Neurocomputing 285 (2018) 10–22

Contents lists available at ScienceDirect

Neurocomputing

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

Sub-optimal scheduling in switched systems with continuous-time dynamics: A gradient descent approach

Tohid Sardarmehni*, Ali Heydari

Department of Mechanical Engineering, Southern Methodist University, Dallas, TX 75205, USA

ARTICLE INFO

Article history:
Received 27 February 2017
Revised 19 December 2017
Accepted 8 January 2018
Available online 3 February 2018

Communicated by Prof. Huaguang Zhang

Keywords:
Optimal control
Switched systems
Policy iteration
Continuous-time dynamics

ABSTRACT

A feedback solution for approximate optimal scheduling of switched systems with autonomous subsystems and continuous-time dynamics is presented. The proposed solution is based on policy iteration algorithm which provides the optimal switching schedule. Algorithms for offline, online, and concurrent implementation of the proposed solution are presented. For online and concurrent training, gradient descent training laws are used and the performance of the training laws is analyzed. The effectiveness of the presented algorithms is verified through numerical simulations.

1. Introduction

In this study, a class of hybrid systems [1] comprised of a finite number of subsystems/modes with continuous-time (CT) dynamics, and a switching rule is considered. In such switched systems, the switching rule assigns the switching instants and active modes [2–4]. Moreover, the modes are considered to be autonomous which means no continuously varying control exists in the modes [4]. Hence, the only control input is switching among modes. This definition of switched systems embraces many interesting engineering problems [5–9].

Due to discontinuous nature of the problem, deriving optimal solutions for control of the switched systems is a challenging task [4,10,11]. From mathematical point of view, the necessary and sufficient condition for optimality is provided by the underlying Hamilton–Jacobi–Bellman (HJB) equation [12]. However, solving the HJB equation analytically is difficult and generally impossible [13]. The existing solutions for the optimal control problems such as calculus of variations or Dynamic Programming (DP) are intractable in highly nonlinear systems [12,14] due to *curse of dimensionality* [12,15,16]. To remedy the mentioned problem in optimal control, Approximate Dynamic Programming (ADP) was introduced which approximates the optimal solution [17,18]. In general, ADP uses function approximators, such as neural networks, to approximate optimal cost-to-go (value function), namely *critic*, and sometimes optimal policy, namely *actor*. The backbone of ADP is application of iterative methods to tune the unknown parameters of the mentioned function approximators in order to approximate the optimal value function which solves the HJB equation [19–23].

Policy iteration (PI) is an iterative algorithm which is frequently used in the ADP methods to find the optimal solutions [14,24–29]. In general, the evolving policies generated by PI algorithm are known to be stabilizing [30]. Mathematically, the proof of convergence in online application of PI algorithm in systems with CT dynamics is either based on satisfaction of Persistency of Excitation (PE) condition [14,27,31,32] or application of carefully selected data along with on-trajectory data [28,29]. The latter is called concurrent training and was developed to relax the PE condition since this condition is restrictive and generally cannot be verified online [33].

The ADP-based solutions for optimal control of the switched systems were studied in [10,11,34–37] for discrete-time (DT) dynamics and in [35,38–40] for CT dynamics. In [35], the problem of optimal scheduling in the switched systems with CT dynamics and homogenous subsystems was solved with a Value Iteration (VI) algorithm. In [39], an optimal scheduler was developed based on a PI algorithm for switched systems with controlled subsystems where the dynamics of the subsystems include both continuous state and control signals. The presented algorithm trained two different neural networks as actor and critic. In another work, an optimal tracking scheduler was developed in [38]. The proposed scheduler formulated a PI algorithm which also trained two

* Corresponding author.
E-mail addresses: tsardarmehni@smu.edu (T. Sardarmehni), aheydari@smu.edu (A. Heydari).

https://doi.org/10.1016/j.neucom.2018.01.003
0925-2312/

neural networks as actor and critic. The output of the actor was a continuous signal which needed to be discretized for scheduling. Hence, the controller used a hard limiter function which receives the continuous output of the actor and discretizes it, to select the proper mode. In [40], a PI algorithm was developed for optimal scheduling in the switched systems where online training law was derived based on recursive least squares. The comparison between the above-mentioned studies and the current one is given in the sequel.

The challenging nature of the switched system control, lack of sufficient theoretical investigations with ADP methods for this problem in CT dynamics, and the large engineering application of this control problem are the main motivations of this paper. The main contributions of the present study are as follows[1].

- Convergence, stability of evolving control policies, and formulation of a PI algorithm for the switched systems are presented.
- A new approach for proof of convergence of PI algorithm in systems with CT dynamics is presented.
- Derivations for extending the results to optimal tracking controllers in the switched systems are given.
- A concurrent training algorithm is presented to implement the PI algorithm for the switched systems.

Before presenting the main results, some comparisons between this paper and recent relevant studies are given. Compared to [35], the present work deals with general class of switched systems with autonomous subsystems rather than special class of switched systems with homogenous subsystems. Meanwhile, the derivations of the approximate optimal solution in the present study is based on PI algorithm. This is one of the differences between this work and [35] which uses VI as another capable learning algorithm. Compared to [39], the present study deals with switched system with autonomous subsystems where in [39] switched systems with controlled subsystems were studied. Compared to [38], the present study explicitly provides the optimal switching schedule. This scheduling is completely different from the procedure used in [38] for assigning active modes through using a hard limiter function to discretize the output of the actor. At last, the present study uses gradient descent for online training with exponential convergence rate. However, in [40] recursive least squares was used for which the convergence rate is not exponential. Meanwhile, the convergence and stability of the PI algorithm for the switched systems are investigated in this paper which were missing in [40].

The rest of the paper is organized as follows. In Section 2, notations are introduced. In Section 3, the optimal switching problem is formulated and the proposed solution is discussed in Section 4. In Section 5, the proposed PI algorithm is introduced and its convergence to the optimal solution is analyzed. The implementation of the proposed algorithm with offline, online, and concurrent training methods are discussed in Section 6 and simulation results are presented in Section 7. At last, Section 8, concludes the paper.

2. Notations

Throughout the paper, $\mathbb{R}$ denotes the real numbers. The set of real n-vectors is denoted with $\mathbb{R}^n$ and set of real $n \times m$ matrices are denoted with $\mathbb{R}^{n\times m}$. The absolute value of a scalar $s \in \mathbb{R}$ is denoted by $|s|$. The Euclidean norm of a vector $v \in \mathbb{R}^n$ is denoted by $\|v\|$ and the 2-norm of a matrix $M \in \mathbb{R}^{n\times m}$ is denoted by $\|M\|$. The transpose operator is denoted by $(.)^T$ and $\lambda_{min}(.)$ represents the minimum eigenvalue of its argument. At last, the gradient of a vector is defined as a column vector and denoted by $\nabla \equiv \frac{\partial}{\partial x}$.

[1] The preliminary results of this research were presented in 2016 International Joint Conference on Neural Networks (IJCNN 2016) [41].

3. Problem formulation

Nonlinear dynamics of a switched system with autonomous subsystems can be presented as

$$\dot{x}(t) = f_\omega\big(x(t)\big), \quad \omega \in \mathcal{V} = \{1, 2, ..., M\}, \quad x(0) = x_0, \tag{1}$$

where $x \in \mathbb{R}^n$ is the state vector and t denotes the time. The active mode in the system is denoted by subscript ω and $f_\omega : \mathbb{R}^n \to \mathbb{R}^n$ denotes the dynamics of the active mode. Meanwhile, the set of all subsystems/modes that can be selected for the operation of the system is denoted by $\mathcal{V}$ and parameter M is the number of subsystems in the system. Considering $\Omega \subset \mathbb{R}^n$ as the region of interest that includes the origin, it is assumed that each mode $f_\omega(.)$ is *Lipschitz* continuous in Ω and there exists at least one mode for which $f_\omega(0) = 0$. For the operation of the system, the active mode ω may be selected by a feedback control law (scheduler) denoted by $v(.)$, such that at each time t only one mode ω is active.

The selected infinite horizon cost function for the switching problem is given by

$$J(x_0) = \int_0^\infty Q(x(\tau))d\tau, \tag{2}$$

where $x_0 = x(0)$ and $Q(.) : \mathbb{R}^n \to \mathbb{R}$ is a positive definite function. The objective is finding a stabilizing switching schedule in a feedback form, i.e., $v(x)$, such that the cost function in (2) is minimized subject to dynamics presented in (1).

4. Proposed solution

Considering (2), and a given policy, the cost-to-go, denoted with $V : \mathbb{R}^n \to \mathbb{R}$, can be defined as

$$V\big(x(t)\big) = \int_t^\infty Q(x(\tau))d\tau. \tag{3}$$

Considering a time interval $[t, t+\delta t]$, one has

$$V\big(x(t)\big) = \int_t^{t+\delta t} Q\big(x(\tau)\big)d\tau + V\big(x(t+\delta t)\big). \tag{4}$$

For notational brevity, hereafter $x = x(t)$ unless otherwise stated. Using the Bellman principle of optimality [12], the optimal cost-to-go, $V^* : \mathbb{R}^n \to \mathbb{R}$, can be defined as

$$V^*(x) = \min_{v(.)} \left(\int_t^{t+\delta t} Q\big(x(\tau)\big)d\tau + V^*\big(x(t+\delta t)\big)\right), \tag{5}$$

where the policy given by $v(.)$ is exploited to propagate the states from t to $t+\delta t$. As $\delta t \to 0$, the optimal switching policy can be given by

$$v^*(x) = \arg\min_{v\in\mathcal{V}} \left(\int_t^{t+\delta t} Q\big(x(\tau)\big)d\tau + V^*\big(x(t+\delta t)\big)\right). \tag{6}$$

In order to derive the infinitesimal form of the Bellman equation, the following assumption is required.

Assumption 1. The value function is continuously differentiable, i.e., $V(.) \in C^1$, [4,39,42].

Remark 1. Besides serving the purpose in derivation of the desired equations, Assumption 1 makes application of function approximators, such as neural networks, possible in order to *uniformly* approximate the value function [43]. In general, differentiability or even continuity of value functions in optimal control problems is not clear [25]. Differentiability and twice differentiability of the value functions are studied in [44–48] for some classes of optimal control systems. For Hybrid and switched systems, continuity of the value functions was investigated in [10,35,49–51] for