J. Math. Anal. Appl. 426 (2015) 978–985

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

On the vanishing discount factor approach for Markov decision processes with weakly continuous transition probabilities

Óscar Vega-Amaya

Departamento de Matemáticas, Universidad de Sonora, Luis Encinas y Rosales s/n, Col. Centro, C.P. 83000, Hermosillo, Sonora, Mexico

ARTICLE INFO

Article history:
Received 8 October 2014
Available online 7 February 2015
Submitted by V. Pozdnyakov

Keywords:
Markov decision processes
Average cost criterion
Vanishing discount factor approach

ABSTRACT

This note deals with average cost Markov decision processes with Borel state and control spaces, possibly unbounded costs and non-compact action subsets under the assumption of weak continuity of the transition law. It provides an elementary proof of the existence of average cost optimal stationary policies using the vanishing discount factor approach.

1. Introduction

The *vanishing discount factor approach* is a general procedure to address average cost (AC) optimal control problems by means of discounted problems when the discount factor tends to one. Its inception goes back to the early years of the discrete-time Markov decision processes (MDPs) theory, but it has been applied to a number of Markovian systems. For a very detailed account up to the early nineties see the survey paper [1], and for more recent results see [4,7,10,13,15,18,20–22]. Since then, besides the discrete-time MDPs, the vanishing discount factor approach has been successfully applied to discrete-time Markov games [19], semi-Markov decision processes [26], risk-sensitive MDPs [2,3,17], unconstrained and constrained continuous-time MDPs [11,12,23], continuous-time Markov control processes [14] and piece-wise deterministic MDPs [5]. The present note focusses on discrete-time MDPs with Borel spaces and weakly continuous transition law, in which the cost function may be unbounded and the action subsets may be noncompact sets as in the paper by Feinberg et al. [7].

Loosely speaking, the vanishing discount factor approach requires to work on two kinds of assumptions. The first one is "continuity/compactness" conditions, which are aimed to ensure either the measurability or lower semicontinuity of the optimal value functions as well as the existence of measurable minimizers. This

E-mail address: ovega@gauss.mat.uson.mx.

http://dx.doi.org/10.1016/j.jmaa.2015.02.007
0022-247X/

is done, of course, by means of measurable selection theorems or Berge's minimum theorems. The second kind of assumptions imposes some condition on the "relative discounted value functions" and essentially asks that such functions do not grow without bound as the discount factor tends to one (see, for instance, [4,10,20,21]). Then, the main idea in the vanishing discount factor approach consists in getting a solution to the average cost optimality inequality as a limit of the relative discounted value functions when the discount factor tends to one. This step is accomplished by means of suitable versions of Fatous' lemma for varying measures (see [9,16,25]).

The recent paper by Feinberg et al. [7] shows the existence of AC optimal stationary policies using the vanishing discount factor approach. They consider discrete-time MDPs with Borel spaces and weakly continuous transition law, in which the cost function may be unbounded and the admissible action subsets may be noncompact sets. More specifically, they assume the one-step cost function satisfies a kind of local inf-compactness property they call $\mathbb{K}$-inf-compactness (see, [6,8]). This kind of compactness condition seems to be the weakest one among all the conditions previously used in the MDP literature. For the second kind of assumptions they use a slightly weaker version of the condition introduced by Schäl [24].

The present note gives a simplified and somewhat elementary proof of the existence of average cost stationary optimal policies under the same assumptions made in the paper [7]. This is done using the concept of lower semicontinuous envelope of functions and an elementary result on the interchange of limits and minima in lieu of a Fatou's lemma for varying measures.

The remainder of the note is organized as follows. Section 2 introduces the Markov decision model and the performance criteria, namely, the average and the discounted cost criteria. The "continuity/compactness" assumption mentioned above is stated in Assumption 3.1, in Section 3; this section also collects several important consequences of Assumption 3.1 on the lower semicontinuity of minima and the existence of measurable minimizer—see Theorem 3.2—and on the discounted optimal control problem—see Theorem 3.3. All these results are borrowed from [7]. The main result of the present note, Theorem 4.5, is stated and proved in Section 4.

2. Performance criteria

The following notation is used throughout the note. Let $\mathbb{R}$ and $\overline{\mathbb{R}}$ denote the real numbers and extended real numbers sets, respectively. Moreover, $\mathbb{N}_0$ and $\mathbb{N}$ stand for the nonnegative and positive integers sets, respectively. For a topological space (S,τ), the Borel σ-algebra generated by the topology τ is denoted by $\mathcal{B}(S)$ and "measurability" will always mean measurability with respect to $\mathcal{B}(S)$. A Borel space Y is a measurable subset of a complete separable metric space endowed with the inherited metric.

Consider the standard Markov decision model $(\mathbf{X},\mathbf{A},\{A(x):x\in\mathbf{X}\},Q,C)$, where the *state space* $\mathbf{X}$ and the *control space* $\mathbf{A}$ are nonempty Borel spaces. The collection $\{A(x):x\in\mathbf{X}\}$ is a family of nonempty measurable subsets of $\mathbf{A}$, where $A(x)$ denotes the *admissible action set* for state $x\in\mathbf{X}$. It is assumed that the *admissible state–action pairs set* $\mathbb{K}=\{(x,a):x\in\mathbf{X},a\in A(x)\}$ belongs to $\mathcal{B}(\mathbf{X}\times\mathbf{A})$. The *transition law* $Q(\cdot|\cdot,\cdot)$ is a stochastic kernel on $\mathbf{X}$ given $\mathbb{K}$, that is, $Q(\cdot|x,a)$ is a probability measure on $\mathbf{X}$ for each $(x,a)\in\mathbb{K}$, and $Q(B|\cdot,\cdot)$ is a measurable function on $\mathbb{K}$ for each $B\in\mathcal{B}(\mathbf{X})$. Finally, the *one-step cost* $C(\cdot,\cdot)$ is a measurable function on $\mathbb{K}$.

The Markov decision model is interpreted as a model of a system that evolves as follows: at each time $n\in\mathbb{N}_0$ the state of the system is observed, say $x_n=x\in\mathbf{X}$, and the controller chooses an action $a_n=a\in A(x)$. As a result of this decision the controller incurs in cost $C(x,a)$ and the system moves to a new state, say $x_{n+1}=y\in\mathbf{X}$, according to the probability measure $Q(\cdot|x,a)$.

Let $\mathbb{H}_0:=\mathbf{X}$ and $\mathbb{H}_n:=(\mathbb{K}\times\mathbf{X})^n$ for $n\in\mathbb{N}$. A *control* or *decision policy* $\pi=\{\pi_n\}$ is a sequence of stochastic rules that choose admissible actions in each decision time; more precisely, each $\pi_n(\cdot|\cdot)$ is a stochastic kernel on $\mathbf{A}$ given $\mathbb{H}_n$ satisfying the constraints