Probability in the Engineering and Informational Sciences, **15**, 2001, 103–133. Printed in the U.S.A.

COMPUTING AVERAGE OPTIMAL CONSTRAINED POLICIES IN STOCHASTIC DYNAMIC PROGRAMMING

LINN I. SENNOTT
Department of Mathematics
Illinois State University
Normal, Illinois 61790-4520
E-mail: sennott@math.ilstu.edu

A stochastic dynamic program incurs two types of cost: a service cost and a quality of service (delay) cost. The objective is to minimize the expected average service cost, subject to a constraint on the average quality of service cost. When the state space S is finite, we show how to compute an optimal policy for the general constrained problem under weak conditions. The development uses a Lagrange multiplier approach and value iteration. When S is denumerably infinite, we give a method for computation of an optimal policy, using a sequence of approximating finite state problems. The method is illustrated with two computational examples.

1. INTRODUCTION

Consider a *stochastic dynamic program* (also known as a *Markov decision chain*). If the system is in state i and action a is chosen, then costs $C(i,a)$ and $D(i,a)$ are incurred. It is desired to minimize the expected average C cost subject to a constraint on the expected average D cost.

To illustrate the theory, optimal policies (rules for system operation) will be computed in two examples. Consider a discrete-time single-server queue, having Bernoulli arrivals and geometrically distributed service times, and for which the action is to choose the probability that a customer completes service in a time slot (i.e., the rate of the geometric distribution). In both examples, the expected average

© 2001 Cambridge University Press 0269-9648/01 $12.50

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

cost of service is to be minimized. In the first example, this is subject to a constraint on the expected number of customers in the system; in the second example, it is subject to a quality of service constraint on the proportion of customers who wait at least L slots for service, where L is a fixed positive integer.

Let us now define the problem. Time is discrete (slotted). The state space S is countable, and associated with each $i \in S$ is a finite (nonempty) set A_i of allowable actions. If the process is in state i at the beginning of a slot and action $a \in A_i$ is chosen, then nonnegative costs $C(i,a)$ and $D(i,a)$ are incurred, and the state at the beginning of the next slot is j with probability $P_{ij}(a)$, where $\sum_{j\in S} P_{ij}(a) = 1$. This specifies the stochastic dynamic program Δ.

A *policy* is a rule for system operation and it may be history dependent and randomized. For the precise definition of a policy, see, for example, Puterman [16] or Sennott [20]. We are particularly interested in *stationary policies* (those that make the same decision each time the process finds itself in a given state). A stationary policy may be identified with a function f such that $f(i) \in A_i$ denotes the unique action chosen in state i. Observe that operating under a stationary policy induces a Markov chain (MC) with costs on S.

For an initial state i and policy θ, define the long-run expected average C cost (average C cost, for short) (respectively, average D cost) as

$$G_\theta(i) = \limsup_{n\to\infty} \frac{1}{n} E_\theta\left[\sum_{t=0}^{n-1} C(X_t, A_t) | X_0 = i\right], \tag{1.1}$$

$$K_\theta(i) = \limsup_{n\to\infty} \frac{1}{n} E_\theta\left[\sum_{t=0}^{n-1} D(X_t, A_t) | X_0 = i\right]. \tag{1.2}$$

Here, X_t is the state of the system at the beginning of slot t, and A_t is the action chosen: both are random quantities. Equations (1.1) and (1.2) involve calculating the expected average cost incurred over n slots, and then taking the limit supremum (in case the limit fails to exist) of the finite horizon averages to obtain the worst-case expected average cost incurred per slot far into the future.

Now, fix a positive constant K, the *constraint.* A policy θ is *feasible* if $K_\theta(i) \le K$, for $i \in S$. The *constrained optimization problem* (COP) is to determine a feasible policy θ satisfying

$$G_\theta(i) = \inf_{\psi \text{ feasible}} G_\psi(i) < \infty, \quad i \in S. \tag{1.3}$$

If there exists a feasible policy, then COP is said to be feasible. A feasible policy realizing the minimum in (1.3) is said to *solve* COP or to be *constrained optimal.*

Sections 2 and 3 present a theory under weak assumptions, when the state space S is finite. Section 4, the centerpiece of the article, gives a method for computing a constrained optimal policy when S is denumerable. Our approach utilizes the approximating sequence method, developed in Sennott [20], for the computation of optimal policies in the unconstrained case. If Δ denotes the original stochastic dynamic program with denumerable state space, then an *approximating sequence* (Δ_N)

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

of finite-state stochastic dynamic programs is defined, such that the constrained problem COP_N in Δ_N can be solved. [The reader may look briefly at the examples in Sections 6 and 7 to get an idea of how (Δ_N) may be defined.] Section 4 presents a set of assumptions under which the resulting sequence of policies converges to a solution of COP. The approach uses a Lagrange multiplier. Section 5 develops sufficient conditions for the assumptions to hold.

To set up the Lagrange multiplier notation, let $\lambda \in [0,\infty)$ and define a new cost structure

$$F(i,a,\lambda) =: C(i,a) + \lambda D(i,a), \quad i \in S, a \in A_i. \tag{1.4}$$

The average cost from (1.4) under a policy θ is

$$\begin{aligned} J_\theta(i,\lambda) &= \limsup_{n\to\infty} \frac{1}{n} E_\theta\left[\sum_{t=0}^{n-1} F(X_t, A_t, \lambda) | X_0 = i\right] \\ &\le G_\theta(i) + \lambda K_\theta(i), \quad i \in S. \end{aligned} \tag{1.5}$$

Define

$$J(i,\lambda) = \inf_\theta J_\theta(i,\lambda), \quad i \in S. \tag{1.6}$$

DEFINITION 1.1: *The policy θ is* λ-optimal *if $J_\theta(i,\lambda) = G_\theta(i) + \lambda K_\theta(i) = J(i,\lambda)$, for $i \in S$.*

Note that for a policy to be optimal, it must realize the infimum in (1.6) and have an equality in the second line of (1.5). The following well-known result (e.g., Beutler and Ross [4]) gives a strategy for solving COP.

PROPOSITION 1.2: *If there exist $\lambda > 0$ and a λ-optimal policy θ such that $G_\theta(i) < \infty$ and $K_\theta(i) = K$, for $i \in S$, then θ solves COP.*

There is a wealth of literature on COP. Derman [6] has an early treatment utilizing linear programming. This approach is further developed in Kallenberg [9], Hordijk and Kallenberg [8], Spieksma [22], Puterman [16], and others. Ross [17] and Beutler and Ross [4] present a Lagrange multiplier approach, which is the grandparent of our approach. The treatment in [4] assumes the existence of a distinguished state 0 such that every stationary policy induces a unichain MC with 0 positive recurrent. This is stronger than our assumptions. In Section 2, we show how far the theory can be developed under the weak assumption that $J(i,\lambda)$ is independent of i. In Section 3, we impose the stronger condition that every λ-optimal stationary policy induces a unichain MC. Some results similar to those in Section 2 are obtained in Makowski and Shwartz [14] under a different framework and with different assumptions.

Sennott [19] extends these ideas to the case of S denumerably infinite. Koole [10] and Ross [18] show that, under certain conditions, there exists a solution to COP for a finite state space that involves randomization in at most one state. An

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

advanced treatment is in Piunovskiy [15]. The work of Altman [1] is a recent treatment; it is assumed for S finite, that every stationary policy induces a unichain MC.

2. COP FOR FINITE STATE SPACES

Throughout this section, it is assumed that the state space is finite. Our goal is to develop the theory under a weak assumption and then add additional assumptions as necessary. For now, we operate under the following assumption. Here, F stands for *finite* (state space).

ASSUMPTION (F): *For $\lambda \in (0,\infty)$, the quantity $J(i,\lambda) \equiv J(\lambda)$, for $i \in S$.*

For a fixed $\lambda > 0$, the Lagrange function is assumed constant. This assumption is in the spirit of [20, Sect. 6.4], where the theory of unconstrained average cost optimization is developed under the assumption that the average cost is constant.

The next result deals with the function $J(\lambda)$. Recall that a *concave function* is one that lies above the chord joining any two points on the graph. A function is concave if and only if its negative is convex.

THEOREM 2.1: *Assume that S is finite and that Assumption (F) holds. Then, the following is true*:

(i) *$J(\lambda)_{\lambda\in[0,\infty)}$ is an increasing, continuous, concave function consisting of finitely many linear segments, each with a nonnegative intercept. (For a different view of this matter, see Figure 1.)*

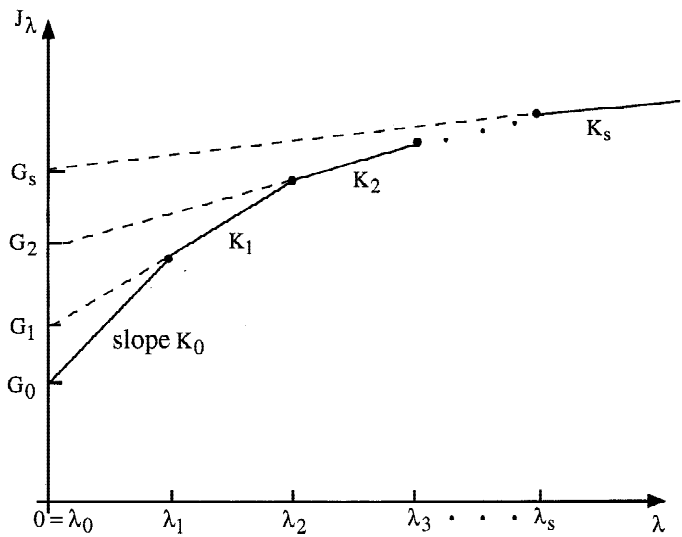


FIGURE 1. Graph of J_λ.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

(ii) *The unique break points* $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_s$ *define the* $s+1$ *intervals* $[\lambda_0, \lambda_1], \ldots, [\lambda_{s-1}, \lambda_s], [\lambda_s, \infty)$, *on each of which* $J(\lambda)$ *is linear. For compact notation, we may refer to the infinite interval as* $[\lambda_s, \lambda_{s+1}]$.

(iii) *Let* $J(\lambda) = G_j + \lambda K_j$, *for* $\lambda \in [\lambda_j, \lambda_{j+1}]$. *Then,* $K_0 > K_1 > \cdots > K_s$ *and* $G_0 < G_1 < \cdots < G_s$.

(iv) *Assume that the stationary policy e is* λ^**-optimal, for some* $\lambda^* \in (\lambda_j, \lambda_{j+1})$. *Then,* $G_e(i) = G_j$ *and* $K_e(i) = K_j$, *for all* $i \in S$, *and hence e is optimal for* $\lambda \in [\lambda_j, \lambda_{j+1}]$. *In this case, we say that e is* essential *on* $[\lambda_j, \lambda_{j+1}]$.

(v) G_0 *is the minimum average C cost. If e is essential on* $[\lambda_0, \lambda_1]$, *then e is average C cost optimal.*

(vi) K_s *is the minimum average D cost. If e is essential on* $[\lambda_s, \infty)$, *then e is average D cost optimal.*

PROOF: Let E be the set of stationary policies. Since the state space and action sets are finite, it follows that E is a finite set. Recall that $e \in E$ induces a MC on S having at least one positive recurrent class. The quantities (1.1) and (1.2) are obtained as limits so that

$$J_e(i, \lambda) = G_e(i) + \lambda K_e(i), \quad i \in S, \lambda \geq 0. \tag{2.1}$$

It is well known (e.g., Derman [6], Puterman [16], or Sennott [20]) that there exists a λ-optimal stationary policy, $\lambda \geq 0$. Hence, it follows that

$$J(i, \lambda) = \min_{e \in E}\{J_e(i, \lambda)\} = \min_{e \in E}\{G_e(i) + \lambda K_e(i)\}, \quad i \in S, \lambda \geq 0. \tag{2.2}$$

Fix $i \in S$. It follows from (2.2) that the function $J(i, \lambda)$, $\lambda \geq 0$, is the minimum of finitely many increasing linear (in λ) functions, each with a nonnegative intercept. This gives the properties in (i) and (ii) for $J(i, \lambda)$, with i fixed. Under Assumption F, all these functions must coincide, for $\lambda > 0$. Hence, they must coincide at $\lambda = 0$. This yields (i) and (ii).

It follows from Webster [23, Cor. 5.1.4] or Denardo [5, p. 201] that the right and left derivatives of $J(\lambda)$ are both decreasing on $(0, \infty)$. However, outside the finite set of break points, $J(\lambda)$ is differentiable, and its derivative is the slope of the line segment. This implies that $K_0 \geq K_1 \geq \cdots \geq K_s$. It is clearly the case that all the slopes are distinct, since if this fails, then the break points are not as in (ii). It then follows from this fact and the fact that $J(\lambda)$ is increasing that the intercepts must be strictly increasing. Hence, (iii) holds.

To prove (iv), assume that the stationary policy e is λ^*-optimal, for some $\lambda^* \in (\lambda_j, \lambda_{j+1})$. This means that $J_e(i, \lambda^*) = J(\lambda^*)$, for $i \in S$. Fix an initial state i. It follows from (2.2) that the linear function $J_e(i, \lambda)$ from (2.1) and the function $J(\lambda)$ must coincide on $[\lambda_j, \lambda_{j+1}]$. Hence, their slopes and intercepts must be equal. Thus, $K_e(i) = K_j$ and $G_e(i) = G_j$. Since this argument may be repeated for every $i \in S$, the result follows.

To prove (v), we set $\lambda = 0$. It then follows from (1.5) and (1.6) that $J(0) = G_0$ is the minimum average C cost. The second statement in (v) then follows from (iv).

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

It remains to prove (vi). Let f be an average D cost optimal stationary policy. We know from (iv) that K_s is the average D cost under a policy that is essential on $[\lambda_s,\infty)$, and hence it follows that $K_f(i) \leq K_s$, for all i. Assume that there exists a state i^* such that $K_f(i^*) < K_s$.

From (2.2), it follows that

$$G_s + \lambda K_s \leq G_f(i^*) + \lambda K_f(i^*), \quad \lambda \in [\lambda_s,\infty). \tag{2.3}$$

Using our assumption, we see that

$$\lambda \leq \frac{G_f(i^*) - G_s}{K_s - K_f(i^*)}, \quad \lambda \in [\lambda_s,\infty). \tag{2.4}$$

This yields a contradiction. Hence, we must have $K_f(i) = K_s$, for all $i \in S$. The second statement in (vi) then follows from (iv). ■

COROLLARY 2.2: *If S is finite and Assumption (F) holds, then COP is feasible if and only if $K_s \leq K$.*

PROOF: This follows immediately from Theorem 2.1(vi) and the definition of feasibility. ■

Consider the following exhaustive set of cases.

Case 1: If $K < K_s$, then COP is infeasible.

Case 2: If $K = K_j$, for some j with $1 \leq j \leq s$, then a stationary policy essential on $[\lambda_j, \lambda_{j+1}]$ solves COP.

Case 3: If $K \geq K_0$, then a stationary policy essential on $[\lambda_0, \lambda_1]$ solves COP.

Case 4: Finally, if $K_j < K < K_{j-1}$, for some j with $1 \leq j \leq s$, then the situation is more complicated. It is discussed fully below.

The result under Case 1 follows from Corollary 2.2. Under Case 2, it follows from Theorem 2.1(iv) and Proposition 1.2. Under Case 3, it follows from Theorem 2.1(iv–v). Thus, Case 4 remains. If K is chosen beforehand, if COP is feasible, and if no C cost optimal policy is feasible, then the probability is essentially 1 that the problem will fall under Case 4.

The remainder of Section 2 discusses solutions to COP under Case 4. Let e be essential on $[\lambda_{j-1}, \lambda_j]$, let f be essential on $[\lambda_j, \lambda_{j+1}]$, and let γ be the break point λ_j. We will refer to (e, f) as an *essential pair* (associated with the break point γ). Both e and f are γ-optimal, and $K_e > K > K_f$.

Two approaches utilizing nonstationary policies are possible when both e and f induce irreducible MC on S. The first approach utilizes a *time-sharing policy* (TSP). This concept was introduced in Ross [18] and further studied in Altman and Shwartz [2,3]. A *cycle* is a first passage to a distinguished state x. The TSP follows e for a certain number of cycles, and then switches to f for a certain number of cycles, and

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

back and forth. A TSP may be constructed to solve COP. The details and proofs are in [3]. Although this approach is appealing, a TSP is a nonstationary policy whose practical implementation may be difficult.

The second approach involves the notion of a *steering policy* (STEP). The possibility that this type of policy might solve COP was presented as an adaptive scheme in Shwartz and Makowski [21], mentioned in Ross [18], and developed further in Ma et al. [11–13]. A STEP operates similarly to a TSP, but the decision (only made in state x) of whether to run the system under e or under f is made on the basis of the current sample average D cost. This may be tracked recursively so that, even though this policy is nonstationary, its implementation is easy.

We will construct a partially randomized policy solving COP. It is well known (e.g., [20, Prop. 6.2.3]) that, for any $\lambda \geq 0$, there exists a stationary policy that is both λ-optimal and expected discount optimal, for discount factors sufficiently close to 1. Such a policy is said to be λ_*-optimal. Our solution is constructed under the following assumption (EP stands for *essential pair*).

ASSUMPTION (F-EP): *Assume that Assumption* (F), *Case 4*, *holds*, *with break point* γ. *Then there exist an essential pair* (e, f) *and a* γ_**-optimal stationary policy* g, *such that the following hold*:

(i) *The Markov chains induced by e and f have a common positive recurrent state x. Let* R_e (*respectively*, R_f) *be the positive recurrent class associated with e* (*respectively*, f) *containing x. Let* $R = R_e \cup R_f$.
(ii) *Every positive recurrent class under the MC induced by g intersects R.*

Let us now assume that Assumption (F-EP) holds, for e, f, and g. For $p \in [0,1]$, the *partially randomized policy* $\psi(p) = \psi(g, p, e, f)$ is defined by

$$\psi(p)(i) = \begin{cases} g(i), & i \notin R \\ e(i), & i \in R_e - R_f \\ f(i), & i \in R_f - R_e \\ \text{chooses } e(i) \text{ w.p. } 1-p \text{ and } f(i) \text{ w.p. } p, & i \in R_e \cap R_f. \end{cases} \tag{2.5}$$

For $p \in (0,1)$, it is easy to see that $\psi(p)$ induces a unichain MC on S with positive recurrent class R. Moreover, $\psi(0)$ (respectively, $\psi(1)$) is unichain with positive recurrent class R_e (respectively, R_f).

The following lemma is similar to results in the literature. Since the proof is short, we give it here.

LEMMA 2.3: *For each* $i \in S$, *the steady-state probability* $\pi_i(\psi(p))$ *is a continuous function of p. The quantities* $G_{\psi(p)}$ *and* $K_{\psi(p)}$ *are continuous functions of p.*

PROOF: For $i \notin R$, we have $\pi_i(\psi(p)) = 0$, for all p, and hence the result holds. Now, focus on R, and let $p_n \to p$. Fix a subsequence of n. We may find a subsequence (for

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

notational ease, continue to denote it by n) and nonnegative limit point δ_i such that $\pi_i(n) =: \pi_i(\psi(p_n)) \to \delta_i$, for $i \in R$. All summations are over $i \in R$. Since $\sum \pi_i(n) = 1$, it follows by taking limits and using the finiteness of R that $\sum \delta_i = 1$. Moreover, we have $\pi_j(n) = \sum P_{ij}(\psi(p_n))\pi_i(n)$, and again taking limits yields $\delta_j = \sum P_{ij}(\psi(p))\delta_i$. Well-known MC theory then implies that $\delta_i = \pi_i(\psi(p))$. Since given any subsequence of n, there exists a subsequence whose steady-state probabilities converge to those for $\psi(p)$, the first statement follows.

Observe that $G_n =: \sum C(i,\psi(p_n))\pi_i(n) = \sum[(1-p_n)C(i,e) + p_n C(i,f)]\pi_i(n)$. Taking limits yields $G_n \to \sum C(i,\psi(p))\pi_i(\psi(p)) = G_{\psi(p)}$. The proof for the D costs is similar. ■

Here is the approach to solving COP under Assumption (F-EP).

PROPOSITION 2.4: *Assume that Assumption (F-EP) holds. Then, for some $p \in (0,1)$, $\psi(p)$ as defined in (2.5) solves COP.*

PROOF: Let e, f, and g be as in Assumption (F-EP). We apply Sennott [20, Thm. 6.4.2] with distinguished state x to obtain the optimality equation

$$J(\gamma) + h(i) = \min_a \left\{ F(i,a,\gamma) + \sum_{j \in S} P_{ij}(a)h(j) \right\}, \quad i \in S. \tag{2.6}$$

For our purposes, the definition of the function h is not important. By [20, Thm. 6.4.2(ii)], g (which is playing the role of f in the theorem) realizes the minimum in (2.6).

Recall that e and f are γ-optimal. It is the case that e realizes the minimum in (2.6) on R_e. To see this, note that (2.6) yields

$$J(\gamma) + h(i) \le F(i,e,\gamma) + \sum_{j \in S} P_{ij}(e)h(j), \quad i \in R_e. \tag{2.7}$$

Assume that (2.7) is a strict inequality for some i_0. Multiply both sides of (2.7) by $\pi_i(e)$, and then sum over $i \in R_e$ and cancel common terms to obtain a contradiction. A similar result holds for f.

Now, define the stationary policy u by

$$u(i) = \begin{cases} g(i), & i \notin R \\ e(i), & i \in R_e - R_f \\ f(i), & i \in R_f - R_e \\ e(i), & i \in R_e \cap R_f. \end{cases} \tag{2.8}$$

The stationary policy w is defined to equal u except on $R_e \cap R_f$, where it will follow f.

By [20, Thm. 6.4.2(iii)], any stationary policy realizing the minimum in (2.6) is average cost optimal. (This is also true for randomized stationary policies.) Both

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

u and w realize the minimum and, hence, are γ-optimal. The randomized policy $\theta(p)$ that chooses u with probability $1-p$ and w with probability p is also γ-optimal. This may be seen by writing out (2.6) for both u and w, then multiplying the u equation by $1-p$ and the w equation by p and summing them. Then, $\theta(p)$ also realizes the minimum in (2.6) and, hence, is optimal. However, $\theta(p)$ coincides with the policy $\psi(p)$ and, hence, $\psi(p)$ is γ-optimal.

We now apply Lemma 2.3 to obtain a value of p such that $K_{\psi(p)} = K$. By Proposition 1.2, the policy $\psi(p)$ then solves COP. ■

The following queuing control problem illustrates the difficulties that can arise if Assumption (F), Case 4, applies but Assumption (F-EP) fails. This example is not a prerequisite to later material and may be omitted without loss of continuity.

Example 2.5: Consider a queuing system with a single server, serving stations A and B. Customers arrive at A (respectively, B) according to a Bernoulli process with probability p (respectively, q), where $p, q \in (0,1]$. The service time at either station follows a geometric distribution with fixed rate $\mu \in (0,1)$. If the server is currently serving a customer at a station, it continues until the service is finished (and the customer then departs the system). When the queue is empty, the server can choose to remain at the current station or to switch stations. If the server switches stations, this happens immediately, before customers arrive. An arriving customer enters if and only if the server is currently at that station and idle, in which case it enters service immediately. Otherwise the customer is lost. For a different view of this matter, see Figure 2.

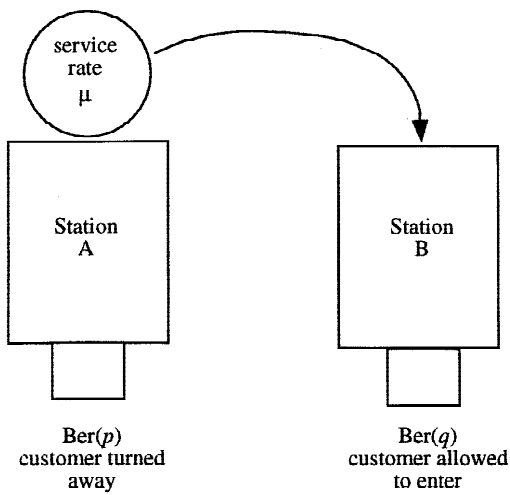


FIGURE 2. Switching to station B.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

Let us formalize this description. There are four system states: 0 (the system is empty and the server is at A), 0^* (the system is empty and the server is at B), 1 (the server is serving a customer at A), and 1^* (the server is serving a customer at B). In state 0, there are decisions A (stay at A) and B (switch to B), and similarly for state 0^*. The costs and transition probabilities are given in Table 1. Here, C (respectively, C^*) is a cost for being at station A (respectively, B) and E is a cost for switching stations. The D cost is a lost opportunity cost associated with the turned-away customers at the other station. The constants C, C^*, H, and E are all positive.

There are four stationary policies: a (always choose A), b (always choose B), c (stay where you are), and d (always switch). Information about the first three policies is as follows:

Policy	Positive Recurrent Class(es)	$J_-(\lambda)$	
a	$\{0,1\}$	$C + \lambda Hq$	
b	$\{0^*,1^*\}$	$C^* + \lambda Hp$	**(2.9)**
c	$\{0,1\}$ $\{0^*,1^*\}$	$C + \lambda Hq$ $C^* + \lambda Hp$	

Table 1. Costs and Transition Probabilities for Example 2.5

State	Decision	C Cost	D Cost	Next State	Probability
1	—	C	Hq	0 1	μ $1-\mu$
1^*	—	C^*	Hp	0^* 1^*	μ $1-\mu$
0	A	C	Hq	0 1	$1-p+p\mu$ $p(1-\mu)$
0	B	C^*+E	Hp	0^* 1^*	$1-q+q\mu$ $q(1-\mu)$
0^*	A	$C+E$	Hq	0 1	$1-p+p\mu$ $p(1-\mu)$
0^*	B	C^*	Hp	0^* 1^*	$1-q+q\mu$ $q(1-\mu)$

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

Policy d induces a positive recurrent class on the whole state space. Let $t = p + q$ and $\delta = (1 - \mu)/\mu$. The steady-state probabilities are easily seen to be

$$\pi_0(d) = \pi_{0^*}(d) = \frac{1}{2 + t\delta}, \qquad \pi_1(d) = \frac{p\delta}{2 + t\delta}, \qquad \pi_{1^*}(d) = \frac{q\delta}{2 + t\delta}. \tag{2.10}$$

Using (2.10), Table 1, and then (2.9), it follows after some algebra that

$$J_d(\lambda) = \frac{2E}{2 + t\delta} + \left(\frac{1 + p\delta}{2 + t\delta}\right) J_a(\lambda) + \left(\frac{1 + q\delta}{2 + t\delta}\right) J_b(\lambda). \tag{2.11}$$

This completes the setup of the model. Let us now assume that $q > p$ and $C^* > C$. Let $\omega =: q - p$, and $U =: C^* - C$. It follows from (2.9) that $J_a(\lambda) < J_b(\lambda)$, for $\lambda \in [0, \gamma)$, where $\gamma = U/H\omega$. We claim that

$$J(\lambda) = \min\{J_a(\lambda), J_b(\lambda)\} = \begin{cases} C + \lambda Hq, & \lambda \in [0, \gamma] \ (a \text{ is essential}) \\ \dfrac{C^* q - Cp}{\omega}, & \lambda = \gamma \\ C^* + \lambda Hp, & \lambda \in [\gamma, \infty) \ (b \text{ is essential}). \end{cases} \tag{2.12}$$

Moreover, c is optimal only at γ and d is not optimal for any λ. The function $J_c(\cdot, \lambda)$ is state dependent and is above the function in (2.12) for some state except at γ, where the two values coincide. It follows from (2.11) that $J_d(\lambda)$ exceeds a convex combination of $J_a(\lambda)$ and $J_b(\lambda)$, and thus $J_d(\lambda) > \min\{J_a(\lambda), J_b(\lambda)\}$.

This verifies that $J(\lambda)$ is given by (2.12). Note that Assumption (F) holds. Let the constraint K satisfy $Hp < K < Hq$. Then, we are under Case 4 with essential pair (a, b), and Assumption (F-EP) fails.

Let us determine the best we can do within the class of randomized stationary policies, with state-dependent randomization factors. When in state 0 (respectively, 0^*), choose A with probability $1 - \alpha$ and B with probability α (respectively, B with probability $1 - \beta$ and A with probability β). Denote this policy by $\theta(\alpha, \beta)$. Assume at least one of α and β nonzero, and let $\sigma =: \alpha(1 + q\delta) + \beta(1 + p\delta)$. An easy calculation yields

$$\pi_0(\theta(\alpha, \beta)) = \frac{\beta}{\sigma}, \qquad \pi_{0^*}(\theta(\alpha, \beta)) = \frac{\alpha}{\sigma},$$

$$\pi_1(\theta(\alpha, \beta)) = \frac{\beta p\delta}{\sigma}, \qquad \pi_{1^*}(\theta(\alpha, \beta)) = \frac{\alpha q\delta}{\sigma}. \tag{2.13}$$

Note that $\alpha = \beta = 1$ yields policy d and agreement with (2.10). If $\alpha = 0$ and $\beta > 0$, then we obtain policy a, which is infeasible. Hence, we may assume that $\alpha > 0$. If $\beta = 0$ and $\alpha > 0$, then we obtain policy b, which is feasible with $G_b = C^*$. We will see that we can do better; hence, let us assume that $\beta > 0$. Policy c need be of no concern, as it does not yield anything new.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

Using (2.13), Table 1, and then (2.9) yields

$$J_{\theta(\alpha,\beta)}(\lambda) = \frac{2E\alpha\beta}{\sigma} + \left(\frac{\beta(1+p\delta)}{\sigma}\right)J_a(\lambda) + \left(\frac{\alpha(1+q\delta)}{\sigma}\right)J_b(\lambda). \tag{2.14}$$

It follows from (2.14) and (2.12), with $\lambda = \gamma$, that

$$J_{\theta(\alpha,\beta)}(\gamma) = J(\gamma) + \frac{2E\alpha\beta}{\sigma}; \tag{2.15}$$

hence, we cannot use Proposition 1.2 to construct a solution to COP.

Now, let $r =: (K - Hp)/(Hq - K)$, and note that $r > 0$. Let $s =: (1 + p\delta)/(1 + q\delta)$ and note that $0 < s < 1$. Let $\eta = \beta/\alpha$. Using (2.14), it follows after some manipulation that

$$K - K_{\theta(\alpha,\beta)} = \frac{(Hq - K)(r - \eta s)}{1 + \eta s} \tag{2.16}$$

and

$$G_{\theta(\alpha,\beta)} = C + \frac{U}{1+\eta s} + \frac{2\beta E}{(1+\eta s)(1+q\delta)}. \tag{2.17}$$

From (2.16), it follows that $\theta(\alpha, \beta)$ is feasible if and only if $\eta \le r/s$. It follows from (2.17) that to minimize $G_{\theta(\alpha,\beta)}$ for a fixed value of β, we should make η as large as possible and, hence, α as small as possible. Combining this with the requirement that $\eta \le r/s$ yields $\alpha = (s/r)\beta$. (Choose $\beta \le r/s$ to ensure that $\alpha \le 1$.)

Denote the resulting policy by $\theta^*(\beta)$. Substituting $\alpha = (s/r)\beta$ into (2.16) and (2.17) yields $K_{\theta^*(\beta)} = K$ and

$$G_{\theta^*(\beta)} = C + \frac{U}{1+r} + \frac{2\beta E}{(1+r)(1+q\delta)}. \tag{2.18}$$

We cannot set $\beta = 0$. However, we can realize within ε of

$$G_{\min} =: C + \frac{U}{1+r}. \tag{2.19}$$

Note that $G_{\min} < C^*$, so we are doing better than feasible policy b.

[Here is how policy d relates to this solution. Substituting $\eta = 1$ into (2.16) shows that any policy with $\alpha = \beta$ (including d for which both quantities equal 1) is feasible if and only if $s \le r$. So let us assume that $s \le r$. Then, substituting $\beta = \eta = 1$ into (2.17), one easily shows that $G_d > G_{\min}$.]

It follows from (2.15) that

$$J_{\theta^*(\beta)}(\gamma) = J(\gamma) + \frac{2\beta E}{(1+r)(1+q\delta)}. \tag{2.20}$$

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

We may solve (2.18) for the last term on the right and substitute this into (2.20) to obtain

$$J(\gamma) = G_{\min} + \gamma K. \tag{2.21}$$

If ψ is any nonstationary feasible policy, then it follows from (2.21) that $G_{\min} \leq G_{\psi}(i)$, for all i. Hence, a policy $\theta^*(\beta)$ that is within ε of optimality within the class of all feasible policies may be constructed.

Remark 2.6: Here, we discuss the example in Haviv [7]. Clearly, Assumption (F) is not true for this example. Let us consider the essential properties of the example.

There is a stationary policy u and states 0 and 1 such that $K_u(0) > K$ and $K_u(1) \leq K$. Moreover, there is a feasible stationary policy d such that $G_u(1) < G_d(1)$. This means that u is infeasible; hence, the minimization in (1.3) will not include u. If the process is in state 1, the controller will operate under d (if it has the lowest average C cost among feasible policies) rather than under u, which makes more sense.

Given this setup, assume that Assumption (F) holds. Under Case 1, COP is infeasible and policy d cannot exist. Thus, assume that one of the other cases applies. If Case 2 applies, then it is easy to see that $G_j \leq G_u(1)$. By Theorem 2.1(iv), any policy e essential on that interval satisfies $G_e(1) = G_j$. Hence, $G_e(1) \leq G_u(1)$, and there is no loss of generality in restricting attention to feasible policies. If Case 3 applies, then any policy e essential on the first interval is feasible and average C cost optimal and, hence, $G_e(1) \leq G_u(1)$.

Now, assume that Assumption (F-EP) holds and let $\psi(p)$ be as in Proposition 2.4. Since $\psi(p)$ is γ-optimal with $K_{\psi(p)} = K$, for all initial states, it again follows that $G_{\psi(p)} \leq G_u(1)$.

Hence, the essential properties of the example in [7] cannot occur.

3. VALUE ITERATION SOLUTION TO COP FOR FINITE STATE SPACES

In this section the state space continues to be finite. The following stronger assumption will enable us to construct a solution to COP using value iteration.

ASSUMPTION (F*): *For $\lambda \in (0,\infty)$, every λ-optimal stationary policy induces a unichain MC with aperiodic positive recurrent class.*

The aperiodicity assumption is not crucial. If it does not hold, then a transformation may be effected so that it will hold in the transformed system. For ease of presentation, we work under the aperiodicity assumption. If e is a λ-optimal stationary policy, then the unique positive recurrent class of the associated MC is denoted R_e.

LEMMA 3.1: *Assume that Assumption (F*) holds. Then, Assumption (F) holds. Moreover, the following hold*:

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

(i) *If policies e and f are λ-optimal stationary policies, then* $R_e \cap R_f \neq \emptyset$.
(ii) *If we are under Case 4, then Assumption (F-EP) holds for any essential pair and any γ-optimal policy.*

PROOF: There exists a λ-optimal stationary policy. Since it induces a unichain MC, it follows that $J(\lambda)$ is constant, and hence Assumption (F) holds.

Now, assume that e and f are λ-optimal, but that $R_e \cap R_f = \emptyset$. Define the stationary policy u by

$$u(i) = \begin{cases} e(i), & i \in S - R_f \\ f(i), & i \in R_f. \end{cases} \tag{3.1}$$

It is readily seen that u is λ-optimal with two positive recurrent classes, violating Assumption (F*). Hence, $R_e \cap R_f \neq \emptyset$.

Now, assume we are under Case 4. Let (e, f) be an essential pair associated with break point γ, and let g be γ-optimal. Then, e and f are γ-optimal; hence, by (i), it follows that $R_e \cap R_f \neq \emptyset$ and Assumption (F-EP)(i) holds. Again by (i), we must have $R_e \cap R_g \neq \emptyset$. Thus, Assumption (F-EP)(ii) and, hence, Assumption (F-EP) hold. ■

We wish to compute a λ-optimal stationary policy using the value iteration algorithm (VIA). See Sennott [20, Sect. 6.6] or Puterman [16]. Fix an arbitrary base point x. For each $\lambda \in (0,\infty)$, Assumption (F*) guarantees that VIA will yield a λ-optimal stationary policy. This policy realizes the minimum in the VIA optimality equation

$$J(\lambda) + r(i,\lambda) = \min_a \left\{ F(i,a,\lambda) + \sum_j P_{ij}(a) r(j,\lambda) \right\}, \quad i \in S, \tag{3.2}$$

where $r(i,\lambda) = \lim_{k\to\infty}[\nu_k(i,\lambda) - \nu_k(x,\lambda)]$ and $J(\lambda)$ is obtained as $\lim_{k\to\infty}[\nu_k(x,\lambda) - \nu_{k-1}(x,\lambda)]$. Here, k is the horizon length, ν is the k horizon value function, and r is the associated relative value function.

These computed policies may then be used in the construction of a solution to COP. The only interesting situation is Case 4, so assume that this case holds.

Let (e, f) be an essential pair associated with break point γ. Assume that e realizes the minimum in (3.2) for some sequence $\lambda_n \to \gamma^-$. (This must be true for some stationary policy, since there are at most finitely many of them.) Assume a similar result for f, for a sequence converging to γ from the right. Then, we say that (e, f) is a VI *essential pair*. The next result tells us that both e and f will be obtained using VIA for γ. This is a *continuity* result for VI essential pairs.

LEMMA 3.2: *Assume that Assumption (F*), Case 4, holds. Let (e, f) be a VI essential pair associated with break point γ. Then, e and f realize the minimum in the VIA optimality equation*

$$J(\gamma) + r(i,\gamma) = \min_a \left\{ F(i,a,\gamma) + \sum_j P_{ij}(a) r(j,\gamma) \right\}, \quad i \in S. \tag{3.3}$$

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

PROOF: We prove the result for e, as the proof for f is similar. Let F_n denote the (1.4) cost and r_n the relative value function associated with λ_n, and let J_n be the corresponding minimum average cost. By assumption, e realizes the minimum in

$$J_n + r_n(i) = \min_a \left\{ F_n(i,a) + \sum_j P_{ij}(a) r_n(j) \right\}, \quad i \in S, \text{ all } n. \tag{3.4}$$

For each n, there exists a λ_n-optimal stationary policy u_n that is also discount optimal, for the discount factor sufficiently close to 1. As there are only finitely many stationary policies, there must exist a stationary policy u and subsequence of n (continue to call it n) such that $u_n = u$ for all members of the subsequence.

By Lemma 3.1(i), there exists $y \in R_e \cap R_u$. Let w_n be the relative value function calculated for the base point y instead of x. Then, we have $r_n = w_n + r_n(y)$. Subtracting the constant $r_n(y)$ from both sides of (3.4) will not affect its solutions, so that e realizes the minimum in

$$J_n + w_n(i) = \min_a \left\{ F_n(i,a) + \sum_j P_{ij}(a) w_n(j) \right\}, \quad i \in S, \text{ all } n. \tag{3.5}$$

It follows from [20, Prop. 6.5.1(iii), with $f = u$, $r = w$, and $x = y$] that

$$w_n(i) = c_{iy}(e) + \lambda_n d_{iy}(e) - J_n m_{iy}(e), \quad i \in S. \tag{3.6}$$

Here, $m_{iy}(e)$ (respectively, $c_{iy}(e), d_{iy}(e)$) is the expected first passage time (respectively, expected first passage C, D cost) from i to y under e. Note that (3.6) also holds for $i = y$ and equals 0.

If we define $s(i) =: c_{iy}(e) + \gamma d_{iy}(e) - J(\gamma) m_{iy}(e)$, then by Theorem 2.1(i), it follows that $w_n \to s$. Note that $s(y) = 0$. Taking the limit of both sides of (3.5) yields

$$J(\gamma) + s(i) = \min_a \left\{ F(i,a,\gamma) + \sum_j P_{ij}(a) s(j) \right\}, \quad i \in S, \tag{3.7}$$

and e realizes the minimum.

Recall that we wish to prove that e realizes the minimum in (3.3). Two more applications of (*) [20, Prop. 6.5.1(iii)] will yield the result. To set this up, let q be a γ-optimal stationary policy that is also discount optimal for discount factor sufficiently close to 1, and let $y^* \in R_e \cap R_q$. Let h_{y^*} be the function in [20, Thm. 6.4.2(i)] with base point y^*. If we subtract $s(y^*)$ from both sides of (3.7), then e still realizes the minimum. An application of (*) (with $f = q$, $r = s - s(y^*)$, and $x = y^*$) yields $s - s(y^*) = h_{y^*}$.

Now, let g be a stationary policy realizing the minimum in (3.3) and let $y^{**} \in R_g \cap R_q$. An application of (*) (with $f = q$, $e = g$, $r = r(\cdot,\gamma) - r(y^{**},\gamma)$, and $x = y^{**}$) yields $r(\cdot,\gamma) - r(y^{**},\gamma) = h_{y^{**}}$. Since $h_{y^{**}} = h_{y^*} + h_{y^{**}}(y^*)$, a bit of algebra shows that $r(\cdot,\gamma) - s = r(y^{**},\gamma) - s(y^*) + h_{y^{**}}(y^*)$; hence, these two functions differ by a constant. Therefore, e must realize the minimum in (3.3). ■

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

The following result strengthens Proposition 2.4.

PROPOSITION 3.3: *Assume that Assumption* (F^*), *Case 4, holds. Let* (e, f) *be a VI essential pair associated with break point* γ. *Then, for some* $p \in (0,1)$, *the randomized policy* $\theta(p)$ *that chooses* e *with probability* $1-p$ *and* f *with probability* p *is* γ*-optimal with* $K_{\theta(p)} = K$, *and hence solves COP.*

PROOF: For $p \in (0,1)$, it is easy to see that $\theta(p)$ induces a unichain MC with $R_{\theta(p)} \supset R_e \cup R_f$. Moreover, $\theta(0) = e$ and $\theta(1) = f$.

The proof of Lemma 2.3 may be modified to show that the steady-state probabilities $\pi_i(\theta(p))$ as well as $G_{\theta(p)}$ and $K_{\theta(p)}$ are continuous functions of p.

We may then obtain $p \in (0,1)$ such that $\theta(p)$ satisfies $K_{\theta(p)} = K$. It follows from Lemma 3.2 that both e and f realize the minimum in (3.3). We may then write out (3.3) for both e and f, multiply the first equation by $1-p$ and the second equation by p, and sum them. Then, $\theta(p)$ also realizes the minimum in (3.3) and, hence, is γ-optimal. It follows from Proposition 1.2 that $\theta(p)$ solves COP. ■

The following situation arises in computation. Let (e, f) be a VI essential pair associated with break point $\gamma = \lambda_j$. Let $\delta \in (\lambda_{j-1}, \lambda_j)$ and let $\omega \in (\lambda_j, \lambda_{j+1})$. We will have computed e, f, δ, ω, $J(\delta)$, K_{j-1}, $J(\omega)$, K_j, and the constrained optimal policy $\theta(p)$. It is desired to find γ and the minimum constrained average cost $G_{\theta(p)}$.

Since $J(\lambda)$ is a continuous function and linear on each of the segments (δ, γ) and (γ, ω), we may set up simultaneous linear equations and solve for γ. This yields

$$\begin{aligned} \gamma &= \frac{G_j - G_{j-1}}{K_{j-1} - K_j}, \\ J(\gamma) &= G_{j-1} + \gamma K_{j-1}, \\ G_{\theta(p)} &= J(\gamma) - \gamma K_{\theta(p)} \cong J(\gamma) - \gamma K. \end{aligned} \tag{3.8}$$

4. COP FOR DENUMERABLE STATE SPACES

In this section, Δ has a denumerably infinite state space S, and it is desired to compute a solution to COP. Several state truncation and approximation schemes are discussed in Altman [1, Chap. 16]. His approaches for defining finite state approximations are similar to ours, but the theory is developed under a strong contraction condition. Our approach is based on the technique developed in Sennott [20]. It is more general and flexible than the approach in [1], but the rates of convergence for the method have not been developed.

Assume that an approximating sequence (Δ_N) has been defined as in [20, Def. 2.5.2]. Informally, this consists of an increasing sequence of finite sets S_N whose union is S. The C and D costs and actions are unchanged on S_N. Given $i \in S_N$ and action a, there is a probability distribution $P_{ij}(a; N)$ on S_N the converges pointwise to the distribution $P_{ij}(a)$ on S. [At this point, the reader may wish to refer to the constructions in Sections 6 and 7.] Let COP_N refer to the constrained problem for Δ_N, with the same positive constraint K.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

ASSUMPTION (C1): *For every N, Assumption* (F^*) *holds in* Δ_N.

ASSUMPTION (C2): *For sufficiently large N,* COP_N *is feasible.*

These assumptions allow us to apply the theory in Section 3. We will assume that Case 4 applies (if not, this development may be modified). We may compute a VI essential pair and then use Proposition 3.3 to find a solution to COP_N. Quantities occurring in Δ_N will be indicated by appending the variable N.

Let $(e(N), f(N))$ be a VI essential pair for Δ_N, associated with break point $\gamma(N)$, and let $\theta(N) = \theta(p(N), N)$ be the constrained optimal policy given by Proposition 3.3. We assume that this setup holds as the background for the remaining assumptions.

We require the notion of a policy θ for Δ that is a limit point of the policies $\theta(N)$. This policy is our candidate to solve COP.

DEFINITION 4.1: *With the above setup, assume that there exist* $p \in [0,1]$, *stationary policies e and f for* Δ, *and a subsequence* N_k *such that the following hold*:

(1) $p(N_k) \to p$.
(2) *For each* $i \in S$, *we have* $e(N_k)(i) = e(i)$ *and* $f(N_k)(i) = f(i)$, *for sufficiently large k* (*dependent on i*).

Then, the randomized stationary policy $\theta = \theta(p)$ *that chooses e with probability* $1 - p$ *and chooses f with probability p is a* limit point *of* $\theta(N)$ (*associated with the subsequence* N_k).

LEMMA 4.2: *The sequence* $\theta(N)$ *has at least one limit point.*

PROOF: The proof follows that of [20, Prop. B.3]. By defining $e(N)$ and $f(N)$ arbitrarily outside of S_N, we may assume that $\theta(N)$ is defined on S. Then, $\theta(N)$ may be identified with a sequence in the compact metric space $[0,1] \times (\Pi_{i \in S} A_i) \times (\Pi_{i \in S} A_i)$. The proof proceeds as on [20, p. 289]. ■

The remaining assumptions are as follows:

ASSUMPTION (C3): *There exists* δ *such that* $\gamma(N) \leq \delta < \infty$ *and* $J(i, \delta) < \infty$, *for* $i \in S$.

ASSUMPTION (C4): *If* N_k *is a subsequence of N such that* $\gamma(N_k) \to \gamma$, *then* $\liminf_{k \to \infty} J(\gamma(N_k), N_k) \leq J(i, \gamma)$, *for* $i \in S$.

ASSUMPTION (C5): *If* θ *is a limit point of* $\theta(N)$ *associated with the subsequence* N_k, *then* $\liminf_{k \to \infty} G_{\theta(N_k)}(N_k) \geq G_\theta(i)$ *and* $\liminf_{k \to \infty} K_{\theta(N_k)}(N_k) \geq K_\theta(i)$, *for* $i \in S$.

The interpretation of these conditions is simple. Assumption (C3) states that the sequence of break points is bounded and the Lagrange function in Δ is finite at some upper bound. Assumption (C4) states that if a subsequence of break points converges to a number γ, then the limit infimum of the corresponding Lagrange values is bounded above by the Lagrange function at γ in Δ. Assumption (C5) states that the limit infimum of the average C (respectively, D) cost associated with solu-

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

tions in Δ_N that converge to a limit point is bounded below by the average C (respectively, D) cost in Δ under the limit point.

Here is the major result of the paper.

THEOREM 4.3: *Assume that Assumptions (C1)–(C5) hold. Then, any limit point θ of the constrained optimal policies in Δ_N is γ-optimal in Δ, for some $\gamma \in [0,\infty)$. Moreover, $G_\theta(i) < \infty$ and $K_\theta(i) = K$, for $i \in S$, and hence θ solves COP.*

PROOF: By Assumptions (C1) and (C2), the above setup, and Proposition 3.3, it follows that

$$J(\gamma(N), N) = G_{\theta(N)}(N) + \gamma(N)K_{\theta(N)}(N). \tag{4.1}$$

Let θ be a limit point associated with the subsequence N_k. By Assumption (C3), it follows that there is a subsequence of N_k (to avoid double subscripts, denote it N_j) and a number γ such that $\gamma(N_j) \to \gamma$.

Take the limit infimum of both sides of (4.1) through values of the subsequence N_j to obtain

$$\begin{aligned} J(i,\gamma) &\geq \liminf_{j\to\infty}\{G_{\theta(N_j)}(N_j) + \gamma(N_j)K_{\theta(N_j)}(N_j)\} \\ &\geq \liminf_{j\to\infty} G_{\theta(N_j)}(N_j) + \gamma\{\liminf_{j\to\infty} K_{\theta(N_j)}(N_j)\} \\ &\geq \liminf_{k\to\infty} G_{\theta(N_k)}(N_k) + \gamma\{\liminf_{k\to\infty} K_{\theta(N_k)}(N_k)\} \\ &\geq G_\theta(i) + \gamma K_\theta(i), \quad i \in S. \end{aligned} \tag{4.2}$$

The first line follows from Assumption (C4). The second and third lines follow from properties of the limit infimum. The last line follows from Assumption (C5). It then follows from (1.5) and (1.6) and Definition 1.1 that θ is γ-optimal.

It follows from Assumption (C3) and the fact that $J(i,\lambda)$ is increasing in λ that $J(i,\gamma)$ (hence, $G_\theta(i)$) is finite. Since $K_{\theta(N)} = K$, for all N sufficiently large, it follows from Assumption (C5) that $K_\theta(i) \leq K$, for all i, so that θ is feasible. Let us consider two cases. First, assume that $\gamma = 0$. Then, θ is a feasible C cost optimal policy and, hence, it solves COP. Now, assume that $\gamma > 0$. Suppose for the moment that $K_\theta(i) < K$, for some i. This implies that $J(i,\gamma) < G_\theta(i) + \gamma K$. However, it follows from (4.2) that $J(i,\gamma) \geq G_\theta(i) + \gamma K$, which yields a contradiction. Hence, $K_\theta(i) = K$, for $i \in S$, and Proposition 1.2 shows that θ solves COP. ■

5. VERIFICATION OF ASSUMPTIONS (C1)–(C5)

Usually, Assumption (C1) may be easily verified. However, direct verification of the other assumptions is typically difficult. General techniques developed in [20] allow us to give sufficient conditions for Assumptions (C2)–(C5), which are quite weak and will hold in numerous systems.

We first discuss an important type of policy. Let z be a distinguished state in S. Assume that g is a (randomized) stationary policy for Δ such that the MC induced

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

by g satisfies $m_{iz}(g) < \infty$, $c_{iz}(g) < \infty$, and $d_{iz}(g) < \infty$, for all $i \in S$. This implies that the induced MC has a single positive recurrent class R_g containing z and that the average C cost (respectively, D cost) is a finite constant G_g (respectively, K_g), for $i \in S$. The policy g (as well as the induced MC) is said to be *z-standard* [20, pp. 143, 301].

If g is any (randomized) stationary policy for Δ, then making the same decisions in Δ_N will create a (randomized) stationary policy $g|N$ for Δ_N. If g is z-standard, then the *conformity* of (Δ_N) with respect to g is defined in [20, Def. C.4.8]. This definition must be extended in an obvious way to cover the D costs. The definition implies that the MC induced by $g|N$ is unichain for sufficiently large N, with z an element of the positive recurrent class. In addition, the steady-state probabilities converge correctly, as do the average C and D costs. For details, see [20, Prop. C.4.9].

PROPOSITION 5.1: *Assume that Assumption (C1) holds. If there exists a z-standard (randomized) stationary policy g such that* $K_g < K$ *and* (Δ_N) *is conforming with respect to g, then Assumptions (C2) and (C3) hold.*

PROOF: Let $\varepsilon =: K - K_g > 0$. Since G_g is finite, we may choose and fix a positive multiplier δ such that $(G_g + 1)\delta^{-1} < \varepsilon/4$.

It follows from conformity that for N exceeding some N_1, we have

$$G_{g|N}(N) < G_g + 1 \quad \text{and} \quad K_{g|N}(N) < K_g + \varepsilon/4. \tag{5.1}$$

By Assumption (C1), it is the case that the Lagrange function in Δ_N is independent of the state. For $N > N_1$, we then have

$$\begin{aligned} J(\delta, N) &\le G_{g|N}(N) + \delta K_{g|N}(N) \\ &\le G_g + 1 + \delta(K_g + \varepsilon/4). \end{aligned} \tag{5.2}$$

If δ is not a break point of the piecewise linear function $J(\lambda, N)$, then let $K_\delta(N)$ be the slope of the corresponding linear segment. If δ is a break point, then it is associated with two slopes. Let $K_\delta(N)$ be the larger one. Theorem 2.1 implies that $K_\delta(N)$ is a decreasing function of δ. Clearly, we have $J(\delta, N) \ge \delta K_\delta(N)$. Combining this with (5.2) yields

$$\begin{aligned} K_\delta(N) &\le (G_g + 1)/\delta + K_g + \varepsilon/4 \\ &\le K_g + \varepsilon/2 \\ &< K. \end{aligned} \tag{5.3}$$

It then follows from Corollary 2.2 that COP_N is feasible by $N > N_1$. Hence, Assumption (C2) holds.

Moreover, it is the case that $\gamma(N) \le \delta$. Note that $J(i, \lambda) \le J_g(i, \lambda) = G_g + \lambda K_g < \infty$, for all λ and i. Hence, Assumption (C3) holds. ■

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

We now give two methods for verifying Assumption (C4).

PROPOSITION 5.2: *Assume that Assumptions (C1)–(C3) hold, and that for each* $\lambda \in (0,\infty)$, *there exists a z-standard* λ*-optimal stationary policy* $g(\lambda)$ *such that* (Δ_N) *is conforming with respect to* $g(\lambda)$. *Then, Assumption (C4) holds.*

PROOF: Let N_k be a subsequence such that $\gamma(N_k) \to \gamma$, and let g be a z-standard γ-optimal policy such that (Δ_N) is conforming with respect to g. Let $\varepsilon > 0$.

For sufficiently large k, it follows that

$$\begin{aligned} J(\gamma(N_k), N_k) &\le J(\gamma + \varepsilon, N_k) \\ &\le J_{g|N_k}(\gamma + \varepsilon, N_k) \\ &= G_{g|N_k}(N_k) + (\gamma + \varepsilon) K_{g|N_k}(N_k). \end{aligned} \tag{5.4}$$

Now take the limit infimum of both sides of (5.4). By conformity, the limit of the right side exists and yields

$$\liminf_{k\to\infty} J(\gamma(N_k), N_k) \le G_g + (\gamma + \varepsilon) K_g. \tag{5.5}$$

Since ε is arbitrary and g is γ-optimal, we have

$$\liminf_{k\to\infty} J(\gamma(N_k), N_k) \le G_g + \gamma K_g = J(\gamma) \tag{5.6}$$

and Assumption (C4) holds. ■

PROPOSITION 5.3: *Assume that Assumptions (C1)–(C3) hold. If for each* $\lambda \in (0,\infty)$, *we have*

$$\nu_n(i, \lambda, N) \le \nu_n(i, \lambda), \quad n \ge 1,\ i \in S_N, \tag{5.7}$$

then Assumption (C4) holds.

PROOF: The condition in (5.7) is that the expected (1.4) cost for operating optimally in Δ_N for n steps cannot be larger than the same cost in Δ.

Let N_k be a subsequence such that $\gamma(N_k) \to \gamma$, and let $\varepsilon > 0$. Fix an initial state i and let k be so large that $i \in S_{N_k}$. Then, for sufficiently large k and an arbitrary policy ψ, it follows that

$$\begin{aligned} J(\gamma(N_k), N_k) &\le J(\gamma + \varepsilon, N_k) \\ &= \lim_{n\to\infty} \frac{\nu_n(i, \gamma + \varepsilon, N_k)}{n} \\ &\le \limsup_{n\to\infty} \frac{\nu_n(i, \gamma + \varepsilon)}{n} \\ &\le \limsup_{n\to\infty} \frac{1}{n} E_\psi \left[\sum_{t=0}^{n-1} F(X_t, A_t, \gamma + \varepsilon) | X_0 = i \right] \\ &= J_\psi(i, \gamma + \varepsilon). \end{aligned} \tag{5.8}$$

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

The second line follows from [20, Thm. 6.4.2(v)]. The third line follows from (5.7), and the fifth line from (1.5).

Since ε and ψ are arbitrary, Assumption (C4) must hold. ■

Finally, we develop a sufficient condition for Assumption (C5). The following lemma sets the stage.

LEMMA 5.4: *Assume that Assumptions (C1) and (C2) hold. Let $\theta = \theta(p)$ be a limit point of $\theta(N)$ associated with the subsequence N_k, and let Γ be the Markov chain induced by θ. Then, the following hold*:

(i) *For i transient or null recurrent in Γ, we have $\pi_i(\theta(N_k)) \to 0$.*
(ii) *Given a positive recurrent class R in Γ and a subsequence of N_k, there exist a subsequence N_j and $b \in [0,1]$ such that $\pi_i(\theta(N_j)) \to b\pi_i(\theta)$, for $i \in R$.*
(iii) *Assume that θ is z-standard and that $\pi_i(\theta(N_k)) \to \pi_i(\theta)$. Then, Assumption (C5) holds.*

PROOF: A result from [20] will be brought to bear to give a short proof. We first construct an approximating sequence (Γ_{N_k}) for Γ following [20, Def. C.4.1]. (Since this construction is solely for the purpose of applying [20, Prop. C.4.3], we may ignore the costs.)

The state space of Γ_{N_k} is S_{N_k}. Associated with state i, define the probability distribution

$$P_{ij}(N_k) = (1 - p(N_k))P_{ij}(e(N_k)) + p(N_k)P_{ij}(f(N_k)), \quad j \in S_{N_k}. \tag{5.9}$$

Note that this is the distribution associated with $\theta(N_k)$. Since the quantity in (5.9) converges to $P_{ij}(\theta)$, this forms an approximating sequence for Γ. An application of [20, Prop. C.4.3], with N replaced by N_k, immediately yields (i) and (ii).

To prove (iii), note that

$$G_{\theta(N_k)}(N_k) = \sum_{i \in S_{N_k}} [(1 - p(N_k))C(i, e(N_k)) + p(N_k)C(i, f(N_k))]\pi_i(\theta(N_k)). \tag{5.10}$$

Then, assuming the hypotheses of (iii), it follows that

$$\begin{aligned} \liminf_{k \to \infty} G_{\theta(N_k)}(N_k) &\geq \sum_{i \in S} [(1 - p)C(i, e) + pC(i, f)]\pi_i(\theta) \\ &= G_\theta \\ &= G_\theta(i), \quad i \in S. \end{aligned} \tag{5.11}$$

The first line follows from [20, Prop. A.2.5], and the last line follows from the delayed renewal reward theorem. The proof for the D costs is similar. ■

The essence of this result is that Assumption (C5) will hold if the steady-state probabilities converge correctly. The next result shows that this holds when the D costs are unbounded outside finite sets.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

PROPOSITION 5.5: *Assume that Assumptions (C1) and (C2) hold and that, for all* $U > 0$, *the set* $B(U) = \{i \in S | D(i,a) \leq U \text{ for some } a\}$ *is finite. Then, the following hold*:

(i) *The MC induced by a limit point* θ *has at least one positive recurrent state* z.
(ii) *If* θ *is z-standard, then Assumption (C5) holds.*

PROOF: Let $B =: B(2K)$ and assume that N is so large that $S_N \supset B$. For $i \in S_N - B$, it is the case that $D(i,a) > 2K$, for all a.

Let the limit point θ be associated with the subsequence N_k. Then,

$$\begin{aligned} K &= \sum_{i \in S_{N_k}} [(1 - p(N_k))D(i, e(N_k)) + p(N_k)D(i, f(N_k))]\pi_i(\theta(N_k)) \\ &\geq \sum_{i \in S_{N_k} - B} [(1 - p(N_k))D(i, e(N_k)) + p(N_k)D(i, f(N_k))]\pi_i(\theta(N_k)) \\ &\geq 2K\left(\sum_{i \in S_{N_k} - B} \pi_i(\theta(N_k))\right) \\ &= 2K\left(1 - \sum_{i \in B} \pi_i(\theta(N_k))\right). \end{aligned} \tag{5.12}$$

Let us assume that all the states of B are transient or null recurrent. Taking the limit of both sides as $k \to \infty$, it follows from Lemma 5.4(i) that $K \geq 2K$, which is a contradiction (since $K > 0$). This proves (i).

Note that (ii) will follow from Lemma 5.4(iii) if we can show that b from Lemma 5.4(ii) always equals 1. Assume that $b < 1$ and let $B =: B[2K/(1-b)]$. Repeating the argument in (5.12) yields

$$K \geq \left(\frac{2K}{1-b}\right)\left(1 - \sum_{i \in B} \pi_i(\theta(N_j))\right). \tag{5.13}$$

Taking the limit of both sides as $j \to \infty$ and using Lemma 5.4(i–ii) yields

$$\begin{aligned} K &\geq \left(\frac{2K}{1-b}\right)\left(1 - b \sum_{i \in B \cap R_\theta} \pi_i(\theta)\right) \\ &\geq \left(\frac{2K}{1-b}\right)(1-b), \end{aligned} \tag{5.14}$$

which is a contradiction. Hence, $b = 1$ and this completes the proof of (ii). ■

6. EXAMPLE WITH AVERAGE NUMBER OF CUSTOMERS CONSTRAINT

In this section, we illustrate the theory by computing constrained optimal policies for a single-server queueing system. Arrivals follow a Bernoulli process, with probability $p \in (0,1)$ of a customer arriving at the beginning of a slot. An arrival is

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

registered during that slot and joins the buffer at the end of that slot. The service time of a customer follows a geometric distribution. The state $i = 0,1,2,\ldots$ is the number of customers in the queue (including the one being served) at the beginning of a slot. If $i \geq 1$, then the server chooses action $a \in \{a_1, a_2, \ldots, a_M\}$, where the actions are the allowable rates of the service time distribution and satisfy $0 < a_1 < \cdots < a_M < 1$. A new rate may be chosen in each slot for which the buffer is nonempty.

Setting $D(i) = i$ and $C(i,a) = C(a) > 0$, for $i \geq 1$, completes the specification of Δ. Then, COP requires us to minimize the average cost of service subject to a constraint on the average number of customers in the system.

To define (Δ_N), let $S_N = \{0,1,\ldots,N\}$. The system can escape S_N only if it is in state N, there is no service completion, and a customer arrives. This happens with probability $(1-a)p$. This *excess probability* is given to N, and all other probabilities are unchanged. This construction defines an *augmentation-type approximating sequence* with excess probability given to N. See [20, pp. 29–30]. Note that $P_{NN-1}(a;N) = a(1-p)$ and $P_{NN}(a;N) = ap + (1-a)(1-p) + (1-a)p = 1 - a(1-p)$. This is not the same as assuming that a customer who arrives when the buffer contains N is lost.

PROPOSITION 6.1: *If $p < a_M$ and $p(1-p)(a_M - p)^{-1} < K$, then Assumptions (C1)–(C5) hold.*

PROOF: It is easy to see that every stationary policy for Δ_N induces an irreducible aperiodic MC on S_N and, hence, (C1) holds.

Let a be a service rate exceeding the arrival rate p, and let $g = g(a)$ be the stationary policy for Δ that always serves at rate a. It follows from [20, Prop. 8.5.1] that g induces a positive recurrent MC on S, with

$$K_{g(a)} = \frac{p(1-p)}{a-p}, \qquad G_{g(a)} = \frac{pC(a)}{a}, \quad a > p. \tag{6.1}$$

The conformity follows from [20, Cor. C.5.4]. Hence, it follows from the hypotheses and Proposition 5.1 applied to $g(a_M)$ that Assumptions (C2) and (C3) hold.

We may give an informal argument that (5.7) holds, since it is clear that the cost of an optimal n-step operation in Δ_N cannot exceed the analogous cost in Δ, for any initial state. Alternatively, we may apply [20, Prop. 3.3.4]. Thus, it follows from Proposition 5.3 that Assumption (C4) holds.

It follows from Proposition 5.5(i) that a limit point θ has at least one positive recurrent state. Since θ induces an irreducible MC on S, the validity of Assumption (C5) follows from Proposition 5.5(ii). ■

The computation takes place in Δ_N and we temporarily omit mention of N. Version 1 of the VIA [20, p. 119] is employed to compute a λ-optimal stationary policy. The VIA equations are given in [20, Eq. (8.11)] with Lagrange multiplier H.

The discussion in Denardo [5] presages the approach we use. The program first computes an interval containing the break point γ. This is done by prompting the user for values of λ, starting with $\lambda = 1$. The slope of $J(1)$ is computed. If it is less

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

than or equal to K, then $\gamma \in [0,1]$, whereas if it is greater than K, then $\gamma \in (1,\infty)$. The user is prompted for another value of λ, and the process continues until an interval (*left*, *right*) containing γ has been found, such that *left* $\neq 0$. The optimal policies *leftopt* and *rightopt* are computed.

Then, bisection is employed on this interval to narrow it further. The bisection terminates when *right* $-$ *left* ≤ 0.005 or when the policies *leftopt* and *rightopt* differ on a single state. Using *leftopt* and *rightopt* as the policies forming an essential pair, bisection is then employed on $[0,1]$ to find a value u for which $K_{\theta(u)} \cong K$. The resulting policy solves COP_N. The program also uses (3.8) to calculate γ, $J(\gamma)$, and $G_{\theta(u)}$.

One may then recompute for larger values of N until the resulting policy is unchanged. (Since we already know the value of γ from the previous calculation, we can hone in on it much more rapidly than in the initial run.) Theorem 4.3 then gives us confidence that a solution to COP has been found. The optimal policy will typically be determined for small N, but obtaining the associated quantities accurate to three decimal places requires larger N.

Keep in mind the assumption that we are always in Case 4. For example, suppose we set $p = 0.4$, $a_1 = 0.6$, $a_2 = 0.8$, $a_3 = 0.9$, and $C(a_1) = 1$, $C(a_2) = 5$, $C(a_3) = 10$. For $g = g(a_3)$, it follows from (6.1) that $K_g = 0.48$. If we set $K = 1.5$ and run the program, it will not work. This is because the policy $g(a_1)$ minimizes the average cost of service and also meets the constraint, and, hence, we are not in Case 4.

The results of various scenarios are displayed in Table 2. A dash in any cell indicates that that entry is the same as in the previous scenario. For Scenario 1, the computations were performed for $N = 30$ and $N = 50$. The system is stable under all three constant rate policies g and the entries in rows 4 and 5 come from (6.1). The constraint is met by $g(0.8)$ and $g(0.9)$, but not by $g(0.6)$. Note that the value of G_θ makes sense.

Let us describe the optimal policy. The buffer content is given in square brackets and the corresponding service rate in parentheses. The optimal policy randomizes in state 1, choosing rate 1 (0.6) with probability 0.699, and rate 2 (0.8) with probability 0.3001. This is an effective service rate of 0.6602. To realize this policy, a randomizing device can be employed; or the sole customer can be served at rate 0.6602, if this option is available. When there are from two to six customers in the system, service should occur at rate 0.8, and when there are seven or more customers in the system, service should occur at rate 0.9. Note that the buffer occupancy for service at maximum rate is to be inferred from the other entries. The row labeled $P(0)$ gives the probability that the system is empty under the optimal policy, and the row labeled $P(\text{ran})$ gives the proportion of time the system must randomize.

Under Scenario 2, the computations were performed for $N = 30$, 50, and 60. The only difference between $N = 30$ and 50 is that for $N = 30$, the value of the randomizing constant is accurate only to two decimal places. In both cases, we do not locate the buffer content for switching to highest service rate. For $N = 60$, we pick this up as occurring at buffer content of 26. This is not of much practical importance,

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

TABLE 2. Results for Section 6

	Scenario							
	1	2	3	4	5	6	7	8
p	0.4	—	—	0.7	0.8	—	0.85	0.9
Service	0.6			0.7	0.75		0.88	0.91
rates	0.8	—	—	0.8	0.9	—	0.95	0.95
	0.9			0.95	0.95		0.98	0.98
Costs	1		1	1	1	1	1	1
	5	—	7	10	15	15	5	15
	15		15	30	30	45	25	30
K_g	1.2			NA	NA		4.25	9.0
	0.6	—	—	2.1	1.6	—	1.275	1.8
	0.48			0.84	1.067		0.981	1.125
G_g	0.667		0.667	1.0	1.0	1.0	0.966	0.989
	2.5	—	3.5	8.75	13.333	13.333	4.474	14.211
	6.667		6.667	22.105	25.263	37.895	21.684	27.551
K	0.75	1.0	—	1.5	2.0	—	1.5	2.0
γ	5.500	1.495	2.310	10.163	6.166	7.674	5.188	7.723
Optimal	[1]	[1, 2] (1)	[1, 2] (1)	[1] (1)	[1] (1)	[1]	[1]	[1] (1)
policy	(0.699,1,2)	[3]	[3]	[2]	[2, 4] (2)	(0.801, 1, 2)	(0.630, 1, 2)	[2]
	[2. 6] (2)	(0.309, 1, 2)	(0.309, 1, 2)	(0.627, 2, 3)	[5]	[2, 8] (2)	[2, 13] (2)	(0.377, 1, 2)
		[4, 25] (2)	[4, 11] (2)		(0.467, 2, 3)			[3] (2)
K_θ	0.750	1.000	1.000	1.500	2.000	2.000	1.500	2.000
G_θ	1.675	0.887	1.007	10.284	10.070	10.255	3.306	9.440
$P(0)$	0.425	0.353	—	0.123	0.060	0.067	0.082	0.036
$P(\text{ran})$	0.429	0.063	—	0.334	0.030	0.339	0.510	0.309

since the probability of this many customers in the system under the optimal policy is exceedingly small, and, in fact, the probability of three or less in the system is 0.983. Scenario 3 is a slight modification of Scenario 2, in which the cost of the middle rate is increased. This computation was performed for $N = 60$.

Scenario 4 has an inexpensive lowest rate equal to the arrival rate. This computation was performed for $N = 50$ and 60. The crucial features of the model are obtained for $N = 30$, but the higher values of N are necessary to obtain the values of K_θ and G_θ accurate to three decimal places. For this scenario, the system will be in state 1 (and using the unstable rate) with probability 0.409.

Scenario 5 has an inexpensive lowest rate less than the arrival rate. This computation was performed for $N = 50$ and 60. The system will be in state 1 (and using the unstable rate) with probability 0.318. Scenario 6 increases the cost of fastest service. Note that the optimal policy changes quite a bit, indicating that optimality is sensitive to changes in the parameters. Now, the system randomizes when there is a single customer present, in order to lessen the probability of moving to a state in which the expensive rate must be employed. The buffer content for this has moved from 6 to 9. Note that G_θ has only increased by 1.8%, even though the cost of fastest service has increased by 50%.

The last two runs were made for $N = 60$.

7. EXAMPLE WITH QUALITY OF SERVICE CONSTRAINT

The system in this section is the same as in Section 6. However, the quantity we constrain is the proportion of customers who wait L or more slots before receiving service, where L is a fixed positive integer.

The assumption that a customer is registered during its arrival slot and joins the queue at the end of that slot still holds. If the customer enters service in the slot following its arrival slot, it has suffered no wait. Otherwise, the "wait clock" will register 1 at the beginning of the slot following the arrival slot and will continue to increase by 1 each slot until the customer enters service.

The program is written for $L = 3$ and we discuss the notation for this case. The general case will be clear. The state space consists of 4-tuples (x_1, x_2, x_3, y) plus a distinguished state *, which indicates that the system is empty. In a tuple state, the server is busy. The quantity x_i, $i = 1,2,3$, equals 1 if there is a customer who has been waiting exactly i slots, and 0 otherwise. The quantity $y \in \{0,1,2,\ldots\}$ is the number of customers that have been waiting four or more slots.

The D cost is $D(x_1, x_2, x_3, y) = x_3$. Note that a cost of 1 is incurred if and only if a customer has been waiting exactly three slots, but no further cost is incurred on this customer. Then, COP requires us to minimize the average cost of service subject to a constraint on the proportion of customers who wait three or more slots for service.

Let $q = 1 - p$. We have $P_{**} = q$ and $P_{*(0000)} = p$. Some examples of the other transition probabilities are as follows: If the system is currently in state (1003) and action a is chosen, then the next state will be (0102) (respectively, (1102), (0103),

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

(1103)) with probability aq (respectively, ap, $(1-a)q$, $(1-a)p$). If the system is currently in state (1010), then the next state will be (0100) (respectively, (1100), (0101), (1101)) with the same probabilities.

To define (Δ_N), let S_N be * together with the tuple states for which $y \leq N$. The system can escape S_N only if it is in a state $(x_1, x_2, 1, N)$ and there is no service completion. The system would transition to $(0, x_1, x_2, N+1)$ with probability $(1-a)q$, and this excess probability is given to state $(0, x_1, x_2, N)$. This is equivalent to assuming that one of the "old" customers is ejected from the system. Similarly, the excess probability $(1-a)p$ is given to state $(1, x_1, x_2, N)$.

PROPOSITION 7.1: *Let $a > p$ and let $g = g(a)$ be the stationary policy for Δ that always serves at rate a. Then, the probability that a customer waits at least L slots for service is*

$$P_g(W \geq L) = \frac{p}{a}\left(\frac{1-a}{1-p}\right)^L, \quad L \geq 1, \tag{7.1}$$

where W is the wait incurred.

PROOF: We outline the proof, which is a straightforward, but tedious, exercise. The steady-state probability π_i of i customers present in the system is given in [20, p. 182]. Let δ_i be the probability that an arriving customer finds i in the system at the beginning of the slot following its arrival slot, and let W_i be the wait of this customer conditioned on finding i. Then

$$P_g(W = k) = \sum_{i=0}^{\infty} P_g(W_i = k)\delta_i, \quad k \geq 0. \tag{7.2}$$

Clearly, $W_0 = 0$. For $i \geq 1$, the distribution of W_i is a negative binomial with parameters (i, a). Moreover, $\delta_i = (1-a)\pi_i + a\pi_{i+1}$. Substituting these quantities into (7.2) and simplifying yields

$$P_g(W = k) = \frac{p(a-p)}{a(1-p)}\left(\frac{1-a}{1-p}\right)^k, \quad k \geq 1. \tag{7.3}$$

Then, an easy calculation gives (7.1). ■

PROPOSITION 7.2: *If $p < a_M$ and $(p/a_M)[(1-a_M)/(1-p)]^L < K$, then Assumptions (C1)–(C5) hold.*

PROOF: The proof is similar to that of Proposition 6.1 and is omitted. ■

The program design is the same as in Section 6 with the content of the procedures changed to reflect the new constraint. Typically, γ is much larger and an initial guess of about $\lambda = 15$ is appropriate.

To understand the behavior of the optimal policy, consider the 16 states of the form (x_1, x_2, x_3, y), with $y \leq 1$. Assume that no new customers enter the system and that service is at constant rate a. For each state, consider the difference between the average D cost incurred to clear the system and the smallest cost, which occurs when

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

TABLE 3. Difference Reasoning

State	Average Penalty	Smallest Penalty	Difference
(1100)	$(1-a)(2+a)$	0	$(1-a)+(1-a^2)$
(1001)	$1-a^2$	0	$1-a^2$
(1010)	$2-a^2$	1	$1-a^2$
(0100)	$1-a$	0	$1-a$
(1000)	$(1-a)^2$	0	$(1-a)^2$

each service lasts exactly one slot. States with a positive difference are ranked in descending order of the magnitude of the difference in Table 3. The controller can make an immediate difference in incurred penalty cost by serving faster in these states, and this is typically what occurs. In the rest of the 16 states, the difference is 0. For example, for state (0,1,1,0), the penalty cost is 2 in each case. In these states, the controller can make no immediate difference by serving faster. If the controller is serving fast, we will see a tendency to back off in some of the 0 difference states. The insights gained from this discussion are referred to as the *difference reasoning*.

The results of various scenarios are displayed in Table 4. The entries in row 4 come from (7.1) with $L = 3$. Denote the three service rates (in increasing order) by 1, 2, and 3. All calculations except Scenario 3 were performed for $N = 50$ and 60. In all scenarios except Scenario 3, the system is stable under all three rates.

Scenario 1 is a lightly loaded system. Neither policy $g(1)$ nor $g(2)$ meets the constraint. Note that the value of G_θ seems reasonable. The optimal policy always serves at the middle rate (indicated by (2)) with the exceptions listed under E. Each exception can be explained using the difference reasoning. In state (0100), the controller randomizes between 0.8 and 0.9, for an effective rate of 0.813. The entries in row 11 are the probabilities under θ that the system is empty, contains one customer (being served), or contains a total of no more than five customers. The entry in row 12 is the probability that the system is in the state in which randomization of the rates must occur. Scenario 2 modifies Scenario 1 by weakening the constraint. Only $g(1)$ fails to meet the constraint. The optimal policy serves at rate 1 with the exceptions listed; these can be explained using the difference reasoning.

We discuss Scenario 3 in a moment. Scenario 4 is a moderately heavily loaded system in which only $g(3)$ meets the constraint. The optimal policy always serves at rate 3 with the exceptions listed. Each exception except (1000) has a difference of 0, as explained in the difference reasoning, making it reasonable to back off in these cases. State (1000) has the smallest positive difference and this is the state in which randomization occurs. By adding $P(0) + P(1) + P(\text{ran})$, we see that 94.8% of the time, the system is either empty, serving a single customer with empty buffer, or serving a customer with one customer waiting one slot. Scenario 5 is similar to Scenario 4.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

TABLE 4. Results for Section 7

	Scenario				
	1	2	3	4	5
p	0.4	—	0.7	0.75	0.85
Service rates	0.6		0.65	0.8	0.9
	0.8	—	0.8	0.9	0.95
	0.9		0.95	0.95	0.98
Costs	1		1	1	1
	5	—	5	5	10
	15		20	15	30
K_g	0.198		1.0	0.48	0.280
	0.019	—	0.259	0.053	0.033
	0.002		0.003	0.006	0.002
G_g	0.667		1.0	0.938	0.944
	2.5	—	4.375	4.167	8.947
	6.667		14.737	11.842	26.020
K	0.01	0.05	—	0.01	0.01
γ	66.203	8.538	21.673	172.723	547.030
Optimal policy	(2)	(1)	(3)	(3)	(3)
	E: (0000) (1)	E: (0100) (0.844, 1, 2)	E: (0000) (0.834, 1, 2)	E: (0000) (2)	E: (0000) (1)
	(0100) (0.868, 2, 3)	(1001) (2)	(0001) (2)	(0001) (2)	(0001) (2)
	(1001) (3)	(1010) (2)	(0010) (2)	(0010) (2)	(0010) (2)
	(1010) (3)	(1100) (2)	(0100) (2)	(1000) (0.183, 2, 3)	(1000) (0.427, 2, 3)
	(1100) (3)		(1000) (2)		
K_θ	0.010	0.050	0.050	0.010	0.010
G_0	1.507	0.856	6.150	6.085	9.338
$P(0)$	0.393	0.351	0.104	0.178	0.085
$P(1)$	0.437	0.390	0.361	0.592	0.535
$P(\leq 5)$	0.9999	0.9941	0.9996	0.9998	0.9999
$P(\text{ran})$	0.024	0.046	0.361	0.178	0.303

The program exhibits *spurious tail behavior*; that is, the computed policy tends to be incorrect for states with larger values of y. Given a fixed finite set of states, as N increases, the computed decisions in those states are guaranteed by Theorem 4.3 to converge to optimal decisions. Thus, we need not be concerned with these spurious values. To eliminate them for a fixed finite set of states, we may both increase N and decrease the tolerance within which $J(\lambda)$ is computed.

This tail behavior is particularly evident in Scenario 3, which is unstable under $g(1)$. The constraint is only met by $g(3)$, but rate 3 is expensive compared to the others. When the system is in state (0000) (i.e., serving a customer with none waiting), then service occurs at effective rate 0.675, an unstable rate. In all other cases, service utilizes a stable rate. The back-off states can be explained using the difference reasoning. In any example with an unstable rate, we need to be alert to the spurious tail behavior and take steps to compensate appropriately, by increasing N and decreasing the tolerance.

We mention a possibility for easing the computational burden. A smaller value of N and looser tolerance can be used to obtain an initial value for γ. The program may then utilize larger N and smaller tolerance to hone in on the optimal policy. Our major purpose has been to prove Theorem 4.3 and to confirm the computational feasibility of this method. Much work remains to be done to reduce the computational effort required to solve COP, especially in networks.

Acknowledgments

The author thanks a referee for encouraging a substantial reworking of an earlier version, and for a careful reading of the revision with thoughtful comments. She would also like to express her appreciation to the editor, Sheldon Ross, for his support of a revision and to Adam Shwartz for helpful comments and references.

References

1. Altman, E. (1999). *Constrained Markov decision processes.* Boca Raton, FL: Chapman & Hall/CRC Press.
2. Altman, E. & Shwartz, A. (1989). Optimal priority assignment: A time sharing approach. IEEE *Transactions on Automatic Control* AC-34: 1098–1102.
3. Altman, E. & Shwartz, A. (1993). Time-sharing policies for controlled Markov chains. *Operations Research* 41: 1116–1124.
4. Beutler, F.J. & Ross, K.W. (1985). Optimal policies for controlled Markov chains with a constraint. *Journal of Mathematical Analysis and Applications* 112: 236–252.
5. Denardo, E.V. (1982). *Dynamic programming, models and applications.* Englewood Cliffs, NJ: Prentice-Hall.
6. Derman, C. (1970). *Finite state Markovian decision processes.* New York: Academic.
7. Haviv, M. (1996). On constrained Markov decision processes. *Operations Research Letters* 19: 25–28.
8. Hordijk, A. & Kallenberg, L.C. (1984). Constrained undiscounted stochastic dynamic programming. *Mathematics of Operations Research* 9: 276–289.
9. Kallenberg, L.C. (1980). Linear programming and finite Markovian control problems. Unpublished Ph.D. dissertation, Leiden University, The Netherlands.
10. Koole, G.M. (1988). Stochastische Dynamische Programmering met Bijvoorwaarden. Unpublished Master's thesis, Leiden University, The Netherlands.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press

11. Ma, D.-J. & Makowski, A.M. (1988). A class of steering policies under a recurrence condition. Proceedings 27th Conference on Decision and Control. Austin, TX.
12. Ma, D.-J., Makowski, A.M., & Shwartz, A. (1990). Stochastic approximation for finite state Markov chains. *Stochastic Processes and Their Applications* 35: 27–45.
13. Makowski, A.M. & Shwartz, A. (1988). Implementation issues for Markov decision processes. In W. Fleming & P.L. Lions (eds.), *Stochastic differential systems, stochastic control theory and applications.* New York: Springer-Verlag, pp. 323–337.
14. Makowski, A.M. & Shwartz, A. (1993). On constrained optimization of the Klimov network and related Markov decision processes. *IEEE Transactions on Automatic Control* AC-38: 354–359.
15. Piunovskiy, A.B. (1997). *Optimal control of random sequences in problems with constraints.* Dordrecht: Kluwer.
16. Puterman, M.L. (1994). *Markov decision processes: Discrete dynamic programming.* New York: Wiley.
17. Ross, K.W. (1985). Constrained Markov decision processes with queueing applications. Unpublished Ph.D. dissertation, University of Michigan, Ann Arbor.
18. Ross, K.W. (1989). Randomized and past-dependent policies for Markov decision processes with multiple constraints. *Operations Research* 37: 474–477.
19. Sennott, L.I. (1993). Constrained average cost Markov decision chains. *Probability in the Engineering and Informational Sciences* 7: 69–83.
20. Sennott, L.I. (1999). *Stochastic dynamic programming and the control of queueing systems.* New York: Wiley.
21. Shwartz, A. & Makowski, A.M. (1986). An optimal adaptive scheme for two competing queues with constraints. In A. Bensoussan & J.L. Lions (eds.), *Analysis and optimization of systems.* New York: Springer-Verlag, pp. 515–532.
22. Spieksma, F.M. (1990). Geometrically ergodic Markov chains and the optimal control of queues. Unpublished Ph.D. dissertation, Leiden University, The Netherlands.
23. Webster, R. (1994). *Convexity.* Oxford: Oxford University Press.

https://doi.org/10.1017/S0269964801151089 Published online by Cambridge University Press