Proceedings of the 10th
World Congress on Intelligent Control and Automation
July 6-8, 2012, Beijing, China

Finite-Horizon Neural Optimal Tracking Control for a Class of Nonlinear Systems with Unknown Dynamics*

Ding Wang, Derong Liu, and Hongliang Li
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing 100190, P. R. China
{ding.wang, derong.liu, hongliang.li}@ia.ac.cn

***Abstract*— A neural-network-based finite-horizon optimal tracking control scheme for a class of unknown nonlinear discrete-time systems is developed. First, the tracking control problem is converted into designing a regulator for the tracking error dynamics under the framework of finite-horizon optimal control theory. Then, with convergence analysis in terms of cost function and control law, the iterative adaptive dynamic programming algorithm is introduced to obtain the finite-horizon optimal controller to make the cost function close to its optimal value within an ε-error bound. Furthermore, in order to implement the algorithm via dual heuristic dynamic programming technique, three neural networks are employed to approximate the error dynamics, the cost function, and the control law, respectively. In addition, a numerical example is given to demonstrate the validity of the present approach.**

***Index Terms*— Adaptive dynamic programming, approximate dynamic programming, finite-horizon optimal tracking control, intelligent control, neural networks.**

I. Introduction

The tracking control problem has been studied by many researchers owing to its wide practical applications [1], [2]. Among that, the finite-horizon optimal tracking control problem is different from the infinite-horizon one. In the former issue, the controlled system must be tracked to a reference trajectory in a finite duration of time. Thereupon, the controller design method for the two cases is also dissimilar. Via appropriate system transformation, the tracking control problem can be converted into the regulator problem, which can be solved under the framework of optimal control theory. However, when dealing with the nonlinear optimal control problem, we often encounter the time-varying Hamilton-Jacobi-Bellman (HJB) equation which is difficult to tackle. Besides, the use of dynamic programming (DP) is usually confined to small dimension problem because of the "curse of dimensionality". Then, by combining DP with artificial neural networks (ANN or NN), the adaptive/approximate dynamic programming (ADP) approach was proposed by Werbos in [3] as a method for solving this optimal control problem forward-in-time.

The ADP approach has currently become a fundamental component of intelligent control [3] and computational intelligence [4]. Much progress has acquired in this field in terms of theory and application [3], [5]–[15]. In the light of [3] and [9], ADP approach was classified into the following schemes: heuristic dynamic programming (HDP), action-dependent HDP (ADHDP), also known as Q-learning, dual heuristic dynamic programming (DHP), ADDHP, globalized DHP (GDHP), and ADGDHP. Thereinto, some methods were employed for handling the optimal tracking control problem based on ADP [14], [15]. Besides, it should be mentioned that Park et al. [16] utilized the multilayer NN to get the optimal tracking neuro-controller for discrete-time nonlinear systems with quadratic cost function. However, [16] did not take the ADP method, while [14] and [15] were aimed at handling the infinite-horizon tracking control problem. In that way, there is still no result to solve the finite-horizon optimal tracking control problem for unknown nonlinear discrete-time systems based on iterative ADP algorithm via DHP technique. In this paper, we will handle this model-free optimal tracking control problem using the DHP technique. The theoretical principle is to introduce the finite-horizon optimal control scheme to deal with the regulation problem, which is converted from the original tracking control problem.

II. Preliminaries

Consider the nonlinear discrete-time systems given by

$$x_{k+1} = f(x_k) + g(x_k)u_p(x_k) \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state vector, $u_p(x_k) \in \mathbb{R}^m$ is the control vector, $f(\cdot)$ and $g(\cdot)$ are differentiable in their argument with $f(0) = 0$. Assume that $f + gu_p$ is Lipschitz continuous on a set Ω in $\mathbb{R}^n$ containing the origin, and that the system (1) is controllable in the sense that there exists a continuous control on Ω that asymptotically stabilizes the system. Note that in the following part, $u_p(x_k)$ is denoted by u_{pk} for simplicity.

For handling the optimal tracking control problem, we should determine the optimal control law u_p^*, which can make the nonlinear system (1) to track a reference trajectory

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

978-1-4673-1398-8/12/$31.00 ©2012 IEEE

$r \in \mathbb{R}^n$ in an optimal manner. The tracking error is defined as

$$e_k = x_k - r. \tag{2}$$

Inspired by the work of [14]–[16], we define the steady control corresponding to the reference trajectory r as

$$u_{dk} = g^{-1}(r)(r - f(r)) \tag{3}$$

where $g^{-1}(r)g(r) = I_m$ and I_m is an $m \times m$ identity matrix. Let

$$u_k = u_{pk} - u_{dk}. \tag{4}$$

Then, considering (1)–(4), we can obtain the error dynamics as follows:

$$\begin{aligned} e_{k+1} = {} & f(e_k + r) + g(e_k + r)g^{-1}(r)(r - f(r)) \\ & - r + g(e_k + r)u_k \end{aligned} \tag{5}$$

where e_k and u_k are regarded as the state and input vector of the error dynamic system, respectively. For simplicity, (5) can be rewritten as

$$e_{k+1} = F(e_k, u_k). \tag{6}$$

Let e_0 be an initial state vector of system (6) and define $\underline{u}_0^{N-1} = (u_0, u_1, \ldots, u_{N-1})$ be a control sequence with which the system (6) gives a trajectory starting from e_0: $e_1 = F(e_0, u_0)$, $e_2 = F(e_1, u_1)$, ..., $e_N = F(e_{N-1}, u_{N-1})$. We call the number of elements in the control sequence $\underline{u}_0^{N-1}$ the length of $\underline{u}_0^{N-1}$ and denote it as $|\underline{u}_0^{N-1}|$. Then, $|\underline{u}_0^{N-1}| = N$. The final state under the control sequence $\underline{u}_0^{N-1}$ is denoted as $e^{(f)}(e_0, \underline{u}_0^{N-1}) = e_N$. Now, let $\underline{u}_k^{N-1} = (u_k, u_{k+1}, \ldots, u_{N-1})$ be a control sequence starting at k with length $N - k$, i.e., $|\underline{u}_k^{N-1}| = N - k$. For solving the finite-horizon optimal tracking control problem, it is desired to find the optimal control sequence which minimizes the following cost function

$$J(e_k, \underline{u}_k^{N-1}) = \sum_{i=k}^{N-1} U(e_i, u_i) \tag{7}$$

where U is the utility function, $U(0,0) = 0$, $U(e_i, u_i) \geq 0$ for $\forall\, e_i, u_i$. Here, the utility function is chosen as the quadratic form as $U(e_i, u_i) = e_i^T Q e_i + u_i^T R u_i$, which can not only force the system state to follow the reference trajectory, but also force the system input to be close to the steady value in maintaining the state to its reference value.

Consequently, the problem of solving the finite-horizon optimal tracking control law u_p^* for system (1) is transformed into seeking the finite-horizon optimal control law u^* for system (6) with respect to (7). Next, we will focus on designing u^* under the framework of finite-horizon optimal control theory. Incidentally, the devised feedback control must be finite-horizon admissible, which is defined as follows.

Definition 1: A control sequence $\underline{u}_k^{N-1}$ is said to be finite-horizon admissible for a state $e_k \in \mathbb{R}^n$ with respect to (7) on Ω if u is continuous on a compact set $\Omega_u \in \mathbb{R}^m$, $u(0) = 0$, $e^{(f)}(e_k, \underline{u}_k^{N-1}) = 0$ and $J(e_k, \underline{u}_k^{N-1})$ is finite.

Let

$$\mathfrak{A}_{e_k} = \{\underline{u}_k : e^{(f)}(e_k, \underline{u}_k) = 0\}$$

be the set of all finite-horizon admissible control sequences of e_k and

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

be the set of all finite-horizon admissible control sequences of e_k with length i. The optimal cost function is denoted as $J^*(e_k)$ and is defined as

$$J^*(e_k) = \inf_{\underline{u}_k} \{J(e_k, \underline{u}_k) : \underline{u}_k \in \mathfrak{A}_{e_k}\}. \tag{8}$$

According to Bellman's optimality principle, $J^*(e_k)$ satisfies the discrete-time HJB (DTHJB) equation

$$J^*(e_k) = \min_{u_k} \{U(e_k, u_k) + J^*(e_{k+1})\}. \tag{9}$$

Meanwhile, the optimal control u^* satisfies the first-order necessary condition, which is formulated as

$$u^*(e_k) = \arg\min_{u_k} \{U(e_k, u_k) + J^*(e_{k+1})\}. \tag{10}$$

Because it is difficult to solve the DTHJB equation (9) directly, we will propose an iterative algorithm to get its solution approximately. Before that, we assume the inverse of the control coefficient matrix $g(e_k + r)$ exists. This make sure that for given e_k, there exists an initial control u_k to transfer e_k to zero in one time step.

III. Finite-Horizon Optimal Tracking Control Scheme Based on Iterative ADP Algorithm

Now, we deal with the finite-horizon optimal tracking control problem for system (1) by using the iterative ADP algorithm. It is equivalent to handle the finite-horizon optimal control problem for system (6).

A. Derivation of the Iterative Algorithm

In this part, we present the iterative ADP algorithm. First, we start with the initial cost function $V_0(\cdot) = 0$ and solve $v_0(e_k)$ as

$$\begin{aligned} v_0(e_k) = {} & \arg\min_{u_k} \{U(e_k, u_k) + V_0(e_{k+1})\} \\ & \text{subject to } F(e_k, u_k) = 0. \end{aligned} \tag{11}$$

Then, we update the cost function as

$$\begin{aligned} V_1(e_k) &= \min_{u_k} \{U(e_k, u_k) + V_0(e_{k+1})\} \\ &= U(e_k, v_0(e_k)), \end{aligned}$$

which can also be written as the following form:

$$\begin{aligned} V_1(e_k) &= \min_{u_k} U(e_k, u_k) \text{ subject to } F(e_k, u_k) = 0 \\ &= U(e_k, v_0(e_k)). \end{aligned} \tag{12}$$

Next, for $i = 1, 2, \ldots$, the algorithm can be carried out between

$$\begin{aligned} v_i(e_k) &= \arg\min_{u_k} \left\{U(e_k, u_k) + V_i(e_{k+1})\right\} \\ &= \arg\min_{u_k} \left\{U(e_k, u_k) + V_i(F(e_k, u_k))\right\} \qquad (13) \end{aligned}$$

and

$$\begin{aligned} V_{i+1}(e_k) &= \min_{u_k} \left\{U(e_k, u_k) + V_i(e_{k+1})\right\} \\ &= U(e_k, v_i(e_k)) + V_i(F(e_k, v_i(e_k))). \qquad (14) \end{aligned}$$

In the following part, we will present the convergence analysis of the iteration between (13) and (14) with the cost function $V_i \to J^*$ and the control law $v_i \to u^*$ as $i \to \infty$. Here, we expand $V_{i+1}(e_k)$ to see what it will be. Considering (12) and (14), we can derive the following expression:

$$\begin{aligned} V_{i+1}(e_k) &= \min_{\underline{u}_k^{k+i}} \sum_{j=0}^{i} U(e_{k+j}, u_{k+j}) \\ &\quad \text{subject to } F(e_{k+i}, u_{k+i}) = 0 \\ &= \min_{\underline{u}_k^{k+i}} \left\{ J\big(e_k, \underline{u}_k^{k+i}\big) : \underline{u}_k^{k+i} \in \mathfrak{A}_{e_k}^{(i+1)} \right\}. \qquad (15) \end{aligned}$$

Using the relationship between (13) and (14), it is important to note that $(v_i(e_k), v_{i-1}(e_{k+1}), \ldots, v_0(e_{k+i}))$ is the finite-horizon admissible control sequence corresponding to $V_{i+1}(e_k)$ with length $i+1$. Thereupon, (15) can be rewritten as

$$V_{i+1}(e_k) = \sum_{j=0}^{i} U(e_{k+j}, v_{i-j}(e_{k+j})). \qquad (16)$$

B. Convergence Analysis of the Iterative Algorithm

Theorem 1: Suppose the set of the finite-horizon admissible control sequences of e_k with length 1 is not null, i.e., $\mathfrak{A}_{e_k}^{(1)} \neq \emptyset$. Define the cost function sequence $\{V_i\}$ as in (14) with $V_0(\cdot) = 0$. Then, we can conclude that $\{V_i\}$ is a monotonically nonincreasing sequence satisfying $V_{i+1}(e_k) \leq V_i(e_k)$ for $\forall i \geq 1$, i.e., $V_1(e_k) = \max\{V_i(e_k) : i = 1, 2, \ldots\}$.

Proof: The theorem can be proved by using mathematical induction.

First, we let $i = 1$. The cost function $V_1(e_k)$ is given in (12) and the finite-horizon admissible control sequence with length 1 is $\hat{\underline{u}}_k^k = (v_0(e_k))$. Now, we show that there exists a finite-horizon admissible control sequence $\hat{\underline{u}}_k^{k+1}$ with length 2 such that $J\big(e_k, \hat{\underline{u}}_k^{k+1}\big) = V_1(e_k)$. Let $\hat{\underline{u}}_k^{k+1} = \big(\hat{\underline{u}}_k^k, 0\big)$, then $\big|\hat{\underline{u}}_k^{k+1}\big| = 2$. Since $e_{k+1} = F(e_k, v_0(e_k)) = 0$ and $\hat{u}_{k+1} = 0$, we have $e_{k+2} = F(e_{k+1}, \hat{u}_{k+1}) = F(0, 0) = 0$. Thus, $\hat{\underline{u}}_k^{k+1}$ is the finite-horizon admissible control sequence with length 2. Since $U(e_{k+1}, \hat{u}_{k+1}) = U(0, 0) = 0$, we obtain

$$\begin{aligned} J\big(e_k, \hat{\underline{u}}_k^{k+1}\big) &= U(e_k, v_0(e_k)) + U(e_{k+1}, \hat{u}_{k+1}) \\ &= U(e_k, v_0(e_k)) \\ &= V_1(e_k). \end{aligned}$$

Note that according to (15), we have

$$V_2(e_k) = \min_{\underline{u}_k^{k+1}} \left\{ J\big(e_k, \underline{u}_k^{k+1}\big) : \underline{u}_k^{k+1} \in \mathfrak{A}^{(2)} \right\}.$$

Then, we can derive that

$$V_2(e_k) \leq J\big(e_k, \hat{\underline{u}}_k^{k+1}\big) = V_1(e_k). \qquad (17)$$

Therefore, the theorem holds for $i = 1$.

Next, assume that the theorem holds for any $i = q - 1$, where q is an integer and $q > 2$. Here, the cost function $V_q(e_k)$ can be formulated as

$$V_q(e_k) = \sum_{j=0}^{q-1} U(e_{k+j}, v_{q-1-j}(e_{k+j})). \qquad (18)$$

Note $\hat{\underline{u}}_k^{k+q-1} = (v_{q-1}(e_k), v_{q-2}(e_{k+1}), \ldots, v_0(e_{k+q-1}))$ is the finite-horizon admissible control sequence corresponding to $V_q(e_k)$ with length q.

Then, for $i = q$, we can construct a control sequence $\hat{\underline{u}}_k^{k+q} = (v_{q-1}(e_k), v_{q-2}(e_{k+1}), \ldots, v_0(e_{k+q-1}), 0)$ with length $q + 1$, under which the error trajectory is given as e_k, $e_{k+1} = F(e_k, v_{q-1}(e_k))$, $e_{k+2} = F(e_{k+1}, v_{q-2}(e_{k+1}))$, $\ldots$, $e_{k+q} = F(e_{k+q-1}, v_0(e_{k+q-1})) = 0$, $e_{k+q+1} = F(e_{k+q}, \hat{u}_{k+q}) = F(0, 0) = 0$. Hence, $\hat{\underline{u}}_k^{k+q}$ is a finite-horizon admissible control sequence with length $q + 1$. Considering $U(e_{k+q}, \hat{u}_{k+q}) = U(0, 0) = 0$, we obtain

$$\begin{aligned} J\big(e_k, \hat{\underline{u}}_k^{k+q}\big) &= U(e_k, v_{q-1}(e_k)) + U(e_{k+1}, v_{q-2}(e_{k+1})) \\ &\quad + \cdots + U(e_{k+q-1}, v_0(e_{k+q-1})) \\ &\quad + U(e_{k+q}, \hat{u}_{k+q}) \\ &= \sum_{j=0}^{q-1} U(e_{k+j}, v_{q-1-j}(e_{k+j})) \\ &= V_q(e_k). \end{aligned}$$

According to (15), we have

$$V_{q+1}(e_k) = \min_{\underline{u}_k^{k+q}} \left\{ J\big(e_k, \underline{u}_k^{k+q}\big) : \underline{u}_k^{k+q} \in \mathfrak{A}_{e_k}^{(q+1)} \right\}.$$

Then, we can derive that

$$V_{q+1}(e_k) \leq J\big(e_k, \hat{\underline{u}}_k^{k+q}\big) = V_q(e_k). \qquad (19)$$

This completes the proof. ■

According to Theorem 1, we derive that the cost function sequence $\{V_i(e_k)\}$ is monotonically nonincreasing. Besides, the quadratic form of the utility function render $V_i(e_k) \geq 0$ for $\forall i \geq 0$, which reveals that the sequence $\{V_i(e_k)\}$ is bounded below. Therefore, the limit of the cost function sequence exists. Here, we denote it as $V_\infty(e_k)$, i.e., $\lim_{i \to \infty} V_i(e_k) = V_\infty(e_k)$.

Theorem 2: Define the cost function sequence $\{V_i\}$ as in (14) with $V_0(\cdot) = 0$. If the state e_k of the error dynamic system is controllable, then J^* is the limit of the cost function sequence $\{V_i\}$, i.e.,

$$V_\infty(e_k) = J^*(e_k). \tag{20}$$

Proof: On the one hand, considering (8) and (15), we can obtain

$$\begin{aligned} J^*(e_k) &= \inf_{\underline{u}_k} \left\{ J(e_k, \underline{u}_k) \colon \underline{u}_k \in \mathfrak{A}_{e_k} \right\} \\ &\le \min_{\underline{u}_k^{k+i-1}} \left\{ J\big(e_k, \underline{u}_k^{k+i-1}\big) : \underline{u}_k^{k+i-1} \in \mathfrak{A}_{e_k}^{(i)} \right\} \\ &= V_i(e_k). \end{aligned}$$

Let $i \to \infty$. Then, we get

$$J^*(e_k) \le V_\infty(e_k). \tag{21}$$

On the other hand, according to the definition of $J^*(e_k)$, for any $\eta > 0$, there exists an admissible control sequence $\underline{\sigma}_k \in \mathfrak{A}_{e_k}$ such that

$$J\big(e_k, \underline{\sigma}_k\big) \le J^*(e_k) + \eta. \tag{22}$$

We suppose $|\underline{\sigma}_k| = q$, which means that $\underline{\sigma}_k \in \mathfrak{A}_{e_k}^{(q)}$. Then, we can acquire that

$$\begin{aligned} V_\infty(e_k) &\le V_q(e_k) \\ &= \min_{\underline{u}_k^{k+q-1}} \left\{ J\big(e_k, \underline{u}_k^{k+q-1}\big) : \underline{u}_k^{k+q-1} \in \mathfrak{A}_{e_k}^{(q)} \right\} \\ &\le J\big(e_k, \underline{\sigma}_k\big). \end{aligned} \tag{23}$$

Combining (23) with (22), we have

$$V_\infty(e_k) \le J^*(e_k) + \eta. \tag{24}$$

Noting that η is chosen arbitrarily in (24), we can obtain that

$$V_\infty(e_k) \le J^*(e_k). \tag{25}$$

Based on (21) and (25), we conclude that $J^*(e_k)$ is the limit of the cost function sequence $\{V_i\}$ as $i \to \infty$, i.e., $V_\infty(e_k) = J^*(e_k)$. ■

According to Theorems 1–2, we have proved that the cost function sequence $\{V_i(e_k)\}$ of the iterative ADP algorithm converges to the optimal cost function $J^*(e_k)$ of the DTHJB equation, i.e., $V_i \to J^*$ as $i \to \infty$. Then, considering (10) and (13), we can conclude the convergence of the corresponding control law sequence, i.e., $\lim_{i\to\infty} v_i(e_k) = u^*(e_k)$.

C. The ε-Optimal Control Algorithm

The aforementioned conclusions imply that we should run the iterative ADP algorithm (11)–(14) until $i \to \infty$ to obtain the optimal cost function $J^*(e_k)$. Then, we can derive a control vector $v_\infty(e_k)$, i.e., the optimal control vector $u^*(e_k)$, based on which we can construct a control sequence $\underline{u}_\infty(e_k) = (v_\infty(e_k), v_\infty(e_{k+1}), \ldots, v_\infty(e_{k+i}), \ldots)$ to make $e_{k+i} \to 0$ as $i \to \infty$. Obviously, $\underline{u}_\infty(e_k)$ has infinite length. However, it is always not practical to acquire $\underline{u}_\infty(e_k)$ because most real world systems need to be effectively controlled within finite time steps. Therefore, in this part, we will propose an ε-optimal control strategy using the iterative ADP algorithm, in order to transfer the error dynamics to zero within finite steps.

Let $\varepsilon > 0$ be any small number, e_k be any controllable state of the error dynamic system, and $J^*(e_k)$ be the optimal value of the cost function sequence $\{V_i(e_k)\}$. According to Theorem 2, it is clear that there exists a finite integer i such that

$$|V_i(e_k) - J^*(e_k)| \le \varepsilon. \tag{26}$$

The length of the optimal control sequence starting from e_k with respect to ε is defined as

$$K_\varepsilon(e_k) = \min\{i \colon |V_i(e_k) - J^*(e_k)| \le \varepsilon\}. \tag{27}$$

According to (13) and (14), the control law corresponding to $V_i(e_k)$ is

$$v_{i-1}(e_k) = \arg\min_{u_k} \left\{ U(e_k, u_k) + V_{i-1}(e_{k+1}) \right\}. \tag{28}$$

It is called the ε-optimal control and is denoted as $\mu_\varepsilon^*(e_k)$.

In this way, we can see that an error ε between $V_i(e_k)$ and $J^*(e_k)$ is introduced into the iterative ADP algorithm. This makes sure that the cost function sequence $\{V_i(e_k)\}$ can converge to its optimal value during finite iteration steps.

However, it is difficult to employ the criterion (26) in practice because the optimal cost function $J^*(e_k)$ is unknown in advance. As a result, we introduce the following criterion to replace (26):

$$|V_i(e_k) - V_{i+1}(e_k)| \le \varepsilon. \tag{29}$$

IV. Implementation of the Iterative Algorithm Using NN-Based DHP Technique

In this section, we implement the iterative ADP algorithm via DHP technique, which is called iterative DHP algorithm for short. In the iterative DHP algorithm, there are three networks, which are model network, critic network and action network. Note that all the networks are chosen as three-layer feedforward NNs. The structure diagram of the iterative DHP algorithm is shown in Fig. 1, where $W = (\partial \hat{e}_{k+1}/\partial e_k)^T$.

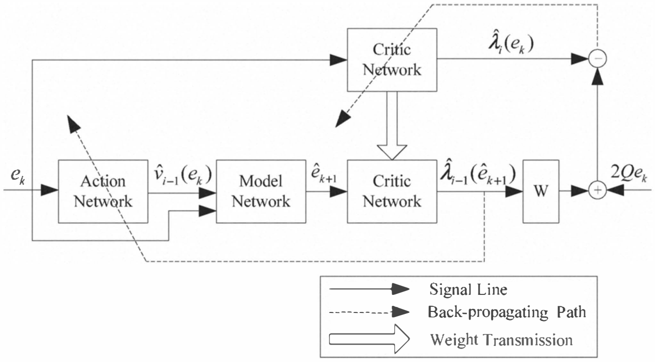

Fig. 1. The structure diagram of the iterative DHP algorithm

A. The Model Network

We design the model network for identifying the error dynamics. After the model network is trained sufficiently, we have

$$F(e_k, u_k) = \omega_m^T \sigma\big(\nu_m^T [e_k^T \;\; u_k^T]^T\big). \tag{30}$$

Taking the partial derivative of both sides of (30) with respect to u_k yields

$$g(e_k + r) = \frac{\partial\Big(\omega_m^T \sigma\big(\nu_m^T [e_k^T \;\; u_k^T]^T\big)\Big)}{\partial u_k}. \tag{31}$$

Hence, we can avoid the requirement of knowing the system dynamics when implementing the algorithm. For one thing, we can use the trained model network to compute the error of the next time step. For another, we can derive the expression of the iterative control law with the help of (31).

B. The Critic Network

The critic network is used to approximate the derivative of the cost function $V_i(e_k)$, which is named as costate function and formulated as $\lambda_i(e_k) = \partial V_i(e_k)/\partial e_k$. Here, we show what $\lambda_i(e_k)$ will be when it is expanded. According to (28), $v_{i-1}(e_k)$ is just the solution of the following equation with respect to u_k:

$$\frac{\partial U(e_k, u_k)}{\partial u_k} + \left(\frac{\partial e_{k+1}}{\partial u_k}\right)^T \frac{\partial V_{i-1}(e_{k+1})}{\partial e_{k+1}} = 0. \tag{32}$$

Then, we have

$$\frac{\partial U(e_k, v_{i-1}(e_k))}{\partial v_{i-1}(e_k)} + \left(\frac{\partial e_{k+1}}{\partial v_{i-1}(e_k)}\right)^T \frac{\partial V_{i-1}(e_{k+1})}{\partial e_{k+1}} = 0. \tag{33}$$

Therefore, considering (33), we can obtain

$$\begin{aligned}\lambda_i(e_k) &= \frac{\partial U(e_k, v_{i-1}(e_k))}{\partial e_k} + \frac{\partial V_{i-1}(e_{k+1})}{\partial e_k} \\ &= 2Qe_k + \left(\frac{\partial e_{k+1}}{\partial e_k}\right)^T \lambda_{i-1}(e_{k+1}). \end{aligned} \tag{34}$$

We denote the output of the critic network as

$$\hat{\lambda}_i(e_k) = \omega_{ci}^T \sigma\big(\nu_{ci}^T e_k\big). \tag{35}$$

In the iteration process, the target function of the critic network is

$$\lambda_i(e_k) = 2Qe_k + \left(\frac{\partial \hat{e}_{k+1}}{\partial e_k}\right)^T \hat{\lambda}_{i-1}(\hat{e}_{k+1}). \tag{36}$$

Then, the error function for training the critic network can be defined as

$$e_{cik} = \hat{\lambda}_i(e_k) - \lambda_i(e_k). \tag{37}$$

Besides, the objective function to be minimized of the critic network is

$$E_{cik} = \frac{1}{2} e_{cik}^T e_{cik}. \tag{38}$$

The weight updating rule for training the critic network is also gradient-based adaptation, so

$$\omega_{ci}(j+1) = \omega_{ci}(j) - \alpha_c \left[\frac{\partial E_{cik}}{\partial \omega_{ci}(j)}\right] \tag{39}$$

$$\nu_{ci}(j+1) = \nu_{ci}(j) - \alpha_c \left[\frac{\partial E_{cik}}{\partial \nu_{ci}(j)}\right] \tag{40}$$

where $\alpha_c > 0$ is the learning rate of the critic network, and j is the inner-loop iteration step for updating the weight parameters.

The training process of action network is omitted here, which can be referred to [11].

Remark 1: According to Theorems 1–2, $V_i \to J^*$ as $i \to \infty$. Since $\lambda_i(e_k) = \partial V_i(e_k)/\partial e_k$, we can conclude that the costate function sequence $\{\lambda_i\}$ is also convergent with $\lambda_i \to \lambda^*$ as $i \to \infty$.

After training the three NNs, we obtain the optimal control input u_k^* for system (5) under the given error bound ε. As a result, we can compute the optimal tracking control input u_{pk}^* for original system (1) by

$$\begin{aligned} u_{pk}^* &= u^*(e_k) + u_{dk} \\ &= u^*(e_k) + g^{-1}(r)(r - f(r)). \end{aligned} \tag{41}$$

The control law u_p^* can make system (1) to track the selected reference trajectory in an optimal manner.

V. Simulation Study

In this section, we illustrate the theoretical results of the iterative DHP algorithm for solving the optimal tracking control problem. The example is derived from [14] with some modifications. Consider the nonlinear system described by

$$x_{k+1} = \begin{bmatrix} 0.2x_{1k}e^{x_{2k}^2} \\ 0.3x_{2k}^3 \end{bmatrix} + \begin{bmatrix} -0.5 & 0 \\ 0 & -1 \end{bmatrix} u_{pk} \tag{42}$$

where $x_k = [x_{1k} \; x_{2k}]^T \in \mathbb{R}^2$ and $u_{pk} = [u_{p1k} \; u_{p2k}]^T \in \mathbb{R}^2$ are the state and control variables, respectively. The parameters of the cost function are chosen as $Q = I$ and $R = I$, where I denotes the identity matrix with suitable dimensions. The state of the controlled system (42) is initialized to be $x_0 = [-0.5 \;\; 1]^T$, while the reference trajectory is selected as $r = [0.5 \;\; -1]^T$.

We set the error bound of the finite-horizon optimal control problem as $\varepsilon = 10^{-4}$ and implement the iterative DHP algorithm at time instant $k = 0$. According to (11), we derive the initial control input of system (5) is $v_0(e_0) = [-0.4e \;\; 0.6]^T$, where $e_0 = [-1 \;\; 2]^T$. Then, we choose three-layer feedforward NNs as model network, critic network and action network with the structures 4–8–2, 2–8–2, and 2–8–2, respectively. The initial weights of three networks are all set to be random in $[-0.1, 0.1]$. First, we train the model network for 500 time steps using 100 data samples under the learning rate $\alpha_m = 0.1$. After the model network is trained sufficiently, its weights are kept unchanged. Then, we train the critic network and action network for 25 iterations (i.e., for $i = 1, 2, \dots, 25$) with each iteration of 1000 training epochs to make sure the given error bound $\varepsilon = 10^{-4}$ is reached. In the training process, the learning rate $\alpha_c = \alpha_a = 0.05$. The convergence process of the costate function sequence of the iterative DHP algorithm is shown in Fig. 2, for $k = 0$. We can see that the iterative costate function sequence converges to the optimal one ultimately, which supports the statement of Remark 1. Note that we have $|V_{24}(e_0) - V_{25}(e_0)| \leq \varepsilon$, which means the length of the optimal control sequence starting from e_0 with respect to ε is $K_\varepsilon(e_0) = 24$. Besides, the ε-optimal control law $\mu_\varepsilon^*(e_0)$ for system (5) can also be obtained after the iteration process.

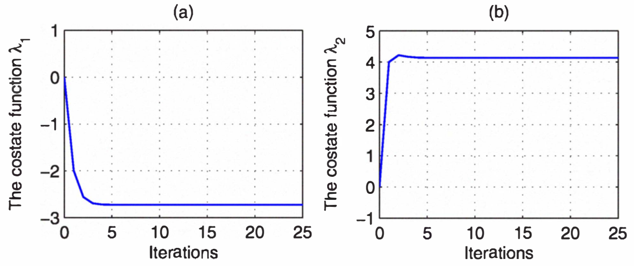

Fig. 2. (a) The convergence process of $\{\lambda_{1i}\}$. (b) The convergence process of $\{\lambda_{2i}\}$.

Now, we apply the derived control law to the error dynamic system (5) for 25 time steps, and obtain the control and error trajectories are shown in Fig. 3(a) and 3(b), respectively. Actually, we can compute the tracking error becomes $e_{24} = [0.2217 \times 10^{-4} \ \ -0.1490 \times 10^{-4}]^T$ after 24 time steps. These results substantiate the excellent performance of the tracking controller derived by the iterative DHP algorithm.

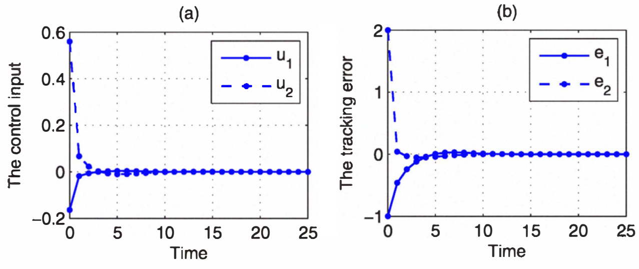

Fig. 3. (a) The control input u of the error dynamic system. (b) The tracking error e.

VI. Conclusion

In this paper, a model-free iterative algorithm is employed to design the finite-horizon near-optimal tracking controller for a class of unknown nonlinear discrete-time systems. Through system transformation, the tracking problem is converted into seeking the finite-horizon optimal control law for the error dynamic system. Then, the iterative ADP algorithm is introduced to deal with the DTHJB equation by using DHP technique. Additionally, the simulation example certified the validity of the tracking control scheme.

References

[1] Z. Liu, H. Zhang, and D. Liu, "Adaptive tracking control of a class of nonlinear time-delay systems with NN actuator saturation compensation," in Proceedings of 8th World Congress on Intelligent Control and Automation, Jinan, P. R. China, pp. 5115–5120, 2010.

[2] F. L. Lewis and V. L. Syrmos, Optimal Control, New York: Wiley, 1995.

[3] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling", in Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches, D. A. White and D. A. Sofge Eds. New York: Van Nostrand Reinhold, 1992, chapter 13.

[4] A. E. Ruano, Intelligent Control Systems Using Computational Intelligence Techniques, London: The Institute of Engineering and Technology, 2008.

[5] F. Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: an introduction," IEEE Computational Intelligence Magazine, vol. 4, no. 2, pp. 39–47, May 2009.

[6] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," IEEE Circuits and Systems Magazine, vol. 9, no. 3, pp. 32–50, July 2009.

[7] 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, vol. 38, no. 4, pp. 943–949, August 2008.

[8] J. Si and Y. T. Wang, "On-line learning control by association and reinforcement," IEEE Transactions on Neural Networks, vol. 12, no. 2, pp. 264–276, March 2001.

[9] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," IEEE Transactions on Neural Networks, vol. 8, no. 5, pp. 997–1007, September 1997.

[10] D. Liu, "Approximate dynamic programming for self-learning control," Acta Automatica Sinica, vol. 31, no. 1, pp. 13–18, January 2005.

[11] D. Wang and D. Liu, "Optimal control for a class of unknown nonlinear systems via the iterative GDHP algorithm," in Proceedings of 8th International Symposium on Neural Networks, Guilin, P. R. China, pp. 630–639, 2011.

[12] F. Y. Wang, N. Jin, D. Liu, and Q. Wei, "Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ε-error bound," IEEE Transactions on Neural Networks, vol. 22, no. 1, pp. 24–36, January 2011.

[13] H. Zhang, Y. Luo, and D. Liu, "Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints," IEEE Transactions on Neural Networks, vol. 20, no. 9, pp. 1490–1503, September 2009.

[14] 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, vol. 38, no. 4, pp. 937–942, August 2008.

[15] T. Dierks and S. Jagannathan, "Optimal tracking control of affine nonliner discrete-time systems with unknown internal dynamics," in Proceedings of Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, P. R. China, pp. 6750–6755, 2009.

[16] Y. M. Park, M. S. Choi, and K. Y. Lee, "An optimal tracking neurocontroller for nonlinear dynamic systems," IEEE Transactions on Neural Networks, vol. 7, no. 5, pp. 1099–1110, September 1996.