Neural Networks 32 (2012) 236–244

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2012 Special Issue

An iterative ϵ-optimal control scheme for a class of discrete-time nonlinear systems with unfixed initial state

Qinglai Wei, Derong Liu *

State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Keywords:
Adaptive dynamic programming
Approximate dynamic programming
ϵ-optimal control
Finite horizon
Neural networks

ABSTRACT

In this paper, a finite horizon iterative adaptive dynamic programming (ADP) algorithm is proposed to solve the optimal control problem for a class of discrete-time nonlinear systems with unfixed initial state. A new ϵ-optimal control algorithm based on the iterative ADP approach is proposed that makes the performance index function iteratively converge to the greatest lower bound of all performance indices within an error ϵ in finite time. The convergence analysis of the proposed ADP algorithm in terms of performance index function and control policy is conducted. The optimal number of control steps can also be obtained by the proposed ϵ-optimal control algorithm for the unfixed initial state. Neural networks are used to approximate the performance index function, and compute the optimal control policy, respectively, for facilitating the implementation of the ϵ-optimal control algorithm. Finally, a simulation example is given to show the effectiveness of the proposed method.

1. Introduction

Strictly speaking, most real-world control systems need to be effectively controlled within a finite time horizon (finite horizon in brief), such as stabilized within a finite horizon. In many theoretical discussions, however, controllers are generally designed to stabilize controlled systems within an infinite time horizon (Dierks, Thumati, & Jagannathan, 2009; Vrabie & Lewis, 2009; Zhang, Wei, & Luo, 2008). The design of finite-time horizon controller faces a major obstacle in comparison with the infinite horizon one. For infinite horizon control problems, Lyapunov theory is popularly used and asymptotic results for the control systems are usually obtained (Landelius, 1997; Zhang, Luo, & Liu, 2009). That is, the system cannot really be stabilized until the time reaches infinity. While for finite horizon control problems, the system must be stabilized to zero within a finite time (Necoara, Kerrigan, Schutter, & Boom, 2007; Uchida & Fujita, 1992). Due to the lack of methodology and the fact that the number of control steps is difficult to determine, the controller design of finite horizon problems is still a challenge to control engineers. On the other hand, optimization is always an important objective for the design of control systems. This is the reason why optimal control has been paid much attention by many researchers for over fifty years and applied to many application domains (Ichihara, 2009; Kioskeridis & Mademlis, 2009; Mao & Cassandras, 2009; Werbos, 2009).

An adaptive dynamic programming (ADP) algorithm was proposed by Werbos (1991), as a powerful methodology for solving optimal control problems forward-in-time. In Prokhorov and Wunsch (1997), ADP approaches were classified into several main schemes: Heuristic Dynamic Programming (HDP), Dual Heuristic Programming (DHP), Action Dependent Heuristic Dynamic Programming (ADHDP), also known as Q-learning, and Action Dependent Dual Heuristic Programming (ADDHP), Globalized-DHP (GDHP) and ADGDHP. Though great progress has been made in ADP research in the optimal control field (Al-Tamimi, Abu-Khalaf, & Lewis, 2008; Kulkarni & Venayagamoorthy, 2010; Liu, Zhang, & Zhang, 2005; Murray, Cox, Lendaris, & Saeks, 2002; Vamvoudakis & Lewis, 2011; Wang, Zhang, & Liu, 2009; Wei, Zhang, & Dai, 2009; Zhang, Wei, & Liu, 2011), discussions about finite horizon optimal control problems are scarce. To the best of our knowledge, only (Wang, Jin, Liu, & Wei, 2011) discussed a finite horizon optimal control problem with fixed initial state. Wei and Liu (2011a, 2011b) proposed an iterative ADP algorithm with unfixed initial state while it requires that the system can reach zero in one step of control to initialize the algorithm which limits the application. So, it is still an open problem how to solve the optimal control problem in a finite horizon with unfixed initial state when the system cannot reach zero directly. This motivates our research.

In this paper, for the first time, we will show how to find an approximate optimal control that makes the performance index function converge to the greatest lower bound of all performance indices within an error according to ϵ (called ϵ-error bound for brief) without the initial condition requirements in Wei and

* Corresponding author. Tel.: +86 10 62557379; fax: +86 10 62557379.
E-mail addresses: qinglaiwei@gmail.com (Q. Wei), derong.liu@ia.ac.cn, derongliu@gmail.com (D. Liu).

0893-6080/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2012.02.027

Liu (2011a, 2011b). It is also shown that the corresponding approximate optimal control (called ϵ-optimal control) can make the performance index function converge to the ϵ-error bound within finite steps where the iterative ADP algorithm is initialized by an arbitrary admissible control sequence. The main contributions of this paper are summarized as follows:

(1) Propose a new proof that the iterative ADP algorithm initialized by an arbitrary admissible control sequence can converge to the optimum.
(2) Prove that the ϵ-optimal control can make the performance index function converge to the greatest lower bound of all performance indices within an error ϵ for unfixed initial state and the initial condition requirements in Wei and Liu (2011a, 2011b) is not needed.
(3) Obtain the length (number of steps) of the ϵ-optimal control.

2. Problem statement

In this paper, we consider the following discrete-time nonlinear systems

$$x_{k+1} = F(x_k, u_k), \quad k = 0, 1, 2, \ldots, \tag{1}$$

where, $x_k \in \mathbb{R}^n$ is the state and $u_k \in \mathbb{R}^m$ is the control vector. Let $x_0 \in \Omega_0$ be the initial state where $\Omega_0 \subset \mathbb{R}^n$ is the domain of initial states. Let the system function $F(x_k, u_k)$ be continuous for $\forall\, x_k, u_k$ and $F(0, 0) = 0$. We will study the optimal control problems for system (1) with finite-horizon and unspecified terminal time. The performance index function for state x_0 under the control sequence $\underline{u}_0^{N-1} = (u_0, u_1, \ldots, u_{N-1})$ is defined as

$$J(x_0, \underline{u}_0^{N-1}) = \sum_{k=0}^{N-1} U(x_k, u_k), \tag{2}$$

where $U(x_k, u_k) \geq 0$, for $\forall\, x_k, u_k$, is the positive semidefinite utility function.

Let $\underline{u}_0^{N-1} = (u_0, u_1, \ldots, u_{N-1})$ be a finite sequence of controls. We call the number of elements in the control sequence $\underline{u}_0^{N-1}$. Define the length of $\underline{u}_0^{N-1}$ as $|\underline{u}_0^{N-1}| = N$. We denote the final state of the trajectory as $x^{(f)}(x_0, \underline{u}_0^{N-1})$, i.e., $x^{(f)}(x_0, \underline{u}_0^{N-1}) = x_N$. For $\forall k \geq 0$, the finite control sequence can be written as $\underline{u}_k^{k+i-1} = (u_k, u_{k+1}, \ldots, u_{k+i-1})$ where $i \geq 1$. The final state can be written as $x^{(f)}(x_k, \underline{u}_k^{k+i-1})$ where $x^{(f)}(x_k, \underline{u}_k^{k+i-1}) = x_{k+i}$. Let $\underline{u}_k$ be an arbitrary finite-horizon admissible control sequence starting at k. Let $\mathfrak{A}_{x_k} = \left\{\underline{u}_k : x^{(f)}\left(x_k, \underline{u}_k\right) = 0\right\}$. Let

$$\mathfrak{A}_{x_k}^{(i)} = \left\{\underline{u}_k^{k+i-1} : x^{(f)}\left(x_k, \underline{u}_k^{k+i-1}\right) = 0,\ |\underline{u}_k^{k+i-1}| = i\right\}$$

be the set of all finite-horizon admissible control sequences of x_k with length i. Then, $\mathfrak{A}_{x_k} = \cup_{1 \leq i < \infty} \mathfrak{A}_{x_k}^{(i)}$. By this notation, a state x_k is controllable if and only if $\mathfrak{A}_{x_k} \neq \emptyset$. Define the optimal performance index function as

$$J^*(x_k) = \min_{\underline{u}_k} \left\{J(x_k, \underline{u}_k) : \underline{u}_k \in \mathfrak{A}_{x_k}\right\}. \tag{3}$$

Then, according to Bellman's principle of optimality (Bellman, 1957), $J^*(x_k)$ satisfies the discrete-time HJB equation

$$J^*(x_k) = \min_{u_k} \left\{U(x_k, u_k) + J^*(F(x_k, u_k))\right\}, \tag{4}$$

and the law of optimal control vector is given by

$$u^*(x_k) = \arg\min_{u_k} \left\{U(x_k, u_k) + J^*(F(x_k, u_k))\right\}. \tag{5}$$

3. Properties of the iterative adaptive dynamic programming algorithm

In this section, a new iterative ADP algorithm is developed to obtain the finite horizon optimal controller for nonlinear systems (1). The goal of the present iterative ADP algorithm is to construct an optimal control law $u^*(x_k), k = 0, 1, \ldots$, which drives the system from an arbitrary initial state x_0 to the singularity 0 within finite time, and simultaneously minimizes the performance index function. Convergence proofs will also be given to show that the performance index function converges to the optimum.

3.1. Derivation of the iterative ADP algorithm

In the iterative ADP algorithm, the performance index function and control policy are updated by recurrent iterations, with the iteration index number i increasing from 0. Assume that x_k is controllable. There exists a finite horizon admissible control sequence $\underline{u}_k^{k+i-1} = \{u_k, u_{k+1}, \ldots, u_{k+i-1}\} \in \mathfrak{A}_{x_k}^{(i)}$ that makes $x^{(f)}(x_k, \underline{u}_k^{k+i-1}) = x_{k+i} = 0$. Let $\underline{v}_k^{N-1} = \{v_k, v_{k+1}, \ldots, v_{N-1}\}$ be an arbitrary admissible sequence in $\mathfrak{A}_{x_k}^{(N-k)}$, where N is an unspecified terminal time. Define $\Phi(x_k)$ as the performance index function constructed by $\underline{v}_k^{N-1}$ which can be expressed by

$$\Phi(x_k) = J(x_k, \underline{v}_k^{N-1}). \tag{6}$$

Let $V_0(x_k) = \Phi(x_k)$ and the iterative performance index function $V_1(x_k)$ can be updated as

$$\begin{aligned} V_1(x_k) &= \min_{u_k} \{U(x_k, u_k) + V_0(x_{k+1})\} \\ &= U(x_k, v_1(x_k)) + V_0(F(x_k, v_1(x_k))), \end{aligned} \tag{7}$$

where the iterative control policy $v_1(x_k)$ is obtained as

$$\begin{aligned} v_1(x_k) &= \arg\min_{u_k} \{U(x_k, u_k) + V_0(x_{k+1})\} \\ &= \arg\min_{u_k} \{U(x_k, u_k) + V_0(F(x_k, u_k))\}. \end{aligned} \tag{8}$$

For $i = 1, 2, \ldots$, the iterative ADP algorithm will calculate the iterative performance index function as

$$\begin{aligned} V_i(x_k) &= \min_{u_k} \{U(x_k, u_k) + V_{i-1}(x_{k+1})\} \\ &= U(x_k, v_i(x_k)) + V_{i-1}(F(x_k, v_i(x_k))) \end{aligned} \tag{9}$$

where the iterative control policy $v_i(x_k)$ is computed as

$$\begin{aligned} v_i(x_k) &= \arg\min_{u_k} \{U(x_k, u_k) + V_{i-1}(x_{k+1})\} \\ &= \arg\min_{u_k} \{U(x_k, u_k) + V_{i-1}(F(x_k, u_k))\}. \end{aligned} \tag{10}$$

Remark 1. The present iterative ADP algorithm (7)–(10) is different from the iterative ADP algorithm proposed in Wei and Liu (2011a, 2011b). In Wei and Liu (2011a, 2011b), it is required that for $\forall x_k \in \mathbb{R}^n$, there exists a control u_k that makes $F(x_k, u_k) = 0$ to initialize the iterative ADP algorithm. It is well known that, for general control systems, especially for nonlinear systems, there may not exist a control that makes $F(x_k, u_k) = 0$ hold for $\forall x_k$. So the initial condition for the iterative ADP algorithm in Wei and Liu (2011a, 2011b) limits its applications. In this paper, the constraints in Wei and Liu (2011a, 2011b) are removed and the proposed iterative ADP algorithm begins with a performance index function constructed by an arbitrary finite horizon admissible control sequence. Thus, we can say that the proposed iterative ADP algorithm in this paper is more effective than prior results.

Theorem 1. *Let x_k be an arbitrary state. Let the iterative performance index function $V_i(x_k)$ be obtained according to (7)–(10). Then, we have*

$$\begin{aligned} V_i(x_k) &= \sum_{j=0}^{i-1} U\left(x_{k+j}, v_{i-j}(x_{k+j})\right) + \Phi(x_{k+i}) \\ &= \min_{\underline{u}_k^{k+i-1}} \left\{\sum_{j=0}^{i-1} U(x_{k+j}, u_{k+j})\right\} + \Phi(x_{k+i}). \end{aligned} \tag{11}$$

Proof. For $i = 0, 1, \ldots$, according to the definition of $V_i(x_k)$ in (7)–(9), we can get

$$\begin{aligned} V_i(x_k) &= \min_{u_k}\{U(x_k, u_k) + V_{i-1}(x_{k+1})\} \\ &= \min_{u_k}\Big\{U(x_k, u_k) + \min_{u_{k+1}}\Big\{U(x_{k+1}, u_{k+1}) \\ &\quad + \cdots + \min_{u_{k+i-1}}\{U(x_{k+i-1}, u_{k+i-1}) + V_0(x_{k+i})\}\cdots\Big\}\Big\} \end{aligned}$$

where $V_0(x_{k+i}) = \Phi(x_{k+i})$. According to the optimality principle, we obtain

$$\begin{aligned} V_i(x_k) &= \min_{\underline{u}_k^{k+i-1}}\Big\{U(x_k, u_k) + U(x_{k+1}, u_{k+1}) \\ &\quad + \cdots + U(x_{k+i-1}, u_{k+i-1}) + V_0(x_{k+i})\Big\} \\ &= \min_{\underline{u}_k^{k+i-1}}\left\{\sum_{j=0}^{i-1} U(x_{k+j}, u_{k+j})\right\} + \Phi(x_{k+i}). \end{aligned} \tag{12}$$

According to (9), we have

$$V_i(x_k) = \sum_{j=0}^{i-1} U\big(x_{k+j}, v_{i-j}(x_{k+j})\big) + \Phi(x_{k+i}). \tag{13}$$

The proof is complete. □

We can see that the optimal performance index function $J^*(x_k)$ in the HJB Eq. (4) is changed to a sequence of iterative performance index functions $V_i(x_k)$. From Theorem 1, $V_i(x_k)$ can be obtained by solving an optimal control problem with terminal constraint. Obviously, $V_i(x_k)$ does not necessarily satisfy the HJB Eq. (4). Convergence analysis of the iterative performance index function $V_i(x_k)$ is required.

3.2. Properties of the iterative ADP algorithm

In the above subsection, we can see that the performance index function $J^*(x_k)$ solved by HJB Eq. (4) is replaced by a sequence of performance index functions $V_i(x_k)$ and the optimal control law $u^*(x_k)$ is replaced by a sequence of control laws $v_i(x_k)$, where $i \geq 1$ is the index of iteration. We can prove that $J^*(x_k)$ defined in (3) is the limit of $V_i(x_k)$ as $i \to \infty$.

Lemma 1 (*Zhang et al., 2008*)**.** *Let the performance index function $V_i(x_k)$ be defined by (9). Let $\underline{\mu}_k = (\mu_k, \mu_{k+1}, \ldots) \in \mathfrak{A}_{x_k}$ be an arbitrary admissible control sequence. Define a new performance index function $P_i(x_k)$ as*

$$P_i(x_k) = U(x_k, \mu_k) + P_{i-1}(x_{k+1}), \tag{14}$$

with $P_0(x_k) = V_0(x_k) = \Phi(x_k)$, then we have $V_i(x_k) \leq P_i(x_k)$, $\forall\, i = 0, 1, \ldots$.

Theorem 2. *Let x_k be an arbitrary state vector and $V_0(x_k) = \Phi(x_k)$ be defined by (6). Then, the performance index function $V_i(x_k)$ obtained by (7)–(10) is a monotonically nonincreasing sequence for $\forall\, i \geq 1$, i.e., $V_{i+1}(x_k) \leq V_i(x_k)$ for $\forall\, i = 0, 1, \ldots$.*

Proof. We prove this conclusion by mathematical induction. First, we let $i = 0$. Let $\underline{\mu}_k \in \mathfrak{A}_{x_k}$ be an arbitrary admissible control sequence. Define the performance index function $P_i(x_k)$ as (14). For $i = 0$, we have

$$P_1(x_k) = U(x_k, \mu_k) + P_0(x_{k+1}) = U(x_k, \mu_k) + \Phi(x_{k+1}). \tag{15}$$

According to the definition of $\Phi(x_k)$ in (6), we have

$$\begin{aligned} \Phi(x_k) - \Phi(F(x_k, \nu_k)) &= J(x_k, \underline{\nu}_k^{N-1}) - J(x_{k+1}, \underline{\nu}_{k+1}^{N-1}) \\ &= U(x_k, \nu_k). \end{aligned} \tag{16}$$

As $\underline{\mu}_k \in \mathfrak{A}_{x_k}$ is arbitrary, let $\nu_k = \mu_k$, and then we can obtain

$$\begin{aligned} V_0(x_k) &= \Phi(x_k) \\ &= U(x_k, \nu_k) + \Phi(F(x_k, \nu_k)) = P_1(x_k) \end{aligned} \tag{17}$$

holds. On the other hand, according to Lemma 1, we have

$$\begin{aligned} V_1(x_k) &= \min_{u_k}\{U(x_k, u_k) + V_0(x_{k+1})\} \\ &\leq P_1(x_k) \\ &= U(x_k, \nu_k) + P_0(x_{k+1}) \\ &= V_0(x_k) \end{aligned} \tag{18}$$

which proves $V_0(x_k) \geq V_1(x_k)$.

Hence, the conclusion holds for $i = 0$. Assume that the conclusion holds for $i = l-1$, where $l = 1, 2, \ldots$. The performance index function $P_{l+1}(x_k)$ is given as

$$\begin{aligned} P_{l+1}(x_k) &= U(x_k, v_{l-1}(x_k)) + U(x_{k+1}, v_{l-2}(x_{k+1})) \\ &\quad + \cdots + U(x_{k+l-1}, v_1(x_{k+l-1})) \\ &\quad + U(x_{k+l}, \mu_{k+l}) + P_0(x_{k+l+1}) \\ &= \sum_{j=0}^{l-1} U\big(x_{k+j}, v_{l-j-1}(x_{k+j})\big) \\ &\quad + U(x_{k+l}, \mu_{k+l}) + \Phi(x_{k+l+1}). \end{aligned} \tag{19}$$

According to Theorem 1, we have the iterative performance index function $V_l(x_k)$ expressed as

$$V_l(x_k) = \sum_{j=0}^{l-1} U\big(x_{k+j}, v_{l-j}(x_{k+j})\big) + \Phi(x_{k+l}). \tag{20}$$

According to (6), we have

$$\Phi(x_{k+l}) = U(x_{k+l}, \nu_{k+l}) + \Phi(F(x_{k+l}, \nu_{k+l})). \tag{21}$$

Let $\nu_{k+l} = \mu_{k+l}$, and then according to (19) and (20), we can get

$$\begin{aligned} V_l(x_k) &= \sum_{j=0}^{l-1} U\big(x_{k+j}, v_{l-j-1}(x_{k+j})\big) + \Phi(x_{k+l}) \\ &= \sum_{j=0}^{l-1} U\big(x_{k+j}, v_{l-j-1}(x_{k+j})\big) + U(x_{k+l}, \nu_{k+l}) + \Phi(x_{k+l+1}) \\ &= P_{l+1}(x_k). \end{aligned} \tag{22}$$

According to Lemma 1, we have $V_{l+1}(x_k) \leq P_{l+1}(x_k)$. Therefore, we can obtain $V_{l+1}(x_k) \leq V_l(x_k)$. □

Remark 2. For the iterative ADP algorithm proposed in Wang et al. (2011) and Wei and Liu (2011a, 2011b), the iterative performance index function $V_i(x_k)$ reaches the max value at $i = 1$. So, the iterative ADP algorithm is not monotonically nonincreasing for $i = 0, 1, \ldots$. While for the iterative ADP algorithm (7)–(10), we have the iterative ADP algorithm monotonically nonincreasing for $\forall i = 0, 1, \ldots$. Therefore, it is another difference between the two iterative ADP algorithms.

From Theorem 2, we know that the performance index function $V_i(x_k) \geq 0$ is a monotonically nonincreasing sequence with lower bound for iteration index $i = 1, 2, \ldots$. Then, there exists a limit of iterative performance index function $V_i(x_k)$. Define the

performance index function $V_\infty(x_k)$ as the limit of the iterative function $V_i(x_k)$, i.e.,

$$V_\infty(x_k) = \lim_{i\to\infty} V_i(x_k). \tag{23}$$

Next, we will prove that the performance index function $V_i(x_k)$ converges to the optimal performance index function $J^*(x_k)$ as $i \to \infty$.

Lemma 2. *Let x_k be an arbitrary controllable state and $\underline{v}_k^{N-1}$ be an arbitrary admissible sequence, where N is an unspecified terminal time. Assume $U(x_k, u_k) \geq 0$ is the positive semidefinite for $\forall x_k, u_k$. Let $\Phi(x_k)$ be defined by (6), and then we have $\Phi(x_k)$ is positive semidefinite for $\forall x_k$.*

Proof. The lemma can be easily proved by mathematical induction and the detailed proofs are omitted here. □

Theorem 3. *Let the performance index function $V_i(x_k)$ be defined by (9). If the system state x_k is controllable, then the performance index function $V_i(x_k)$ converges to the optimal performance index function $J^*(x_k)$ as $i \to \infty$, i.e.,*

$$V_i(x_k) \to J^*(x_k). \tag{24}$$

Proof. According to the definition of $J^*(x_k)$ in (3), we have $J^*(x_k) \leq V_i(x_k)$. Let $i \to \infty$, and then we have

$$J^*(x_k) \leq V_\infty(x_k). \tag{25}$$

Next, as $P_q(x_k) - J^*(x_k) \geq 0$ and $\underline{\mu}_k$ is arbitrary, then taking $\underline{\mu}_k^{k+q-1} = \underline{u}_k^{*k+q-1}$ into $P_i(x_k)$ in (14), we can obtain $\Phi(x_{k+q}) - \sum_{j=q}^{N-1} U(x_{k+j}, u^*_{k+j}) \geq 0$, where N is the unspecified terminal time. As $\underline{\mu}_k$ is an admissible control sequence, we have $x_{k+q} \to 0$ as $q \to \infty$. According to Lemma 2, we know that $\Phi(x_{k+q}) \to 0$ as $q \to \infty$. Let $\epsilon > 0$ be an arbitrary positive number. There exists a finite horizon admissible control sequence η_q such that

$$P_q(x_k) \leq J^*(x_k) + \epsilon. \tag{26}$$

On the other hand, according to Lemma 1, for any finite horizon admissible control η_q, we have

$$V_\infty(x_k) \leq V_q(x_k) \leq P_q(x_k). \tag{27}$$

Combining (26) and (27), we have $V_\infty(x_k) \leq J^*(x_k) + \epsilon$. As ϵ is an arbitrary positive number, we have

$$V_\infty(x_k) \leq J^*(x_k). \tag{28}$$

According to (25) and (28), we have

$$V_\infty(x_k) = J^*(x_k). \tag{29}$$

The proof is complete. □

4. The ϵ-optimal control algorithm

In the previous section, we proved that the iterative performance index function $V_i(x_k)$ converges to the optimal performance index function $J^*(x_k)$ as $i \to \infty$. This means that if we want to obtain the optimal performance index function $J^*(x_k)$, we should run the iterative ADP algorithm (7)–(10) for $i \to \infty$. But unfortunately, it is untenable to run the algorithm for infinite number of times. For finite horizon optimal control, the infinite horizon ADP algorithm may not be effective. First, the infinite horizon optimal control makes the iterative performance index function converge to the optimum as $i \to \infty$, and the optimal control law is also convergent to the optimum as $i \to \infty$. While for the finite horizon optimal control problem, for different initial state x_k, we should adopt a different optimal control law. Second, the number of steps of optimal control for the infinite horizon optimal control is infinite. While for the finite horizon control problem, for different initial states, the optimal step number is also different. Hence, to overcome this difficulty, a new ϵ-optimal control algorithm is established in this section.

4.1. Derivation of the ϵ-optimal control algorithm

In this subsection, we will introduce our method for iterative ADP with the consideration of the length of control sequences. For different x_k, we will use different i for the length of optimal control sequence. For a given error bound $\epsilon > 0$, the number i will be chosen so that the error between $J^*(x_k)$ and $V_i(x_k)$ is bounded with ϵ.

Theorem 4. *Let $\epsilon > 0$ be any small number and x_k be any controllable state. Let the performance index function $V_i(x_k)$ be defined by (9) and $J^*(x_k)$ be the optimal performance index function. Then, there exists a finite i satisfying*

$$|V_i(x_k) - J^*(x_k)| \leq \epsilon. \tag{30}$$

Definition 1. Let x_k be a controllable state vector. Let $\epsilon > 0$ be a small positive number. The approximate length of optimal control with respect to ϵ is defined as

$$K_\epsilon(x_k) = \min\{i : |V_i(x_k) - J^*(x_k)| \leq \epsilon\}. \tag{31}$$

Remark 3. An important property we must point out. For the iterative ADP algorithm (7)–(10), we have proved that for arbitrary initial performance index function $V_0(x_k) = \Phi(x_k)$, the iterative performance index function $V_i(x_k) \to J^*(x_k)$ as $i \to \infty$. For the finite horizon iterative ADP algorithm, the length $K_\epsilon(x_k)$ is different for different initial performance index function $\Phi(x_k)$, which makes it difficult to obtain the ϵ-optimal control law and the approximate length.

Next, we will show that if we give a constraint for the initial performance index function $V_0(x_k)$, we can get that $K_\epsilon(x_k)$ is unique.

Theorem 5. *Let $\underline{u}_{k+1}^{k+l*} = \{u^*_{k+1}, \ldots, u^*_{k+l}\}$ be the optimal control sequence and $\Phi^*(x_{k+1}) = J(x_{k+1}, \underline{u}_{k+1}^{k+l*})$. If we let $V_0(x_{k+1}) = \Phi^*(x_{k+1})$, then we have*

$$V_i(x_k) = J(x_k, \underline{u}_k^{k+l+i-1*}). \tag{32}$$

Proof. According to Theorem 1, the iterative performance index function $V_i(x_k)$ can be expressed as in (12). As $\Phi^*(x_{k+1}) = J^*(x_{k+1}, \underline{u}_{k+1}^{k+l-1*})$, we have

$$\Phi^*(x_{k+1}) = \min_{\underline{u}_{k+1}^{k+l}} \sum_{j=1}^{l} U(x_{k+j}, u_{k+j}). \tag{33}$$

Putting $\Phi^*(x_{k+1})$ into (12), we can obtain

$$\begin{aligned} V_i(x_k) &= \min_{\underline{u}_k^{k+l+i-1}} \left\{ \sum_{j=0}^{l+i-1} U(x_{k+j}, u_{k+j}) \right\} \\ &= J(x_k, \underline{u}_k^{k+l+i-1*}). \end{aligned} \tag{34}$$

The proof is complete. □

From Theorem 5, we can see that if we can find an optimal control sequence $\underline{u}_{k+1}^{k+l*}$, then we can obtain the optimal control law and the optimal control length for the state x_k. While according to Theorem 3, we know that it requires to run the iterative ADP algorithm (7)–(10) for an infinite number of times to obtain $J^*(x_{k+1})$ which is impossible to realize in the real applications. Therefore, we give an ϵ-optimal control algorithm to obtain the approximate optimal performance index function and control law. Before introducing the ϵ-optimal control algorithm, the following definition and lemma are necessary.

Definition 2. Let x_k be a controllable state vector and ϵ be a positive number. For $i = 1, 2, \ldots$, define the set

$$\mathcal{T}_i^{(\epsilon)} = \{x_k \in \mathcal{T}_\infty : K_\epsilon(x_k) \le i\}. \tag{35}$$

When $x_k \in \mathcal{T}_i^{(\epsilon)}$, to find the optimal control sequence which has performance index less than or equal to $J^*(x_k) + \epsilon$, we only need to consider the control sequences $\underline{u}_k$ with length $|\underline{u}_k| \le i$. The set $\mathcal{T}_i^{(\epsilon)}$ has the following properties.

Lemma 3 (*Wang et al., 2011*)**.** *Let $\epsilon > 0$ and $i = 1, 2, \ldots$. Then,*

(i) $x_k \in \mathcal{T}_i^{(\epsilon)}$ *if and only if* $V_i(x_k) \le J^*(x_k) + \epsilon$;
(ii) $\mathcal{T}_i^{(\epsilon)} \subseteq \mathcal{T}_i$;
(iii) $\mathcal{T}_i^{(\epsilon)} \subseteq \mathcal{T}_{i+1}^{(\epsilon)}$;
(iv) $\cup_i \mathcal{T}_i^{(\epsilon)} = \mathcal{T}_\infty$;
(v) *If* $\epsilon > \delta > 0$, *then* $\mathcal{T}_i^{(\epsilon)} \supseteq \mathcal{T}_i^{(\delta)}$.

Next, we will introduce the ϵ-optimal control iterative ADP algorithm. First, let $\widehat{\underline{u}}_0^{K-1} = (u_0, u_1, \ldots, u_{K-1})$ be an arbitrary finite horizon admissible control sequence and the corresponding state sequence is $\widehat{\underline{x}}_0^K = (x_0, x_1, \ldots, x_K)$ where $x_K = 0$. We can see that the initial control sequence $\widehat{\underline{u}}_0^{K-1} = (u_0, u_1, \ldots, u_{K-1})$ may not be optimal which means the initial number of control steps K may not be optimal and also the law of the initial control sequence $\widehat{\underline{u}}_0^{K-1}$ may not be optimal. In the following, we will show how the number of control steps and the control law are both optimized in the iterative ADP algorithm simultaneously.

For the state x_{K-1}, we have $F(x_{K-1}, u_{K-1}) = 0$. Then we run the iterative ADP algorithm proposed in Wang et al. (2011) and Wei and Liu (2011a, 2011b) at x_{K-1} until

$$\left|V_{l_0}(x_{K-1}) - J^*(x_{K-1})\right| \le \epsilon \tag{36}$$

holds where $l_0 > 0$ is a positive integer. This means $x_{K-1} \in \mathcal{T}_{l_0}^{(\epsilon)}$ and the number of optimal control steps $K_\epsilon(x_{K-1}) = l_0$.

Then, considering $x_{K-j}, j = 1, 2, \ldots, K$, we have $F(x_{K-j}, u_{K-j}) = x_{K-j+1}$. For x_{K-j}, if

$$\left|V_{l_{j-1}}(x_{K-j}) - J^*(x_{K-j})\right| \le \epsilon \tag{37}$$

holds, then we say $x_{K-j} \in \mathcal{T}_{l_{j-1}}^{(\epsilon)}$, and $v_{l_{j-1}}(x_{K-j})$ is the corresponding ϵ-optimal control law. If not, $x_{K-j} \notin \mathcal{T}_{l_{j-1}}^{(\epsilon)}$ and then we run the iterative ADP algorithm as

$$v_{l_{j-1}+1}(x_{K-j}) = \arg\min_{u_{K-j}} \left\{U(x_{K-j}, u_{K-j}) + V_{l_{j-1}}(x_{K-j+1})\right\} \tag{38}$$

and

$$V_{l_{j-1}+1}(x_{K-j}) = U(x_{K-j}, v_{l_{j-1}+1}(x_{K-j})) + V_{l_{j-1}}(F(x_{K-j}, v_{l_{j-1}+1}(x_{K-j}))). \tag{39}$$

For $i = 1, 2 \ldots$, the iterative ADP algorithm between

$$v_{l_{j-1}+i+1}(x_{K-j}) = \arg\min_{u_{K-j}} \left\{U(x_{K-j}, u_{K-j}) + V_{l_{j-1}+i}(x_{K-j+1})\right\} \tag{40}$$

and

$$V_{l_{j-1}+i+1}(x_{K-j}) = U(x_{K-j}, v_{l_{j-1}+i}(x_{K-j})) + V_{l_{j-1}+i}(F(x_{K-j}, v_{l_{j-1}+i+1}(x_{K-j}))) \tag{41}$$

until the following inequality

$$\left|V_{l_j}(x_{K-j}) - J^*(x_{K-j})\right| \le \epsilon \tag{42}$$

holds where $l_j > 0$ is a positive integer. So we can obtain $x_{K-j} \in \mathcal{T}_{l_j}^{(\epsilon)}$ and the number of optimal control steps $K_\epsilon(x_{K-j}) = l_j$.

4.2. Properties of the ϵ-optimal control algorithm

We can see that an error ϵ between $J^*(x_k)$ and $V_i(x_k)$ is introduced into the iterative ADP algorithm which makes the performance index function $V_i(x_k)$ converge within finite iteration step i. In this subsection, we will show that the corresponding control is also an effective control that makes the performance index function reach the optimal within error bound ϵ. According to Lemma 3, we have the following theorem.

Theorem 6. *Let $\epsilon > 0$ and $i = 0, 1, 2, \ldots$. If $x_k \in \mathcal{T}_i^{(\epsilon)}$, then for any $x_k' \in \mathcal{T}_i^{(\epsilon)}$, we have the following inequality*

$$|V_i(x_k') - J^*(x_k')| \le \epsilon. \tag{43}$$

Proof. The theorem can be easily proved by contradiction. Assume the conclusion is false. Then for some $x_k' \in \mathcal{T}_i^{(\epsilon)}$, we have

$$V_i(x_k') > J^*(x_k') + \epsilon. \tag{44}$$

So we can get

$$K_\epsilon(x_k') = \min\{j : |V_j(x_k') - J^*(x_k')| \le \epsilon\} > i. \tag{45}$$

Then, according to Definition 2, we can obtain $x_k' \notin \mathcal{T}_i^{(\epsilon)}$ which contradicts the assumption of $x_k' \in \mathcal{T}_i^{(\epsilon)}$. So the conclusion holds. □

Corollary 1. *For $i = 1, 2, \ldots$, let $\mu_\epsilon^i(x_k)$ be expressed as*

$$\mu_\epsilon^i(x_k) = \arg\min_{u_k} \{U(x_k, u_k) + V_{i-1}(F(x_k, u_k))\} \tag{46}$$

that makes the performance index function (30) hold for $x_k \in \mathcal{T}_i^{(\epsilon)}$. Then for $x_k' \in \mathcal{T}_i^{(\epsilon)}$, $\mu_\epsilon^i(x_k')$ satisfies

$$|V_i(x_k') - J^*(x_k')| \le \epsilon. \tag{47}$$

Then, we have the following theorem.

Theorem 7. *For $i = 1, 2, \ldots$, if we let $x_k \in \mathcal{T}_i^{(\epsilon)}$ and $\mu_\epsilon^i(x_k)$ be expressed in (46), then $F(x_k, \mu_\epsilon^i(x_k)) \in \mathcal{T}_{i-1}^{(\epsilon)}$. In other words, if $K_\epsilon(x_k) = i$, then we have $K_\epsilon(F(x_k, \mu_\epsilon^i(x_k))) \le i - 1$.*

Proof. The detailed proof can be seen in Wei and Liu (2011a, 2011b). □

Corollary 2. *For $i = 0, 1, \ldots$, let $\mu_\epsilon^i(x_k)$ be expressed in (46) where the performance index function $|V_i(x_k) - J^*(x_k)| \le \epsilon$. Then for any $x_k' \in \mathcal{T}_i^{(\epsilon)}$, we have the following inequality*

$$|V_i(x_k') - J^*(x_k')| \le \epsilon. \tag{48}$$

Now we look at the optimal control problem with respect to performance index function. If the initial state x_0 is fixed, we will show that if we choose x_0 to run the iterative index function we can obtain the ϵ-optimal control.

Theorem 8. *Let x_0 be the fixed initial state, $\mu_\epsilon^i(x_0)$ satisfies (46) at $k = 0$. If $x_k, k = 0, 1, \ldots$, is the state under the control law $\mu_\epsilon^i(x_k)$, then we have $|V_i(x_k) - J^*(x_k)| \le \epsilon$ for any k.*

Proof. For the system (1) with respect to the performance index function (2), we have $x_0 \in \mathcal{T}_i^{(\epsilon)}$ and $K_\epsilon(x_0) = i$. Then for small ϵ, there exists an ϵ-optimal control law $\mu_\epsilon^i(x_0)$ which stabilizes the system (1) within finite time N and minimizes the performance index function (2). Then obviously, we have $x_i \in \mathcal{T}_0, x_{i-1} \in \mathcal{T}_1^{(\epsilon)}, \ldots, x_0 \in \mathcal{T}_i^{(\epsilon)}$ where $0 = \mathcal{T}_0 \subseteq \mathcal{T}_1^{(\epsilon)} \subseteq \cdots \subseteq \mathcal{T}_i^{(\epsilon)}$. So according to Theorem 7 and Corollary 2, we have that the ϵ-optimal control law μ_ϵ^i obtained by the initial state x_0 using the iterative ADP algorithm is effective for the states $x_1, x_2, \ldots, x_i$. The proof is complete. □

We can see that if we choose x_0 to run the iterative index function we can obtain the ϵ-optimal control. While if the initial state x_0 is unfixed, then we do not know which one should be used to implement the iterative ADP algorithm. In the next section, we will solve this problem.

4.3. The ϵ-optimal control algorithm for unfixed initial state

For a lot of practical nonlinear systems, the initial state x_0 cannot be fixed. Instead, the initial state belongs to a set and we define the domain of initial states as Ω_0 where $\Omega_0 \subseteq \mathbb{R}^n$. Then, we have $x_0 \in \Omega_0$. For this case, if we only choose one state $x_0^{(i)} \in \Omega_0$ to run the iterative ADP algorithm and get corresponding ϵ-optimal control μ_ϵ^i, then the ϵ-optimal control μ_ϵ^i may not be ϵ-optimal for all $x_0 \in \Omega_0$ because there may exist a state $x_0^{(j)} \in \Omega_0$ such that $x_0^{(i)} \in \mathcal{T}_i^{(\epsilon)}$ while $x_0^{(j)} \in \mathcal{T}_j^{(\epsilon)} \setminus \mathcal{T}_i^{(\epsilon)}$ where $j > i$. If we let

$$I = \max\left\{i : x_0 \in \mathcal{T}_i^{(\epsilon)} \text{ s.t. } x_0 \in \Omega_0\right\} \tag{49}$$

then according to Corollary 2, we should find the initial state $x_0 \in \mathcal{T}_I^{(\epsilon)}$ to obtain the most effective ϵ-optimal control. Thus, the next job is to obtain the state $x_0 \in \mathcal{T}_I^{(\epsilon)}$. For this case, there are two methods which are the "entire state space searching method" and the "partial state space searching method" to obtain the ϵ-optimal control $\mu_\epsilon^i(x_k)$ for $k = 0, 1, \ldots$.

(I) *Entire state space searching method.*

Choosing randomly an array of enough states

$$X = \left(x^{(1)}, x^{(2)}, \ldots, x^{(Q)}\right) \tag{50}$$

from the entire initial state space Ω, where $Q > 0$ is a positive integer number. First, we solve (7) where $x_k = x^{(1)}, x^{(2)}, \ldots, x^{(Q)}$, respectively and $V_0(x_{k+1})$ is the converged iterative performance index function obtained by (36)–(42) at x_{k+1}. If for $0 \le j_1 \le Q$ and $x_k = x^{(j_1)} \in X$, the inequality

$$|V_1(x^{(j_1)}) - J^*(x^{(j_1)})| \le \epsilon \tag{51}$$

holds, then we have $x^{(j_1)} \in \mathcal{T}_1^{(\epsilon)}$. We record the performance index function V_1 and let

$$X_1 = \left\{x^{(j_1)} \in X : |V_1(x^{(j_1)}) - J^*(x^{(j_1)})| \le \epsilon\right\}. \tag{52}$$

We can repeat the process (51)–(52) for iteration index $i = 1, 2, \ldots,$ to solve (9), where $x_k \in X \setminus \{X_1 \cup \cdots \cup X_{i-1}\}$. If for $0 \le j_i \le Q$ and $x_k = x^{(j_i)} \in X \setminus \{X_1 \cup \cdots \cup X_{i-1}\}$, the inequality

$$|V_i(x^{(j_i)}) - J^*(x^{(j_i)})| \le \epsilon \tag{53}$$

holds, then we have $x^{(j_i)} \in \mathcal{T}_i^{(\epsilon)}$. We record the performance index function V_i and let

$$X_i = \left\{x^{(j_i)} \in X \setminus \{X_1 \cup X_2 \cup \cdots \cup X_{i-1}\} : |V_i(x^{(j_i)}) - J^*(x^{(j_i)})| \le \epsilon\right\}. \tag{54}$$

For the initial state x_0, if $|V_i(x_0) - J^*(x_0)| \le \epsilon$ holds, then the ϵ-optimal performance index function is obtained and the corresponding control law is the ϵ-optimal control law μ_ϵ^i.

Remark 4. The structure of the entire state space searching method is clear and simple which is based on the idea of dynamic programming. This is the merit of the entire space searching method. But it also possesses serious shortcomings. First, the array of states X in (50) should include enough state points which is distributed for the entire initial state space Ω. Second, for each state point $x^{(j_i)} \in X$, the iterative algorithm (50)–(53) should run one time and then record X_i in (54). So the computational complexity is huge. Especially for neural network implement, it means the neural network should be trained at every state point for the entire state space to obtain the optimal control and the "curse of dimensionality" cannot be avoided. Therefore, the entire state space searching method is very difficult to apply to the optimal control problem of real-world systems.

(II) *Partial state space searching method*:

In the partial state space searching method, it is not necessary to search the entire state space to obtain the optimal control. Instead, only the boundary of the domain of initial states Ω is searched to obtain the ϵ-optimal control which overcomes the difficulty of the "curse of dimensionality" effectively.

Theorem 9. *Let $\Omega_0 \subseteq \mathbb{R}^n$ be the domain of initial states and the initial state $x_0 \in \Omega_0$. If Ω_0 is a convex set on $\mathbb{R}^n$, then $x_0^{(I)}$ is a boundary point of Ω_0 where I is defined in (49).*

Proof. The theorem can be proved by contradiction. Assume that $x_0^{(I)}$ is a interior point of Ω_0. Without loss of generality, let the point be $x_a = x_0^{(I)}$. Make a beeline between the origin and x_a. Let the point of intersection between the beeline and the boundary of the set Ω_0 be x_b. Let the point of intersection between the extended line and the boundary of the set Ω_0 be x_c. The situation of x_a, x_b and x_c is shown in Fig. 1. As $x_0^{(I)}$ is an interior point of convex set Ω_0, according to the property of convex set, for $\forall x_0^{(j)} \in \Omega_0, j = 0, 1, \ldots,$ there exists a positive real number $0 \le \lambda \le 1$ that makes

$$x_0^{(j_a)} = \lambda\, x_0^{(j_b)} + (1 - \lambda)\, x_0^{(j_c)} \tag{55}$$

hold, where j_a, j_b and j_c are nonnegative integer numbers.

If we let $x_a = x_0^{(I)} = x_0^{(j_a)}$, $x_b = x_0^{(j_b)}$ and $x_c = x_0^{(j_c)}$, then we have

$$x_a = \lambda\, x_b + (1 - \lambda)\, x_c. \tag{56}$$

If we assume that $x_a \in \mathcal{T}_a^{(\epsilon)}$, $x_b \in \mathcal{T}_b^{(\epsilon)}$ and $x_c \in \mathcal{T}_c^{(\epsilon)}$, then we have

$$\mathcal{T}_c^{(\epsilon)} \subseteq \mathcal{T}_a^{(\epsilon)} \tag{57}$$

because $x_a = x_0^{(I)}$ where I is expressed in (49). Then we can obtain

$$I = K_\epsilon(x_a) \ge K_\epsilon(x_c) = c. \tag{58}$$

On the other hand, as x_c is the point of intersection between the extended beeline and the boundary of the set Ω_0, obviously the point x_c is farther from the origin. So we have

$$K_\epsilon(x_a) \le K_\epsilon(x_c) = c \tag{59}$$

which is a contradiction to (58). So $x_0^{(I)}$ cannot be expressed as (55). Then the assumption is false and therefore $x_0^{(I)}$ must be the boundary point of the set Ω_0. □

Remark 5. Theorem 9 gives an important property of the optimal control law. It means that if the initial set Ω_0 is convex, it is not necessary to search all the state points of the set. Instead, it only requires to search the boundary of the set and therefore the computational burden is much reduced.

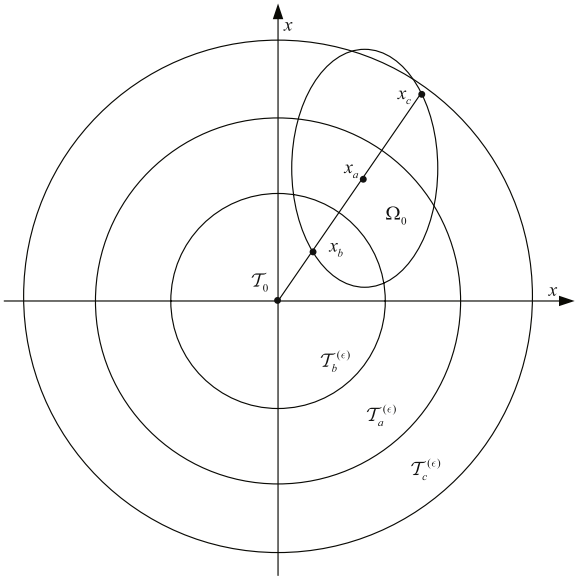

Fig. 1. The situation outline of the sates x_a, x_b and x_c.

4.4. The expressions of the ϵ-optimal control algorithm

In Wang et al. (2011), we analyzed the ϵ-optimal control iterative ADP algorithm when the initial state is fixed. In Wei and Liu (2011a, 2011b), we give an iterative ADP algorithm for unfixed initial state while it requires the control system can reach zero directly. In this paper, we propose a new ϵ-optimal control iterative ADP algorithm for unfixed initial state, while the strict initial condition in Wei and Liu (2011a, 2011b), can be omitted. In summary, the finite horizon ϵ-optimal control problem with finite time can be separated into four cases.

Case 1. The initial state x_0 is fixed and for any state $x_k \in \mathbb{R}^n$, there exits a control $u_k \in \mathbb{R}^m$ that stabilizes the state to zero directly (proposed in Wang et al. (2011)).

Case 2. The initial state $x_0 \in \Omega_0$ is unfixed and for any state $x_k \in \mathbb{R}^n$, there exits a control $u_k \in \mathbb{R}^m$ that stabilizes the state to zero directly (proposed in Wei and Liu (2011a, 2011b)).

Case 3. The initial state x_0 is fixed and $\exists x_k \in \mathbb{R}^n$ such that $F(x_k, u_k) = 0$ is no solution for $\forall u_k \in \mathbb{R}^m$ (proposed in Wang et al. (2011)).

Case 4. The initial state $x_0 \in \Omega_0$ is unfixed and $\exists x_k \in \mathbb{R}^n$ such that $F(x_k, u_k) = 0$ is no solution for $\forall u_k \in \mathbb{R}^m$ (proposed in this paper).

We can see that Cases 1–3 are special cases of Case 4. Therefore, we can say that the proposed iterative ADP algorithm is the most effective one. Given the preparations, we now summarize the iterative ADP algorithms as follows:

Step01. Give the initial state space Ω_0, the max iterative number $i_{\max}$ and the computation precision ϵ.

Step02. Let $\bar{\Omega}_0$ be the boundary of the domain of initial states Ω_0. Grid the set $\bar{\Omega}_0$ into $\bar{P}$ subsets which are expressed as $\bar{\Omega}_0^{(1)}$, $\bar{\Omega}_0^{(2)}, \ldots, \bar{\Omega}_0^{(\bar{P})}$ where $\bar{\Omega}_0 = \cup_{j=1}^{\bar{P}} \bar{\Omega}_0^{(j)}$ and $\bar{P} > 0$ is a positive integer number. For $j = 1, 2, \ldots, \bar{P}$, let X_0 be expressed as $X_0 = (x^{(1)}, \ldots, x^{(\bar{P})})$, and then $x_0^{(j)}$ satisfies $x_0^{(j)} \in \bar{\Omega}_0^{(j)}$.

Step03. For $j = 1, 2, \ldots, \bar{P}$, let $x_0 = x_0^{(j)}$ and loop (36)–(42).

Step04. For $x_0 = x_0^{(j)}$, obtain $x_0^{(j)} \in \mathcal{T}_{i_j}^{(\epsilon)}$. Record the performance index function $V_{i_j}(x_0^{(j)})$, and the control law $\mu_\epsilon^{i_j}(x_0^{(j)})$.

Step05. Let I be expressed as (49), we get $x_0^{(\bar{j})} \in \mathcal{T}_I^{(\epsilon)}$ and $K_\epsilon(x_0^{(\bar{j})}) = I$.

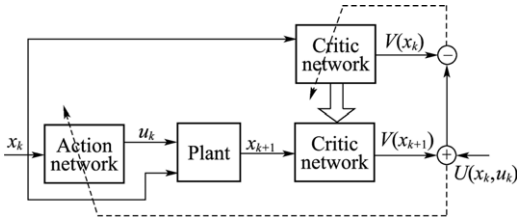

Fig. 2. The structure diagram of the algorithm.

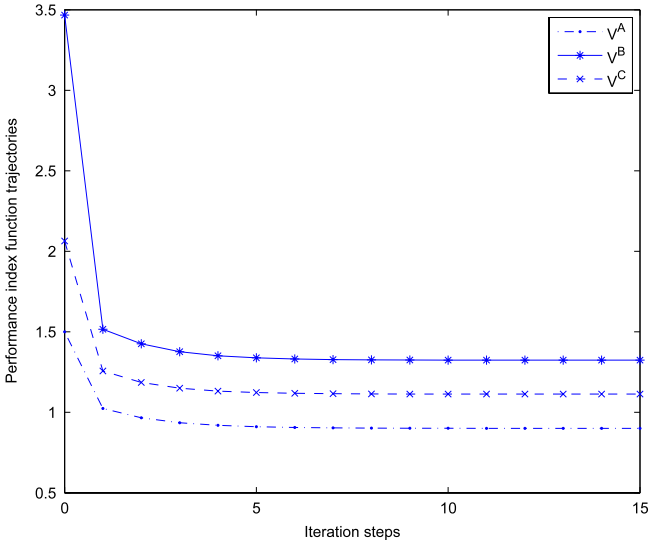

Fig. 3. The convergence of performance index functions.

Step06. Record the corresponding performance index function $V_I(x_0^{(\bar{j})})$, and the control law $\mu_\epsilon^I(x_0^{(\bar{j})})$.

Step07. Stop.

5. Neural network implementation for the ϵ-optimal control scheme

Assume that the number of hidden layer neurons is denoted by l, the weight matrix between the input layer and hidden layer is denoted by V, and the weight matrix between the hidden layer and output layer is denoted by W. Then, the output of three-layer NN is represented by:

$$\hat{F}(X, V, W) = W^T \sigma(V^T X) \tag{60}$$

where $\sigma(V^T X) \in R^l$, $[\sigma(z)]_i = \frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}}$, $i = 1, \ldots, l$, are the activation functions.

The NN estimation error can be expressed by

$$F(X) = F(X, V^*, W^*) + \varepsilon(X) \tag{61}$$

where V^*, W^* are the ideal weight parameters, and $\varepsilon(X)$ is the reconstruction error.

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

The training rules of the neural network can be seen in Si and Wang (2001) and Wei and Liu (2011a).

6. Simulation study

To evaluate the performance of our iterative ADP algorithm, we give an example with quadratic utility functions for numerical

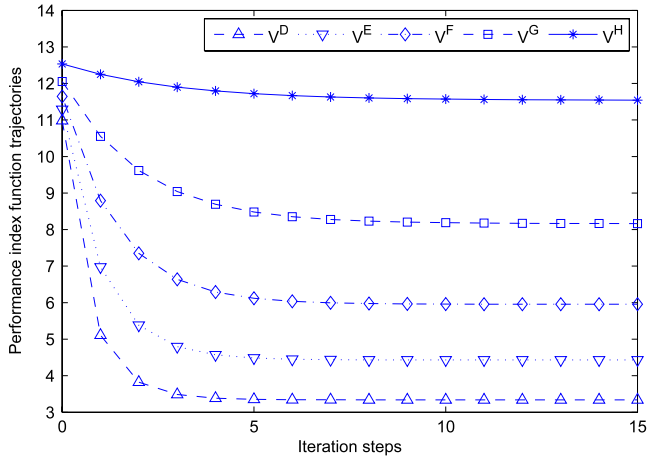

Fig. 4. The convergence of performance index functions.

experiment. Our example is also used in Wang et al. (2011) and Wei and Liu (2011a, 2011b). We consider the system

$$x_{k+1} = F(x_k, u_k) = x_k + \sin(0.1x_k^2 + u_k), \tag{62}$$

where $x_k, u_k \in \mathbb{R}$, and $k = 0, 1, 2, \ldots$. The domain of initial states is expressed as

$$\Omega_0 = \{x_0 | 0.8 \leq x_0 \leq 1.5\}. \tag{63}$$

The performance index function is in quadratic form with a finite-time horizon that is expressed as (2) with $U(x_k, u_k) = x_k^T Q x_k + u_k^T R u_k$, where the matrix $Q = R = E$ and E denotes the identity matrix with suitable dimensions.

We can see that for the initial state $0.8 \leq x_0 \leq 1$, there exists a control $u_0 \in \mathbb{R}$ that makes $x_1 = F(x_0, u_0) = 0$. Thus the situation then belongs to Case 2. While for the initial state $1 < x_0 \leq 1.5$, there does not exist a control $u_0 \in \mathbb{R}$ that makes $x_1 = F(x_0, u_0) = 0$. Thus the situation then belongs to Case 4. Then we will compute the ϵ-optimal control law for $0.8 \leq x_0 \leq 1$ and $1 < x_0 \leq 1.5$ separately. The computation error of the iterative ADP is given as $\epsilon = 10^{-6}$. The critic network and the action network are chosen as three-layer BP neural networks with the structure 1–8–1 and 1–8–1, respectively. For $0.8 \leq x_0 \leq 1$, we run the iterative ADP algorithm for Case 2. The search step is 0.1 from $x_k = 0.8$ to $x_k = 1$. We illustrate the convergence of performance index functions at 3 points which are $x_A = 0.8$, $x_B = 0.9$ and $x_C = 1$. The corresponding convergence trajectories are V^A, V^B and V^C which are showed in Fig. 3, respectively.

For $1 < x_0 \leq 1.5$, we run the iterative ADP algorithm for Case 4. The search step is 0.1 from $x_k = 1$ to $x_k = 1.5$. There are 4 state points which are $x_D = 1.1$, $x_E = 1.2$, $x_F = 1.3$, $x_G = 1.4$ and $x_H = 1.5$. For each state point, we should give a finite horizon admissible control sequence as the initial control sequence. For convenience, the length of all the initial control sequence is 2. The control sequences are ${}_D\hat{\underline{u}}_0^1 = (-\sin^{-1}(0.3) - 0.121, -\sin^{-1}(0.8) - 0.064)$, ${}_E\hat{\underline{u}}_0^1 = (-\sin^{-1}(0.4) - 0.144, -\sin^{-1}(0.8) - 0.064)$, ${}_F\hat{\underline{u}}_0^1 = (-\sin^{-1}(0.5) - 0.169, -\sin^{-1}(0.8))$, ${}_G\hat{\underline{u}}_0^1 = (-\sin^{-1}(0.6) - 0.196, -\sin^{-1}(0.8))$ and ${}_H\hat{\underline{u}}_0^1 = (-\sin^{-1}(0.7) - 0.225, -\sin^{-1}(0.8))$. The corresponding state trajectories are ${}_D\hat{\underline{x}}_0^2 = (1.1, 0.8, 0)$, ${}_E\hat{\underline{x}}_0^2 = (1.2, 0.8, 0)$, ${}_F\hat{\underline{x}}_0^2 = (1.3, 0.8, 0)$, ${}_G\hat{\underline{x}}_0^2 = (1.4, 0.8, 0)$, ${}_H\hat{\underline{x}}_0^2 = (1.5, 0.8, 0)$.

We run the iterative ADP algorithm for Case 4 at state points x_D, x_E, x_F, x_G and x_H. For each iterative step, the critic network and the action network are also trained for 1000 steps under the learning rate $\alpha = 0.05$ so that the given neural network accuracy $\varepsilon = 10^{-8}$ is reached. After 15 iterative steps, we obtain the performance index function trajectories shown in Fig. 3. The corresponding convergent trajectories of the performance index functions are V^D, V^E, V^F, V^G and V^G which are shown in Fig. 4.

From the simulation results we have $x_A \in \mathcal{T}_5^{(\epsilon)}$, $x_B \in \mathcal{T}_5^{(\epsilon)}$, $x_C \in \mathcal{T}_6^{(\epsilon)}$, $x_D \in \mathcal{T}_6^{(\epsilon)}$, $x_E \in \mathcal{T}_6^{(\epsilon)}$, $x_F \in \mathcal{T}_6^{(\epsilon)}$, $x_G \in \mathcal{T}_7^{(\epsilon)}$ and $x_H \in \mathcal{T}_7^{(\epsilon)}$ and we have $I = 7$. To show the effectiveness of the optimal control, we arbitrarily choose 3 state points in Ω_0 such as $x_\alpha = 0.8$, $x_\beta = 1$ and $x_\gamma = 1.5$. Applying the optimal control law of $\mu_\epsilon^7(x_H)$ to the 3 state points, we obtain the following results exhibited Figs. 5 and 6.

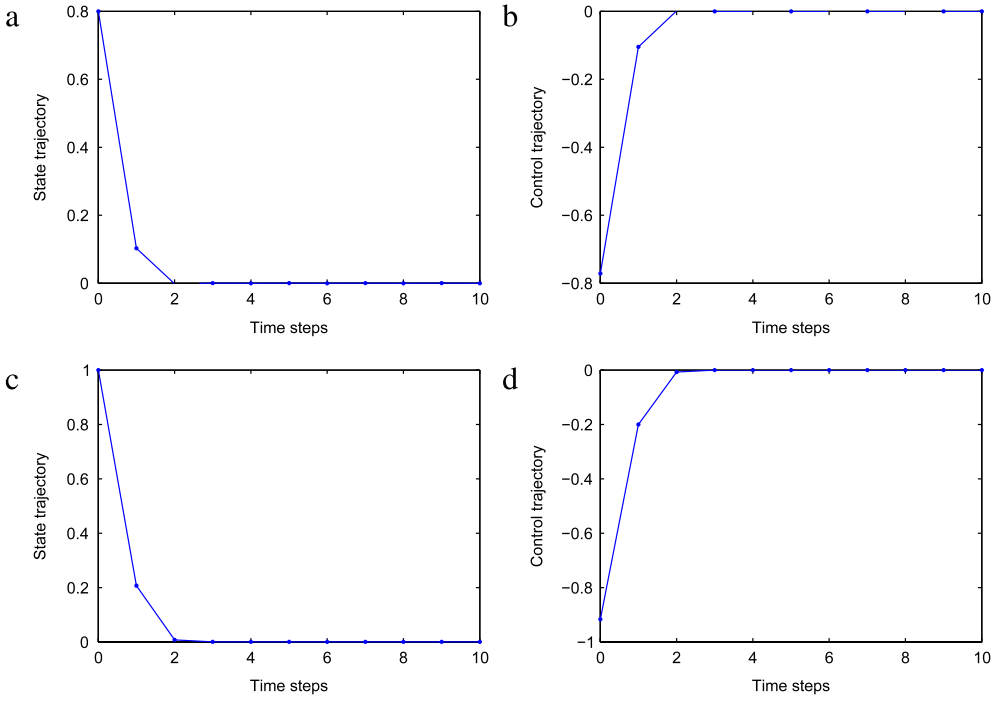

Fig. 5. Simulation results. (a) State trajectory for $x_\alpha = 0.8$. (b) Control trajectory for $x_\alpha = 0.8$. (c) State trajectory for $x_\beta = 1$. (d) Control trajectory for $x_\beta = 1$.

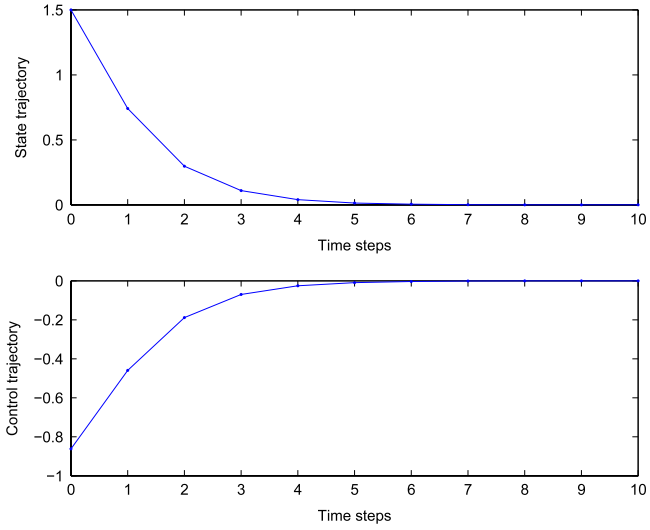

Fig. 6. Simulation results for the state $x_\gamma = 1.5$. (a) State trajectories. (b) Control trajectories.

7. Conclusions

In this paper we developed an effective iterative ADP algorithm for finite-horizon ϵ-optimal control of discrete-time nonlinear systems with unfixed initial state. The iterative ADP algorithm can be implemented by an arbitrary admissible control sequence while the initial constraint which requires the system to reach zero directly is removed. Convergence of the performance index function of the iterative ADP algorithm is proved and the ϵ-optimal number of control steps can also be obtained. Neural networks are used to implement the iterative ADP algorithm. Finally, a simulation example is given to illustrate the performance of the proposed algorithm.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 60904037, 60921061, and 61034002, in part by the Beijing Natural Science Foundation under Grant 4102061, in part by China Postdoctoral Science Foundation under Grant 201104162.

References

Al-Tamimi, A., Abu-Khalaf, M., & Lewis, F. L. (2008). Adaptive critic designs for discrete-time zero-sum games with application to H_∞ control. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, *37*(1), 240–247.

Bellman, R. E. (1957). *Dynamic programming*. Princeton, New Jersey: Princeton University Press.

Dierks, T., Thumati, B. T., & Jagannathan, S. (2009). Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence. *Neural Networks*, *22*(5–6), 851–860.

Ichihara, H. (2009). Optimal control for polynomial systems using matrix sum of squares relaxations. *IEEE Transactions on Automatic Control*, *54*(5), 1048–1053.

Kioskeridis, I., & Mademlis, C. (2009). A unified approach for four-quadrant optimal controlled switched reluctance machine drives with smooth transition between control operations. *IEEE Transactions on Automatic Control*, *24*(1), 301–306.

Kulkarni, R. V., & Venayagamoorthy, G. K. (2010). Bio-inspired algorithms for autonomous deployment and localization of sensor nodes. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, *40*(6), 663–675.

Landelius, T. (1997). Reinforcement learning and distributed local model synthesis. *Ph.D. Dissertation*. Linkoping University. Sweden.

Liu, D., Zhang, Y., & Zhang, H. (2005). A self-learning call admission control scheme for CDMA cellular networks. *IEEE Transactions on Neural Networks*, *16*(5), 1219–1228.

Mao, J., & Cassandras, C. G. (2009). Optimal control of multi-stage discrete event systems with real-time constraints. *IEEE Transactions on Automatic Control*, *54*(1), 108–123.

Murray, J. J., Cox, C. J., Lendaris, G. G., & Saeks, R. (2002). Adaptive dynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, *32*(2), 140–153.

Necoara, I., Kerrigan, E. C., Schutter, B. D., & Boom, T. (2007). Finite-horizon min–max control of max-plus-linear systems. *IEEE Transactions on Automatic Control*, *52*(6), 1088–1093.

Prokhorov, D. V., & Wunsch, D. C. (1997). Adaptive critic designs. *IEEE Transactions on Neural Networks*, *8*(5), 997–1007.

Si, J., & Wang, Y. T. (2001). On-line learning control by association and reinforcement. *IEEE Transactions on Neural Networks*, *12*(2), 264–276.

Uchida, K., & Fujita, M. (1992). Finite horizon H^∞ control problems with terminal penalties. *IEEE Transactions on Automatic Control*, *37*(11), 1762–1767.

Vamvoudakis, K. G., & Lewis, F. L. (2011). Multi-player non-zero-sum games: online adaptive learning solution of coupled Hamilton–Jacobi equations. *Automatica*, *47*(8), 1556–1569.

Vrabie, D., & Lewis, F. L. (2009). Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Networks*, *22*(3), 237–246.

Wang, F. Y., Jin, N., Liu, D., & Wei, Q. (2011). Adaptive dynamic programming for finite horizon optimal control of discrete-time nonlinear systems with ε-error bound. *IEEE Transactions on Neural Networks*, *22*(1), 24–36.

Wang, F. Y., Zhang, H., & Liu, D. (2009). Adaptive dynamic programming: an introduction. *IEEE Computational Intelligence Magazine*, *4*(2), 39–47.

Wei, Q., & Liu, D. (2011a). Finite horizon optimal control of discrete-time nonlinear systems with unfixed initial state using adaptive dynamic programming. *Journal of Control Theory & Applications*, *9*(1), 123–133.

Wei, Q., & Liu, D. (2011b). Optimal control for discrete-time nonlinear systems with unfixed initial state using adaptive dynamic programming. In *Proceeding of international joint conference on neural networks. IJCNN 2011. San Jose, USA. July* (pp. 61–67).

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

Werbos, P. J. (1991). A menu of designs for reinforcement learning over time. In W. T. Miller, R. S. Sutton, & P. J. Werbos (Eds.), *Neural networks for control*. Cambridge: MIT Press.

Werbos, P. J. (2009). Intelligence in the brain: a theory of how it works and how to build it. *Neural Networks*, *22*(3), 200–212.

Zhang, H. G., Luo, Y. H., & Liu, D. (2009). The RBF neural network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraint. *IEEE Transactions on Neural Networks*, *20*(9), 1490–1503.

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

Zhang, H. G., Wei, Q. L., & Luo, Y. H. (2008). 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*, *38*(4), 937–942.