Adaptive Dynamic Programming with Stable Value Iteration Algorithm for Discrete-Time Nonlinear Systems*

Qinglai Wei and Derong Liu
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing 100190, P. R. China
Emails: qinglai.wei@ia.ac.cn, derong.liu@ia.ac.cn

Abstract—In this paper, a new stable value iteration adaptive dynamic programming (ADP) algorithm, named “θ-ADP” algorithm, is proposed for solving the optimal control problems of infinite horizon discrete-time nonlinear systems. By introducing a parameter θ in the iterative ADP algorithm, it is proved that any of iterative control obtained in the proposed algorithm can stabilize the nonlinear system which overcomes the disadvantage of traditional value iteration algorithms. Neural networks are used to approximate the performance index function and compute the optimal control policy, respectively, for facilitating the implementation of the iterative θ-ADP algorithm. Finally, a simulation example is given to illustrate the performance of the proposed method.

I. INTRODUCTION

DYNAMIC programming is a very useful tool in solving optimization and optimal control problems [3], [4]. However, it is often computationally untenable to run true dynamic programming due to the backward numerical process required for its solution, i.e., as a result of the well-known “curse of dimensionality” [4]. Adaptive dynamic programming (ADP) is an effective approach to solve optimal control problems. ADP was proposed by Werbos [12] to solve optimal control problems forward-in-time. Recursive methods are also used in ADP to obtain the solution of HJB equation indirectly and have received more and more attentions [1], [8]–[11], [13]–[15]. There are two main recursive ADP algorithms which are policy iteration algorithm and value iteration algorithm [6].

In the policy iteration algorithms of ADP, to obtain the performance index functions and iterative controls iteratively, an initial admissible control sequence of the system must be required. But, unfortunately, the admissible control sequence for the nonlinear system is also difficult to obtain. Thus, the initial conditions for the controller greatly limit the applications of the policy iteration algorithms.

Value iteration algorithms for the optimal control problems of discrete-time nonlinear systems were given in [5]. In 2007, Al-Tamimi and Lewis [2] studied the deterministic discrete-time affine nonlinear systems. A value iteration algorithm, which was referred to as HDP, is proposed for finding the optimal control law. For the value iteration algorithm of ADP, the initial admissible control sequence for the nonlinear system can be avoided. While for each of the iterative control $u_i(x_k)$, $i = 0, 1, \ldots$, the stability of the system can not be guaranteed. This means only the converged $u^*(x_k)$ can be used to control the nonlinear system, and all the iterative controls $u_i(x_k)$, $i = 0, 1, \ldots$, are invalid. So the computation efficiency of the value iteration ADP method is very low. Till now, all the value iteration algorithms are implemented off-line which limit their applications very much. So we can see that due to the lack of methodology, both the policy iteration and value iteration algorithms of ADP are very limited to the applications. To overcome these difficulties, a new ADP method must be proposed and this motivates our research.

In this paper, a new stable value iteration algorithm, called “iterative θ-ADP algorithm”, is proposed to solve optimal control problems for discrete-time nonlinear systems. By introducing a parameter θ in the proposed iterative ADP algorithm, the condition of the initial admissible controller is avoided. It can also be proved that the iterative performance index function converges to the optimum and any of the iterative controls can stabilize the nonlinear system under some mild assumptions. First, in the proposed iterative θ-ADP algorithm, the HJB equation for discrete-time nonlinear systems is derived. In order to solve this HJB equation, the new iterative θ-ADP algorithm is introduced to obtain the iterative controls and the iterative performance index functions. Second, it will show that the iterative θ-ADP algorithm can obtain the approximate optimal control that makes the performance index function converge to the optimum. Third, it will show that any of the iterative controls can stabilize the nonlinear system.

II. PROBLEM STATEMENT

In this paper, we will study the following deterministic discrete-time systems

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

*This work was supported in part by the National Natural Science Foundation of China under Grants 60904037 and 61034002, by Beijing Natural Science Foundation under Grant 4102061, and by China Postdoctoral Science Foundation under Grant 201104162.
where $x_k \in \mathbb{R}^n$ is the $n$-dimensional state vector and $u_k \in \mathbb{R}^m$ is the $m$-dimensional control vector. Let $x_0$ be the initial state and $F(x_k, u_k)$ be the system function.

Let $\bar{u}_k = (u_k, u_{k+1}, \ldots)$ be an arbitrary sequence of controls from $k$ to $\infty$. The performance index function for state $x_0$ under the control sequence $\bar{u}_0 = (u_0, u_1, \ldots)$ is defined as

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

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

The goal of this paper is to find an optimal control scheme which stabilizes the system (1) and simultaneously minimizes the performance index function (2). For convenience of analysis, results of this paper are based on the following assumptions.

**Assumption 1:** The system (1) is controllable and the function $F(x_k, u_k)$ is Lipschitz continuous for $\forall x_k, u_k$.

**Assumption 2:** The system state $x_k = 0$ is an equilibrium state of system (1) under the control $u_k = 0$, i.e., $F(0, 0) = 0$.

**Assumption 3:** The feedback control $u_k = u(x_k)$ satisfies $u_k = u(x_k) = 0$ for $x_k = 0$.

**Assumption 4:** The utility function $U(x_k, u_k)$ is a continuous positive definite function for any $x_k$ and $u_k$.

**Remark 2.1:** Assumption 1 means for $\forall x_k \in \mathbb{R}^n$, there exists a stable control sequence that moves the state $x_k$ to zero and the system (1) cannot jump to infinity by any one step of finite control. Assumptions 2 and 3 mean that when the system state $x_k$ reaches the equilibrium state of the system, the control input becomes zero and the state of system (1) is kept at zero. Assumptions 1–3 are general assumptions for most control problems of nonlinear systems which can easily be verified in real control systems. Assumption 4 means the utility function $U(x_k, u_k) = 0$ if and only if $x_k = u_k = 0$ and $U(x_k, u_k) \to \infty$ as $\|x_k\|, \|u_k\| \to \infty$. For example, all the quadratic form utility functions with the expression $x_k^TQx_k + u_k^TRu_k$ for $Q, R > 0$ and the state-dependent quadratic form with the expression $x_k^TQ(x_k)x_k + u_k^TR(x_k)u_k$ for $Q(x_k), R(x_k) > 0$ satisfy this property. Therefore, Assumptions 1–4 are not strong assumptions and can easily be satisfied for real control systems.

As the system (1) is controllable, there exists a stable control sequence $\bar{u}_k = \{u_k, u_{k+1}, \ldots\}$ that moves $x_k$ to zero. Let $\mathfrak{U}_k$ denote the set which contains all the stable control sequences and let $\mathfrak{A}_k$ be the set of the stable control law. Then, the optimal performance index function can be defined as

$$J^*(x_k) = \min \{ J(x_k, \bar{u}_k) : \bar{u}_k \in \mathfrak{A}_k \}.$$  \hspace{1cm} (3)

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

$$J^*(x_k) = \min_{u_k} \{ U(x_k, u_k) + J^*(F(x_k, u_k)) \}. \hspace{1cm} (4)$$

Define the law of optimal control as

$$u^*(x_k) = \arg \min_{u_k} \{ U(x_k, u_k) + J^*(F(x_k, u_k)) \}. \hspace{1cm} (5)$$

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

$$J^*(x_k) = U(x_k, u^*(x_k)) + J^*(F(x_k, u^*(x_k))). \hspace{1cm} (6)$$

III. PROPERTIES OF THE ITERATIVE $\theta$-ADAPTIVE DYNAMIC PROGRAMMING ALGORITHM

In this section, the iterative $\theta$-ADP algorithm will be proposed to obtain the optimal controller for nonlinear systems.

A. Derivation of the iterative $\theta$-ADP algorithm

In the proposed iterative $\theta$-ADP algorithm, the performance index function and control law are updated by recurrent iteration, with the iteration index $i$ increasing from 0 to infinity.

The following definition is necessary to begin the algorithm.

**Definition 3.1:** Let

$$\Psi_{x_k} = \left\{ \left. \Psi(x_k) : \Psi(x_k) > 0, \text{ and } \exists \nu(x_k) \in \mathcal{U}_k, \Psi(F(x_k, \nu(x_k))) < \Psi(x_k) \right\} \right.$$  \hspace{1cm} (7)

be the set of initial positive definite functions.

For $\forall x_k \in \mathbb{R}^n$, let the initial function $\Psi(x_k)$ be an arbitrary function that satisfies $\Psi(x_k) \in \Psi_{x_k}$. Then, for $\forall x_k \in \mathbb{R}^n$, let the initial performance index function $V_0(x_k) = \theta \Psi(x_k)$, where $\theta > 0$ is a finite positive constant. The iterative control law $v_0(x_k)$ can be computed as follows:

$$v_0(x_k) = \arg \min_{u_k} \{ U(x_k, u_k) + V_0(x_k+1) \}$$

$$= \arg \min_{u_k} \{ U(x_k, u_k) + V_0(F(x_k, u_k)) \} \hspace{1cm} (8)$$

where $V_0(x_k+1) = \theta \Psi(x_k+1)$. The performance index function can be updated as

$$V_1(x_k) = U(x_k, v_0(x_k)) + V_0(F(x_k, v_0(x_k))). \hspace{1cm} (9)$$

For $i = 1, 2, \ldots$ and for $\forall x_k$, the iterative ADP algorithm will iterate between

$$v_i(x_k) = \arg \min_{u_k} \{ U(x_k, u_k) + V_i(x_k+1) \}$$

$$= \arg \min_{u_k} \{ U(x_k, u_k) + V_i(F(x_k, u_k)) \} \hspace{1cm} (10)$$

and performance index function

$$V_{i+1}(x_k) = \min_{u_k} \{ U(x_k, u_k) + V_i(x_k+1) \}$$

$$= U(x_k, v_i(x_k)) + V_i(F(x_k, v_i(x_k))). \hspace{1cm} (11)$$

B. Properties of the iterative $\theta$-ADP algorithm

In the above, we can see that the optimal performance index function $J^*(x_k)$ is replaced by a sequence of iterative performance index functions $V_i(x_k)$ and the optimal control law $u^*(x_k)$ is replaced by a sequence of iterative control laws $v_i(x_k)$, where $i \geq 0$ is the iteration index. As (11) is not an HJB equation for $\forall i \geq 0$, generally, the iterative performance index function $V_i(x_k)$ is not optimal. However, we can prove that $J^*(x_k)$ is the limit of $V_i(x_k)$ as $i \to \infty$. 

Lemma 3.1: Let $\eta_i(x_k), i = 0, 1, \ldots$, be an arbitrary control law, and $v_i(x_k)$ be expressed as (10). For $\forall x_k$, define $V_{i+1}(x_k)$ as (11) and $A_{i+1}(x_k)$ as

$$A_{i+1}(x_k) = U(x_k, \eta_i(x_k)) + A_i(x_{k+1}).$$

(12)

where $A_0(x_k) = V_0(x_k) = \theta \Psi(x_k)$, then $V_i(x_k) \leq A_i(x_k)$.

Corollary 3.1: Let $\mu(x_k) \in \mathcal{A}_k$ be an arbitrary stable control law, and for $i = 0, 1, \ldots$, define $V_{i+1}(x_k)$ as (11). For $\forall x_k$, define a new performance index function $P_i(x_k)$ as

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

(13)

with $P_0(x_k) = V_0(x_k) = \theta \Psi(x_k)$, then $V_i(x_k) \leq P_i(x_k)$.

Theorem 3.1: Let $x_k$ be an arbitrary state vector. The iterative performance index function $V_{i+1}(x_k)$ and iterative control law $v_i(x_k)$ are obtained by (8)–(11). If Assumptions 1–4 hold, then for any finite $i = 0, 1, \ldots$, there exists a finite $\theta > 0$ that makes

$$V_{i+1}(x_k) \leq V_i(x_k)$$

(14)

hold.

Proof: To obtain the conclusion, we will show that for arbitrary finite $i = 0, 1, \ldots$, there exists a finite $\theta_i > 0$ that makes (14) hold. We prove this by mathematical induction.

First, we let $i = 0$. Let $\mu(x_k) \in \mathcal{A}_0$ be an arbitrary stable control law. Define the performance index function $P_1(x_k)$ as (13). For $i = 0$, we have

$$P_1(x_k) = U(x_k, \mu(x_k)) + P_0(x_{k+1})$$

$$= U(x_k, \mu(x_k)) + \theta \Psi(F(x_k, \mu(x_k))).$$

(15)

According to Definition 3.1, there exists a stable control law $\bar{v}_k = \bar{v}(x_k)$ which satisfies

$$\Psi(x_k) - \Psi(F(x_k, \bar{v}(x_k))) > 0.$$

(16)

As $\bar{v}(x_k)$ is a stable control law, we have that the utility function $U(x_k, \bar{v}(x_k))$ is finite. Then, there exists a finite $\theta_0 > 0$ which satisfies

$$\theta_0 \Psi(x_k) - \Psi(F(x_k, \bar{v}(x_k))) \geq U(x_k, \bar{v}(x_k)).$$

(17)

As $\mu(x_k) \in \mathcal{A}_k$ is arbitrary, we can let $\mu(x_k) = \bar{v}(x_k)$. If we let $\theta = \theta_0$ and $P_0(x_k) = V_0(x_k) = \theta_0 \Psi(x_k)$, then we can get

$$\theta_0 \Psi(x_k) \geq U(x_k, \bar{v}(x_k)) + \theta_0 \Psi(F(x_k, \bar{v}(x_k))) = P_1(x_k)$$

(18)

hold. According to Corollary 3.1, we have

$$V_1(x_k) = \min_k \{U(x_k, \eta_k) + P_0(x_{k+1})\}$$

$$\leq P_1(x_k) = U(x_k, \bar{v}(x_k)) + P_0(x_{k+1}).$$

(19)

As $P_0(x_k) = V_0(x_k) = \theta_0 \Psi(x_k)$, then there exists a finite constant $\theta_0 > 0$ which satisfies

$$V_0(x_k) = \theta_0 \Psi(x_k) \geq P_1(x_k) \geq V_1(x_k),$$

(20)

which proves

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

(21)

Hence, the conclusion holds for $i = 0$. Assume that for $i = l - 1$, there exists a finite $\theta_{l-1}$ that makes (14) hold. Now, we consider the situation for $i = l$. According to Lemma 3.1, for $\forall \theta_i > 0$, the iterative performance index function $V_l(x_k)$ can be expressed as

$$V_l(x_k) = \sum_{j=0}^{l-1} U(x_{k+j}, v_j(x_{k+j})) + \theta_i \Psi(x_{k+l}),$$

(22)

where $v_l(x_k)$ is the iterative control law satisfying (10), and $V_0(x_k) = F(x_k, \bar{v}(x_k))$. Define the performance index function $P_{l+1}(x_k)$ as

$$P_{l+1}(x_k) = U(x_k, v_{l-1}(x_k)) + U(x_{k+1}, v_{l-2}(x_{k+1})) + \cdots + U(x_{k+l-1}, v_0(x_{k+l-1}))$$

$$+ U(x_{k+l}, \mu(x_{k+l})) + P_0(x_{k+l+1})$$

$$= \sum_{j=0}^{l-1} U(x_{k+j}, v_j(x_{k+j})) + U(x_{k+l}, \mu(x_{k+l})) + \theta_l \Psi(x_{k+l+1}),$$

(23)

where $\mu(x_{k+l}) \in \mathcal{A}_{k+l}$. According to Definition 3.1, there exists a stable control law $\bar{v}(x_{k+l}) \in \mathcal{A}_{k+l}$ which satisfies

$$\Psi(x_{k+l}) - \Psi(F(x_{k+l}, \bar{v}(x_{k+l}))) > 0.$$

(24)

So, we have that there exists a finite $\theta_l$ satisfies

$$\theta_l(\Psi(x_{k+l}) - \Psi(F(x_{k+l}, \bar{v}(x_{k+l}))) \geq U(x_{k+l}, \bar{v}(x_{k+l})).$$

(25)

Let $\mu(x_{k+l}) = \bar{v}(x_{k+l})$, and $\theta_l = \theta_l$. Then, according to (22) and (23) we can get

$$V_l(x_k) = \sum_{j=0}^{l-1} U(x_{k+j}, v_{l-1}(x_{k+j})) + \theta_l \Psi(x_{k+l})$$

$$\geq \sum_{j=0}^{l-1} U(x_{k+j}, v_{l-1}(x_{k+j})) + U(x_{k+l}, \bar{v}(x_{k+l}))$$

$$+ \theta_l \Psi(x_{k+l+1})$$

$$= P_{l+1}(x_k).$$

(26)

According to Lemma 3.1, we have $V_{l+1}(x_k) \leq P_{l+1}(x_k)$. Therefore, we can obtain

$$V_{l+1}(x_k) \leq V_l(x_k).$$

(27)

The mathematical induction is completed.

On the other hand, as $i$ is finite, if we let $\bar{\theta} = \max\{\theta_0, \theta_1, \ldots, \theta_i\}$, then we can choose an arbitrary finite $\theta$ that satisfies $\theta \geq \bar{\theta}$ to make (14) hold and the proof is completed.

Theorem 3.2: Let $x_k$ be an arbitrary state vector. If Assumptions 1–4 hold and there exists a control law $\bar{v}(x_k) \in \mathcal{A}_k$ which satisfies (7), that makes the following limit

$$\lim_{x_k \to 0} \frac{U(x_k, \bar{v}(x_k))}{\Psi(x_k) - \Psi(F(x_k, \bar{v}(x_k)))}$$

exist, then for $\forall i = 0, 1, \ldots$, there exists a finite $\theta > 0$ that makes (14) hold.
Proof: According to (25) in Theorem 3.1, we can see that for any finite \(i = 0, 1, \ldots\), the parameter \(\theta_i\) should satisfy
\[
\theta_i \geq \frac{U(x_{k+i}, \bar{v}(x_{k+i}))}{\Psi(x_{k+i}) - \Psi(F(x_{k+i}, \bar{v}(x_{k+i}))}
\] (29)
to make (14) hold. Let \(i \to \infty\), we have
\[
\lim_{i \to \infty} \theta_i \geq \lim_{i \to \infty} \frac{U(x_{k+i}, \bar{v}(x_{k+i}))}{\Psi(x_{k+i}) - \Psi(F(x_{k+i}, \bar{v}(x_{k+i}))}
\] (30)
We can see that if the limit of the right hand in (30) exists, then we have \(\theta_\infty = \lim_{i \to \infty} \theta_i\) can be defined. Therefore, if we define
\[
\bar{\theta} = \max\{\theta_0, \theta_1, \ldots, \theta_\infty\},
\] (31)
then \(\bar{\theta}\) can be well defined. Hence, we can choose an arbitrary finite \(\theta\) which satisfies
\[
\theta \geq \bar{\theta},
\] (32)
to make (14) hold.

On the other hand, \(\bar{v}(x_k) \in \mathfrak{A}_k\) is a stable control law. We have \(x_k \to 0\) as \(k \to \infty\) under the stable control sequence \((\bar{v}_k, v_{k+1}, \ldots)\), where \(v_k = \bar{v}(x_k)\) for \(\forall k = 0, 1, \ldots\), Then, if (28) holds, according to (30) and (31), there exists a finite \(\theta\) that makes the following inequality
\[
\theta \geq \lim_{x_k \to 0} \frac{U(x_k, \bar{v}(x_k))}{\Psi(x_k) - \Psi(F(x_k, \bar{v}(x_k)))}
\] (33)
hold. The proof is completed.

There is an important property we must point out. If we want to obtain \(\theta\) according to Theorem 3.2, it is unnecessary to take all the stable control laws in \(\mathfrak{A}_k\) to determine the limit of (28). Actually, if there exists one stable control law \(\bar{v}(x_k)\), \(\in \mathfrak{A}_k\) that makes the limit (28) exist, then we can say that the optimal control law \(u^*(x_k)\) makes the limit (28) exist. The following theorem will show this property.

Theorem 3.3: Let \(x_k\) be an arbitrary state vector. Let Assumptions 1–4 hold. Let \(\bar{v}(x_k) \in \mathfrak{A}_k\) be an arbitrary stable control law that makes the limit (28) exist. If let \(\theta^*\) be expressed as
\[
\theta^* = \lim_{x_k \to 0} \frac{U(x_k, u^*(x_k))}{\Psi(x_k) - \Psi(F(x_k, u^*(x_k))},
\] (34)
then we have
\[
\theta^* = \inf \left\{ \lim_{x_k \to 0} \frac{U(x_k, \bar{v}(x_k))}{\Psi(x_k) - \Psi(F(x_k, \bar{v}(x_k))} \right\}
\] (35)
Proof: As \(\bar{v}(x_k) \in \mathfrak{A}_k\) is an arbitrary stable control law, then according to (35), we have
\[
\theta^* \geq \inf \left\{ \lim_{x_k \to 0} \frac{U(x_k, \bar{v}(x_k))}{\Psi(x_k) - \Psi(F(x_k, \bar{v}(x_k))} \right\}.
\] (36)
As \(\bar{v}(x_k)\) is a stable control law, we know that \(\bar{v}(x_k)\) can stabilize the system \(F(x_k, \bar{v}(x_k))\) to zero as \(k \to \infty\). Without loss of generality, let \(F(x_{N-1}, \bar{v}(x_{N-1})) = x_N\), where \(x_N = 0\) and \(N \to \infty\).

According to Bellman’s principle, we can define a performance index function as
\[
\mathcal{P}(x_{N-1}) = U(x_{N-1}, \bar{v}(x_{N-1})) + \mathcal{P}(x_N)
= U(x_{N-1}, \bar{v}(x_{N-1}))
\] (37)
where \(\mathcal{P}(x_N) = 0\) as \(x_N = 0\). Then, the optimal performance index function can be expressed as
\[
J^*(x_{N-1}) = U(x_{N-1}, u^*(x_{N-1})) + J^*(x_N)
= U(x_{N-1}, u^*(x_{N-1}))
\] (38)
where \(J^*(x_{N-1}) = 0\) as \(x_N = 0\). Then, we have
\[
J^*(x_{N-1}) = U(x_{N-1}, u^*(x_{N-1}))
\leq U(x_{N-1}, \bar{v}(x_{N-1})) = \mathcal{P}(x_N).
\] (39)
On the other hand, we have that \(F(x_{N-1}, v(x_{N-1})) = F(x_{N-1}, u^*(x_{N-1})) = x_N = 0\) where \(N \to \infty\). So, for \(\forall v(x_{N-1}) \in \mathfrak{A}_{N-1}\), we have
\[
\theta^* \leq \frac{U(x_{N-1}, \bar{v}(x_{N-1}))}{\Psi(x_{N-1}) - \Psi(F(x_{N-1}, \bar{v}(x_{N-1}))}
\] (40)
As \(\bar{v}(x_k)\) is an arbitrary stable control law and \(N \to \infty\), then we can obtain
\[
\theta^* \leq \lim_{N \to \infty} \frac{U(x_{N-1}, \bar{v}(x_{N-1}))}{\Psi(x_{N-1}) - \Psi(F(x_{N-1}, \bar{v}(x_{N-1}))}
\leq \inf \left\{ \lim_{N \to \infty} \frac{U(x_{N-1}, \bar{v}(x_{N-1}))}{\Psi(x_{N-1}) - \Psi(F(x_{N-1}, \bar{v}(x_{N-1}))} \right\}
\leq \inf \left\{ \lim_{x_k \to 0} \frac{U(x_k, \bar{v}(x_k))}{\Psi(x_k) - \Psi(F(x_k, \bar{v}(x_k))} \right\}.
\] (41)
Therefore, we have (35) holds.

From Theorem 3.1 to Theorem 3.3, we can see that if there exists a finite \(\theta\) that makes (14) hold, then we have \(V_i(x_k) \geq 0\) is a nonincreasing sequence with lower bound for iteration index \(i = 0, 1, \ldots\). Then, we can derive the following theorem.

Theorem 3.4: Let \(x_k\) be an arbitrary state vector. Define the performance index function \(V_\infty(x_k)\) as the limit of the iterative function \(V_i(x_k)\), i.e.,
\[
V_\infty(x_k) = \lim_{i \to \infty} V_i(x_k).
\] (42)
Then, we have
\[
V_\infty(x_k) = \min_{u_k} \{U(x_k, u_k) + V_\infty(x_{k+1})\}.
\] (43)
Proof: Let \(\mu(x_k)\) be an arbitrary stable control law. According to Theorem 3.1, for \(\forall i\), we have
\[
V_\infty(x_k) \leq V_{i+1}(x_k) \leq U(x_k, \mu(x_k)) + V_i(x_{k+1}).
\] (44)
Let \(i \to \infty\), we have
\[
V_\infty(x_k) \leq U(x_k, \mu(x_k)) + V_\infty(x_{k+1}).
\] (45)
According to (53) and (56), we have

\[ V_p(x_k) - \epsilon \leq V_\infty(x_k) \leq V_p(x_k). \] (47)

Then, we let

\[ V_p(x_k) = \min_{u_k} \{ U(x_k, u_k) + V_{p-1}(x_{k+1}) \} \]

\[ = U(x_k, v_{p-1}(x_k)) + V_{p-1}(x_{k+1}). \] (48)

Hence

\[ V_\infty(x_k) \geq U(x_k, v_{p-1}(x_k)) + V_{p-1}(x_{k+1}) - \epsilon \]

\[ \geq \min_{u_k} \{ U(x_k, u_k) + V_\infty(x_{k+1}) \} - \epsilon. \] (49)

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

\[ V_\infty(x_k) \geq \min_{u_k} \{ U(x_k, u_k) + V_\infty(x_{k+1}) \}. \] (50)

Combining (46) and (50) we have

\[ V_\infty(x_k) = \min_{u_k} \{ U(x_k, u_k) + V_\infty(x_{k+1}) \}, \] (51)

which proves the conclusion of this theorem.

**Theorem 3.5:** Let the performance index function \( V_i(x_k) \) be defined by (11) where \( \theta \) satisfies (32). 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.,

\[ V_i(x_k) \to J^*(x_k). \] (52)

**Proof:** According to the definition of \( J^*(x_k) \) in (3), we have \( J^*(x_k) \leq V_i(x_k) \). Then, let \( i \to \infty \), we have

\[ J^*(x_k) \leq V_\infty(x_k). \] (53)

Let \( \epsilon > 0 \) be an arbitrary positive number, and we have that

\[ V_q(x_k) \leq P_q(x_k) \leq J^*(x_k) + \epsilon. \] (54)

On the other hand, according to Theorem 3.1, we have

\[ V_\infty(x_k) \leq V_q(x_k) \] (55)

holds. Combining (54) and (55), we have \( V_\infty(x_k) \leq J^*(x_k) + \epsilon \). As \( \epsilon \) is arbitrary positive number, we have

\[ V_\infty(x_k) \leq J^*(x_k). \] (56)

According to (53) and (56), we have

\[ V_\infty(x_k) = J^*(x_k). \] (57)

The proof is completed.

As is known, the stability property of control systems is a most basic and necessary property for any control systems. So, in the following part, we will give the stability analysis for the system (1) under the iterative ADP algorithm (8)–(11).

**Theorem 3.6:** Let \( x_k \) be an arbitrary controllable state. For \( i = 0, 1, \ldots \), if Assumptions 1–4 hold and the iterative performance index function \( V_i(x_k) \) and iterative control law \( v_i(x_k) \) are defined by (8)–(11) where \( \theta \) satisfies (32), then we have for all \( i \), \( v_i(x_k) \) is an asymptotically stable control law for system (1).

**Proof:** The Theorem will be proved by two steps.

1) Show that for all \( i = 0, 1, \ldots \), the iterative performance index function \( V_i(x_k) \) is a positive definite function.

For the iterative \( \theta \)-ADP algorithm, for \( i = 0 \), we have \( V_0(x_k) = \theta \Psi(x_k) \). According to Assumption 4 of this paper, we have \( V_0(x_k) \) is a positive definite function for \( i = 0 \).

Assume that for \( i \), \( V_i(x_k) \) is a positive definite function. Then for \( i + 1 \), we have (11) holds. Then let \( x_k = 0 \), we can get

\[ V_{i+1}(0) = U(0, v_i(0)) + V_i(F(0, v_i(0))). \] (58)

According to Assumption 1–4, we have \( v_i(0) = 0 \), \( F(0, v_i(0)) = 0 \), \( U(0, v_i(0)) = 0 \). As \( V_i(x_k) \) is a positive definite function, we have \( V_i(0) = 0 \). Then we have \( V_{i+1}(0) = 0 \).

On the other hand, let \( \|x_k\| \to \infty \), as \( U(x_k, u_k) \) is a positive definite function, \( V_{i+1}(x_k) \to \infty \). So, \( V_{i+1}(x_k) \) is a positive definite function. The mathematical induction is completed.

2) Show that \( v_i(x_k) \) is an asymptotically stable control law for system (1).

As for all \( i = 0, 1, \ldots \), the iterative performance index function \( V_i(x_k) \) is a positive definite function, then according to (11), we have

\[ V_i(F(x_k, v_i(x_k)) - V_{i+1}(x_k) = -U(x_k, v_i(x_k)) \leq 0. \] (59)

According to Theorem 3.1, we have \( V_{i+1}(x_k) \leq V_i(x_k) \) for all \( i \geq 0 \). Then, we can obtain

\[ V_i(F(x_k, v_i(x_k)) - V_{i+1}(x_k) \leq V_i(F(x_k, v_i(x_k)) - V_i(x_k) \]

\[ = -U(x_k, v_i(x_k)) \leq 0. \] (60)

According to Lyapunov theorem [7], we have for all \( i = 0, 1, \ldots \), the iterative performance index function \( V_i(x_k) \) is a Lyapunov function. Therefore, the conclusion is proved.

**IV. SIMULATION STUDY**

In this section, an example is provided to demonstrate the effectiveness of the proposed control scheme. Our example is chosen as the example in [9], [15] with modifications. We consider the following nonlinear system

\[ x_{k+1} = f(x_k) + g(x_k)u_k, \] (61)

where \( x_k = [x_{1k}, x_{2k}]^T \) and \( u_k = [u_{1k}, u_{2k}]^T \) are the state and control variables, respectively. The system functions are given as

\[ f(x_k) = \begin{bmatrix} 0.8x_{1k} \exp(x_{2k}^2) \\ 0.9x_{2k}^2 \end{bmatrix}, \quad g(x_k) = \begin{bmatrix} -0.2 & 0 \\ 0 & -0.2 \end{bmatrix}. \]

The initial state is \( x_0 = [1, -1]^T \). The performance index function is in quadratic form expressed as \( J(x_0, u_0^\infty) = \)}
the neural network training error becomes less than 10^-8.

Neural networks are used to implement the iterative $\theta$-ADP algorithm. The critic network and the action network are chosen as three-layer BP neural networks with the structures of 2–8–1 and 2–8–2, respectively. For each iterative step, the critic network and the action network are trained for 2000 iteration steps using the learning rate of $\alpha = 0.01$ so that the neural network training error becomes less than 10^-8. To show the effectiveness of the iterative ADP algorithm, we choose four $\theta$s (include $\theta = 3.5, 5, 7, 10$) to initialize the algorithm. Let the algorithm run for 35 iterative steps for the different $\theta$ to obtain the optimal performance index function. The convergence curves of the performance index functions are shown in Fig. 1.

We apply the optimal control law to the system for $T_f = 20$ time steps and obtain the following results. The optimal state and control trajectories are shown in Fig. 2(a) and Fig. 2(b), respectively.

V. CONCLUSION

In this paper, we propose an effective stable value iteration algorithm to find the infinite horizon optimal control for discrete-time nonlinear systems. In the proposed algorithm, any of the iterative control is stable for the nonlinear system which makes the proposed algorithm be used both on-line and off-line. Convergence analysis of the performance index function for the iterative ADP algorithm is proved and the stability proofs are also given. Finally, a simulation example is given to illustrate the performance of the proposed algorithm.

REFERENCES


