DENUMERABLE CONSTRAINED MARKOV DECISION PROCESSES AND FINITE APPROXIMATIONS

Eitan Altman
INRIA
Centre Sophia Antipolis
06565 Valbonne Cedex, France

Submitted: December 1990
Revised: December 1991, September 1992

Abstract

The purpose of this paper is two fold. First to establish the Theory of discounted constrained Markov Decision Processes with a countable state and action spaces with general multi-chain structure. Second, to introduce finite approximation methods.

We define the occupation measures and obtain properties of the set of all achievable occupation measures under the different admissible policies. We establish the optimality of stationary policies for the constrained control problem, and obtain a LP with a countable number of decision variables through which stationary optimal policies are computed.

Since for such a LP one cannot expect to find an optimal solution in a finite number of operations, we present two schemes for finite approximations and establish the convergence of optimal values and policies for both the discounted and the expected average cost, with unbounded cost.

Sometimes it turns to be easier to solve the problem with infinite state space than the problem with finite yet large state space. Based on the optimal policy for the problem with infinite state space, we construct policies which are almost optimal for the problem with truncated state space. This method is applied to obtain an ϵ-optimal policy for a problem of optimal priority assignment under constraints for a system of K finite queues.

Keywords: Constrained Markov Decision Processes, countable state space, Linear Programming approach, Finite approximations.

1 INTRODUCTION

In recent years growing attention was given to solving constrained MDPs (Markov Decision Processes, also known as controlled Markov chains). Such problems frequently arise in computer networks and data communications. Lazar [22] and Hordijk and Spieksma [15] have considered flow control problems with constraints in queueing networks, e.g. maximize throughputs, with constraints on delays of different traffic types. Nain and Ross [23], Altman and Shwartz [1], Ross and Chen [28] considered optimal priority assignment under constraints in communication networks; a typical problem there is to minimize delay of non-interactive traffic, with constraints on delays of interactive traffic. The theory for solving general constrained MDPs with expected average cost was developed by Hordijk and Kallenberg [19], [20] who solve the case of finite state and finite action spaces using methods based on Linear Programming (LP), and by Beutler and Ross [8] who considered a single constraint for the case of compact action space. An LP approach for solving the discounted problem is given in Kallenberg [20] and Altman and Shwartz [4]. The LP approach was generalized by Altman and Shwartz [2] and Spieksma [31] to the countable state case, with expected average cost. The Lagrange approach in [8] for treating the case of a single constraint was generalized to the countable state space by Sennott [24], [25] who considers both the expected average and the discounted cost.

The first goal of this paper is to establish the theory of constrained MDPs with several constraints for the discounted cost with a general multi-chain structure and with a countable state space. We show that a stationary optimal policy exists, and that the control problem can be reduced to an LP with an infinite number of decision variables and constraints. The second goal of the paper is to obtain methods that enable the numerical solution of the control problem.

It is interesting to note that the LP obtained in in [2] and [31] for the expected average cost (under the unichain assumption) is the same as the one obtained by substituting one for the discount factor in the new LP (for the discounted cost). However, unlike the LP for the expected average cost, the parameters of the new LP depend on the initial state, and thus the optimal value and policies depend on the initial state. The fact that the optimal policies depend on the initial state for the discounted cost was exhibitted in Example 5.3 in [3], and is in contrast to unconstrained problems.

In order to reduce the control problem to an LP, we first establish the linear representation of the cost as a function of the "occupation measure" (which generalizes the notion of "state-action frequencies" [13] used for the expected average cost). Under some conditions, it is known for the single chain case, that the set of occupation measures achieved by all admissible policies is equal to the set of occupation measures achieved by the stationary policies. Moreover, it is compact, convex, and its extreme points correspond to stationary deterministic policies (see [2, 9, 13, 19]). We show that these properties also hold for the multi-chain case with a countable state space and discounted cost. This allows us to establish the optimality of a stationary policy.

Since the number of states and actions in the control problem (and the number of decision variables in the corresponding LP) is not finite, the question of approximations arises. If a finite approximation of the state space is used, a finite LP can be applied to obtain an optimal policy for the approximating model (e.g. [19, 4]). The issue of approximating MDPs by other finite state MDPs was investigated in a few papers of White (e.g. [33]), Cavazos-Cadena [12], Hernandez-Lerma [16] and Thomas and Stengos [32]. Since in our case additional constraints are introduced in the control problem, the methods described in previous papers (e.g. [12, 16, 32, 33]) are not applicable, since they are all based on dynamic programming techniques. Unlike the unconstrained problem it is known [19] that there may not exist a stationary deterministic optimal policy for the constrained problem. This, and the fact that constrained MDPs are usually solved using LP methods, imply that new approaches for approximating constrained MDPs should be used.

Finite state approximations techniques may serve other purposes than to solve a problem with a countable number of decision variables. In some problems in queueing networks it is possible to obtain optimal policies for the countable state case, whereas there are no explicit solutions for similar finite state problems. This is the case e.g. in the problem of optimal priority assignment where N infinite queues compete for the attention of a single server. A simple index rule (the "μc" rule, see [7] and references therein) is known to minimize a given weighted sum of the expected waiting times. When additional constraints are added, there exists an optimal policy that multiplexes between different strict priority rules ([1, 23, 28]). The theory on finite approximations of MDPs which we establish can thus be used to approximate the solution to the case of large queues' sizes by the known solution for infinite queues' sizes. We demonstrate that in Section 9.

We introduce in this paper two approximating schemes, for which we establish conditions for the convergence of optimal values and policies. Moreover, we obtain conditions for the stability of the (almost) optimal policy; i.e. we show that an (almost) optimal policy for the original problem is almost optimal for a finite approximation of the problem.

The structure of the paper is as follows: after presenting the model notation and assumptions in Section 2, we present some properties of the occupation measures in Section 3, and relate them to the cost. Conditions for the existence of a stationary optimal policy are presented in Section 4. In Section 5 we present the LP and establish the relation to the original control problem. In Section 6 we obtain a key Theorem for approximation. In Section 7 we then present a finite approximation scheme based on replacing the countable state space with a finite state space. In Section 8 we introduce another general scheme for finite approximations for constrained problems, based on arbitrarily picking the probabilities of choosing the different actions in all but a finite number N of states. Then the probabilities in the remaining N states are chosen so as to optimize the restricted problem. The problem of optimal priority assignment for N competing finite queues is finally treated in Section 9.

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, $\mathbf{A}(y)$. With some abuse of notation, $\mathbf{X} \times \mathbf{A}$ will denote all possible pairs $(y, a) : y \in \mathbf{X}, a \in \mathbf{A}(y)$. 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. (A more detayled definition and construction of the MDP can be found e.g. in [30]).

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 $P^g_{xy} := \sum_{a \in \mathbf{A}} p^g_{a|x} \mathcal{P}_{xay}$.

The class of *stationary deterministic policies* $U(SD)$ is a subclass of $U(S)$ and, with some abuse of notation, every $g \in U(SD)$ is identified with a mapping $g : \mathbf{X} \to \mathbf{A}$, so that $p^g_{\cdot|x} = \delta_{g(x)}(\cdot)$ is concentrated at the point $g(x)$ in $\mathbf{A}$ for each x.

For any (finite or countable) set B, let $M(B)$ denote the set of probability measures on B endowed with the topology of weak convergence $\tau(B)$. Note that $U(S)$ can be identified with the set $\prod_{y \in \mathbf{X}} M(\mathbf{A}(y))$. Clearly $U(S)$ is compact with respect to the product topology $\prod_{y \in \mathbf{X}} \tau(\mathbf{A}(y))$.

2.2 The constrained problem

Let $C(x,u)$ and $D(x,u) := \{D^k(x,u)\ ,\ 1 \le k \le K\}$ 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

$$D^k(x,u) \le V_k\ , \quad k = 1, 2, \dots, K \tag{1}$$

The constrained control problem is:

(**COP**): Find a feasible $v \in U$ that minimizes $C(x,u)$.

$C(x,u)$ 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. Let $0 < \beta \le 1$ be a discount factor. We assume throughout the paper that $\forall u \in U,\ x \in \mathbf{X}$, $E_x^u c(X_s, A_s)$ and $E_x^u d^k(X_s, A_s)$, $k = 1, ..., K$ exist. We shall use the following *normalized discounted* cost functionals from $\mathbf{X} \times U$ to

$\mathbb{R}$ (see [4]).

$$
\begin{aligned}
C_\beta^t(x,u) &:= \left(\sum_{r=0}^t \beta^r\right)^{-1} E_x^u \left[\sum_{r=0}^t \beta^r c(X_r,A_r)\right] \\
D_\beta^{k,t}(x,u) &:= \left(\sum_{r=0}^t \beta^r\right)^{-1} E_x^u \left[\sum_{r=0}^t \beta^r d^k(X_r,A_r)\right] \qquad k=1,\dots,K
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
C_\beta(x,u) &:= \overline{\lim}_{t\to\infty} C_\beta^t(x,u) \\
D_\beta^k(x,u) &:= \overline{\lim}_{t\to\infty} D_\beta^{k,t}(x,u) \qquad k=1,\dots,K
\end{aligned}
\tag{3}
$$

When $\beta=1$, (2-3) reduce to the well-known definition of the expected average cost, since in that case $\sum_{r=0}^t \beta^r = t+1$. For $\beta<1$ the definition (3) are slightly different than the *standard expected discounted* costs (see e.g. [27])

$$
\begin{aligned}
\tilde{C}_\beta(x,u) &:= E_x^u \left[\sum_{r=0}^\infty \beta^r c(X_r,A_r)\right] \\
\tilde{D}_\beta^k(x,u) &:= E_x^u \left[\sum_{r=0}^\infty \beta^r d^k(X_r,A_r)\right] \qquad k=1,\dots,K
\end{aligned}
\tag{4}
$$

If $c(\cdot,\cdot)$ is bounded from either below or from above, then $\tilde{C}_\beta(x,u)$ is well defined and

$$
C_\beta(x,u) = (1-\beta)\tilde{C}_\beta(x,u) = (1-\beta)\sum_{r=0}^\infty \beta^r E_x^u c(X_r,A_r) \tag{5}
$$

(and similarly for $D_\beta^k(x,u)$). However, the cost (3) is defined for cases for which the limit as $t\to\infty$ of the partial sums (till time t) in (4) does not exist. An advantage of using the new definition (3) is that the additional constant of $(1-\beta)$ that appears in (5) enables to obtain the same LP for both the discounted and expected average cost for solving COP (see e.g. [4] for the finite case, and Section 5 in this paper for the countable case). This enables to reduce problems of continuity, sensitivity [4] and singular perturbations [5] of 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 [14].

2.3 Occupation measure

Given a discount factor $0 < \beta \leq 1$, define the collections $\{\bar{f}_{sa}^{t,\beta}(x,u;y,a)\}_{y,a}$ and $\{\bar{f}_{s}^{t,\beta}(x,u;y)\}_y$ by

$$\bar{f}_{sa}^{t,\beta}(x,u;y,a) := [\sum_{r=0}^{t} \beta^r]^{-1} \sum_{r=0}^{t} \beta^r P_x^u(X_r = y, A_r = a) \tag{6}$$

$$\bar{f}_{s}^{t,\beta}(x,u;y) := [\sum_{r=0}^{t} \beta^r]^{-1} \sum_{r=0}^{t} \beta^r P_x^u(X_r = y) \tag{7}$$

By $\bar{f}_{sa}^{\beta}(x,u) := \{\bar{f}_{sa}^{\beta}(x,u;y,a)\}_{ya}$ we denote a generic accumulation point of $\bar{f}_{sa}^{t,\beta}(x,u)$. Similarly, we define $\bar{f}_{s}^{\beta}(x,u) := \{\bar{f}_{s}^{\beta}(x,u;y)\}_y$ to be a generic accumulation point of $\bar{f}_{s}^{t,\beta}(x,u)$. These quantities are known as the "occupation measures" (see e.g. [9]) and for the case of $\beta = 1$ they are also known as the *state-action frequencies*. For $\beta < 1$ there is a single accumulation point

$$\bar{f}_{sa}^{\beta}(x,u;y,a) = [1-\beta] \sum_{t=0}^{\infty} \beta^t P_x^u(X_t = y, A_t = a) \tag{8}$$

$$\bar{f}_{s}^{\beta}(x,u;y) = [1-\beta] \sum_{t=0}^{\infty} \beta^t P_x^u(X_t = y) \tag{9}$$

For each β, t, x and u, $\bar{f}_{sa}^{t,\beta}(x,u)$ can be considered as a probability measure over $\mathbf{X} \times \mathbf{A}$. For $\beta < 1$, $\bar{f}_{sa}^{\beta}(x,u)$ and $\bar{f}_{s}^{\beta}(x,u)$ are also probability measures, and hence $\sum_{y,a} \bar{f}_{sa}^{\beta}(x,u;y,a) = \sum_y \bar{f}_{s}^{\beta}(x,u;y) = 1$. In order to ensure that this property holds for the expected average case i.e. $\beta = 1$, we need assumptions A3(i) or A3(ii) defined below.

2.4 Assumptions and Notation

The following assumptions are used frequently in the paper:

A1: Under any $g \in U(SD)$, $\mathbf{X}$ contains a single ergodic class, and absorption into the positive recurrent class takes place in finite expected time.

A1': Under any $g \in U(SD)$, $\mathbf{X}$ consists of a single ergodic class (with no transient states).

A2: there exists some policy u (not necessarily an optimal policy) such that

$$D^k(x,u) < V_k \ , \quad k = 1, 2, \ldots, K$$

A3(i) For a given initial state x, the set $\{\bar{f}^{t;1}_{sa}(x,u)\}$, $u \in U(S)$ is tight.
A3(ii) The family of stationary probabilities $\{\pi^u(\cdot)\}$, $\pi^u \in M(\mathbf{X})$, corresponding to policies $u \in U(S)$, is tight.
It is known that {A1 and A3(i)} imply A3(ii); moreover under A1', A3(i) and A3(ii) are equivalent (see [2] Section 4, [18] Sections 10, and [31] p. 171). Other sufficient conditions for tightness can be found in [18] Section 11 and [2] Section 4.

Let $0 \in \mathbf{X}$ be a recurrent state under $g \in U(S)$. With the convention that $\inf \emptyset = \infty$, define $\eta(1) \stackrel{\text{def}}{=} \inf\{t \geq 0 : X_t = 0\}$, $\eta(k+1) \stackrel{\text{def}}{=} \inf\{t > \eta(k) : X_t = 0\}$, where $\eta(k) = \infty$ implies $\eta(k+1) = \infty$. Define the following assumption:

$$\mathbf{A4(u)}: \ E^u_x \left[\sum_{r=0}^{\eta(1)-1} |c(X_r, A_r)| \right] = \infty \ implies \ E^u_x \left[\sum_{r=\eta(1)}^{\eta(2)-1} |c(X_r, A_r)| \right] = \infty,$$

for a given $u \in U(S)$;
A4: A4(u) holds for all $u \in U(S)$. (Note that A1' implies A4).

The following notation is used below: $\delta_a(x)$ is the Kronecker delta function. 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 Occupation measures and cost

We relate below the cost to the occupation measure.

Lemma 3.1 *For each instantaneous cost $c(\cdot,\cdot)$, $y \in \mathbf{X}, a \in \mathbf{A}$, $u \in U$*

$$C_\beta(x,u) = \sum_{y \in \mathbf{X}, a \in \mathbf{A}} c(y,a) \bar{f}_{sa}^{\beta}(x,u;y,a) \tag{10}$$

holds for $\beta < 1$ provided that either

(i) c is bounded from below or from above, or

(ii) $\{c(X_s, A_s)\}_s$ are uniformly integrable with respect to P^u.

Proof: We first prove (i). Assume first that $c(\cdot,\cdot) \geq M$, for some $M \in \mathbb{R}$. If we consider $[c(\cdot,\cdot) - M]$ as a measure and further consider $\left(\sum_{r=0}^{t} \beta^r\right) \bar{f}_{sa}^{t,\beta}(x,u)$ as a RV on $\mathbf{X} \times \mathbf{A}$ then we obtain from the Monotone Convergence Theorem

$$\lim_{t\to\infty} \left(\sum_{r=0}^{t} \beta^r\right) \sum_{y,a} [c(y,a) - M] \bar{f}_{sa}^{t,\beta}(x,u) = (1-\beta)^{-1} \sum_{y,a} [c(y,a) - M] \bar{f}_{sa}^{\beta}(x,u)$$

It then follows that

$$C_\beta(x,u) = \lim_{t\to\infty} \sum_{y,a} c(y,a) \bar{f}_{sa}^{t,\beta}(x,u)$$

$$= M + (1-\beta) \lim_{t\to\infty} \left(\sum_{r=0}^{t} \beta^r\right) \sum_{y,a} [c(y,a) - M] \bar{f}_{sa}^{t,\beta}(x,u) = \sum_{y,a} c(y,a) \bar{f}_{sa}^{\beta}(x,u)$$

A similar proof holds for the case that $c(\cdot,\cdot)$ is bounded above.

The proof of (ii) is the same as for the expected average cost, see [2] Lemma 2.2. ■

Next we quote a similar representation for the expected average case, which is known to hold under condition similar to (ii) of the Lemma above, but in general need not hold under conditions similar to (i).

Lemma 3.2 *([2] Lemma 2.2 and 2.3) Let $\beta = 1$. Let $u \in U$ be such that a single limit exists to $\bar{f}_{sa}^{t,1}(x,u)$. Then for any instantaneous cost $c(\cdot,\cdot)$, $y \in \mathbf{X}, a \in \mathbf{A}$, (10) holds provided that A1 and A3(i) hold and one of the following is true:*

(i) $\{c(X_s, A_s)\}_s$ are uniformly integrable with respect to P^u,

(ii) A4(u) holds, c is bounded from either below or from above, and $u \in U(S)$.

Let L_x^β denote the set $\{\bar{f}_{sa}^\beta(x,u)\}$ achieved by all policies in U, $L_x^\beta(SD)$ the set achieved by all policies in $U(SD)$ and $L_x^\beta(S)$ the set achieved by all policies in $U(S)$. The following Theorem states the "completeness" of stationary policies i.e. $L_x^\beta = L_x^\beta(S)$, as well as the compactness of L_x^β (with respect to the topology $\tau(\mathbf{X}\times\mathbf{A})$ defined in the previous section).

Theorem 3.1 *For any $0<\beta\le 1$, L_x^β is convex. Assume either $\beta<1$ or $\{A1$ and $A3(i)\}$. Then $L_x^\beta(S)=L_x^\beta$ is compact and is equal to the convex hull of $L_x^\beta(SD)$.*

For proving the theorem we need the following Lemma

Lemma 3.3 *Choose some $g\in U(S)$ and a state y. Define $g_a\in U(S)$ to be the policy that chooses always action a when in state y, and otherwise behaves exactly like g. Then for every $0<\beta\le 1$ there exists a probability measure α on $\mathbf{A}(y)$ such that*
(i) For any cost function c which is bounded below,

$$C_\beta(x,g)=\sum_{a\in\mathbf{A}(y)}\alpha(a)C_\beta(x,g_a)$$

(ii)

$$\bar{f}^\beta(x,g)=\sum_{a\in\mathbf{A}(y)}\alpha(a)\bar{f}^\beta(x,g_a)$$

Proof: For $\beta=1$ see e.g. Key Lemma in [31] p. 168.

Define the stopping times $\sigma(y)\stackrel{\text{def}}{=}\inf_{r>0}\{X_r=y\}$, $y\in\mathbf{X}$, with the convention that $\inf\{\emptyset\}=\infty$. Denote

$$\alpha(a)\stackrel{\text{def}}{=}\frac{p^g_{a|y}\left(1-E_y^{g_a}\beta^{\sigma(y)}\right)}{\sum_{a'}p^g_{a'|y}(1-E_y^{g_{a'}}\beta^{\sigma(y)})}$$

Consider an arbitrary immediate cost function $c(\cdot,\cdot)$, and define

$$W_g^\beta(x,y)\stackrel{\text{def}}{=}(1-\beta)E_x^g\left[\sum_{r=0}^{\sigma(y)-1}\beta^r c(X_r,A_r)\right]$$

The cost $C_\beta(x,g)$ can be expressed as

$$C_\beta(x,g) = W_g^\beta(x,y) + (1-\beta)E_x^g\left[\beta^{\sigma(y)}\sum_{r=0}^{\infty}\beta^r c(X_{r+\sigma(y)}, A_{r+\sigma(y)})\right] = W_g^\beta(x,y) + C_\beta(y,g)E_x^g\beta^{\sigma(y)}$$

In particular, we have for $x = y$:

$$C_\beta(y,g) = W_g^\beta(y,y) + C_\beta(y,g)E_y^g\beta^{\sigma(y)}$$

Hence we obtain:

$$C_\beta(y,g) = \frac{W_g^\beta(y,y)}{1 - E_y^g\beta^{\sigma(y)}} = \frac{\sum_a p_{a|y}^g W_{g_a}^\beta(y,y)}{\sum_{a'} p_{a'|y}^g(1 - E_y^{g_{a'}}\beta^{\sigma(y)})}$$

$$= \sum_a C_\beta(y,g_a)\frac{p_{a|y}^g\left(1 - E_y^{g_a}\beta^{\sigma(y)}\right)}{\sum_{a'} p_{a'|y}^g(1 - E_y^{g_{a'}}\beta^{\sigma(y)})} = \sum_a \alpha(a)C_\beta(y,g_a)$$

which establishes the proof for the case $x = y$. For $x \neq y$,

$$C_\beta(x,g) = W_g^\beta(x,y) + C_\beta(y,g)E_x^g\beta^{\sigma(y)}$$

$$= \sum_a \alpha(a)\left[W_g^\beta(x,y) + E_x^g\beta^{\sigma(y)}C_\beta(y,g_a)\right] = \sum_a \alpha(a)C_\beta(x,g_a)$$

since for $x \neq y$, $W_g^\beta(x,y) = W_{g_a}^\beta(x,y)$. This establishes (i). (ii) is then obtained by choosing the immediate cost to be $c(y',a') = 1\{y' = y, a' = a\}$ and applying (i). ■

Proof of Theorem 3.1: The convexity of L_x^β follows from the fact that for any sequence of policies $u(1), u(2), ... \in U$, initial state x, time t and any distribution α on the set of integers, there exists a policy ("Markov" policy) v such that

$$\sum_{k=1}^{\infty}\alpha_k P_x^{u(k)}(X_t = \bullet, A_t = \bullet) = P_x^v(X_t = \bullet, A_t = \bullet)$$

(see e.g. Theorem 13.2 [18]).

Next we prove that $L_x^\beta = L_x^\beta(S)$. For the case $\beta = 1$, the latter is given in [2] Cor. 3.3. Let $\beta < 1$. ([9] presents another more complex proof for the infinite case, but for a single recurrent class).

Choose any policy $u \in U$ and define the set $\alpha \in \mathbb{R}^{|\mathbf{X}\times\mathbf{A}|}$ as

$$\alpha_y^a \stackrel{\text{def}}{=} \frac{\bar{f}_{sa}^{\beta}(x,u;y,a)}{\bar{f}_s^{\beta}(x,u;y)}$$

whenever the denominator is nonzero. When it is zero, α_y^a is any arbitrary real positive number such that $\sum_a \alpha_y^a = 1$. Define the stationary policy g as $p_{a|y}^g \stackrel{\text{def}}{=} \alpha_y^a$. We show below that $\bar{f}_{sa}^{\beta}(x,u) = \bar{f}_{sa}^{\beta}(x,g)$. Note that $P_{vy}^g = \sum_a P_{vay}\alpha_y^a$.

$$\begin{aligned}
\bar{f}_s^{\beta}(x,u;y) &= (1-\beta)\delta_x(y) + \beta(1-\beta)\textstyle\sum_{r=1}^{\infty}\beta^{r-1}P_x^u(X_r=y) \\
&= (1-\beta)\delta_x(y) + \beta(1-\beta)\textstyle\sum_{r=0}^{\infty}\sum_{v\in\mathbf{X}}\sum_{a\in\mathbf{A}}\beta^r P_x^u(X_r=v,A_r=a)P_{vay} \\
&= (1-\beta)\delta_x(y) + \beta\textstyle\sum_{v\in\mathbf{X}}\sum_{a\in\mathbf{A}}P_{vay}\bar{f}_{sa}^{\beta}(x,u;v,a) \qquad (11)\\
&= (1-\beta)\delta_x(y) + \beta\textstyle\sum_{v\in\mathbf{X}}\sum_{a\in\mathbf{A}}P_{vay}\bar{f}_s^{\beta}(x,u;v)\alpha_y^a \\
&= (1-\beta)\delta_x(y) + \beta\textstyle\sum_{v\in\mathbf{X}}P_{vy}^g\bar{f}_s^{\beta}(x,u;v)
\end{aligned}$$

In matrix notation the above equation yields

$$\bar{f}_s^{\beta}(x,u)[I-\beta P^g] = (1-\beta)\delta_x$$

where δ_x stands for the row vector whose xth entry is equal to one and all other entries are zero. Since $\beta < 1$, $[I-\beta P^g]$ is invertible (see Appendix, Lemma 10.1) and hence

$$\bar{f}_s^{\beta}(x,u) = (1-\beta)\delta_x[I-\beta P^g]^{-1} \qquad (12)$$

Since (12) clearly holds for the case $u = g$, we have

$$\bar{f}_s^{\beta}(x,u) = \bar{f}_s^{\beta}(x,g)$$

It then follows that

$$\bar{f}_{sa}^{\beta}(x,u;y,a) = \bar{f}_s^{\beta}(x,u;y)\alpha_y^a = \bar{f}_s^{\beta}(x,g;y)p_{a|y}^g = \bar{f}_{sa}^{\beta}(x,g;y,a)$$

and hence $L_x^{\beta}(S) = L_x^{\beta}$.

The compactness of L_x^β is established by showing that $\bar{f}_{sa}^\beta(x,u) : U(S) \to L_x^\beta$ is continuous (since $U(S)$ is compact in $\prod_{y \in \mathbf{X}} \tau(\mathbf{A}(y))$). To show this, note first that for $u \in U(S)$, $P_{xy}^u = \sum_a \mathcal{P}_{xay} p_{a|x}^u$ is continuous in u and hence the transition probability matrix P^u is continuous in u. (P^u is an element of the space $[M(\mathbf{X})]^{\mathbf{X}}$ endowed with the product topology $[\tau(\mathbf{X})]^{\mathbf{X}}$). We claim that $\bar{f}_s^\beta(x,u) : U(S) \to L_x^\beta$ is continuous in u. For the case of $\beta < 1$ this follows from Lemma 10.2 in the appendix since $\bar{f}_s^\beta(x,u) = (1-\beta)\sum_{r=0}^\infty \beta^r [P^u]^r$. For $\beta = 1$, A1 and A3(i) imply A3(ii) (see [2] Lemma 4.1), and so indeed the steady-state probability $\bar{f}_s^1(x,u) = \pi^u$, $u \in U(S)$ is continuous in u ([18] Lemma 10.2 p. 83). Finally, since $\bar{f}_{sa}^\beta(x,u;y,a) = \bar{f}_s^\beta(x,u;y) p_{a|y}^u$ it follows that $\bar{f}_{sa}^\beta(x,u) : U(S) \to L_x^\beta$ is continuous, which establishes the compactness of L_x^β.

Next we show that $L_x^\beta(S)$ is equal to the convex hull of $L_x^\beta(SD)$. Since it is compact, by the Krein-Milman theorem it is the convex hull of its extreme points. Choose some $g \in U(S)$. Suppose that g is not deterministic. Then there exists a state y where at least two different actions have positive probabilities to be chosen by g. But then by Lemma 3.3, g is not an extreme point of L_x^β. ■

Remarks:
(i) For the case $\beta = 1$, all the statements in Theorem 3.1 for $L_x^1(S)$ except for $L_x^1(S) = L_x^1$ remain valid with A3(ii) replacing assumption A3(i).
(ii) For the case $\beta = 1$, Theorem 3.1 is proven in [2] Thm 5.1 under the stronger assumptions A1' and A3(i) using a different approach.

4 Optimality of Stationary Policies

Theorem 4.1 *Assume either $\beta < 1$, or A1, A3(i) and A4 holds. Further assume that the immediate costs are bounded below. Then*
(i) for each policy $u \in U$, there exists a stationary policy g such that

$$C_\beta(x,g) \le C_\beta(x,u), \qquad D_\beta^k(x,g) \le D_\beta^k(x,u), \quad k = 1,2,...,K \tag{13}$$

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: For the case $\beta = 1$ Altman and Shwartz establish (i) in [2] Thm. 2.8 (and its proof). Borkar proves in [10] that when restricting COP to $U(S)$, there exists a stationary optimal policy, if COP is feasible. (An alternative proof is given in Cor. 5.4. in [2] with A1' replacing A1). Combining these facts establishes (ii).

For the case of $\beta < 1$, (i) follows from Lemma 3.1 and Theorem 3.1. The proof of (ii) is exactly the same as the proof for the case $\beta = 1$, see [2] Cor. 5.4. (which is based on the fact that L_x^β is compact and on showing that the costs $C_\beta(x,u)$ and $D_\beta^k(x,u)$ are lower semicontinuous functions of $\bar{f}_{sa}^\beta(x,u)$). ■

In the expected average case, the tightness assumption A3(i) in Theorem 4.1 can be replaced by some structure on the immediate costs, that makes it optimal to use policies for which tightness holds. $c(\cdot,\cdot)$ is said to be "V-almost monotone" if there exists an increasing collection of compact (finite) subsets K_i of $\mathbf{X} \times \mathbf{A}$ such that $\cup_i K_i = \mathbf{X} \times \mathbf{A}$, and such that the cost function $c(y,a)$ satisfies

$$\lim_{i\to\infty} \inf\{c(y,a); (y,a) \notin K_i\} \geq V.$$

Theorem 4.2 *Assume A1 and A4, and $\beta = 1$. Assume $c(\cdot,\cdot)$ is V_0-monotone for some constant V_0, and $d^k(\cdot,\cdot)$ is V_k-monotone, $1 \leq k \leq K$. If there exists a policy $u' \in U'$ such that $C(x,u') \leq V_0$ and $D^k(x,u') \leq V_k$, $1 \leq k \leq K$, then an optimal policy for COP exists within U(S).*

Proof: The sufficiency of the stationary policies under the conditions of the Theorem is established by combining Theorem 3.2 and Lemma 4.6 in [2]. Borkar proves in [10] that when restricting COP to $U(S)$, there exists a stationary optimal policy, if COP is feasible. Combining these facts establishes the Theorem. ■

Remarks:
(i) Borkar [10] has established recently for the average cost under general ergodic conditions that

the policy that is optimal among $U(S)$ can be chosen such that the number of randomizations is not greater than the number of constraints plus one. A simple alternative proof for that result (under stronger conditions) is given in Remark 3 in the end of Section 7. This proof establishes this property for discounted costs as well.

(ii) A sufficient condition for a cost c to be V-almost monotone for all V, is that for any V, there exists a finite set $W_V \subset \mathbf{X} \times \mathbf{A}$ such that $c(y,a) > V$ for all $(y,a) \notin W_V$.

5 Equivalent Infinite Linear Programming

We show below for the discounted cost criteria that COP is equivalent to a LP with countable number of decision variables and a countable number of constraints. Such equivalence is known in the finite case (see [19] for $\beta = 1$ and [4] for $\beta < 1$), and it is then used as a method for computing stationary optimal policies. A similar equivalence was shown to hold also for the countable case with $\beta = 1$, see [2], [31]; moreover the LP for solving that case is obtained from the LP below just by substituting in it $\beta = 1$.

Consider the following LP:

$\mathbf{LP}_\beta$: Find the infimum of $\mathcal{C}(z) := \sum_{y,a} c(y,a) z(y,a)$ subject to:

$$\sum_{y,a} z(y,a)\,[\delta_v(y) - \beta P_{yav}] = [1-\beta]\delta_v(x) \qquad v \in \mathbf{X} \tag{14}$$

$$\mathcal{D}^k(z) := \sum_{y,a} d^k(y,a) z(y,a) \le V_k \qquad 1 \le k \le K \tag{15}$$

$$\sum_{y,a} z(y,a) = 1 \tag{16}$$

$$z(y,a) \ge 0 \tag{17}$$

Define $g(z)$ to be any stationary policy such that $p_{a|y} = z(y,a)[\sum_a z(y,a)]^{-1}$ whenever the denominator is nonzero.

Theorem 5.1 *Assume that $\beta < 1$ and let c and d^k, $k = 1, ..., K$ be bounded from below or from above.*

(1.1) For every policy u, $\zeta \stackrel{\text{def}}{=} \bar{f}^{\beta}_{sa}(x,u)$ satisfies (14), (16) and (17).

(1.2) $\mathcal{C}(\zeta) = C_\beta(x,u)$ and $\mathcal{D}^k(\zeta) = D^k_\beta(x,u) \quad 1 \le k \le K$. Consequently, if u is feasible for COP_β then ζ satisfies (15).

(2) Choose any ζ that satisfies (14) (16) and (17). Then

(2.1) $\bar{f}^{\beta}_{sa}(x, g(\zeta)) = \zeta$.

(2.2) $\mathcal{C}(\zeta) = C_\beta(x, g(\zeta))$ and $\mathcal{D}^k(\zeta) = D^k_\beta(x, g(\zeta)) \quad 1 \le k \le K$.

(2.3) Assume that ζ satisfies also (15). Then $g(\zeta)$ is a feasible policy for COP_β.

(3.1) COP is feasible iff LP_β is.

(3.2) If LP_β is feasible then there exists some z^* that achieves the optimal value of LP_β; the optimal value for COP is equal to $\mathcal{C}(z^*)$; $g(z^*)$ is an optimal policy for COP.

(3.3) Suppose that g^* is an optimal policy for COP. Then $z^* \stackrel{\text{def}}{=} \bar{f}^{\beta}_{sa}(x, g^*)$ is an optimal solution for LP_β (and thus achieves its infimum).

Proof: (1.1) follows immediately from (11). Lemma 3.1 then implies (1.2).

To prove (2.1), define the vector $\zeta_s \in \mathbb{R}^{|\mathbf{X}|}$ by $\zeta_s(y) \stackrel{\text{def}}{=} \sum_a \zeta(y,a)$. (14) implies (in vector notation) that

$$\zeta_s = (1-\beta)\delta_x [I - \beta P^{g(\zeta)}]^{-1} \tag{18}$$

where δ_x stands for the row vector whose xth entry is equal to one and all other entries are zero. (It is shown in Lemma 10.1 that indeed $I - \beta P$ is invertible for any stochastic matrix P defined on $\mathbf{X} \times \mathbf{X}$). Hence for each y, (12) implies that

$$\sum_a \bar{f}^{\beta}_{sa}(x, g(\zeta); y, a) = \sum_a \zeta(y,a)$$

(2.1) is then established by the definition of $g(\zeta)$. Lemma 3.1 now implies (2.2) and (2.3). Finally, statements (3) of the Theorem follow from statements (1) and (2) and Theorem 4.1. ∎

6 Key Theorems for Approximation

In this Section we discuss approximations of constrained Markov Processes. Assume that for a given initial state x and any policy u we have a sequence of approximations $C_n(x,u)$ and $D_n^k(x,u)$, $n = 1, 2, \ldots$ of the costs $C(x,u)$ and $D^k(x,u)$ with $k = 1, 2, \ldots, K$. 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 is omitted from the notation in the following sections), where as $C_n(x,u)$ and $D_n^k(x,u)$, may be **arbitrary functions** from $\mathbf{X} \times U$ to $\mathbb{R}^+$. (They could be e.g. finite horizon costs, they could be defined as discounted costs with discount factor that changes in time, or costs related to other transition probabilities and immediate costs).

Consider the following sequence of problems:

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

$C_n(x) := \inf_{u \in U}\{C_n(x,u); D_n^k(x,u) \leq V_k,\ k = 1, 2, \ldots, K\}$. Let $C(x)$ be the optimal value of COP. We shall assume that $\lim_{n\to\infty} C_n(x,u) = C(x,u)$ and $\lim_{n\to\infty} D_n^k(x,u) = D^k(x,u)$ for some large set of policies and a given initial state x. We are then interested in the following three questions.

1. convergence of optimal values: When does $\lim_{n\to\infty} C_n(x) = C(x)$?

2. convergence of optimal policies: When is the limit of the policies which are optimal (or "almost" optimal) for COP_n, optimal for COP? This question is related to the problem of approximating the optimal policy for COP by a policy which is (almost) optimal for some COP_n.

3. Stability of the optimal policy When is an (almost) optimal policy for COP almost optimal for COP_n?

Note that COP_n may not have any optimal policy even if it is feasible. Moreover, there may not exist any ϵ-optimal stationary policy for COP_n. In Theorem 6.1 below, we answer the first two questions. Its proof is based on ideas from Theorem 6.1 in [4] (which deals with sensitivity of COP to the discount factor β in the finite case). In Theorem 6.2 we then use the optimal policy for COP to construct almost optimal policies for COP_n. We apply these Theorems in the following section to obtain an Algorithm for finite state approximations of COP.

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

$$C_m(x, u^m) \leq C_m(x, u), \qquad D^k_m(x, u^m) \leq D^k_m(x, u), \quad k = 1, 2, \ldots, K. \tag{19}$$

Theorem 6.1 *Assume*
(1) A2 and that $c(\cdot, \cdot)$ and $d^k(\cdot, \cdot)$ are bounded below.
(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 $u \in \Gamma$ and a given initial state x, uniformly in $u \in \Gamma$.
(3) $\beta < 1$, or $\{\beta = 1$ and A1, A3(i) hold as well as A4 $\}$.
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

$$C(x, w(n)) \leq C(x, r(n)), \qquad D^k(x, w(n)) \leq D^k(x, r(n)), \quad k = 1, 2, \ldots, K. \tag{20}$$

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

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

Then w is optimal for COP.

Proof: We begin by establishing (i). We first show that

$$\underline{\lim}_{n \to \infty} C_n(x) \geq C(x) \tag{21}$$

Assume (21) does not hold. Then there exists some $\epsilon > 0$ such that for every $N > 0$ there exists some $m > N$ and a policy r^m which is feasible for COP_m and

$$C_m(x, r^m) < C(x) - \epsilon \tag{22}$$

Let $u^m \in \Gamma$ be some feasible policy for COP, that satisfies

$$C_m(x, u^m) \leq C_m(x, r^m), \qquad D^k_m(x, u^m) \leq D^k_m(x, r^m), \quad k = 1, 2, ..., K \tag{23}$$

Theorem 4.1 and A2 imply that there exists a policy $v \in U(S)$ and some positive real number η such that $D^k(v) < V_k - \eta$, for all $1 \leq k \leq K$. Choose some $0 < \alpha < 0.5$ such that $\alpha C(x, v) < \epsilon/8$, and $\delta > 0$ that satisfies $0 < \delta < min\{\frac{\epsilon}{2}\ , \eta\alpha\}$. Fix N such that for all $u \in \Gamma$ and for all $m > N$

$$|C_m(x, u) - C(x, u)| < \delta \tag{24}$$

$$|D^k_m(x, u) - D^k(x, u)| < \delta \qquad k = 1, 2, ..., K \tag{25}$$

We then obtain

$$C(x, u^m) \leq C_m(x, u^m) + \delta \leq C_m(x, r^m) + \delta \leq C(x) + \delta - \epsilon \tag{26}$$

$$D^k(x, u^m) \leq D^k_m(x, u^m) + \delta \leq D^k_m(x, r^m) + \delta \leq V_k + \delta \tag{27}$$

From Theorem 3.1 it follows that there exists some stationary policy u' such that

$$\bar{f}^\beta_{sa}(x, u') = (1 - \alpha)\bar{f}^\beta_{sa}(x, u^m) + \alpha \bar{f}^\beta_{sa}(x, v) \tag{28}$$

It follows from (27) and (28) that u' is feasible for COP since by Lemma 3.1 or Lemma 3.2

$$D^k(x, u') = \bar{f}^\beta_{sa}(x, u') \cdot d^k \leq (1 - \alpha)(V_k + \delta) + \alpha(V_k - \eta)$$

$$= V_k + (1 - \alpha)\delta - \alpha\eta < V_k + \delta - \alpha\eta < V_k$$

From (26) and (28) we obtain by Lemma 3.1 or Lemma 3.2

$$C(x, u') = \bar{f}^\beta_{sa}(x, u') \cdot c = (1 - \alpha)C(x, u^m) + \alpha C(x, v)$$

$$< (1 - \alpha)(C(x) + \delta - \epsilon) + \epsilon/8 \leq (1 - \alpha)C(x) + (\delta - \epsilon)/2 + \epsilon/8 < (1 - \alpha)C(x) - \epsilon/8$$

Since α can be chosen arbitrarily small, it follows that $C(x, u') < C(x)$. But this contradicts the definition of $C(x)$, which proves (21).

Next we prove that

$$\overline{\lim_{n \to \infty}}\, C_n(x) \leq C(x) \tag{29}$$

Let $u^* \in U$ be a policy which satisfies

$$C(x,u) = c \cdot \bar{f}^{\beta}_{sa}(x,u), \qquad D^k(x,u) = d^k \cdot \bar{f}^{\beta}_{sa}(x,u), \quad k = 1,...,K, \tag{30}$$

and is ϵ-optimal for COP (its existence follows from Theorem 4.1 when choosing a stationary policy). (The scalar products above are defined in the end of Section 2). Choose a policy $u(\epsilon) \in U$ that satisfies the linear representation (30) and such that

$$\bar{f}^{\beta}_{sa}(x,u(\epsilon)) = (1-\epsilon)\bar{f}^{\beta}_{sa}(x,u^*) + \epsilon \bar{f}^{\beta}_{sa}(x,v) \tag{31}$$

$$\lim_{n\to\infty} D^k_n(x,u(\epsilon)) = D^k(x,u(\epsilon)), \ k = 1,...,K, \tag{32}$$

$$\lim_{n\to\infty} C_n(x,u(\epsilon)) = C(x,u(\epsilon)) \tag{33}$$

(the existence of a stationary policy satisfying these requirements follows from Theorem 3.1, and Lemma 3.1 or Lemma 3.2). It follows that

$$\overline{\lim_{n\to\infty}} D^k_n(x,u(\epsilon)) = D^k(x,u(\epsilon)) = (1-\epsilon)D^k(x,u^*) + \epsilon D^k(x,v)$$

$$\leq (1-\epsilon)V_k + \epsilon(V_k - \eta) = V_k - \epsilon\eta \tag{34}$$

for all $1 \leq k \leq K$. Therefore there exists some integer $N(\epsilon)$ such that for every ϵ and $n \geq N(\epsilon)$, $u(\epsilon)$ is feasible for COP_n. Since $u(\epsilon)$ may not be optimal for COP_n, we clearly have

$$\overline{\lim_{n\to\infty}} C_n(x) \leq \overline{\lim_{n\to\infty}} C_n(x,u(\epsilon)) = C(x,u(\epsilon)) \leq (1-\epsilon)(C(x)+\epsilon) + \epsilon C(x,v) \tag{35}$$

where the last inequality follows from (31) and the linear representation of the cost. Since this holds for every ϵ, (29) follows. This completes the proof of (i).

Next we prove (ii). From Lemma 10.3 in the appendix we have $\lim_{i\to\infty} \bar{f}^{\beta}_{sa}(x,w(i_n)) = \bar{f}^{\beta}_{sa}(x,w)$. Hence for any $k = 1,...,K$,

$$\begin{aligned} 0 &= \lim_{i\to\infty} \left[D^k_{i_n}(x,w(i_n)) - D^k(x,w(i_n)) \right] \\ &= \lim_{i\to\infty} \left[D^k_{i_n}(x,w(i_n)) - \bar{f}^{\beta}_{sa}(x,w(i_n)) \cdot d^k \right] \end{aligned}$$

$$\begin{aligned}
&\leq \varliminf_{i\to\infty} \left[D^k_{i_n}(x, w(i_n)) - \bar{f}^\beta_{sa}(x, w) \cdot d^k\right] \\
&= \varliminf_{i\to\infty} \left[D^k_{i_n}(x, w(i_n)) - D^k(x, w)\right] \\
&\leq \left[V_k - D^k(x, w)\right]
\end{aligned}$$

where the first inequality follows from Fatou's Lemma, and the second one follows from the fact that COP_n is feasible for all large enough n (this follows from (34)). Hence w is feasible for COP. From the same argument it also follows that

$$0 \leq \varliminf_{i\to\infty} \left[C^k_{i_n}(x, w(i_n)) - C(x, w)\right] = C(x) - C(x, w)$$

Since $C(x)$ is the optimal value of COP, it follows that $C(x, w) = C(x)$.

■

Remarks: (i) The role of the uniform convergence of the cost (assumption (2) of the theorem) in approximation methods appears already in previous literature on approximating models of non-constrained MDPs (see e.g. Lemma 5.4 in [17]).
(ii) Note that the convergence of the costs need not be uniform for (29) to hold.

Next we construct a policy which is almost optimal for COP_n for all n large enough.

Theorem 6.2 *Let $-M$, $M \geq 0$ be a lower bound on the immediate costs c. Under the conditions of Theorem 6.1, $u(\epsilon)$ defined above (31), is $\hat{\epsilon}$-optimal for COP_n for all n large enough, where $\hat{\epsilon} = \epsilon[C(x, v) + M + 3]$.*

Proof: For all n large enough, COP_n is feasible (this follows from (34)); moreover,

$$\begin{aligned}
C_n(x, u(\epsilon)) &\leq C(x, u(\epsilon)) + \epsilon \leq (1-\epsilon)(C(x) + \epsilon) + \epsilon C(x, v) + \epsilon && (36)\\
&\leq C_n(x) + 2\epsilon - \epsilon C(x) + \epsilon C(x, v) + \epsilon \leq C_n(x) + \epsilon[C(x, v) + M + 3] && (37)
\end{aligned}$$

where the first and second inequality in (36) follows from (35), and (37) follows from Theorem 6.1 (i). ■

7 Approximating the dynamics, finite approximations

In this section we consider the problem of approximating the Controlled Markov Chain (CMC) which is characterized by the dynamics (i.e. the transition probabilities) $\mathcal{P}_{xay}$ by a sequence of Controlled Markov Chains CMC_n governed by the dynamics $\{\mathcal{P}_{xay}(n)\}$, $n = 1, 2, ...$ We denote by $C_n(x, u)$ and $D_n(x, u)$ the costs (given in (3)) under policy u and initial state x corresponding to CMC_n and discount factor β, $0 < \beta \leq 1$ (β is omitted from the notation). We shall construct the CMC_n such that for all $x, y \in \mathbf{X}$, $a \in \mathbf{A}$ $\lim_{n\to\infty} \mathcal{P}_{xay}(n) = \mathcal{P}_{xay}$. We then show that the construction ensures that $\lim_{n\to\infty} C_n^\beta(x, u) = C^\beta(x, u)$ and $\lim_{n\to\infty} D_n^{k,\beta}(x, u) = D^{k,\beta}(x, u)$, $k = 1, ..., K$ uniformly in $u \in U(S)$ which will be shown to imply the conditions of Theorem 6.1 and hence some convergence properties of the optimal values and policies. We shall often use the notation CMC_∞ for CMC.

Introduce the following approximation scheme **FA**:

(i) For each $n = 0, ...$ the state space is decomposed in two disjoint classes of states: E^n, which contains a finite number of states, and T^n. The set of actions $\mathbf{A}(y)$ do not depend on y.

(ii) Under any stationary policy u, E^n is a recurrent class, T^n is a transient class, and absorption into the positive recurrent class takes place in finite expected time from any initial state.

(iii) $E_n \subset E_{n+1}$, $n = 1, ...$; $E_0 := \{\emptyset\}$; $E_\infty = \mathbf{X}$.

(iv) The following holds:

$$\mathcal{P}_{xay}(n) \begin{cases} \geq \mathcal{P}_{xay} & x, y \in E_n \\ = 0 & x \in E_n,\ y \notin E_n \\ = 1\{y = 1\} & x \notin E_n \end{cases} \tag{38}$$

where state “1” is an arbitrary state in E_1.

(v) There is some partial order on $\mathbf{X}$. For any x, $P_{xa\bullet}(n)$ are stochastically nondecreasing in n, $n = 1, 2, ..., \infty$. This is equivalent to the following (see e.g. [26] p. 256): for any function $h : \mathbf{X} \to \mathbb{R}$ which is nondecreasing w.r.t. the partial order, we have for $1 \leq n \leq m \leq \infty$:

$$\sum_{y\in\mathbf{X}} \mathcal{P}_{xay}(n)h(y) \leq \sum_{y\in\mathbf{X}} \mathcal{P}_{xay}(m)h(y) \tag{39}$$

(vi) For any n, $P_{xa\bullet}(n)$ are stochastically nondecreasing in x, $n = 1, 2, ..., \infty$. This is equivalent to the following (see e.g. [26] p. 256): for any function $h : \mathbf{X} \to \mathbb{R}$ which is nondecreasing w.r.t. the partial order, we have for $x \le z$ (w.r.t. the partial order):

$$\sum_{y \in \mathbf{X}} \mathcal{P}_{xay}(n) h(y) \le \sum_{y \in \mathbf{X}} \mathcal{P}_{zay}(n) h(y) \tag{40}$$

There are two general applications of finite approximations, where **FA** is used:

First application: the objective is to construct an approximating CMC_n. In that case **FA**(v) can be achieved by choosing

$$\mathcal{P}_{xay}(n) \begin{cases} = \mathcal{P}_{xay} & x \in E_n,\ y \in E_{n-1} \\ \ge \mathcal{P}_{xay} & x \in E_n,\ y \in E_n \backslash E_{n-1} \\ = 0 & x \in E_n,\ y \notin E_n \\ = 1\{y = 1\} & x \notin E_n \end{cases} \tag{41}$$

where state "1" is an arbitrary state in E_1. **FA** in general, and (41) in particular, imply that a stationary optimal policy can be found for the approximating dynamics using a finite LP [19], where the state space is composed of $\{x\} \cup E_n$.

Second Application: the objective is to use the original CMC in order to approximate CMC_n for n large enough. This is useful in control of queueing networks, where often the control problem with infinite buffers is much easier than the control problems with finite buffers. The state space in problems with infinite buffers is typically multi-dimensional with elements $\mathbf{x} = (x_1, ..., x_N) \in \mathbb{N}^{\mathrm{N}}$ representing the vector of queues' length, where as for the case of finite buffers of sizes $\mathbf{R} = (R_1, ..., R_N)$ the state space is $\mathbf{X}(\mathbf{R}) = \prod_{j=1}^{N} \{0, 1, ..., R_j\}$. The partial order could then be the componentwise order, i.e. $\mathbf{x} \ge \mathbf{y}$ iff $x_i \ge y_i$ for all $i = 1, ..., N$. **FA** is then typically satisfied, and $\mathcal{P}_{xay}(n)$ has the form of (41). In that case, (v) is typically satisfied since the finiteness of the buffers inhibits in some states to go to larger states, where as transitions to lower states stay unchanged. (vi) is also quite natural in queueing systems, and its intuitive meaning is that if in system 1 there are at least as many customers than in system 2, which is identical, then after one transition we still have at least as many customers in system 1 as in 2.

For $u \in U(S)$, denote $P_{xy}^u(n) := \sum_a \mathcal{P}_{xay}(n) p_{a|x}^u$ and let $\bar{f}^\beta(n; x, u)$ denote the corresponding occupation measure.

Theorem 7.1 *Consider a sequence of finite approximations COP_n obtained by applying* **FA**. *Assume A2, and { A1, A3(i) and A4 } or $\beta < 1$. Assume moreover that*

(i) both $c(y) = c(y,a)$ and $d^k(y) = d^k(y,a)$, $k = 1,...,K$ are only functions of the state, are nondecreasing in **X** *and are bounded below.*

(ii) $C(x,u)$ and $D^k(x,u)$, $k = 1,...,K$ are continuous in $u \in U(S)$. Then

(1) $\lim_{n\to\infty} C_n(x,u) = C(x,u)$ and $\lim_{n\to\infty} D^k_n(x,u) = D^k(x,u)$, $k = 1,...,K$, uniformly in $u \in U(S)$; $\lim_{n\to\infty} C_n(x) = C(x)$.

(2) Choose some $\epsilon > 0$. Let u^* be an ϵ-optimal (or optimal) policy for COP that satisfies (30) (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 (30), (31), (32) and (33) is $\hat{\epsilon}$-optimal for COP_n, where $\hat{\epsilon}$ is given in Theorem 6.2.

(3) Let $w(n)$ be a stationary 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 6.1 (ii) to the policies $w(n)$ is optimal for COP.

Proof: We show that

$$\lim_{n\to\infty} C_n(x,u) = C(x,u)$$

uniformly in $U(S)$. Note that it follows from Theorem 4.1 (i) and the ergodic structure obtained by using **FA** that $\Gamma = U(S)$ indeed satisfies (19).

Let $h : \mathbf{X} \to \mathbb{R}$ be a nondecreasing function. (39) and (40) imply that for any $u \in U(S)$ and $N = 2$,

$$\begin{aligned}
\sum_{y\in\mathbf{X}} [P^u(n)]^N_{xy} h(y) &= \sum_{y\in\mathbf{X}} [P^u(n)]_{xy} ([P^u(n)]^{N-1} h)_y \\
&\leq \sum_{y\in\mathbf{X}} [P^u(m)]_{xy} ([P^u(n)]^{N-1} h)_y \qquad (42) \\
&\leq \sum_{y\in\mathbf{X}} [P^u(m)]^N_{xy} h(y), \qquad 1 \leq n \leq m \leq \infty, \forall x \in \mathbf{X} \qquad (43)
\end{aligned}$$

$$\sum_{y\in\mathbf{X}} [P^u(n)]^N_{xy} h(y) = \sum_{y\in\mathbf{X}} [P^u(n)]_{xy} ([P^u(n)]^{N-1} h)_y$$

$$\leq \sum_{y \in \mathbf{X}} [P^u(n)]_{zy} ([P^u(n)]^{N-1} h)_y \tag{44}$$

$$= \sum_{y \in \mathbf{X}} [P^u(n)]^N_{zy} h(y), \qquad x \leq z, \forall n \in \mathbb{N}. \tag{45}$$

(42) and (44), with $N = 2$, follow since, by (45), $([P^u(n)]^{N-1}h)_y$ is nondecreasing in y. (43) follows since $([P^u(n)]^{N-1}h)_y \leq ([P^u(m)]^{N-1}h)_y$. The inequalities above easily extends by induction to any $N \in \mathbb{N}$.

It follows that for any $0 < \beta \leq 1$, $\bar{f}^\beta_s(n; x, u)$ is stochastically nondecreasing in n, and so $\bar{f}^\beta_{sa}(n; x, u) \cdot c$ are nondecreasing in n. In particular,

$$\lim_{n \to \infty} \bar{f}^\beta_{sa}(n; x, u) \cdot c \leq \bar{f}^\beta_{sa}(x, u) \cdot c \tag{46}$$

and similarly with d^k, $k = 1, ..., K$. We shall show that in fact these inequalities are achieved with strict equality. Let $\mathcal{K} \in \mathbf{X}$ be a finite set that satisfies: if $y \in \mathcal{K}$ and $z \leq y$ then $z \in \mathcal{K}$. Let $h : \mathbf{X} \to \mathbb{R}$ be given by $h(y) = 1\{y \notin \mathcal{K}\}$. It follows that $\bar{f}^\beta_s(n; x, u) \cdot h \leq \bar{f}^\beta_s(x, u) \cdot h$ and hence $\{\bar{f}^\beta_s(n; x, u)\}_n$ as well as $\{\bar{f}^\beta_{sa}(n; x, u)\}_n$ are tight. Therefore, by arguments similar to those in Lemma 10.3, $\lim_{n \to \infty} \bar{f}^\beta_{sa}(n; x, u) = \bar{f}^\beta_{sa}(x, u)$. By Fatou's Lemma we thus obtain

$$\lim_{n \to \infty} \bar{f}^\beta_{sa}(n; x, u) \cdot c \geq \bar{f}^\beta_{sa}(x, u) \cdot c \tag{47}$$

and combining that with (46) yields

$$\lim_{n \to \infty} C_n(x, u) = \lim_{n \to \infty} \bar{f}^\beta_{sa}(n; x, u) \cdot c = \bar{f}^\beta_{sa}(x, u) \cdot c = C(x, u)$$

and similarly with $D^k(x, u)$, $k = 1, ..., K$. Next we show that this convergence is uniform in $U(S)$. $c \cdot \bar{f}^\beta_{sa}(n; x, u)$ is continuous in $u \in U(S)$ (for $n = \infty$ this follows from assumption (ii)). Since $L^\beta_x(S)$ is compact (Theorem 3.1), and since $\bar{f}^\beta_{sa}(n; x, u) \cdot c$ is monotone in n, the uniform convergence follows (see [29] p. 150 Thm. 7.13). The theorem now follows from Theorems 6.1 and 6.2. ∎

Remarks: (1) The assumption that the immediate cost are nondecreasing is quite natural in queueing systems whenever the cost represents quantities as delays, holding costs or blocking probabilities.

(2) A sufficient condition for Assumption (ii) in the theorem is the uniformly integrability of the immediate costs c and d^k w.r.t. $\bar{f}^{\beta}_{sa}(x,u)$, $u \in U(S)$. Indeed, it follows from Lemma 10.3 in the appendix that for every $1 \leq n \leq \infty$, $\bar{f}^{\beta}_{sa}(n;x,u)$ is continuous in $u \in U(S)$. This implies by the uniform integrability, that $c \cdot \bar{f}^{\beta}_{sa}(n;x,u)$ is also continuous in $u \in U(S)$. Other sufficient conditions for this continuity (uniform μ-geometric ergodicity or uniform μ-geometric recurrence) can be found in [31] p. 97-98.

(3) From Theorem 7.1 (3) we deduce that if the conditions of the Theorem 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).

8 Second approximation scheme

We present below a general scheme for finite approximations for constrained problems, based on arbitrarily picking the probabilities of choosing the different actions in all but a finite number N of states. Then the probabilities in the remaining N states are chosen so as to optimize the constrained problem. We present conditions under which this scheme indeed approximates the optimal policy for the original constrained problem.

Without loss of generality, we shall assume that the state space is given by $\mathbf{X} = \{0,1,2,....\}$. Let w be any stationary policy. Let $U(n)$ be the set of stationary policies that behave exactly like w in states $y \geq n$, i.e. $p^u_{a|y} = p^w_{a|y}$ for all $y \geq n$, $a \in \mathbf{A}$ and $u \in U(n)$. Denote by COP_n the restriction of COP to $U(n)$.

Theorem 8.1 *Assume A2 and that both $c(\cdot,\cdot)$ and $d^k(\cdot,\cdot)$, $k = 1,...,K$ are bounded below. Further assume that $C(x,u)$ and $D^k(x,u)$, $k = 1,...,K$ are continuous in $u \in U(S)$ (see Remark (2) in the previous Section). If either (a) $\beta < 1$ or (b) A1, A3(ii), A4 hold, then $\lim_{n\to\infty} C_n(x) = C(x)$.*

Proof: Clearly $\underline{\lim}_{n\to\infty} C_n(x) \geq C(x)$. We prove below that $\overline{\lim}_{n\to\infty} C_n(x) \leq C(x)$ which establishes the proof. Let u^* be any policy which is optimal for COP. Theorem 4.1 and Assumption A2 imply that there exists a stationary policy v and some positive real number η such that $D^k(v) < V_k - \eta$, for all $1 \le k \le K$. Choose a stationary policy $u(\epsilon)$ such that

$$\bar{f}^{\beta}_{sa}(x, u(\epsilon)) = (1-\epsilon)\bar{f}^{\beta}_{sa}(x, u^*) + \epsilon \bar{f}^{\beta}_{sa}(x, v) \tag{48}$$

It follows from (48) and the linear representation of the cost that

$$C(x, u(\epsilon)) = (1-\epsilon)C(x) + \epsilon C(x, v)$$

and

$$D^k(x, u(\epsilon)) \le (1-\epsilon)V_k + \epsilon(V_k - \eta) = V_k - \epsilon\eta, \qquad 1 \le k \le K. \tag{49}$$

Let $u(\epsilon, n) \in U(n)$ be the stationary policy given by $p^{u(\epsilon,n)}_{a|y} = p^{u(\epsilon)}_{a|y}$, $y < n$, $a \in \mathbf{A}$. Since $u(\epsilon, n) \to u(\epsilon)$, it follows from the continuity of $C(x, u)$ and $D^k(x, u)$, $k = 1, ..., K$ in $u \in U(S)$ that

$$\lim_{n\to\infty} C(x, u(\epsilon, n)) = C(x, u(\epsilon))$$

and

$$\lim_{n\to\infty} D^k(x, u(\epsilon, n)) = D^k(x, u(\epsilon)), \quad k = 1, ..., K$$

Choose some $0 < \delta < \epsilon\eta$. We see from (49) that there exists an integer $N(\epsilon, \delta)$ such that for any $n > N(\epsilon, \delta)$,

$$C(x, u(\epsilon, n)) < C(x, u(\epsilon)) + \delta$$

$$D^k(x, u(\epsilon, n)) < V_k + \delta - \epsilon\eta < V_k, \quad k = 1, ..., K \tag{50}$$

and hence $u(\epsilon, n)$ is feasible for COP. But then, since $u(\epsilon, n) \in U(n)$, we have

$$C_n(x) \le C(x, u(\epsilon, n)) < C(x, u(\epsilon)) + \delta = (1-\epsilon)C(x) + \epsilon C(x, v) + \delta. \tag{51}$$

Since ϵ and δ can be chosen arbitrarily small, we obtain $\overline{\lim}_{n\to\infty} C_n(x) \le C(x)$, which establishes the proof. ∎

Theorem 8.2 *Assume that the conditions of Theorem 8.1 hold. Then for any n, there exists a stationary optimal policy among $U(n)$ (defined above Theorem 8.1) for COP_n. Let $w(n)$*

be a stationary optimal policy for COP_n if COP_n is feasible, otherwise let it be an arbitrary stationary policy. Let w be an arbitrary accumulation point of $w(n)$ (see Theorem 6.1 (ii)). Then w is optimal for COP.

Proof: It follows from (50) that for all large enough n, COP_n is feasible. $w(n)$ are thus feasible policies for both COP_n and COP for n large. It follows from the continuity of $C(x,u)$ and $D^k(x,u)$, $k = 1,...,K$ in $u \in U(S)$, and from Theorem 8.1, that $C(x) = \lim_{n\to\infty} C_n(x) = \lim_{n\to\infty} C(x, w(n)) = C(x, w)$, which establishes the Proof. ■

9 Application to a queueing model

Consider the following discrete time system ([1], [2] Section 6, [23], [28], [31]). 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 large 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 noise 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).

All previous research on this constrained model assumed infinite buffers, for which optimal policies with a simple structure exist ([1], [2] Section 6, [23], [28], [31]). There is however no known solution for this constrained (or even unconstrained) problem in case that the buffers are finite. In that case, an arriving packet that finds its buffer full is lost. Our finite approximations techniques developed in previous sections enable to obtain almost optimal policies for the case that the finite buffers are large enough.

We begin by considering the model with infinite buffers. 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 $(\lambda_1, ..., \lambda_N)$. 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. 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.

We assume the standard stability condition on the traffic intensity $\rho := \sum_{i=1}^N \lambda_i/\mu_i < 1$. 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 buffer. Let $|\mathbf{G}| = L$. For the unconstrained control problem, there exists an optimal policy within $\mathbf{G}$; it is the so called "μc rule", for which the priorities are set according to increasing order of the $\mu_i c_i$ (see [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. (The exact definition can be found in [1]). It is shown in [1] that

$$\bar{f}^1_{sa}(x, \hat{\alpha}) = \sum_{j=1}^{L} \alpha_j \bar{f}^1_{sa}(x, g_j)$$

and

$$C_1(x,\hat{\alpha}) = c \cdot \bar{f}^1_{sa}(x,\hat{\alpha}) = \sum_{j=1}^{L} \alpha_j c \cdot \bar{f}^1_{sa}(x,g_j) = \sum_{j=1}^{L} \alpha_j C_1(x,g_j). \tag{52}$$

For a given $\delta > 0$, consider the following **LP**: find $\alpha \in \mathbb{R}^L$ that

$$\text{minimize } \sum_{j=1}^{L} \alpha_j C_1(x,g_j)$$

$$\text{subject to } \sum_{j=1}^{L} \alpha_j D_1^k(x,g_j) \le V_k - \delta, \qquad k = 1,...,K$$

$$\sum_{j=1}^{L} \alpha_j = 1, \qquad \alpha_i \ge 0, \ i = 1,...,L$$

The quantities $C_1(x,g_j)$ and $D_1^k(x,g_j)$ can be obtained as in [23]. Let $\alpha^*(\delta)$ be the solution of LP with a given δ. Then $\hat{\alpha}^*(0)$ is an optimal policy for COP_{queues}. Note that $\alpha^*(0)$ may be chosen so as to have at most $K+1$ non-zero elements, which means that at most $K+1$ different g_j's need to be multiplexed under the optimal policy $\hat{\alpha}^*(0)$. Under A2, 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 [1]).

In the following Theorem we consider a sequence of problems COP^n_{queues} for the systems with buffers of sizes $\mathbf{R}^n = (R^n_1,...,R^n_N)$, $1 \le n \le \infty$, where COP^∞_{queues} is the one with all buffers infinite, and $\mathbf{R}^n < \mathbf{R}^{n+1}$ (the strict inequality holds for at least one component). Assume without loss of generality that the initial state x satisfies $x \le \mathbf{R}^1$.

Theorem 9.1 *Assume A2 and* $\beta = 1$. *Then*
(i) $\lim_{n\to\infty} C_n(x) = C(x)$.
(ii) Choose some $0 < \epsilon < 1$. *Let* $u(\epsilon)$ *be the PTS policy with*

$$\alpha = \epsilon\alpha^*(\delta^*) + (1-\epsilon)\alpha^*(0).$$

Then for all n *large enough,* $u(\epsilon)$ *is* $\hat{\epsilon}$*-optimal for* COP^n_{queues}, *where* $\hat{\epsilon} = \epsilon[C(x,\hat{\alpha}^*(\delta^*)) + 3]$.
(iii) In case that there are no constraints, for every $\epsilon > 0$ *there exists some* $N(\epsilon)$ *such that the "*μc *rule" is* ϵ*-optimal for* COP^n_{queues} *for all* $n \ge N(\epsilon)$.

Proof: It is shown in [2] p. 804 that A1 and A3(i) hold. It follows from [21] Corollary 5.1.1 that $c(x)$ and $d^k(x)$, $k = 1, ..., K$ are uniformly integrable w.r.t. $\bar{f}_s^1(x, u)$, $u \in U(S)$. Hence $C(x, u)$ and $D^k(x, u)$, $k = 1, ..., K$ are continuous in $u \in U(S)$. (An alternative proof of this continuity is obtained by combining Thm. 9.1 p. 143 in [31] and the first line in [31] p. 98). The conditions of **FA** are easily seen to hold for this problem. (52) implies that (30), (31), (32) and (33) are satisfied under $u(\epsilon)$. The proof of (i) and (ii) then follows from Theorem 7.1. Note that A4 is satisfied since there are no transient states under any $u \in U(S)$ (see [21]). In the absence of constraints, $\alpha^*(\delta^*) = \alpha^*(0)$ is the "μc" rule and thus (ii) implies (iii). ∎

Remark: It follows from Theorem 4.1 that the stationary policies are optimal for COP_{queues} with infinite buffers also for the case that $\beta < 1$. Since the "μc rule is known to be optimal for the discounted unconstrained problem with infinite buffers, it seems that by appropriately multiplexing between policies in **G**, using randomization, one can obtain an optimal policy for COP_{queues} for $\beta < 1$ using a similar LP as above. An ϵ-optimal policy can then be obtained for the case of finite buffers and $\beta < 1$, as in Theorem 9.1.

10 Appendix

We present below three Lemmas. The proof of the first two is given in [6].

Lemma 10.1 *Let P be a stochastic matrix on $\mathbf{X} \times \mathbf{X}$. Then $(I - \beta P)^{-1} = \sum_{j=0}^{\infty} \beta^j P^j$, i.e.*

$$(\sum_{j=0}^{\infty} \beta^j P^j)(I - \beta P) = I = (I - \beta P) \sum_{j=0}^{\infty} \beta^j P^j \tag{53}$$

where $P^0 \stackrel{\text{def}}{=} I$ is the identity matrix.

Lemma 10.2 *Let P_n, $n = 1, ...$ and P be stochastic matrices on $\mathbf{X} \times \mathbf{X}$ such that $\lim_{n \to \infty} P_n = P$ (the convergence is componentwise or equivalently in the product topology $\tau(\mathbf{X})^{\mathbf{X}}$). Then*

$$\lim_{n \to \infty} (1 - \beta) \sum_{j=0}^{\infty} \beta^j [P_n]^j = (1 - \beta) \sum_{j=0}^{\infty} \beta^j [P]^j, \qquad 0 < \beta < 1.$$

Lemma 10.3 *Assume either A1 and A3(ii), or $\beta < 1$. Let $w(n)$ and w be stationary policies such that $\lim_{n\to\infty} w(n) = w$. Then*

$$\lim_{n\to\infty} \bar{f}^1_{sa}(x, w(n)) = \bar{f}^1_{sa}(x, w). \tag{54}$$

Proof: It easily follows that the transition probabilities converge pointwise: $\lim_{n\to\infty} P^{w(n)} = P^w$. For $\beta < 1$, (54) then follows from Lemma 10.2. For $\beta = 1$, this follows from the continuity of π^u in $u \in U(S)$, which holds by A3(ii) see [18] p. 82. ∎

References

[1] E. Altman and A. Shwartz, "Optimal priority assignment: a time sharing approach", *IEEE Trans. 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: Criteria and Policies", *Annals of Oper. Res.*, **28**, special issue on Markov Decision Processes, Eds. O. Hernandez-Lerma and J. B. Lasserre, pp. 101-134, 1991.

[4] E. Altman and A. Shwartz, "Sensitivity of constrained Markov Decision Problems", *Annals of Oper. Res.*, **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 Trans. Auto. Control.*

[6] E. Altman, "Denumerable constrained Markov Decision Problems and finite approximations", technical report No. 1568, INRIA-Sophia, France , 1991.

[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", *Math. Anal. Appl.* **112**, 236-252, 1985.

[9] V. S. Borkar, "A convex analytic approach to Markov decision processes", *Probab. Th. Rel. Fields*, Vol. 78, pp. 583-602, 1988.

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

[11] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York, 1968.

[12] R. Cavazos-Cadena, "Finite-state approximations for denumerable state discounted Markov Decision Processes", *J. Appl. Math. Optim.* **14** pp. 27-47, 1986.

[13] C. Derman, *Finite State Markovian Decision Processes,* Academic Press, 1970.

[14] V. A. Gaitsgory and A. A. Pervozvanskii, "Perturbation Theory for Mathematical Programming Problems", *JOTA*, 389-410, 1986.

[15] A. Hordijk and F. Spieksma, "Constrained admission control to a queueing system" *Adv. Appl. Prob.* Vol. 21, pp. 409-431, 1989.

[16] O. Hernandez-Lerma, "Finite state approximations for denumerable multidimensional - state discounted Markov decision processes", *J. Math. Anal. Appl.*, **113** pp. 382-389, 1986.

[17] O. Hernandez-Lerma, *Adaptive Control of Markov Processes*, Springer Verlag, 1989.

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

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

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

[21] A. M. Makowski and A. Shwartz, "Recurrence properties of a system of competing queues, with applications", EE Pub. 627, June 1987.

[22] A. Lazar, "Optimal flow control of a class of queueing networks in equilibrium", *IEEE Trans. Auto. Cont.*, Vol 28 no. 11, pp. 1001-1007, 1983.

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

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

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

[26] S. Ross, *Stochastic Processes*, John Wiley & Sons, New York, 1983.

[27] S. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, 1970.

[28] K. W. Ross and B. Chen, "Optimal scheduling of interactive and non interactive traffic in telecommunication systems", *IEEE Trans. on Auto. Control,* Vol. 33 No. 3 pp. 261-267, 1988.

[29] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1985.

[30] Schäl M., "Conditions for optimality in dynamic programming and for the limit of the n-stage optimal policies to be optimal", *Z. Wahrscheinlichkeitsth.*, Vol. 32, pp. 179-196, 1975.

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

[32] 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.

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