Automatica 48 (2012) 1825–1832

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming[☆]

Ding Wang [a], Derong Liu [a,1], Qinglai Wei [a], Dongbin Zhao [a], Ning Jin [b]

[a] State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, PR China

[b] Department of Electrical and Computer Engineering, University of Illinois, Chicago, IL 60607, USA

ARTICLE INFO

Article history:
Received 31 October 2010
Received in revised form
22 January 2012
Accepted 8 February 2012
Available online 25 June 2012

Keywords:
Adaptive critic designs
Adaptive dynamic programming
Approximate dynamic programming
Globalized dual heuristic programming
Intelligent control
Neural network
Optimal control

ABSTRACT

An intelligent-optimal control scheme for unknown nonaffine nonlinear discrete-time systems with discount factor in the cost function is developed in this paper. The iterative adaptive dynamic programming algorithm is introduced to solve the optimal control problem with convergence analysis. Then, the implementation of the iterative algorithm via globalized dual heuristic programming technique is presented by using three neural networks, which will approximate at each iteration the cost function, the control law, and the unknown nonlinear system, respectively. In addition, two simulation examples are provided to verify the effectiveness of the developed optimal control approach.

1. Introduction

The main difference between optimal control of linear systems and nonlinear systems lies in that the latter often requires solving the nonlinear Hamilton–Jacobi–Bellman (HJB) equation instead of the Riccati equation (Abu-Khalaf & Lewis, 2005; Al-Tamimi, Lewis, & Abu-Khalaf, 2008; Primbs, Nevistic, & Doyle, 2000; Wang, Zhang, & Liu, 2009). For example, the discrete-time HJB (DTHJB) equation is more difficult to deal with than Riccati equation because it involves solving nonlinear partial difference equations. Although there were some methods that did not need to solve the HJB equation directly (e.g., Beard, Saridis, & Wen, 1997; Chen, Edgar, & Manousiouthakis, 2004), they were limited to handle some special classes of systems or they needed to perform very complex calculations. On the other hand, dynamic programming (DP) has been a useful technique in solving optimal control problems for many years (Bellman, 1957). However, it is often computationally untenable to run DP to obtain optimal solutions due to the "curse of dimensionality" (Bellman, 1957). Moreover, the backward direction of search precludes the application of DP in real-time control.

Artificial neural networks (ANN or NN) are an effective tool to implement intelligent control due to the properties of nonlinearity, adaptivity, self-learning, fault tolerance, and universal approximation of input–output mapping (Jagannathan, 2006; Werbos, 1992, 2008, 2009). Thus, it has been used for universal function approximation in adaptive/approximate dynamic programming (ADP) algorithms, which were proposed in Werbos (1992, 2008, 2009) as a method to solve optimal control problems forward-in-time. There are several synonyms used for ADP including "adaptive dynamic programming" (Lewis & Vrabie, 2009; Liu & Jin, 2008; Murray, Cox, Lendaris, & Saeks, 2002; Wang et al., 2009), "approximate dynamic programming" (Al-Tamimi et al., 2008; Werbos, 1992), "neuro-dynamic programming" (Bertsekas & Tsitsiklis, 1996), "neural dynamic programming" (Si & Wang, 2001), "adaptive critic designs" (Prokhorov & Wunsch, 1997), and "reinforcement learning" (Watkins & Dayan, 1992).

As an effective intelligent control method, in recent years, ADP and the related research have gained much attention from researchers (Balakrishnan & Biega, 1996; Balakrishnan, Ding, & Lewis, 2008; Dierks, Thumati, & Jagannathan, 2009; Jagannathan &

☆ This work was supported in part by the National Natural Science Foundation of China under Grants 60874043, 60904037, 60921061, and 61034002, and by Beijing Natural Science Foundation under Grant 4102061. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Gang Tao under the direction of Editor Miroslav Krstic.

E-mail addresses: ding.wang@ia.ac.cn (D. Wang), derong.liu@ia.ac.cn (D. Liu), qinglai.wei@ia.ac.cn (Q. Wei), dongbin.zhao@ia.ac.cn (D. Zhao), njin@uic.edu (N. Jin).

[1] Tel.: +86 10 62557379; fax: +86 10 62650912.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.05.049

He, 2008; Vamvoudakis & Lewis, 2010; Venayagamoorthy, Harley, & Wunsch, 2002; Venayagamoorthy, Wunsch, & Harley, 2000; Vrabie & Lewis, 2009; Yen & Delima, 2005; Zhang, Luo, & Liu, 2009; Zhang, Wei, & Liu, 2011). According to Prokhorov and Wunsch (1997) and Werbos (1992), ADP approaches were classified into several main schemes: heuristic dynamic programming (HDP), action-dependent HDP (ADHDP; note the prefix "action-dependent" (AD) used hereafter), also known as Q-learning (Watkins & Dayan, 1992), dual heuristic dynamic programming (DHP), ADDHP, globalized DHP (GDHP), and ADGDHP. Al-Tamimi et al. (2008) derived a significant result that applied the HDP iteration algorithm to solve the DTHJB equation of affine nonlinear discrete-time systems.

In this paper, we will tackle the optimal control problem for unknown nonlinear discrete-time systems using iterative ADP algorithm via GDHP technique (iterative GDHP algorithm for brief). Though great progress has been made for ADP in optimal control field, to the best of our knowledge, there is still no result to solve this problem by using the iterative GDHP algorithm. Additionally, the outputs of critic network of the GDHP technique contain not only the cost function but also its derivatives. This is different from HDP and DHP and is very important because the information associated with the cost function is as useful as the knowledge of its derivatives. Though the structure of the GDHP technique is somewhat complicated, it is expected to bring remarkable advantage when compared with simple ADP strategies. These motivate our research.

This paper is organized as follows. In Section 2, we present the formulation of the problem. In Section 3, we develop the optimal control scheme based on iterative ADP algorithm with convergence analysis, and then present the corresponding NN implementation of the iterative GDHP algorithm. In Section 4, two examples are given to demonstrate the effectiveness of the present control strategy. In Section 5, concluding remarks are given.

2. Problem statement

Here, we make the assumption that the state of the controlled system is available for measurement.

In this paper, we will study the nonlinear discrete-time systems described by

$$x_{k+1} = F(x_k, u_k), \quad k = 0, 1, 2, \dots, \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state and $u_k = u(x_k) \in \mathbb{R}^m$ is the control vector. Let x_0 be the initial state. The system function $F(x_k, u_k)$ is continuous for $\forall x_k, u_k$ and $F(0, 0) = 0$. Hence, $x = 0$ is an equilibrium state of system (1) under the control $u = 0$.

Definition 1. A nonlinear dynamical system is said to be stabilizable on a compact set $\Omega \in \mathbb{R}^n$, if for all initial states $x_0 \in \Omega$, there exists a control sequence $u_0, u_1, \dots, u_i \in \mathbb{R}^m$, $i = 0, 1, \dots$, such that the state $x_k \to 0$ as $k \to \infty$.

It is desired to find the control law $u_k = u(x_k)$ which minimizes the infinite horizon cost function given by

$$J(x_k) = \sum_{p=k}^{\infty} \gamma^{p-k} U(x_p, u_p), \tag{2}$$

where U is the utility function, $U(0, 0) = 0$, $U(x_p, u_p) \geq 0$ for $\forall x_p, u_p$, and γ is the discount factor with $0 < \gamma \leq 1$. In this paper, the utility function is chosen as the quadratic form $U(x_p, u_p) = x_p^T Q x_p + u_p^T R u_p$, where Q and R are positive definite matrices with suitable dimensions.

For optimal control problems, the designed feedback control must not only stabilize the system on Ω but also guarantee that (2) is finite, i.e., the control must be admissible.

Definition 2. A control $u(x)$ is said to be admissible with respect to (2) on Ω if $u(x)$ is continuous on a compact set $\Omega_u \in \mathbb{R}^m$, $u(0) = 0$, u stabilizes (1) on Ω, and $\forall x_0 \in \Omega$, $J(x_0)$ is finite.

Note that Eq. (2) can be written as

$$\begin{aligned} J(x_k) &= x_k^T Q x_k + u_k^T R u_k + \gamma \sum_{p=k+1}^{\infty} \gamma^{p-k-1} U(x_p, u_p) \\ &= x_k^T Q x_k + u_k^T R u_k + \gamma J(x_{k+1}). \end{aligned} \tag{3}$$

According to Bellman's optimality principle, the optimal cost function $J^*(x_k)$ satisfies the DTHJB equation

$$J^*(x_k) = \min_{u_k} \left\{ x_k^T Q x_k + u_k^T R u_k + \gamma J^*(x_{k+1}) \right\}. \tag{4}$$

Besides, the optimal control u^* can be expressed as

$$u^*(x_k) = \arg\min_{u_k} \left\{ x_k^T Q x_k + u_k^T R u_k + \gamma J^*(x_{k+1}) \right\}. \tag{5}$$

By substituting (5) into (4), the DTHJB equation becomes

$$J^*(x_k) = x_k^T Q x_k + u^{*T}(x_k) R u^*(x_k) + \gamma J^*(x_{k+1}). \tag{6}$$

It should be noticed that Definitions 1 and 2 are the same for linear systems. Moreover, when dealing with linear quadratic regulator problems, the DTHJB equation reduces to the Riccati equation which can be efficiently solved. For the general nonlinear case, however, it is considerably difficult to cope with the DTHJB equation directly. Therefore, we will develop an iterative ADP algorithm to solve it in the next section, based on Bellman's optimality principle and the greedy iteration approach.

3. Neuro-optimal control scheme based on iterative ADP algorithm via the GDHP technique

3.1. Derivation of the iterative algorithm

First, we start with the initial cost function $V_0(\cdot) = 0$ and obtain the law of the single control vector $v_0(x_k)$ as follows:

$$v_0(x_k) = \arg\min_{u_k} \left\{ x_k^T Q x_k + u_k^T R u_k + \gamma V_0(x_{k+1}) \right\}. \tag{7}$$

Then, we update the cost function as

$$V_1(x_k) = x_k^T Q x_k + v_0^T(x_k) R v_0(x_k). \tag{8}$$

Next, for $i = 1, 2, \dots$, the algorithm iterates between

$$v_i(x_k) = \arg\min_{u_k} \left\{ x_k^T Q x_k + u_k^T R u_k + \gamma V_i(x_{k+1}) \right\} \tag{9}$$

and

$$V_{i+1}(x_k) = x_k^T Q x_k + v_i^T(x_k) R v_i(x_k) + \gamma V_i(F(x_k, v_i(x_k))). \tag{10}$$

In the above recurrent iteration, i is the iteration index, while k is the time index. The cost function and control law are updated until they converge to the optimal ones. In the following, we will present the convergence proof of the iteration between (9) and (10) with the cost function $V_i \to J^*$ and the control law $v_i \to u^*$ as $i \to \infty$.

3.2. Convergence analysis of the iterative algorithm

The convergence analysis provided here is an extension of that given in Al-Tamimi et al. (2008).