Neurocomputing 171 (2016) 379–386

Contents lists available at ScienceDirect

Neurocomputing

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

Infinite-time stochastic linear quadratic optimal control for unknown discrete-time systems using adaptive dynamic programming approach ☆

Tao Wang [a,b], Huaguang Zhang [a,*], Yanhong Luo [a]

[a] School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, PR China
[b] Department of Computer and Mathematics Teaching, Shenyang Normal University, Shenyang, Liaoning 110034, PR China

ARTICLE INFO

Article history:
Received 22 March 2015
Received in revised form
30 May 2015
Accepted 24 June 2015
Communicated by Rongni Yang
Available online 4 July 2015

Keywords:
Stochastic linear quadratic optimal control
Adaptive dynamic programming
Control gain matrix
Stochastic algebra equation
Neural networks

ABSTRACT

In this paper, an adaptive dynamic programming (ADP) algorithm based on value iteration (VI) is proposed to solve the infinite-time stochastic linear quadratic (SLQ) optimal control problem for the linear discrete-time systems with completely unknown system dynamics. Firstly, the SLQ control problem is converted into the deterministic problem through system transformation and then an iterative ADP algorithm is introduced to solve the optimal control problem with convergence analysis. Secondly, for the implementation of the iteration algorithm, a neural network (NN) is used to identify the unknown system and then the other two NNs are employed to approximate the cost function and the control gain matrix. Lastly, the effectiveness of the iterative ADP approach is illustrated by two simulation examples.

1. Introduction

The stochastic linear quadratic (SLQ) optimal control problem was pioneered by Wonham [1] and had rapid development in both theory and application [2–6], which played an important role in modern control theory. Compared with the deterministic case, it is more complicated because the process of solving the stochastic algebra equation (SAE) for the SLQ optimal control problem is more complex than the Riccati equation (RE) [7–9]. Some intrinsic relations between SAE and RE are discussed in [10]. After the generalized RE (GRE) was introduced, the feasibility of the SLQ optimal control problem is equal to the solvability of the GRE [11]. It becomes easier and easier for solving the SLQ optimal control problem because of the introduction of linear matrix inequalities, semidefinite programming, Lagranger multiplier theorem and the common iterative method [12–14]. However, the algorithms mentioned above are all based on the prerequisite that the knowledge of the system dynamics must be known in advance. When the system dynamics is completely unknown, the methods are all invalid. So it presents a challenge to solve the SLQ optimal control problem without the knowledge of the system model.

Adaptive dynamic programming (ADP) as a very powerful tool in solving the deterministic optimal control problem gains comprehensive attentions [15–18] and the rapid development has been made recently [19–24]. Werbos classified ADP as follows: heuristic dynamic programming (HDP); dual heuristic programming (DHP); action-dependent heuristic dynamic programming (ADHDP); action-dependent dual heuristic programming (ADDHP) [15]. The value iteration (VI) ADP algorithm with the convergence proof was proposed solving the optimal control problem for the general nonlinear discrete-time systems [7]. Vrabie [25] adopted the policy iteration (PI) HDP scheme to approximate the optimal control for the partly unknown continuous-time systems. Based on [25], an intelligent optimal control scheme based on ADP for the completely unknown nonlinear discrete-time system was developed [26].

In this paper, we will tackle the SLQ optimal control problem for the completely unknown stochastic linear discrete-time systems. It is clear that solving the optimal control problem through the solution of the SAE needs the knowledge of the system dynamics. In order to avoid solving directly the SAE, the VI ADP algorithm with convergence proof is introduced. For the implementation of the algorithm for the unknown systems, three neural networks (NNs) are employed to approximate the system model, the cost function and the control gain matrix. At last, the comparison between the ADP algorithm and the analytical algorithm is given to indicate the validity of the ADP algorithm.

☆This work was supported by the National Natural Science Foundation of China (61433004) and the Fundamental Research Funds for the Central Universities (N140404004).

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2015.06.053
0925-2312/

This paper is organized as follows. In Section 2, we present the formulation of the problem and the well-posedness of the SLQ optimal control problem. In Section 3, the iterative ADP algorithm with convergence proof is given. In Section 4, the neural network implementation for the iterative scheme is presented. In Section 5, two simulation examples and the convergence analysis are provided to demonstrate the effectiveness of the iterative scheme. In Section 6, the concluding remarks are given.

2. Problem statement and preliminaries

Consider the following stochastic linear discrete-time systems as follows:

$$x(t+1)=(Ax(t)+Bu(t))+(Cx(t)+Du(t))\omega(t) \tag{1}$$

where $x(t)\in R^n$ is the system state vector, $u(t)\in R^m$ is the control input vector and $x(0)=x_0$ is the system initial state vector. $A,C\in R^{n\times n}$ and $B,D\in R^{n\times m}$ are the system constant real matrices. Let $(\Omega,\mathcal{F},\mathcal{P},\mathcal{F}_t)$ be a given filtered complete probability space with a standard one-dimensional Brownian motion $\omega(t)$ on $\{0,1,2,\cdots\}$ (with $\omega(0)=0$), where $\mathcal{F}_t=\sigma\{\omega(t)|t=0,1,2,\cdots\}$ denotes the σ-algebra generated by $\omega(t)$. We assume that the initial state x_0 is independent of $\omega(t)$, $t=0,1,2,\ldots$.

It is desired to find a control input $u:R^n\to R^m$ to minimize the following quadratic cost function given by

$$J(x_0,u)=E\sum_{t=0}^{\infty}[x'(t)Qx(t)+u'(t)Ru(t)] \tag{2}$$

where $Q\in R^{n\times n}$ and $R\in R^{m\times m}$ are respectively positive semidefinite real matrix and positive definite real matrix, $E(\cdot)$ denotes the mathematical expectation.

Definition 1 (*Mustapha and Zhou [12]*). A feedback control $u(t)$ is called mean-square stabilizing at $x_0$if there exists a linear feedback control $u(t)=Kx(t)$ (with K a constant real matrix) such that the corresponding state of the system (1) with respect to every initial state x_0 satisfies $\lim_{t\to\infty}E[x'(t)x(t)]=0$, where K is called a mean-square stabilizing control gain matrix.

Definition 2 (*Mustapha and Zhou [12]*). The system (1) is called mean-square stabilizable if there exists a mean-square stabilizing feedback control for the system (1).

Definition 3 (*Liu et al. [14]*). A feedback control $u(t)$ is said to be admissible if $u(t)$ satisfies the following: (1) $u(t)$ is a $\mathcal{F}_t$-adapted and measurable stochastic process; (2) $E\sum_{t=0}^{\infty}\|u(t)\|^2<+\infty$; (3) $u(t)$ is mean-square stabilizing. Let U_{ad} denote an admissible control set which contains all admissible controls.

The SLQ optimal control problem is how to choose an admissible control to make the cost function (2) reach the minimum value for the system (1) with respect to the initial state x_0, namely

$$V^*(x_0)=\inf_{u\in U_{ad}}\{J(x_0,u)\} \tag{3}$$

Definition 4 (*Liu et al. [14]*). The SLQ optimal control problem is well posed if

$$-\infty<V^*(x_0)<+\infty,\quad \forall x_0\in R^n$$

For the SLQ control problem (1) and (2) we will find the optimal control only in the class of linear feedback controls, whose forms are given by

$$u(t)=Kx(t),\quad K\in R^{m\times n} \tag{4}$$

Taking (4) into (1) the corresponding system model is given by

$$x(t+1)=(A+BK)x(t)+(C+DK)x(t)\omega(t) \tag{5}$$

and the system (5) for $V=A+BK$ and $W=C+DK$ is simplified as

$$x(t+1)=Vx(t)+Wx(t)\omega(t) \tag{6}$$

Let $X(t)=E(x(t)x'(t))$, then the system (6) by taking expectation on $x(t+1)x'(t+1)$ is converted into the deterministic difference equation

$$X(t+1)=VX(t)V'+WX(t)W' \tag{7}$$

and the cost function (2) is transformed into

$$J(X_0,K)=\mathrm{tr}\left\{\sum_{t=0}^{\infty}(Q+K'RK)X(t)\right\} \tag{8}$$

where $X_0=E(x(0)x'(0))$ denotes the initial state of the system (7).

Before solving the SLQ optimal control problem, it is necessary to know whether it is well posed. First, we provide the following lemma.

Lemma 1. *If the feedback control $u(t)=Kx(t)$ is admissible, then the SLQ control problem is well posed and the corresponding cost function with respect to the initial state x_0 is $J(x_0,u)=x_0'Px_0$, where the symmetric matrix P satisfies the following SAE:*

$$P=(A+BK)'P(A+BK)+(C+DK)'P(C+DK)+Q+K'RK \tag{9}$$

Proof. Let the feedback control $u(t)=Kx(t)$ be admissible, the cost function (2) is converted into

$$J(x_0,u)=E\sum_{t=0}^{\infty}x'(t)(Q+K'RK)x(t)$$

According to the fact that

$$\begin{aligned}
&E\sum_{t=0}^{\infty}(x'(t+1)Px(t+1)-x'(t)Px(t))\\
&\quad=E\sum_{t=0}^{\infty}[(Vx(t)+Wx(t)\omega(t))'P(Vx(t)+Wx(t)\omega(t))-x'(t)Px(t)]\\
&\quad=E\sum_{t=0}^{\infty}x'(t)(V'PV+W'PW-P)x(t)
\end{aligned}$$

where the matrix P satisfies the SAE, we can see that

$$\begin{aligned}
J(x_0,u)&=E\sum_{t=0}^{\infty}x'(t)(P-V'PV-W'PW)x(t)\\
&=-E\sum_{t=0}^{\infty}(x'(t+1)Px(t+1)-x'(t)Px(t))\\
&=x_0'Px_0-\lim_{t\to\infty}E(x'(t)Px(t))
\end{aligned}$$

Since the feedback control $u(t)$ is admissible, we can obtain $\lim_{t\to\infty}E(x'(t)Px(t))=0$ and $J(x_0,u)=x_0'Px_0$.□

Remark 1. When the system (1) is mean-square stabilizable, the solution of the SAE will make the well-posedness of the SLQ control problem.

In order to make sure the well-posedness of the SLQ optimal control problem, we make the following assumption.

Assumption 1. The system (1) is mean-square stabilizable.