Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Nearly Optimal Control Scheme for Discrete-Time Nonlinear Systems With Finite Approximation Errors Using Generalized Value Iteration Algorithm ⋆

Qinglai Wei, Derong Liu

The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (Tel: +86-10-82544761; Fax: +86-10-82544799; emails: qinglai.wei@ia.ac.cn, derong.liu@ia.ac.cn).

Abstract: In this paper, a new generalized value iteration algorithm is developed to solve infinite horizon optimal control problems for discrete-time nonlinear systems. The idea is to use iterative adaptive dynamic programming (ADP) to obtain the iterative control law which makes the iterative performance index function reach the optimum. The generalized value iteration algorithm permits an arbitrary positive semi-definite function to initialize it, which overcomes the disadvantage of traditional value iteration algorithms. When the iterative control law and iterative performance index function in each iteration cannot be accurately obtained, a new design method of the convergence criterion for the generalized value iteration algorithm with finite approximation errors is established to make the iterative performance index functions converge to a finite neighborhood of the lowest bound of all performance index functions. Simulation results are given to illustrate the performance of the developed algorithm.

Keywords: Adaptive dynamic programming, approximate dynamic programming, nonlinear system, optimal control, reinforcement learning.

1. INTRODUCTION

Dynamic programming is an important technique in handling optimal control problems. However, due to the "curse of dimensionality", the optimal solutions cannot be obtained directly by dynamic programming (Bellman [1957]). Adaptive dynamic programming (ADP), proposed by Werbos [1977] and [1991], has demonstrated the capability to find the optimal control policy and solve the HJB equation in a principled way. Iterative methods are primary tools in ADP to obtain the solution of HJB equation indirectly and have attracted increasing attention (Heydari and Balakrishnan [2013]; Liu et al. [2013]; Liu and Wei [2014]; Zhang et al. [2011]).

Value iteration algorithms are one class of the most primary and important iterative ADP algorithms (Wei et al. [2009]; Wei and Liu [2012]; Yang and Jagannathan [2012]). Value iteration algorithms of ADP are given in Bertsekas and Tsitsiklis [1996]. In 2008, Al-Tamimi et al. studied a value iteration algorithm for discrete-time affine nonlinear systems (Al-Tamimi et al. [2008]). Starting from a zero initial performance index function, it is proven that the iterative performance index function is a non-decreasing sequence and bounded, which makes the iterative performance index function converge to the optimum as the iteration index increases to infinity. In recent years, value iteration algorithms have attracted more and more researchers (Liu et al. [2012]; Wei and Liu [2013a]; Wei and Liu [2013b]; Zhang et al. [2008]). But it is known that the previous value iteration algorithms, i.e., traditional value iteration algorithms in brief, are required to start from a zero initial condition. Other initial conditions are seldom discussed. On the other hand, most previous discussions on ADP required that the approximation structure could approximate the iterative performance index function accurately. But for most real-world control systems, the accurate performance index function cannot be achieved. Hence, ADP algorithms with approximation errors are important to discuss. Although in several papers (Liu and Wei [2013a]; Wei and Liu [2014]), the convergence properties of ADP algorithm were discussed, in these papers a uniform approximation error was required to build these convergence criteria. However, the uniform approximation error is generally difficult to obtain. To the best of our knowledge, all the convergence criteria in the previous papers were difficult to obtain and there are no discussions on how to design a convergence criterion that makes the iterative ADP algorithms converge. This motivates our research.

In this paper, a new discrete-time generalized value iteration algorithm with finite approximation errors will be constructed. First, the detailed generalized value iteration algorithm is described. It permits an arbitrary positive

⋆ This work was supported in part by the National Natural Science Foundation of China under Grants 61034002, 61233001, 61273140, 61304086, and 61374105, in part by Beijing Natural Science Foundation under Grant 4132078, and in part by the Early Career Development Award of SKLMCCS.

semi-positive function to initialize the developed algorithm, which overcomes the disadvantage of traditional value iteration algorithms. Second, the convergence properties for the finite-approximation-error based generalized value iteration algorithm are analyzed. We emphasized that for the first time a new "design method of the convergence criterion" for the generalized value iteration algorithm with finite approximation errors is established. It permits that the developed generalized value iteration algorithm designs a suitable approximation error adaptively to make the iterative performance index function converge to a finite neighborhood of the optimal performance index function. Finally, simulation results are given to show the effectiveness of the developed iterative ADP algorithm.

2. PROBLEM FORMULATION

In this paper, the following discrete-time nonlinear system is considered

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

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$ and x_0 is the initial state.

Let $\underline{u}_k = (u_k, u_{k+1}, \ldots)$ be a sequence of controls from k to ∞. The performance index function is defined as

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

where $U(x_k, u_k) > 0$, for $\forall\, x_k, u_k \neq 0$, is the utility function. In this paper, we aim to find an optimal control scheme that minimizes the performance index function (2). The following assumption is necessary for the analysis of the developed ADP algorithm.

Assumption 1. The system (1) is controllable; $x_k = 0$ is a unique equilibrium state, i.e., $F(0, 0) = 0$; $u(x_k) = 0$ for $x_k = 0$; $U(x_k, u_k)$ is positive definite.

Define the control sequence set as $\underline{\mathfrak{U}}_k = \big\{\underline{u}_k \colon \underline{u}_k = (u_k, u_{k+1}, \ldots), \forall u_{k+i} \in \mathbb{R}^m, i = 0, 1, \ldots\big\}$. Then, for arbitrary control sequence $\underline{u}_k \in \underline{\mathfrak{U}}_k$, the optimal performance index function can be defined as

$$J^*(x_k) = \inf_{\underline{u}_k} \Big\{ J(x_k, \underline{u}_k) \colon \underline{u}_k \in \underline{\mathfrak{U}}_k \Big\}. \tag{3}$$

According to Bellman's principle of optimality, $J^*(x_k)$ satisfies the discrete-time Hamilton-Jacobi-Bellman (HJB) equation

$$J^*(x_k) = \inf_{u_k} \Big\{ U(x_k, u_k) + J^*(F(x_k, u_k)) \Big\}. \tag{4}$$

Then, the law of optimal single control can be expressed as

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

Hence, the HJB equation (4) can be written as

$$J^*(x_k) = U(x_k, u^*(x_k)) + J^*(F(x_k, u^*(x_k))). \tag{6}$$

3. GENERALIZED VALUE ITERATION ALGORITHM WITH FINITE APPROXIMATION ERRORS

In this section, a new generalized value iteration algorithm is developed to obtain the optimal control law for nonlinear systems (1). Approximation errors of the iterative performance index functions and iterative control laws are considered. New convergence property analysis methods will be established. The new design method of the convergence criterion will be developed.

3.1 Derivation of the Generalized Value Iteration Algorithm With Finite Approximation Errors

The developed generalized value iteration algorithm is updated by iterations, with the iteration index i increasing from 0 to ∞. For $\forall x_k$, let the initial function $\hat{V}_0(x_k) = \Psi(x_k)$, where $\Psi(x_k) \geq 0$ is a positive semi-definite function. The iterative control law $\hat{v}_0(x_k)$ can be computed as follows:

$$\hat{v}_0(x_k) = \arg\min_{u_k} \Big\{ U(x_k, u_k) + \hat{V}_0(x_{k+1}) \Big\} + \rho_0(x_k) \tag{7}$$

where $\hat{V}_0(x_{k+1}) = \Psi(x_{k+1})$ and the performance index function can be updated as

$$\hat{V}_1(x_k) = U(x_k, \hat{v}_0(x_k)) + \hat{V}_0(F(x_k, \hat{v}_0(x_k))) + \pi_0(x_k), \tag{8}$$

where $\rho_0(x_k)$ and $\pi_0(x_k)$ are finite approximation error functions. For $i = 1, 2, \ldots$, the iterative ADP algorithm will iterate between

$$\begin{aligned} \hat{v}_i(x_k) &= \arg\min_{u_k} \Big\{ U(x_k, u_k) + \hat{V}_i(x_{k+1}) \Big\} + \rho_i(x_k) \\ &= \arg\min_{u_k} \Big\{ U(x_k, u_k) + \hat{V}_i(F(x_k, u_k)) \Big\} + \rho_i(x_k) \end{aligned} \tag{9}$$

and

$$\hat{V}_{i+1}(x_k) = U(x_k, \hat{v}_i(x_k)) + \hat{V}_i(F(x_k, \hat{v}_i(x_k))) + \pi_i(x_k), \tag{10}$$

where $\rho_i(x_k)$ and $\pi_i(x_k)$ are finite approximation error functions of the iterative control and iterative performance index function, respectively. In next subsection, it will be proven that for $i \to \infty$, the iterative performance index function $V_i(x_k)$ and the iterative control law $v_i(x_k)$ converge to the optimal ones.

3.2 Properties of the Generalized Value Iteration Algorithm With Finite Approximation Errors

For the generalized value iteration algorithm (7)–(10), if for $\forall\, i = 0, 1, \ldots$, the iterative performance index function and the iterative control law can accurately be obtained, then the algorithm is reduced to the following equations

$$\begin{aligned} v_i(x_k) &= \arg\min_{u_k} \Big\{ U(x_k, u_k) + V_i(F(x_k, u_k)) \Big\}, \\ V_{i+1}(x_k) &= \min_{u_k} \Big\{ U(x_k, u_k) + V_i(F(x_k, u_k)) \Big\} \\ &= U(x_k, v_i(x_k)) + V_i(F(x_k, v_i(x_k))), \end{aligned} \tag{11}$$

where $V_0(x_k) = \Psi(x_k)$ is an arbitrary positive semi-definite function. In Liu and Wei [2013b], it is shown that iterative performance index function converges to the optimum. As the existence of the approximation errors, the convergence may not hold. The following lemma will show this property.

Lemma 1. For $i = 1, 2, \ldots$, Let $\Upsilon_i(x_k)$ be the target iterative performance index function, which is expressed as

$$\Upsilon_i(x_k) = \min_{u_k} \Big\{ U(x_k, u_k) + \hat{V}_{i-1}(x_{k+1}) \Big\}, \tag{12}$$

where $\hat{V}_i(x_k)$ is defined in (10). If the initial iterative performance index function $\hat{V}_0(x_k) = \Upsilon_0(x_k) = \Psi(x_k)$ and for

$\forall i = 1, 2, \ldots$, there exists a uniform finite approximation error ζ that satisfies

$$\hat{V}_i(x_k) - \Upsilon_i(x_k) \leq \zeta, \tag{13}$$

then we have

$$\hat{V}_i(x_k) - V_i(x_k) \leq i\zeta. \tag{14}$$

Proof. The details of the proof can be seen in Liu and Wei [2013a] and omitted here.

Thus, a new analysis method will be developed. To facilitate analysis, the expressions of the approximation error are transformed. For $\forall i = 1, 2, \ldots$, there exists a finite constant $\vartheta_i > 0$ that makes

$$\hat{V}_i(x_k) \leq \vartheta_i \Upsilon_i(x_k) \tag{15}$$

hold. From (15), it can be seen that the iterative performance index function $\hat{V}_i(x_k)$ is upper bounded by $\vartheta_i \Upsilon_i(x_k)$. If the convergence properties of $\Upsilon_i(x_k)$ are analyzed for different ϑ_i, then the convergence of $\hat{V}_i(x_k)$ can be justified. Thus, in the following, the convergence properties of the upper bound will be discussed.

Theorem 1. For $\forall i = 1, 2, \ldots$, let $\Upsilon_i(x_k)$ be expressed as in (12) and $\hat{V}_i(x_k)$ be expressed as in (10). If for $\forall i = 1, 2, \ldots$, there exists $0 < \vartheta_i < 1$ that makes (15) hold, then we have that the iterative performance index function is convergent.

Proof. If $0 < \vartheta_i < 1$, according to (15), we have $0 \leq \hat{V}_i(x_k) < \Upsilon_i(x_k)$. Using mathematical induction, we can prove that for $\forall i = 1, 2, \ldots$, the following inequality

$$0 < \hat{V}_i(x_k) < V_i(x_k) \tag{16}$$

holds. According to Liu and Wei [2013b], we have $V_i(x_k) \to J^*(x_k)$. Then for $\forall i = 0, 1, \ldots$, $\hat{V}_i(x_k)$ is upper bounded and

$$0 < \lim_{i\to\infty} \hat{V}_i(x_k) < \lim_{i\to\infty} V_i(x_k) = J^*(x_k). \tag{17}$$

The proof is completed.

Next, we will analyze the situation of $1 \leq \vartheta_i < \infty$.

Theorem 2. For $\forall i = 1, 2, \ldots$, let $\Upsilon_i(x_k)$ be expressed as (12) and $\hat{V}_i(x_k)$ be expressed as (10). Let $0 < \varphi_i < \infty$ be a constant that makes

$$V_i(F(x_k, u_k)) \leq \varphi_i U(x_k, u_k) \tag{18}$$

hold. If Assumption 1 holds and for $\forall i = 1, 2, \ldots$, there exists $1 \leq \vartheta_i < \infty$ that makes (15) hold, then we have

$$\hat{V}_i(x_k) \leq \vartheta_i \Bigg(1 + \sum_{j=1}^{i-1} \Bigg(\vartheta_{i-1}\vartheta_{i-2}\cdots\vartheta_{i-j+1}(\vartheta_{i-j} - 1) \times \frac{\varphi_{i-1}\varphi_{i-2}\cdots\varphi_{i-j}}{(\varphi_{i-1}+1)(\varphi_{i-2}+1)\cdots(\varphi_{i-j}+1)}\Bigg)\Bigg) V_i(x_k), \tag{19}$$

where we define $\sum_{j}^{i}(\cdot) = 0$, for $\forall j > i$, $i, j = 0, 1, \ldots$, and $\vartheta_{i-1}\vartheta_{i-2}\cdots\vartheta_{i-j+1}(\vartheta_{i-j} - 1) = (\vartheta_{i-1} - 1)$, for $j = 1$.

Proof. The theorem can be proven by mathematical induction. First, let $i = 1$ and then (12) becomes

$$\begin{aligned}\Upsilon_1(x_k) &= \min_{u_k}\left\{U(x_k, u_k) + \hat{V}_0(x_{k+1})\right\} \\ &= V_1(x_k).\end{aligned} \tag{20}$$

According to (15), we have $\hat{V}_1(x_k) \leq \vartheta_1 V_1(x_k)$. Thus, the conclusion holds for $i = 1$. Assume that (19) holds for $i = l - 1$, where $l = 2, 3, \ldots$. Then, for $i = l$, we can obtain (19).

Then, according to (15), we can obtain (19). The mathematical induction is completed.

From (19), we can see that for $\forall i = 0, 1, \ldots$, there exists an error between the $\hat{V}_i(x_k)$ and $V_i(x_k)$. As $i \to \infty$, the bound of the approximation errors may increase to infinity. Thus, in the following, we will give the convergence properties of the iterative ADP algorithm (7)–(10) using error bound method. Before presenting the next theorem, the following lemma is necessary.

Lemma 2. Let $\{b_i\}$, $i = 1, 2, \ldots$ be a sequence of positive number. Let $0 < \lambda_i < \infty$ be a bounded positive constant for $\forall i = 1, 2, \ldots$ and let $a_i = \lambda_i b_i$. If $\sum_{i=1}^{\infty} b_i$ is finite, then we have that $\sum_{i=1}^{\infty} a_i$ is finite.

Proof. As for $\forall i = 1, 2, \ldots$, λ_i is finite, if we let $\bar{\lambda} = \sup\{\lambda_1, \lambda_2, \ldots\}$, then we have that

$$\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} \lambda_i b_i \leq \bar{\lambda} \sum_{i=1}^{\infty} b_i \tag{20}$$

is finite.

Theorem 3. Let $\hat{V}_i(x_k)$ be expressed as (19). If for $\forall i = 1, 2, \ldots$, the inequality

$$1 \leq \vartheta_{i+1} \leq q_i \frac{\varphi_i + 1}{\varphi_i} \tag{21}$$

holds, where q_i is an arbitrary constant which satisfies $\frac{\varphi_i}{\varphi_i + 1} < q_i < 1$, then as $i \to \infty$, the iterative performance index function $\hat{V}_i(x_k)$ of the generalized value iteration algorithm converges to a finite neighborhood of $J^*(x_k)$.

Proof. For (19) in Theorem 2, if we let

$$\Delta_i = \sum_{j=1}^{i-1} \Bigg(\vartheta_{i-1}\vartheta_{i-2}\cdots\vartheta_{i-j+1}(\vartheta_{i-j} - 1) \times \frac{\varphi_{i-1}\varphi_{i-2}\cdots\varphi_{i-j}}{(\varphi_{i-1}+1)(\varphi_{i-2}+1)\cdots(\varphi_{i-j}+1)}\Bigg), \tag{22}$$

$$a_{ij} = \frac{\vartheta_{i-1}\vartheta_{i-2}\cdots\vartheta_{i-j}\varphi_{i-1}\varphi_{i-2}\cdots\varphi_{i-j}}{(\varphi_{i-1}+1)(\varphi_{i-2}+1)\cdots(\varphi_{i-j}+1)}, \tag{23}$$

and

$$b_{ij} = \frac{\vartheta_{i-1}\vartheta_{i-2}\cdots\vartheta_{i-j+1}\varphi_{i-1}\varphi_{i-2}\cdots\varphi_{i-j}}{(\varphi_{i-1}+1)(\varphi_{i-2}+1)\cdots(\varphi_{i-j}+1)}, \tag{24}$$

where $i = 1, 2, \ldots$, and $j = 1, 2, \ldots, i-1$, then we have $\Delta_i = \sum_{j=1}^{i-1} a_{ij} - \sum_{j=1}^{i-1} b_{ij}$. We know that if $\sum_{j=1}^{i-1} a_{ij}$ and $\sum_{j=1}^{i-1} b_{ij}$ are both finite as $i \to \infty$, then $\lim_{i\to\infty} \Delta_i$ is finite. According to (24), we have $\frac{b_{ij}}{b_{i(j-1)}} = \frac{\vartheta_{i-j+1}\varphi_{i-j}}{(\varphi_{i-j}+1)}$. If $\frac{b_{ij}}{b_{i(j-1)}} \leq q_{i-j} < 1$, then we can get $\vartheta_{i-j+1} \leq q_{i-j}\frac{\varphi_{i-j}+1}{\varphi_{i-j}}$. Let $\ell = i - j$ and then we can obtain

$$\vartheta_{\ell+1} \leq q_\ell \frac{\varphi_\ell + 1}{\varphi_\ell}, \tag{25}$$

$$
\begin{aligned}
\Upsilon_l(x_k) =& \min_{u_k}\Big\{U(x_k,u_k)+\hat{V}_{l-1}(F(x_k,u_k))\Big\}\\
\le& \min_{u_k}\Bigg\{U(x_k,u_k)+\vartheta_{l-1}\Bigg(1+\sum_{j=1}^{l-2}\Bigg(\vartheta_{l-2}\vartheta_{l-3}\cdots\vartheta_{l-j}(\vartheta_{l-j-1}-1)\frac{\varphi_{l-2}\varphi_{l-3}\cdots\varphi_{l-j-1}}{(\varphi_{l-2}+1)(\varphi_{l-3}+1)\cdots(\varphi_{l-j-1}+1)}\Bigg)\Bigg)\\
&\times V_{l-1}(x_k)\Bigg\}\\
\le& \min_{u_k}\Bigg\{\Bigg(1+\varphi_{l-1}\Bigg(\sum_{j=1}^{l-1}(\vartheta_{l-1}\vartheta_{l-2}\cdots\vartheta_{l-j+1}(\vartheta_{l-j}-1)\frac{\varphi_{l-2}\varphi_{l-3}\cdots\varphi_{l-j}}{(\varphi_{l-1}+1)(\varphi_{l-2}+1)\cdots(\varphi_{l-j}+1)}\Bigg)\Bigg)U(x_k,u_k)\\
&+\Bigg(\frac{\varphi_{l-1}\vartheta_{l-1}}{\varphi_{l-1}+1}\Bigg(1+\sum_{j=1}^{l-2}(\vartheta_{l-2}\vartheta_{l-3}\cdots\vartheta_{l-j}(\vartheta_{l-j-1}-1))\frac{\varphi_{l-2}\varphi_{l-3}\cdots\varphi_{l-j-1}}{(\varphi_{l-2}+1)(\varphi_{l-3}+1)\cdots(\varphi_{l-j-1}+1)}\Bigg)\\
&+\frac{1}{\varphi_{l-1}+1}\Bigg)V_{l-1}(x_k)\Bigg\}\\
=&\Bigg(1+\sum_{j=1}^{l-1}(\vartheta_{l-1}\vartheta_{l-2}\cdots\vartheta_{l-j+1}(\vartheta_{l-j}-1))\frac{\varphi_{l-1}\varphi_{l-2}\cdots\varphi_{l-j}}{(\varphi_{l-1}+1)(\varphi_{l-2}+1)\cdots(\varphi_{l-j}+1)}\Bigg)\min_{u_k}\Big\{U(x_k,u_k)+V_{l-1}(x_k)\Big\}\\
=&\Bigg(1+\sum_{j=1}^{l-1}(\vartheta_{l-1}\vartheta_{l-2}\cdots\vartheta_{l-j+1}(\vartheta_{l-j}-1))\frac{\varphi_{l-1}\varphi_{l-2}\cdots\varphi_{l-j}}{(\varphi_{l-1}+1)(\varphi_{l-2}+1)\cdots(\varphi_{l-j}+1)}\Bigg)V_l(x_k) \qquad (19)
\end{aligned}
$$

where $\ell = 1,2,\ldots,i-1$. Let $i\to\infty$ and we can obtain (21). Let $q=\sup\{q_1,q_2,\ldots\}$ and we have $0<q<1$. We can obtain

$$\sum_{j=1}^{i-1} b_{ij} \le \sum_{j=1}^{i-1}\left(\frac{\varphi_{i-1}+1}{\varphi_{i-1}}\right)q^{j-1}. \qquad (26)$$

As $\dfrac{\varphi_i}{\varphi_i+1}<q<1$ and φ_{i-1} is finite for $\forall\, i=1,2,\ldots$, let $i\to\infty$ and we have $\lim\limits_{i\to\infty}\sum\limits_{j=1}^{i-1} b_{ij}$ is finite.

On the other hand, for $\forall\, i=1,2,\ldots$ and for $\forall j=1,2,\ldots,i-1$, we have $a_{ij}=\vartheta_{i-j}b_{ij}$. As for $\forall\, i=1,2,\ldots$ and for $\forall j=1,2,\ldots,i-1$, $1\le\vartheta_{i-j}<\infty$ is finite, according to Lemma 2, we have $\lim\limits_{i\to\infty}\sum\limits_{j=1}^{i-1} a_{ij}$ must be finite. Therefore, we can obtain $\lim\limits_{i\to\infty}\Delta_i$ is finite. According to Liu and Wei [2013b], we have $\lim\limits_{i\to\infty} V_i(x_k)=J^*(x_k)$. Hence, the iterative performance index function $\hat{V}_i(x_k)$ is convergent to a bounded neighborhood of the optimal performance index function $J^*(x_k)$. The proof is completed.

Combining Theorems 1 and 3, the convergence criterion of the generalized value iteration algorithm with finite approximation errors can be established.

Theorem 4. If Assumption 1 holds and for $\forall\, i=0,1,\ldots$, the inequality

$$0<\vartheta_{i+1}\le q_i\frac{\varphi_i+1}{\varphi_i} \qquad (27)$$

holds, where $0<q_i<1$ is an arbitrary constant, then the iterative performance index function $\hat{V}_i(x_k)$ in the generalized value iteration algorithm converges to a finite neighborhood of the optimal performance index function $J^*(x_k)$, as $i\to\infty$.

We can see that if we can obtain φ_i, then we can design the approximation error to make $\hat{V}_i(x_k)$ converge. The following theorem will give an effective way to obtain φ_i. Define Ω_{φ_i} as

$$\Omega_{\varphi_i}=\Big\{\varphi_i|\varphi_i U(x_k,u_k)\ge V_i(F(x_k,u_k))\Big\}. \qquad (28)$$

Theorem 5. Let $\mu(x_k)$ be an arbitrary admissible control law of the nonlinear system (1), i.e.,

$$P_{i+1}(x_k)=U(x_k,\mu(x_k))+P_i(x_{k+1}) \qquad (29)$$

where $P_0(x_k)=V_0(x_k)=\Psi(x_k)$. If there exists a constant $\tilde{\varphi}_i$ that satisfies

$$\tilde{\varphi}_i U(x_k,u_k)\ge P_i(F(x_k,u_k)), \qquad (30)$$

then we have $\tilde{\varphi}_i\in\Omega_{\varphi_i}$.

Proof. As $\mu(x_k)$ is an arbitrary admissible control law, we have $P_i(x_k)\ge V_i(x_k)$. If $\tilde{\varphi}_i$ satisfies (30), then can get

$$\tilde{\varphi}_i U(x_k,u_k)\ge P_i(F(x_k,u_k))\ge V_i(F(x_k,u_k)). \qquad (31)$$

The proof is completed.

From Theorem 5, we know that if we obtain an admissible control law $\mu(x_k)$, then φ_i can be estimated. The method to obtain the admissible control law can be seen in Liu and Wei [2014] and omitted here.

Remark 1. One property should be pointed out. First, the developed value iteration algorithm of ADP in this paper is different from the traditional value iteration algorithms (Al-Tamimi et al. [2008] and Wei et al. [2009]). For the traditional value iteration algorithms, the initial performance index function is required to be zero. In this paper, the initial performance index function can be an arbitrary positive semi-definite function. On the other hand, the developed value iteration algorithm in this paper is also different from Liu and Wei [2013a] and Wei and Liu [2014]. In Liu and Wei [2013a] and Wei and Liu [2014], it requires a uniform approximation error to

construct the convergence criterion. In this paper, the approximation error ϑ_i can be different for different i. This makes the convergence analysis in this paper different from our previous papers.

3.1 Summary of the Generalized Value Iteration Algorithm With Finite Approximation Errors

Now, we summarize the generalized value iteration algorithm with finite approximation errors in Algorithm 1.

Algorithm 1 Generalized value iteration algorithm with finite approximation errors

Initialization:

Choose randomly an array of initial states x_0;
Choose a semi-positive definite function $\Psi(x_k) \geq 0$;
Choose a convergence precision ζ;
Choose an admissible control law $\mu(x_k)$;
Give a sequence $\{q_i\}$, $i = 0, 1, \ldots$, where $0 < q_i < 1$;
Give two constants $0 < \varsigma < 1$, $0 < \varrho < 1$.

Iteration:

1: Let the iteration index $i = 0$;
2: Let $V_0(x_k) = \Psi(x_k)$ and obtain φ_0 by (30);
3: Compute $\hat{v}_i(x_k)$ by (9) and obtain $\hat{V}_{i+1}(x_k)$ by (10);
4: Obtain ϑ_{i+1} by (15). If ϑ_{i+1} satisfies (27), then estimate φ_{i+1} by (30), and goto next step. Otherwise decrease $\rho_i(x_k)$ and $\pi_i(x_k)$, i.e., $\rho_i(x_k) = \varsigma\rho_i(x_k)$ and $\pi_i(x_k) = \varrho\pi_i(x_k)$, respectively. Goto Step 3;
5: If $|\hat{V}_{i+1}(x_k) - \hat{V}_i(x_k)| \leq \zeta$, then the optimal performance index function is obtained and goto Step 6; else let $i = i + 1$ and goto Step 3;
6: **return** $\hat{v}_i(x_k)$ and $\hat{V}_i(x_k)$.

Remark 2. Generally, in iterative ADP algorithms, the difference between $\hat{V}_i(x_k)$ and $\Gamma_i(x_k)$ is obtained, i.e.,

$$\hat{V}_i(x_k) - \Gamma_i(x_k) = \zeta_i(x_k). \tag{32}$$

where $\zeta_i(x_k)$ is the approximation error function. According to the definition of ϑ_i in (15) and the convergence criterion (27), we can easily obtain the following convergence criterion

$$\zeta_i(x_k) \leq \frac{1}{\varphi_{i-1} + 1}\hat{V}_i(x_k). \tag{33}$$

4. SIMULATION STUDIES

We now examine the performance of the developed algorithm in a torsional pendulum system in Liu and Wei [2014]. The dynamics of the pendulum is as follows

$$\begin{bmatrix} x_{1(k+1)} \\ x_{2(k+1)} \end{bmatrix} = \begin{bmatrix} 0.1x_{2k} + x_{1k} \\ -0.49\sin(x_{1k}) + 0.98x_{2k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} u_k, \tag{34}$$

where $x_{1k} = \theta_k$ and $x_{2k} = \omega_k$. Let the initial state be $x_0 = [1, -1]^T$. We choose the $p = 10000$ states. Let the structures of the critic and action networks be 2–12–1 and 2–12–1. The neural network training method can be seen in Liu and Wei [2014] and omitted here. To illustrate the effectiveness of the algorithm, we also choose four different initial performance index functions which are expressed by $\Psi^j(x_k) = x_k^T P_j x_k$, $j = 1, \ldots, 4$. Let $P_1 = 0$. Let P_2–P_4 be initialized by arbitrary positive definite matrices with the forms $P_2 =$ $[2.35, 3.31; 3.31, 9.28]$, $P_3 = [5.13, -5.72; -5.72, 15.13]$, $P_4 = [100.78, 5.96; 5.96, 20.51]$, respectively. Let $q_i = 0.9999$ for $\forall\, i = 0, 1, \ldots$, and let $\varsigma = \varrho = 0.5$. Initialized by $\Psi^j(x_k)$, $j = 1, \ldots, 4$, the developed algorithm with finite approximation errors is implemented. The trajectories of φ's with $\Psi^1(x_k)$–$\Psi^4(x_k)$ are presented in Figs. 1(a)–(d), respectively.

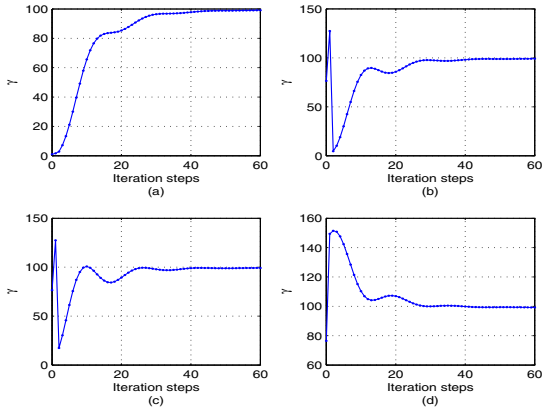

Fig. 1. The trajectories of φ's with $\Psi^1(x_k)$–$\Psi^4(x_k)$. (a)

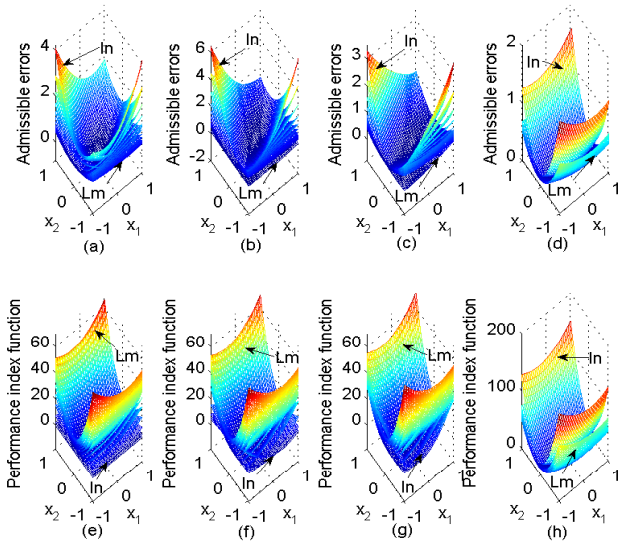

Fig. 2. The curves of the admissible errors and the iterative performance index functions with $\Psi^1(x_k)$–$\Psi^4(x_k)$. (a) Admissible errors with $\Psi^1(x_k)$. (b) Admissible errors with $\Psi^2(x_k)$. (c) Admissible errors with $\Psi^3(x_k)$. (d) Admissible errors with $\Psi^4(x_k)$. (e) Performance index function with $\Psi^1(x_k)$. (f) $\Psi^2(x_k)$. (g) Performance index function with $\Psi^3(x_k)$. (h) Performance index function with $\Psi^4(x_k)$.

According to φ's, the curved surfaces of the admissible errors with $\Psi^1(x_k)$–$\Psi^4(x_k)$ are shown in Figs. 2(a)–(d), and the iterative performance index functions are shown in Figs. 2(e)–(h) where "In" denotes initial iteration and "Lm" denotes limiting iteration.

From Figs. 1–2, it can be seen that for different initial performance index functions $\Psi^1(x_k)$–$\Psi^4(x_k)$, the iterative performance index functions by the generalized value iteration algorithm can converge to a finite neighborhood of the optimal one. The corresponding iterative controls

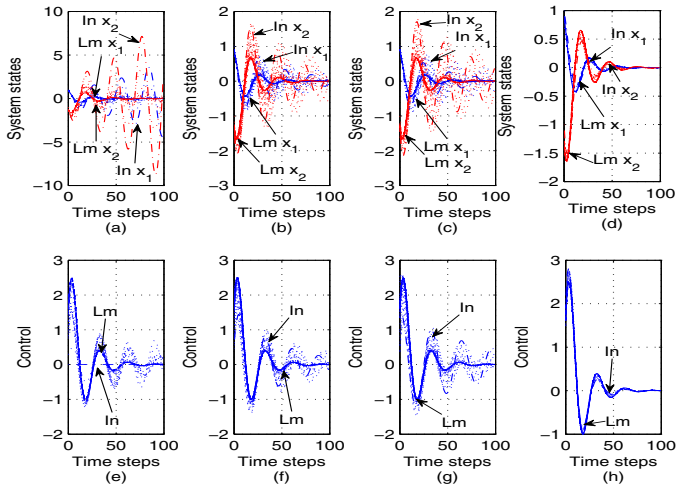

Fig. 3. Iterative trajectories of states and controls with $\Psi^1(x_k)$–$\Psi^4(x_k)$. (a) States with $\Psi^1(x_k)$. (b) States with $\Psi^2(x_k)$. (c) States with $\Psi^3(x_k)$. (d) States with $\Psi^4(x_k)$. (e) Controls with $\Psi^1(x_k)$. (f) Controls with $\Psi^2(x_k)$. (g) Controls with $\Psi^3(x_k)$. (h) Controls with $\Psi^4(x_k)$.

and iterative states are shown in Figs. 3, which are also convergent. Therefore, the effectiveness of the developed generalized value iteration algorithm with finite approximation errors can be proven.

5. CONCLUSION

In this paper, a new generalized value iteration algorithm is developed to solve infinite horizon optimal control problems for discrete-time nonlinear systems. The developed generalized value iteration algorithm of ADP permits an arbitrary positive semi-definite function to initialize the algorithm, which overcomes the disadvantage of traditional value iteration algorithms. Considering the approximation errors, for the first time a new "design method of the convergence criterion" for the generalized value iteration algorithm with finite approximation errors is established to make the iterative performance index function converge to a finite neighborhood of the optimal performance index function. Finally, simulation results are given to illustrate the performance of the developed algorithm.

REFERENCES

A. Al-Tamimi, F.L. Lewis, and M. Abu-Khalaf. Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 38(4): 943–949, 2008.

D.P. Bertsekas, and J.N. Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, Belmont, MA, 1996.

R.E. Bellman, *Dynamic Programming.* Princeton University Press, Princeton, New Jersey, 1957.

A. Heydari, and S.N. Balakrishnan. Finite-horizon control-constrained nonlinear optimal control using single network adaptive critics. *IEEE Transactions on Neural Networks and Learning Systems*, 24(1): 145–157, 2013.

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

D. Liu, Y. Huang, D. Wang, and Q. Wei. Neural network observer-based optimal control for unknown nonlinear systems using adaptive dynamic programming. *International Journal of Control*, 86(9): 1554–1566, 2013.

D. Liu and Q. Wei. Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems. *IEEE Transactions on Cybernetics*, 43(2): 779–789, 2013a.

D. Liu and Q. Wei. Generalized adaptive dynamic programming algorithm for discrete-time nonlinear systems: convergence and stability analysis. In *Proceedings of Third IEEE International Conference on Information Science and Technology*, Yangzhou, China, 134–141, 2013b.

D. Liu and Q. Wei. Policy iterative adaptive dynamic programming algorithm for discrete-time nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 25(3): 621–634, 2014.

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

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

Q. Wei and D. Liu. Numerical adaptive learning control scheme for discrete-time nonlinear systems. *IET Control Theory & Applications*, 7(11): 1472–1486, 2013a.

Q. Wei and D. Liu. A novel iterative θ-Adaptive dynamic programming for discrete-time nonlinear systems. *IEEE Transactions on Automation Science and Engineering*, 2013b. Article in Press. DOI: 10.1109/TASE.2013.2280974

Q. Wei and D. Liu. Data-driven neuro-optimal temperature control of water gas shift reaction using stable iterative adaptive dynamic programming. *IEEE Transactions on Industrial Electronics*, 2014. Article in Press. DOI: 10.1109/TIE.2014.2301770

P.J. Werbos. Advanced forecasting methods for global crisis warning and models of intelligence. *General Systems Yearbook*, 22: 25–38, 1977.

P.J. Werbos. A menu of designs for reinforcement learning over time. W.T. Miller, R.S. Sutton, and P.J. Werbos, editors. *Neural Networks for Control.* MIT Press, Cambridge, 1991.

Q. Yang and S. Jagannathan. Reinforcement learning controller design for affine nonlinear discrete-time systems using online approximators. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 42(2): 377–390, 2012.

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

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