Soft Comput (2014) 18:349–357
DOI 10.1007/s00500-013-1062-2

METHODOLOGIES AND APPLICATION

Approximate optimal solution of the DTHJB equation for a class of nonlinear affine systems with unknown dead-zone constraints

Dehua Zhang · Derong Liu · Ding Wang

Published online: 11 June 2013

Abstract In this paper, an optimal control scheme of a class of unknown discrete-time nonlinear systems with dead-zone control constraints is developed using adaptive dynamic programming (ADP). First, the discrete-time Hamilton–Jacobi–Bellman (DTHJB) equation is derived. Then, an improved iterative ADP algorithm is constructed which can solve the DTHJB equation approximately. Combining with Riemann integral, detailed proofs of existence and uniqueness of the solution are also presented. It is emphasized that this algorithm allows the implementation of optimal control without knowing internal system dynamics. Moreover, the approach removes the requirements of precise parameters of the dead-zone. Finally, simulation studies are given to demonstrate the performance of the present approach using neural networks.

Keywords Nonlinear affine system · Dead-zone · DTHJB · Neural networks · Riemann integral

1 Introduction

Dead-zone, backlash, saturation, hysteresis and actuator nonlinearities are very common in most practical industrial control systems. Because of the nonanalytical nature of these actuator nonlinearities and the fact that their accurate nonlinear mathematical functions are difficult to know, such systems present a challenge for practitioners. Thus, there have been many discussions on this subject. For example, backlash, hysteresis and saturation nonlinearities were considered in Tao and Kokotovic (1995a, b, 1996), Zhang et al. (2009), Bernstein (1995), Saberi et al. (1996) and Sussmann et al. (1994), respectively.

Dead-zone is one of the most important nonlinearities in many industrial processes, which can severely affect the system's performance. The study of control systems with dead-zone nonlinearities has been the focus of researchers for many years. Several solutions for deriving control laws considering the dead-zone phenomena can be found in Tao and Kokotovic (1994, 1995c), Gao and Rastko (2006), Recker et al. (1991), Selmic and Lewis (2000), Lewis et al. (1999), Xu et al. (2005), Wang et al. (2004), Ma and Yang (2010) and Zhang and Ge (2007, 2008). In Xu et al. (2005), affine nonlinear systems with dead-zone input were first investigated. The study of adaptive control for systems with unknown dead-zone was initiated by Recker et al. (1991), where an adaptive scheme was proposed for the case of full state measurement. In recent years, there are some works of handling dead-zone nonlinearities from the perspective of adaptive control (Tao and Kokotovic 1994, 1995c; Wang et al. 2004; Ma and Yang 2010; Zhang and Ge 2007, 2008). In fact the exact solution of the Hamilton–Jacobi–Bellman (HJB) equation is generally impossible to obtain for nonlinear systems especially with dead-zone constraints. To overcome the difficulty, recursive methods are employed to obtain the solution of HJB equation indirectly, such as the iterative adaptive dynamic programming (ADP).

ADP is a very useful tool in solving optimization and optimal control problems by employing the principle of

Communicated by G. Acampora.

D. Zhang · D. Liu (✉) · D. Wang
State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
e-mail: derong.liu@ia.ac.cn

D. Zhang
e-mail: dhuazhang@163.com

D. Wang
e-mail: ding.wang@ia.ac.cn

optimality, which is expressed as "An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision" (Bellman 1957). In recent years, ADP using universal function approximators has received much attention from many researchers in order to obtain approximate solutions of the HJB equation (Abu-Khalaf and Lewis 2005; Balakrishnan et al. 2008; Al-Tamimi et al. 2007; Al-Tamimi and Lewis 2008; Liu et al. 2001, 2005; Wei et al. 2009). Typically, these methods use fuzzy logic systems or neural networks (NNs) to parameterize the unknown nonlinearities.

According to Werbos (1992) and Prokhorov and Wunsch (1997), ADP approaches were classified into several main schemes: heuristic dynamic programming (HDP), action-dependent HDP (ADHDP), dual heuristic dynamic programming (DHP), ADDHP, globalized DHP (GDHP), and ADGDHP. In Seiffertt et al. (2001), HJB equations were motivated and proven on time scales. The authors coupled the calculus of time scales with stochastic control via ADP algorithm and pointed out three significant directions for the investigation of ADP on time scales.

In the present paper, we study this problem through ADP in order to solve the discrete-time HJB (DTHJB) equations and then the optimal control problems for general affine nonlinear discrete-time systems with dead-zone control constraints.

In summary, the main contributions of this paper are as follows.

1. Present a framework for optimal control of nonlinear systems with unknown dead-zone constraints.
2. Develop a novel nonquadratic functional to deal with dead-zone control constraints of nonlinear discrete-time systems and derive the corresponding DTHJB equation.
3. Utilize two NNs to approximate the dead-zone and the corresponding nonlinear control system, respectively, and thus solve the corresponding unknown model problems.
4. The two simulation examples demonstrate that the iterative value function sequence converges to the optimal value function, and show the validity of the proposed algorithm.

Specifically, we use a model network to approximate the nonlinear system dynamics, which renders the iterative ADP algorithm suitable to unknown plants. In addition, the dead-zone nonlinearity will be approximated using NN. It is the first time to use HDP technique to solve the optimal control problem with dead-zone constraints.

2 Problem statement and preliminaries

Consider a class of affine nonlinear discrete-time system $\Sigma(f,g)$ with dead-zone control inputs described by

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

$$u_{ik} = \Phi(v_{ik}) = \begin{cases} \Phi_r(v_{ik}) & v_{ik} \geqslant b_r \\ 0 & b_l < v_{ik} < b_r \\ \Phi_l(v_{ik}) & v_{ik} \leqslant b_l, \end{cases} \tag{2}$$

where $x_k = x(k) \in \mathbb{R}^n$ is the state vector, and $f: \mathbb{R}^n \to \mathbb{R}^n$, $g: \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are differentiable with respect to their arguments with $f(0) = 0$. In addition, $u_{ik} = \Phi(v_{ik})$ is the output of dead-zone where $\Phi(\cdot)$ is an abstract mathematical description of the dead-zone, $v_{ik} = v_i(x_k)$ is the input vector to dead-zone, $\Phi_r(v_{ik})$ and $\Phi_l(v_{ik})$ are unknown nonlinear functions for $v_{ik} \in [b_r, +\infty)$ and for $v_{ik} \in (-\infty, b_l]$, respectively. We denote $\Omega_u = \{u_k | u_k = [u_{1k}, u_{2k}, \ldots, u_{mk}]^T \in \mathbb{R}^m, u_{ik} = \Phi(v_{ik}),\ \ i = 1, \ldots, m\}$.

As stated in Tao and Kokotovic (1994), the simplest linear symmetric dead-zone model is only a static simplification of diverse physical phenomena with negligible fast dynamics. Actually, in many industrial processes the dead-zone parameters are unknown, but its model is monotonic. Thus the dead-zone inverse function exists.

The key features of the control problems investigated in this paper are:

1. We have the input/output data of the dead-zone, namely prior knowledge of the dead-zone control.
2. Assume that $f + gu$ is Lipschitz continuous on a set $\Omega \subseteq \mathbb{R}^n$ containing the origin, and that the system $\Sigma(f,g)$ is controllable in the sense that there exists at least a continuous control law on Ω that asymptotically stabilizes the system. This assures not only the existence of solution of the DTHJB equation but also the corresponding optimal control.
3. The dead-zone and the system models cannot be acquired exactly.
4. The dead-zone is continuous monotonic.

Definition 1 (cf. Zhang et al. 2009) Stabilizable system: a nonlinear dynamical system is said to be stabilizable on a compact set $\Omega \subseteq \mathbb{R}^n$, if for all initial conditions $x_0 \in \Omega$, there is a control input $u_k \in \mathbb{R}^m$, such that the state $x_k \to 0$ as $k \to +\infty$.

It is desired to find the optimal control action u_k for the system $\Sigma(f,g)$ which minimizes the infinite-horizon value function given by

$$V(x_k) = \sum_{i=k}^{\infty} \gamma^{i-k} U(x_i, u_i), \tag{3}$$

where $U(x_i, u_i)$ is the utility function and is positive definite, i.e., $U(0, 0) = 0$ for $x_i = 0$, $u_i = 0$, and $U(x_i, u_i) > 0$ for $\forall x_i \neq 0, u_i \neq 0$. γ is the discount factor with $0 < \gamma \leqslant 1$. The utility function usually can be expressed as

$$U(x_i, u_i) = x_i^T Q x_i + W(u_i), \tag{4}$$

where Q and $W(u_i)$ are positive definite as above, and for unconstrained control inputs, a common choice for $W(u_i)$ is $W(u_i) = u_i^T R u_i$, where $R > 0, R \in \mathbb{R}^{m \times m}$.

But for constrained control inputs, inspired by Abu-Khalaf and Lewis (2005) and Lyshevski (1998) who introduced a generalized nonquadratic function when dealing with bounded controls, we can define

$$\begin{aligned} W(u_i) &= 2\int_0^{u_i} \Phi^{-T}(v_i) R dv_i, \\ \Phi(v_i) &= [\phi(v_{1i}), \phi(v_{2i}), \ldots, \phi(v_{mi})]^T, \end{aligned} \tag{5}$$

where $\Phi(\cdot)$ is a constrained and continuous monotonic function, and R is positive definite and assumed to be symmetric for simplicity. Since $\phi(\cdot)$ is a monotonic odd function and R is positive definite, $W(u)$ is positive definite. For the general optimal control problems, the controllers need to be stable and also to guarantee that (3) is finite, i.e., the control must be admissible (Abu-Khalaf and Lewis 2005).

Definition 2 Admissible control: a control u_k is said to be admissible with respect to (3) on Ω if u_k is continuous on a compact set $\Omega \subseteq \mathbb{R}^n$, $u_0 = 0$, u_k stabilizes (3) on Ω, and $\forall x_0 \in \Omega$, $V(x_0)$ is finite.

Equation (3) can be rewritten as follows:

$$\begin{aligned} V(x_k) &= \sum_{i=k}^{\infty} \gamma^{i-k}\left(x_i^T Q x_i + W(u_i)\right) \\ &= x_k^T Q x_k + W(u_k) \\ &\quad + \gamma \sum_{i=k+1}^{\infty} \gamma^{i-k-1}\left(x_i^T Q x_i + W(u_i)\right) \\ &= x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V(x_{k+1}). \end{aligned} \tag{6}$$

According to Bellman's optimality principle, the optimal value function $V^*(x_k)$ satisfies the following DTHJB equation:

$$V^*(x_k) = \min_{u_k}\left\{ x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^*(x_{k+1}) \right\}. \tag{7}$$

The optimal control u_k^* at time k is the u_k which achieves the aforementioned minimum, i.e.,

$$u_k^* = \arg\min_{u_k}\left\{ x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^*(x_{k+1}) \right\}. \tag{8}$$

Note that the DTHJB equation develops backward in time, and u_k^* satisfies the first-order necessary condition, which is given by the gradient of the right-hand side of (7) with respect to u_k as

$$\frac{\partial}{\partial u_k}\left(x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^*(x_{k+1}) \right) = 0. \tag{9}$$

Therefore, we have

$$u_k^* = \Phi\left(-\frac{\gamma}{2} R^{-1} \left(\frac{\partial x_{k+1}}{\partial u_k} \right)^T \frac{\partial V^*(x_{k+1})}{\partial x_{k+1}} \right). \tag{10}$$

By substituting (10) into (7), the DTHJB can be expressed as

$$V^*(x_k) = x_k^T Q x_k + 2\int_0^{u_k^*} \Phi^{-T}(v_k) R dv_k + \gamma V^*(x_{k+1}), \tag{11}$$

which is the optimal value function corresponding to the optimal control policy u_k^*.

Equations (10) and (11) are called the best optimized implementation of dynamic programming problems. In the general nonlinear dynamic case, the HJB equation cannot be solved analytically due to the well-known "curses of the dimensionality". Therefore, in the following sections, we will present how the HDP algorithm works with the definition of Riemann integral to solve approximately the optimal control problem with dead-zone constraints.

3 The approximate solution of the DTHJB equation for general nonlinear systems

As said above, the DTHJB equation is generally nonanalitical for nonlinear systems and the iterative ADP algorithm is a good method to solve it. Therefore, in order to overcome the difficulty in solving the DTHJB equation, we employ recursive ADP method combined with Riemann integral to obtain its approximate solution. Also we demonstrate the existence and uniqueness proofs of the solution in this section.

3.1 The iterative ADP algorithm

In this subsection, we develop the iterative ADP algorithm, based on the Bellman's principle of optimality, the Riemann integral, and the greedy HDP algorithm. This improved algorithm can avoid using the complex mathematical analysis method and ensure the calculating accuracy of the DTHJB Eq. (7). The algorithm can be performed as follows.

For any $x_k \in \mathbb{R}^n$ at time k, the first step is to assign an arbitrary nonnegative constant to the initial value function $V^0(x_k)$, e.g., $V^0(\cdot) = 0$, then we can search the corresponding control vector u_k^0, which optimizes the value function $V^1(x_k)$:

$$\begin{aligned} u_k^0 &= \arg\min_{u_k}\{V^1(x_k)\} \\ &= \arg\min_{u_k}\left\{x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^0(x_{k+1})\right\}. \end{aligned} \tag{12}$$

Once we find the corresponding optimal control vector u_k^0, the value function in the next step $V^1(x_k)$ can be updated as follows

$$\begin{aligned} V^1(x_k) &= \min_{u_k}\left\{x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^0(x_{k+1})\right\} \\ &= x_k^T Q x_k + 2\int_0^{u_k^0} \Phi^{-T}(v_k) R dv_k. \end{aligned} \tag{13}$$

In this way, for the iteration index $i = 1, 2, \ldots,$ the algorithm can be executed between

$$\begin{aligned} u_k^i &= \arg\min_{u_k}\{V^{i+1}(x_k)\} \\ &= \arg\min_{u_k}\left\{x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^i(x_{k+1})\right\} \end{aligned} \tag{14}$$

and

$$\begin{aligned} V^{i+1}(x_k) &= \min_{u_k}\left\{x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^i(x_{k+1})\right\} \\ &= x_k^T Q x_k + 2\int_0^{u_k^i} \Phi^{-T}(v_k) R dv_k + \gamma V^i(x_{k+1}) \end{aligned} \tag{15}$$

until both the value function and the corresponding control law converge to their respective optimal values.

Note that the integral term in the iterative algorithm cannot be solved for the unknown dead-zone nonlinear model. To get around such difficulty, we introduce the Riemann integral. Based on the Jordan measure, the Riemann integral is defined by taking the limit of the Riemann sum. Thus, the integral term in the above DTHJB equation can easily be solved when expressed as the following form

$$\int_0^{u_k} \Phi^{-T}(v_k) R dv_k = \lim_{\max(\Delta v_\ell)\to 0} \sum_{\ell=1}^{n^*} (\Phi^{-T}(v_k^\ell) R \Delta v_\ell), \tag{16}$$

where v_k^0 is an arbitrary point in the interval Δv_ℓ, and the subinterval $\Delta v_\ell = u_k/n^*$. The value $\max(\Delta v_\ell)$ is called the maximum of a partition of the interval $[0, u_k]$ into subintervals Δv_ℓ.

So far, the DTHJB equation has been solved through the improved iterative ADP algorithm and the Riemann integral. In the next subsection, we will prove the solution's existence and uniqueness

3.2 Existence and uniqueness of the solution of the DTHJB equation

Theorem 1 *Consider the iterative value function sequence* $\{V^i(x_k)\}$ *in* (15) *and its corresponding controls* $\{u_k^{\mathrm{i}}\}$ *in* (14), *respectively. If the nonlinear system is stabilizable on a compact set* Ω *and there is an admissible control sequence* $\{u_k^{\mathrm{i}}\}$, *then the derived DTHJB Eq.* (7) *has a unique solution and it can be obtained through the iteration between* (14) *and* (15), *i.e.,* $V^\infty(x_k) = V^*(x_k),\ u_k^\infty = u_k^*$.

Proof

1. Existence

   For any admissible control input ξ_k^i, the corresponding value function $\Gamma^i(x_k)$ is updated by

$$\Gamma^{i+1}(x_k) = x_k^T Q x_k + W(\xi_k^i) + \gamma\Gamma^i(x_{k+1}). \tag{17}$$

Let $\Gamma^0(\cdot) = V^0(\cdot)$ be any nonnegative constant. For convenience, let the constant be zero. So

$$\begin{aligned} \Gamma^{i+1}(x_k) - \Gamma^i(x_k) &= \gamma\left(\Gamma^i(x_{k+1}) - \Gamma^{i-1}(x_{k+1})\right) \\ &= \gamma^2\left(\Gamma^{i-1}(x_{k+2}) - \Gamma^{i-2}(x_{k+2})\right) \\ &= \gamma^3\left(\Gamma^{i-2}(x_{k+3}) - \Gamma^{i-3}(x_{k+3})\right) \\ &\ \ \vdots \\ &= \gamma^i\left(\Gamma^1(x_{k+i}) - \Gamma^0(x_{k+i})\right) \\ &= \gamma^i\Gamma^1(x_{k+i}). \end{aligned} \tag{18}$$

Thus, $\Gamma^{i+1}(x_k)$ can be rewritten as the following form

$$\begin{aligned} \Gamma^{i+1}(x_k) &= \Gamma^i(x_k) + \gamma^i\Gamma^1(x_{k+i}) \\ &= \Gamma^{i-1}(x_k) + \gamma^{i-1}\Gamma^1(x_{k+i-1}) + \gamma^i\Gamma^1(x_{k+i}) \\ &= \Gamma^1(x_k) + \gamma\Gamma^1(x_{k+1}) + \cdots \\ &\quad + \gamma^{i-1}\Gamma^1(x_{k+i-1}) + \gamma^i\Gamma^1(x_{k+i}) \\ &= \sum_{j=0}^{i} \gamma^j\Gamma^1(x_{k+j}). \end{aligned} \tag{19}$$

Combining with $\Gamma^i(k) \geqslant 0$, we can further get

$$\Gamma^{i+1}(x_k) = \sum_{j=0}^{i} \gamma^j \left(x_{k+j}^T Q x_{k+j} + W(\xi_{k+j}) \right) \leqslant \sum_{j=0}^{\infty} \gamma^j \left(x_{k+j}^T Q x_{k+j} + W(\xi_{k+j}) \right). \quad (20)$$

Since the nonlinear system is stabilizable, that is, the state $x_k \to 0$ as $k \to \infty$, we have

$$\Gamma^{i+1}(x_k) \leqslant \sum_{j=0}^{\infty} \gamma^j \left(x_{k+j}^T Q x_{k+j} + W(\xi_{k+j}) \right) \leqslant C, \quad \forall i, \quad (21)$$

where C is a nonnegative constant. On the other hand, $V^{i+1}(x_k)$ is the minimum of the right hand side of (15) with respect to the corresponding control input u_k. Thus,

$$V^{i+1}(x_k) \leqslant \Gamma^{i+1}(x_k) \leqslant C, \quad \forall i. \quad (22)$$

In conclusion, the limit of the value function sequence $\{V^i\}$ exists.

2. Uniqueness

First, we will prove that $\{V^i(x_k)\}$ is a nondecreasing sequence. Define a new value function at u_k^{i+1}

$$\Psi^{i+1}(x_k) = x_k^T Q x_k + W(u_k^{i+1}) + \gamma \Psi^i(x_{k+1}), \quad (23)$$

with the initial value $\Psi^0(\cdot) = V^0(\cdot)$.

For $i = 0$, $V^1(x_k) = x_k^T Q x_k + W(u_k^0) + \Psi^0(x_k) \geqslant \Psi^0(x_k)$.

We assume that $V^i(x_k) \geqslant \Psi^{i-1}(x_k)$ for $i - 1$. Then for i, we have $V^{i+1}(x_k) = x_k^T Q x_k + W(u_k^i) + \gamma V^i(x_{k+1}) \geqslant x_k^T Q x_k + W(u_k^i) + \gamma \Psi^{i-1}(x_{k+1}) = \Psi^i(x_k)$. Combining with (22), we get

$$V^{i+1}(x_k) \geqslant \Psi^i(x_k) \geqslant V^i(x_k). \quad (24)$$

Next we will present how to prove the uniqueness of the solution using the above property.

According to (15),

$$V^i(x_k) \leqslant x_k^T Q x_k + W(u_k) + \gamma V^{i-1}(x_{k+1}) \leqslant x_k^T Q x_k + W(u_k) + \gamma V^{\infty}(x_{k+1}).$$

When $i \to \infty$,

$$V^{\infty}(x_k) \leqslant x_k^T Q x_k + W(u_k) + \gamma V^{\infty}(x_{k+1}), \quad \forall u_k,$$

which suggests that

$$V^{\infty}(x_k) \leqslant \min_{u_k}\{x_k^T Q x_k + W(u_k) + \gamma V^{\infty}(x_{k+1})\}. \quad (25)$$

On the other hand,

$$V^{\infty}(x_k) \geqslant V^i(x_k) = \min_{u_k}\{x_k^T Q x_k + W(u_k) + \gamma V^{i-1}(x_{k+1})\}, \quad \forall i.$$

Similarly, when $i \to \infty$,

$$V^{\infty}(x_k) \geqslant \min_{u_k}\{x_k^T Q x_k + W(u_k) + \gamma V^{\infty}(x_{k+1})\}. \quad (26)$$

Combining (25) with (26), we can conclude that

$$V^{\infty}(x_k) = \min_{u_k}\{x_k^T Q x_k + W(u_k) + \gamma V^{\infty}(x_{k+1})\}. \quad (27)$$

Given the above, we have proved the existence and uniqueness of the solution of the DTHJB equation. The solution can be obtained through the iterative algorithm.

Remark 1 As we all know, the optimal control of nonlinear discrete time systems is often reduced to solving the nonlinear DTHJB equation. From above, we can see that $V^{\infty}(x_k)$ is the unique solution of the DTHJB equation, which indicates that both the value function and the corresponding control law sequence converge to the optimal values, respectively, i.e. $V^{\infty}(x_k) = V^*(x_k)$, $u_k^{\infty} = u_k^*$.

4 NN implementation

It is well known that NNs can be employed to approximate any continuous function on prescribed compact sets (Liu et al. 2010; Zhang et al. 2010), so it is natural to use NNs to approximate the nonlinear system and dead-zone model. Here, we use the multi-layered back-propagation (BP) NNs. Figure 1 shows the structural diagram of the iterative algorithm.

As for (10), we could build an inverse dead-zone NNs using the prior knowledge of the dead-zone control. Then, for $i = 0, 1, 2, \ldots$, the iterative algorithm can be implemented between

$$u_k^i = \arg\min_{u_k}\left\{ x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v) R dv + \gamma V^i(x_{k+1}) \right\} = \Phi\left(-\frac{\gamma}{2} R^{-1} \left(\frac{\partial x_{k+1}}{\partial u_k^i} \right)^T \frac{\partial V^i(x_{k+1})}{\partial x_{k+1}} \right), \quad (28)$$

and

$$V^{i+1}(x_k) = \min_{u_k}\left\{ x_k^T Q x_k + 2\int_0^{u_k} \Phi^{-T}(v_k) R dv_k + \gamma V^i(x_{k+1}) \right\} = x_k^T Q x_k + 2 \lim_{\max(\Delta v_\ell) \to 0} \sum_{\ell=1}^{n^*} (\Phi^{-T}(v_k^\ell) R \Delta v_\ell) + \gamma V^i(x_{k+1}), \quad (29)$$

where $\Delta v_\ell = u_k / n^*$.

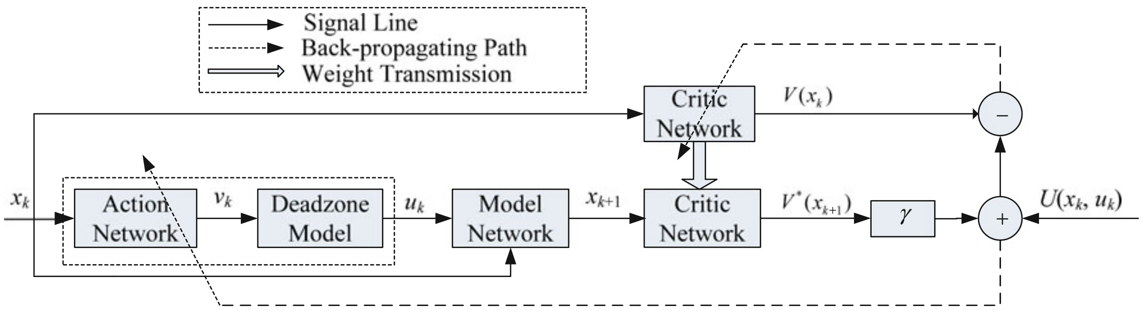


Fig. 1 Flowchart of the proposed iterative algorithm with dead-zone

In the above iterative algorithm, i is the iterative index of the value function and the control law, while k is the time index. The value function and the control law are updated until they both converge to their respective optimal values. In the above flowchart, the critic network estimates the optimal value function $V^*(x_k)$, and is trained forward in time. While the action network is used to approximate the optimal control input.

To approximate the nonlinear value function and nonlinear dynamic system, we choose backpropagation algorithm to train artificial neural networks and mean square error is used as a measure of how well the neural network has learnt. The nonlinear activation function in the critic network is chosen as logistic function which can make its output a positive number, while in the model network and the action network we adopt tansig function. For the structure and other parameters of our NNs, we will present them in the next two specific examples.

5 Simulation studies

In this section, the effectiveness of the proposed iterative algorithm is demonstrated for nonlinear affine systems with unknown dead-zone control input by using two examples.

Example 1 (*Nonlinear affine system*) First of all the present algorithm is implemented for the example identical to the one in Chen and Jagannathan (2008) except the dead-zone control. We consider the following system:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 0.05x_{2k} + x_{1k} \\ -0.0005x_{1k} - 0.0335x_{1k}^3 + x_{2k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.05 \end{bmatrix} u_k, \tag{30}$$

$$u_k = \Phi(v_k) = \begin{cases} v_k - 0.1 & v_k \geqslant 0.1 \\ 0 & -0.1 < v_k < 0.1 \\ v_k + 0.1 & v_k \leqslant -0.1. \end{cases} \tag{31}$$

We choose three-layer three-layer BP NNs as model network, dead-zone network, critic network, and action network with the structures 3–12–2, 1–12–1, 2–8–1, and 2–8–1, respectively. The initial weights of the four networks are all set to be random in [−1,1]. It should be noted that the model identification of the dead-zone and the dynamic system is implemented first under the learning rate $l_m = 0.05$, and then their weights are kept unchanged. After that, we use the trained model network and the dead-zone network to train the critic network and the action network with the learning rates $l_c = l_a = 0.05$, for $i = 50$ iteration steps, where $n^* = 1{,}000$. The discount factor is chosen as $\gamma = 0.35$. Enough iteration steps should be implemented to guarantee the solution accuracy of the DTHJB equation. The value function is defined as (4), where $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $R = 0.05$.

Then for the given initial state $x_0 = [1, \ -1]^T$, we apply the iterative algorithm to the controlled nonlinear system and obtain the iterative value function as in Fig. 2 which verified the presented theory. From Figs. 3 and 4, we can see that the system states, the value function and the corresponding control sequence converge to their respective optimal values quickly.

Example 2 (*The pendulum*) The second example studied here is the pendulum swinging up and balancing control problem with dead-zone constraints. This example is chosen from Si and Wang (2001) with some modifications. Here we consider the following nonlinear affine discrete-time system:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 0.1x_{2k} + x_{1k} \\ -0.49\sin(x_{1k}) + 0.8x_{2k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} u_k. \tag{32}$$

The dead-zone nonlinear model is as follows

$$u_k = \Phi(v_k) = \begin{cases} (1 - 0.3\sin(v_k))(v_k - 0.2) & v_k \geqslant 0.2 \\ 0 & -0.2 < v_k < 0.2 \\ (0.8 - 0.2\cos(v_k))(v_k + 0.2) & v_k \leqslant -0.2. \end{cases} \tag{33}$$

We choose NNs as above and apply the algorithm to the plant with the initial state [1; −1]. The simulation parameters and value function are defined the same as in

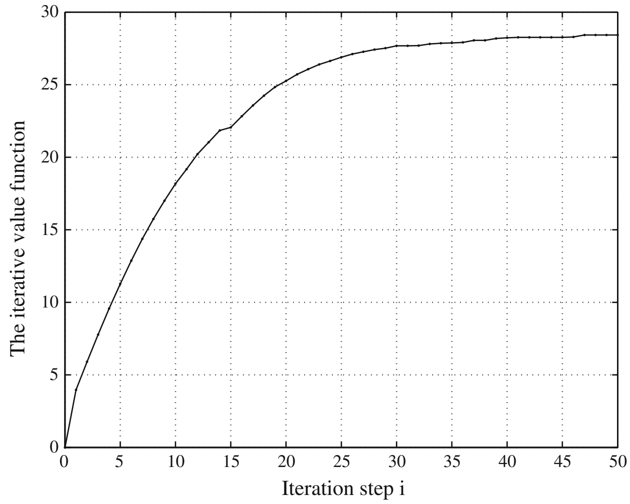


Fig. 2 The value function with iteration step i

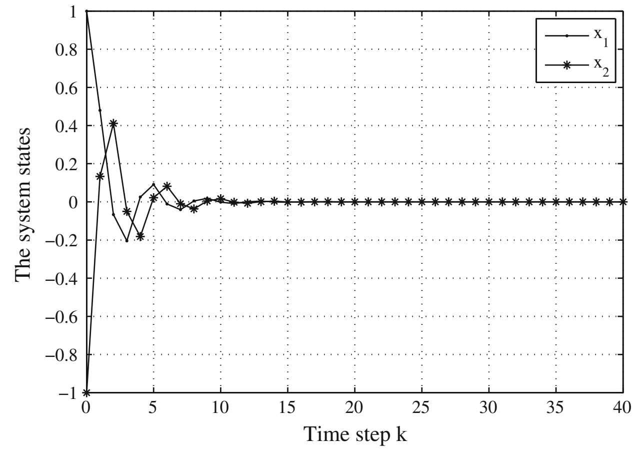


Fig. 3 The system state vector

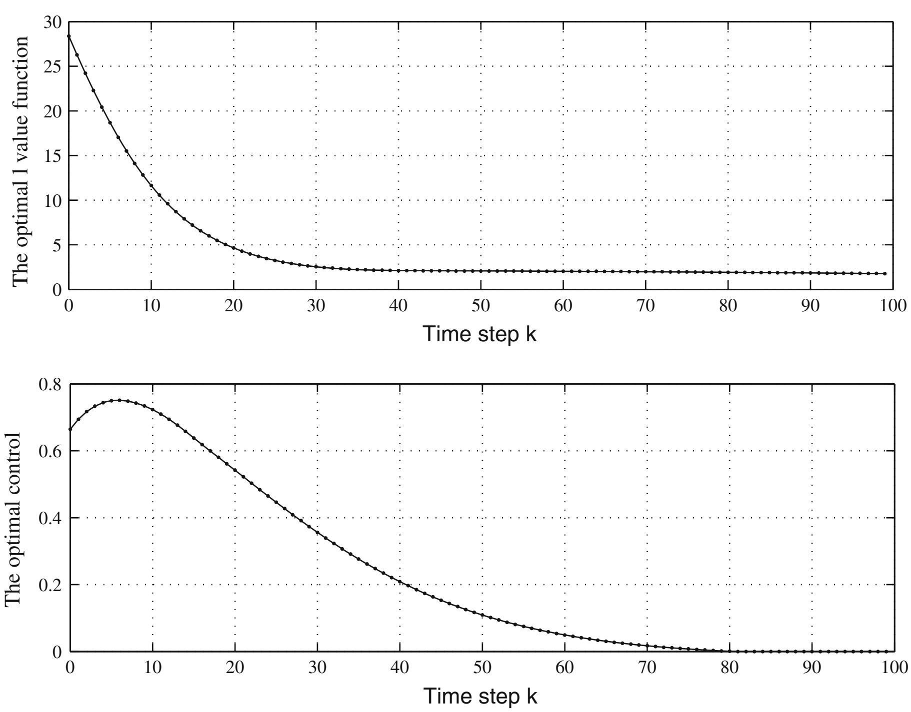
Example 1. Then we can obtain the following simulation results. Figure 5 shows the value function with successive approximation whereas Fig. 6 demonstrates the value function with the updating time step. From Fig. 7 we can see that the state trajectories converge to the equilibrium point quickly. Figures 8 and 9 display the dead-zone input and its output with the time step iterations.

From the results, we can see that the algorithm has been proven to be effective. The value function and the control converge to their optimal values and the system states also achieve the static equilibrium quickly. Note that the method is independent on the mathematical models of nonlinear system and dead-zone. The models in the paper act as the role of providing data of the plant.

6 Conclusions

Based on ADP technique and Rimann integral, we derived the DTHJB equation and solved approximately the equation through the improved iterative algorithm. We also demonstrated proofs of the existence and the uniqueness of the solution. Inspired by the fact that NNs can approximate arbitrary nonlinear functions, we established dead-zone network and the model network, and thus obtained the approximate solution for the optimal control problem of

Fig. 4 The optimal value function and the corresponding optimal control with time step k



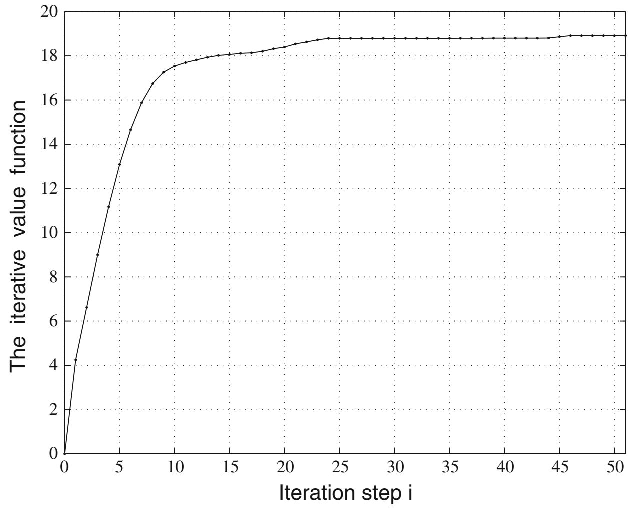


Fig. 5 The value function with iteration step *i*

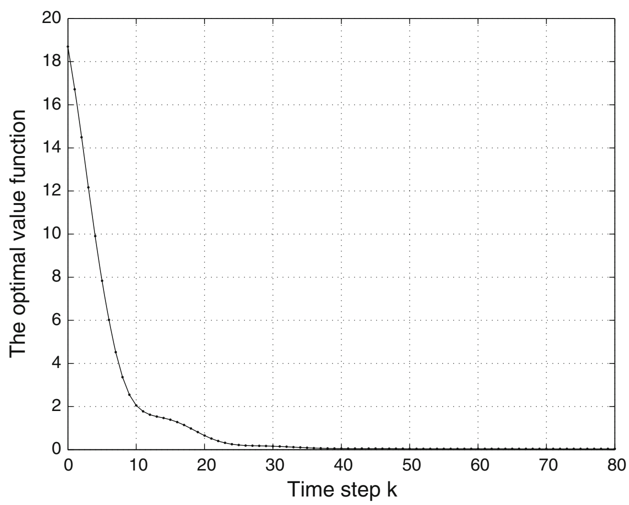


Fig. 6 The optimal value function with time step *k*

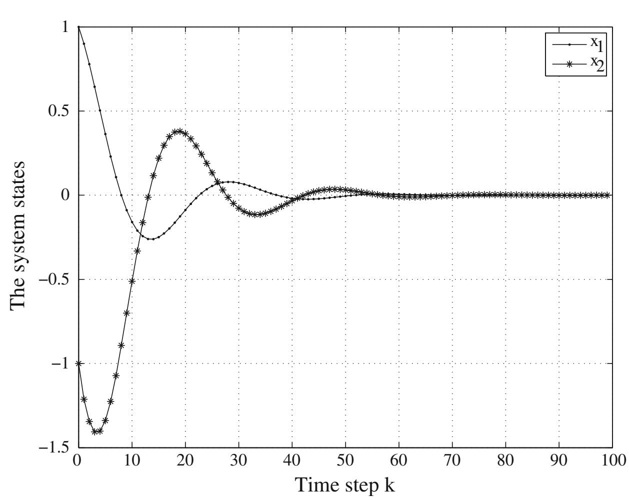


Fig. 7 The system state vector

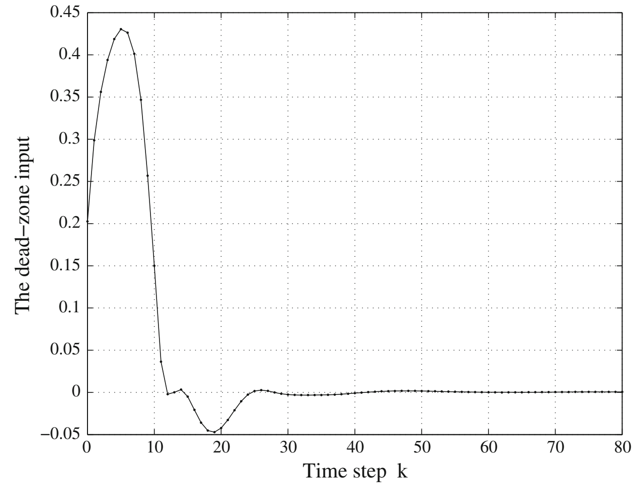


Fig. 8 The dead-zone input

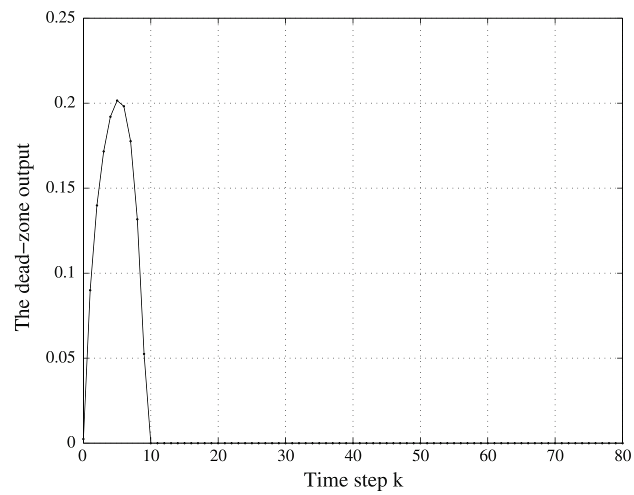


Fig. 9 The dead-zone output

nonlinear affine system with dead-zone constraints. The correctness and validity of the theoretical analysis and the algorithm were demonstrated by simulation results.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China under Grants 61034002, 61233001, and 61273140.

References

Abu-Khalaf M, Lewis FL (2005) Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. Automatica 41(5):779–791

Al-Tamimi A, Abu-Khalaf M, Lewis FL (2007) Adaptive critic designs for discrete-time zero-sum games with application to H_∞ control. IEEE Trans Syst Man Cybern B 37(1):240–247

Al-Tamimi A, Lewis FL (2008) Discrete-time nonlinear HJB solution using approximate dynamic programming: convergence proof. IEEE Trans Syst Man Cybern B 38(4):943–949

Balakrishnan SN, Ding J, Lewis FL (2008) Issues on stability of ADP feedback controllers for dynamical systems. IEEE Trans Syst Man Cybern B 38(4):913–917

Bellman RE (1957) Dynamic programming. Princeton University Press, Princeton

Bernstein DS (1995) Optimal nonlinear, but continuous, feedback control of systems with saturating actuators. Int J Control 62(5):1209–1216

Chen Z, Jagannathan S (2008) Generalized Hamilton–Jacobi–Bellman formulation-based neural network control of affine nonlinear discrete-time systems. IEEE Trans Neural Netw 19(1):90–106

Gao W, Rastko RS (2006) Neural network control of a class of nonlinear systems with actuator saturation. IEEE Trans Neural Netw 17(1):147–156

Lewis FL, Tim WK, Wang LZ, Li ZX (1999) Deadzone compensation in motion control systems using adaptive fuzzy logic control. IEEE Trans Control Syst Technol 7(6):731–742

Liu DR, Xiong XX, Zhang Y (2001) Action-dependent adaptive critic designs. In: Proceeding of the International Joint Conference on Neural Networks, Washington, pp 990–995

Liu DR, Zhang Y, Zhang HG (2005) A self-learning call admission control scheme for CDMA cellular networks. IEEE Trans Neural Netw 16(5):1219–1228

Liu ZW, Zhang HG, Zhang QL (2010) Novel stability analysis for recurrent neural networks with multiple delays via line integral-type L-K functional. IEEE Trans Neural Netw 21(11):1710–1718

Lyshevski SE (1998) Optimal control of nonlinear continuous-time systems: design of bounded controllers via generalized nonquadratic functional. In: Proceedings of American control conference, Philadelphia, pp 205–209

Ma HJ, Yang GH (2010) Adaptive output control of uncertain nonlinear systems with non-symmetric dead-zone input. Automatica 46(2):413–420

Prokhorov DV, Wunsch DC (1997) Adaptive critic designs. IEEE Trans Neural Netw 8(5):997–1007

Recker D, Kokotovic P, Rhode D, Winkelman J (1991) Adaptive nonlinear control of systems containing a dead-zone. In: Proceedings of the 30th IEEE conference on decision and control, Brighton, pp 2111–2115

Saberi A, Lin Z, Teel A (1996) Control of linear systems with saturating actuators. IEEE Trans Autom Control 41(3):368–378

Seiffertt J, Sanyal S, Wunsch DC (2001) Hamilton–Jacobi–Bellman equations and approximate dynamic programming on time scales. IEEE Trans Syst Man Cybern B 38(4):918–923

Selmic RR, Lewis FL (2000) Deadzone compensation in motion control systems using neural networks. IEEE Trans Autom Control 45(4):602–613

Si J, Wang Y (2001) Online learning control by association and reinforcement. IEEE Trans Neural Netw 12(2):264–276

Sussmann H, Sontag ED, Yang Y (1994) A general result on the stabilization of linear systems using bounded controls. IEEE Trans Autom Control 39(12):2411–2425

Tao G, Kokotovic PV (1994) Adaptive control of plants with unknown dead-zones. IEEE Trans Autom Control 39(1):59–68

Tao G, Kokotovic PV (1995) Adaptive control of systems with unknown output backlash. IEEE Trans Autom Control 40(2):326–330

Tao G, Kokotovic PV (1995) Continuous-time adaptive control of systems with unknown backlash. IEEE Trans Autom Control 40(6):1083–1087

Tao G, Kokotovic PV (1995) Discrete-time adaptive control of systems with unknown deadzones. Int J Control 61(1):1–17

Tao G, Kokotovic PV (1996) Adaptive control of systems with actuator and sensor nonlinearities. John Wiley, New York

Wang XS, Su CY, Hong H (2004) Robust adaptive control of a class of nonlinear systems with unknown dead-zone. Automatica 40(3):407–413

Wei QL, Zhang HG, Dai J (2009) Model-free multiobjective approximate dynamic programming for discrete-time nonlinear systems with general performance index functions. Neurocomputing 72(7–9):1839–1848

Werbos PJ (1992) Approximate dynamic programming for real-time control and neural modeling. In: White DA, Sofge DA (eds) Handbook of intelligent control: neural, fuzzy, and adaptive approaches, New York

Xu JX, Xu J, Lee TH (2005) Iterative learning control for systems with input deadzone. IEEE Trans Autom Control 50(9):1455–1459

Zhang TP, Ge SS (2007) Adaptive neural control of MIMO nonlinear state time-varying delay systems with unknown dead-zones and gain signs. Automatica 43(6):1021–1033

Zhang TP, Ge SS (2008) Adaptive dynamic surface control of nonlinear systems with unknown deadzone in pure feedback form. Automatica 44(7):1895–1903

Zhang H, Luo Y, Liu D (2009) Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. IEEE Trans Neural Netw 20(9):1490–1503

Zhang HG, Liu ZW, Huang GB, Wang ZS (2010) Novel weighting-delay-based stability criteria for recurrent neural networks with time-varying delay. IEEE Trans Neural Netw 21(1):91–106