IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received February 2, 2020, accepted February 11, 2020, date of publication February 20, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975391

Data-Driven Nonlinear Near-Optimal Regulation Based on Multi-Dimensional Taylor Network Dynamic Programming

QI-MING SUN[1], **CHAO ZHANG**[2], **NAN-YUN JIANG**[3], **JING-JING YU**[4], **AND LEI XU**[1]

[1]College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China
[2]School of Electrical Engineering and Automation, Henan Institute of Technology, Xinxiang 453003, China
[3]Department of Economics and Management, Nanjing Tech University, Nanjing 210096, China
[4]China Railway Shanghai Design Institute Group Co., Ltd., Shanghai 200070, China

Corresponding author: Qi-Ming Sun (sqm1122345@126.com)

This work was supported in part by the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China, under Grant 19KJB510034, in part by the Shuishan Scientific Research Foundation of Nanjing Forestry University under Grant 163070112, in part by the High-Level Talent Research Foundation of Henan Institute of Technology under Grant KQ1863, and in part by the Special Research and Promotion Program of Henan Province under Grant 202102110274.

ABSTRACT Using the data-driven control formulation, an iterative dynamic programming approach which is based on a multi-dimensional Taylor network is established to design the near optimal regulation of discrete-time nonlinear systems. For discrete-time general nonlinear systems, the iterative adaptive dynamic programming algorithm is developed and proved to guarantee the property of convergence and optimality. Three networks are constructed, namely, the identification network, critic network and action network. Moreover, a globalized dual heuristic programming technique with detailed implementation is developed. The cost function and its derivative can be approximated by this novel architecture. Besides, without the consideration of the system dynamics, this technique can learn the near-optimal control law simultaneously and adaptively. In addition, this technique greatly improves the existing results of the iterative adaptive dynamic programming algorithm in terms of reducing the requirement of the control matrix. Furthermore, because of the approach that is based on the multi-dimensional Taylor network, the amount of calculation needed is also greatly reduced. The simulation experiment is described to illustrate the effectiveness of the data-driven optimal regulation method proposed in this paper.

INDEX TERMS Adaptive dynamic programming, data-driven control, multi-dimensional Taylor network, nonlinear control.

I. INTRODUCTION

A wide range of applications involve optimal control in engineering technology. To optimize the performance index of the controlled system, the controller design is the basis of the optimal control research [1]. Therefore, optimal control has become one of the main topics of modern control theory [2]. Unlike the optimal control problem of linear systems, the optimal control problems of nonlinear systems usually require solving nonlinear Hamilton-Jacobi-Bellman (HJB) equations [3]. However, it is very difficult to solve nonlinear partial differential equations, even though some equations cannot be solved under certain conditions. Therefore, the emergence of dynamic programming provides a new method for optimal control [4], [5]. A novel iterative two stage dual heuristic programming is proposed to solve the optimal control problems for a class of discrete time that is switched nonlinear systems subject to actuators saturation [6]. In the past few years, adaptive-based methods have been well developed [7]–[9], such as heuristic dynamic programming (HDP) [10], dual heuristic dynamic programming (DHP) [11], and globalized dual heuristic dynamic programming (GDHP) [12]. In literature [13], it is discussed that a robust adaptive control scheme based on a cascaded structure with a full state feedback controller with integrator terms as inner control loop and computed torque as an outer control loop for flexible joint robots. Based on the adaptive laws and finite-time stability theory, a nonsingular terminal sliding

The associate editor coordinating the review of this manuscript and approving it for publication was Sun Junwei.

mode control is designed in literature [14]. Literature [15] proposes a new output-constrained robust adaptive controller for a class of uncertain multi-input multi-output (MIMO) nonlinear systems. These methods which are simple to implement, do not require the controlled object model. Because of these advantages, the adaptive dynamic programming method is well developed [16]. It is concerned with a novel generalized policy iteration algorithm for solving optimal control problems for discrete-time nonlinear systems [17]. In literature [18], a simple, effective method is given for designing the autonomous memristor chaotic systems.

In the current technological development of large data, with the research on data-driven thinking and the study of learning algorithms, the adaptive dynamic programming (ADP) algorithm has become an effective means of optimizing design and intelligent control [19]–[21]. For the introduction of this algorithm, a large amount of researches on the ADP algorithm have emerged [22]–[25]. A novel data-driven robust approximate optimal tracking control scheme is proposed for unknown general nonlinear systems by using the ADP method [26]. A new iterative ADP method is proposed to solve a class of continuous-time nonlinear two-player zero-sum differential games [27]. In literature [28], it is showed how to implement ADP methods using only measured input output data from the system. An online adaptive policy learning algorithm (APLA) based on ADP is proposed in literature [29]. The first proposed HDP algorithm based on greedy iteration was reported in reference [30]. It mainly studies the infinite time optimal control design. In the basic ADP algorithm, it is generally necessary to construct two networks, namely, a critic network and an action network. The critic network is used to approximate the cost function, and the action network is used to approximate the control function [31]. Therefore, the training of the action network relies on the system dynamics in the existing iterative ADP algorithms. However, the structure of HDP cannot directly output the derivative function information of the cost function. Moreover, in general, network construction is mainly based on neural networks (NN) [32]. Memristor-Based Neural Network Circuit is discussed in literature [33]. As is known, a NN with a larger hidden node number tends to make the control more complex and increases the computation burden for the control system. Moreover, the NN neuron has exponential functions, which contribute to the complexity of the calculation and cause the NN control to fail to meet the real-time requirements.

Recently, the multi-dimensional Taylor network has been proposed. The multi-dimensional Taylor network (MTN) is a new structure [34]–[40]. The MTN that is a simple function of the state, input and is easy to analyze and solve for its polynomials. The MTN is good at approximating nonlinear dynamical systems, even unstable ones, as polynomials approach infinity well and can also accurately express the polynomial dynamical systems. In addition, the MTN only involves multiplication and addition; thus, its simple computation makes desirable real-time control possible. Due to the unique structural characteristic of MTN, its output is a linear combination of finite difference and product of its input, which is more suitable for computer implementation. Besides, the discrete MTN eliminates the dependence on the system model and reduces its design complexity. The adaptive controller based on MTN was proposed in [36]. However, the parameters of that controller are fixed. Then, an MTN tracking control scheme is proposed for a class of stochastic nonlinear systems with unknown input dead-zone [37]. And it is investigated the problem of adaptive MTN control for SISO uncertain stochastic non-linear systems [38].

To address the above issues, a data-driven approximate optimal control method for discrete-time nonlinear systems based on multidimensional Taylor network dynamic programming is proposed. The main contributions of the proposed control schemes are as followings:

1 Three networks, namely, the identification network, critic network and action network, were constructed based on the MTN. The parameter selection algorithm and the detailed control process are given.

2 The convergence of the algorithm is proved. The adaptive algorithm proposed in this paper can be guaranteed to be convergent under infinite time conditions.

3 The simulation experiment proves the effectiveness of the optimal control method proposed in this paper.

II. PROBLEM DESCRIPTION

Consider the following discrete nonlinear systems:

$$\begin{cases} \boldsymbol{x}(k+1) = \boldsymbol{F}(\boldsymbol{x}(k), \boldsymbol{u}(k)) \\ \boldsymbol{y}(k) = \boldsymbol{x}(k) \end{cases} \tag{1}$$

where $\boldsymbol{x}(k) = [x_1(k), x_2(k), \cdots, x_n(k)]^{\mathrm{T}} \in \Omega_x \subset \mathbf{R}^n$ is the state vector of the system;

$\boldsymbol{u}(k) = [u_1(k), u_2(k), \cdots, u_m(k)]^{\mathrm{T}} \in \Omega_u \subset \mathbf{R}^m$ is the control vector of the system; and

$\boldsymbol{y}(k) = [y_1(k), y_2(k), \cdots, y_n(k)]^{\mathrm{T}} \in \Omega_y \subset \mathbf{R}^n$ is the output vector of the system.

In addition, when $k = 0$, $\mathbf{x}(0) = [x_1(0), x_2(0), \cdots, x_n(0)]^{\mathrm{T}}$ is the initial vector of the system.

Here, the following assumptions are made:

Assumption 1: System (1) is controllable, that is, there is a set of control laws to stabilize the system.

Assumption 2: The nonlinear mapping $\boldsymbol{F}(\cdot)$ is Lipschitz continuous within the set Ω_x, and $\boldsymbol{F}(\mathbf{0}, \mathbf{0}) = 0$. This outcome shows that $\boldsymbol{x}(0) = \mathbf{0}$ is a balanced state of the system under the control law $\boldsymbol{u}(0) = \mathbf{0}$.

Control target: In the infinite time domain, designing the output feedback control law $\boldsymbol{u}(\boldsymbol{y})$ to stabilize the system from the initial state to the equilibrium state and minimize the cost function at the same time.

Cost function:

$$J(\boldsymbol{x}(k)) = \sum_{p=k}^{\infty} \gamma^{p-k} U(\boldsymbol{x}(p), \boldsymbol{u}(p)) \tag{2}$$

where U is the utility function, $U(0,0) = 0$, and for $\forall \boldsymbol{x}(k)$, $\forall \boldsymbol{u}(k)$, $U(\boldsymbol{x}(k), \boldsymbol{u}(k)) \geq 0$.

Utility factor $\gamma \in (0, 1]$.

Quadratic utility function:

$$U(\boldsymbol{x}(k), \boldsymbol{u}(k)) = \boldsymbol{x}^{\mathrm{T}}(k)P\boldsymbol{x}(k) + \boldsymbol{u}^{\mathrm{T}}(k)Q\boldsymbol{u}(k)$$

is chosen in this paper, where P and Q are Positive definite matrices.

The cost function provides a standard for us to evaluate the effect of learning. For a controllable system, through a control scheme from allowable solutions, the cost function of the system is continuously reduced during the process of movement, and that is the control target.

Set the optimal cost function

$$J^*(\boldsymbol{x}(k)) = \min \sum_{p=k}^{\infty} \gamma^{p-k} U(\boldsymbol{x}(p), \boldsymbol{u}(p)) \tag{3}$$

Further available

$$\begin{aligned} &J^*(\boldsymbol{x}(k)) \\ &= \min \left\{ U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma \sum_{p=k+1}^{\infty} \gamma^{p-k-1} U(\boldsymbol{x}(p), \boldsymbol{u}(p)) \right\}, \end{aligned}$$

Thus, $J^*(\boldsymbol{x}(k))$ satisfies the discrete time HJB equation

$$J^*(\boldsymbol{x}(k)) = \min \left\{ U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma J^*(\boldsymbol{x}(k+1)) \right\} \tag{4}$$

The corresponding optimal control $\boldsymbol{u}^*(k)$ is

$$\boldsymbol{u}^*(k) = \arg\min \left\{ U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma J^*(\boldsymbol{x}(k+1)) \right\} \tag{5}$$

This finding shows that the next time state vector $\boldsymbol{x}(k+1)$ is required to solve the optimal control $\boldsymbol{u}^*(k)$ at the current moment. However, this requirement is impossible at the current moment. Therefore, an iterative algorithm based on the multidimensional Taylor network is proposed in this paper to obtain an approximate solution.

III. MULTI-DIMENSIONAL TAYLOR NETWORK

The MTN can approximate any nonlinear functions with a finite point of discontinuity. Neat structure is the merit of MTN, whose parameters are easy to adjust.

The detailed application of the MTN can be found in [18]–[20].

Let

$$z(k) = [z_1(k), z_2(k), \ldots z_{n_z}(k)] \tag{6}$$

The basic structure of MTN is shown in Fig 1.

In other words, there exists a set of parameter vectors $\boldsymbol{W}_j(k) = [w_{j1}(k), w_{j2}(k), \cdots, w_{jN(n_z,t)}(k)]$ such that the output of MTN $O_{ut\,jn}(k)$ can be expressed as

$$O_{ut\,jn}(k) = \sum_{i=1}^{N(n_z,t)} w_{ji}(k) \prod_{s=1}^{n} z_i^{\lambda_{s,i}}(k). \tag{7}$$

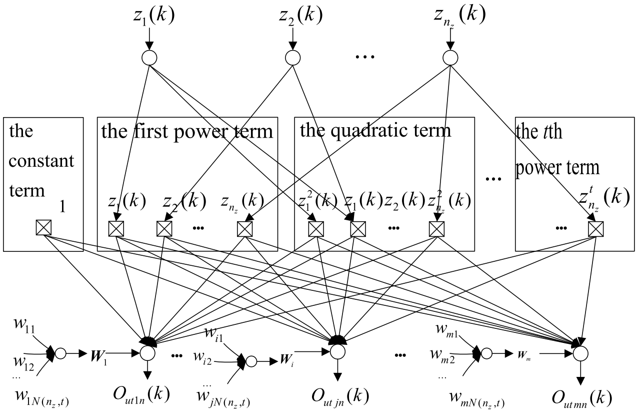

FIGURE 1. Basic structure of MTN.

where $N(n_z, t)$ is the total number of the expansion, $w_{ji}(k)$ is the weight of the *ith* product term, $\lambda(s, i)$ is the power of $z_s(k)$ in the ith product term, and $\sum_{s=1}^{n} \lambda_{s,i} \leq t$.

Setting $\eta(z(k)) = [1, z_1(k), z_2(k), \ldots, z_{n_z}(k), \ldots, z_1^2(k), z_1(k)z_2(k), \ldots, z_{n_z}^t(k)]^{\mathrm{T}}$

we obtain

$$O_{ut\,jn}(k) = W_j(k) \cdot \eta(z(k)). \tag{8}$$

IV. ITERATIVE ALGORITHM CONVERGENCE ANALYSIS

To prove the convergence of iterative algorithms, two basic sequences $\{V_i(\boldsymbol{x}(k))\}$ and $\{\boldsymbol{v}_i(\boldsymbol{x}(k))\}$ are constructed here.

$\{V_i(\boldsymbol{x}(k))\}$ denotes the cost function sequence, and $\{\boldsymbol{v}_i(\boldsymbol{x}(k))\}$ denotes the approximate optimal control laws. $\boldsymbol{v}$ is a vector, and the number of elements is the same as the number of elements in the control vector, $i = 0, 1, \cdots$.

Set $V_0(\cdot) = 0$, and $\boldsymbol{v}_0(\boldsymbol{x}(k)) = \arg\min\limits_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k))\}$

The iterative process is as follows:

$$\begin{aligned} &\boldsymbol{v}_i(\boldsymbol{x}(k)) \\ &\quad = \arg\min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_i(\boldsymbol{x}(k+1))\} \\ &\quad = \arg\min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_i(\boldsymbol{F}(\boldsymbol{x}(k), \boldsymbol{u}(k)))\} \end{aligned} \tag{9}$$

$$\begin{aligned} &V_{i+1}(\boldsymbol{x}(k)) \\ &\quad = \min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_i(\boldsymbol{x}(k+1))\} \\ &\quad = U(\boldsymbol{x}(k), \boldsymbol{v}_i(\boldsymbol{x}(k))) + \gamma V_i(\boldsymbol{F}(\boldsymbol{x}(k), \boldsymbol{v}_i(\boldsymbol{x}(k)))) \end{aligned} \tag{10}$$

Stop until $V_i \to J^*$ and $\boldsymbol{v}_i \to \boldsymbol{u}^*$.

Here, the following two lemmas are used.

Lemma 1 (Monotonicity): The cost function sequence $\{V_i(\boldsymbol{x}(k))\}$ is as shown in equation (10), where $V_0(\cdot) = 0$. With the control law sequence $\{\boldsymbol{v}_i(\boldsymbol{x}(k))\}$ given by equation (9), $\{V_i(\boldsymbol{x}(k))\}$ is a monotone non-decreasing sequence.

In other words, $\forall i$, $0 \leq V_i(\boldsymbol{x}(k)) \leq V_{i+1}(\boldsymbol{x}(k))$.

Lemma 2 (Boundedness): If the system is controllable and the cost function sequence $\{V_i(\boldsymbol{x}(k))\}$ is given by equation (10), there is an upper bound D such that $\forall i$, $0 \leq V_i(\boldsymbol{x}(k)) \leq D$.

Theorem 1: The cost function sequence $\{V_i(\boldsymbol{x}(k))\}$ is given by equation (10); the control law sequence $\{\boldsymbol{v}_i(\boldsymbol{x}(k))\}$ is given by equation (9); and $V_0(\cdot) = 0$; then the cost function sequence $\{V_i(\boldsymbol{x}(k))\}$ converges to $J^*(\boldsymbol{x}(k))$.

In other words, when $i \to \infty$, $V_i(\boldsymbol{x}(k)) \to J^*(\boldsymbol{x}(k))$, and $\{\boldsymbol{v}_i(\boldsymbol{x}(k))\}$ converges to the optimal control law $\boldsymbol{u}^*(\boldsymbol{x}(k))$. In other words, $\lim\limits_{i\to\infty} \boldsymbol{v}_i(\boldsymbol{x}(k)) = \boldsymbol{u}^*(\boldsymbol{x}(k))$.

Proof: According to Lemmas 1 and 2, the cost function sequence $\{V_i(\boldsymbol{x}(k))\}$ is monotone non-decreased and bounded. Thus, there is a limit $V_\infty(\boldsymbol{x}(k))$, that is, $\lim\limits_{i\to\infty} V_i(\boldsymbol{x}(k)) = V_\infty(\boldsymbol{x}(k))$.

Additionally, for $\forall \boldsymbol{u}(k)$, $\forall i$, as seen from equation (10):

$$V_i(\boldsymbol{x}(k)) \leq U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_{i-1}(\boldsymbol{x}(k+1)) \tag{11}$$

According to Lemma 1, for $\forall i$, $V_i(\boldsymbol{x}(k)) \leq V_\infty(\boldsymbol{x}(k))$. Thus, we can obtain

$$V_i(\boldsymbol{x}(k)) \leq U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_\infty(\boldsymbol{x}(k+1)) \tag{12}$$

When $i \to \infty$,

$$V_\infty(\boldsymbol{x}(k)) \leq U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_\infty(\boldsymbol{x}(k+1)) \tag{13}$$

Furthermore, $\boldsymbol{u}(k)$ of the formula 13 is arbitrary, then we can obtain

$$V_\infty(\boldsymbol{x}(k)) \leq \min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_\infty(\boldsymbol{x}(k+1))\} \tag{14}$$

However, for $\forall i$

$$V_i(\boldsymbol{x}(k)) = \min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_{i-1}(\boldsymbol{x}(k+1))\} \tag{15}$$

and $V_i(\boldsymbol{x}(k)) \leq V_\infty(\boldsymbol{x}(k))$; thus, we can obtain

$$V_\infty(\boldsymbol{x}(k)) \geq \min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_{i-1}(\boldsymbol{x}(k+1))\} \tag{16}$$

Set $i \to \infty$, and according to (14) and (16)

$$V_\infty(\boldsymbol{x}(k)) = \min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_\infty(\boldsymbol{x}(k+1))\} \tag{17}$$

Similarly, $\boldsymbol{v}_\infty(\boldsymbol{x}(k))$ is the limit of $\boldsymbol{v}_i(\boldsymbol{x}(k))$.

In other words, $\lim\limits_{i\to\infty} \boldsymbol{v}_i(\boldsymbol{x}(k)) = \boldsymbol{v}_\infty(\boldsymbol{x}(k))$.

According to (9) and (10)

$$\begin{aligned} V_\infty(\boldsymbol{x}(k)) &= \min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_\infty(\boldsymbol{x}(k+1))\} \\ &= U(\boldsymbol{x}(k), \boldsymbol{v}_\infty(k)) + \gamma V_\infty(\boldsymbol{F}(\boldsymbol{x}(k), \boldsymbol{v}_\infty(k))) \end{aligned} \tag{18}$$

where

$$\boldsymbol{v}_\infty(k) = \arg\min_{\boldsymbol{u}(k)} \{U(\boldsymbol{x}(k), \boldsymbol{u}(k)) + \gamma V_\infty(\boldsymbol{x}(k+1))\} \tag{19}$$

known by (18) and (3), with (19) and (4), we can obtain

$$V_\infty(\boldsymbol{x}(k)) = J^*(\boldsymbol{x}(k)), \quad \textit{and } \boldsymbol{v}_\infty(k) = \boldsymbol{u}^*(\boldsymbol{x}(k)).$$

Thus,

$$\lim_{i\to\infty} V_i(\boldsymbol{x}(k)) = J^*(\boldsymbol{x}(k)),$$
$$\lim_{i\to\infty} \boldsymbol{v}_i(k) = \boldsymbol{u}^*(\boldsymbol{x}(k)).$$

The proof is completed.

V. ITERATIVE ALGORITHM AND ITS IMPLEMENTATION

For the general HJB equation of the nonlinear system is difficult to solve, the optimal control law and the optimal iterative function can be obtained via an iterative algorithm in principle.

However, because the controlled system is unknown, the construction of the system dynamics $\{V_i(\boldsymbol{x}(k))\}$ and $\{\boldsymbol{v}_i(\boldsymbol{x}(k))\}$ is required. A dynamic iterative implementation based on the multidimensional Taylor network is proposed in this section. The model mainly includes the construction of three networks, namely, the identification network, critic network and action network.

A. IDENTIFICATION NETWORK

Before implementing the iterative control process, an identification network must be built to ensure that the control process does not require dynamic information of the system. The weight vectors of the identification network are $\omega_m \in R^{N_m}$, where N_m is a number of the multi-dimensional Taylor network identification network expansion items.

The output of the identification network:

$$\hat{x}(k+1) = \omega_m^{\mathrm{T}} \eta_m(x(k)^{\mathrm{T}}, \hat{\boldsymbol{v}}_i^{\mathrm{T}}(\boldsymbol{x}(k))) \tag{20}$$

where $\eta_m(\cdot)$ is the expansion items of the multi-dimensional Taylor network identification network.

The error function of the identification network is

$$e_m(k) = \hat{x}(k+1) - x(k+1) \tag{21}$$

The training objective function is

$$E_m(k) = \frac{1}{2} e_m(k)^{\mathrm{T}} e_m(k) \tag{22}$$

The weight vectors of the identification network are updated via the gradient method:

$$\omega_m^{(j+1)} = \omega_m^{(j)} - \alpha_m \left[\frac{\partial E_m(k)}{\partial \omega_m^{(j)}} \right] \tag{23}$$

where $\alpha_m > 0$ is the model learning rate, and j is the training weight iteration index.

When the identification network is fully trained and the weights are no longer changed, the training of both the critic network and the action network is started.

B. CRITIC NETWORK

The role of the critic network is to approximate the cost function $V_i(\boldsymbol{x}(k))$ and the partial derivative of the cost function $\frac{\partial V_i(\boldsymbol{x}(k))}{\partial \boldsymbol{x}(k)}$.

Let $\lambda_i(x(k))$ be a co-function.

Thus:

$$\lambda_i(x(k)) := \frac{\partial V_i(\boldsymbol{x}(k))}{\partial \boldsymbol{x}(k)}. \tag{24}$$

According to Theorem 1, when $i \to \infty$, we have

$$V_i(\boldsymbol{x}(k)) = J^*(\boldsymbol{x}(k)). \tag{25}$$

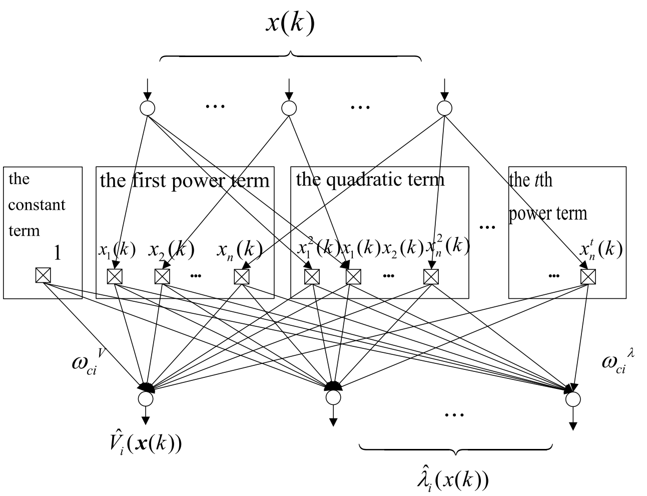

FIGURE 2. The architecture network.

Thus, because

$$\lambda_i(x(k)) = \frac{\partial V_i(\boldsymbol{x}(k))}{\partial \boldsymbol{x}(k)} \tag{26}$$

when $i \to \infty$, the co-function is also convergent. Therefore,

$$\lambda_i(x(k)) \to \lambda^*(x(k)) \tag{27}$$

Set N_c to be the number of multi-dimensional Taylor network critic network expansion items. Thus, the weight vector of the critic network is $\omega_c \in R^{N_c}$. At the ith iteration, the output of the critic network is

$$\begin{bmatrix}\hat{V}_i(\boldsymbol{x}(k))\\ \hat{\lambda}_i(x(k))\end{bmatrix} = \begin{bmatrix}\omega_{ci}^{V\mathrm{T}}\\ \omega_{ci}^{\lambda\mathrm{T}}\end{bmatrix}\eta\left(\boldsymbol{x}(k)\right) = \omega_{ci}^{\mathrm{T}}\eta\left(\boldsymbol{x}(k)\right) \tag{28}$$

where $\omega_{ci} = \left[\omega_{ci}^{V}, \omega_{ci}^{\lambda}\right]$. Expanding the equation, we can obtain

$$\hat{V}_i(\boldsymbol{x}(k)) = \omega_{ci}^{V\mathrm{T}}\eta_c\left(\boldsymbol{x}(k)\right) \tag{29}$$

$$\hat{\lambda}_i(x(k)) = \omega_{ci}^{\lambda\mathrm{T}}\eta_c\left(\boldsymbol{x}(k)\right) \tag{30}$$

where $\eta_c(\cdot)$ is the expansion items of the multi-dimensional Taylor network critic network.

Although the introduction of co-functions increases the amount of computation to a certain extent, the cost function can be output directly. Furthermore, the control effect is also improved to some extent.

The structure is shown in Fig. 2.

The training objective of the critic network consists of two parts, namely, the cost function and the co-function, i.e.,

$$V_i(\boldsymbol{x}(k)) = U(\boldsymbol{x}(k), \hat{\boldsymbol{v}}_{i-1}(\boldsymbol{x}(k))) + \gamma\hat{V}_{i-1}(\hat{\boldsymbol{x}}(k+1)) \tag{31}$$

$$\begin{aligned}\lambda_i(\boldsymbol{x}(k)) &= 2Q\boldsymbol{x}(k) + 2\left(\frac{\partial\hat{\boldsymbol{v}}_{i-1}(\boldsymbol{x}(k))}{\partial\boldsymbol{x}(k)}\right)^{\mathrm{T}} R\hat{\boldsymbol{v}}_{i-1}(\boldsymbol{x}(k))\\ &\quad+\gamma\left(\frac{\partial\hat{\boldsymbol{x}}(k+1)}{\partial\boldsymbol{x}(k)} + \frac{\partial\hat{\boldsymbol{x}}(k+1)}{\partial\hat{\boldsymbol{v}}_{i-1}(\boldsymbol{x}(k))}\frac{\partial\hat{\boldsymbol{v}}_{i-1}(\boldsymbol{x}(k))}{\partial\boldsymbol{x}(k)}\right)^{\mathrm{T}}\\ &\quad\times\hat{\lambda}_{i-1}(\boldsymbol{x}(k+1))\end{aligned} \tag{32}$$

The training errors include two parts:

$$\begin{aligned}e_{ci}^{V}(k) &= \hat{V}_i(\boldsymbol{x}(k)) - V_i(\boldsymbol{x}(k)) \quad \text{and}\\ e_{ci}^{\lambda}(k) &= \hat{\lambda}_i(x(k)) - \lambda_i(x(k)).\end{aligned} \tag{33}$$

Setting $E_{ci}^{V}(k) = \frac{1}{2}e_{ci}^{V\mathrm{T}}(k)e_{ci}^{V}(k)$ and

$$E_{ci}^{\lambda}(k) = \frac{1}{2}e_{ci}^{\lambda\mathrm{T}}(k)e_{ci}^{\lambda}(k),$$

the objective function is

$$E_{ci}(k) = (1-\beta)\,E_{ci}^{V}(k) + \beta E_{ci}^{\lambda}(k) \tag{34}$$

The weight of the critic network is updated by the gradient descent method to obtain

$$\omega_c^{(j+1)} = \omega_c^{(j)} - \alpha_c\left[(1-\beta)\,\frac{\partial E_{ci}^{V}(k)}{\partial\omega_{ci}^{(j)}} + \beta\frac{\partial E_{ci}^{\lambda}(k)}{\partial\omega_{ci}^{(j)}}\right] \tag{35}$$

where $\alpha_c > 0$ is the learning rate of the critic network, j is the training weight iteration index, and $0 < \beta < 1$ is a constant used to reflect the weight of $E_{ci}^{V}(k)$ and $E_{ci}^{\lambda}(k)$.

C. ACTION NETWORK

The role of the multidimensional Taylor network action network is to approximate the optimal control law. Set N_a to be the number of multi-dimensional Taylor network action network expansion items. Thus, the weight vector of the action network is $\omega_a \in R^{N_a}$, and the output of the critic network is

$$\hat{v}_{i-1}(\boldsymbol{x}(k)) = \omega_{a(i-1)}^{\mathrm{T}}\eta_a(x(k)) \tag{36}$$

where $\eta_a(\cdot)$ is the expansion items of the multi-dimensional Taylor network action network.

Set the error function

$$ea(i-1)(k) = \hat{V}_{i-1}(\boldsymbol{x}(k+1)) - S(k) \tag{37}$$

where $S(k) = 0$ is the target value of $\hat{V}_{i-1}(\boldsymbol{x}(k+1))$.

Thus, the Objective function is

$$Ea(i-1)(k) = \frac{1}{2}e_{a(i-1)}^{\mathrm{T}}(k)ea(i-1)(k) \tag{38}$$

To minimize the objective function, the weight of the action network is adjusted using the gradient method:

$$\omega_{a(i-1)}^{(j+1)} = \omega_{a(i-1)}^{(j)} - \alpha_a\left[\frac{\partial E_{a(i-1)}(k)}{\partial\omega_{a(i-1)}^{(j)}}\right] \tag{39}$$

where $\alpha_a > 0$ is the learning rate of the action network, and j is the training weight iteration index.

The traditional control methods too much rely on the dynamic information of the controlled object. In the process of training the action network, it is necessary to use the direct information of the control matrix or rely on the neural network to express it. However, using traditional control strategies such as the basic framework, the control method proposed in this paper can guarantee the convergence of iterative algorithms. Moreover, the proposed method can relax the dynamic requirements of the system, making it easier to achieve the control effects.

The basic control structure is shown in Fig. 3.

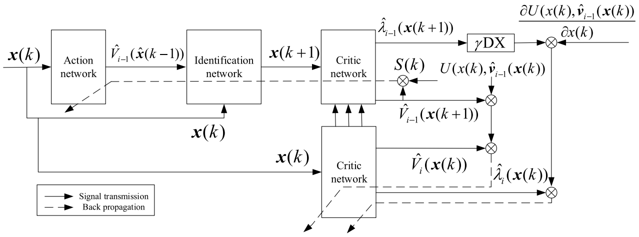
$\boldsymbol{x}(k)$
Action network
$\hat{V}_{i-1}(\hat{\boldsymbol{x}}(k-1))$
Identification network
$\boldsymbol{x}(k+1)$
Critic network
$\hat{\lambda}_{i-1}(\boldsymbol{x}(k+1))$
γDX
$\frac{\partial U(x(k),\hat{\boldsymbol{v}}_{i-1}(\boldsymbol{x}(k))}{\partial x(k)}$
$S(k)$
$U(x(k),\hat{\boldsymbol{v}}_{i-1}(\boldsymbol{x}(k))$
$\hat{V}_{i-1}(\boldsymbol{x}(k+1))$
$\boldsymbol{x}(k)$
$\boldsymbol{x}(k)$
Critic network
$\hat{V}_i(\boldsymbol{x}(k))$
$\hat{\lambda}_i(\boldsymbol{x}(k))$
Signal transmission
Back propagation

FIGURE 3. The architecture of iterative dynamic programming.

D. CONTROL PROCESS

Suppose $x(k)$ is in any control state, and $J^*(x(k))$ is the optimal cost function. According to theorem 1, when the iteration index $i \to \infty$, $V_i(x(k)) \to J^*(x(k))$.

However, it is impossible to perform an iterative algorithm infinitely. As a result, the error ε is introduced so that

$$\left|J^*(x(k)) - V_i(x(k))\right| \leq \varepsilon \tag{40}$$

to guarantee the cost function can converge after a finite number of iterations. This approximation in the practical sense can meet the needs of the general engineering design.

However, in the general situation, the optimal cost function $J^*(x(k))$ is unknown in advance. It is difficult to use the error as a stopping criterion. Therefore, the following stopping criterion is used here:

$$|V_{i+1}(x(k)) - V_i(x(k))| \leq \varepsilon. \tag{41}$$

Theorem 2: For nonlinear systems (1) and cost functions (2), in the iterative process, the convergence criteria (40) and (41) are equivalent.

Proof: If $|J^*(x(k)) - V_i(x(k))| \leq \varepsilon$ is established, then we have the following:

$$J^*(x(k)) \leq V_i(x(k)) + \varepsilon \tag{42}$$

According to Theorem 1, the following inequality is established:

$$V_i(x(k)) \leq V_{i+1}(x(k)) \leq J^*(x(k)) \tag{43}$$

In other words, $V_i(x(k)) \leq V_{i+1}(x(k)) \leq V_i(x(k))+\varepsilon$; thus, we can obtain

$$0 \leq V_{i+1}(x(k)) - V_i(x(k)) \leq \varepsilon \tag{44}$$

Therefore, equation (41) is established.

Moreover, according to theorem 1,

$$|V_{i+1}(x(k)) - V_i(x(k))| \to 0.$$

In other words, $V_i(x(k)) \to J^*(x(k))$. Thus, for any small ε, we have $|V_{i+1}(x(k)) - V_i(x(k))| \leq \varepsilon$.

When $i \to \infty$,

$$\left|J^*(x(k)) - V_i(x(k))\right| \leq \varepsilon \text{ is established.}$$

The Proof is completed.

A design criterion is provided by theorem 2. Therefore, in practical applications, applying the control strategy proposed in this paper can obtain more reasonable control effects.

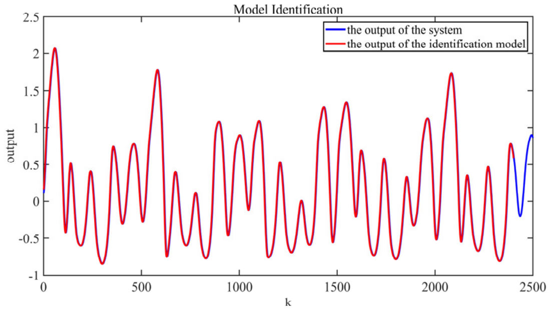

FIGURE 4. Identification model.

Theorem 2 validates the equivalence between formula 40 and formula 41. And the important role of the Theorem 2 lies in that it provides practical design criteria of approximate optimal regulation for discrete-time nonlinear systems using iterative MTN dynamic programming method.

VI. SIMULATION EXAMPLE

To validate the controller proposed in this paper, consider the following nonlinear system:

$$\begin{cases} x_1(k+1) = 0.68 \cdot x_1(k) + 0.1 \cdot x_2(k) \\ x_2(k+1) = 0.93 \cdot x_2(k) - 0.23 \cdot x_1^2(k) \cdot x_2(k) \\ -0.16 \cdot x_1^2(k) + u(k) + 0.1 \cdot (x_1^3(k)+x_2(k)) \cdot \dfrac{1-e^{-u(k)}}{1+e^{-u(k)}} \\ y(k) = x_1(k) \end{cases} \tag{45}$$

Setting $\boldsymbol{x}(k) = [x_1(k), x_2(k)]^{\mathrm{T}}$, the utility function is as follows:

$$U(\boldsymbol{x}(k), u(k)) = \boldsymbol{x}^{\mathrm{T}}(k) \cdot \boldsymbol{x}(k) + u^{\mathrm{T}}(k) \cdot u(k)$$

The initial parameters of the system are

$$x_1(0) = 0, x_2(0) = 0.15.$$

When the excitation function is

$$u_r(k) = 0.8 \cdot \sin(k/17) + 0.7 \cdot \cos(k/80) + 0.4 \cdot \sin(k/27)$$

the curve of the identification model by the identification network proposed in this paper is shown in Fig. 4.

For the method proposed in this paper, when the dimension n of the controller is equal to 2, the unit step response curve is as shown in Fig. 5.

Alternatively, the BP neural network self-adaption reconstitution algorithm gives the unit step response curve shown in Fig. 5.

Fig. 5 shows that the data-driven approximate optimal control method based on the multi-dimensional Taylor network has a faster response.

To verify the follow-up response performance of the controller, when $k = 10$, with the input curve overlaying a sinusoidal signal, Fig. 10 shows the response curves.

Fig. 6 reveals that the data-driven approximate optimal control method tracks the desired signal more quickly. And

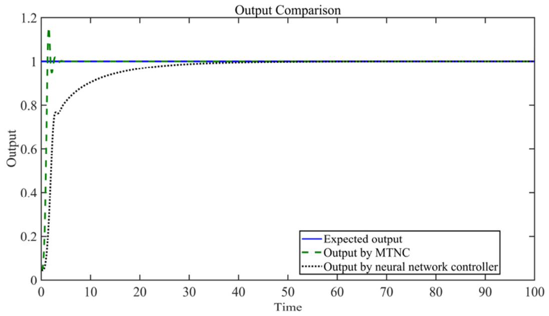

FIGURE 5. Output comparison.

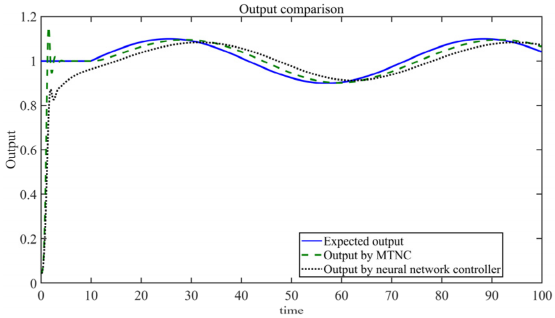

FIGURE 6. Output comparison.

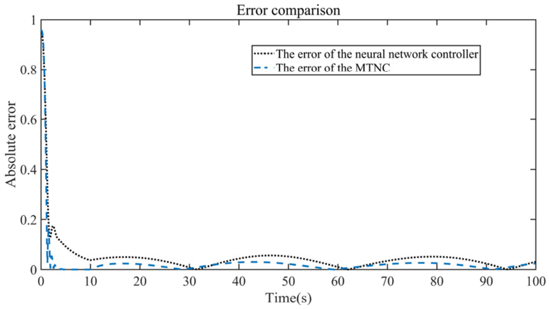

FIGURE 7. Error comparison.

convergence proof of the proposed algorithm has been described in detail in literature [39] and literature [40].

The algorithm proposed in this paper and the neural network algorithm are all differential adjustment algorithm. That is, the error occurs first in the system, and then the controller acts to reduce the error and eventually to zero. Under this control mechanism, the faster response speed can ensure that the system fluctuates within a relatively small error range. To illustrate this problem, Fig. 7 shows the absolute error curves of the two algorithms.

It can be seen from the figure that the absolute error of the algorithm proposed in this paper is always smaller than that of the contrast algorithm. Thank you again for your valuable comments.

VII. CONCLUSION

For discrete time nonlinear systems, which are based on a data-driven approach, an approximate optimal iterative dynamic programming method based on MTN was proposed in this paper. Moreover, the convergence of the iterative algorithm was proved. Based on MTN, three networks were constructed: the identification network, critic network and action network. As the construction and execution of the control network do not require dynamic information of the controlled system here, the dependence on the structure of the controlled object model will be greatly reduced. The effectiveness of the method proposed in this paper was verified by the simulation results. Compared with the traditional NN control method, the iterative algorithm proposed here based on MTN has faster response speed.

The current research focuses on theoretical analysis in this paper. The convergence of the algorithm under time-infinite conditions was proved. Directions of further research include determination of how to promote the results to a limited time and how to prove the convergence of iterative algorithm in finite time. In addition, an approach to combine theoretical methods with practice requires further study.

REFERENCES

[1] Y. Huang, "Optimal guaranteed cost control of uncertain non-linear systems using adaptive dynamic programming with concurrent learning," *IET Control Theory Appl.*, vol. 12, no. 8, pp. 1025–1035, May 2018.

[2] T. Jiang, C. Zhang, and Q.-M. Sun, "Green job shop scheduling problem with discrete whale optimization algorithm," *IEEE Access*, vol. 7, pp. 43153–43166, 2019.

[3] Y. Zhang, S. Li, and X. Jiang, "Near-optimal control without solving HJB equations and its applications," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7173–7184, Sep. 2018.

[4] P. J. Werbos, *Approximate Dynamic Programming of Real-Time Control and Neural Modeling. Handbook of Intelligent Control*. New York, NY, USA: Van Nostrand, 1992.

[5] R. E. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[6] H. Zhang, C. Qin, and Y. Luo, "Neural-Network-Based constrained optimal control scheme for discrete-time switched nonlinear system using dual heuristic programming," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 3, pp. 839–849, Jul. 2014.

[7] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controller," *IEEE Control Syst.*, vol. 32, no. 6, pp. 76–105, Nov. 2012.

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

[9] T. H. Jiang and C. Zhang, "Adaptive discrete cat swarm optimization algorithm for the flexible job shop problem," *Int. J. Bio-Inspired Comput.*, vol. 13, no. 3, pp. 199–208, 2019.

[10] 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 Trans. Syst. Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 937–942, Aug. 2008.

[11] X. Xu, Z. Hou, C. Lian, and H. He, "Online learning control using adaptive critic designs with sparse kernel machines," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 5, pp. 762–775, May 2013.

[12] X. Zhong, Z. Ni, and H. He, "Gr-GDHP: A new architecture for globalized dual heuristic dynamic programming," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3318–3330, Oct. 2017.

[13] L. Le-Tien and A. Albu-Schaffer, "Robust adaptive tracking control based on state feedback controller with integrator terms for elastic joint robots with uncertain parameters," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 6, pp. 2259–2267, Nov. 2018.

[14] J. Sun, Y. Wu, G. Cui, and Y. Wang, "Finite-time real combination synchronization of three complex-variable chaotic systems with unknown parameters via sliding mode control," *Nonlinear Dyn.*, vol. 88, no. 3, pp. 1677–1690, Feb. 2017.

[15] K. Sachan and R. Padhi, ''Output-constrained robust adaptive control for uncertain nonlinear MIMO systems with unknown control directions,'' *IEEE Control Syst. Lett.*, vol. 3, no. 4, pp. 823–828, Oct. 2019.

[16] J. Si and Y. T. Wang, ''Online learning control by association and reinforcement,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 12, no. 2, pp. 264–276, Mar. 2001.

[17] D. Liu, Q. Wei, and P. Yan, ''Generalized policy iteration adaptive dynamic programming for discrete-time nonlinear systems,'' *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 12, pp. 1577–1591, Dec. 2015.

[18] J. Sun, X. Zhao, J. Fang, and Y. Wang, ''Autonomous memristor chaotic systems of infinite chaotic attractors and circuitry realization,'' *Nonlinear Dyn.*, vol. 94, no. 4, pp. 2879–2887, Aug. 2018.

[19] 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 Trans. Neural Netw.*, vol. 20, no. 9, pp. 1490–1503, Sep. 2009.

[20] C. H. Hu, P. P. Li, and Z. Pan, ''Phenotyping of poplar seedling leaves based on a 3D visualization method,'' *Int. J. Agricult. Biol. Eng.*, vol. 11, no. 6, pp. 145–151, 2018.

[21] C. Hu, Z. Pan, and P. Li, ''A 3D point cloud filtering method for leaves based on manifold distance and normal estimation,'' *Remote Sens.*, vol. 11, no. 2, p. 198, Jan. 2019.

[22] 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 Trans. Autom. Sci. Eng.*, vol. 9, no. 3, pp. 628–634, Jul. 2012.

[23] F. Y. Wang, N. Jin, D. R. Liu, and Q. L. Wei, ''Adaptive dynamic programming for finite-horizon optimal of discrete-time nonlinear systems with ϵ-error bound,'' *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 24–36, Sep. 2011.

[24] D. Wang, D. Liu, Q. Wei, D. Zhao, and N. Jin, ''Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming,'' *Automatica*, vol. 48, no. 8, pp. 1825–1832, Aug. 2012.

[25] J. Zhang and Q. Sun, ''Prescribed performance adaptive neural output feedback dynamic surface control for a class of strict-feedback uncertain nonlinear systems with full state constraints and unmodeled dynamics,'' *Int. J. Robust Nonlinear Control*, vol. 30, no. 2, pp. 459–483, Oct. 2019, doi: 10.1002/rnc.4769.

[26] H. Zhang, L. Cui, X. Zhang, and Y. Luo, ''Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method,'' *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2226–2236, Dec. 2011.

[27] H. Zhang, Q. Wei, and D. Liu, ''An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games,'' *Automatica*, vol. 47, no. 1, pp. 207–214, Jan. 2011.

[28] F. L. Lewis and K. G. Vamvoudakis, ''Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data,'' *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 14–25, Feb. 2011.

[29] H. Zhang, C. Qin, B. Jiang, and Y. Luo, ''Online adaptive policy learning algorithm for H_∞ state feedback control of unknown affine nonlinear discrete-time systems,'' *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2706–2718, Dec. 2014.

[30] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, ''Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof,'' *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 943–949, Aug. 2008.

[31] T. Dierks, B. T. Thumati, and S. Jagannathan, ''Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence,'' *Neural Netw.*, vol. 22, nos. 5–6, pp. 851–860, Jul. 2009.

[32] P. He and S. Jagannathan, ''Reinforcement learning neural-network-based controller for nonlinear discrete-time systems with input constraints,'' *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 425–436, Apr. 2007.

[33] J. Sun, G. Han, Z. Zeng, and Y. Wang, ''Memristor-based neural network circuit of full-function pavlov associative memory with time delay and variable learning rate,'' *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2019.2951520.

[34] B. Zhou and H. S. Yan, ''Financial time series forecasting based on wavelet and multi-dimensional Taylor network dynamics model,'' (in Chinese), *Syst. Eng.-Theory Pract.*, vol. 33, no. 10, pp. 2654–2662, 2013.

[35] H.-S. Yan and A.-M. Kang, ''Asymptotic tracking and dynamic regulation of SISO non-linear system based on discrete multi-dimensional Taylor network,'' *IET Control Theory Appl.*, vol. 11, no. 10, pp. 1619–1626, Jun. 2017.

[36] A.-M. Kang and H.-S. Yan, ''Stability analysis and dynamic regulation of multi-dimensional Taylor network controller for SISO nonlinear systems with time-varying delay,'' *ISA Trans.*, vol. 73, pp. 31–39, Feb. 2018.

[37] Q. M. Sun and H. S. Yan, ''Optimal adjustment control of SISO nonlinear systems based on multi-dimensional Taylor network only by output feedback,'' *Adv. Mater. Res.*, vols. 1049–1050, pp. 1389–1391, Oct. 2014.

[38] Y. Han, S. Zhu, and S. Yang, ''Adaptive multi-dimensional Taylor network tracking control for a class of stochastic nonlinear systems with unknown input dead-zone,'' *IEEE Access*, vol. 6, pp. 34543–34554, 2018.

[39] Y.-Q. Han and H.-S. Yan, ''Adaptive multi-dimensional Taylor network tracking control for SISO uncertain stochastic non-linear systems,'' *IET Control Theory Appl.*, vol. 12, no. 8, pp. 1107–1115, May 2018.

[40] Q.-M. Sun and H.-S. Yan, ''Multi-dimensional Taylor network modelling and optimal control of SISO nonlinear systems for tracking by output feedback,'' *IMA J. Math. Control Inf.*, Jul. 2019, doi: 10.1093/imamci/dnz020.

QI-MING SUN received the B.S. degree from the Tianjin University of Technology, Tianjin, China, in 2010, and the Ph.D. degree in control theory and control engineering from Southeast University, Nanjing, China, in 2018. Since January 2019, he has been a Lecturer with the College of Information Science and Technology, Nanjing Forestry University. His major research interest is in multi-dimensional taylor network (MTN) optimal control of nonlinear systems.

CHAO ZHANG received the B.S. degree in automation from the Zhongyuan University of Technology, Zhengzhou, China, in 2007, the M.S. degree in control science and engineering from the University of Science and Technology Beijing, Beijing, China, in 2010, and the Ph.D. degree in control theory and control engineering from Southeast University, Nanjing, China, in 2018.

Since April 2014, he has been a Lecturer with the School of Electrical Engineering and Automation, Henan Institute of Technology, Xinxiang, China. He is the author of four books and more than 20 articles. His research interests include nonlinear systems, adaptive control, and MTN optimal control.

NAN-YUN JIANG received the B.S. degree in electric science and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2004, and the M.S. and Ph.D. degrees in control theory and control engineering from Southeast University, Nanjing, China, in 2007 and 2018, respectively.

Since April 2007, she has been a Lecturer with the Industrial Engineering Department, School of Economics and Management, Nanjing Tech University. Her research interests include flexible manufacturing systems, computer integrated manufacturing systems (CIMS), and production planning and scheduling.

JING-JING YU received the master's degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2012. She joined China Railway Shanghai Design Institute Group Co., Ltd., in 2012. She is currently the Director of electric power with the Department of Electrical Design Department.

LEI XU received the B.S. degree in automation from the Henan University of Technology, Henan, China, in 2012, and the M.S. degree in control theory and control engineering from Nanjing Forestry University, Nanjing, China, in 2015.

Since July 2015, he has been a Lecturer with the College of Information Science and Technology, Nanjing Forestry University. His research interests include automatic control, electrical control systems, and high-voltage electrical design.