J. Math. Anal. Appl. 435 (2016) 321–337

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Near optimality of quantized policies in stochastic control under weak continuity conditions ☆

Naci Saldi *, Serdar Yüksel, Tamás Linder

Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada

ARTICLE INFO

Article history:
Received 13 May 2015
Available online 22 October 2015
Submitted by X. Zhang

Keywords:
Stochastic control
Quantization
Approximation
Partially observed Markov decision processes

ABSTRACT

This paper studies the approximation of optimal control policies by quantized (discretized) policies for a very general class of Markov decision processes (MDPs). The problem is motivated by applications in networked control systems, computational methods for MDPs, and learning algorithms for MDPs. We consider the finite-action approximation of stationary policies for a discrete-time Markov decision process with discounted and average costs under a weak continuity assumption on the transition probability, which is a significant relaxation of conditions required in earlier literature. The discretization is constructive, and quantized policies are shown to approximate optimal deterministic stationary policies with arbitrary precision. The results are applied to the fully observed reduction of a partially observed Markov decision process, where weak continuity is a much more reasonable assumption than more stringent conditions such as strong continuity or continuity in total variation.

1. Introduction

In this paper, we study the finite-action approximation of optimal control policies for a discrete time Markov decision process (MDP) with Borel state and action spaces, under discounted and average cost criteria. Various stochastic control problems may benefit from such an investigation.

The optimal information transmission problem in networked control systems is one such example. In many applications to networked control, the perfect transmission of the control actions to an actuator is infeasible when there is a communication channel of finite capacity between a controller and an actuator. Hence, the actions of the controller must be quantized to facilitate reliable transmission to an actuator. Although, the problem of optimal information transmission from a plant/sensor to a controller has been

☆ This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery Grant Program. Parts of this work were presented at 53rd IEEE Conference on Decision and Control (CDC), Los Angeles, CA, 2014.

* Corresponding author.
E-mail addresses: nsaldi@mast.queensu.ca (N. Saldi), yuksel@mast.queensu.ca (S. Yüksel), linder@mast.queensu.ca (T. Linder).

http://dx.doi.org/10.1016/j.jmaa.2015.10.008
0022-247X/

studied extensively [25], much less is known about the problem of transmitting actions from a controller to an actuator. Such transmission schemes usually require a simple encoding/decoding rule since an actuator does not have the computational/intelligence capability of a controller to use complex algorithms. Therefore, time-invariant uniform quantization is a practically useful encoding rule for controller–actuator communication.

The investigation of the finite-action approximation problem is also useful in computing near optimal policies and learning algorithms for MDPs. In a recent work [22], we consider the development of *finite-state* approximations for obtaining near optimal policies. However, to establish constructive control schemes, one needs to quantize the action space as well. Thus, results on approximate optimality of finite-action models pave the way for practical computation algorithms which are commonly used for finite-state/action MDPs. One other application regarding approximation problems is on learning a controlled Markov chain using simulations. If one can ensure that learning a control model with only finitely many control actions is sufficient for approximate optimality, then it is easier to develop efficient learning algorithms which allow for the approximate computation of finitely many transition probabilities. In particular, results developed in the learning and information theory literature for conditional kernel estimations [11] (with control-free models) can be applied to transition probability estimation for MDPs.

Motivated as above, in this paper we investigate the following approximation problem: For uncountable Borel state and action spaces, under what conditions can the optimal performance (achieved by some optimal stationary policy) be arbitrarily well approximated if the controller action set is restricted to be finite? We show that quantized stationary policies obtained by uniform quantization of the action space can approximate optimal policies with arbitrary precision for an MDP with an unbounded one-stage cost function, under a weak continuity assumption on the transition probability.

Various approximation results, which are somewhat related to our work, have been established for MDPs with Borel state and action spaces in the literature along the theme of computation of near optimal policies. For rather complete surveys of these techniques, we refer the reader to [1,5–7,16–19,24] and the references therein. With the exception of [17], these works assume in general restrictive continuity conditions on the transition probability. In [17], the authors considered an approximation problem in which all the components of the original model are *allowed* to vary in the reduced model (varying only the action space corresponds to the setup considered in this paper). Under weak continuity of the transition probability, [17] established the convergence of the reduced models to the original model for the discounted cost when the one-stage cost function is bounded. In this paper we allow the one-stage cost function to be unbounded. In addition, we also study the approximation problem for the challenging average cost case. Hence, our results can be applied to a wider range of stochastic systems. However, analogous with [17], the price we pay for imposing weaker assumptions is that we do not obtain explicit performance bounds in terms of the rate of the quantizer used in the approximations.

In [21] we solved a variant of this problem for the discounted cost under the following assumptions: (i) the action space is compact, (ii) the transition probability is setwise continuous in the action variable, and (iii) the one stage cost function is bounded and continuous in the action variable. The average cost was also considered under some additional restrictions on the ergodicity properties of Markov chains induced by deterministic stationary policies. In this paper we consider a similar problem for systems where the transition probability is weakly continuous in the state-action variables. An important motivation for considering these conditions comes from the fact that for the fully observed reduction of a partially observed MDP (POMDP), the setwise continuity of the transition probability in the action variable is a prohibitive condition even for simple systems such as the one described in Example 2.1 in the next section.

Organization: In Section 2 we give definitions and the problem formulation. The main result for discounted cost is stated and proved in Section 3. In Section 4 an analogous approximation result is obtained for the average cost criterion. In Section 5 the results for the discounted cost are applied to the fully observed reduction of POMDPs via appealing to results by Feinberg et al. [9]. Section 6 concludes the paper.