



Optimal scheduling for reference tracking or state regulation using reinforcement learning

Ali Heydari

Mechanical Engineering Department, South Dakota School of Mines and Technology, Rapid City, SD 57701, United States

Received 11 March 2014; received in revised form 5 November 2014; accepted 17 November 2014

Abstract

The problem of optimal control of autonomous nonlinear switching systems with *infinite-horizon* cost functions, for the purpose of tracking a family of reference signals or regulation of the states, is investigated. A reinforcement learning scheme is presented which *learns* the solution and provides scheduling between the modes in a feedback form without enforcing a mode sequence or a number of switching. This is done through a value iteration based approach. The convergence of the iterative learning scheme to the optimal solution is proved. After answering different analytical questions about the solution, the learning algorithm is presented. Finally, numerical analyses are provided to evaluate the performance of the developed technique in practice.

© 2014 The Franklin Institute. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Optimal scheduling between different modes/subsystems in control of *switching systems* is a challenging problem in controls engineering discipline and numerous research papers have emerged in the literature within the last decade in this regard, [1–14]. The reason for this attention and the vast effort in solving these problems is the fact that many real-world control problems can be classified as switching problems, including problems in mechanical and aerospace systems [15,16], electronics [17], chemical processes [18], and bioengineering [12,19]. Conventional optimal control methods fail to provide solution to switching problems, generally. It should be noted that the solution includes discrete *decisions*, in switching

E-mail address: ali.heydari@sdsmt.edu

<http://dx.doi.org/10.1016/j.jfranklin.2014.11.008>

0016-0032/© 2014 The Franklin Institute. Published by Elsevier Ltd. All rights reserved.

problems. In other words, the solution provides optimal decisions, in terms of suitable switching between the modes.

One of the most attractive approaches to solving optimal switching problems is freezing the *mode sequence*, i.e., the order of active modes, as well as the *number of switches*, and optimizing the *switching times*. Note that once the mode sequence and the number of switches are fixed, the unknown is the switching time. Nonlinear programming is an approach followed by different researchers [1–7], in which, the gradient of the cost function with respect to the switching instant is utilized for optimizing the switching instant/time with a pre-selected mode sequence and number of switching. Some ideas were presented in [5,6] for admitting free mode sequence conditions. In [5], a two stage optimization algorithm was developed where in one stage the switching time is updated and at another stage the mode sequence is modified. In a recent study, another nonlinear programming based solution was proposed in [20] for the case of free mode sequence. Nonlinear programming based methods generally lead to open loop solutions for a given/fixed initial condition. Each time the initial condition is changed, another set of numerical calculations are required to be conducted in order to find the new optimal switching times. The dependency of the solutions on the selected initial conditions leads to the limitation that, for example in finding the optimal switching between gears in a manual transmission car in order to accelerate to a desired speed, a calculated solution will be valid for implementation only if the initial speed of the car is exactly the same as the one using which the problem was numerically solved. Otherwise the solution will not take the car to the desired speed. In [3] the validity of the results was extended to different initial conditions within a pre-selected set through determining the switching parameter such that it minimizes the worst possible cost for all trajectories starting in the selected set of initial states. Discretization of state space is an approach followed by [8], in order to end up with a finite number of choices, and dynamic programming was used for solving the problem. Refs. [21,22] investigated the use of (relaxed/approximate) dynamic programming for different problems including optimal switching. Genetic algorithm is also another approach for finding a numerical solution for a given initial condition [9]. An optimization scheme was developed in [19] to find both the optimal mode sequence and the switching time for positive linear systems.

The demonstrated potential of Reinforcement Learning (RL) and Approximate Dynamic Programming (ADP) in solving conventional optimal control problems [23–37] motivated the author of this study to utilize ADP for solving optimal switching problems in the past. The results were solutions to problems with fixed switching sequence [38], free switching sequence with autonomous subsystems [39], free switching sequence with controlled subsystems [40], and the applications of the developed ideas to multi-therapeutic treatment of HIV disease [12] and aerospace vehicles [16]. All these developments, however, deal with problems with fixed-final-time, i.e., problems with finite-horizon cost functions. Many real-world problems, however, have infinite horizon, e.g., regulation of a system with on-off actuators. The motivation behind this work is providing a solution for such problems.

In a simultaneous, but independent research, Refs. [13,41] proposed a different ADP based solution to switching problems. The number of functions needed to be learned at each training iteration grows exponentially with the number of iterations and soon becomes prohibitive, in the proposed method. Moreover, in these developments the training is done for a single selected initial condition. Another investigation for solving switching problems using ADP was reported in [14]. The differences, compared with this study, are the approach and the point that the initial conditions are assumed to be known a priori in that study.

Download English Version:

<https://daneshyari.com/en/article/4974878>

Download Persian Version:

<https://daneshyari.com/article/4974878>

[Daneshyari.com](https://daneshyari.com)