Automatica 50 (2014) 2624–2632

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Adaptive dynamic programming and optimal control of nonlinear nonaffine systems☆

Tao Bian [1], Yu Jiang, Zhong-Ping Jiang

Control and Networks Lab, Department of Electrical and Computer Engineering, Polytechnic School of Engineering, New York University, Brooklyn, NY 11201, USA

ARTICLE INFO

Article history:
Received 3 September 2013
Received in revised form
5 March 2014
Accepted 26 May 2014
Available online 11 September 2014

Keywords:
Adaptive optimal control
Adaptive dynamic programming
Nonaffine systems

ABSTRACT

In this paper, a novel optimal control design scheme is proposed for continuous-time nonaffine nonlinear dynamic systems with unknown dynamics by adaptive dynamic programming (ADP). The proposed methodology iteratively updates the control policy online by using the state and input information without identifying the system dynamics. An ADP algorithm is developed, and can be applied to a general class of nonlinear control design problems. The convergence analysis for the designed control scheme is presented, along with rigorous stability analysis for the closed-loop system. The effectiveness of this new algorithm is illustrated by two simulation examples.

1. Introduction

Adaptive control is a method that aims to control dynamic systems with unknown parameters by estimating the system parameters or updating the controller parameters via adaptive laws. The adaptive control method has been extensively studied for both linear and nonlinear systems (Åström & Wittenmark, 1995; Ge, Hang, Lee, & Zhang, 2002; Krstic, Kanellakopoulos, & Kokotovic, 1995; Marino, 1995; Tao, 2004) as well as nonaffine systems (Ge, Hang, & Zhang, 1999; Ge & Zhang, 2003; Ge, Zhang, & Lee, 2004). However, in the past literature, stability is the main concern for the closed-loop system, and optimality properties are not addressed systematically.

In this paper, we address the adaptive optimal control design for a class of nonlinear nonaffine systems by means of ADP. ADP is a non-model-based method that can directly approximate the optimal control policy via online learning. The idea of ADP can be traced back to the seminal work of Werbos (1968) that brings reinforcement learning and dynamic programming (DP) together to approximate the (generally intractable) solution of the Bellman's equation.

Over the past two decades, attention has mainly been given to ADP-related control design for Markov decision processes (MDP) (Bertsekas & Tsitsiklis, 1996; Bradtke & Michael, 1994; Powell, 2007; Singh, Jaakkola, Littman, & Szepesvári, 2000; Sutton, 1988; Sutton & Barto, 1998; Tsitsiklis, 1994; Watkins & Dayan, 1992) and discrete-time feedback control systems (Al-Tamimi, Lewis, & Abu-Khalaf, 2008; Lewis & Vrabie, 2009; Wang, Liu, Wei, Zhao, & Jin, 2012; Wang, Zhang, & Liu, 2009). However, the state spaces considered in MDP are finite or countable, and the stability issue is usually overlooked. Moreover, due to the difference between discrete and continuous-time Bellman's equations, the existing ADP methods for discrete-time systems cannot be extended to the continuous-time case directly. On the other hand, most existing ADP-based control design for continuous-time systems (Bhasin et al., 2013; Jiang & Jiang, 2012a; Murray, Cox, Lendaris, & Saeks, 2002; Vamvoudakis & Lewis, 2010; Vrabie & Lewis, 2009; Vrabie, Pastravanu, Abu-Khalaf, & Lewis, 2009) only considered the affine case, i.e., the control input is assumed to appear linearly in the standard state equation of dynamic systems. This assumption is quite restrictive and does not address the situation where the control input appears nonlinearly. For example, for the human glucose control system, it is found that the decreasing rate of the glucose level is not linearly related to the amount of the insulin produced by pancreas all the time (Muniyappa, Lee, Chen, & Quon, 2008). Thus, for the wide applicability of ADP, it is necessary and important to develop the ADP algorithm for nonaffine nonlinear systems. To the authors' best knowledge, only a few papers (Doya,

☆ This work has been supported in part by the National Science Foundation grants DMS-0906659, ECCS-1101401, and ECCS-1230040. The material in this paper was not presented at any IFAC conference. This paper was recommended for publication in revised form by Associate Editor Shuzhi Sam Ge under the direction of Editor Miroslav Krstic.

E-mail addresses: tbian@nyu.edu (T. Bian), yu.jiang@nyu.edu (Y. Jiang), zjiang@nyu.edu (Z.-P. Jiang).

[1] Tel.: +1 718 260 3779; fax: +1 718 260 3906.

http://dx.doi.org/10.1016/j.automatica.2014.08.023
0005-1098/

2000; Liu, Huang, Wang, & Wei, 2013) have studied the ADP design for continuous-time nonaffine systems. In Doya (2000), a temporal difference (TD) learning-based algorithm for continuous-time nonaffine systems was first developed. However, no convergence analysis was given for the proposed algorithm. Liu et al. (2013) investigated the ADP-based control design for continuous-time nonaffine systems. However, the cost considered is in a restrictive form, and only a result on the uniform ultimate boundedness property is obtained.

In this paper, a new ADP methodology is developed to solve the optimal control design problem for continuous-time nonaffine systems with unknown dynamics. Since the control input appears nonlinearly, the policy improvement step cannot be given in the form as in Vamvoudakis and Lewis (2010) and Vrabie and Lewis (2009). To address this problem, we introduce another set of basis functions in the learning process. Furthermore, we propose a policy iteration (PI) method for continuous-time nonaffine systems by combining the idea of successive approximation (Bellman, 1961; Howard, 1960; Leake & Liu, 1967) with Lyapunov function-based stability analysis. A comprehensive convergence analysis for the proposed ADP algorithm is also given.

The remainder of this paper is organized as follows. In Section 2, the optimal control problem is formulated; some preliminaries are also introduced. In Section 3, both model-based and non-model-based methods are presented to iteratively approximate the optimal control. The convergence analysis of the proposed ADP methodology is considered in Section 4. Simulation results are given in Section 5 to illustrate the effectiveness of the proposed ADP algorithm. Finally, the conclusion is drawn in Section 6.

Notations. Throughout the paper, we use $\mathbb{R}$ to denote the set of real numbers. $\mathbb{R}^+$ denotes the set of nonnegative real numbers. I_n denotes the identity matrix of dimension n. The interior of a set $\mathbb{A}$ is denoted as $\mathring{\mathbb{A}}$. $\|\cdot\|$ denotes the Euclidean norm for vectors, or the induced matrix norm for matrices.

2. Mathematical preliminaries

2.1. Problem formulation

Consider the continuous-time nonaffine dynamical system described by

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0, \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the system state, $u : \mathbb{R}^+ \to \mathbb{R}^m$ is a piecewise continuous control input, and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is a locally Lipschitz function with $f(0, 0) = 0$.

The cost associated with system (1) is given by

$$J(x_0; u) = \int_0^\infty r(x(t), u(t))dt, \tag{2}$$

where $r : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable and positive definite function.

Throughout this paper, we assume the cost (2) has a unique absolute minimum for each $x_0 \in \mathbb{R}^n$ with respect to u. Moreover, we assume

$$\lim_{\|x_0\|\to\infty} J(x_0; u) = \infty, \tag{3}$$

for any piecewise continuous control input u.

Problem 1. Consider system (1) and cost (2). The optimal control problem consists in finding an optimal controller u^*, such that the closed-loop system is globally asymptotically stable (GAS) at the origin, and satisfies

$$J(x_0; u) \geq J(x_0; u^*),$$

for any piecewise continuous controller u and $x_0 \in \mathbb{R}^n$.

2.2. Optimal control design

It is well known that Problem 1 can be solved by using DP (Bellman, 1957). Before giving the optimality condition, we first introduce some preliminaries.

Assumption 1. There exists uniquely a continuous function $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$, such that for any $x, p \in \mathbb{R}^n$, we have

$$H(x, p, c(x, p)) < H(x, p, q), \quad \forall q \in \mathbb{R}^m, \ q \neq c(x, p),$$

where

$$H(x, p, q) := p^T f(x, q) + r(x, q), \quad \forall x, p \in \mathbb{R}^n, \ \forall q \in \mathbb{R}^m.$$

Then, we define the Hamilton–Jacobi–Bellman (HJB) equation as

$$H(x, V_x^T(x), u_c(x)) = 0, \tag{4}$$

where $V \in \mathbb{V}, \mathbb{V} = \{V : \mathbb{R}^n \to \mathbb{R} \mid V$ is positive definite and continuously differentiable $\}$, $V_x(x) = \partial V(x)/\partial x \in \mathbb{R}^{1\times n}$, and

$$u_c(x) := \arg\min_{v\in\mathbb{R}^m}\{H(x, V_x^T(x), v)\}, \quad \forall x \in \mathbb{R}^n. \tag{5}$$

Next, the concept of admissible control policy is introduced.

Definition 1. For system (1), a control policy $u_c : \mathbb{R}^n \to \mathbb{R}^m$ is called *admissible* with respect to the cost (2), if

- u_c is piecewise continuous;
- system (1) with $u = u_c(x)$ is GAS at the origin;
- $\int_0^\infty r(x, u_c(x))dt < \infty$ for all $x_0 \in \mathbb{R}^n$.

Denote $\mathbb{D}$ as the set of all admissible control policies. In Afanasiev, Kolmanovskii, and Nosov (1996), the following sufficient optimality condition is given:

Theorem 1 (*Afanasiev et al., 1996*)**.** *Consider system (1). Suppose there exists $V \in \mathbb{V}$, such that (4) and (5) are satisfied, and $u_c \in \mathbb{D}$. Then $V(x) = J(x; u^*)$, and $u^*(t) = u_c(x(t))$.*

Denote the functions V and u_c satisfying Theorem 1 as V^* and u_c^*, respectively. If Assumption 1 is also satisfied, then the pair (V^*, u_c^*) is unique, and u_c^* is continuous.

The HJB equation (4) is a nonlinear partial differential equation, and therefore its analytical solution is difficult to be found in most cases. In the following section, we give iteration-based methods to numerically approximate the optimal control policy.

3. Policy iteration-based optimal control design

3.1. Systems with known dynamics

In this subsection, a PI method for nonaffine continuous-time systems (Bellman, 1961; Leake & Liu, 1967) is adopted to provide a numerical solution for Problem 1. Different from the successive approximation method given in Bellman (1961), Howard (1960) and Leake and Liu (1967), a comprehensive stability analysis is included in our PI method.

Given an initial control policy $u_c^0 \in \mathbb{D}$, the PI method can be summarized in the following algorithm:

Algorithm 1. Policy iteration for nonaffine continuous-time systems

1. (**Policy evaluation**) Solve for V^j from

$$H(x, (V_x^j(x))^T, u_c^j(x)) = 0, \quad V^j(0) = 0, \tag{6}$$

where $u_c^j \in \mathbb{D}$.