Policy iteration optimal tracking control for chaotic systems by using an adaptive dynamic programming approach*

Wei Qing-Lai(魏庆来)†, Liu De-Rong(刘德荣), and Xu Yan-Cai(徐延才)

The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, *China*

(Received 22 July 2014; revised manuscript received 9 October 2014; published online 30 January 2015)

A policy iteration algorithm of adaptive dynamic programming (ADP) is developed to solve the optimal tracking control for a class of discrete-time chaotic systems. By system transformations, the optimal tracking problem is transformed into an optimal regulation one. The policy iteration algorithm for discrete-time chaotic systems is first described. Then, the convergence and admissibility properties of the developed policy iteration algorithm are presented, which show that the transformed chaotic system can be stabilized under an arbitrary iterative control law and the iterative performance index function simultaneously converges to the optimum. By implementing the policy iteration algorithm via neural networks, the developed optimal tracking control scheme for chaotic systems is verified by a simulation.

Keywords: adaptive critic designs, adaptive dynamic programming, approximate dynamic programming, neuro-dynamic programming

PACS: 05.45.Gg **DOI:** 10.1088/1674-1056/24/3/030502

1. Introduction

The control of chaotic systems has been the focus of control research in the past decade.[1–7] For most control methods of chaotic systems, such as impulsive control methods[2–4,6] and adaptive synchronization control methods,[8–10] only the stability properties of the chaotic systems are considered. The optimality property is an important performance of chaotic control systems. As is well known, dynamic programming is a very useful tool in solving optimal control problems.[11] However, due to the difficulties related to solving the time-varying Hamilton–Jacobi–Bellman (HJB) equations, the closed-loop optimal feedback control can hardly be obtained. Approximate optimal solutions of the optimal problem have consequently attracted much attention.[12–16] Among those approximate approaches, the adaptive dynamic programming (ADP) algorithm, proposed by Werbos,[17,18] has played an important role in seeking the approximate solutions of dynamic programming problems.[19–21] There are several synonyms used for ADP, including adaptive critic designs,[22] adaptive dynamic programming,[23–27] approximate dynamic programming,[28] neural dynamic programming,[29] and reinforcement learning.[30] In Refs. [22] and [28], ADP approaches were classified into several main schemes, which are heuristic dynamic programming (HDP), action-dependent HDP (ADHDP), dual heuristic dynamic programming (DHP), action-dependent DHP (ADDHP), which is also called Q-learning,[31] globalized DHP (GDHP), and action-dependent GDHP (ADGDHP). Iterative methods are also used in ADP to obtain the solution of the HJB equation indirectly and they have received increasing attention.[32–40] There are two main iterative ADP algorithms,[41] which are based on value and policy iterations, respectively.

Value iteration algorithms for the optimal control of discrete-time nonlinear systems were given in Ref. [42]. Al-Tamimi and Lewis[43] studied the deterministic discrete-time affine nonlinear systems and a value iteration algorithm, which is referred to as HDP and proposed for finding the optimal control law. Starting from $J^{[0]}(\boldsymbol{x}_k) \equiv 0$, the value iteration algorithm is updated by

$$\begin{cases} \boldsymbol{u}^{[i]}(\boldsymbol{x}_k) = -\dfrac{1}{2}\boldsymbol{R}^{-1}\boldsymbol{g}(\boldsymbol{x}_k)^{\mathrm{T}}\dfrac{\partial \boldsymbol{J}^{[i]}(\boldsymbol{x}_{k+1})}{\partial \boldsymbol{x}_{k+1}}, \\ J^{[i+1]}(\boldsymbol{x}_k) = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}_k + \boldsymbol{u}^{[i]\mathrm{T}}(\boldsymbol{x}_k)\boldsymbol{R}\boldsymbol{u}^{[i]}(\boldsymbol{x}_k) + J^{[i]}(\boldsymbol{x}_{k+1}), \end{cases} \tag{1}$$

for $i = 0, 1, 2, \ldots$, where $\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k) + \boldsymbol{g}(\boldsymbol{x}_k)\boldsymbol{u}^{[i]}(\boldsymbol{x}_k)$. In Ref. [43], it was proved that $J^{[i]}(\boldsymbol{x}_k)$ is a nondecreasing sequence and bounded, and hence converges to $J^*(\boldsymbol{x}_k)$. Zhang *et al.* applied a value iteration algorithm for optimal tracking problems.[36] Song *et al.* successfully implemented value iteration algorithms to find the optimal tracking control law for Hénon chaotic systems.[44] For most previous value iteration algorithms of ADP, the initial performance index function is chosen as zero.[36,43,44] For each of the iterative controls $\boldsymbol{u}^{[i]}(\boldsymbol{x}_k)$, $i = 0, 1, \ldots$, the stability of the system cannot be guaranteed. This means that only the converged $\boldsymbol{u}^*(\boldsymbol{x}_k)$ can be used to control the nonlinear system, and all the iterative controls $\boldsymbol{u}^{[i]}(\boldsymbol{x}_k)$, $i = 0, 1, \ldots$ may be invalid. So the computation efficiency of the value iteration ADP method is very low.

Policy iteration algorithms for the optimal control of continuous-time systems with continuous state and action

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61034002, 61233001, 61273140, 61304086, and 61374105) and the Beijing Natural Science Foundation, China (Grant No. 4132078).

†Corresponding author. E-mail: qinglai.wei@ia.ac.cn

© 2015 Chinese Physical Society and IOP Publishing Ltd

spaces were given in Refs. [12] and [32]. In Ref. [12], Murray *et al.* proved that for continuous-time affine nonlinear systems, each of the iterative controls stabilizes the nonlinear systems by using the policy iteration algorithms. This is the great merit of the policy iteration algorithms. However, almost all of the discussions on the policy iteration algorithms are for continuous-time control systems.[35] In Ref. [45], the policy iteration algorithm for discrete-time systems was displayed and the stability and convergence properties were discussed. To the best of our knowledge, there is still no discussion focused on the policy iteration algorithms for the optimal tracking controls of discrete-time chaotic systems. This motives our research.

In this paper, inspired by Ref. [45], we develop a policy iteration ADP algorithm for the first time to solve the optimal tracking controls of discrete-time chaotic systems. First, by system transformations, the optimal tracking problem is transformed into an optimal regulation problem. Next, a policy iteration algorithm for discrete-time chaotic systems is described. Then, the convergence and admissibility properties of the developed policy iteration algorithm are presented, which show that the transformed chaotic system can be stabilized under an arbitrary iterative control law and simultaneously guarantee the iterative performance index function to converge to the optimum. By employing neural networks, the developed policy iteration algorithm is implemented. To justify the effectiveness of the developed algorithm, the simulation results using the policy iteration algorithm are compared with the results obtained by the traditional value iteration algorithm.

This paper is organized as follows. In Section 2, we present the problem formulation and preliminaries. In Section 3, the optimal chaotic tracking control scheme based on the policy iteration ADP algorithm is given, where the corresponding convergence and admissibility properties will also be presented. In Section 4, a simulation example is given to demonstrate the effectiveness of the developed optimal tracking control scheme. In Section 5, the conclusion is given.

2. Problem formulation and preliminaries

Consider the following MIMO chaotic dynamic system:

$$\begin{aligned} x_{1,k+1} &= f_1(\boldsymbol{x}_k) + \textstyle\sum_{j=1}^{m} g_{1j} u_{j,k}, \\ &\vdots \\ x_{n,k+1} &= f_n(\boldsymbol{x}_k) + \textstyle\sum_{j=1}^{m} g_{nj} u_{j,k}, \end{aligned} \tag{2}$$

where $\boldsymbol{x}_k = [x_{1,k}, \ldots, x_{n,k}]^{\mathrm{T}} \in \mathbb{R}^n$ is the system state vector which is assumed to be available from measurement. Let $\boldsymbol{u}_k = [u_{1,k}, \ldots, u_{m,k}]^{\mathrm{T}} \in \mathbb{R}^m$ be the control input and g_{ij} ($i = 1, \ldots, n$, $j = 1, \ldots, m$) be the constant control gain. If we denote $\boldsymbol{f}(\boldsymbol{x}_k) = [f_1(\boldsymbol{x}_k), \ldots, f_n(\boldsymbol{x}_k)]^{\mathrm{T}}$ and $\boldsymbol{g} = \begin{bmatrix} g_{11} & \cdots & g_{1m} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nm} \end{bmatrix}$, then the chaotic dynamic system (2) can be rewritten as

$$\boldsymbol{x}_{k+1} = F(\boldsymbol{x}_k, \boldsymbol{u}_k) = \boldsymbol{f}(\boldsymbol{x}_k) + \boldsymbol{g}\boldsymbol{u}_k. \tag{3}$$

It should be pointed out that system (3) denotes a large class of chaotic systems, such as the Hénon systems[46] and the new discrete chaotic system proposed in Ref. [47]. In an optimal tracking problem, the control objective is to design optimal control $\boldsymbol{u}(\boldsymbol{x}_k)$ for system (2) such that the system state $\boldsymbol{x}_k$ tracks the specified desired trajectory $\boldsymbol{\eta}_k \in \mathbb{R}^n$, $k = 0, 1, \ldots$, where $\boldsymbol{\eta}_k$ satisfies

$$\boldsymbol{\eta}_{k+1} = \boldsymbol{\sigma}(\boldsymbol{\eta}_k), \tag{4}$$

and $\boldsymbol{\sigma}$ is a given function. We define the tracking error as

$$\boldsymbol{z}_k = \boldsymbol{x}_k - \boldsymbol{\eta}_k, \tag{5}$$

and define the following quadratic performance index:

$$J(\boldsymbol{z}_0, \underline{\boldsymbol{u}}_0) = \sum_{k=0}^{\infty} \left\{ \boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{z}_k + (\boldsymbol{u}_k - \boldsymbol{u}_{\mathrm{e},k})^{\mathrm{T}} \boldsymbol{R} (\boldsymbol{u}_k - \boldsymbol{u}_{\mathrm{e},k}) \right\}, \tag{6}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{R} \in \mathbb{R}^{m \times m}$ are positive definite matrices and $\underline{\boldsymbol{u}}_0 = (\boldsymbol{u}_0, \boldsymbol{u}_1, \ldots)$. Let

$$U(\boldsymbol{z}_k, \boldsymbol{w}_k) = \boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{z}_k + \boldsymbol{w}_k^{\mathrm{T}} \boldsymbol{R} \boldsymbol{w}_k \tag{7}$$

be the utility function, where $\boldsymbol{w}_k = \boldsymbol{u}_k - \boldsymbol{u}_{\mathrm{e},k}$. Let $\boldsymbol{u}_{\mathrm{e},k}$ denote the expected control introduced for analytical purpose, which can be given as

$$\boldsymbol{u}_{\mathrm{e},k} = \boldsymbol{g}^{-1}(\boldsymbol{\eta}_k)(\boldsymbol{\sigma}(\boldsymbol{\eta}_k) - \boldsymbol{f}(\boldsymbol{\eta}_k)), \tag{8}$$

where $\boldsymbol{g}^{-1}(\boldsymbol{\eta}_k)\boldsymbol{g}(\boldsymbol{\eta}_k) = I$ and $I \in \mathbb{R}^{m \times m}$ is the identity matrix. Combining Eqs. (2) and (3), we obtain

$$\begin{aligned} \boldsymbol{z}_{k+1} &= \mathcal{F}(\boldsymbol{z}_k, \boldsymbol{w}_k) \\ &= f(\boldsymbol{z}_k + \boldsymbol{\eta}_k) - \boldsymbol{\sigma}(\boldsymbol{\eta}_k) + \boldsymbol{g}(\boldsymbol{z}_k + \boldsymbol{\eta}_k) \\ &\quad \times \left(\boldsymbol{w}_k + \boldsymbol{g}^{-1}(\boldsymbol{\eta}_k)(\boldsymbol{f}(\boldsymbol{\eta}_k) - \boldsymbol{\sigma}(\boldsymbol{\eta}_k)) \right). \end{aligned} \tag{9}$$

We will study optimal tracking control problems for system (2). The goal is to find an optimal tracking control scheme which tracks the desired trajectory $\boldsymbol{\eta}_k$ and simultaneously minimizes the performance index function (6). The optimal performance index function is defined as

$$J^*(z_k) = \inf_{\underline{\boldsymbol{w}}_k} \{ J(\boldsymbol{z}_k, \underline{\boldsymbol{w}}_k) \}, \tag{10}$$

where $\underline{\boldsymbol{w}}_k = (\boldsymbol{w}_k, \boldsymbol{w}_{k+1}, \ldots)$. According to Bellman's principle of optimality, $J^*(\boldsymbol{z}_k)$ satisfies the discrete-time HJB equation

$$J^*(\boldsymbol{z}_k) = \inf_{\boldsymbol{w}_k} \{ U(\boldsymbol{z}_k, \boldsymbol{w}_k) + J^*(\mathcal{F}(\boldsymbol{z}_k, \boldsymbol{w}_k)) \}. \tag{11}$$

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

$$\boldsymbol{w}^*(\boldsymbol{z}_k) = \arg\inf_{\boldsymbol{w}_k} \{ U(\boldsymbol{z}_k, \boldsymbol{w}_k) + J^*(\boldsymbol{z}_{k+1}) \}. \tag{12}$$

Hence, the HJB equation (11) can be written as

$$J^*(z_k) = U(z_k, w^*(z_k)) + J^*(z_{k+1}). \tag{13}$$

From Eqs. (4)–(13), we can see that the optimal tracking problem for chaotic system (3) is transformed into an optimal regulation problem for chaotic system (9). The objective of this paper is to construct an optimal tracking controller such that the chaotic system state x_k tracks the reference signal η_k, i.e., $z_k \to 0$, as $k \to 0$. For convenience of analysis, the results of this paper are based on the following assumptions.

Assumption 1 The system (9) is controllable and the function $\mathcal{F}(z_k, w_k)$ is Lipschitz continuous for $\forall z_k, w_k$.

Assumption 2 The system state $z_k = 0$ is an equilibrium state of system (9) under the control $v = 0$, i.e., $F(0,0) = 0$.

Assumption 3 The feed-back control $w_k = w(z_k)$ satisfies $w(z_k) = 0$ for $z_k = 0$.

Assumption 4 The utility function $U(z_k, w_k)$ is a continuous positive definite function of z_k and w_k.

Generally, $J^*(z_k)$ is unknown before the control error $w_k \in \mathbb{R}^m$ is considered. If we adopt the traditional dynamic programming method to obtain the optimal performance index function, then we have to face "the curse of dimensionality". This makes the optimal control nearly impossible to obtain by the HJB equation. So, in the next section, an effective policy iteration ADP algorithm will be developed to obtain the optimal chaotic tracking control iteratively.

3. Optimal chaotic tracking control scheme based on policy iteration ADP algorithm

In this section, a new discrete-time policy iteration algorithm is developed to obtain the optimal tracking control law for chaotic systems (2). The goal of the present iterative ADP algorithm is to construct an optimal control law $w^*(z_k)$, $k = 0, 1, \ldots$, which moves an arbitrary initial state x_0 to the desired trajectory η_k, and simultaneously minimizes the performance index function. Convergence and admissibility properties will be analyzed.

3.1. The policy iteration algorithm

For optimal control problems, the developed control scheme must not only stabilize the control systems but also make the performance index function finite, i.e., the control law must be admissible.[43] We can define Ω_w as the set of the admissible control laws, which is expressed as

$$\Omega_w = \left\{ u(z_k) \middle| \lim_{k\to\infty} F(z_k, w(z_k)) = 0,\ u(0) = 0,\ \text{and} \right.$$
$$\left. \sum_{j=0}^{\infty} U(z_{k+j}, w(z_{k+j})) < \infty \right\}. \tag{14}$$

Based on the above definition, we can introduce our policy iteration algorithm. Let i be the iteration index that increases from 0 to infinity. Let $w^{[0]}(z_k) \in \Omega_w$ be an arbitrary admissible control law. For $i = 0$, let $J^{[0]}(z_k)$ be the iterative performance index function constructed by $w^{[0]}(z_k)$, which satisfies the following generalized HJB (GHJB) equation:

$$J^{[0]}(z_k) = U(z_k, w^{[0]}(z_k)) + J^{[0]}(\mathcal{F}(z_k, w^{[0]}(z_k))). \tag{15}$$

For $\forall i = 1, 2, \ldots$, the developed policy iteration algorithm will iterate between

$$w^{[i]}(z_k) = \arg\min_{w_k} \left\{ U(z_k, w_k) + J^{[i-1]}(z_{k+1}) \right\}$$
$$= \arg\min_{w_k} \left\{ U(z_k, w_k) + J^{[i-1]}(\mathcal{F}(z_k, w_k)) \right\}, \tag{16}$$

and

$$J^{[i]}(z_k) = U(z_k, w^{[i]}(z_k)) + J^{[i]}(\mathcal{F}(z_k, w^{[i]}(z_k))). \tag{17}$$

From Eqs. (15)–(17), we can see that the procedure of policy iteration is obviously different from value iteration (1). Thus, the convergence analysis for the value iteration algorithms in Refs. [36] and [43] is invalid for the policy iteration. Therefore, a new analysis will be given in this paper.

3.2. Properties of the algorithm

In Ref. [12], the convergence and admissibility properties of the continuous-time policy iteration algorithm are discussed. This shows that any of the iterative control laws can stabilize the transformed chaotic system. For the discrete-time systems, the proofs of the continuous-time policy iteration algorithms are invalid. Hence, a new analysis will be developed. We will show that for the discrete-time policy iteration algorithm, any of the iterative control laws can also stabilize the transformed system.

Theorem 1 For $i = 0, 1, \ldots$, let $J^{[i]}(z_k)$ and $w^{[i]}(z_k)$ be obtained by Eqs. (15)–(17), where $w^{[0]}(z_k) \in \Omega_w$ is an arbitrary admissible control law. Then, for $\forall i = 0, 1, \ldots$, the iterative control law $w^{[i]}(z_k)$ stabilizes the chaotic system (9).

Proof According to Eq. (17), we have

$$J^{[i]}(z_k) = \sum_{t=0}^{\infty} U(z_{k+t}, w^{[i]}(z_{k+t})). \tag{18}$$

Let $z_k = 0$, according to Assumption 3, we have $w^{[i]}(z_k) = 0$. According to Assumption 2, we also have $z_{k+1} = \mathcal{F}(z_k, w^{[i]}(z_k)) = 0$ and furthermore $w^{[i]}(z_{k+1}) = 0$. Then according to the mathematical induction and Assumption 4, we have $\forall U(z_{k+j}, w^{[i]}(z_{k+j})) = 0$, $j = 0, 1, \ldots$, which shows $J^{[i]}(z_k) = 0$ for $z_k = 0$.

On the other hand, according to Assumption 4, the utility function $U(z_k, w_k)$ is positive definite for $\forall z_k, w_k$. Then we can get $U(z_k, w_k) \to \infty$ as $z_k \to \infty$, which means $J^{[i]}(z_k) \to \infty$. Hence, $J^{[i]}(z_k)$ is positive definite for $\forall i = 0, 1, \ldots$.

According to Eqs. (15) and (17), for $\forall i = 0, 1, \ldots$,

$$J^{[i]}(z_{k+1}) - J^{[i]}(z_k) = -U(z_k, w^{[i]}(z_k)) < 0 \tag{19}$$

holds. Then for $\forall i = 0, 1, \ldots$, $J^{[i]}(z_k)$ is a Lyapunov function. Thus $w^{[i]}(z_k)$ is a stable control law. The proof is completed.

Theorem 2 For $i = 0, 1, \ldots$, let $J^{[i]}(z_k)$ and $w^{[i]}(z_k)$ be obtained by Eqs. (15)–(17). For $\forall z_k \in \mathbb{R}^n$, the iterative performance index function $J^{[i]}(z_k)$ is a monotonically non-increasing sequence for $\forall i \geq 0$, i.e.,

$$J^{[i+1]}(z_k) \leq J^{[i]}(z_k). \tag{20}$$

Proof For $i = 0, 1, \ldots$, define a new performance index function $\Upsilon^{[i+1]}(z_k)$ as

$$\Upsilon^{[i+1]}(z_k) = U(z_k, w^{[i+1]}(z_k)) + J^{[i]}(\mathcal{F}(z_k, w^{[i+1]}(z_k))), \tag{21}$$

where $w^{[i+1]}(z_k)$ is obtained by Eq. (16). According to Eq. (21), for $\forall z_k$, we can obtain

$$\begin{aligned} \Upsilon^{[i+1]}(z_k) &= U(z_k, w^{[i+1]}(z_k)) + J^{[i]}(z_{k+1}) \\ &= \min_{w_k} \left\{ U(z_k, w_k) + J^{[i]}(z_{k+1}) \right\} \\ &\leq U(z_k, w^{[i]}(z_k)) + J^{[i]}(z_{k+1}) \\ &= J^{[i]}(z_k). \end{aligned} \tag{22}$$

The inequality (20) will be proven by the mathematical induction. According to Theorem 1, for $i = 0, 1, \ldots$, $w^{[i+1]}(z_k)$ is a stable control law. Then, $z_k \to 0$, for $\forall k \to \infty$. Without loss of generality, let $z_N = 0$, where $N \to \infty$. We can obtain

$$J^{[i+1]}(z_N) = \Upsilon^{[i+1]}(z_N) = J^{[i]}(z_N) = 0. \tag{23}$$

First, we let $k = N - 1$. According to Eq. (16), we have

$$\begin{aligned} & w^{[i+1]}(z_{N-1}) \\ &= \arg\min_{w_{N-1}} \left\{ U(z_{N-1}, w_{N-1}) + J^{[i]}(z_N) \right\} \\ &= \arg\min_{w_{N-1}} \left\{ U(z_{N-1}, w_{N-1}) + J^{[i+1]}(z_N) \right\}. \end{aligned} \tag{24}$$

According to Eqs. (17) and (24), we can obtain

$$\begin{aligned} J^{[i+1]}(z_{N-1}) &= U(z_{N-1}, w^{[i+1]}(z_{N-1})) + J^{[i+1]}(z_N) \\ &= \min_{w_{N-1}} \left\{ U(z_{N-1}, w_{N-1}) + J^{[i+1]}(z_N) \right\} \\ &= \min_{w_{N-1}} \left\{ U(z_{N-1}, w_{N-1}) + J^{[i]}(z_N) \right\} \\ &\leq U(z_{N-1}, w^{[i]}(z_{N-1})) + J^{[i]}(z_N) \\ &= J^{[i]}(z_{N-1}). \end{aligned} \tag{25}$$

So, the conclusion holds for $k = N - 1$. Assume that the conclusion holds for $k = l + 1$, $l = 0, 1, \ldots$. For $k = l$, we have

$$\begin{aligned} J^{[i+1]}(z_l) &= U(z_l, w^{[i+1]}(z_l)) + J^{[i+1]}(z_{l+1}) \\ &\leq U(z_l, w^{[i+1]}(z_l)) + J^{[i]}(z_{l+1}) \\ &= \Upsilon^{[i+1]}(z_l). \end{aligned} \tag{26}$$

According to Eq. (22), for $\forall z_l$, we have

$$\Upsilon^{[i+1]}(z_l) \leq J^{[i]}(z_l). \tag{27}$$

According to inequalities (26) and (27), for $i = 0, 1, \ldots$, inequality (20) holds for $\forall z_k$. The proof is completed.

Corollary 1 For $i = 0, 1, \ldots$, let $J^{[i]}(z_k)$ and $w^{[i]}(z_k)$ be obtained by Eqs. (15)–(17), where $w^{[0]}(z_k)$ is an arbitrary admissible control law. Then for $\forall i = 0, 1, \ldots$, the iterative control law $w^{[i]}(z_k)$ is admissible.

From Theorem 2, we can see that the iterative performance index function $J^{[i]}(z_k)$ is monotonically non-increasing and lower bounded. Hence as $i \to \infty$, the limit of $J^{[i]}(z_k)$ exists, i.e.,

$$J^{[\infty]}(z_k) = \lim_{i \to \infty} J^{[i]}(z_k). \tag{28}$$

We can then derive the following theorem.

Theorem 3 For $i = 0, 1, \ldots$, let $J^{[i]}(z_k)$ and $w^{[i]}(z_k)$ be obtained by Eqs. (15)–(17). The iterative performance index function $J^{[i]}(z_k)$ then converges to the optimal performance index function $J^*(z_k)$ as $i \to \infty$, i.e.,

$$\lim_{i \to \infty} J^{[i]}(z_k) = J^*(z_k), \tag{29}$$

which satisfies the HJB equation (11).

Proof The theorem can be proven in two steps. Let $\mu(z_k) \in \Omega_w$ be an arbitrary admissible control law. Define a new performance index function $P(z_k)$, which satisfies

$$P(z_k) = U(z_k, \mu(z_k)) + P(z_{k+1}). \tag{30}$$

For $i = 0, 1, \ldots$, let $J^{[i]}(z_k)$ and $w^{[i]}(z_k)$ be obtained by Eqs. (15) and (16). We will then prove

$$J^{[\infty]}(z_k) \leq P(z_k). \tag{31}$$

The statement (31) can be proven by a mathematical induction. As $\mu(z_k)$ is an admissible control law, we have $z_k \to 0$ as $k \to \infty$. Without loss of generality, let $z_N = 0$, where $N \to \infty$. According to Eq. (30), we have

$$\begin{aligned} P(z_k) = \lim_{N \to \infty} \{ & U(z_k, \mu(z_k)) + U(z_{k+1}, \mu(z_{k+1})) + \cdots \\ & + U(z_{N-1}, \mu(z_{N-1})) + P(z_N) \}, \end{aligned} \tag{32}$$

where $z_N = 0$. According to Eq. (28), the iterative performance index function $J^{[\infty]}(z_k)$ can be expressed as

$$\begin{aligned} & J^{[\infty]}(z_k) \\ &= \lim_{N \to \infty} \left\{ U(z_k, w^{[\infty]}(z_k)) + U(z_{k+1}, w^{[\infty]}(z_{k+1})) \right. \\ &\quad \left. + \cdots + U(z_{N-1}, w^{[\infty]}(z_{N-1})) + J^{[\infty]}(z_N) \right\} \\ &= \lim_{N \to \infty} \left\{ \min_{w_k} \left\{ U(z_k, w_k) + \min_{w_{k+1}} \left\{ U(z_{k+1}, w_{k+1}) \right. \right. \right. \\ &\quad \left. \left. \left. + \cdots + \min_{w_{N-1}} \left\{ U(z_{N-1}, w_{N-1}) + J^{[\infty]}(z_N) \right\} \right\} \right\} \right\}. \end{aligned} \tag{33}$$

Since $\boldsymbol{w}^{[\infty]}(z_k)$ is an admissible control law, we can obtain $z_N = 0$ when $N \to \infty$, which means $J^{[\infty]}(z_N) = P(z_N) = 0$ as $N \to \infty$. For $N-1$, according to Eq. (43), we can obtain

$$\begin{aligned} P(z_{N-1}) &= U(z_{N-1}, \boldsymbol{\mu}(z_{N-1})) + P(z_N) \\ &\geq \min_{\boldsymbol{w}_{N-1}} \{U(z_{N-1}, \boldsymbol{w}_{N-1}) + P(z_N)\} \\ &= \min_{\boldsymbol{w}_{N-1}} \left\{U(z_{N-1}, \boldsymbol{w}_{N-1}) + J^{[\infty]}(z_N)\right\} \\ &= J^{[\infty]}(z_{N-1}). \end{aligned} \tag{34}$$

Assume that the statement holds for $k = l+1, l = 0, 1, \ldots$, then for $k = l$, we have

$$\begin{aligned} P(z_l) &= U(z_l, \boldsymbol{\mu}(z_l)) + P(z_{l+1}) \\ &\geq \min_{\boldsymbol{w}_l} \{U(z_l, \boldsymbol{w}_l) + P(z_{l+1})\} \\ &\geq \min_{\boldsymbol{w}_l} \left\{U(z_l, \boldsymbol{w}_l) + J^{[\infty]}(z_{l+1})\right\} \\ &= J^{[\infty]}(z_l). \end{aligned} \tag{35}$$

Hence for $\forall z_k$, $k = 0, 1, \ldots$, the inequality (31) holds. The mathematical induction is completed.

Next, we will prove that the iterative performance index function $J^{[i]}(z_k)$ converges to the optimal performance index function $J^*(z_k)$ as $i \to \infty$, i.e.,

$$\lim_{i \to \infty} J^{[i]}(z_k) = J^*(z_k), \tag{36}$$

which satisfies the HJB equation (11).

According to the definition of $\Upsilon^{[i+1]}(z_k)$ in Eq. (21), we have

$$\begin{aligned} \Upsilon^{[i+1]}(z_k) &= U(z_k, \boldsymbol{w}^{[i+1]}(z_k)) + J^{[i]}(z_{k+1}) \\ &= \min_{\boldsymbol{w}_k} \left\{U(z_k, \boldsymbol{w}_k) + J^{[i]}(z_{k+1})\right\}. \end{aligned} \tag{37}$$

According to Eq. (26), we obtain

$$J^{[i+1]}(z_k) \leq \Upsilon^{[i+1]}(z_k) \tag{38}$$

for $\forall z_k$. Let $i \to \infty$. We obtain

$$\begin{aligned} J^{[\infty]}(z_k) &= \lim_{i \to \infty} J^{[i]}(z_k) \leq J^{[i+1]}(z_k) \leq \Upsilon^{[i+1]}(z_k) \\ &\leq \min_{\boldsymbol{w}_k} \left\{U(z_k, \boldsymbol{w}_k) + J^{[i]}(z_{k+1})\right\}. \end{aligned} \tag{39}$$

Thus we can obtain

$$J^{[\infty]}(z_k) \leq \min_{\boldsymbol{w}_k} \{U(z_k, \boldsymbol{w}_k) + J^{[\infty]}(z_{k+1})\}. \tag{40}$$

Let $\varepsilon > 0$ be an arbitrary positive number. Since $J^{[i]}(z_k)$ is non-increasing for $i \geq 1$ and $\lim_{i \to \infty} J^{[i]}(z_k) = J^{[\infty]}(z_k)$, there exists a positive integer p such that

$$J^{[p]}(z_k) - \varepsilon \leq J^{[\infty]}(z_k) \leq J^{[p]}(z_k). \tag{41}$$

Hence, we can obtain

$$\begin{aligned} J^{[\infty]}(z_k) &\geq U(z_k, \boldsymbol{w}^{[p]}(z_k)) + J^{[p]}(\mathcal{F}(z_k, \boldsymbol{w}^{[p]}(z_k))) - \varepsilon \\ &\geq U(z_k, \boldsymbol{w}^{[p]}(z_k)) + J^{[\infty]}(\mathcal{F}(z_k, \boldsymbol{w}^{[p]}(z_k))) - \varepsilon \\ &\geq \min_{\boldsymbol{w}_k} \{U(z_k, \boldsymbol{w}_k) + J^{[\infty]}(z_{k+1})\} - \varepsilon. \end{aligned}$$

Since ε is arbitrary, we have

$$J^{[\infty]}(z_k) \geq \min_{\boldsymbol{w}_k} \{U(z_k, \boldsymbol{w}_k) + J^{[\infty]}(z_{k+1})\}. \tag{42}$$

Combining Eqs. (40) and (42), we can obtain

$$J^{[\infty]}(z_k) = \min_{\boldsymbol{w}_k} \{U(z_k, \boldsymbol{w}_k) + J^{[\infty]}(z_{k+1})\}. \tag{43}$$

According to the definition of $J^*(z_k)$ in Eq. (40), for $\forall i = 0, 1, \ldots$, we have

$$J^{[i]}(z_k) \geq J^*(z_k). \tag{44}$$

Let $i \to \infty$, we can then obtain

$$J^{[\infty]}(z_k) \geq J^*(z_k). \tag{45}$$

On the other hand, as $\mu(z_k) \in \Omega_w$ is arbitrary, if we let $\boldsymbol{w}^{[i]}(z_k) = w^*(z_k)$, then according to inequality (31), we have

$$J^{[\infty]}(z_k) \leq J^*(z_k). \tag{46}$$

According to inequalities (45) and (46), we can obtain Eq. (36). The proof is completed.

Corollary 2 Let $z_k \in \mathbb{R}^n$ be an arbitrary state vector. If Theorem 3 holds, then the iterative control law $\boldsymbol{w}^{[i]}(z_k)$ converges to the optimal control law as $i \to \infty$, i.e., $\boldsymbol{w}^*(z_k) = \lim_{i \to \infty} \boldsymbol{w}^{[i]}(z_k)$.

3.3. Constructing the iterative performance index function

In the policy iteration algorithm (15)–(17), for $\forall i = 0, 1, \ldots$, we should construct an iterative performance index function $J^{[i]}(z_k)$ to satisfy Eq. (17). In this subsection, we will give an effective method to construct the iterative performance index function.

Let $\Psi(z_k)$ be a positive semi-definite function. Introducing a new iteration index $j = 0, 1, \ldots$. Define a new performance index function $J^{[i]}_{[j]}(z_k)$. For $i = 0$ and $j = 0$, let $J^{[0]}_{[0]}(z_k) = \Psi(z_k)$. For $i = 0$ and $j = 0, 1, \ldots$, let

$$J^{[0]}_{[j+1]}(z_k) = U(z_k, \boldsymbol{w}^{[0]}(z_k)) + J^{[0]}_{[j]}(z_{k+1}). \tag{47}$$

For $i = 1, 2, \ldots$ and $j = 0, 1 \ldots$, let

$$J^{[i]}_{[j+1]}(z_k) = U(z_k, \boldsymbol{w}^{[i]}(z_k)) + J^{[i]}_{[j]}(z_{k+1}), \tag{48}$$

where $J^{[i]}_0(z_k) = \lim_{j \to \infty} J^{[i-1]}_{[j]}(z_k)$. Then we can derive the following theorem.

Theorem 4 Let $\Psi(z_k) \geq 0$ be an arbitrary semi-positive definite function. Let $J^{[i]}_{[j]}(z_k)$, $i = 0, 1, \ldots$, $j = 0, 1 \ldots$ be the

iterative performance index function updated by Eqs. (47) and (48), respectively. Then, for $\forall i = 0, 1, \ldots$, we have

$$J^{[i]}(z_k) = \lim_{j\to\infty} J^{[i]}_{[j]}(z_k), \tag{49}$$

where $J^{[i]}(z_k)$ satisfies Eqs. (15) and (17), respectively.

Proof According to Eq. (47), we have

$$\begin{cases} J^{[0]}_{[j+1]}(z_k) - J^{[0]}_{[j]}(z_{k+1}) = U(z_k, w^{[0]}(z_k)), \\ J^{[0]}_{[j]}(z_{k+1}) - J^{[0]}_{[j-1]}(z_{k+2}) = U(z_{k+1}, w^{[0]}(z_{k+1})), \\ \vdots \\ J^{[0]}_{[1]}(z_{k+i}) - J^{[0]}_{[0]}(z_{k+i+1}) = U(z_{k+i}, w^{[0]}(z_{k+i})). \end{cases} \tag{50}$$

We can then obtain

$$J^{[0]}_{[j+1]}(z_k) = \sum_{l=0}^{j} U(z_{k+l}, w^{[0]}(z_{k+l})) + \Psi(z_{k+i+1}). \tag{51}$$

Let $j \to \infty$. We can obtain

$$\begin{aligned} \lim_{j\to\infty} J^{[0]}_{[j+1]}(z_k) &= \sum_{l=0}^{\infty} U(z_{k+l}, w^{[0]}(z_{k+l})) \\ &\quad + \lim_{j\to\infty} \Psi(z_{k+j+1}). \end{aligned} \tag{52}$$

Since $w^{[0]}(z_k)$ is an admissible control law, $\sum_{j=0}^{\infty} U(z_{k+j}, w^{[0]}(z_{k+j}))$ is finite. For $j \to \infty$, we have $z_{k+j+1} \to 0$ and hence $\Psi(z_{k+i+1}) = 0$. As for an arbitrary finite z_k, $\Psi(z_{k+j+1})$ is finite, then for $\forall j = 0, 1, 2, \ldots$, $J^{[0]}_{[j+1]}(z_k)$ is finite. Hence $\lim_{j\to\infty} J^{[0]}_{[j+1]}(z_k)$ is finite, which means that Equation (49) holds for $i = 0$. According to the mathematical induction, Equation (49) holds for $\forall i = 0, 1, \ldots$. The proof is completed.

4. Simulation

To evaluate the performance of our policy iteration algorithm, we choose an example with quadratic utility functions for numerical experiments. Consider the following chaotic system:[4]

$$\begin{aligned} \dot{x}_1 &= \theta_1(x_2 - h(x_1)) + u_1, \\ \dot{x}_2 &= x_1 - x_2 + x_3 + u_2, \\ \dot{x}_3 &= -\theta_2 x_2 + u_3, \end{aligned} \tag{53}$$

where $h(x_1) = m_1 x_1 + (m_0 - m_1)(|x_1 + \theta_3| - |x_1 - \theta_3|)/2$, $\theta_1 = 9$, $\theta_2 = 14.28$, $\theta_3 = 1$, $m_0 = -1/7$, and $m_1 = 2/7$. The state trajectory of the chaotic system is given in Fig. 1. According to Euler's discretization method, the continuous time chaotic system can be represented as follows:

$$\begin{aligned} x_{1,k+1} &= x_{1,k} + \Delta T(\theta_1(x_{2,k} - h(x_{1,k})) + u_{1,k}), \\ x_{2,k+1} &= x_{2,k} + \Delta T(x_{1,k} - x_{2,k} + x_{3,k} + u_{2,k}), \\ x_{3,k+1} &= x_{3,k} + \Delta T(-\theta_2 x_{2,k} + u_{3,k}), \end{aligned} \tag{54}$$

where $\Delta T = 0.1$. Let the desired orbit be $\eta_k = [\sin(k), 0.5\cos(k), 0.6\sin(k)]^{\mathrm{T}}$ and the initial state be selected as $[1, -1, 1.5]^{\mathrm{T}}$. The utility function is defined as $U(z_k, w_k) = z_k^{\mathrm{T}} Q z_k + w_k^{\mathrm{T}} R w_k$, where $Q = R = I_3$. Neural networks are used to implement the developed policy iteration algorithm.

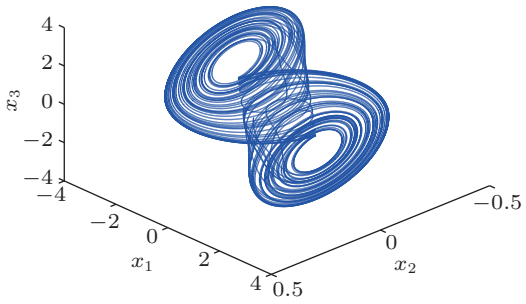


Fig. 1. (color online) The chaotic attractor of system (54).

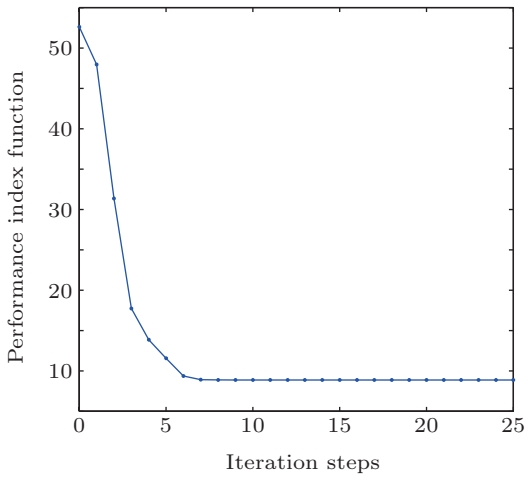


Fig. 2. (color online) The performance index function trajectory in the policy iteration algorithm.

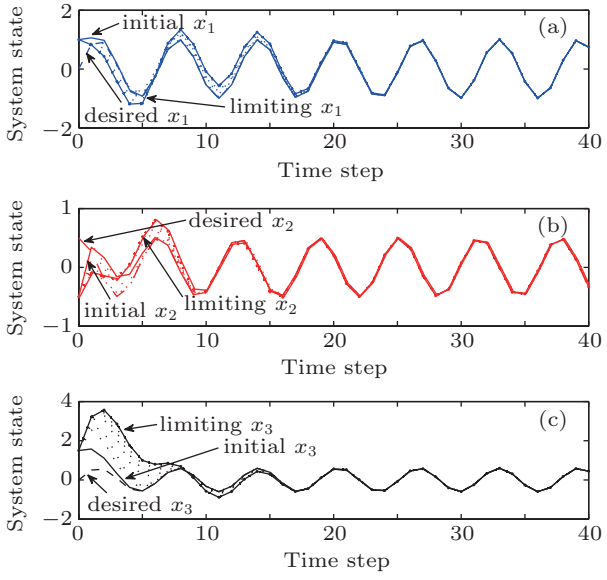


Fig. 3. (color online) The trajectories of system states x_1, x_2, and x_3 in the policy iteration algorithm.

The critic network and the action network are chosen as three-layer BP neural networks with the structures of 3–10–1 and 3–10–3, respectively. For each iteration step, the critic network and the action network are trained for 1000 steps using a learning rate of $\alpha = 0.02$ so that the neural network training error becomes less than 10^{-5}. The training methods of the neural networks have been described in Refs. [30] and [45], which are omitted here. The performance index function is shown in Fig. 2. The system states are shown in Fig. 3. The state errors are given in Fig. 4. The iterative controls are shown in Fig. 5. In the simulation results, we let "initial" and "limiting" denote the initial iteration and the limiting iteration, respectively.

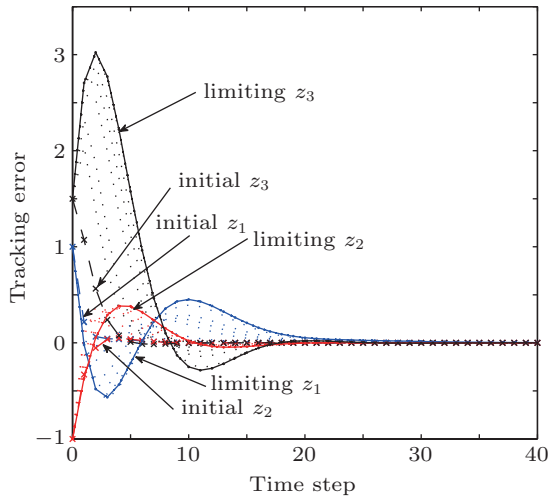


Fig. 4. (color online) The trajectories of tracking errors z_1, z_2, and z_3 in the policy iteration algorithm.

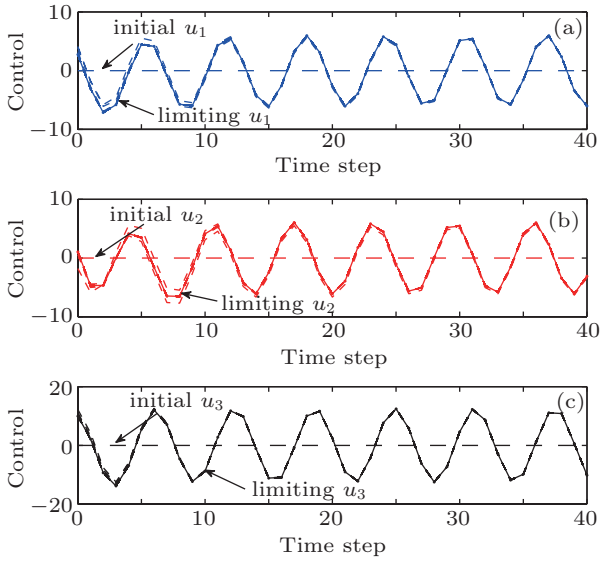


Fig. 5. (color online) The trajectories of controls u_1, u_2, and u_3 in the policy iteration algorithm.

From the simulation results, we can see that the iterative performance index function $J^{[i]}(\boldsymbol{z}_k)$ is monotonically non-increasing and converges to the optimal performance index function. For $\forall i = 0, 1, \ldots$, the regulation system (9) is stable under the iterative control law $\boldsymbol{w}^{[i]}(\boldsymbol{z}_k)$. To show the

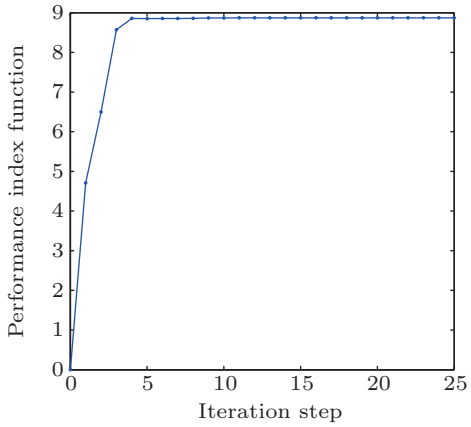


Fig. 6. (color online) The performance index function trajectory in the value iteration algorithm.

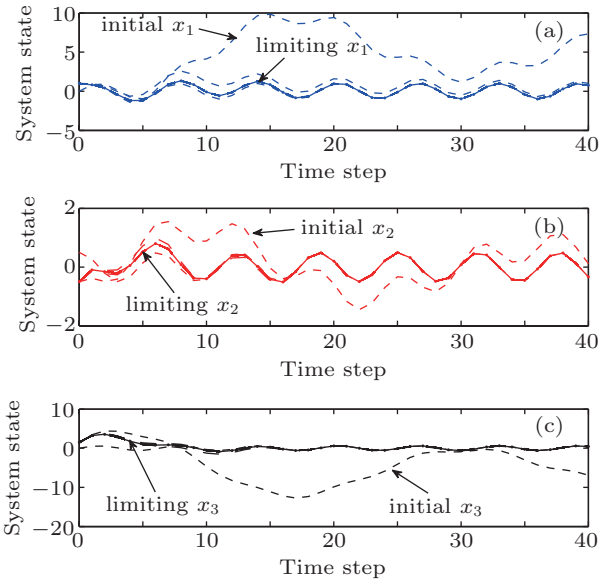


Fig. 7. (color online) The trajectories of system states x_1, x_2, and x_3 in the value iteration algorithm.

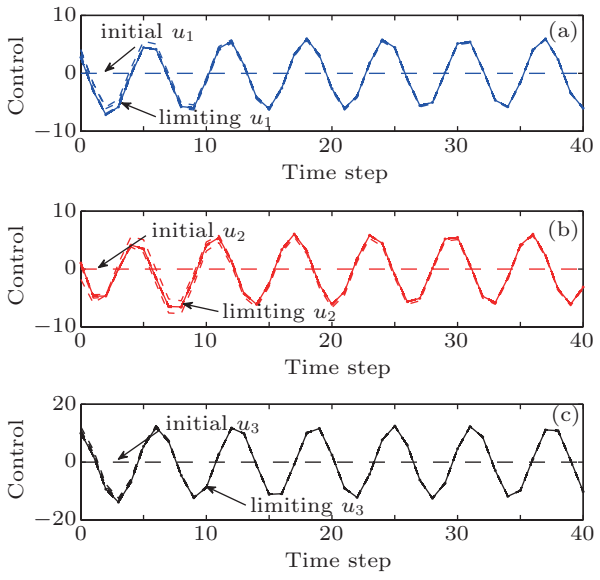


Fig. 8. (color online) The trajectories of controls u_1, u_2, and u_3 in the value iteration algorithm.

effectiveness of the developed algorithm, the results of our developed policy iteration algorithm are compared with the results using a value iteration algorithm.[36] Let the value iteration algorithm be implemented by Eq. (1), where the initial performance index function $J^{[0]}(z_k) \equiv 0$. All of the parameters are set the same as those for implementing the polity iteration algorithm. The performance index function is shown in Fig. 6. The system states are shown in Fig. 7. The iterative controls are shown in Fig. 8.

From Fig. 6, we can see that the iterative performance index function is monotonically non-decreasing and converges to the optimum. The convergence properties of value and policy iteration algorithms are inherently different. On the other hand, from Figs. 7 and 8, we can see that there exit unstable system states and control laws in the value iteration algorithm. From Figs. 3 and 5, we can see that the chaotic system (54) can track the desired trajectories with the policy iteration algorithm, which illustrates the effectiveness of the developed algorithm.

5. Conclusion

We have developed an optimal tracking control method for chaotic systems by policy iteration. By system transformations, the optimal tracking problem is transformed into an optimal regulation one. Next, the policy iteration algorithm for the transformed chaotic systems are presented. The convergence and admissibility properties of the developed policy iteration algorithm are presented, which show that the transformed chaotic system can be stabilized under the present iterative control law and the iterative performance index function simultaneously converges to the optimum. Finally, the effectiveness of the developed optimal tracking control scheme for chaotic systems is verified by a simulation.

References

[1] Liu H, Yu H and Xiang W 2012 *Chin. Phys. B* **21** 120505
[2] Ma T and Fu J 2011 *Chin. Phys. B* **20** 050511
[3] Ma T, Fu J and Sun Y 2010 *Chin. Phys. B* **19** 090502
[4] Ma T, Zhang H anf Fu J 2008 *Chin. Phys. B* **17** 4407
[5] Yu X, Ren Z and Xu C 2014 *Chin. Phys. B* **23** 040201
[6] Zhang H, Ma T, Fu J and Tong S 2009 *Chin. Phys. B* **18** 3751
[7] Zheng Z, Zheng P and Ge H 2014 *Chin. Phys. B* **23** 020503
[8] Chen S and Lu J 2002 *Chaos Soliton. Fract.* **14** 643
[9] Zhang H, Huang W, Wang Z and Chai T 2006 *Phys. Lett. A* **350** 363
[10] Zhang H, Wang Z and Liu D 2004 *Int. J. Bifur. Chaos* **14** 3505
[11] Li X, Yang K and Wang Y 2011 *Chin. Phys. B* **20** 074301
[12] Murray J, Cox C, Lendaris G and Saeks R 2002 *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* **32** 140
[13] Wei Q and Liu D 2012 *Neural Networks* **32** 236
[14] Wei Q, Wang D and Zhang D 2013 *Neural Computing & Applications* **23** 1851
[15] Liu D and Wei Q 2013 *IEEE Transactions on Cybernetics* **43** 779
[16] Kiumarsi B, Lewis F, Modares H, Karimpour A and Naghibi-Sistani MB 2014 *Automatica* **50** 1167
[17] Werbos P 1977 *General Systems Yearbook* **22** 25
[18] Werbos P 1991 *Neural Networks for Control* (Miller W, Sutton R and Werbos P, Ed) (Boston: MIT Press) p. 67
[19] Heydari A and Balakrishnan S 2013 *IEEE Transactions on Neural Networks and Learning Systems* **24** 145
[20] Wei Q and Liu D 2013 *IET Control Theory & Applications* **7** 1472
[21] Wei Q and Liu D 2014 *Neural Computing & Applications* **24** 1355
[22] Prokhorov D and Wunsch D 1997 *IEEE Transactions on Neural Networks* **8** 997
[23] Wei Q, Wang F, Liu D and Yang X 2014 *IEEE Transactions on Cybernetics* **44** 2820
[24] Wei Q and Liu D 2015 *Neurocomputing* **149** 106
[25] Wei Q, Zhang H, Liu D and Zhao Y 2010 *Acta Automatica Sinica* **36** 121
[26] Wei Q, Liu D, Shi G and Liu Y 2015 *IEEE Transactions on Industrial Electronics*, accept article in press
[27] Wei Q, Liu D and Xu Y 2015 *Soft Computing*, article in press
[28] Werbos P 1992 *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches* (White D and Sofge D, Ed) (New York: Van Nostrand Reinhold) chapter 13
[29] Enns R and Si J 2003 *IEEE Transactions on Neural Networks* **14** 929
[30] Si J and Wang Y 2001 *IEEE Transactions on Neural Networks* **12** 264
[31] Wei Q, Liu D and Shi G 2015 *IEEE Transactions on Industrial Electronics*, accept article in press
[32] Abu-Khalaf M and Lewis F 2005 *Automatica* **41** 779
[33] Wei Q, Zhang H and Dai J 2009 *Neurocomputing* **72** 1839
[34] Zhang H, Song R, Wei Q and Zhang T 2011 *IEEE Transactions on Neural Networks* **22** 1851
[35] Zhang H, Wei Q and Liu D 2011 *Automatica* **47** 207
[36] Zhang H, Wei Q and Luo Y 2008 *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **38** 937
[37] Song R, Xiao W and Wei Q 2014 *Chin. Phys. B* **23** 050504
[38] Wei Q and Liu D 2014 *IEEE Transactions on Industrial Electronics* **61** 6399
[39] Wei Q and Liu D 2014 *IEEE Transactions on Automation Science and Engineering* **11** 1020
[40] Wei Q and Liu D 2014 *IEEE Transactions on Automation Science and Engineering* **11** 1176
[41] Lewis F, Vrabie D and Vamvoudakis K 2012 *IEEE Control Systems* **32** 76
[42] Lincoln B and Rantzer A 2006 *IEEE Transactions on Automatic Control* **51** 1249
[43] Al-Tamimi A, Lewis F and Abu-Khalaf M 2008 *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics* **38** 943
[44] Song R, Xiao W, Sun C and Wei Q 2013 *Chin. Phys. B* **22** 090502
[45] Liu D and Wei Q 2014 *IEEE Transactions on Neural Networks and Learning Systems* **25** 621
[46] Héno M 1976 *Communications in Mathematical Physics* **50** 69
[47] Lu J, Wu X, Lü J and Kang L 2004 *Chaos Soliton. Fract.* **22** 311