Optimal control laws for time-delay systems with saturating actuators based on heuristic dynamic programming

Ruizhao Song, Huaguang Zhang, Yanhong Luo, Qinglai Wei

1. Introduction

Saturation nonlinearity is unavoidable in most of the actuators. Due to the nonanalytic nature of the actuator nonlinear dynamics and the fact that the exact actuator nonlinear functions are unknown, such systems present a challenge to control engineers. So the control of systems with saturating actuators has been the focus of many researchers for many years. Several methods for deriving control laws considering the saturation phenomena can be found in [4,13]. On the other hand, time delays often occur in the transmission or material between different parts of systems [16]. So, the research on the optimal control for nonlinear time-delay systems with saturating actuators is beneficial to practical application.

It is well known that there are many methods for designing stable control for nonlinear systems [10,21,25]. However, stability is only a bare minimum requirement in a system design. Ensuring optimality guarantees the stability of the nonlinear systems. Dynamic programming is a very useful tool in solving optimization and optimal control problems by employing the principle of optimality. However, solving the associated Hamilton–Jacobi–Bellman (HJB) equation demands a large (rather infeasible) number of computations and storage space dedicated to this purpose [22]. Since ADP algorithms have made great progress in the optimal control field [9,23,26], to the best of our knowledge, it is still an open problem how to solve the optimal control problem for time-delay systems with saturating actuators based on ADP algorithms. This motivates our research. First, the HJB equation for time-delay systems with saturating actuators is derived using a new iterative ADP algorithm.

In this paper a new iterative heuristic dynamic programming (HDP) algorithm is proposed to solve the optimal control problem for a class of nonlinear discrete time-delay systems with saturating actuators. Considering the saturation nonlinearity in the actuators, a nonquadratic performance index function is introduced. In the meantime, a state modification is used to deal with the obstacle induced by time delays. In the new iterative HDP algorithm the local and global optimization searching processes are developed to solve the optimal feedback control problem with convergence analysis. In the presented iterative HDP algorithm, two neural networks are used to facilitate the implementation of the iterative algorithm. Finally, two simulation examples are given to demonstrate the convergence and feasibility of the proposed optimal control scheme.
In order to solve this HJB equation, two optimization searching processes are proposed. One is local optimization searching process which is used to modify the states for every iteration, and the other is global optimization searching process which is used to get the optimal control law. We also give the convergence proof for the new iterative HDP algorithm. At last, two networks are used to implement the iterative HDP algorithm. They are critic neural network and action neural network which approximate the performance index function and compute the corresponding control law, respectively.

This paper is organized as follows. In Section 2, we present the problem formulation. In Section 3, the optimal control scheme is developed based on iterative HDP algorithm and the convergence proof is given. In Section 4, the neural network implementation for the optimal control scheme is discussed. In Section 5, two examples are given to demonstrate the effectiveness of the proposed control scheme. In Section 6, the conclusion is given.

2. Problem formulation

Consider a class of affine nonlinear discrete time-delay systems with saturating actuators as follows:

$$\begin{aligned}
    & x_{k+1} = f(x_k, u_k) + g(x_k, u_k)u_k, \\
    & x_k = \lambda_k, \\
\end{aligned}$$

where $x_{k, \cdots, k, \cdots, k} \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $f(x_{k, \cdots, k, \cdots, k}) \in \mathbb{R}^m$, $g(x_{k, \cdots, k, \cdots, k}) \in \mathbb{R}^{m \times m}$, $\lambda_k$ describes the initial condition. Set $0 = \sigma_0 < \sigma_1 < \cdots < \sigma_m$, $i = 1, \ldots, m$ is positive integer number. Here assume that the system is controllable on $\Omega \subset \mathbb{R}^n$. Assume that $f$ and $g$ are all Lipschitz continuous functions. We denote $\Omega_k = \{u_k | u_k = [u_1(1) \cdots u_m(1)]^T \in \mathbb{R}^m, u_i(i) \leq \sigma_i, i = 1, \ldots, m\}$, where $\Omega_k$ is the saturating bound for the $k$th actuator. Let $\mathbf{U} = \text{diag}(\Omega_1 \cdots \Omega_m)$.

This paper is desired to find an optimal control law for the system (1), which minimizes a generalized nonquadratic performance index function

$$J(x_k, u_k) = \sum_{i=k}^{\infty} [Q(x_i) + W(u_i)],$$

where $Q(x_k), W(u_k)$ are positive definite, and $U(x_k, u_k) = Q(x_k) + W(u_k)$ is the utility function denoting the cost incurred in going from $k$ to $k+1$ using control $u_k$.

Note that, for optimal control problems, the state feedback control laws $u_k$ not only stabilize the system (1) on $\Omega$, but also guarantee (2) is finite. Such controls are defined to be admissible [11].

Definition 1 (Admissible control). A control law $u_k(x_k)$ is defined to be admissible with respect to (2) on $\mathbb{R}^n$ if $u_k(x_k)$ is continuous on $\mathbb{R}^n$, $u(0) = 0$, and for $\forall x_k \in \mathbb{R}^n$, $u(x_k)$ stabilizes (1) on $\mathbb{R}^n$, and $\forall x_k \in \mathbb{R}^n, x_k(u_k(i))$ is finite, where $u_k(i) = (u_0, u_1, \ldots)$ and $u_k = U(x_k)$, $k = 0, 1, \ldots$

In this paper, we mainly discuss the optimal control for discrete time-delay system with saturating actuators. Inspired by [11,15], we let

$$Q(x_k) = X_{k-\sigma}^T Q X_{k-\sigma}$$

and

$$W(u_k) = 2 \int_0^{u_k} \phi^{-1}(\mathbf{T}^{-1}s)\mathbf{UR}ds,$$

where $X_{k-\sigma} = [x_{k-\sigma}, x_{k-\sigma-1}, \ldots, x_{k-\sigma-m}]$, $Q$, $R$ are positive definite, and we assume that $R$ is diagonal for simplicity of analysis, $s \in \mathbb{R}^m$, $\phi \in \mathbb{R}^m$, $\phi^{-1} = \psi^{-1}(\psi(u_1(1)) \cdots \psi(u_m(m)))$, $\psi$ is a bounded single mapping function satisfying $|\psi(\cdot)| \leq 1$ and belonging to $C^p$ ($p \geq 1$) and $L_2$. Moreover, it is a monotonic increasing odd function with its first derivative bounded by a constant $M$. We know that the hyperbolic tangent function $\psi(\cdot) = \tanh(\cdot)$ is one example of such function. Noting that $W(u_k)$ is assured to be positive definite by the definition above, because $\psi^{-1}(\cdot)$ is a monotonic odd function and $R$ is positive definite.

Let $J^*(x_k) = \min_{u_k} J(x_k, u_k)$ denote the optimal performance index function, and let $u^*_k$ denote the corresponding optimal control law. According to Bellman’s principle of optimality, the optimal performance index function $J^*(x_k)$ should satisfy the following HJB equation:

$$J^*(x_k) = \min_{u_k} \sum_{i=k}^{\infty} [Q(x_i) + W(u_i)] = \min_{u_k} [Q(x_k) + W(u_k) + J^*(x_{k+1})],$$

and the optimal controller $u_k^*$ should satisfy

$$u_k^* = \text{argmin}_{u_k} \sum_{i=k}^{\infty} [Q(x_i) + W(u_i)] = \text{argmin}_{u_k} \{Q(x_k) + W(u_k) + J^*(x_{k+1})\}. \quad (6)$$

The optimal control problem for the nonlinear discrete time-delay system with saturating actuators can be solved if the optimal performance index function $J^*(x_k)$ can be obtained from (5). However, there is currently no quite effective method for solving this performance index function for the nonlinear discrete time-delay system with saturating actuators. Therefore, in the following part we will discuss how to utilize the iterative HDP algorithm to seek the approximate optimal control solution.

3. The optimal control based on iterative HDP algorithm

3.1. Derivation of the iterative HDP algorithm

First, for any given initial state $x_k$ and initial control policy $u_k$, we start with initial iteration performance index function $J^0((\cdot), \cdot) = 0$. Then we find the control vector $u_k^0$ as follows:

$$u_k^0 = \text{argmin}_{u_k} \{Q(x_k) + W(u_k) + J^0(x_{k+1})\},$$

and the performance index function is updated as

$$J^1(x_k) = \min_{u_k} \{Q(x_k) + W(u_k) + J^0(x_{k+1})\} = \min_{u_k} \left( J^0(x_k) + \int_0^{u_k} \phi^{-1}(\mathbf{T}^{-1}s)\mathbf{UR}ds + J^0(x_{k+1}) \right), \quad (7)$$

where $x_1, \ldots, x_k$ are obtained under the action $u_k$, and $x_{k+1} = f(x_{k-m}, \ldots, x_{k-1}) + g(x_{k-m}, \ldots, x_{k-1})u_k^0$. Moreover, for $i = 1, 2, \ldots$ the iterative HDP algorithm iterates between

$$u_k^i = \text{argmin}_{u_k} \{Q(x_k) + W(u_k) + J^i(x_{k+1})\} = \min_{u_k} \left( J^i(x_k) + \int_0^{u_k} \phi^{-1}(\mathbf{T}^{-1}s)\mathbf{UR}ds + J^i(x_{k+1}) \right), \quad (10)$$

and

$$J^{i+1}(x_k) = \min_{u_k} \{Q(x_k) + W(u_k) + J^i(x_{k+1})\} = \min_{u_k} \left( J^i(x_k) + \int_0^{u_k} \phi^{-1}(\mathbf{T}^{-1}s)\mathbf{UR}ds + J^i(x_{k+1}) \right), \quad (11)$$

where $x_1, \ldots, x_k$ are obtained under the action $u_k^{i-1}$, and

$$x_{k+1} = f(x_{k-m}, \ldots, x_{k-1}) + g(x_{k-m}, \ldots, x_{k-1})u_k^{i-1}. \quad (12)$$

We further compute the control law $u_k^i$ from Eq. (10):

$$u_k^i = \mathbf{U}^{-1} \left[ \frac{\partial J^i(x_{k+1})}{\partial x_{k+1}} \right]. \quad (13)$$
Proof. The system is controllable if and only if \(\phi^{-1}(\mathbf{x}_k)\) is an admissible control input sequence. Therefore there exists an upper bound \(Y\) such that
\[
p^i+1(\mathbf{x}_k) \leq \lim_{i \to \infty} \sum_{j=0}^i L^{-i}(\mathbf{x}_{k+j}) \leq Y, \quad \forall i.
\]
(21)

Thus from the definition of cost function and Lemma 1 we can obtain
\[
0 \leq f^{i+1}(\mathbf{x}_k) \leq p^{i+1}(\mathbf{x}_k) \leq Y, \quad \forall i. \quad \Box
\]
(22)

Theorem 2. Define the cost function sequence \(f^i\) as in (11) with \(f^0(\cdot) = 0\), the control law sequence \(u_k^i\) as in (10). Then we can conclude that \(f^i\) is nondecreasing sequence satisfying \(f^{i+1}(\mathbf{x}_k) \geq f(x_k), \forall i, k\).

Proof. We define a new sequence \(\Phi(x_k)\) as follows:
\[
\Phi(x_k) = (X_{k-\sigma})^T Q X_{k-\sigma} + 2 \int_{0}^{t_k} \phi^{-1}(\mathbf{U}_k) s \mathbf{U}_k ds + 1,\]
(23)
with \(\Phi^0(\cdot) = f^0(\cdot) = 0\). And \(f^{i+1}(\mathbf{x}_k)\) is updated by (11).

In the following part, we prove \(\Phi(x_k) \leq f^{i+1}(\mathbf{x}_k)\) by mathematical induction.

First, we prove it holds for \(i = 0\). Notice that
\[
f^1(x_k) - \Phi(x_k) = (X_{k-\sigma})^T Q X_{k-\sigma} + 2 \int_{0}^{t_k} \phi^{-1}(\mathbf{U}_k) s \mathbf{U}_k ds \geq 0.
\]
(24)
Thus for \(i = 0\), we get
\[
f^1(x_k) \geq \Phi(x_k).
\]
(25)

Second, we suppose that \(f^i(x_k) \geq \Phi^{i+1}(x_k)\), for \(i = 0, \ldots, n, \forall k\). Then for \(i, n\), since
\[
f^{i+1}(x_k) - \Phi(x_k) = (X_{k-\sigma})^T Q X_{k-\sigma} + 2 \int_{0}^{t_k} \phi^{-1}(\mathbf{U}_k) s \mathbf{U}_k ds + f^i(x_k+1)
\]
(26)
and
\[
\Phi(x_k) = (X_{k-\sigma})^T Q X_{k-\sigma} + 2 \int_{0}^{t_k} \phi^{-1}(\mathbf{U}_k) s \mathbf{U}_k ds + \Phi^{i+1}(x_k+1)
\]
(27)
so we can get
\[
f^{i+1}(x_k) - \Phi(x_k) = f^i(x_k+1) - \Phi^{i+1}(x_k+1) \geq 0.
\]
(28)
Therefore, by the mathematical induction, we have
\[
f^{i+1}(x_k) \geq \Phi(x_k), \quad \forall i.
\]
(29)

In addition, from Lemma 1 we know that \(f^i(x_k) \leq \Phi(x_k)\). Therefore we have
\[
f^i(x_k) \leq \Phi(x_k) \leq f^{i+1}(x_k),
\]
(30)
which proves that \([f^i(x_k)]\) is a nondecreasing sequence bounded by (22). Hence we conclude that \([f^i(x_k)]\) is a convergent sequence as \(i \to \infty\). \(\Box\)

Theorem 3. According to Theorem 1, there is a limit for the performance index function sequence \([f^i(x_k)]\) when \(i \to \infty\). Without loss of generality, we define \(\lim_{i \to \infty} f^i(x_k) = f(x_k)\), accordingly, \(\lim_{i \to \infty} u_k^i = u_k\). Then, for any discrete time step \(k\), the limit \(f(x_k)\) is the “optimal” performance index function, i.e.,
\[
f^1(x_k) = \min_{u_k} (Q(x_k) + W(u_k) + f^1(x_{k+1})).
\]
(31)
Proof. For any \( i \), the performance index function sequence satisfies
\[
J^{i+1}(x_k) = \min_{u_k} (Q(x_k) + W(u_k) + J^i(x_{k+1})).
\]
Combining with \( J^{i+1}(x_k) \leq \lim_{i \to \infty} J^i(x_k) \) and \( \lim_{i \to \infty} J^i(x_k) = J^*(x_k) \), we can obtain
\[
J^*(x_k) \geq \min_{u_k} (Q(x_k) + W(u_k) + f(x_{k+1})).
\]
Let \( i \to \infty \), we have
\[
J^*(x_k) \geq Q(x_k) + W(u^*_k) + J^*(x_{k+1}).
\]
On the other hand, for any \( i \) and \( u^i_k \), we have
\[
J^{i+1}(x_k) \leq Q(\hat{x}_k) + W(\hat{u}^i_k) + f(x_{k+1}),
\]
where \( \hat{x}_{k+1} \) is under the action of \( \hat{u}^i_k \). As \( f(x_k) \leq f^{i+1}(x_k) \), we can obtain
\[
J^{i+1}(x_k) \leq Q(\hat{x}_k) + W(\hat{u}^i_k) + f(x_{k+1}).
\]
Let \( i \to \infty \), we have
\[
J^*(x_k) \leq Q(\hat{x}_k) + W(\hat{u}^*_k) + J^*(x_{k+1}).
\]
Since \( \hat{u}^i_k \) in the above equation is chosen arbitrarily. We have
\[
J^*(x_k) \leq Q(x_k) + W(u^*_k) + J^*(x_{k+1}).
\]
Thus, we have
\[
J^*(x_k) \leq Q(x_k) + W(u^*_k) + J^*(x_{k+1}) = \min_{u_k} (Q(x_k) + W(u_k) + J^*(x_{k+1})).
\]
Therefore, we can conclude that the performance index function sequence \( \{J^i\} \) converges to the optimal performance index function of the discrete-time HJB equation, i.e., \( J^i \to J^* \) as \( i \to \infty \). Simultaneously, we can conclude that the corresponding control law sequence \( \{u^i_k\} \) converges to the optimal control law \( u^*_k \) as \( i \to \infty \). □

In the following part, we give the detailed procedure about how to implement the iterative HDP algorithm.

4. The implementation of iterative HDP algorithm

In this paper, the local state modification and global optimization processes are used to implement the iterative HDP algorithm. Fig. 1 depicts the state modified implementation procedure for every iteration. Note that the local iterative index \( j \) starts from \( j = 1 \).

In the case of linear systems the cost and control laws are quadratic and linear, respectively. In the nonlinear case, this is not necessarily true and therefore one needs to use a parametric structure or a neural network to approximate both \( f(x_k) \) and \( u_k \). Therefore, as is standard, in order to implement the HDP iterations on Eqs. (13) and (14), we employ neural networks for function approximation.

Assume the number of hidden layer neurons is denoted by \( l \), the weight matrix between the input layer and the hidden layer is denoted by \( V \), the weight matrix between the hidden layer and the output layer is denoted by \( W \), then the output of three-layer NN is represented by
\[
\hat{F}(X,V,W) = W^T \sigma(V^TX),
\]
where \( \sigma(V^TX) \in \mathbb{R}^l \) is the activation function, the output layer neuron uses linear activation function. The NN estimation can be expressed by
\[
\hat{F}(X) = F(X,V^*,W^*) + \alpha(X),
\]
where \( V^*, W^* \) are the ideal weight parameters, \( \alpha(X) \) is the reconstruction error.

Here, there are two networks, which are critic network and action network, respectively. All the neural networks are chosen as three-layer feedforward network. The whole structure diagram is shown in Fig. 2.

Remark 1. For iterative HDP algorithm, neural networks are used to approximate the performance index function (11) and the corresponding optimal control policy (10), respectively, for facilitating the implementation of the iterative HDP algorithm. Thus it is not necessary to the detailed expressions of the performance index function and the optimal control law like the DP method. Therefore, the computation burden is much released. As the neural network probably converges to the local minimum point, therefore the optimal solution cannot ensure to be a global one.

4.1. The critic network

The critic network is used to approximate the performance index function \( f(x_k) \). The output of the critic network is denoted as
\[
\hat{f}^j(x_k) = (w_{1k})^T \sigma((w_{2k})^T X_k),
\]
where \( \sigma(V^TX) \in \mathbb{R}^l \) is the activation function, the output layer neuron uses linear activation function. The NN estimation can be expressed by
\[
\hat{f}(X,V^*,W^*) = F(X,V^*,W^*),
\]
where \( \sigma(V^TX) \in \mathbb{R}^l \) is the activation function, the output layer neuron uses linear activation function. The NN estimation can be expressed by
\[
\hat{f}(X) = F(X,V^*,W^*),
\]
where \( \sigma(V^TX) \in \mathbb{R}^l \) is the activation function, the output layer neuron uses linear activation function. The NN estimation can be expressed by
\[
\hat{f}(X) = F(X,V^*,W^*) + \alpha(X),
\]
where \( V^*, W^* \) are the ideal weight parameters, \( \alpha(X) \) is the reconstruction error.

Here, there are two networks, which are critic network and action network, respectively. All the neural networks are chosen as three-layer feedforward network. The whole structure diagram is shown in Fig. 2.
where $u_k^i$ is the target function which can be described by
\begin{align}
    u_k^i &= \mathcal{U}(\phi \left( -\frac{1}{2} (\mathcal{U}R)^{-1} g^T (x_{k-\sigma_1}, \cdots, x_{k-\sigma_m}) \frac{\partial f}{\partial x_{k+1}} \right)) \label{51}.
\end{align}

The weights in the critic network are updated to minimize the following cost error measure:
\begin{align}
    Ed_k^i &= \frac{1}{2} (ed_k^i)^T ed_k^i \label{52}.
\end{align}

The weights updating algorithm is similar to the one for the critic network. By the gradient descent rule, we can obtain
\begin{align}
    \Delta w_k^i &= a \left[ -\frac{\partial Ed_k^i}{\partial w_k^i} \right] \label{54},
\end{align}
and
\begin{align}
    \frac{\partial Ed_k^i}{\partial w_k^i} &= \frac{\partial Ed_k^i}{\partial u_k^i} \cdot \frac{\partial u_k^i}{\partial w_k^i}. \label{55}
\end{align}

and $a$ is the learning rate of action network, and the weight update rule for $\nu_k^i$ is the same as $w_k^i$.

5. Simulation study

In this section, two examples are provided to demonstrate the effectiveness of the control scheme proposed in this paper. The examples used in this section are the one appeared in [2,12] with modification.

5.1. Example 1

Consider the following two-order nonlinear discrete time-delay system with saturating actuators:
\begin{align}
    x_{k+1} &= f(x_k, x_{k-2}) + g(x_k, x_{k-2}) u_k, \quad k \geq 0, \\
    x_k &= \lambda_k, \quad k = 0, -1, -2, \label{56}
\end{align}
where $f(x_k, x_{k-2}) = [0.2x_k \exp(-0.1x_{k-2})^2, 0.3x_k(x_{k-2})^3]$, and assume that the control constraint is set to $|u_k| \leq 0.3$.

Define the performance index functional as
\begin{align}
    J(x_k) &= \sum_{k=0}^{\infty} \left( x_{k-\sigma}^T Q x_{k-\sigma} + 2 \int_0^{u_k} \tanh^{-1}(\mathcal{U}^{-1} s) \mathcal{U} R ds \right), \label{57}
\end{align}
where $x_{k-\sigma} = [x_k; \cdots; x_{k-\sigma}]$, $x_k = [x_k(1); x_k(2)]$, $Q = I_m$, $R = 0.5$. The initial condition $x_k = [1.50, 5]^T$, $k = 0, -1, -2$. Moreover, we implement the algorithm at the time step $k = 2$.

In the presented iterative HDP algorithm, we choose three-layer neural networks as the critic network and the action network with the structure 2–8–1 and 4–8–1. The initial weight matrices are chosen randomly from $[-0.1, 0.1]$. The critic network and the action network are trained with the learning rates $\lambda_a = \lambda_c = 0.01$. After 5000 iterative steps we get the curves of states, control input and performance index function, and the action NN requires $4 \times 8 + 1 \times 8 = 40$ bytes and the critic NN requires $2 \times 8 + 1 \times 8 = 24$ bytes. The memory requirement of the iterative HDP algorithm is about 5064 bytes for action NN, critic NN and the sequence of performance index function. The whole computing process takes 17.463457 s. The state trajectories are given as Fig. 3, and the corresponding control curve is given as Fig. 4. The convergence curve of the performance index function is shown in Fig. 5.
To more effectively demonstrate the efficiency of the proposed approach, we adopt conventional DP method to obtain the performance index value for system (56). Firstly, we discretize the control space at intervals of 0.006, i.e. $u_k$ has 100 choices, $8_k$. We compute three steps, so we assume that the state $x_5 = 0$ and the control $u_5 = 0$ at time step $k = 5$, accordingly $J(x_5) = 0$. Then we apply DP method. The detail of the DP method can be seen in [22], and the detail is omitted here. In the process of DP method, we should record all the corresponding performance index values of all $u_k$. So the requirement is about $1600^3$ bytes. After 472.786592 s the value of the performance index function for three steps is obtained as 36.6200.

From the above analysis, we can see that the iterative HDP algorithm get around the problem of large number of computations and storage space appearing in the DP method.

5.2. Example 2

Consider the following three-order nonlinear discrete time-delay system with saturating actuators:

$$\begin{align*}
    x_{k+1} &= f(x_k, x_{k-2}) + g(x_k, x_{k-2})u_k, \quad k \geq 0, \\
    x_k &= \lambda_k, \quad k = 0, -1, -2, \\
\end{align*}$$

(58)

where

$$\begin{align*}
    f(x_k, x_{k-2}) &= \begin{bmatrix}
        0.2x_{k-2}(1)\exp(x_{k}(2))^2 \\
        0.3(x_{k}(2))^3 \\
        x_{k}(3)
    \end{bmatrix}, \\
    g(x_k, x_{k-2}) &= \begin{bmatrix}
        0 \\
        -0.2 \\
        0
    \end{bmatrix}.
\end{align*}$$

$X_k = [x_k(1); x_k(2); x_k(3)]$, and assume that the control constraint is set to $|u_k| \leq 0.3$, $\lambda_k = [1.5 \; 0.5 \; 1]^T$. 

To more effectively demonstrate the efficiency of the proposed approach, we adopt conventional DP method to obtain the performance index value for system (56). Firstly, we discretize the control space at intervals of 0.006, i.e. $u_k$ has 100 choices, $8_k$. We compute three steps, so we assume that the state $x_5 = 0$ and the control $u_5 = 0$ at time step $k = 5$, accordingly $J(x_5) = 0$. Then we apply DP method. The detail of the DP method can be seen in [22],
The performance function index is the same as Example 1, where $Q=I_n$, $R=0.5$. We choose three-layer neural networks as the critic network and the action network with the structure $3\times8\times1$ and $6\times8\times1$. The initial weight matrices are chosen randomly from $[-0.1, 0.1]$. The critic network and the action network are trained with the learning rates $\alpha_c = \alpha_a = 0.01$. After 18,521788 s, we get the simulation results. The state trajectories are shown in Fig. 6, and the corresponding control curve is given as Fig. 7. The convergence curve of the performance index function is shown in Fig. 8. It is clear from the simulations that the new iterative algorithm in this paper is very effective.

6. Conclusion

In this paper, we have developed an effective algorithm to resolve optimal control problem for nonlinear time-delay system with saturating actuators. First we have defined a performance index function for time-delay system. Then a novel iterative HDP algorithm has been developed which contains two optimization processes. Two neural networks have been used to facilitate the implementation of the iterative algorithm. Simulation study has demonstrated the effectiveness of the proposed optimal control algorithm.

References


Rui Zhuo Song is currently working towards Ph.D. degree in Northeastern University from 2008. His research interests include neural-networks-based control, nonlinear control, fuzzy control, approximate dynamic programming and their industrial application.

Huaguang Zhang (SM’04) received the B.S. degree and the M.S. degree in control engineering from Northeast Dianli University of China, Jinlin City, China, in 1982 and 1985, respectively. He received the Ph.D. degree in thermal power engineering and automation from Southeast University, Nanjing, China, in 1991. He joined the Department of Automatic Control, Northeastern University, Shenyang, China, in 1992, as a Postdoctoral Fellow for two years. Since 1994, he has been a Professor and Head of the Institute of Electric Automation, School of Information Science and Engineering, Northeastern University, Shenyang, China. His main research interests are fuzzy control, stochastic system control, neural networks based control, nonlinear control, and their applications. He has authored and coauthored over 200 journal and conference papers, four monographs and co-invented 20 patents.

Dr. Zhang is an Associate Editor of Automatica, IEEE Transactions on Fuzzy Systems, IEEE Transactions on Systems, Man, and Cybernetics—Part B and Neurocomputing, respectively. He was awarded the Outstanding Youth Science Foundation Award from the National Natural Science Foundation Committee of China in 2003. He was named the Cheung Kong Scholar by the Education Ministry of China in 2005.
Yanhong Luo (M’09) received the B.S. degree in automation control, the M.S. degree and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2003, 2006, and 2009, respectively. She is currently working in Northeastern University as an Associate Professor. Her research interests include fuzzy control, neural networks adaptive control, approximate dynamic programming, and their industrial application.

Qinglai Wei received the B.S. degree in automation control and the M.S. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2002 and 2005, respectively, and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2009. He is currently a Postdoctoral Fellow with the Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include neural-networks-based control, non-linear control, adaptive dynamic programming and their industrial application.