ASYMPTOTIC PROPERTIES OF CONSTRAINED MARKOV DECISION PROCESSES *

Eitan Altman
INRIA, Centre Sophia Antipolis
06565 Valbonne Cedex, France
Tel: 93 96 76 37.

October 1992

Abstract

We present in this paper several asymptotic properties of constrained Markov Decision Processes (MDPs) with a countable state space. We treat both the discounted and the expected average cost, with unbounded cost. We are interested in (1) the convergence of finite horizon MDPs to the infinite horizon MDP, (2) convergence of MDPs with a truncated state space to the problem with infinite state space, (3) convergence of MDPs as the discount factor goes to a limit. In all these cases we establish the convergence of optimal values and policies. Moreover, based on the optimal policy for the limiting problem, we construct policies which are almost optimal for the other (approximating) problems. Based on the convergence of MDPs with a truncated state space to the problem with infinite state space, we show that an optimal stationary policy exists such that the number of randomisations it uses is less or equal to the number of constraints plus one. We finally aply the results to a dynamic scheduling problem.

Keywords: Constrained Markov Decision Processes, countable state space, finite horizon, infinite horizon, finite approximations, asymptotic properties, dynamic scheduling.

*This work was partially supported by the Chateaubriand fellowship from the French embassy in Israel and by the European Grant BRA-QMIPS of CEC DG XIII

1 INTRODUCTION

In recent years growing attention was given to solving constrained MDPs (Markov Decision Processes). Such problems frequently arise in computer networks and data communications, see Lazar [18], Spieksma and Hordijk [14], Nain and Ross [19] and Altman and Shwartz [1]. The theory for solving constrained MDPs was developed by Hordijk and Kallenberg [16], Kallenberg [17], Beutler and Ross [8], Altman and Shwartz [2, 4], Altman [6], Spieksma [22], Sennott [20, 21] and Borkar [9].

We are interested in (1) the convergence of finite horizon MDPs to the infinite horizon MDP, (2) convergence of MDPs with a truncated state space to the problem with infinite state space, (3) convergence of MDPs as the discount factor goes to a limit. In all these cases we establish the convergence of optimal values and policies. Moreover, based on the optimal policy for the limiting problem, we construct policies which are almost optimal for the other (approximating) problems.

In the finite horizon case, the need for approximation arrizes since there are no known techniques to solve such problems, unlike the infinite horizon case. Of special importance is therefore to construct a policy that is almost optimal for the finite horizon case, based on a policy that is optimal (or almost optimal) for the infinite horizon case.

When the number of states and actions in the control problem is not finite, the question of finite approximation of the state is of interest, since the only general solution known for constrained MDPs is through a LP (Linear Programming) with infinite number of decision variables (see Altman and Shwartz [2], Altman [6] and Spieksma [22]). In the finite case, however, a finite LP can be applied to obtain an optimal policy (see Hordijk and Kallenberg [16] and Altman and Shwartz [4]). The issue of approximating MDPs by other finite state MDPs was investigated in a few papers of White (e.g. [24]), Cavazos-Cadena [10], Hernandez-Lerma [12] and Thomas and Stengos [23]. Since in our case additional constraints are introduced in the control the methods used in tha above papers are not applicable, as they are all based on dynamic programming techniques. In Altman [6], a general theory for approximation for constrained MDPs was developed and applied for finite approximations. The cost in [6] is assumed to be non decreasing and some stochastic monotonicity assumptions on the transition probabilities are made. In this paper, in Section 5, we generalize these results to costs that are not necessarily monotone, and to transition probabilities without the stochastic monotonicity property. Yet, we restrict to transitions to "nearest neighbors", i.e., from each state we assume that it is possible to move in one step to only a finite number of states.

The third issue that we treat is the convergence of MDPs as the discount factor goes to a limit. This is of special interest for the case that the discount factor tends to one. Indeed, in control problems of queueing systems without constraints, an optimal policy for the expected average cost has often been obtained as the limit of optimal policies for the discounted cost, as the discount factor goes to one. By proving the continuity of MDPs in the discount factor, we thus establish the validity of this approach to constrained MDPs as well.

We finally illustrate the above ideas by the problem of optimal priority assignment where N infinite queues compete for the attention of a single server.

Previous results on asymptotic properties of constrained MDPs have been obtained by Altman and Shwartz [4] and Altman and Gaitsgory [5] for the case of finite state space. They obtain conditions for the continuity of the optimal value and policy of MDPs in the immediate cost, transition probabilities and discount factor. These results are obtained by applying the theory of stability of finite LPs. This method is however not extendible to the case of infinite state space. An alternative method is used by Altman and Shwartz [4] that does not rely on a LP approach in order to obtain the convergence of the optimal value as the horizon tends to infinity for the case of discounted cost, and finite state space. The theory of convergence of finite state constrained MDPs is applied in Altman and Shwartz [3, 4] to solve adaptive control problems. A characterization and a study of cases were the optimal values and policies of MDPs are not continuous is presented by Altman and Gaitsgory in [5].

The structure of the paper is as follows: after presenting the model notation and assumptions in Section 2, we cite in Section 3 key Theorems for approximation. We analyze the finite horizon problem in Section 4, and then in Section 5 we introduce and analyze the finite state approximation. The continuity of MDPs in the discounted cost is studied in Section 6. The asymptotic results are applied in Section 7 to a problem of optimal dynamic schedulng.

2 Model and Assumptions

2.1 The model

We consider a discrete time MDP with a countable state space $\mathbf{X}$ and a countable action space $\mathbf{A}$. However, we assume that at each state $y \in \mathbf{X}$ there is only a finite number of available actions, $\mathbf{A}(y)$. The probability to go from state x to y given that action a is used, is given by the transition probabilities $\mathcal{P} = \{\mathcal{P}_{xay}\}$. A policy u in the policy space U is a sequence $u = \{u_0, u_1, ...\}$, where u_t is a probability over $\mathbf{A}$ conditioned on the whole history of actions and states before t, plus the current state at time t. Each initial state x and policy u define a probability measure P_x^u on which the the state process $\{X_t\}_{t=0}^{\infty}$ and the action process $\{A_t\}_{t=0}^{\infty}$ are defined. Let E_x^u be the expectation under P_x^u.

Let $u \in U(M), U(S)$ and $U(SD)$ denote the class of Markov policies, stationary (randomized) policies and stationary deterministic policies, respectively . A stationary policy $g \in U(S)$ is characterized by a single conditional distribution $p^g_{\bullet|x} := g(\bullet \mid X_t = x)$ over $\mathbf{A}$, so that $p^g_{\mathbf{A}|x} = 1$; under g, X_t becomes a Markov chain with stationary transition probabilities, given by $P^g_{xy} := \sum_{a \in \mathbf{A}} p^g_{a|x} \mathcal{P}_{xay}$. For $g \in U(S)$, let $\Pi^g := \lim_{n \to \infty} [P^g]^n$. If all rows of Π^g are equal, we denote them by π^g. A more detailed definition and construction of the MDP can be found e.g. in [13].

2.2 The constrained MDP

Let $C(x, u)$ and $D(x, u) := \{D^k(x, u)\ ,\ 1 \leq k \leq K\}$ be cost functions associated with each policy u and an initial state x. The real vector $V := \{V_k\ ,\ k = 1, ..., K\}$ is held fixed hereafter. Call a policy u *feasible* if

$$D^k(x, u) \leq V_k\ , \quad k = 1, 2, \ldots, K \tag{2.1}$$

The constrained control problem is:

COP: Find $v \in U$ that minimizes $C(x, u)$ over all policies $u \in U$ that are feasible. We shall define in particular the constrained optimization problem restricted to a subclass $\overline{U} \subset U$

of policies:

COP$(\overline{U})$: Find $v \in \overline{U}$ that minimizes $C(x,u)$ over all policies $u \in \overline{U}$ that are feasible.

$C(x,u)$ and $D^k(x,u)$ will stand for either the discounted or the expected average cost, defined below. Let $c(x,a)$, $d(x,a) := \{d^k(x,a)\ ,\ k = 1,...,K\}$ be real ($\mathbb{R}^K$ resp.) valued instantaneous cost functions, i.e. costs per state-action pair. We shall assume throughout the paper that they are uniformly bounded from below. Let $0 < \beta \le 1$ be a discount factor. We define two versions of discounted cost functionals from $\mathbf{X} \times U$ to $\mathbb{R}$ (see Altman and Shwartz [4]).

$$\begin{aligned} C_\beta^t(x,u) &:= (\textstyle\sum_{s=0}^t \beta^s)^{-1} E_x^u \left[\sum_{s=0}^t \beta^s c(X_s,A_s)\right] \\ D_\beta^{k,t}(x,u) &:= (\textstyle\sum_{s=0}^t \beta^s)^{-1} E_x^u \left[\sum_{s=0}^t \beta^s d^k(X_s,A_s)\right] \qquad k=1,...,K \end{aligned} \tag{2.2}$$

$$\begin{aligned} \tilde{C}_\beta^t(x,u) &:= (1-\beta) E_x^u \left[\textstyle\sum_{s=0}^t \beta^s c(X_s,A_s)\right] \\ \tilde{D}_\beta^{k,t}(x,u) &:= (1-\beta) E_x^u \left[\textstyle\sum_{s=0}^t \beta^s d^k(X_s,A_s)\right] \qquad k=1,...,K \end{aligned} \tag{2.3}$$

$$\begin{aligned} C_\beta(x,u) &:= \overline{\lim}_{t\to\infty} C_\beta^t(x,u) \\ D_\beta^k(x,u) &:= \overline{\lim}_{t\to\infty} D_\beta^{k,t}(x,u) \qquad k=1,...,K \end{aligned} \tag{2.4}$$

When $\beta = 1$, (2.4) reduce to the well-known definition of the expected average cost, since in that case $\sum_{s=0}^t \beta^s = t+1$. For $\beta < 1$ the definition (2.4) are a normalized version of the *standard expected discounted* costs

$$C_\beta(x,u) := (1-\beta) E_x^u \left[\sum_{s=0}^{\infty} \beta^s c(X_s,A_s)\right] = \lim_{t\to\infty} C_\beta^t(x,u) = \lim_{t\to\infty} \tilde{C}_\beta^t(x,u) \tag{2.5}$$

and similarly for $D_\beta(x,u)$. The addition of a normalizing factor, e.g. in (2.2) is important when one is interested in continuity of the (finite horizon) cost in the discount parameter, especially at $\beta = 1$, see Section 6. Another advantage of using the new definition (2.4) is that it enables to obtain the same LP for both the discounted and expected average cost for solving COP (see e.g. Altman [6]). This enables to reduce problems of continuity, sensitivity and singular perturbations of finite constrained MDPs in the discount parameter β as $\beta \to 1$, to the corresponding problems in Mathematical Programs for which the theory is well known (see e.g. Altman and Shwartz [4]).

Finally, for $u \in U(M)$, define $c(u_t) = \{c(y, u_t), y \in \mathbf{X}\}$, where $c(y, u_t) = \sum_a c(y, a) p^{u_t}_{a|y}$. In particular, if $u \in U(S)$ then $c(u) := c(u_t)$ (for any $t \geq 1$).

2.3 Assumptions and Notation

Let $\mu : \mathbf{X} \to \mathbb{R}$ be some positive function. Following Dekker and Hordijk [11] and Spieksma [22], define the μ-norm of vectors $\pi \in \mathbb{R}^{\mathbf{X}}$ and of matrices $P \in \mathbb{R}^{\mathbf{X}\times\mathbf{X}}$ as

$$\begin{aligned} \|\pi\|_\mu &= \sup_{x\in\mathbf{X}} \mu_x^{-1} |\pi_x| \\ \|P\|_\mu &= \sup_{x\in\mathbf{X}} \mu_x^{-1} \sum_{y\in\mathbf{X}} |P_{xy}| \mu_y \end{aligned} \tag{2.6}$$

For a subset $M \subset \mathbf{X}$, let ${}_M P$ be the taboo matrix corresponding to P, i.e.

$${}_M P_{xy} = \begin{cases} P_{xy}, & y \notin M \\ 0, & y \in M \end{cases} \tag{2.7}$$

A MDP is said to be μ-uniform geometrically recurrent (μ-UGR), if a finite set M and a $\zeta < 1$ exist, such that for any $u \in U(SD)$,

$$\|{}_M P^u\|_\mu \leq \zeta. \tag{2.8}$$

The following assumptions are used frequently in the paper:

B1: Under any $g \in U(SD)$, $\mathbf{X}$ contains a single ergodic aperiodic class (and possibly also transient states).

B2: there exists some policy u (not necessarily an optimal policy) such that

$$D^k(x, u) < V_k \ , \quad k = 1, 2, \ldots, K. \tag{2.9}$$

B3: There exists some function $\mu \in \mathbf{X} \to \mathbb{R}$ such that $\mu_y > 0,\ y \in \mathbf{X}$;

B3(i): $\mu_y \geq 1,\ y \in \mathbf{X}$ and the MDP is μ-UGR;

B3(ii): if $\beta < 1$ then $\exists\alpha$, $0 < \alpha < 1$ such that for all $u \in U(SD)$, $\|\beta P^u\|_\mu \le \alpha$.

B3(iii): $\exists R < \infty$ such that $\|\max_a c(\bullet, a)\|_\mu \le R$, $\left\|\max_a d^k(\bullet, a)\right\|_\mu \le R$, $k = 1, ..., K$.

B4: For any x, a, $\mathcal{P}_{xa\bullet}$ has finite support (that may depend on x, a).

We shall often need the following result, extending B3(i) and B3(ii) to $U(M)$.

Lemma 2.1 *For any $u \in U(M)$ and any $k = 0, 1, ...$*

$$\begin{aligned} B3(i) \quad &\Rightarrow \quad \|_M P^{u_k}\|_\mu \le \zeta \\ B3(ii) \quad &\Rightarrow \quad \|\beta P^{u_k}\|_\mu \le \alpha \end{aligned} \tag{2.10}$$

Proof: For any $u \in U(M)$ and $k = 0, 1, ...$

$$\|P^{u_k}\|_\mu = \sup_{x\in\mathbf{X}} \mu_x^{-1} \sum_{y\in\mathbf{X}} |\sum_a \mathcal{P}_{xay} p^{u_k}_{a|x}| \mu_y \le \sup_{x\in\mathbf{X}} \mu_x^{-1} \sum_{y\in\mathbf{X}} |\max_a \mathcal{P}_{xay}| \mu_y = \|P^w\|_\mu \tag{2.11}$$

where $w \in U(SD)$ is a policy that achieves the max in (2.11). Hence $\|\beta P^{u_k}\|_\mu \le \|\beta P^w\|_\mu \le \alpha$ if B3(ii) holds. The proof of the claim for B3(i) follows similarly. ■

The following notation is used below: For any set B, $1\{B\}$ is the indicator function of the set, $|B|$ the cardinality of this set (if B is finite then $|B|$ is the number of elements in B). For vectors D and V in $\mathbb{R}^K$, the notation $D < V$ stands for $D_k < V_k$, $k = 1, 2, \ldots, K$, with a similar convention for matrices. For two matrices ζ, Q of appropriate dimensions, $\zeta \cdot Q$ stands for summation over common indices (scalar product). Q^T denotes the transposed of the matrix Q.

3 Key Theorems for Approximation

In this section we quote and adapt some results from [6] needed later for approximations of constrained MDPs. Assume that for any initial state x and policy u we have a sequence of approximations $C_n(x, u)$ and $D_n^k(x, u)$, $n = 1, 2, ...$ of the costs $C(x, u)$ and $D^k(x, u)$ with

$k = 1, 2, ..., K$. Below $C(x, u)$ and $D^k(x, u)$ will stand for either the expected average cost $\beta = 1$ or for the discounted cost $\beta < 1$ (β will be omitted from the notation), where as $C_n(x, u)$ and $D^k_n(x, u)$, may be arbitrary functions from $\mathbf{X} \times U$ to $\mathbb{R}^+$.

Let $W \subset U$ be a class of policies that contains the stationary policies. Consider the following sequence of problems:

$\mathbf{COP_n(W)}$: Find $C_n^W(x)$ which is given by:

$C_n^W(x) := \inf_{u \in W}\{C_n(x, u); D^k_n(x, u) \leq V_k,\ k = 1, 2, ..., K\}$.

Let $C^W(x)$ be the optimal value of COP(W). In the above definitions, when $W = U$ then we omit W from the notation. Note that COP_n may not have any optimal policy even if it is feasible.

Let $\Gamma(W) \subset W$ be a set of policies which includes the stationary policies $U(S)$ and is sufficient for $COP_n(W)$ in the following sense. For any $m = 1, 2, ...$ and $u \in W$, there exists some $u^m \in \Gamma(W)$, such that

$$C_m(x, u^m) \leq C_m(x, u), \qquad D^k_m(x, u^m) \leq D^k_m(x, u), \quad k = 1, 2, ..., K. \tag{3.1}$$

Theorem 3.1 *Assume (1) B2,*

(2) $\lim_{n\to\infty} C_n(x, u) = C(x, u)$ *and* $\lim_{n\to\infty} D^k_n(x, u) = D^k(x, u)$ *for every policy* $u \in \Gamma(W)$, *uniformly in* $u \in \Gamma(W)$,

(3) $\beta < 1$, *or { B1 and B3(i) } hold.*

Then (i) $\lim_{n\to\infty} C_n(x) = C(x)$.

(ii) Choose a sequence $\xi_n \to 0$. *Let* $r(n)$ *be a* ξ_n*-optimal policy for* COP_n *if* COP_n *is feasible, otherwise let it be an arbitrary stationary policy. Let* $w(n)$ *be a stationary policy that satisfies*

$$C(x, w(n)) \leq C(x, r(n)), \qquad D^k(x, w(n)) \leq D^k(x, r(n)), \quad k = 1, 2, ..., K \tag{3.2}$$

(see Lemma 3.1 below). Let w *be an arbitrary accumulation point of* $w(n)$, $n = 1, 2, ...$ *i.e. there exists a subsequence* $\{n_i\}_{i=1}^{\infty}$ *such that for all* $y \in \mathbf{X}, a \in \mathbf{A}$,

$$\lim_{i\to\infty} p^{w(n_i)}_{a|y} = p^w_{a|y}. \tag{3.3}$$

Then w *is optimal for* COP.

Proof: The proof of the theorem is given in Altman [6] Thm. 6.1 for the special case that $W = U$ is the class of all policies. However, the proof carries in a straight forward way to the case that $W \neq U$ provided that $U(S) \subset W$. The conditions of Thm. 6.1 are however slightly different for $\beta = 1$, and we show below that they hold indeed. It follows from Proposition 5.1 (vi) p. 97 of Flos [22] and Lemma 2.1 that the family $\{[P^u]^n_{x,\bullet}\}_n$ is tight (since B3(i) is assumed for $\beta = 1$) for any given $u \in U(S)$ and initial state x. This implies condition A3(i) in [6]. The condition A4 in [6] is implied by the fact that under any stationary policy and for any given initial state x, the total expected cost till the hitting of some recurrent state in M, is finite (this follows since by [22] p. 25 Table 1 and Lemma 2.5, if a Markov chain is $\mu - UGR$ then it is "$\mu - BSRS(M)$", see definition in [22] p. 19). ∎

Lemma 3.1 *Assume either $\beta < 1$, or \{ B1 and B3(i) \}. Then*
(i) for each policy $u \in U$, there exists a stationary policy g such that

$$C_\beta(x, g) \leq C_\beta(x, u), \qquad D^k_\beta(x, g) \leq D^k_\beta(x, u), \quad k = 1, 2, ..., K \tag{3.4}$$

and hence the stationary policies are "sufficient" for COP.
(ii) if COP is feasible then an optimal policy for COP exists within U(S).

Proof: Follows immediately from [6] Thm. 4.1. The same arguments as in the proof of the last theorem imply that the conditions of Thm 4.1 in [6] hold. ∎

Next we construct a policy which is almost optimal for $COP_n(W)$ for all n large enough. Let $u^* \in W$ be a policy for which
(1) the following limit exists for all y and a

$$\overline{f}(x, u; y, a) := \lim_{t\to\infty} [\sum_{s=0}^{t} \beta^s]^{-1} \sum_{s=0}^{t} \beta^s P^u_x(X_s = y, A_s = a); \tag{3.5}$$

we shall use the notation $\overline{f}(x, u) = \{\overline{f}(x, u; y, a\}_{y,a}$;
(2) the following holds:

$$C(x, u) = c \cdot \overline{f}(x, u), \qquad D^k(x, u) = d^k \cdot \overline{f}(x, u), \quad k = 1, ..., K, \tag{3.6}$$

(3) u^* is ϵ-optimal for $COP(W)$.

Let $u(\epsilon) \in W$ be the policy that satisfies (3.5), the linear representation (3.6) and such that

$$\overline{f}(x, u(\epsilon)) = (1-\epsilon)\overline{f}(x, u^*) + \epsilon \overline{f}(x, v) \tag{3.7}$$

where v is a stationary policy that satisfies assumption **B2** (it follows from Lemma 3.1 that indeed, if (2.9) holds for some u then there exists a stationary policy v for which it also holds. The existence of the policy $u(\epsilon)$ under the conditions of Theorem 3.1 is established in Altman [6] in the proof of Theorem 6.1 there).

Theorem 3.2 *([6] Thm. 6.2) Let $-\mathcal{M}$, $\mathcal{M} \geq 0$ be a lower bound on the immediate costs c. Under the conditions of Theorem 3.1, $u(\epsilon)$ defined in (3.7) is $\hat{\epsilon}$-optimal for $COP_n(W)$ for all n large enough, where $\hat{\epsilon} = \epsilon[C(x, v) + \mathcal{M} + 3]$.*

4 The finite horizon case

Let COP_n denote the COP with finite horizon costs $C_\beta^n(x, u)$ and $D_\beta^n(x, u)$. Let $C_\beta^n(x)$ be the optimal value for COP_n.

Theorem 4.1 *Assume B2, B3(ii), B3(iii) and $\beta < 1$. Then*

(1) $\lim_{n\to\infty} C_\beta^n(x) = C_\beta(x)$.

(2) Choose some $\epsilon > 0$. Let u^* be an ϵ-optimal (or optimal) policy for COP that satisfies (3.6) (e.g. any ϵ-optimal stationary policy). There exists some $N(\epsilon)$ such that for all $n \geq N(\epsilon)$, the policy $u(\epsilon)$ satisfying (3.6) and (3.7) is $\hat{\epsilon}$-optimal for COP_n, where $\hat{\epsilon}$ is given in Theorem 3.2.

(3) Choose a sequence $\xi_n \to 0$. Let $r(n)$ be a ξ_n-optimal policy for COP_n if COP_n is feasible, otherwise let it be an arbitrary stationary policy. The stationary policy w obtained by applying the limiting procedure in Theorem 3.1 (ii) to the policies $r(n)$ is optimal for COP.

Proof: We begin by showing that $\left\| C_\beta^n(\bullet, u) - C_\beta(\bullet, u) \right\|_\mu \to 0$ uniformly in $U(M)$. The restriction to $U(M)$ follows by identifying $U(M)$ with $\Gamma(U)$ (required in Theorem 3.1). More precisely, it follows from Theorem 13.2 in [15] that for any $n = 1, 2, ...$ and $u \in U$, there exists some $u^n \in U(M)$, such that

$$C_n(x, u^n) = C_n(x, u), \qquad D_n^k(x, u^n) = D_n^k(x, u), \quad k = 1, 2, ..., K, \tag{4.1}$$

and hence $U(M)$ can be chosen as $\Gamma(U)$. We have by Lemma 2.1 for any $u \in U(M)$

$$\left\| \tilde{C}_\beta^n(\bullet, u) - C_\beta(\bullet, u) \right\|_\mu \leq (1-\beta) \left\| \sum_{k=n}^{\infty} \beta^{k+1} P^{u_0} P^{u_1} \cdots P^{u_k} \right\|_\mu \sup_{k>n} \|c(u_k)\|_\mu \leq \frac{(1-\beta) R \alpha^{n+1}}{1-\alpha} \tag{4.2}$$

$$C_\beta^n(\bullet, u) - \tilde{C}_\beta^n(\bullet, u) = (1-\beta) C_\beta^n(\bullet, u) \left([1-\beta]^{-1} - \sum_{k=0}^{n} \beta^k \right) = \beta^{n+1} C_\beta^n(\bullet, u) \tag{4.3}$$

and hence

$$\left\| C_\beta^n(\bullet, u) - \tilde{C}_\beta^n(\bullet, u) \right\|_\mu = \left\| \beta^{n+1} C_\beta^n(\bullet, u) \right\|_\mu \leq \frac{\beta^{n+1} R}{1-\alpha} \tag{4.4}$$

Combining (4.2) and (4.4) we get

$$\left\| C_\beta^n(\bullet, u) - C_\beta(\bullet, u) \right\|_\mu \leq \frac{(\beta^{n+1} + \alpha^{n+1}) R}{1-\alpha} \tag{4.5}$$

and similarly for $D_\beta^n(\bullet, u)$. Hence the conditions of Theorem 3.1 and 3.2 are satisfied, which establishes the proof. ■

Next we treat the case where B3(i) holds (i.e. the uniform μ-geometrically recurrence) instead of B3(ii). This turns to be not only an alternative condition for the convergence of the discounted (finite horizon) cost, but also a condition for the convergence of the expected average cost, $\beta = 1$. We shall sometimes omit β from the notation below. We show that the optimal value and an optimal policy for COP (among all policies) are given as limits of the optimal value and policy for the finite horizon constrained problem, when restricting to the subclass of stationary policies. We thus consider $W = U(S)$, and denote $C_n^W(x)$ the optimal value for $COP_n(W)$. We know from Lemma 3.1 that there exists an optimal stationary policy for COP (among all policies) if it is feasible (but not necessarily for COP_n).

Theorem 4.2 *Assume $\{B1$ or $\beta < 1\}$, B2, B3(i) and B3(iii). Then*

(1) $\lim_{n\to\infty} C_n^W(x) = C(x)$.

(2) Choose some $\epsilon > 0$. Let $u^* \in U(S)$ be an ϵ-optimal (or optimal) policy for COP. There exists some $N(\epsilon)$ such that for all $n \geq N(\epsilon)$, the stationary policy $u(\epsilon)$ satisfying (3.7) is $\hat{\epsilon}$-optimal for COP_n, where $\hat{\epsilon}$ is given in Theorem 3.2.

(3) Choose a sequence $\xi_n \to 0$. Let $w(n)$ be a ξ_n-optimal stationary policy for $COP_n(W)$ if $COP_n(W)$ is feasible, otherwise let it be an arbitrary stationary policy. The stationary policy w obtained by applying the limiting procedure in Theorem 3.1 (ii) to the policies $w(n)$ is optimal for COP.

Proof: Note that both u^* and $u(\epsilon)$ satisfy (3.6) since they are stationary. B3(i) implies that the MDP is uniform μ-geometrically ergodic, i.e. $\exists \rho > 0$, $0 < \sigma < 1$, s.t.

$$\begin{aligned} \|[P^u]^n - \Pi^u\|_\mu &\leq \rho\sigma^n, \ \forall n \in \mathbb{N}, \\ \|P^u\|_\mu &\leq \rho \end{aligned} \tag{4.6}$$

for all $u \in U(SD)$ (see Key Theorem II, p. 108 in Spieksma [22]). This easily extends to all $u \in U(S)$ by Lemma 2.1. Moreover, B3(i) implies that $\exists r > 0$ s.t.

$$\|\Pi^u\|_\mu \leq r, \quad \text{and} \quad \|[P^u]^n\|_\mu \leq r, \quad \text{uniformly in } n \in \mathbb{N} \ \text{ and u} \in \text{U(S)}. \tag{4.7}$$

This follows from Proposition 5.1 p. 97 in Spieksma [22] for $u \in U(SD)$ and generalizes readily to $U(S)$ due to Lemma 2.1.

Consider first $\beta < 1$. For all $u \in U(S)$, $n \in \mathbb{N}$,

$$\left\| \tilde{C}_\beta^n(\bullet, u) - C_\beta(\bullet, u) \right\|_\mu \leq (1-\beta) \left\| \sum_{k=n+1}^{\infty} \beta^k [P^u]^k \right\|_\mu \sup_{k>n} \|c(u_k)\|_\mu \leq rR\beta^{n+1} \tag{4.8}$$

It follows from (4.3) that

$$\left\| C_\beta^n(\bullet, u) - \tilde{C}_\beta^n(\bullet, u) \right\|_\mu = \left\| \beta^{n+1} C_\beta^n(\bullet, u) \right\|_\mu \leq \frac{\beta^{n+1} rR}{1-\beta} \tag{4.9}$$

Combining (4.8) and (4.9) we get

$$\left\| C_\beta^n(\bullet, u) - C_\beta(\bullet, u) \right\|_\mu \le \frac{2\beta^{n+1} r R}{1-\beta} \tag{4.10}$$

and similarly with $D_\beta(\bullet, u)$.

Similarly, we have for $\beta = 1$, for all $u \in U(S)$, $n \in \mathbb{N}$,

$$\|C_1^n(\bullet, u) - C_1(\bullet, u)\|_\mu \le \frac{\sum_{k=0}^n \left\| [P^u]^k - \Pi^u \right\|_\mu}{n+1} \|c(u)\|_\mu \le \frac{\rho R \sum_{k=0}^n \sigma^k}{n+1} \le \frac{\rho R}{(n+1)(1-\sigma)}, \tag{4.11}$$

and similarly with $D_\beta(\bullet, u)$, where ρ and σ are given in (4.6). Hence the conditions of Theorem 3.1 and 3.2 are satisfied, which establishes the proof. ■

The following Corollary immediately follows from (4.2) and (4.8).

Corollary 4.1 *Let $\beta < 1$. Then Theorems 4.1 and 4.2 hold also for the finite horizon discounted cost $\tilde{C}_\beta^n(x, u)$ replacing $C_\beta^n(x, u)$, and similarly with $\tilde{D}_\beta^n(x, u)$.*

5 Finite approximations

In this section we consider the problem of approximating the MDP which is characterized by the original transition probabilities $\mathcal{P}_{xay}$ by a sequence MDP_m with transition probabilities $\{\mathcal{P}_{xay}(m)\}$, $m = 1, 2, \ldots$. We denote by $C_\beta(m; x, u)$ and $D_\beta(m; x, u)$ the costs (given in (2.4)) under policy u and initial state x corresponding to MDP_m and discount factor β $(0 < \beta \le 1)$. Define similarly $C_\beta^n(m; x, u)$ and $D_\beta^n(m; x, u)$ the finite horizon costs given in (2.2). Let $C_\beta(m; x)$ denote the optimal value for COP_m (which is the constrained problem coresponding to MDP_m).

We shall construct the MDP_m such that for all $x, y \in \mathbf{X}$, $a \in \mathbf{A}$ $\lim_{m\to\infty} \mathcal{P}_{xay}(m) = \mathcal{P}_{xay}$. We then show that the construction ensures that $\lim_{m\to\infty} C_\beta(m; x, u) = C_\beta(x, u)$ and $\lim_{m\to\infty} D_\beta^k(m; x, u) = D_\beta^k(x, u)$, $k = 1, \ldots, K$ uniformly in $u \in U(S)$ which will be shown to imply the conditions of Theorem 3.1 and 3.2, and hence the convergence of the optimal values and policies.

Introduce the following approximation scheme **FAP**:

(i) For all $m = 1, 2, ...$, the state space is decomposed in two disjoint classes of states: E^m, which contains a finite number of states, and T^m.

(ii) For all m large enough the following holds: under any stationary policy u, E^m contains one recurrent class (plus possibly some transient states), T^m is a transient class, and absorption into the positive recurrent class takes place in finite expected time from any initial state.

(iii) $E_m \subset E_{m+1}$, $m = 1, ...$; $E_\infty = \lim_{m\to\infty} E_m = \mathbf{X}$.

(iv) There is some partial order on $\mathbf{X}$, and the following holds:

$$\mathcal{P}_{xay}(m) \begin{cases} = \mathcal{P}_{xay} & x \in E_m,\ y \in E_{m-1} \\ \geq \mathcal{P}_{xay} & x \in E_m,\ y \in E_m \backslash E_{m-1} \\ = 0 & x \in E_m,\ y \notin E_m \\ = 1\{y = 1\} & x \notin E_m \end{cases} \tag{5.1}$$

where 1 is some arbitrary state in E_1. Moreover, for every $m > 1$ and each $y \in E_m \backslash E_{m-1}$ and $x \in E_{m-1}$, we have $x \leq y$ w.r.t. the partial order on $\mathbf{X}$.

For $u \in U(S)$, denote $P^u_{xy}(m) := \sum_a \mathcal{P}_{xay}(m) p^u_{a|x}$

Theorem 5.1 *Consider a sequence of finite approximations MDP_m obtained by applying* **FAP**. *Assume B2, B3(i), B3(iii), B4 and $\{B1$ or $\beta < 1\}$. Assume that μ (defined in assumption B3) is nondecreasing in $\mathbf{X}$ with respect to the partial order introduced in* **FAP** *(iv). Then*

(1) $\lim_{m\to\infty} C_\beta(m; x) = C_\beta(x)$.

(2) Choose some $\epsilon > 0$. Let u^* be an ϵ-optimal (or optimal) policy for COP that satisfies (3.6) (e.g. any ϵ-optimal stationary policy). There exists some $N(\epsilon)$ such that for all $m \geq N(\epsilon)$, the policy $u(\epsilon)$ satisfying (3.6) and (3.7) is $\hat{\epsilon}$-optimal for COP_m, where $\hat{\epsilon}$ is given in Theorem 3.2.

(3) Let $w(m)$ be an optimal stationary policy for COP_m if COP_m is feasible, otherwise let it be an arbitrary stationary policy. The stationary policy w obtained by applying the limiting procedure in Theorem 3.1 (ii) to the policies $w(m)$ is optimal for COP.

Proof: We first note that $\Gamma(U)$, for which we need to show the uniform convergence (Theorem

3.1 condition (2)) can be chosen as the class of stationary policies (follows from Lemma 3.1 applied to COP_m). Now

$$\begin{aligned} |C_\beta(m;x,u) - C_\beta(x,u)| \;\; &\leq \;\; \left|C_\beta(m;x,u) - C^n_\beta(m;x,u)\right| \\ &\quad + \left|C^n_\beta(m;x,u) - C^n_\beta(x,u)\right| \qquad (5.2) \\ &\quad + \left|C^n_\beta(x,u) - C_\beta(x,u)\right| \end{aligned}$$

By (4.10) or (4.11), $\left|C^n_\beta(x,u) - C_\beta(x,u)\right|$ converges to zero as $n \to \infty$ uniformly in $u \in U(S)$.

Next we show that this also holds for the first term in (5.2). Without loss of generality, assume that $1 \in M$ (1 is the state defined in **FAP** (iv), and M is defined in (2.8)). We have for any $u \in U(S)$ for $\beta = 1$

$$\begin{aligned} \|{}_M P^u(m)\|_\mu \;\; &= \;\; \sup_{x \in E_m} \mu_x^{-1} \sum_{y \notin M} |P^u_{xy}(m)|\mu_y \\ &= \;\; \max_{x \in E_m} \mu_x^{-1} \left[\sum_{y \in E_{m-1} \setminus M} P^u_{xy}\mu_y + \sum_{y \in E_m \setminus E_{m-1} \setminus M} P^u_{xy}(m)\mu_y \right] \\ &\leq \;\; \max_{x \in E_m} \mu_x^{-1} \left[\sum_{y \in E_{m-1} \setminus M} P^u_{xy}\mu_y + \sum_{y \in \mathbf{X} \setminus E_{m-1} \setminus M} P^u_{xy}\mu_y \right] \\ &= \;\; \max_{x \in E_m} \mu_x^{-1} \sum_{y \in \mathbf{X} \setminus M} P^u_{xy}\mu_y \\ &\leq \;\; \|{}_M P^u\|_\mu \qquad (5.3) \end{aligned}$$

We show that this implies that the first term in (5.2) converges to 0 as $n \to \infty$ uniformly in $U(S)$ and in m. Consider a new MDP^* which is identical to the original MDP, except that in each state an extra action is added that allows to reach state 1 in one step w.p.1. Let $U(S^*)$ be the set of stationary policies in MDP^*. Let $\overline{P}^u$ and $\overline{C}_\beta(x,u)$ correspond to the transition probabilities and cost associated with MDP^* (where $u \in U(S^*)$). It is easily seen that MDP^* is also μ-UGR with the same M and ζ as the original one. For each policy $u \in U(S)$ and each $m = 1, 2, ...$, there exists $w(u,m) \in U(S^*)$ such that $P^u(m) = \overline{P}^{w(u,m)}$;

hence $C^n_\beta(m;x,u) = \overline{C}^n_\beta(x,w(u,m))$ and $C_\beta(m;x,u) = \overline{C}_\beta(x,w(u,m))$. Thus for all m large enough, MDP_m are $\mu - UGR$ with the the same M and ζ as the original one and are uniform μ-geometrically ergodic with σ, ρ and r (see (4.6) and (4.7)) that do not depend on m. The conditions of Theorem 4.2 therefore hold for MDP_m. It thus follows from (4.10) and (4.11) that for all m large enough, $\left|C^n_\beta(m;x,u) - C_\beta(m;x,u)\right|$ converges to zero as $n \to \infty$ uniformly in $u \in U(S)$ and in m.

Choose some ϵ. Since both the first and the last term converge to zero uniformly in $u \in U(S)$ for any m (large enough), it follows that $\exists n(\epsilon)$ that does not depend on m, such that

$$\left|C_\beta(m;x,u) - C^{n(\epsilon)}_\beta(m;x,u)\right| < \epsilon/2, \qquad \left|C^{n(\epsilon)}_\beta(x,u) - C_\beta(x,u)\right| < \epsilon/2 \tag{5.4}$$

for all m large enough. B4 implies that for any n and x, $\exists L(n,x)$ s.t. $\forall m > L(n,x)$, $\forall u \in U(S)$ (5.4) holds and

$$\left|C^n_\beta(m;x,u) - C^n_\beta(x,u)\right| = 0. \tag{5.5}$$

Combining this with (5.4) we obtain $\forall u \in U(S)$, $\forall \epsilon > 0$,

$$|C(m;x,u) - C(x,u)| < \epsilon \tag{5.6}$$

for all $m \geq L(n(\epsilon),x)$, and hence as $m \to \infty$, $C(m;x,u)$ converges to $C(x,u)$ uniformly in $u \in U(S)$. We obtain similarly the uniform convergence of $D(m;x,u)$. Thus the conditions of Theorem 3.1 and 3.2 are satisfied, which establishes the proof. ∎

Theorem 5.2 *Consider a sequence of finite approximations MDP_m obtained by applying* **FAP**. *Assume B2, B3(ii), B3(iii), B4 and $\beta < 1$. Assume that μ (defined in assumption B3) is non-decreasing in* $\mathbf{X}$ *with respect to the partial order introduced in* **FAP** *(iv). Then the statements of Theorem 5.1 hold.*

Proof: Here again we use the inequality (5.2). By (4.5), $\left|C^n_\beta(x,u) - C_\beta(x,u)\right|$ converges to zero as $n \to \infty$ uniformly in $u \in U(S)$.

It follows from (5.3) with $M = \emptyset$ that $\|\beta P^u(m)\|_\mu \leq \|\beta P^u\|_\mu \leq \alpha$. By (4.5), this implies that the first term in (5.2) converges to 0 as $n \to \infty$ uniformly in $U(S)$ and in m. By the same argument as in the proof of Theorem 5.1 it follows from (4.5) that for all m large enough, $\left|C_\beta^n(m; x, u) - C_\beta(m; x, u)\right|$ converges to zero as $n \to \infty$ uniformly in $u \in U(S)$ and in m. The rest of the proof is the same as the proof of Theorem 5.1. ■

From Theorem 5.1 (3) or Theorem 5.2 we deduce that if the conditions of one of the Theorems hold, then an optimal stationary policy exists for which the number of randomizations required, for either the average or the discounted cost, is at most the number of constraints plus one. This follows by a simple limit argument since for the approximating MDPs this holds true ([4] p. 9).

6 Convergence in the discount factor

Let COP_β denote the COP with cost $C_\beta(x, u)$ and $D_\beta(x, u)$, where $0 < \beta \leq 1$. Let $C_\beta(x)$ be the optimal value for COP_β.

Theorem 6.1 *Let $\gamma \to \beta$, $0 < \gamma \leq 1, 0 < \beta \leq 1$. Assume B2, B3(iii) and either $\{\beta = 1$, B1 and B3(i) $\}$ or $\{\beta < 1$ and B3(i) or B3(ii) hold $\}$. Then*
(1) $\lim_{\gamma \to \beta} C_\gamma(x) = C_\beta(x)$.
(2) Choose some $\epsilon > 0$. Let u^* be an ϵ-optimal (or optimal) policy for COP_β that satisfies (3.6) (e.g. any ϵ-optimal stationary policy). There exists some $\Delta(\epsilon) > 0$ such that for all γ such that $|\gamma - \beta| < \Delta(\epsilon)$, the policy $u(\epsilon)$ satisfying (3.6) and (3.7) is $\hat{\epsilon}$-optimal for COP_γ, where $\hat{\epsilon}$ is given in Theorem 3.2.
(3) Let $\gamma(m)$ be a subsequence converging to β. Choose a sequence $\xi_m \to 0$. Let $r(m)$ be a ξ_m-optimal policy for $COP_{\gamma(m)}$ if $COP_{\gamma(m)}$ is feasible, otherwise let it be an arbitrary stationary policy. The stationary policy w obtained by applying the limiting procedure in Theorem 3.1 (ii) to the policies $r(m)$ is optimal for COP_β.

Proof: Here again, $\Gamma(U)$, for which we need to show the uniform convergence (Theorem 3.1 condition (2)) can be taken as the class of stationary policies (see Lemma 3.1). Assume first $0 < \beta < 1$ and B3(ii). For any γ such that $\gamma < \beta/\alpha$ (where α is the constant in assumption B3(ii)) we have

$$\begin{aligned}
\|C_\gamma(\bullet, u) - C_\beta(\bullet, u)\|_\mu &= \left\| \sum_{j=0}^{\infty} \left[(1-\gamma)\gamma^j - (1-\beta)\beta^j\right] [P^u]^j c(u) \right\|_\mu \\
&\leq \sum_{j=0}^{\infty} \left|(1-\gamma)\gamma^j - (1-\beta)\beta^j\right| \left(\frac{\alpha}{\beta}\right)^j R \\
&\leq R \sum_{j=0}^{\infty} \left|(\gamma^j - \beta^j)\right| \left(\frac{\alpha}{\beta}\right)^j = R \left| \frac{1}{1-\alpha} - \frac{1}{1-\alpha\beta^{-1}\gamma} \right|
\end{aligned}$$

Next we consider condition B3(i) instead of B3(ii). For $\beta < 1$ we have

$$\begin{aligned}
\|C_\gamma(\bullet, u) - C_\beta(\bullet, u)\|_\mu &= \left\| \sum_{j=0}^{\infty} \left[(1-\gamma)\gamma^j - (1-\beta)\beta^j\right] [P^u]^j c(u) \right\|_\mu \\
&\leq \sum_{j=0}^{\infty} \left|(1-\gamma)\gamma^j - (1-\beta)\beta^j\right| rR \\
&\leq rR \sum_{j=0}^{\infty} \left|(\gamma^j - \beta^j)\right| = rR \left| \frac{1}{1-\beta} - \frac{1}{1-\gamma} \right|
\end{aligned}$$

where r is given in (4.7). If $\beta = 1$ we have

$$\begin{aligned}
\|C_\gamma(\bullet, u) - C_\beta(\bullet, u)\|_\mu &= \left\| \left[\sum_{j=0}^{\infty} (1-\gamma)\gamma^j [P^u]^j - \Pi^u \right] c(u) \right\|_\mu \\
&= \left\| \sum_{j=0}^{\infty} (1-\gamma)\gamma^j \left[[P^u]^j - \Pi^u \right] c(u) \right\|_\mu \\
&\leq R \sum_{j=0}^{\infty} (1-\gamma)\gamma^j \rho\sigma^j = \frac{\rho R(1-\gamma)}{1-\gamma\sigma}
\end{aligned}$$

where σ and ρ are given in (4.6). Hence under the conditions of the Theorem, for any $\beta \in (0,1]$, $C_\gamma(x,u)$ converges to $C_\beta(x,u)$ uniformly in $U(S)$, and similarly with $D_\gamma(x,u)$. Thus the conditions of Theorem 3.1 and 3.2 are satisfied, which establishes the proof. ■

7 Application to a queueing model

Consider the following discrete time system (Altman and Shwartz [1], [2] Section 6, Nain and Ross [19], Spieksma [22]). Packets of information of N different types, such as data files, video and voice signals, compete for access to some shared resource. Each type of arriving packets waits in a buffer till it gets access to the resource. At the beginning of each time slot, priority is given to one of the traffic types according to some prespecified decision rule, and the packet is served for one unit of time. Service problems and errors due to noises are modeled by allowing the service to fail with positive (class dependent) probability. If the service is successful, the packet disappears from the system; otherwise, it remains in the queue. The problem COP_{queues} is to find a scheduling policy that minimizes a linear combination of the average delays of some types of traffic (typically, of the noninteractive types) subject to constraints on (linear combination of) average delays of other types (typically the interactive traffic).

At time t, M_t^i customers arrive to queue i, $1 \leq i \leq N$. Arrival vectors $M_t = (M_t^1, ..., M_t^N)$ are independent from slot to slot and form a renewal sequence with finite means λ_i. During a time slot (t,t+1) a customer from any class i, $1 \leq i \leq N$ may be served, according to some policy, which is a prespecified dynamic priority assignment. If served, with probability μ_i it completes its service and leaves the system; otherwise it remains in its queue. (μ_i should not be confused with the μ-norm introduced in Section 2. We use the same notation for both since both are widely used in Literature with this notation). A generic element of the state is given by $x = (x^1, x^2, ..., x^N)$ and it represents an N dimensional vector of the different queues' sizes. Throughout we restrict to non-idling policies.

Assume $\sum_{i=1}^N \lambda_i/\mu_i < 1$. Consider the linear cost function $c(x,a) := \sum_{i=1}^N c_i x^i$ and $d^k(x,a) = \sum_{i=1}^N d_i^k x^i$ for $1 \leq k \leq K$, where c_i and d_i^k are non-negative constants. Thus the costs $C(x,u)$ and $D^k(x,u)$ are related to linear combinations of expected average length of the different queues, and COP_{queues} has the form: find $u \in U$ that minimizes $C(x,u)$ s.t. $D^k(x,u) \leq V_k,\ k = 1, ..., K$, where V_k are given constants. Consider the expected average cost.

By Little's law these quantities are proportional to the respective waiting times in the different queues.

Let $\mathbf{G} = \{g_j\}$ be the set of all strict priority rules. A strict priority rule is a policy for which each type of customer has an index, and a customer of a given type is served only if there are no customers with lower priority in the system, and if it is the first in its queue. Let $|\mathbf{G}| = L$. For the unconstrained control problem, there exists an optimal policy within $\mathbf{G}$; it is the well known "μc" rule, for which the priorities are set according to increasing order of the $\mu_i c_i$ (see Baras et al. [7]). Thus, the queue for which $\mu_i c_i$ is the largest has the highest priority, and so on. Optimal policies for COP_{queues} are obtained by time multiplexing between the different g_j's. More specifically, define an L dimensional vector parameter $\alpha = (\alpha_1, \alpha_2, ..., \alpha_L)$, where α is a probability measure. Define a "cycle" as the time between two consecutive instants that the system is empty. During any cycle, a fixed g_j is used. A PTS policy $\hat{\alpha}$ is defined as a policy that chooses different policies g_j in such a way that the relative average number of cycles during which g_j was used is equal to α_j, as t goes to infinity (see definition in [1]). It is shown in [1] that

$$\overline{f}^1(x, \hat{\alpha}) = \sum_{j=1}^{L} \alpha_j \overline{f}^1(x, g_j) \tag{7.1}$$

and

$$C_1(x, \hat{\alpha}) = c \cdot \overline{f}^1(x, \hat{\alpha}) = \sum_{j=1}^{L} \alpha_j c \cdot \overline{f}^1(x, g_j) = \sum_{j=1}^{L} \alpha_j C_1(x, g_j). \tag{7.2}$$

where $\overline{f}$ is defined in (3.5). For a given $\delta > 0$, consider the following **LP**: find $\alpha \in \mathbb{R}^L$ that

$$\begin{gathered} \text{minimize } \sum_{j=1}^{L} \alpha_j C_1(x, g_j) \\ \text{subject to } \sum_{j=1}^{L} \alpha_j D_1^k(x, g_j) \leq V_k - \delta, \qquad k = 1, ..., K \\ \textstyle\sum_{j=1}^{L} \alpha_j = 1, \qquad \alpha \geq 0 \end{gathered} \tag{7.3}$$

$C_1(x, g_j)$ and $D_1^k(x, g_j)$ can be obtained as in Nain and Ross [19]. Let $\alpha^*(\delta)$ be the solution of LP with a given δ. Altman and Shwartz [1] show that $\hat{\alpha}^*(0)$ is an optimal policy for COP_{queues}.

Under B2, it can be shown that there exists some $\delta^* > 0$ such that $\hat{\alpha}^*(\delta^*)$ is feasible for COP_{queues} (this follows from the fact that the PTS policies are sufficient for COP_{queues}, see Altman and Shwartz [1]).

In the following Theorem we consider
(1) a sequence of problems COP^m_{queues} for the systems with queues of sizes $\mathbf{R}^m = (R^m_1, ..., R^m_N)$, $1 \le m \le \infty$, where COP^∞_{queues} is the one with all queues infinite, and $\mathbf{R}^m < \mathbf{R}^{m+1}$ (where strict inequality holds in at least one component). Assume without loss of generality that the initial state x satisfies $x \le \mathbf{R}^1$.
(2) Convergence in the discount factor.

Theorem 7.1 . *Assume B2 and* $\beta = 1$. *Then*
(i) $\lim_{m\to\infty} C(m; x) = C(x)$, $\lim_{\gamma\to 1} C_\gamma(x) = C_1(x)$.
(ii) Choose some $0 < \epsilon < 1$. *Let* $u(\epsilon)$ *be the PTS policy with*

$$\alpha = \epsilon\alpha^*(\delta^*) + (1 - \epsilon)\alpha^*(0). \tag{7.4}$$

Let $\hat{\epsilon} = \epsilon[C(x, \hat{\alpha}^*(\delta^*)) + 3]$. *Then for all* m *large enough* $u(\epsilon)$ *is* $\hat{\epsilon}$*-optimal for* COP^m_{queues} *(i.e. with finite queues of sizes* $\mathbf{R}^m$*); for all* β *close enough to 1,* $u(\epsilon)$ *is* $\hat{\epsilon}$*-optimal for* COP^β_{queues}.
(iii) In case that there are no constraints, for all m *large enough the* μc *rule is* ϵ*-optimal for* COP^m_{queues}*; for all* β *close enough to 1, it is* ϵ*-optimal for* COP^β_{queues}.

Proof: B3(i) and B3(iii) are satisfied by Spieksma [22] Theorem 9.1. Moreover, the function μ has the form

$$\mu(x) = \prod_{i=1}^{N} (1 + z_i)^{x^i} \tag{7.5}$$

where z are some numbers with $z_i > 0, \forall i$. It is easily seen that this implies (iv) in the finite approximation scheme **FAP**, and that the other features of **FAP** are satisfied as well. B1 is satisfied (see [2]) and B4 clearly holds too. The theorem then follows by applying Theorems 5.1 and 6.1. ∎

References

[1] E. Altman and A. Shwartz, "Optimal priority assignment: a time sharing approach", *IEEE Transactions on Automatic Control* Vol. AC-34 No. 10, pp. 1089-1102, 1989.

[2] E. Altman and A. Shwartz, "Markov decision problems and state-action frequencies," *SIAM J. Control and Optimization.* **29**, No. 4, pp. 786-809, 1991

[3] E. Altman and A. Shwartz, "Adaptive control of constrained Markov chains", *IEEE Transactions on Automatic Control*, **36**, No. 4, pp. 454-462, 1991.

[4] E. Altman and A. Shwartz, "Sensitivity of constrained Markov Decision Problems", *Annals of Operations Research*, **32**, pp. 1-22, 1991.

[5] E. Altman and V. A. Gaitsgory, "Stability and Singular Perturbations in Constrained Markov Decision Problems", to appear in *IEEE Transactions on Automatic Control.*

[6] E. Altman, "Denumerable constrained Markov Decision Processes and finite approximations", submitted to *MOR*, 1992.

[7] J. S. Baras, D. -J. Ma, and A. M. Makowski, "K competing queues with geometric service requirements and linear costs: the μc rule is always optimal," *Systems and Control Letters,* **6** No. 3 pp. 173-180, August 1985.

[8] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint", *J. Mathematical Analysis and Applications* **112**, 236-252, 1985.

[9] V. S. Borkar, "Ergodic control of Markov Chains with constraints – the general case", manuscript.

[10] R. Cavazos-Cadena, "Finite-state approximations for denumerable state discounted Markov Decision Processes", *J. Applied Mathematics and Optimization* **14** pp. 27-47, 1986.

[11] R. Dekker and A. Hordijk, "Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards", *Mathematics of Operations Research*, **13**, pp. 395-421, 1988.

[12] O. Hernandez-Lerma, “Finite state approximations for denumerable multidimensional - state discounted Markov decision processes”, *J. Mathematical Analysis and Applications* **113** pp. 382-389, 1986.

[13] K. Hinderer, *Foundation of Non-Stationary Dynamic Programming with Discrete Time Parameter*, vol. 33, Lecture Notes in Operations Research and Mathematical Systems, Springer-Verlag, Berlin, 1970.

[14] A. Hordijk and F. Spieksma, “Constrained admission control to a queuing system” *Advances of Applied Probability* Vol. 21, pp. 409-431, 1989.

[15] A. Hordijk, *Dynamic Programming and Markov Potential Theory*, Second Edition, Mathematical Centre Tracts 51, Mathematisch Centrum, Amsterdam, 1977.

[16] A. Hordijk and L. C. M. Kallenberg, “Constrained undiscounted stochastic dynamic programming”, *Mathematics of Operations Research,* **9**, No. 2, May 1984.

[17] L. C. M. Kallenberg, *Linear Programming and Finite Markovian Control Problems*, Mathematical Centre Tracts 148, Amsterdam, 1983.

[18] A. Lazar, “Optimal flow control of a class of queuing networks in equilibrium”, *IEEE Transactions on Automatic Control*, Vol. 28 no. 11, pp. 1001-1007, 1983.

[19] P. Nain and K. W. Ross, ”Optimal Priority Assignment with hard Constraint,” *Transactions on Automatic Control,* Vol. 31 No. 10, pp. 883-888, October 1986.

[20] L. I. Sennott, “Constrained discounted Markov decision chains”, *Probability in the Engineering and Informational Sciences*, **5**, pp. 463-475, 1991.

[21] L. I. Sennott, “Constrained average cost Markov decision chains”, to appear in *Probability in the Engineering and Informational Sciences.*

[22] F. M. Spieksma, *Geometrically Ergodic Markov Chains and the Optimal Control of Queues*, Ph.D. thesis, 1990, University of Leiden.

[23] L. C. Thomas and D. Stengos, “Finite State Approximation Algorithms for Average Cost Denumerable State Markov Decision Processes”, *OR Spectrum*, **7**, pp. 27-37, 1985.

[24] D. J. White, "Finite State Approximations for Denumerable State Infinite Horizon Discounted Markov Decision Processes with Unbounded Rewards", *J. Mathematical Analysis and Applications* **86**, pp. 292-306, 1982.