{\mathrm{T}}(t)\boldsymbol{\tau}(t)) \leq$$
$$(-1+\alpha^2)((\tilde{w}^{\mathrm{out}}(t))^{\mathrm{T}}\psi(\boldsymbol{z}(t)))^{\mathrm{T}}(\tilde{w}^{\mathrm{out}}(t))^{\mathrm{T}}\psi(\boldsymbol{z}(t)) +$$
$$(-1+\lambda+\alpha^2\lambda)\tilde{\boldsymbol{e}}^{\mathrm{T}}(t)\tilde{\boldsymbol{e}}(t) \tag{30}$$

holds.

As α satisfies (22), so

$$\alpha^2 \leq 1 \tag{31}$$

and

$$\alpha^2 \leq \frac{1-\lambda}{\lambda} \tag{32}$$

Thus we obtain

$$\Delta V(t) \leq 0 \tag{33}$$

It is concluded that the identification error $\tilde{\boldsymbol{e}}(t+1)$ satisfies stability in the sense of Lyapunov, and the weights estimation error $\tilde{w}^{\mathrm{out}}$ is bounded. □

Remark 2. From the analyses above, it can be concluded that the data-based identifier in this paper is a development of traditional neural network model. First, the input/output data are collected into the database. Then, the data in database are used to train the identifier. Furthermore, the weight update law is designed to make the identifier asymptotically stable, and the comprehensive theoretical analysis is given.

2.2 The M network

Based on Assumption 1, it is clear that $\boldsymbol{u}_s(t)$ is existent. So according to the input/output data $(\boldsymbol{\eta}(t), \boldsymbol{u}_s(t), \boldsymbol{\eta}(t+1))$ of the actual system, we will introduce the M network to be as the approximator of $G_d(\cdot)$. So we define

$$\hat{\boldsymbol{u}}_s(t) = (w^M(t))^{\mathrm{T}}\psi(z^{\boldsymbol{\eta}}(t)) \tag{34}$$

where $z^{\boldsymbol{\eta}}(t) = [\boldsymbol{\eta}^{\mathrm{T}}(t), \boldsymbol{\eta}^{\mathrm{T}}(t+1)]^{\mathrm{T}}$. We define the error function of the model network as

$$\tilde{\boldsymbol{u}}_s(t) = \hat{\boldsymbol{u}}_s(t) - \boldsymbol{u}_s(t) \tag{35}$$

Define the error measure as:

$$E^M(t) = \frac{1}{2}\tilde{\boldsymbol{u}}_s^{\mathrm{T}}(t)\tilde{\boldsymbol{u}}_s(t) \tag{36}$$

The M network is updated as follows:

$$w^M(t+1) = w^M(t) + \Delta w^M(t) =$$
$$w^M(t) - \alpha^M\frac{\partial E^M(t)}{\partial w^M(t)} \tag{37}$$

where α^M is the learning rate. The weights update rule is similar as (18), so if the appropriate learning rate is selected, then $\hat{\boldsymbol{u}}_s(t)$ asymptotically converges to $\boldsymbol{u}_s(t)$.

2.3 Derivation of the optimal tracking control scheme

With the above preparation, based on ADP algorithm, a novel optimal tracking control method will be proposed in this subsection. In HJB equation (10), the expressions of $M(\boldsymbol{e}(t))$ and $N(\boldsymbol{v}(t))$ are not explicit, so we firstly define

$$M(\boldsymbol{e}(t)) = \boldsymbol{e}^{\mathrm{T}}(t)Q_1\boldsymbol{e}(t) \tag{38}$$

and

$$N(\boldsymbol{v}(t)) = 2\int_0^{\boldsymbol{v}(t)} (\phi^{-1}(s))^{\mathrm{T}}R\mathrm{d}s \tag{39}$$

where $\phi(\cdot) \in C^p$ is the one-to-one bounded function, $p \geq 1$ and $L_2(\Omega)$. $\phi^{-1}(\boldsymbol{v}(t)) = [\varphi^{-1}(\boldsymbol{v}_1(t)), \cdots, \varphi^{-1}(\boldsymbol{v}_m(t))]^{\mathrm{T}}$. The boundedness of the first derivative is $\bar{M}$, and it is an increasing monotonic function. $\varphi(\cdot)$ satisfies $\beta^{\min} \leq \varphi(\cdot) \leq \beta^{\max}$.

Remark 3. In [5], the saturating upper and lower bound norms for the actuators are the same, so the hyperbolic tangent is selected as the activation function. While in this paper, the lower bound norm $|\beta^{\min}|$ and the upper bound norm $|\beta^{\max}|$ are not necessarily equivalent, so the odd function tanh(·) cannot very well solve the problem. Thus a new function $\varphi(\boldsymbol{x}) = \dfrac{\mathrm{e}^{\boldsymbol{x}} - \mathrm{e}^{-\boldsymbol{x}}}{\frac{1}{\beta^{\max}}\mathrm{e}^{\boldsymbol{x}} + (-\frac{1}{\beta^{\min}})\mathrm{e}^{-\boldsymbol{x}}}$ is used in this paper. If $\beta^{\min} = -5$ and $\beta^{\max} = 2$, then the figure of $\varphi(\boldsymbol{x})$ is as in Fig. 1.

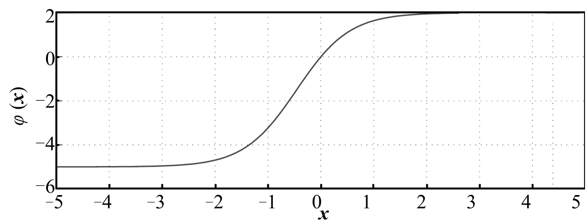

Fig. 1 The trajectory of $\varphi(\boldsymbol{x})$

In this paper, we propose an iterative ADP algorithm.

The iterative ADP algorithm starts with the initial iterative performance index functional $H^{[0]}(\cdot) = 0$. Then for $i = 0, 1, \cdots$, we define the iterative control vector $\boldsymbol{v}^{[i]}(t)$ as follow:

$$\boldsymbol{v}^{[i]}(t) = \arg\inf_{\boldsymbol{v}(t)}\left\{M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^{[i]}(\boldsymbol{e}(t+1))\right\} \tag{40}$$

and

$$H^{[i+1]}(\boldsymbol{e}(t)) = \inf_{\boldsymbol{v}(t)}\left\{M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^{[i]}(\boldsymbol{e}(t+1))\right\} \tag{41}$$

where $N(\boldsymbol{v}(t))$ is defined as in (39) and the saturated bounds are $\boldsymbol{\beta}_j^{\min} \leq \boldsymbol{v}_j \leq \boldsymbol{\beta}_j^{\max}$, in which $\boldsymbol{\beta}_j^{\min} = \boldsymbol{u}_j^{\min} - \boldsymbol{u}_{sj}$ and $\boldsymbol{\beta}_j^{\max} = \boldsymbol{u}_j^{\max} - \boldsymbol{u}_{sj}$, $j = 1, \cdots, m$.

Then, from the equation (40), we have $\boldsymbol{v}^{[i]}(t)$ as

$$\boldsymbol{v}^{[i]}(t) = \phi\left(-\frac{1}{2}R^{-1}\left(\frac{\partial\boldsymbol{e}(t+1)}{\partial\boldsymbol{v}^{[i]}(t)}\right)^{\mathrm{T}}\frac{\partial H^{[i]}(\boldsymbol{e}(t+1))}{\partial\boldsymbol{e}(t+1)}\right) \tag{42}$$

and

$$H^{[i+1]}(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{[i]}(t)) + H^{[i]}(\boldsymbol{e}(t+1)) \tag{43}$$

2.4 Convergence analysis of the iterative ADP algorithm

This part is about the convergence and stabilization analyses.

Lemma 2. Let $\{\boldsymbol{\varsigma}^{[i]}(t)\}$ be any sequence of control policies, and $\{\boldsymbol{v}^{[i]}(t)\}$ be the policies as expressed in (40). Let $H^{[i+1]}$ be as in (41) and $\Lambda^{[i]}$ is updated as

$$\Lambda^{[i+1]}(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{\varsigma}^{[i]}(t)) + \Lambda^{i}(\boldsymbol{e}(t+1)) \quad (44)$$

If $H^{[0]}(\cdot) = \Lambda^{[0]}(\cdot) = 0$, then $H^{[i+1]}(\boldsymbol{e}(t)) \le \Lambda^{[i+1]}(\boldsymbol{e}(t))$, $\forall\, i$.

Proof. The fact can be obtained from (40). As $H^{[i+1]}(\boldsymbol{e}(t))$ is achieved by $\boldsymbol{v}^{i}(t)$, while $\Lambda^{[i+1]}(\boldsymbol{e}(t))$ is obtained by any control input. □

Theorem 2. Let $H^{[i+1]}(\boldsymbol{e}(t))$ be defined as in (41), then $0 \le H^{[i+1]}(\boldsymbol{e}(t)) \le Y$, $\forall\, i$ holds, where Y is the bound.

Proof. Let $\{\boldsymbol{\gamma}^{[i]}(t)\}$ be any admissible control, $P^{[0]}(\cdot) = 0$ and $P^{[i+1]}(\boldsymbol{e}(t))$ be

$$P^{[i+1]}(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{\gamma}^{[i]}(t)) + P^{[i]}(\boldsymbol{e}(t+1)) \quad (45)$$

From (45), we obtain

$$P^{[i]}(\boldsymbol{e}(t+1)) = M(\boldsymbol{e}(t+1)) + N(\boldsymbol{\gamma}^{[i-1]}(t+1)) + P^{[i-1]}(e(t+2)) \quad (46)$$

So we have

$$\begin{aligned} P^{[i+1]}(\boldsymbol{e}(t)) = \\ & M(\boldsymbol{e}(t)) + N(\boldsymbol{\gamma}^{[i]}(t)) + M(\boldsymbol{e}(t+1)) + \\ & N(\boldsymbol{\gamma}^{[i-1]}(t+1)) + \cdots + \\ & M(\boldsymbol{e}(t+i)) + N(\boldsymbol{\gamma}^{[0]}(t+i)) = \\ & \sum_{j=0}^{i}\{M(\boldsymbol{e}(t+j)) + N(\boldsymbol{\gamma}^{[i-j]}(t+j))\} \end{aligned} \quad (47)$$

Note that $\{\boldsymbol{\gamma}^{[i]}(t)\}$ is admissible, so

$$P^{[i+1]}(\boldsymbol{e}(t)) \le \lim_{i\to\infty}\sum_{j=0}^{i}\{M(\boldsymbol{e}(t+j)) + N(\boldsymbol{\gamma}^{[i-j]}(t+j))\} \le Y, \quad \forall\, i \quad (48)$$

holds.

Thus we can obtain

$$0 \le H^{[i+1]}(\boldsymbol{e}(t)) \le P^{[i+1]}(\boldsymbol{e}(t)) \le Y, \quad \forall\, i \quad (49)$$

□

Theorem 3. Let $\{H^{[i+1]}(\boldsymbol{e}(t))\}$ be as in (41) and $H^{[0]}(\cdot) = 0$. Then we have that $\{H^{[i+1]}(\boldsymbol{e}(t))\}$ satisfies $H^{[i+1]}(\boldsymbol{e}(t)) \ge H^{[i]}(\boldsymbol{e}(t))$, $\forall\, i$.

Proof. Define $\{\Phi^{[i]}(\boldsymbol{e}(t))\}$ as

$$\Phi^{[i]}(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{[i]}(t)) + \Phi^{[i-1]}(\boldsymbol{e}(t+1)) \quad (50)$$

with $\Phi^{[0]}(\cdot) = 0$.

Then, we will prove that $\Phi^{[i]}(\boldsymbol{e}(t)) \le H^{[i+1]}(\boldsymbol{e}(t))$.

First, for $i = 0$, we have

$$H^{[1]}(\boldsymbol{e}(t)) - \Phi^{[0]}(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{[0]}(t)) \ge 0 \quad (51)$$

So for $i = 0$, we have

$$H^{[1]}(\boldsymbol{e}(t)) \ge \Phi^{[0]}(\boldsymbol{e}(t)) \quad (52)$$

Second, suppose that $H^{[i]}(\boldsymbol{e}(t)) \ge \Phi^{[i-1]}(\boldsymbol{e}(t))$, for $i-1$, $\forall\, \boldsymbol{e}(t)$. As

$$H^{[i+1]}(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{[i]}(t)) + H^{[i]}(\boldsymbol{e}(t+1)) \quad (53)$$

and

$$\Phi^{[i]}(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{[i]}(t)) + \Phi^{[i-1]}(\boldsymbol{e}(t+1)) \quad (54)$$

so we can get

$$H^{[i+1]}(\boldsymbol{e}(t)) - \Phi^{[i]}(\boldsymbol{x}(t)) = H^{[i]}(\boldsymbol{e}(t+1)) - \Phi^{[i-1]}(\boldsymbol{e}(t+1)) \ge 0 \quad (55)$$

Therefore, we have

$$H^{[i+1]}(\boldsymbol{e}(t)) \ge \Phi^{[i]}(\boldsymbol{e}(t)), \ \ \forall\, i \quad (56)$$

From Lemma 2 we have $H^{[i]}(\boldsymbol{e}(t)) \le \Phi^{[i]}(\boldsymbol{e}(t))$. So

$$H^{[i]}(\boldsymbol{e}(t)) \le \Phi^{[i]}(\boldsymbol{e}(t)) \le H^{[i+1]}(\boldsymbol{e}(t)) \quad (57)$$

holds, and the proof is completed. □

Hence $\{H^{[i]}(\boldsymbol{e}(t))\}$ is convergent, for $i \to \infty$.

In this paper, we can define $\lim_{i\to\infty} H^{[i]}(\boldsymbol{e}(t)) = H^{*}(\boldsymbol{e}(t))$, accordingly $\lim_{i\to\infty} \boldsymbol{v}^{[i]}(t) = \boldsymbol{v}^{*}(t)$, and $\boldsymbol{u}^{*}(t) = \boldsymbol{v}^{*}(t) + \boldsymbol{u}_s(t)$.

Theorem 4. The limit $H^{*}(\boldsymbol{e}(t))$, $\forall\, t$, satisfies

$$H^{*}(\boldsymbol{e}(t)) = \inf_{\boldsymbol{v}(t)}\{\, M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^{*}(\boldsymbol{e}(t+1))\} \quad (58)$$

Proof. For any i, the performance index functional sequence satisfies

$$H^{[i+1]}(\boldsymbol{e}(t)) = \inf_{\boldsymbol{v}(t)}\left\{M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^{[i]}(\boldsymbol{e}(t+1))\right\} \quad (59)$$

Combining with $H^{[i+1]}(\boldsymbol{e}(t)) \le \lim_{i\to\infty} H^{[i+1]}(\boldsymbol{e}(t))$, $\forall\, i$, and $\lim_{i\to\infty} H^{[i+1]}(\boldsymbol{e}(t)) = H^{*}(\boldsymbol{e}(t))$, we can obtain

$$H^{*}(\boldsymbol{e}(t)) \ge \inf_{\boldsymbol{v}(t)}\left\{M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^{[i]}(\boldsymbol{e}(t+1))\right\} = M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{[i]}(t)) + H^{[i]}(\boldsymbol{e}(t+1)) \quad (60)$$

Let $i \to \infty$, we have

$$H^{*}(\boldsymbol{e}(t)) \ge M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{*}(t)) + H^{*}(\boldsymbol{x}(t+1)) \quad (61)$$

On the other hand, for $\forall\, i$ and any admissible control input sequence $\boldsymbol{\gamma}^{[i]}(t)$, we obtain

$$H^{[i+1]}(\boldsymbol{e}(t)) \le M(\boldsymbol{e}(t)) + N(\boldsymbol{\gamma}^{[i]}(t)) + H^{[i]}(\boldsymbol{e}(t+1)) \quad (62)$$

As $H^{[i]}(\boldsymbol{e}(t)) \le H^{[i+1]}(\boldsymbol{e}(t))$, we can obtain

$$H^{[i]}(\boldsymbol{e}(t)) \le M(\boldsymbol{e}(t)) + N(\boldsymbol{\gamma}^{[i]}(t)) + H^{[i]}(\boldsymbol{e}(t+1)) \quad (63)$$

Since $\boldsymbol{\gamma}^{[i]}(t)$ is selected arbitrarily, so

$$H^{[i]}(\boldsymbol{e}(t)) \le M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{[i]}(t)) + H^{[i]}(\boldsymbol{e}(t+1)) \quad (64)$$

holds.

Let $i \to \infty$, then

$$H^{*}(\boldsymbol{e}(t)) \le M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^{*}(t)) + H^{*}(\boldsymbol{e}(t+1)) \quad (65)$$

So, we have

$$H^*(\boldsymbol{e}(t)) = M(\boldsymbol{e}(t)) + N(\boldsymbol{v}^*(t)) + H^*(\boldsymbol{e}(t+1)) = \inf_{\boldsymbol{v}(t)} \{ M(\boldsymbol{e}(t)) + N(\boldsymbol{v}(t)) + H^*(\boldsymbol{e}(t+1))\} \tag{66}$$

□

Thus, $H^{[i]}(\boldsymbol{e}(t)) \to H^*(\boldsymbol{e}(t))$ as $i \to \infty$. Simultaneously, we have $\boldsymbol{v}^{[i]}(t) \to \boldsymbol{v}^*(t)$ as $i \to \infty$. Therefore, the optimal tracking control of system (1) is $\boldsymbol{u}^*(t) = \boldsymbol{v}^*(t) + \boldsymbol{u}_s(t)$.

In the next section, the detailed implemention of the optimal tracking control method will be given.

3 Implementation of the iterative ADP algorithm

In this paper, the neural networks are used is as the function approximator. The neural network implementation process is shown in Fig. 2.

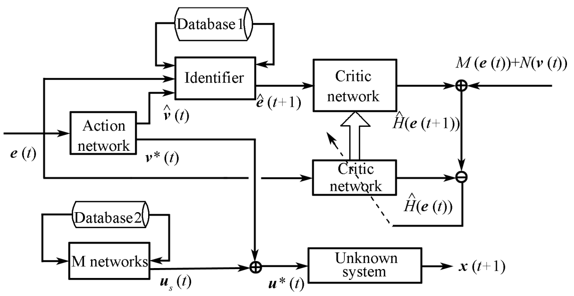

Fig. 2 The structural diagram of the networks

In Fig. 2, the data-based identifier module is introduced in Subsection 2.1, which is used to build the system dynamics. The M network module is given in Subsection 2.2, which is used to obtain $\boldsymbol{u}_s(t)$. The action network is used to approximate the control error. The two critic networks are used to approximate $H(\boldsymbol{e}(t))$ and $H(\boldsymbol{e}(t+1))$, respectively. The update methods of action network and critic network modules will be shown in the following part.

3.1 The critic network

The approximator of $H^{[i+1]}(\boldsymbol{e}(t))$ is the critic network output, which is expressed as

$$\hat{H}^{[i+1]}(\boldsymbol{e}(t)) = (W_c^{[i+1]})^{\mathrm{T}}\psi(\boldsymbol{e}(t)) \tag{67}$$

The error of the critic network is

$$ec^{[i]}(t) = \hat{H}^{[i+1]}(\boldsymbol{e}(t)) - H^{[i+1]}(\boldsymbol{e}(t)) \tag{68}$$

Define

$$Ec^{[i]}(t) = \frac{1}{2}(ec^{[i]}(t))^2 \tag{69}$$

So, we have

$$wc^{[i+1]}(t) = wc^{[i]}(t) + \Delta wc^{[i]}(t) \tag{70}$$

where

$$\Delta wc^{[i]}(t) = \alpha_c \left[-\frac{\partial Ec^{[i]}(t)}{\partial wc^{[i]}(t)}\right] \tag{71}$$

and

$$\frac{\partial Ec^{[i]}(t)}{\partial wc^{[i]}(t)} = \frac{\partial Ec^{[i]}(t)}{\partial \hat{H}^{[i]}(\boldsymbol{e}(t))}\frac{\partial \hat{H}^{[i]}(\boldsymbol{e}(t))}{\partial wc^{[i]}(t)} \tag{72}$$

$wc^{[i]}(t)$ is the weight vector and the learning rate $\alpha_c > 0$.

3.2 The action network

The approximator of $\boldsymbol{v}(t)$ is the action network, which is defined as

$$\hat{\boldsymbol{v}}^{[i]}(t) = (W_a^{[i]})^{\mathrm{T}}\psi(\boldsymbol{e}(t)) \tag{73}$$

The error is

$$ea^{[i]}(t) = \hat{\boldsymbol{v}}^{[i]}(t) - \boldsymbol{v}^{[i]}(t) \tag{74}$$

So, we obtain

$$Ea^{[i]}(t) = \frac{1}{2}(ea^{[i]}(t))^2 \tag{75}$$

where

$$wa^{[i+1]}(t) = wa^{[i]}(t) + \Delta wa^{[i]}(t) \tag{76}$$

in which

$$\Delta wa^{[i]}(t) = \alpha_a \left[-\frac{\partial Ea^{[i]}(t)}{\partial wa^{[i]}(t)}\right] \tag{77}$$

and

$$\frac{\partial Ea^{[i]}(t)}{\partial wa^{[i]}(t)} = \frac{\partial Ea^{[i]}(t)}{\partial ea^{[i]}(t)}\frac{\partial ea^{[i]}(t)}{\partial \boldsymbol{v}^{[i]}(t)}\frac{\partial \boldsymbol{v}^{[i]}(t)}{\partial wa^{[i]}(t)} \tag{78}$$

and $\alpha_a > 0$.

3.3 The procedure of the ADP algorithm

In this subsection, the iterative algorithm for optimal tracking control problem is summarized as:

Step 1. Give the maximal iterative step $i_{\max}$ and the initial state $\boldsymbol{x}(0)$. Give the iterative accuracy ϱ and the desired trajectory $\boldsymbol{\eta}(t)$.

Step 2. Practically collect large offline data ($\boldsymbol{e}(t), \boldsymbol{v}(t), \boldsymbol{e}(t+1)$) to train the identifier and use data ($\boldsymbol{\eta}(t), \boldsymbol{u}_s(t), \boldsymbol{\eta}(t+1)$) to train the M network.

Step 3. From the iterative index $i = 0$ compute $H^{[i+1]}(\boldsymbol{e}(t))$ and $\boldsymbol{v}^{[i]}(t)$.

Step 4. If $|H^{[i+1]}(\boldsymbol{e}(t)) - H^{[i]}(\boldsymbol{e}(t))| \le \varrho$, then stop the iterative algorithm.

Step 5. If $i > i_{\max}$, then stop the iterative algorithm.

Step 6. Using $\boldsymbol{u}^*(t)$ obtain the optimal trajectory of the unknown system (1).

Step 7. Stop.

4 Simulation study

We consider the discrete-time nonlinear system as follows[22]:

$$\boldsymbol{x}(t+1) = f(\boldsymbol{x}(t)) + g(\boldsymbol{x}(t))\boldsymbol{u}(t) \tag{79}$$

where $-0.7 \le \boldsymbol{u}_i(t) \le 0.7$, $i = 1, 2$.

$$f(\boldsymbol{x}(t)) = \begin{bmatrix} 0.2\boldsymbol{x}_1(t)\exp(\boldsymbol{x}_2(t))^2\boldsymbol{x}_2(t) \\ 0.3(\boldsymbol{x}_2(t))^2\boldsymbol{x}_1(t) \end{bmatrix}$$

and

$$g(\boldsymbol{x}(t)) = \begin{bmatrix} -0.2 & 0 \\ 0 & -0.2 \end{bmatrix}$$

The desired trajectory $\boldsymbol{\eta}(t)$ is generated by the following expression

$$\boldsymbol{\eta}(t+1) = A\boldsymbol{\eta}(t) \tag{80}$$

with

$$A = \begin{bmatrix} \cos wT & \sin wT \\ -\sin wT & \cos wT \end{bmatrix}, \quad \boldsymbol{\eta}(0) = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$$

where $T = 0.1\,\mathrm{s}$, $w = 0.8\pi$.

For the unknown system (79) and desired orbit (80), the data-based identifier of $e(t+1)$ is firstly established according to Subsection 2.1. The maximal iterative step $i_{\max}$ is 100. The data-based identifier structure is 4-8-2. From $[-0.15,\ 0.15]$, the initial weights are selected, $\alpha_m = 0.02$. Then the identification error trajectories are obtained as in Fig. 3. We can see that the identification errors converge to zero asymptotically. For obtaining the steady control (4), the M network presented in Subsection 2.2 is used. For M network the learning rate is $w^M = 0.2$, the trajectories of the steady control $\boldsymbol{u}_s$ are shown in Fig. 4. Based on the above preparation works, the optimal ADP tracking algorithm is going to be implemented. The three-layer BP neural networks are used to approximate the critic network and the action network, respectively. After 3 000 iterative steps, the performance index functional is monotone and increasing as in Fig. 5. Finally it converges to the optimal one.

The optimal tracking state trajectories of system (79) are also achieved after 50 time steps. The dotted lines in Fig. 6 and Fig. 7 are the states of system (79). The solid lines in Fig. 6 and Fig. 7 are the states of system (80). Obviously, the states of system (79) can fast track the desired trajectory (80). The tracking error trajectories are shown in Fig. 8.

The control error trajectories are shown in Fig. 9. It is clear that the tracking error is asymptotically stable and the new optimal tracking control policy for the unknown system is very feasible.

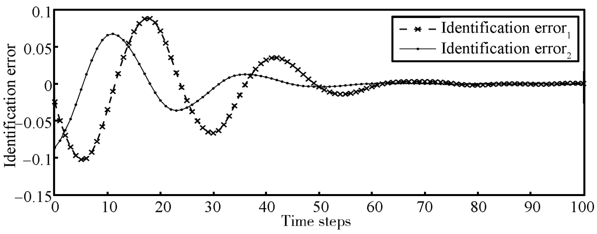

Fig. 3 The identification error trajectories

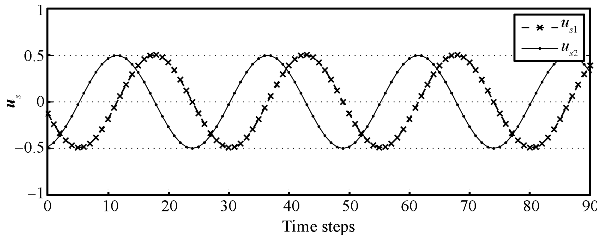

Fig. 4 The trajectories of $\boldsymbol{u}_s$

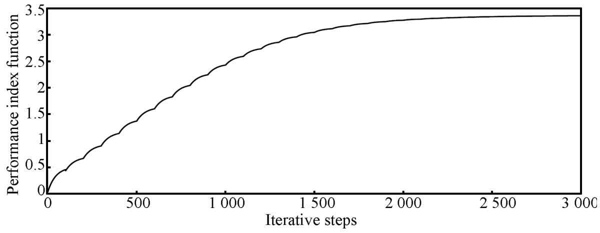

Fig. 5 The performance index function trajectories

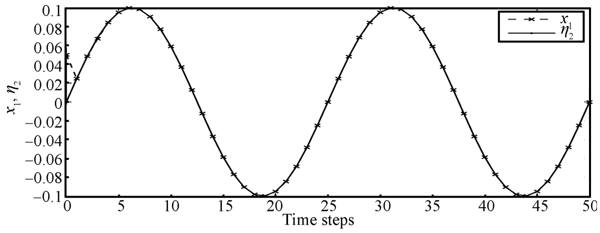

Fig. 6 The trajectories of x_1 and η_1

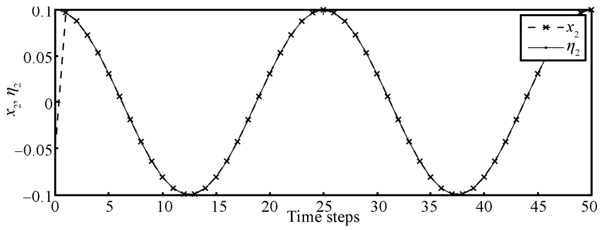

Fig. 7 The trajectories of x_2 and η_2

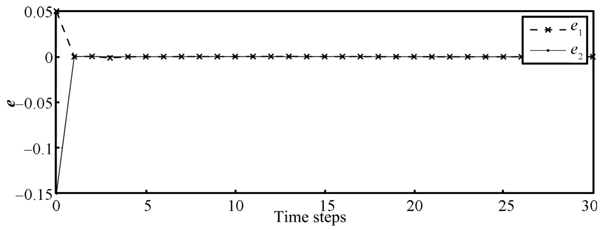

Fig. 8 The tracking error trajectories

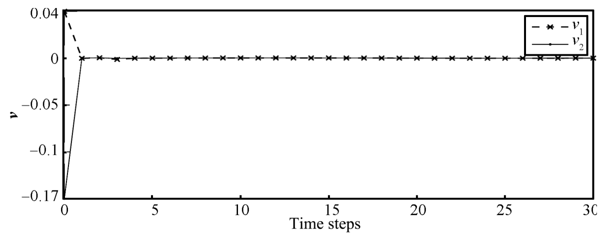

Fig. 9 The tracking control error trajectories

5 Conclusion

This paper proposed an effective optimal tracking control scheme for a class of unknown nonlinear systems with actuator saturation. First, a data-based identifier was constructed for the unknown nonlinear system. Then the M network was introduced to obtain the expression of the steady control. Next, a new nonquadratic performance functional was proposed, and a novel ADP algorithm was developed. The simulation works have shown the effectiveness of the proposed optimal tracking control method for the unknown systems with actuator saturation.

References

1 Zhang X, Zhang H G, Sun Q Y, Luo Y H. Adaptive dynamic programming-based optimal control of unknown nonaffine nonlinear discrete-time systems with proof of convergence. *Neurocomputing*, 2012, **91**: 48−55

2 Hou Z S, Wang Z. From model-based control to data-driven control: survey, classification and perspective. *Information Sciences*, 2013, **235**: 3−35

3 Shen D, Hou Z S. Iterative learning control with unknown

control direction: a novel data-based approach. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2237−2249

4 Lai C Y, Xiang C, Lee T H. Data-based identification and control of nonlinear systems via piecewise affine approximation. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2189−2200

5 Zhang H G, Luo Y H, Liu D R. Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. *IEEE Transactions on Neural Networks*, 2009, **20**(9): 1490−1503

6 Saberi A, Lin Z L, Teel A R. Control of linear systems with saturating actuators. *IEEE Transactions on Automatic Control*, 1996, **41**(3): 368−378

7 Bernstein D S. Optimal nonlinear, but continuous, feedback control of systems with saturating actuators. *International Journal of Control*, 1995, **62**(5): 1209−1216

8 Sussmann H J, Sontag E D, Yang Y. A general result on the stabilization of linear systems using bounded controls. *IEEE Transactions on Automatic Control*, 1994, **39**(12): 2411−2425

9 Bellman R E. *Dynamic Programming*. Princeton, NJ: Princeton University Press, 1957

10 Bryson A E, Ho Y C. *Applied Optimal Control: Optimization, Estimation, and Control*. New York, NY: Hemisphere-Wiley, 1975

11 Liu D R, Xiong X X, Zhang Y. Action-dependent adaptive critic designs. In: Proceedings of the 2001 International Joint Conference on Neural Networks. Washington, D.C.: IEEE, 2001, **2**: 990−995

12 Zhang H G, Song R Z, Wei Q L, Zhang T Y. Optimal tracking control for a class of nonlinear discrete-time systems with time delays based on heuristic dynamic programming. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 1851−1862

13 Wei Q L, Liu D R. An iterative ϵ-optimal control scheme for a class of discrete-time nonlinear systems with unfixed initial state. *Neural Networks*, 2012, **32**: 236−244

14 Song R Z, Zhang H G, Luo Y H, Wei Q L. Optimal control laws for time-delay systems with saturating actuators based on heuristic dynamic programming. *Neurocomputing*, 2010, **73**(16−18): 3020−3027

15 Song R Z, Zhang H G. The finite-horizon optimal control for a class of time-delay affine nonlinear system. *Neural Computing and Applications*, 2013, **22**(2): 229−235

16 Liu Z W, Zhang H G. Delay-dependent stability for systems with fast-varying neutral-type delays via a PTVD compensation. *Acta Automatica Sinica*, 2010, **36**(1): 147−152

17 Wei Q L, Zhang H G, Cui L L. Data-based optimal control for discrete-time zero-sum games of 2-D systems using adaptive critic designs. *Acta Automatica Sinica*, 2009, **35**(6): 682−692

18 Zhang H G, Wei Q L, Liu D R. An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games. *Automatica*, 2011, **47**(1): 207−214

19 Fu J, He H B, Zhou X M. Adaptive learning and control for MIMO system based on adaptive dynamic programming. *IEEE Transactions on Neural Networks*, 2011, **22**(7): 1133−1148

20 He H B, Ni Z, Fu J. A three-network architecture for on-line learning and optimization based on adaptive dynamic programming. *Neurocomputing*, 2012, **78**(1): 3−13

21 Wei Q L, Zhang H G, Liu D R, Zhao Y. An optimal control scheme for a class of discrete-time nonlinear systems with time delays using adaptive dynamic programming. *Acta Automatica Sinica*, 2010, **36**(1): 121−129

22 Zhang H G, Wei Q L, Luo Y H. A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, **38**(4): 937−942

23 Wei Q L, Zhang H G, Dai J. Model-free multiobjective approximate dynamic programming for discrete-time nonlinear systems with general performance index functions. *Neurocomputing*, 2009, **72**(7−9): 1839−1848

24 Wang D, Liu D R, Wei Q L, Zhao D B, Jin N. Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming. *Automatica*, 2012, **48**(8): 1825−1832

25 Zhang H G, Cui L L, Zhang X, Luo Y H. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2226−2236

26 Liu D R, Wang D, Zhao D B, Wei Q L, Jin N. Neural-network-based optimal control for a class of unknown discrete-time nonlinear systems using globalized dual heuristic programming. *IEEE Transactions on Automation Science and Engineering*, 2012, **9**(3): 628−634

SONG Rui-Zhuo Postdoctoral fellow at the School of Automation and Electrical Engineering, University of Science and Technology Beijing. She received her Ph. D. degree in control theory and control engineering from Northeastern University in 2012. Her research interest covers neural-networks-based control, nonlinear control, wireless sensor networks, adaptive dynamic programming and their industrial application. Corresponding author of this paper. E-mail: ruizhuosong@163.com

XIAO Wen-Dong Received his bachelor and Ph. D. degrees from Northeastern University in 1990 and 1995, respectively. Currently he is a professor at the School of Automation and Electrical Engineering, University of Science and Technology Beijing. Previously he held various academic and research positions with Northeastern University, China, POSCO Technical Research Laboratories, South Korea, Nanyang Technological University, Singapore, and the Institute for Infocomm Research, A*STAR, Singapore. His research interest covers information fusion, wireless localization and tracking, information-driven resource management, wireless sensor networks, and internet of things.
E-mail: wendongxiao68@163.com

SUN Chang-Yin Professor at the School of Automation and Electrical Engineering, University of Science and Technology Beijing. He received his master and Ph. D. degrees in electrical engineering from Southeast University in 2001 and 2003, respectively. His research interest covers intelligent control, flight control, pattern recognition, and optimal theory. E-mail: changyinsun@163.com