A Novel Infinite-Time Optimal Tracking Control Scheme for a Class of Discrete-Time Nonlinear Systems via the Greedy HDP Iteration Algorithm

Huaguang Zhang, *Senior Member, IEEE*, Qinglai Wei, and Yanhong Luo

***Abstract*—In this paper, we aim to solve the infinite-time optimal tracking control problem for a class of discrete-time nonlinear systems using the greedy heuristic dynamic programming (HDP) iteration algorithm. A new type of performance index is defined because the existing performance indexes are very difficult in solving this kind of tracking problem, if not impossible. Via system transformation, the optimal tracking problem is transformed into an optimal regulation problem, and then, the greedy HDP iteration algorithm is introduced to deal with the regulation problem with rigorous convergence analysis. Three neural networks are used to approximate the performance index, compute the optimal control policy, and model the nonlinear system for facilitating the implementation of the greedy HDP iteration algorithm. An example is given to demonstrate the validity of the proposed optimal tracking control scheme.**

***Index Terms*—Convergence, greedy heuristic dynamic programming (HDP) iteration, infinite time, neural network, optimal tracking control.**

I. Introduction

THE OPTIMAL tracking problem of nonlinear systems has always been the key focus in the control field in the last several decades. Traditional optimal tracking control is mostly implemented by feedback linearization [9] or plant inversion [6]. However, on one hand, the controller designed by feedback linearization technique is only effective in the neighborhood of the equilibrium point. On the other hand, the exact inversion model of the plant is quite difficult to obtain, if not impossible. Therefore, it is necessary to study the direct optimal tracking control approach for the original nonlinear system. However, the earlier optimal control laws for nonlinear systems are mostly open-loop [1]. The difficulty for closed-loop optimal feedback control lies in solving the time-varying Hamilton-Jacobi-Bellman (HJB) equation which is usually too hard to solve analytically. To overcome the difficulty, some methods are proposed in [5], [8] and [15] for continuous-time systems and [17] for the finite-time tracking problem. To the best of our knowledge, there are no results on the infinite-time optimal tracking control for discrete-time (DT) nonlinear systems. For infinite-time optimal tracking problem, the main difficulty lies in the definition of performance index, because the performance index indices defined for finite-time tracking problem become invalid in infinite horizon. To overcome this difficulty, the definition of a new type of performance index will be introduced in this paper.

The approximate dynamic programming (ADP) algorithm, a powerful technique to solve optimal control problems, has been studied by many researchers [4], [7], [10]–[14], [16], [18]. However, most results are focus on the optimal regulation problem. There have been no results discussing how to use ADP to solve the infinite-time optimal tracking control problem for nonlinear systems, much less for DT systems. It is the first time that the optimal tracking control problem is solved by ADP algorithm, especially by the greedy Heuristic Dynamic Programming (HDP) iteration algorithm. In this paper, we will first transform the tracking problem into an optimal regulation problem, and then the greedy HDP iteration algorithm can be properly introduced to deal with this regulation problem. The main contributions of this paper include: 1) the proposal of a new type of performance index for infinite-time optimal tracking problems and 2) the development of an optimal tracking control law by tracking error and the transformation of the optimal tracking problem into an optimal regulation problem for DT nonlinear system.

II. Problem Formulation

Consider a class of affine nonlinear systems of the form

$$x(k+1) = f(x(k)) + g(x(k))\,u(x(k)) \tag{1}$$

where $x(k) \in \Re^n$, $f(x(k)) \in \Re^n$, $g(x(k)) \in \Re^{n\times m}$, and the input $u(x(k)) \in \Re^m$. Here, assume that the system is strongly controllable on $\Omega \subset \Re^n$.

For infinite-time optimal tracking problems, the control objective is to design optimal control $u(x(k))$ for (1), such that the state $x(k)$ tracks the specified desired trajectory $\eta(k) \in \Re^n$, $k = 0, 1, \ldots$, where we assume that there exists a function $\phi : \Re^n \to \Re^n$ satisfies $\eta(k+1) = \phi(\eta(k))$. We further assume that the mapping between the state $x(k)$ and the desired trajectory $\eta(k)$ is one-to-one. In the following part, for simplicity, $u(x(k))$ is denoted by $u(k)$.

Define the tracking error as follows:

$$\overline{z}(k) = x(k) - \eta(k). \tag{2}$$

Manuscript received August 30, 2007; revised January 26, 2008. This work was supported in part by the National Natural Science Foundation of China under Grants 60534010, 60572070, 60774048, and 60728307, by the Program for Changjiang Scholars and Innovative Research Groups of China under Grant 60521003, and by the National High Technology Research and Development Program of China under Grant 2006AA04Z183. This paper was recommended by Guest Editor F. L. Lewis.

The authors are with the School of Information Science and Engineering, and the Key Laboratory of Integrated Automation of Process Industry Northeastern University, Shenyang 110004, China (e-mail: hgzhang@ieee.org; qinglai_wei@163.com; neuluo@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2008.920269

For the time-invariant optimal tracking problems of linear systems, the performance index is generally defined as the following quadratic form:

$$J(\overline{z}(0),u)=\sum_{k=0}^{\infty}\Big\{\overline{z}^{\mathrm{T}}(k)\overline{Q}\overline{z}(k)+(u(k+1)-u(k))^{\mathrm{T}}\times R(u(k+1)-u(k))\Big\}. \quad (3)$$

However, for time-variant tracking problems in nonlinear environment, the problem is much more complex, and the aforementioned performance index may be invalid, i.e., $J(\overline{z}(0),u)$ calculated by (3) may be infinity because the control $u(k)$ depends on the desired trajectory $\eta(k)$.

To solve this problem, we present the following performance index, which is derived from [17] and [20]:

$$J(\overline{z}(0),w)=\sum_{k=0}^{\infty}\Big\{\overline{z}^{\mathrm{T}}(k)\overline{Q}\overline{z}(k)+(w(k)-w(k-1))^{\mathrm{T}}\times R(w(k)-w(k-1))\Big\} \quad (4)$$

where $\overline{Q}\in\Re^{n\times n}$ and $R\in\Re^{m\times m}$ are both diagonal positive definite matrices and $w(k)$ is defined as follows:

$$w(k)=u(k)-u_e(k) \quad (5)$$

and $w(k)=0$ for $k<0$. $u_e(k)$ denotes the expected control, which can be given as follows:

$$u_e(k)=g^{-1}(\eta(k))(\eta(k+1)-f(\eta(k))) \quad (6)$$

where $g^{-1}(\eta(k))g(\eta(k))=I$ and $I\in\Re^{m\times m}$ is the identity matrix.

In (4), the first term means the tracking error and the second term means the difference of the control error. However, there still exists a problem that if only the tracking error and the difference of the control error are considered in the performance index, the system may oscillate. For example, the difference of the control error may be small but the error between the real tracking control and the expected tracking control may be large.

Therefore, it is necessary to define a new type of performance index for solving this kind of infinite-time optimal tracking problem. We propose a new type of quadratic performance index as follows:

$$J(\overline{z}(0),u,w)=\sum_{k=0}^{\infty}\Big\{\overline{z}^{\mathrm{T}}(k)\overline{Q}\overline{z}(k)+(u(k)-u_e(k))^{\mathrm{T}}\times S(u(k)-u_e(k))+(w(k)-w(k-1))^{\mathrm{T}}R(w(k)-w(k-1))\Big\} \quad (7)$$

where $S\in\Re^{m\times m}$ is also a diagonal positive definite matrix and the other parameters are the same as those of (4).

Compared with (4), a new term $(u(k)-u_e(k))^{\mathrm{T}}S(u(k)-u_e(k))$ is added to represent the error between the real and expected tracking controls. Via this performance index, not only the tracking error and the difference of the control error are considered but also the error between the real and expected tracking controls; therefore, the oscillation can be well prevented.

After the definition of such new type of performance index, in the following section, we mainly discuss the optimal tracking control scheme based on the greedy HDP iteration algorithm.

III. Optimal Tracking Control Scheme Based on Greedy HDP Iteration Algorithm

A. System Transformation

For simplicity, denote $v(k)=w(k)-w(k-1)$. Noticing that $v(0)=w(0)$, thus, we can obtain

$$w(k)=v(k)+v(k-1)+\cdots+v(0). \quad (8)$$

Define $z(k)=\left[\overline{z}^{\mathrm{T}}(k),\eta^{\mathrm{T}}(k)\right]^{\mathrm{T}}$. Thus, (7) can be written as follows:

$$J(z(k),v)=\sum_{t=k}^{\infty}\Big\{z^{\mathrm{T}}(t)Qz(t)+v^{\mathrm{T}}(t)Rv(t)+[v(t)+v(t-1)+\cdots+v(0)]^{\mathrm{T}}\times S[v(t)+v(t-1)+\cdots+v(0)]\Big\}. \quad (9)$$

where $Q=\begin{bmatrix}\overline{Q} & 0^{n\times n}\\ 0^{n\times n} & 0^{n\times n}\end{bmatrix}$. According to (5) and (6), we can obtain

$$\overline{z}(k+1)=-\phi(\eta(k))+f(\overline{z}(k)+\eta(k))+g(\overline{z}(k)+\eta(k))(v(k)+v(k-1)+\ldots+v(0))-g(\overline{z}(k)+\eta(k))g^{-1}(\eta(k))\times(f(\eta(k))-\phi(\eta(k))). \quad (10)$$

Considering $z(k+1)=\left[\overline{z}^{\mathrm{T}}(k+1),\eta^{\mathrm{T}}(k+1)\right]^{\mathrm{T}}$, we have

$$z(k+1)=\overline{f}(z(k))+\overline{g}(z(k))(v(k)+\ldots+v(0)) \quad (11)$$

where

$$\overline{f}(z(k))=\big[\big(-\phi(\eta(k))-g(\overline{z}(k)+\eta(k))g^{-1}(\eta(k))\times(f(\eta(k))-\phi(\eta(k)))+f(\overline{z}(k)+\eta(k))\big)^{\mathrm{T}},\phi^{\mathrm{T}}(\eta(k))\big]^{\mathrm{T}}$$
$$\overline{g}(z(k))=\left[g^{\mathrm{T}}(\overline{z}(k)+\eta(k)),0^{m\times n}\right]^{\mathrm{T}}.$$

Therefore, the optimal tracking problem of (1) is transformed into the optimal regulation problem of (11) with respect to (9). Here, the optimal controller is designed by the state feedback of the transformed system (11). Thus, our next work is to design a controller $v(z(k))$ for (11) to regulate $\overline{z}(k)$ and simultaneously to guarantee (9) is finite, i.e., admissible control [3].

B. Derivation of the Greedy HDP Iteration Algorithm

In this section, we will design the optimal tracking controller using the greedy HDP iteration algorithm.

According to the Bellman optimality principle and (9), the HJB equation comes out to be as follows:

$$J^*(z(k))=\min_{v(k)}\Big\{z^{\mathrm{T}}(k)Qz(k)+v^{\mathrm{T}}(k)Rv(k)+[v(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}\times S[v(k)+v(k-1)+\cdots+v(0)]+J^*(z(k+1))\Big\} \quad (12)$$

where $J^*(z(k))$ is the optimal performance index of the optimal tracking problem.

In the greedy HDP iteration algorithm, the performance index and control policy are updated by recurrent iteration, with the iteration number i increasing from 0 to ∞.

First, we start with initial performance index $J_0(z(k)) = 0$, and the control $v_0(k)$ can be computed as follows:

$$\begin{aligned} v_0(k) = \arg\min_{v(k)} \Big\{ & z^{\mathrm{T}}(k)Qz(k) + v^{\mathrm{T}}(k)Rv(k) \\ & + [v(k) + v(k-1) + \cdots + v(0)]^{\mathrm{T}} \\ & \times S\,[v(k) + v(k-1) + \cdots + v(0)] \\ & + J_0\,(z(k+1)) \Big\}. \end{aligned} \tag{13}$$

As there is no constraint for the system, we can obtain

$$v_0(k) = -(R+S)^{-1} S\,(v(k-1) + \cdots + v(0))\,. \tag{14}$$

Then, we update the performance index as follows:

$$\begin{aligned} J_1\,(z(k)) = & z^{\mathrm{T}}(k)Qz(k) + (v(k-1) + \cdots + v(0))^{\mathrm{T}} \\ & \times S(R+S)^{-1}R(R+S)^{-1} \\ & \times S\,(v(k-1) + \cdots + v(0)) \\ & + \big(-(R+S)^{-1}S\,(v(k-1) + \cdots + v(0)) \\ & \quad + v(k-1) + \cdots + v(0)\big)^{\mathrm{T}} \\ & \times S\,\big(-(R+S)^{-1}S\,(v(k-1) + \cdots + v(0)) \\ & \quad + v(k-1) + \cdots + v(0)\big)\,. \end{aligned} \tag{15}$$

Thus, for $i = 1, 2, \ldots$, the greedy HDP iteration algorithm can be used to implement the iteration between the following:

$$\begin{aligned} v_i(k) = -\frac{1}{2}(R+S)^{-1} \Bigg(& 2S\,(v(k-1) + \cdots + v(0)) \\ & + \overline{g}^{\mathrm{T}}\,(z(k))\,\frac{dJ_i\,(z(k+1))}{dz(k+1)} \Bigg) \end{aligned} \tag{16}$$

and

$$\begin{aligned} J_{i+1}\,(z(k)) = & z^{\mathrm{T}}(k)Qz(k) \\ & + \frac{1}{4}\Bigg(2S\,(v(k-1) + \cdots + v(0)) \\ & \quad + \overline{g}^{\mathrm{T}}\,(z(k))\,\frac{dJ_i\,(z(k+1))}{dz(k+1)}\Bigg)^{\mathrm{T}} \\ & \times (R+S)^{-1}R(R+S)^{-1} \\ & \times \Bigg(2S\,(v(k-1) + \cdots + v(0)) \\ & \quad + \overline{g}^{\mathrm{T}}\,(z(k))\,\frac{dJ_i\,(z(k+1))}{dz(k+1)}\Bigg) \\ & + \Bigg(-\frac{1}{2}(R+S)^{-1} \\ & \quad \times \Bigg(2S\,(v(k-1) + \cdots + v(0)) \\ & \quad\quad + \overline{g}^{\mathrm{T}}\,(z(k))\,\frac{dJ_i\,(z(k+1))}{dz(k+1)}\Bigg) \\ & \quad + v(k-1) + \cdots + v(0)\Bigg)^{\mathrm{T}} \\ & \times S\Bigg(-\frac{1}{2}(R+S)^{-1} \\ & \quad \times \Bigg(2S\,(v(k-1) + \cdots + v(0)) \\ & \quad\quad + \overline{g}^{\mathrm{T}}\,(z(k))\,\frac{dJ_i\,(z(k+1))}{dz(k+1)}\Bigg) \\ & \quad + v(k-1) + \cdots + v(0)\Bigg) \\ & + J_i\,(z(k+1))\,. \end{aligned} \tag{17}$$

In the following section, we present a proof of convergence of the iteration between (16) and (17), with the performance index $J_i(z(k)) \to J^*(z(k))$ and $v_i(k) \to v^*(k)$, as $i \to \infty$, $\forall k$.

Lemma 1 (cf., [2]): Let $\tilde{v}_i(k)$, $k = 0, 1, \ldots$ be any sequence of control and $v_i(k)$ is expressed as (16). Define $J_{i+1}(z(k))$ as (17) and $\Lambda_{i+1}(z(k))$ as follows:

$$\begin{aligned} \Lambda_{i+1}(z(k)) = & z^{\mathrm{T}}(k)Qz(k) + \tilde{v}_i^{\mathrm{T}}(k)R\tilde{v}_i(k) \\ & + [\tilde{v}_i(k) + v(k-1) + \cdots + v(0)]^{\mathrm{T}} \\ & \times S[\tilde{v}_i(k) + v(k-1) + \cdots + v(0)] + \Lambda_i\,(z(k+1))\,. \end{aligned} \tag{18}$$

If $J_0(z(k)) = \Lambda_0(z(k)) = 0$, then, $J_i(z(k)) \le \Lambda_i(z(k))$, $\forall i$.

In order to prove the convergence of the performance index, the following theorem is also necessary.

Theorem 1: Let the sequence $\{J_i(z(k))\}$ be defined by (17). If $\overline{z}(k)$ for the system (10) is strongly controllable, then there is an upper bound Y such that $0 \le J_i(z(k)) \le Y$, $\forall i$.

Proof: Let $\bar{v}(k)$, $k = 0, 1, \ldots$ be any stabilizing and admissible control input. Define a new sequence $\{P_i(z(k))\}$ as follows:

$$\begin{aligned} P_{i+1}\,(z(k)) = & z^{\mathrm{T}}(k)Qz(k) + \bar{v}^{\mathrm{T}}(k)R\bar{v}(k) \\ & + [\bar{v}(k) + v(k-1) + \cdots + v(0)]^{\mathrm{T}} \\ & \times S[\bar{v}(k) + v(k-1) + \cdots + v(0)] + P_i(z(k+1)) \end{aligned} \tag{19}$$

with $P_0(z(k)) = J_0(z(k)) = 0$, $J_i(z(k))$ is updated by (17). Thus, we can obtain

$$\begin{aligned} P_{i+1}\,(z(k)) - P_i\,(z(k)) &= P_i\,(z(k+1)) - P_{i-1}\,(z(k+1)) \\ &\vdots \\ &= P_1\,(z(k+i)) - P_0\,(z(k+i))\,. \end{aligned} \tag{20}$$

Because $P_0(z(k+i)) = 0$, we have

$$\begin{aligned} P_{i+1}\,(z(k)) &= P_1\,(z(k+i)) + P_i\,(z(k)) \\ &= P_1\,(z(k+i)) + P_1\,(z(k+i-1)) + P_{i-1}\,(z(k)) \\ &= P_1\,(z(k+i)) + P_1\,(z(k+i-1)) \\ &\quad + P_1\,(z(k+i-2)) + \cdots + P_1\,(z(k)) \\ &= \sum_{j=0}^{i} P_1\,(z(k+j))\,. \end{aligned} \tag{21}$$

According to (19), (21) can be written as follows:

$$\begin{aligned}P_{i+1}(z(k)) &= \sum_{j=0}^{i}\Big\{z^{\mathrm{T}}(k+j)Qz(k+j)+\bar{v}^{\mathrm{T}}(k+j)\\ &\quad\times R\bar{v}(k+j)+[\bar{v}(k+j)+\bar{v}(k+j-1)\\ &\quad+\cdots+\bar{v}(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}\\ &\quad\times S[\bar{v}(k+j)+\bar{v}(k+j-1)+\cdots+\\ &\quad\times\bar{v}(k)+v(k-1)+\cdots+v(0)]\Big\}\\ &\le\sum_{j=0}^{\infty}\Big\{z^{\mathrm{T}}(k+j)Qz(k+j)+\bar{v}^{\mathrm{T}}(k+j)\\ &\quad\times R\bar{v}(k+j)+[\bar{v}(k+j)+\bar{v}(k+j-1)\\ &\quad+\cdots+\bar{v}(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}\\ &\quad\times S[\bar{v}(k+j)+\bar{v}(k+j-1)+\cdots+\\ &\quad\times\bar{v}(k)+v(k-1)+\cdots+v(0)]\Big\}. \end{aligned}\tag{22}$$

Note that the control input $\bar{v}(k)$, $k=0,1,\ldots$ is an admissible control; thus, we can obtain

$$\forall i: P_{i+1}(z(k))\le\sum_{j=0}^{\infty}P_1(z(k+j))\le Y. \tag{23}$$

From Lemma 1, we have

$$\forall i: J_{i+1}(z(k))\le P_{i+1}(z(k))\le Y. \tag{24}$$

■

With Lemma 1 and Theorem 1, the following main theorem can be derived.

Theorem 2: Define the sequence $\{J_i(z(k))\}$ as (17), with $J_0(z(k))=0$. Then, $\{J_i(z(k))\}$ is a nondecreasing sequence in which $J_{i+1}(z(k))\ge J_i(z(k))$, $\forall i$, and converges to the optimal performance index of the DT HJB, i.e., $J_i(z(k))\to J^*(z(k))$ as $i\to\infty$.

Proof: For the convenience of analysis, define a new sequence $\{\Phi_i(z(k))\}$ as follows:

$$\begin{aligned}\Phi_{i+1}(z(k)) &= z^{\mathrm{T}}(k)Qz(k)+v_{i+1}^{\mathrm{T}}(k)Rv_{i+1}(k)\\ &\quad+[v_{i+1}(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}\\ &\quad\times S[v_{i+1}(k)+v(k-1)+\cdots+v(0)]+\Phi_i(z(k+1))\end{aligned}\tag{25}$$

with $v_i(k)$ obtained by (16) and $\Phi_0(z(k))=J_0(z(k))=0$. $J_i(z(k))$ is updated by (17).

In the following section, we prove $\Phi_i(z(k))\le J_{i+1}(z(k))$ by mathematical induction.

First, we prove that it holds for $i=0$. Noting that

$$\begin{aligned}J_1(z(k))-\Phi_0(z(k)) &= z^{\mathrm{T}}(k)Qz(k)+v_0^{\mathrm{T}}(k)Rv_0(k)\\ &\quad+[v_0(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}\\ &\quad\times S[v_0(k)+v(k-1)+\cdots+v(0)]\\ &\ge 0.\end{aligned}\tag{26}$$

Thus, for $i=0$, we can get

$$J_1(z(k))\ge\Phi_0(z(k)). \tag{27}$$

Second, we assume it holds for $i-1$, i.e., $J_i(z(k))\ge\Phi_{i-1}(z(k))$, $\forall z(k)$. Then, for i, because

$$\begin{aligned}\Phi_i(z(k)) &= z^{\mathrm{T}}(k)Qz(k)+v_i^{\mathrm{T}}(k)Rv_i(k)\\ &\quad+[v_i(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}\\ &\quad\times S[v_i(k)+v(k-1)+\cdots+v(0)]+\Phi_{i-1}(z(k+1))\end{aligned}\tag{28}$$

$$\begin{aligned}J_{i+1}(z(k)) &= z^{\mathrm{T}}(k)Qz(k)+v_i^{\mathrm{T}}(k)Rv_i(k)\\ &\quad+[v_i(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}\\ &\quad\times S[v_i(k)+v(k-1)+\cdots+v(0)]+J_i(z(k+1))\end{aligned}\tag{29}$$

thus, we can obtain

$$J_{i+1}(z(k))-\Phi_i(z(k))=J_i(z(k))-\Phi_{i-1}(z(k))\ge 0 \tag{30}$$

i.e.,

$$\Phi_i(z(k))\le J_{i+1}(z(k)). \tag{31}$$

Therefore, the mathematical induction proof is completed.

Moreover, from Lemma 1, we know that $J_i(z(k))\le\Phi_i(z(k))$; therefore, we can obtain

$$J_i(z(k))\le\Phi_i(z(k))\le J_{i+1}(z(k)) \tag{32}$$

which proves that $\{J_i(z(k))\}$ is a nondecreasing sequence bounded by (24). Hence, we conclude that $J_i(z(k))\to J^*(z(k))$ as $i\to\infty$. ■

C. Procedure of the Algorithm

Now, we summarize the greedy HDP iteration algorithm for the nonlinear optimal tracking control problem as follows.

- Step 1) Give $x(0)$, $i_{\max}$, ε, desired trajectory $\eta(k)$, and control sequence $u(0),u(1),\ldots,u(k-1)$.
- Step 2) Compute $z(k)$ according to (2) and $\eta(k)$. Compute $v(0),\ldots,v(k-1)$ according to (5) and (8).
- Step 3) Set $i=0$, $J_0(z(k))=0$.
- Step 4) Compute $v_0(k)$ by (14) and the performance index $J_1(z(k))$ by (15).
- Step 5) Set $i=i+1$.
- Step 6) Compute $v_i(k)$ by (16) and the corresponding performance index $J_{i+1}(z(k))$ by (17).
- Step 7) If $|J_{i+1}(z(k))-J_i(z(k))|<\varepsilon$, then go to Step 9); else, go to Step 8).
- Step 8) If $i>i_{\max}$, then go to Step 9); otherwise, go to Step 5).
- Step 9) Stop.

If the optimal tracking control policy $v(k)$ is obtained under the given accuracy ε, then, we can compute the tracking control input for the original nonlinear system (1) by $u(k)=v(k)+v(k-1)+\cdots+v(0)-g^{-1}(\eta(k))(f(\eta(k))-\eta(k+1))$.

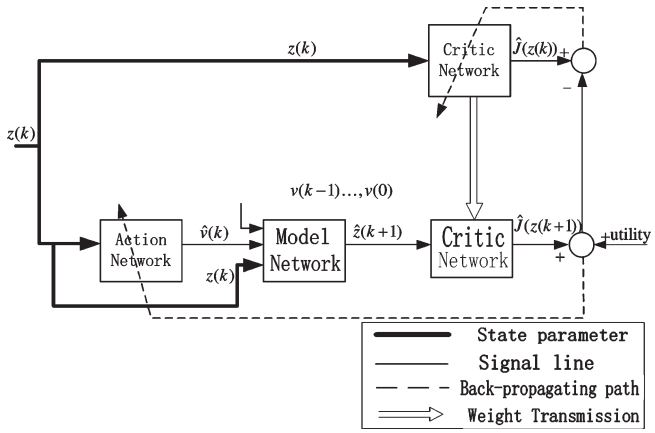

Fig. 1. Structure diagram of the greedy HDP iteration algorithm.

D. Neural Network Implementation for the Tracking Control Scheme

In the case of linear systems, the performance index is quadratic and the control policy is linear. In the nonlinear case, this is not necessarily true, and therefore, we use neural networks to approximate $v_i(k)$ and $J_i(z(k))$. There are three neural networks needed to implement the algorithm, which are the critic, model, and action networks. All the neural networks are chosen as three-layer feedforward networks. The structure diagram for running the greedy HDP iteration algorithm is shown in Fig. 1. The utility term in the figure denotes $z^{\mathrm{T}}(k)Qz(k)+\hat{v}^{\mathrm{T}}(k)R\hat{v}(k)+[\hat{v}(k)+v(k-1)+\cdots+v(0)]^{\mathrm{T}}S[\hat{v}(k)+v(k-1)+\cdots+v(0)]$. The gradient descent rule is adopted for the weight update rules of each neural network, and the details can be seen in [19], and the analysis of the neural network can be referred to [21], which is omitted here.

IV. Simulation Study

In this section, an example is provided to demonstrate the effectiveness of the tracking control scheme proposed in this paper. Our example is a modification of example 2 in [2] by extending the control to a vector.

Consider the following affine nonlinear system:

$$x(k+1)=f\left(x(k)\right)+g\left(x(k)\right)u(k) \tag{33}$$

where

$$\begin{aligned} x(k)&=\begin{bmatrix}x_1(k) & x_2(k)\end{bmatrix}^{\mathrm{T}}\\ u(k)&=\begin{bmatrix}u_1(k) & u_2(k)\end{bmatrix}^{\mathrm{T}}\\ f\left(x(k)\right)&=\begin{bmatrix}0.2x_1(k)\exp\left(x_2^2(k)\right)\\ 0.3x_2^3(k)\end{bmatrix}\\ g\left(x(k)\right)&=\begin{bmatrix}-0.2 & 0\\ 0 & -0.2\end{bmatrix}. \end{aligned}$$

The given initial state $x(0)=[1.5 \quad 1]^{\mathrm{T}}$ and the desired trajectory is set to $\eta(k)=[\sin(k/20)+(\pi/2) \quad 0.5\cos(k/20)]^{\mathrm{T}}$. In order to demonstrate the advantage of (7), a comparison on the tracking performance for two different performance indexes is presented. For the convenience of comparison, we define an evaluation function by $\mathrm{PER}=\sum_0^{T_f}\overline{z}^{\mathrm{T}}(k)\overline{z}(k)$, which means the sum of the square of tracking error during the running time, where T_f is the running-time steps.

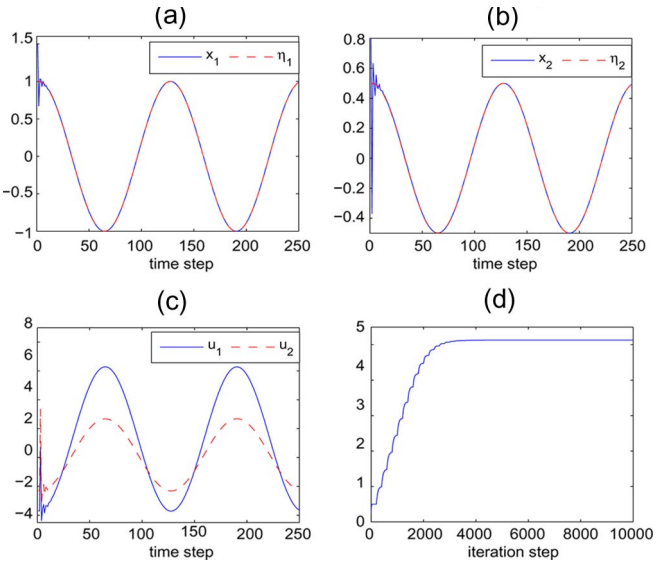

Fig. 2. Simulation results for Case 1. (a) State trajectory x_1 and desired trajectory η_1. (b) State trajectory x_2 and desired trajectory η_2. (c) Optimal tracking control curves. (d) Convergence of performance index.

Case 1: The performance index is defined by (4), where $\overline{Q}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ and $R=\begin{bmatrix}0.2 & 0\\ 0 & 0.2\end{bmatrix}$. The system is first transformed as (11). Moreover, we implement the algorithm at the time instant $k=0$. The critic, action, and model networks are chosen as three-layer neural networks with the structures 4-8-1, 4-8-2, and 6-8-4, respectively. The initial weights of the action, critic, and model networks are all set to be random in $[-1, 1]$. We take 1000 groups of sampling data to train the network. For each group of data with the learning rate $\alpha=0.01$, we train 4000 steps to reach the given accuracy $\varepsilon=10^{-6}$. After the training of the model network is completed, the weights keep unchanged. Then, the critic and action networks are trained for 10 000 iteration steps with the learning rate $\alpha=0.01$ so that the given accuracy $\varepsilon=10^{-6}$ is reached. Then, we apply the optimal tracking policy to the system for $T_f=250$ time steps and obtain the simulation results as in Fig. 2. In this case, with $T_f=250$, we can obtain the evaluation function value of the proposed tracking control scheme PER $=4.2958$.

Case 2: Define the performance index as (7) where $S=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$. All the other parameters are set the same as Case 1. Moreover, we also implement the algorithm at the time instant $k=0$. The simulation results are shown in Fig. 3. In this case, with $T_f=250$, we can obtain the evaluation function value of the proposed tracking control scheme PER $=2.1424$.

We can see that the tracking performance of Case 2 is much better than that of Case 1, although in both cases, the system states finally track the desired trajectories. In Case 1, both the states and control inputs oscillate seriously, whereas in Case 2, the oscillation is much slighter, and the evaluation function value is much smaller than that in Case 1. Hence, we can conclude that the control scheme proposed in this paper does quite satisfyingly solve the nonlinear tracking problem, and the optimal tracking controller obtained through the performance index defined in this paper has shown better performance.

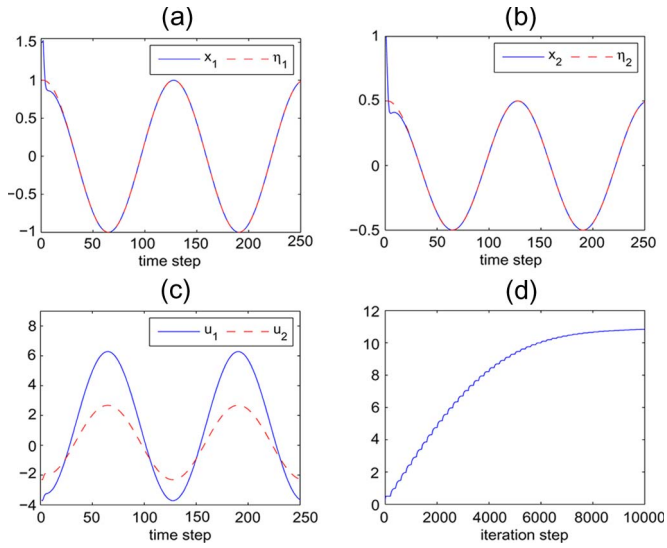

Fig. 3. Simulation results for Case 2. (a) State trajectory x_1 and desired trajectory η_1. (b) State trajectory x_2 and desired trajectory η_2. (c) Optimal tracking control curves. (d) Convergence of performance index.

V. Conclusion

In this paper, we propose an infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems. Via system transformation, the tracking problems are transformed as regulation problems. Then the greedy HDP iteration algorithm is introduced to deal with the regulation problems. The simulation studies have demonstrated the outstanding performance of the proposed tracking control scheme.

References

[1] Z. Aganovic and Z. Gajic, "The successive approximation procedure for finite-time optimal control of bilinear systems," *IEEE Trans. Autom. Control*, vol. 39, no. 9, pp. 1932–1935, Sep. 1994.

[2] A. Al-Tamimi and F. L. Lewis, "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof," in *Proc. IEEE Symp. Approximate Dynam. Program. Reinforcement Learn.*, Apr. 2007, pp. 38–43.

[3] R. Beard, "Improving the closed-loop performance of nonlinear systems," Ph.D. dissertation, Rensselaer Polytech. Inst., Troy, NY, 1995.

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

[5] T. Cimen and S. P. Banks, "Nonlinear optimal tracking control with application to super-tankers for autopilot design," *Automatica*, vol. 40, no. 11, pp. 1845–1863, Nov. 2004.

[6] S. Devasiad, D. Chen, and B. Paden, "Nonlinear inversion-based output tracking," *IEEE Trans. Autom. Control*, vol. 41, no. 7, pp. 930–942, Jul. 1996.

[7] S. Ferrari and R. F. Stengel, "Online adaptive critic flight control," *J. Guid. Control Dyn.*, vol. 27, no. 5, pp. 777–786, Sep./Oct. 2004.

[8] D. Gao, G. Tang, and B. Zhang, "Approximate optimal tracking control for a class of nonlinear systems with disturbances," in *Proc. 6th World Congr. Intell. Control Autom.*, Dalian, China, 2006, vol. 1, pp. 521–525.

[9] I. J. Ha and E. G. Gilbert, "Robust tracking in nonlinear systems," *IEEE Trans. Autom. Control*, vol. AC-32, no. 9, pp. 763–771, Sep. 1987.

[10] J. Huang, "An algorithm to solve the discrete HJI equation arising in the L_2-gain optimization problem," *Int. J. Control*, vol. 72, no. 1, pp. 49–57, Jan. 1999.

[11] N. Jin, D. Liu, T. Huang, and Z. Pang, "Discrete-time adaptive dynamic programming using wavelet basis function neural networks," in *Proc. IEEE Symp. Approximate Dynam. Program. Reinforcement Learn.*, Apr. 2007, pp. 135–142.

[12] D. Liu, X. Xiong, and Y. Zhang, "Action-dependent adaptive critic designs," in *Proc. INNS-IEEE Int. Joint Conf. Neural Netw.*, Washington DC, Jul. 2001, vol. 2, pp. 990–995.

[13] D. Liu and H. Zhang, "A neural dynamic programming approach for learning control of failure avoidance problems," *Int. J. Intell. Control Syst.*, vol. 10, no. 1, pp. 21–32, Mar. 2005.

[14] D. Liu, Y. Zhang, and H. Zhang, "A self-learning call admission control scheme for CDMA cellular networks," *IEEE Trans. Neural Netw.*, vol. 16, no. 5, pp. 1219–1228, Sep. 2005.

[15] T. W. Mclain, C. A. Bailry, and R. W. Beard, "Synthesis and experimental testing of a nonlinear optimal tracking controller," in *Proc. IEEE Amer. Control Conf.*, Jun. 1999, vol. 4, pp. 2847–2851.

[16] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 32, no. 2, pp. 140–153, May 2002.

[17] Y.-M. Park, M.-S. Choi, and K. Y. Lee, "An optimal tracking neuro-controller for nonlinear dynamic systems," *IEEE Trans. Neural Netw.*, vol. 7, no. 5, pp. 1009–1110, Sep. 1996.

[18] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1007, Sep. 1997.

[19] J. Si and Y.-T. Wang, "On-line learning control by association and reinforcement," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 264–276, Mar. 2001.

[20] M. L. Stefano, "Optimal design of PID regulators," *Int. J. Control*, vol. 33, no. 4, pp. 601–616, 1981.

[21] Q. Yang and S. Jagannathan, "Online reinforcement learning neural network controller design for nanomanipulation," in *Proc. IEEE Symp. Approximate Dynam. Program. Reinforcement Learn.*, Apr. 2007, pp. 225–232.

Huaguang Zhang (SM'04) received the B.S. and M.S. degrees in control engineering from Northeastern Electric Power University, Jilin, China, in 1982 and 1985, respectively, and the Ph.D. degree in thermal power engineering and automation in Southeastern University, Nanjing, China, in 1991.

He joined the Automatic Control Department, Northeastern University, Shenyang, China, in 1992, as a Postdoctoral Fellow. He is now a Professor with the School of Information Science and Engineering, and the Key Laboratory of Integrated Automation of Process Industry of National Education Ministry, Northeastern University. His main research interests are fuzzy control, chaos control, neural-network-based control, nonlinear control, signal processing, and their industrial application.

Dr. Zhang received the "Excellent Youth Science Foundation Award," nominated by the China Natural Science Foundation Committee, in 2003. He was named the Changjiang Scholar by the China Education Ministry in 2005. He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, PART B since 2007.

Qinglai Wei received the B.S. degree in automation control and the M.S. degree in control theory and control engineering in 2002 and 2006, respectively, from Northeastern University, Shenyang, China, where he is currently working toward the Ph.D. degree.

His research interests include neural-network-based control, nonlinear control, approximate dynamic programming, and their industrial application.

Yanhong Luo received the B.S. degree in automation control and the M.S. degree in control theory and control engineering in 2003 and 2006, respectively, from Northeastern University, Shenyang, China, where she is currently working toward the Ph.D. degree.

Her research interests include fuzzy control, neural networks adaptive control, approximate optimal control, and their industrial application.