European Journal of Operational Research 000 (2017) 1–11

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

Customizing exponential semi-Markov decision processes under the discounted cost criterion

Bora Çekyay*

Department of Industrial Engineering, Doğuş University, İstanbul, Turkey

ARTICLE INFO

Article history:
Received 3 April 2017
Accepted 12 September 2017
Available online xxx

Keywords:
Markov processes
Markov decision processes
Optimal maintenance
Infinite server queue
Customization

ABSTRACT

The uniformization technique is a widely used method for establishing the existence of optimal policies with certain monotonicity properties. This technique converts a semi-Markov decision process with exponential sojourn times (ESMDP) into an equivalent discrete-time Markov decision process by defining some fictitious jumps. This study proposes a new device, called customization, which can convert a given ESMDP into another equivalent ESMDP whose formulation possibly simplifies mathematical analysis. The customization technique uses the fictitious jump idea to establish the equivalence under deterministic stationary policies just like the uniformization technique. However, it allows the transition rates of the new ESMDP to be different. Moreover, it can be applied even when the transition rates of the initial ESMDP are unbounded. This flexibility can be very useful in analyzing the problems where the uniformization is not applicable or not so helpful. We analyze a complex optimal replacement problem and an infinite server queueing problem with unbounded transition rates to demonstrate the applicability and advantages of customization.

1. Introduction

Markov decision process (MDP) models have been successfully applied to many diverse fields such as management sciences, economics, and even ecology. MDPs are used for sequential decision making when the successive states of the considered system are uncertain and dependent on the successive decisions of the decision maker. An MDP model consists of decision epochs, states, actions, costs (rewards), and transition probabilities. At each decision epoch, the decision maker has to choose an action after observing the state. The chosen action generates a cost (reward) and determines the state at the next decision epoch through the transition probabilities. The solution of an MDP model provides the actions that optimize a criterion which is a function of the total cost (reward) in the long-run.

An MDP can generally be formulated in two different ways. In the first one, the process is assumed to be observed continuously, and the decisions are allowed at any time. This type of decision processes with exponentially distributed sojourn times between successive state transitions are called continuous-time Markov decision processes (CTMDP). In the second way of defining an MDP, the process is assumed to be observed at some discrete time points, and the decisions are allowed only at these time points, which are called decision epochs. The decision epochs can be some predetermined time points chosen by the decision maker, or some random time points when certain events occur. The MDP models with predetermined decision epochs are called discrete-time Markov decision processes (DTMDP). If the durations between decision epochs are exponentially distributed, this MDP model is referred to as a semi-Markov decision process with exponential sojourn times (ESMDP). Puterman (2005), Hu and Yue (2007), and Guo and Hernández-Lerma (2009) are excellent references on MDPs. We refer the interested readers to these books for detailed and mathematically rigorous treatments of MDP models. It should be remarked that the time between successive decision epochs could be generally distributed as well. This type of models is called semi-Markov decision processes, which are out of the scope of this study.

It is interesting that CTMDPs and ESMDPs with bounded transition rates can be converted to equivalent DTMDPs after a transformation called *uniformization*. The equivalence between ESMDPs and DTMDPs is first recognized by Lippman (1975), and its formal mathematical description is given by Serfozo (1979). Kakumanu (1977) formulates a similar equivalence between CTMDPs and DTMDPs. With the uniformization technique, all transition rates of the original MDP are converted to a single scalar, which is larger than the biggest transition rate. The equality of the value functions of the initial and final MDP processes is guaranteed by altering

* Corresponding author.
E-mail addresses: cekyay@gmail.com, bcekyay@dogus.edu.tr

https://doi.org/10.1016/j.ejor.2017.09.016
0377-2217/

the costs (rewards) appropriately, and by defining fictitious jumps from any state back to itself. Furthermore, uniformization results for SMDPs with general sojourn times are proposed by Beutler and Ross (1987).

The uniformization technique gives a relatively simple discrete-time process which is generally supposed to be easier to analyze mathematically. It is very useful especially in proving the existence of optimal policies with certain monotonicity properties. On the other hand, converting all transition rates into the same value is not so helpful in proving (or in identifying) some monotonic properties especially when the state space is multidimensional (for a related example, see Section 3.1). Moreover, the uniformization technique is not applicable when the transition rates are unbounded, which is unavoidable in some important applications, such as infinite server queues and queues with impatient customers (for related examples, see Section 3.2 and Bhulai, Brooms, & Spieksma, 2014).

This study extends the idea used in the uniformization technique to develop a new device which is applicable in case of unbounded transition rates, and which provides more flexibility to decision makers in proving structural properties of MDP models. We call this new device *customization* and introduce it for ESMDPs with the expected total discounted cost criterion by focusing only on stationary deterministic policies.

It should be remarked that when the transition rates are unbounded, the standard stability condition for SMDPs, guaranteeing that an infinite number of decision epochs is impossible within finite time, does not hold (Puterman, 2005). Since it is applicable to ESMDPs with unbounded transition rates, the customization technique has a potential to be useful in the applications for which the stability condition does not hold. We present a related example in Section 3.2, where the ESMDP (with the expected total discounted cost criterion) violating the stability condition is reduced to a DTMDP (with the expected total cost criterion) by using the customization. In contrast to ESMDP literature, the general results covering the unbounded transition rates are more well established in the CTMDP literature (see, especially, Chapter 6 in Guo & Hernández-Lerma (2009)). Moreover, Bhulai et al. (2014) propose a smoothed rate truncation method which is promising in proving structural properties of CTMDPs with unbounded transition rates. The authors also underline the necessity of a general method, that is applicable under unbounded transition rates, in the MDP literature. We believe that the customization technique will help fill this gap, at least for ESMDPs.

The customization technique converts a given ESMDP to another equivalent ESMDP by changing the transition rates. The equivalence of these ESMDPs is guaranteed by modifying the costs appropriately, and by defining some fictitious jumps. In contrast to the uniformization, the transition rates of the new ESMDP obtained after customization do not have to be the same. Each of them can be chosen by the decision maker independent of the others. The only restriction is that the new transition rate corresponding to a state and a chosen action cannot be less than the non-fictitious portion of the initial transition rate corresponding to the same state and action. To explain more clearly, let $\lambda_a(i)$, $\overline{\lambda}_a(i)$, and $p_a(i;\, i)$ be the initial transition rate, the transition rate after customization, and the probability of a fictitious jump in state i when action a is chosen, respectively. In the customization technique, $\overline{\lambda}_{a_1}(i_1)$ and $\overline{\lambda}_{a_2}(i_2)$ do not have to be the same for two different state-action pairs $(i_1,\, a_1)$ and $(i_2,\, a_2)$. They, however, must satisfy $\overline{\lambda}_{a_1}(i_1) \geq \lambda_{a_1}(i_1)(1 - p_{a_1}(i_1; i_1))$ and $\overline{\lambda}_{a_2}(i_2) \geq \lambda_{a_2}(i_2)(1 - p_{a_2}(i_2; i_2))$. Since the final transition rates do not have to be the same, the decision maker has the chance of obtaining different value functions (by changing the final transition rates) which are more convenient to prove different results. More importantly, the customization is applicable even when the transition rates are unbounded. We believe that the customization will be helpful to establish some new and so far unproven results in the literature if it is used properly.

The rest of the paper is organized as follows. In Section 2, we present the main customization result for ESMDPs under the expected discounted cost criterion. Section 3 focuses on the examples showing the advantages and the applicability of the customization.

Throughout the paper, $\mathbb{N}$ denotes the set of nonnegative integers, $\mathbb{R}^+$ denotes the set of nonnegative real numbers, and I is the indicator function defined as

$$I\{\text{condition}\} = \begin{cases} 1 & \text{if condition holds,} \\ 0 & \text{otherwise.} \end{cases}$$

2. Customizing ESMDPs

In this section, we present the main customization result for ESMDPs with the expected total discounted cost criterion by considering only stationary deterministic policies. We will start with defining an ESMDP Y with possibly unbounded transition rates. First, we will show that the fictitious jumps in Y can be eliminated without changing the optimal stationary deterministic policy. The ESMDP obtained after customization will be denoted by $\overline{Y}$. Then, the equivalence between Y and $\overline{Y}$ regarding the expected total discounted cost will be proved by using the fictitious jump elimination result.

We consider an ESMDP $Y = (S, A, r, \lambda, p, \alpha)$ with a countable state space S, and an arbitrary action space A, where all costs are continuously discounted by a factor $\alpha > 0$. After the decision process arrives at a state $i \in S$, an action $a \in A$ is chosen and a lump-sum cost $r_a(i)$ (possibly unbounded) is paid. The process remains in state i for a random sojourn time which is exponentially distributed with parameter $\lambda_a(i) > 0$ (possibly unbounded), and it jumps to state $j \in S$ with probability $p_a(i;\, j)$. Unless otherwise specified, we assume that fictitious jumps are allowed, but absorbing states are not allowed in our setting, i.e., $0 \leq p_a(i;\, i) < 1$. We let T_n be the time of the nth jump of Y, which is to state Y_n. The sojourn times of Y can be defined as $S_n = T_n - T_{n-1}$ for $n \geq 1$, where $T_0 = 0$.

It is assumed that all costs are lump-sum costs incurred at the beginning of sojourn intervals. This is not a limitation since it is well-known that if some cost rates are given, the corresponding lump-sum costs can be computed easily.

Let f be a deterministic stationary policy choosing action $f(i) \in A$ when the process is in state i. We let $r_a^+(i) = \max\{r_a(i), 0\}$ and $r_a^-(i) = \max\{-r_a(i), 0\}$, and define

$$W_{f,\alpha}^+(i) = \lim_{N\to\infty} E_f\left[\sum_{n=0}^{N} e^{-\alpha T_n} r_{a_n}^+(Y_n) | Y_0 = i\right] = E_f^i\left[\sum_{n=0}^{\infty} e^{-\alpha T_n} r_{a_n}^+(Y_n)\right]$$

and

$$W_{f,\alpha}^-(i) = \lim_{N\to\infty} E_f\left[\sum_{n=0}^{N} e^{-\alpha T_n} r_{a_n}^-(Y_n) | Y_0 = i\right] = E_f^i\left[\sum_{n=0}^{\infty} e^{-\alpha T_n} r_{a_n}^-(Y_n)\right],$$

where $a_n = f(Y_n)$. Both of the above limits exist since the summands are nonnegative. The expected total discounted cost associated with the deterministic stationary policy f is defined as

$$W_{f,\alpha}(i) = W_{f,\alpha}^+(i) - W_{f,\alpha}^-(i), \tag{1}$$

which is well defined whenever $W_{f,\alpha}^+(i)$ or $W_{f,\alpha}^-(i)$ is finite. Therefore, we impose the following assumption throughout the paper, which assures that $W_{f,\alpha}(i)$ is well defined.

Assumption 2.1. For every deterministic stationary policy f and $i \in S$, either $W_{f,\alpha}^+(i)$ or $W_{f,\alpha}^-(i)$ is finite.

Please cite this article as: B. Çekyay, Customizing exponential semi-Markov decision processes under the discounted cost criterion, European Journal of Operational Research (2017), https://doi.org/10.1016/j.ejor.2017.09.016