Neural Network Based Finite Horizon Optimal Control for a Class of Nonlinear Systems with State Delay and Control Constraints

Xiaofeng Lin, Nuyun Cao, Yuzhang Lin

Abstract—In this paper, a new finite horizon iterative ADP algorithm is used to solve a class of nonlinear systems with state delay and control constraints problem and finite time $\epsilon$-optimal control is obtained. First of all, a new performance index function is designed to deal with the control constraints, the discrete nonlinear systems HJB equation with state delay is presented. Second, the iterative process and mathematical proof of the convergence is illustrated for the proposed finite horizon ADP algorithm. Approximate optimal control is obtained by introducing an error bound $\epsilon$. Two BP neural networks are developed to approximate control law function and performance index function in our ADP algorithm. Finally, comparing simulation cases are used to verify the effectiveness of the method proposed in this paper.

I. INTRODUCTION

TIME delay, control constraints (saturation, dead-zone, backlash, etc.) usually occur in practical control systems.

Strictly speaking, delay exists in all of the actual systems. Due to time delay, controller adjustment time of the system will increase and could not be able to response the change of the systems in time. Such problems cause great difficulties to analysis and design the controller for time delay systems [1-2]. On the other hand, control constraints will reduce the system's dynamic performance, and even affect the stability of the system [3-4]. Therefore, it is quite important to study the optimal control problems with time delay and control constraints system. However, most of the research results are about system stability or just considering linear system [5-7]. There is no command way to deal with time delay and control constraints optimal control problems for nonlinear systems.

Adaptive dynamic programming (ADP) is an effective method to deal with optimal control problem for nonlinear systems, it is solved by iterative to get an approximate solution of HJB equation, and avoids the "curse of dimensionality" in using dynamic programming, and has achieved much attention [8-11].

However, for practical systems, a limited period of time is required to achieve control. $\epsilon$-ADP algorithm provides an effective way to deal with finite time optimal control by introducing an $\epsilon$-error bound [12-13]. In [14], finite horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems is obtained by finite time ADP. X.Lin uses this kind of finite horizon ADP algorithm to get the optimal control for nonlinear systems with saturating actuators, and obtain the $\epsilon$-optimal control within finite time steps [15]. But it does not consider time delay which may always happen in actual systems. This motivates our research.

In this paper, we will use the new finite horizon ADP algorithm to get $\epsilon$-optimal control for a class of nonlinear systems with time delay and control constraints. It is organized as follows, In Section II, the state delay and control constraints optimal control problem is stated and the performance index function is defined. In Section III, the iterative process and the convergence of our new ADP algorithm are derived. In Section IV, $\epsilon$-optimal control is obtained in finite time steps, and BP neural network is used to implement our finite horizon ADP algorithm. In the last section V, comparing simulation examples are given to show the effectiveness of our method.

II. PROBLEM STATEMENT

In this study, we will consider a class of discrete-time affine nonlinear systems with state delay as follows

$$\begin{align*}
x(t+1) &= f(x(t), x(t-\sigma)+g(x(t), x(t-\sigma))u(t), t = 0,1,\ldots, \\
x(t) &= \phi(t), t = -\sigma, -\sigma+1,\ldots, 0
\end{align*}$$

(1)

Where $x(t), x(t-\sigma) \in \mathbb{R}^n$ are state vectors, $u(t) \in \mathbb{R}^m$ is control input vector. $\sigma > 0$ is the state delay and it is an integer. Assume that $f(\cdot), g(\cdot), x(\cdot), x(t-\sigma)u(\cdot)$ is Lipschitz continuous for $x, u$ on a set $\Omega \subset \mathbb{R}^n$. The initial state is $x(s), s = -\sigma, -\sigma+1,\ldots, 0$, and the control set is

$$\Omega = \{u(k) = [u_1(k), u_2(k), \ldots, u_m(k)]^T \in \mathbb{R}^m : |u_i(k)| \leq \pi_i, i = 1,\ldots,m\}$$

where $|u_i(k)| \leq \pi_i$ is the control constraints we introduce. The finite horizon optimal control problem for nonlinear systems with state delay and control constraints is to minimize an opportunity performance index function with a finite control sequence

$$u^N = (u(0), u(1), \ldots, u(N-1))$$

from the initial state.

A. Definition of the Performance index function

Performance index function is used to evaluate the control systems, which can be defined according to the needs of the actual system. For state delay optimal control problem without constrains, the performance index function usually select a quadratic function, that is,

$$J(x(0)) = \sum_{j=0}^{N-1} (x(i-\sigma)^TQx(i-\sigma) + u(i)^TRu(i))$$

(2)

Where $U = x(i)^TQx(i) + u(i)^TRu(i))$ is the utility function, and $Q, R > 0$ are two constant matrixes with appropriate
dimensions.

In our case, if we still use this performance index function, then the optimal control we get will exceed the constraints, it may not be able to guarantee to get the optimal system performance, even leads to system instability.

Inspired by Refs. [16], we define a new function as follows

\[
J(x(0)) = \sum_{i=0}^{N-1} (x(i-\sigma)^T Q x(i-\sigma) + 2 \int_0^{x(i)} \Theta^T (\Lambda^{-1}) R dx) \tag{3}
\]

Where \(\Theta^T(u(i)) = [\Theta^T(u_1(i)), \Theta^T(u_2(i)), \ldots, \Theta^T(u_n(i))]^T\), \(|u_i| \leq \Lambda, u \in R^n, \Theta \in R^n\), let \(R\) be a diagonal positive definite, \(\theta(*)\) is a bounded monotonically increasing odd function belongs to \(C^r (p \geq 1)\) and \(L_2(\Omega)\) with \(|\theta'| \leq 1\), the first derivative is a bounded constant, let \(\theta(*) = \tanh(*)\), then \(\Theta^T(u(i)) = [\Theta^T(u_1(i)), \Theta^T(u_2(i)), \ldots, \Theta^T(u_n(i))]^T\).

Thus, (3) could be written as (4)

\[
J(x(0)) = \sum_{i=0}^{N-1} (x(i-\sigma)^T Q x(i-\sigma) + 2 \int_0^{x(i)} \Theta^T (\Lambda^{-1}) R dx) \tag{4}
\]

Let \(\mathcal{M}(u(i)) = 2 \int_0^{x(i)} (\tanh^{-1}(v/\Lambda))^2 R d v\), then the performance index function is

\[
J(x(0)\omega_{N-1}) = \sum_{k=0}^{N-1} (x(k-\sigma)^T Q x(k-\sigma) + \mathcal{M}(u(k))) \tag{5}
\]

B. Discrete time HJB equation with state delay and control constraints

Assume that the state \(x(k)\) is finite time controllable (defined in Refs. [12]), then the finite time optimal performance index function can be written as

\[
J^*(x(k)) = \inf_{u \in \Psi_{x(k)}} \{ J(x(k), u_k): u_k \in \Psi_{x(k)} \} \tag{6}
\]

Where \(\Psi_{x(k)}\) be all the admissible control sets.

Notice that (5) could be written as

\[
J(x(k)\omega_{N-1}) = \sum_{i=k}^{N-1} (x(i-\sigma)^T Q x(i-\sigma) + \mathcal{M}(u(i))) \tag{7}
\]

By Bellman’s principle of optimality, the optimal performance function satisfies the discrete-time HJB equation

\[
J^*(x(k)) = \min_{u_{k+1}} \{ x^T (k-\sigma) Q x(k-\sigma) + \mathcal{M}(u(k)) + J^*(x(k+1)) \} \tag{8}
\]

Optimal control \(u^*\) satisfies

\[
\frac{\partial J^*(x(k))}{\partial u(k)} = 0, \text{ that is}
\]

\[
\frac{\partial \mathcal{M}(u(k))}{\partial u(k)} + \frac{\partial J^*(x(k+1))}{\partial u(k)} = 0
\]

Where \(\mathcal{M}(u(k)) = 2 \int_0^{x(k)} (\tanh^{-1}(v/\Lambda))^2 R d v\), thus,

\[
\frac{\partial J^*(x(k))}{\partial u(k)} = 2 \tanh^{-1}(\Lambda^{-1} u(k)) R, \text{ then}
\]

\[
\frac{\partial \mathcal{M}(u(k))}{\partial u(k)} = 2 \Lambda^{-1} u(k) R, \text{ then}
\]

\[
\frac{\partial J^*(x(k))}{\partial u(k)} = 2 \Lambda^{-1} u(k) R + \frac{\partial J^*(x(k+1))}{\partial u(k)}
\]

Let \(u^*(k) = \text{ArgMin} \{ x^T (k-\sigma) Q x(k-\sigma) + \mathcal{M}(u(k)) + J^*(x(k+1)) \}

\]

III. Iterative Process and Convergence Proof for ADP

A. Iterative Process for ADP algorithm

For any state \(x(k), k > \sigma\), the cost function sequence (i.e. performance index function) \(\{V^i(x(k))\}\) and the control law function sequence \(\{u^i(k)\}\) are updated by recursive iteration.

i=0, the initial cost function \(V_0(x(0)) = 0\).

i=1, the cost function satisfies

\[
V_1(x(k)) = \min_{u(k)} \{ x^T (k-\sigma) Q x(k-\sigma) + \mathcal{M}(u(k)) \}
\]

s.t. \(f(x(k), x(k-\sigma)) + g(x(k), x(k-\sigma)) u(k) = 0\)

Where

\[
\text{argmin}_{u(k)} \{ x^T (k-\sigma) Q x(k-\sigma) + \mathcal{M}(u(k)) \}
\]

For \(i=2,3,\ldots\), we have

\[
V_i(x(k)) = \min_{u(k)} \{ x^T (k-\sigma) Q x(k-\sigma) + \mathcal{M}(u(k)) + V_{i-1}(x(k+1)) \}
\]

and

\[
\text{argmin}_{u(k)} \{ x^T (k-\sigma) Q x(k-\sigma) + \mathcal{M}(u(k)) + V_{i-1}(x(k+1)) \}
\]

In the following part, we will prove the convergence of the above ADP algorithm, that is to say, when the iteration step \(i\) goes to \(\infty\), then we could get the optimal performance index function \(V^*\) and the optimal control \(u^*\) by iteration of equation (11) and (14).
B. Convergence Proof for ADP algorithm

\[ V_i(x(k)) = \min_{\sigma_i} \left\{ J(x(k)|Q(x(k-1)+\mathcal{M}(u(k+1)) \right\} \]
\[ = \min_{\sigma_i} \left\{ x(k)^T(Qx(k-1)+\mathcal{M}(u(k+1)) \right\} \]
\[ = \min_{\sigma_i} \left\{ x(k)^T(Qx(k-1)+\mathcal{M}(u(k+1)) \right\} \]
\[ = \min \left\{ x(k)^T(Qx(k-1)+\mathcal{M}(u(k+1)) \right\} \]

Where

\[ V_i(x(k+i)) = \min_{\sigma_i} \left\{ J(x(k+i)|Q(x(k+i-1)+\mathcal{M}(u(k+i))) \right\} \]

s.t. \( f(x(k+i),x(k+i-1)) + g(x(k+i),x(k+i-1))u(k+i) = 0 \)

Thus we have

\[ V_{i+1}(x(k)) = \min_{\sigma_i} \left\{ J(x(k)|Q(x(k)+\mathcal{M}(u(k))) \right\} \]

Theorem 1

For system (1), let \( x(k), \sigma \geq 0 \) be an arbitrary state, then \( \{ V_i(x(k)) \} \) is a monotonically non-increasing sequence, i.e., for all \( i \geq 1, V_i(x(k)) \leq V_{i+1}(x(k)) \).

Proof. We will prove this theorem by mathematical induction.

(i) For \( i = 1 \), let \( \theta_{k+1} = \{ u_k \} \), then

\[ x(k+1) = f(x(k),x(k-\sigma)) + g(x(k),x(k-\sigma))u(k) = 0 \]

\[ \hat{V}_1(x(k),\theta_{k+1}) = x(k)^T(Qx(k)+\mathcal{M}(u(k))) \]

Where

\[ V_1(x(k)) = \min_{\sigma_i} \left\{ x(k)^T(Qx(k)+\mathcal{M}(u(k))) \right\} \]

So we have

\[ \hat{V}_1(x(k),\theta_{k+1}) = V_1(x(k)) \]

By (15), we can obtain

\[ V_i(x(k)) = \min_{\sigma_i} \left\{ J(x(k)|Q(x(k)+\mathcal{M}(u(k))) \right\} \]

\[ = \min_{\sigma_i} \left\{ x(k)^T(Qx(k)+\mathcal{M}(u(k))) \right\} \]

\[ \leq V_{i+1}(x(k)) \]

So that \( V_i(x(k)) \leq V_{i+1}(x(k)) \) holds.

(ii) Assume that Theorem 1 holds for \( i \geq 2 \). Then we have \( u_{k+i+1} = (u_k, u_{k+1}, \cdots, u_{k+i-1}) \), thus

\[ V_i(x(k)) = \min_{\sigma_i} \left\{ x(k+i)^T(Qx(k+i)+\mathcal{M}(u_{k+i})) \right\} \]

For \( i + 1 \), let \( \theta_{k+i+1} = (u_k, u_{k+1}, \cdots, u_{k+i-1}) \), then we have

\[ x(k+i-1|\sigma) = x(k+i-1|\sigma) + \mathcal{M}(u_{k+i-1}) \]

\[ V_{i+1}(x(k),\theta_{k+i+1}) = \min_{\sigma_i} \left\{ x(k+i-1|\sigma) \right\} \]

\[ = \min_{\sigma_i} \left\{ x(k+i-1|\sigma) \right\} \]

By (15), we can obtain

\[ V_{i+1}(x(k)) = \min_{\sigma_i} \left\{ J(x(k)|Q(x(k)+\mathcal{M}(u(k+i))) \right\} \]

\[ = \min_{\sigma_i} \left\{ x(k+i)^T(Qx(k+i)+\mathcal{M}(u(k+i))) \right\} \]

\[ \leq V_{i+1}(x(k)) \]

Theorem 1 holds by (i)(ii).

Theorem 2

Let \( V_{\infty}(x(k)) = \lim_{i \to \infty} V_i(x(k)) \), then for any state \( x(k), \sigma \geq 0 \), \( V_{\infty}(x(k)) \) satisfies HJB equation, that is

\[ J'(x(k))V_{\infty}(x(k)) = \min_{\sigma_i} \left\{ x(k)^T(Qx(k)+\mathcal{M}(u(k))) \right\} \]

Proof. Let \( \hat{u} \) be an admissible control, then we have

\[ V_{\infty}(x(k)) \leq x(k)^T(Qx(k)+\mathcal{M}(u(k))) \]

By Theorem 1, \( V_{\infty}(x(k)) = \lim_{i \to \infty} V_i(x(k)) \leq V_{i+1}(x(k)) \)

With (16) and (17), we could obtain

\[ V_{\infty}(x(k)) \leq x(k)^T(Qx(k)+\mathcal{M}(u(k))) + V_{\infty}(x(k+1)) \]

On the other side, for any \( \epsilon > 0 \), there exists a positive integer \( p \) which satisfies \( V_{\infty}(x(k))-\epsilon \leq V_{\infty}(x(k)) \), then

\[ V_{\infty}(x(k)) \geq V_{\infty}(x(k)) - \epsilon \]

Since \( \epsilon \) is arbitrary, we have

\[ V_{\infty}(x(k)) = \min_{\sigma_i} \left\{ x(k)^T(Qx(k)+\mathcal{M}(u(k))) + V_{\infty}(x(k+1)) \right\} \]

Theorem 2 holds. It is clear that \( J'(x(k)) = V_{\infty}(x(k)) \), the convergence of our ADP algorithm is completely proved.

IV. \( \epsilon \)-OPTIMAL CONTROL AND BP NEURAL NETWORK IMPLEMENTATION FOR ADP ALGORITHM

A. \( \epsilon \)-Optimal Control

Theorem 1-2 has shown the convergence of the iterative ADP algorithm for nonlinear systems with state delay and control constraints, that is to say, the optimal index function and optimal control will obtain as iteration step goes to infinity. However, in practical control systems, it is impossible to run all the time. In order to get a finite horizon control, we introduce an error bound \( \epsilon \), then an approximate \( \epsilon \)-optimal control will be obtained within finite time [12].

Let \( x(k), \sigma \geq 0 \) be any controllable state, \( \epsilon > 0 \) be any small number, then the \( \epsilon \)-optimal iteration step for finite horizon ADP algorithm can be defined as

\[ K(x(k)) = \min \left\{ i : J'(x(k)) - V_{\infty}(x(k)) \leq \epsilon \right\} \]

the \( \epsilon \)-optimal control is defined as

\[ u_{\epsilon}(k) = u_i(k) \]

where \( u_i(k) \) is an optimal control at the iteration step \( i \).

Theorem 3: let \( \epsilon > 0 \) be any small positive number, then

\[ V_{\infty}(x(k)) - J'(x(k)) \leq \epsilon \Rightarrow V_{\infty}(x(k)) - V_{\infty}(x(k)) \leq \epsilon \]

Proof:

(i) \( V_{\infty}(x(k)) - J'(x(k)) \leq \epsilon \Rightarrow V_{\infty}(x(k)) - V_{\infty}(x(k)) \leq \epsilon \)
Since $|V(x(k)) - J'(x(k))| \leq \varepsilon$, so we have,

$$V'(x(k)) \leq J'(x(k)) + \varepsilon$$

Thus,

$$J'(x(k)) \leq V'_c(x(k)) \leq V(x(k))$$

Then

$$0 \leq V(x(k)) - V'_c(x(k)) \leq \varepsilon$$

Hence

$$|V'_c(x(k)) - V(x(k))| \leq \varepsilon$$

(ii) $V'_c(x(k)) = V_c(x(k))$ since $\varepsilon$ is an any small positive number, so we have $|V'_c(x(k)) - V(x(k))| \to 0$, that is $V'(x(k)) \to J'(x(k))$. By theorem 1 and theorem 2, we can get that $\lim_{k \to \infty} V'(x(k)) = J'(x(k))$. According to the definition of limit, for $\forall \varepsilon > 0$, there exists an integer $I \in N$, when $i > I$, $|V(x(k)) - J'(x(k))| \leq \varepsilon$ holds.

Together with (i) and (ii), we could get

$$|V'_c(x(k)) - V(x(k))| \leq \varepsilon \Rightarrow |V(x(k)) - J'(x(k))| \leq \varepsilon$$

Theorem 3 will help to achieve program implementation for our finite horizon ADP algorithm, since the optimal performance index function $J'(x(k))$ is unknown for the optimality criteria inequality $|V'(x(k)) - V(x(k))| \leq \varepsilon$, according to Theorem 3, we use the new criteria inequality $|V'_c(x(k)) - V(x(k))| \leq \varepsilon$ instead.

B. BP Neural Network implementation

The solution of the HJB equation is required for nonlinear optimal control, which is usually unable to obtain the exact solution. BP neural network is used to approximate the control law function $u(x)$ and performance index function $V'(x(k))$ due to its strong nonlinear capacity [11]. Figure 1 shows the structure of our iterative ADP algorithm using BP neural network.

![Structure diagram of the iterative ADP algorithm using BP NN](Image)

Where critic BP Network is used to approximate $V'(x(k))$ and action BP network is used to approximate $u(x)$. The critic network is designed to approximate $V'(x(k))$, the input of the network is $x(k)$, thus the output is

$$V_c(x(k)) = \rho(V'(x(k)))$$

Thus,

$$V'_c(x(k)) = x^T(k)Qx(k) + u^T(k)Ru(k) + J(x(k))$$

Then the approximate for critic network error is

$$e_c(k) = V'_c(x(k)) - V_c(x(k))$$

Our aim is to minimized the following function

$$E_c(k) = \frac{1}{2} e^T_c(k)e_c(k)$$

By the same way, we could design the action network in order to approximate $u(k)$ . The input of the network are $x(k), x(k-\sigma)$, while the output comes to

$$\hat{u}(x(k)) = W^T_c \rho(V_c(x(k)))$$

The target of the output of the action network is given by (9). The approximate error for action network will be

$$e_u(k) = \hat{u}(k) - u(k)$$

Our aim is to minimized the following function

$$E_u(k) = \frac{1}{2} e^T_u(k)e_u(k)$$

V. SIMULATION STUDY

In this study, we will consider a class of nonlinear systems with state delay $\sigma = 2$

$$x(t+1) = f(x(t), x(t-2)+g(x(t), x(t-2))u(t), t \geq 0$$

$$x(t) = \phi(t), -2 \leq t \leq 0$$

$$f(x(t), x(t-2)) = \begin{bmatrix} -0.5x_1(k-2) \\ \sin(0.8x_1(k-2) - x_2(k)) + 1.8x_2(k-2) \end{bmatrix},$$

$$g(x(t), x(t-2)) = \begin{bmatrix} -0.1x_1(k-2) \\ 0 \\ 0 - 0.8x_2(k-2) \end{bmatrix},$$

the performance index function is defined in (4), where

$$Q = R = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.3 \end{bmatrix}, A = \begin{bmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{bmatrix} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix},$$

i.e. the control constraints is $|p_1| \leq 0.2, |p_2| \leq 0.2$. The parameter selection for our finite horizon ADP algorithm is list as follows,

The initial states $\phi(t) = [1, -1]^T$, the initial time step is $k = 3$, error bound $\varepsilon = 10^{-5}$. Learning rates for action BP network and critic BP network are $\alpha = \beta = 0.01$, the network structures are 4-10-2 and 2-12-1, respectively, the initial weight for the networks are set to be random in [-0.01, 0.01]. Simulation results are shown in Figure 2-4.
Figure 2 shows the convergence process of performance index function, in which the $V_i(x(k))$ is decreasing with the increase of the number of iteration steps, while the $\varepsilon$-optimal iteration step $K_s(x(k))=15$, that is to say, the performance index function converges to optimal using our finite horizon ADP algorithm with an error bound $\varepsilon=10^{-4}$.

Figure 3 and figure 4 show the state and control trajectory for the state delay system in 30 time steps with the given control constraints $|u_1| \leq 0.2, |u_2| \leq 0.2$. The results show that the system could control in stable within finite time, at the same time, the control keeps under the constraints.

Let’s change a control constraints for $|u_1| \leq 0.1, |u_2| \leq 0.1$, for this case $\Lambda = \begin{bmatrix} \pi & 0 \\ 0 & \pi \end{bmatrix} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$, instead, one should notice that the definition of the performance index function has changed. The parameter selection for our finite horizon ADP algorithm stays the same for this case, and run our algorithm again to get state and control trajectory as shown in figure 5 and figure 6. Results shows that the system could control in stable within 30 time steps and control, however, control components $|u_1|$ is lager than 0.2 and 0.1 for this time.

Comparing with the above three experiment cases, we finally could draw a conclusion that the method we proposed in this paper works well for nonlinear systems with state delay and control constraints.

VI. CONCLUSION

1) We developed an $\varepsilon$-optimal control for a class of
nonlinear discrete time systems with state delay and control constraints based on finite horizon ADP algorithm.

2) Research innovations: a new type of performance index function was defined to deal with control constraints, discrete time HJB equation with state delay and control constraints was derived. Iteration ADP algorithm was used to overcome the difficulties of nonlinear systems, mathematical analysis for convergence was proved.

3) Two BP networks were used to get control law function and performance index function for finite time ADP algorithm, i.e. action BP network and critic BP network. Simulation results show the effectiveness of our method.

4) Future work. Other control constraints could be considered further, such as dead zone, backslash, etc. Besides, other neural network structure could be introduced, such as echo state network for finite horizon ADP algorithm [17].

REFERENCES


