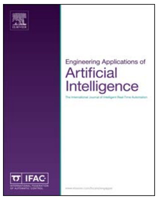
Engineering Applications of Artificial Intelligence 63 (2017) 1–19

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Numerical stability improvements of state-value function approximations based on RLS learning for online HDP-DLQR control system design

Ernesto F.M. Ferreira [a,*], Patrícia H.M. Rêgo [b], João V.F. Neto [a]

[a] UFMA - Cidade Universitaria Dom Delgado, Brazil
[b] UEMA - Cidade Universitaria Paulo VI, Brazil

ARTICLE INFO

Article history:
Received 10 June 2016
Received in revised form
27 January 2017
Accepted 22 April 2017

Keywords:
Heuristic dynamic programming
Discrete linear quadratic regulator
Recursive least squares
QR Decomposition
Numerical stability

ABSTRACT

In order to overcome numerical stability problems that inherently occur in the recursive least-squares (RLS)-based adaptive dynamic programming paradigms for online optimal control design, a novel method to promote improvements in the state-value function approximations for online algorithms of the discrete linear quadratic regulator (DLQR) control system design is proposed. The algorithms resulting from that methodology are embedded in actor-critic architectures based on heuristic dynamic programming (HDP). The proposed solution is grounded on unitary transformations and $\mathcal{QR}$ decomposition, which are integrated in the critic network, to improve the RLS learning efficiency for online realization of the HDP-DLQR control design. In terms of numerical stability and computational cost, the developed learning strategy is designed to provide computational performance improvements, which aim at making possible the real time implementations of optimal control design methodology based upon actor-critic reinforcement learning paradigms. The convergence behavior and numerical stability of the proposed online algorithm are evaluated by computational simulations in two multiple-input and multiple-output models that represent a fourth order RLC circuit with two input voltages and two controllable voltage levels, and a doubly-fed induction generator with six inputs and six outputs for wind energy conversion systems.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

During several decades, the traditional dynamic programming has been recognized as computationally intractable for solving large-scale multi-stage optimal control problems because of the computational complexity associated with the Hamilton-Jacobi-Bellman (HJB) equation solution, referred to as “curse of dimensionality” (Lendaris, 2009). This phenomenon makes the online controllers design unfeasible, as the dynamic programming require a “backward in time” procedure to solve HJB equation.

In contrast, the approximate dynamic programming (ADP), through the well-known adaptive critic designs (ACD) (Werbos, 2012), proposes an online design feasibility of the optimal control systems that combines function approximation and simulation methods, such as temporal differences, to bypass the curse of dimensionality (Bertsekas, 2012). Such an approach employs a “forward in time” mechanism which searches for an optimal control policy (optimal feedback control) by adapting two parametric structures, i.e., an action network and a critic network to approximate the HJB equation solution. The action network computes control actions while the critic network learns to approximate the value function. This function evaluates the effect of the control on its future performance, and it provides guidance on how to improve the control law. Those structures have been used as a model-free learning framework that solves optimal decision problems in real time, without requiring knowledge of the full system dynamics model.

ADP has shown numerous applications in online optimal control design, such as high performance aircrafts or missile systems (Bertsekas et al., 2000), autopilot (Ferrari and Stengel, 2004; Chuan, 2005), and robotics (Khan and Lewis, 2012). A new method for online optimal control systems design for wind power and solar system based on ADP paradigms has been developed (Fonseca Neto et al., 2013). Successful applications of ADP are presented in Weber et al. (2008), and references therein, including learning to control mobile robots, autonomous helicopters, industrial PH, oxidation plants, and air traffic management. An overview of ADP techniques and their advances focused on adaptive optimal control is presented in Khan and Lewis (2012). Important surveys are given in Lewis and Vrabie (2009), Lin et al.

[*] Corresponding author.
E-mail address: efmf86@gmail.com (E.F.M. Ferreira).

http://dx.doi.org/10.1016/j.engappai.2017.04.017
0952-1976/© 2017 Elsevier Ltd. All rights reserved.

(2009), Wang et al. (2009) and Buşoniu et al. (2011).

In ADP-based control design methodologies, among the proposed iterative algorithms to estimate the value function parameters, recursive least-squares (RLS) is one of the most successful. Such favorable outcome is mainly due to its robustness to cope with time variations in the regression parameters and fast convergence speed when compared with stochastic gradient methods. In the reinforcement learning (RL) and online optimal control system design context (Martijn van Otterlo and Wiering, 2012), the main references supporting the development of adaptive critic schemes (Werbos, 1992), and training algorithms based on recursive least-squares (Wittemmak, 1989), are presented. In terms of RLS learning for solving the HJB-Riccati equation in optimal control problems, which are solved by the reinforcement learning paradigms, the authors Rêgo et al. (2013) developed methods and algorithms based on RLS training for the online design of discrete linear quadratic regulator (DLQR)-type optimal controllers (Queiroz et al., 2015), and their evaluations of the feasibility obtained through wind generator models (Queiroz et al., 2015), and multivariable dynamic systems (Fonseca Neto and Lopes, 2011; Ferreira et al., 2016).

Thus, the development of novel algorithms (Silva et al., 2014; Queiroz et al., 2014), for the online optimal control system applications is made viable. According to the methods proposed in Daniel and Gajski (2009) and Lee and Seshia (2015), and aiming at attending to the specific demand of a given process such as: the generation of alternative energies or industrial processes, the algorithms for the online design of optimal control methods developed by Fonseca Neto et al. (2013) may be embedded in devices with a control architecture dedicated to a specific dynamic system.

Broadly, the training methods based on the least squares (LS) learning and their applications are highlighted. In reference Chen (1995), the RLS learning is customized as a training algorithm of a radial basis functions (RBF)-type network in the time series modeling and prediction. In Gokhale and Nawghare (2004), the modified RLS algorithm is used for training a neural network using piecewise linear functions. A convergence study of the LS learning was developed by Marcet and Sargent (1989), where the authors investigate the LS properties in stochastic linear models. In Xu et al. (2002), the authors present the development of RLS-based training algorithms to solve problems of linear value function approximation via RL, such as: learning algorithm of the multistage and temporal difference type.

Furthermore, a large number of developments and applications of RLS methods can be found in the literature, some of which deal with neural network training problems (see e.g., Dua and Zhai, 2008; Yeh et al., 2010; Yeh and Su, 2012). In Chua et al. (2008) and Cheng et al. (2013), the authors explore RLS methods to solve actor-critic reinforcement learning problems.

Recently, there has been a great deal of interest in issues related to numerical stability of RLS methods in the context of reinforcement learning and approximate dynamic programming for online optimal control. Some work toward this study can be found in our previous paper, Rêgo et al. (2013), in which we presented a proposal to solve, via UDU^T factorization, numerical problems that arise due to covariance matrix ill-conditioning of the RLS approach for approximating value function in the online DLQR optimal control and heuristic dynamic programming (HDP) framework.

It has been observed that the online design becomes unstable during the RLS estimation process for determining the decision actions, Rêgo et al. (2013). Such situation is characterized as a numerical stability problem associated with the RLS covariance matrix originating in the nature or formation of regressor vectors, which are governed by the process, generating linearly dependent or weakly linearly independent vectors when associated with reinforcement learning and optimal control. Thus, the proposed solution method in Rêgo et al. (2013) is seen as an improvement in the RLS estimation process of DLQR optimal decision policies, since the UDU^T factorization circumvents problems related to loss of positivity of the RLS covariance matrix.

Motivated by that research result, the present paper is concerned with developing a novel solution procedure for numerical instability problems inherent in the implementation of RLS-based HDP methods for online DLQR optimal control. The proposed solution method is based on unitary transformations and $\mathcal{QR}$ decomposition, which are employed in the critic network, to improve the RLS learning efficiency for realization of online DLQR control design. The resulting method is called RLS_{μ}-$\mathcal{QR}$-HDP-DLQR algorithm. The convergence and numerical stability of the proposed algorithm are evaluated with respect to the behavior of the covariance matrix of the RLS estimation process for the online solution of the algebraic Riccati (ARE)-type HJB equation underlying the DLQR problem.

In general, the importance of these investigations is in the scope of applications of reinforcement learning for online optimal control design via real time solution of HJB equation. A large body of research about ADP-LQR designs has reported theoretical results of stability and convergence for several adaptive critic schemes (see, e.g., Bradtke et al., 1994; Landelius and Knutsson, 1996; Al-Tamimi et al., 2007; Vamvoudakis and Lewis, 2010; Jiang and Jiang, 2010; Lee et al., 2010; Feng et al., 2016). However, numerical stability of ADP-based optimal control methods is an issue that has been little discussed in literature.

From a numerical stability analysis point of view, we show the need for better implementations of RLS-based HDP methodology for online DLQR optimal control. It has been observed that the problem of numerical stability loss occurs mainly during the steady state behavior of the dynamic system, while during the transient state behavior, one has enough information to span the basis of regressor vectors. Herein, a new method is presented dealing with numerical issues. The proposed method not only enhances numerical stability but also substantially reduces the computational effort spent on approximating the DLQR cost function. This allows for further real-time applications of ADP-based optimal controllers.

With that purpose in mind, this work is structured as follows: Section 2 provides, briefly, the fundamentals to assemble the framework of online optimal control design which is supported in Bellman equation formulation, greedy policy iteration principle, and function approximation. Section 3 presents the method of online optimal control system design, which is based on the adaptive critic approach, where besides the controller gain being self-adjustable, the value function computation is fully independent of plant model. Numerical stability and computational cost issues of the RLS_{μ}-HDP-DLQR and RLS_{μ}-$\mathcal{QR}$-HDP-DLQR algorithms with respect to the RLS estimation problem, the calculation problem of the matrix Φ inverse, and the RLS-$\mathcal{QR}$ decomposition for approximating the DLQR state-value function, under greedy policy iterations, are also discussed in this section. In addition, simulation results that verify the performance of the proposed algorithm in this work are presented in Section 4. Such results are analyzed under the viewpoint of convergence and numerical stability of the RLS methods for approximating DLQR state-value function. Finally, conclusions and comments are contained in Section 5.

2. ADP-DLQR framework

The following topics assemble the framework of online optimal control design which is based on discrete LQR policy via solution

Download English Version:

https://daneshyari.com/en/article/4942682

Download Persian Version:

https://daneshyari.com/article/4942682

Daneshyari.com