CHAPTER 13

MONOTONE OPTIMAL POLICIES FOR LEFT-SKIP-FREE MARKOV DECISION PROCESSES

Shaler Stidham Jr. and Richard R. Weber

13.1 Introduction

In a previous paper (Stidham and Weber [9]), we considered a variety of models for optimal control of the service rate in a queueing system, in which the objective is to minimize the limiting average expected cost per unit time. By standard techniques, we showed how to convert such a problem into an equivalent problem in which the objective is to minimize the expected total (undiscounted) cost until the first entrance into state zero. Under weak assumptions on the one-stage (service plus holding) costs and transition probabilities, we showed that an optimal policy is monotonic, that is, a larger service rate is used in larger states. In contrast to previous models in the literature on control of queues, we assumed that the holding cost was nondecreasing, but not necessarily convex, in the state. A common assumption in all the models was that services take place one at a time, so that the state transitions are *skip-free to the left*: a one-step transition from state i to a state $j < i - 1$ is impossible. Many queueing models have this property, including all birth–death models, as well as a variety of M/GI/1-type models, including models with batch arrivals, phase-type service times, and *LCFS-PR* queue discipline.

Julian Keilson introduced the concept of skip-free transitions in a Markov process in [2,3]. (He used the terms *skip-free in the negative (positive) direction* where we use *skip-free to the left (right)*.) In these pioneering works, he exploited the skip-free property in the context of descriptive models for both discrete-state and continuous-state Markov processes. Exploitation of the skip-free property in control models (that is, Markov decision processes) may be found, for example, in Wijngaard and Stidham [11], who developed efficient algorithms for calculating optimal policies in right-skip-free MDPs, in addition to the already-mentioned paper by Stidham and Weber [9], which focuses on left-skip-free MDPs.

The significance of left-skip-free transitions in the queueing control models of [9] is that the problem of optimally moving the system from state i to state 0 can be decomposed into two separate problems: first, to move the system optimally from state i to state $i - 1$, and then to move the system optimally from state $i - 1$ to state 0. This decomposition was exploited in two ways in [9] to prove monotonicity of an optimal policy: (1) to facilitate simple coupling or pairwise-switch arguments in the case of additive transitions; and (2) to construct a proof based on downward induction on the state variable in the case of nonadditive transitions.

In the present chapter, we extend the analysis in [9] to a general class of Markov decision processes with left-skip-free transitions. We focus on the problem in which the objective is to minimize the expected total cost until the first entrance into state zero. The equivalence to the minimum-average-cost problem follows by arguments similar to those in [9] and will be left to the reader. Our basic results are presented in the next section, followed by applications in Section 13.3.

13.2 The Model

Consider a discrete-time Markov decision process on the state space $S = \{0, 1, \ldots\}$. The action space is $A = [0, \bar{a}]$, where $\bar{a} < \infty$. When the process is in state $i \in S$ and action $a \in A$ is taken, the process makes a transition to state j with probability $p_{ij}(a)$. We assume that transitions are *left skip free*. That is, $p_{ij}(a) = 0$, $0 \le j < i - 1$, $a \in A$. There is a one-period cost $c(i, a)$. We assume that $c(i, a)$ is nonnegative and that $p_{ij}(a)$ and $c(i, a)$ are both continuous functions of $a \in A$ for each $0 \le i \le j + 1$. The objective is to minimize the expected total cost until the system first reaches state 0, from each starting state $i \ge 1$.

Let $v(i)$ denote the minimum expected total cost until the system first reaches state 0, starting from state $i, i \ge 1$. Then it follows from the general theory of Markov decision processes (cf. [7,6,1]) that v satisfies the following optimality equation ($i \ge 1$; $v(0) = 0$):

$$v(i) = \min_{a \in A}\left\{c(i, a) + \sum_{j=i-1}^{\infty} p_{ij}(a)v(j)\right\}. \tag{13.1}$$

Moreover, a stationary optimal policy may be constructed by choosing in state i ($i \ge 1$) an action $a \in A$, denoted $a(i)$, that achieves the minimum in (13.1). To resolve ties, we shall select the largest such action a.

Our goal is to find sufficient conditions under which this optimal policy is monotonic, that is, $a(i)$ is nondecreasing in $i \ge 1$. We shall find it convenient to work with an equivalent transformation of the optimality equation (13.1) that exploits the left-skip-free transition structure. To this end, let $z(j, i)$ denote the minimum expected total cost until the system first enters state i, starting from state $j, j > i \ge 0$. (Again, $z(j, i)$ is well defined, since the costs are non-negative.) Thus $z(j, 0) = v(j)$, and it follows from the left-skip-free transition structure of the system that

$$v(j) = z(j, k) + v(k), \quad j > k \ge 0. \tag{13.2}$$

(Because the transitions are left skip free, the system must visit each state $k < j$ on its way from state j to state 0. An optimal policy must therefore minimize both the cost to go from j to k and the cost to go from k to 0.) Now subtract $v(i - 1)$ from both sides of (13.1) and use (13.2) to obtain

$$z(i, i-1) = \min_{a \in A}\left\{c(i, a) + \sum_{j=i}^{\infty} p_{ij}(a)z(j, i-1)\right\},$$

which in turn is equivalent to

$$z(i, i-1) = \min_{a \in A}\left\{c(i, a) + \sum_{j=i}^{\infty} g_{ij}(a)z(j, j-1)\right\}, \tag{13.3}$$

where $g_{ij}(a) := \sum_{k=j}^{\infty} p_{ik}(a)$. Note that the optimal action $a(i)$ for sate i also attains the minimum on the right-hand side of (13.3).

Recall (see [9,10]) that a function $f(i, a)$ ($i \ge 0$, $a \in A$) is said to be *submodular* in (i, a) if $f(i, a_2) - f(i, a_1)$ is nonincreasing in $i \ge 0$, for each $a_2 \in A$, $a_1 \in A$, such that $a_2 > a_1$. The largest minimizer,

$$a(i) := \max\left\{a \in A : f(i, a(i)) = \min_{a \in A} f(i, a)\right\},$$

of such a function is nondecreasing in i. Thus, to show that $a(i)$ is nondecreasing in i, it suffices to show that the quantity in brackets on the right-hand side of (13.3),

$$J(i, a) := c(i, a) + \sum_{j=i}^{\infty} g_{ij}(a)z(j, j-1),$$

is submodular. Below we present a lemma which gives sufficient conditions for the function $J(i, a)$ to be submodular. We shall use the following conditions:

(C1) $c(i, a)$ is submodular in (i, a).
(C2) $g_{i,i+m}(a)$ is submodular in (i, a), for each $m \ge 0$.
(C3) $g_{i,i+m}(a)$ is non-increasing in $a \in A$ for each $i \ge 1$, $m \ge 0$.

Lemma 13.1. Assume Conditions (C1)–(C3) and suppose $z(k, k-1)$ is nonnegative and nondecreasing in $k \ge 1$. Then $J(i, a)$ is submodular in (i, a) and hence $a(i)$ is nondecreasing in $i \ge 1$.

Proof. Let $f(i, a) := \sum_{j=i}^{\infty} g_{ij}(a)z(j, j-1)$. Let $a_2 > a_1$. Then

$$\begin{aligned}
f(i, a_2) - f(i, a_1) &= \sum_{j=i}^{\infty}(g_{ij}(a_2) - g_{ij}(a_1))z(j, j-1) \\
&= \sum_{m=0}^{\infty}(g_{i,i+m}(a_2) - g_{i,i+m}(a_1))z(i+m, i+m-1) \\
&\ge \sum_{m=0}^{\infty}(g_{i+1,i+1+m}(a_2) - g_{i+1,i+1+m}(a_1))z(i+m, i+m-1) \\
&\ge \sum_{m=0}^{\infty}(g_{i+1,i+1+m}(a_2) - g_{i+1,i+1+m}(a_1))z(i+1+m, i+m) \\
&= f(i+1, a_2) - f(i+1, a_1).
\end{aligned}$$

The first inequality follows from (C2), since $z(i+m, i+m-1) \geq 0$. The second inequality follows from (C3), since $z(i+m, i+m-1) \leq z(i+1+m, i+m)$. Thus, $f(i, a)$ is submodular. Submodularity of $J(i, a)$ now follows form (C1), since the sum of submodular functions is submodular. ■

In what follows we shall assume that (C1)–(C3) hold and make further assumptions on $c(i, a)$ and $g_{ij}(a)$ that will enable us to prove that $z(j, j-1)$ is nondecreasing. (Nonnegativity of $z(j, j-1)$ follows immediately from the assumption that $c(i, a) \geq 0$ for all i, a.) It should come as no surprise that monotonicity of $c(i, a)$ in i will be a basic assumption.

(C4) $c(i, a)$ is nondecreasing in $i \geq 1$, for all $a \in A$.

Our first result is for the special case of additive transitions. In this case, a simple coupling argument establishes monotonicity of $z(j, j-1)$. To this end we introduce the following condition, which is a strengthening of (C2).

(C2′) There exists a function $q_m(a)$ such that $p_{i,i-1+m}(a) = q_m(a)$, independent of $i \geq 1$, for all $m \geq 0$, $a \in A$.

Additive transitions, of course, are characteristic of many models of queues and inventories, in which state transitions are due to inputs and/or outputs of, for example, customers, work to be done, or work in process.

Theorem 13.1. Assume (C1), (C2′), (C3), and (C4). Then $J(i, a)$ is submodular and hence a monotonic policy is optimal; that is, $a(i)$ is nondecreasing in $i \geq 1$.

Proof. For a stationary policy π, let $z^\pi(i, i-1)$ denote the expected total cost to go from state i to state $i-1$ following policy π. Then $z(i, i-1) = \min_\pi z^\pi(i, i-1)$. Now consider an arbitrary $i \geq 1$ and suppose π is optimal. That is, π takes action $a(j)$ whenever the process is in state j, $j \geq 1$. Let π' be a stationary policy that takes action $a(j+1)$ whenever the process is in state j, $j \geq i$. It follows from (C2′) that we can couple each sample path starting from state $i+1$, following π, with a corresponding sample path starting from state i, following π', in such a way that the former sample path is in state $j+1$ if and only if the latter sample path is in state j ($j \geq i$). Since π takes the same action in each state $j+1$ (namely, $a = a(j+1)$) that π' takes in state j, and $c(j+1, a) \geq c(j, a)$ by (C4), it follows that the total cost along the former sample path is at least as large as the total cost along the latter sample path, and hence $z^\pi(i+1, i) \geq z^{\pi'}(i, i-1)$. Therefore, we conclude that

$$z(i+1, i) = z^\pi(i+1, i) \geq z^{\pi'}(i, i-1) \geq z(i, i-1),$$

thus establishing that $z(i, i-1)$ is nondecreasing in $i \geq 1$. Since (C2′) implies (C2), the theorem now follows from (C1)–(C3) and Lemma 13.1. ■

Now we turn our attention to the more general case of nonadditive transitions. Here we shall establish monotonicity of $z(i, i-1)$ by a downward induction on i.

To this end, we shall find it convenient to work with another transformed optimality equation that is equivalent to (13.1) and (13.3).

First observe that $z(i, i-1)$, $i \geq 1$, satisfies (13.3) if and only if, for all $a \in A$,

$$p_{i,i-1}(a) z(i, i-1) \leq c(i, a) + \sum_{j=i+1}^{\infty} g_{ij}(a) z(j, j-1),$$

or equivalently,

$$z(i, i-1) \leq \left[c(i, a) + \sum_{j=i+1}^{\infty} g_{ij}(a) z(j, j-1) \right] \Big/ p_{i,i-1}(a),$$

with equality for a value of $a \in A$ that achieves the minimum in (13.3). (Here we interpret the right-hand side of the inequality as being equal to $+\infty$ if $p_{i,i-1}(a) = 0$.) Thus (13.3) is equivalent to the following optimality equation:

$$z(i, i-1) = \min_{a \in A} \left\{ \tilde{c}(i, a) + \sum_{j=i+1}^{\infty} \tilde{g}_{ij}(a) z(j, j-1) \right\}, \tag{13.4}$$

where $\tilde{c}(i, a) := c(i, a)/p_{i,i-1}(a)$ and $\tilde{g}_{ij}(a) := g_{ij}(a)/p_{i,i-1}(a)$. Once again, the optimal action $a(i)$ for state i also attains the minimum on the right-hand side of (13.3). Note that, in contrast to (13.3), (13.4) expresses $z(i, i-1)$ recursively in terms of $z(i+1, i), z(i+2, i+1), \ldots$, which makes it ideally suited for a proof by downward induction that $z(i, i-1)$ is nondecreasing in i.

We shall use the following condition:

(C5) There exists a nonnegative function $b(j)$, $j \geq 1$, such that

$$\tilde{c}(i, a) - \tilde{c}(i-1, a) \geq \sum_{j=i+1}^{\infty} (\tilde{g}_{i-1,j-1}(a) - \tilde{g}_{ij}(a))^+ b(j), \quad \text{for all } i \geq 2,\ a \in A.$$

Remark. By itself, (C5) is innocuous. It is satisfied, for example, by $g(j) \equiv 0$ when $\tilde{c}(i, a) - \tilde{c}(i-1, a) \geq 0$. In our applications, however, we shall also require that $b(j)$ be an upper bound for $z(j, j-1)$ (cf. Lemma 13.2) below), in which case (C5) becomes nontrivial. In that context, it is the key to our proof that $z(j, j-1)$ is monotonic. It may thus be viewed as a weak sufficient condition for a generalization of stochastic monotonicity of the transitions, since $z(j, j-1)$ is a (cost) generalization of the mean first passage time from state j to $j-1$.

The following lemma will form the basis for the inductive step in our proof that $z(i, i-1)$ is nondecreasing in i.

Lemma 13.2. Assume (C5). Let $i \geq 2$ be given and suppose $0 \leq z(j, j-1) \leq b(j)$ and $z(j+1, j) \geq z(j, j-1)$, for all $j \geq i$. Then $z(i, i-1) \geq z(i-1, i-2)$.

Proof. Since $0 \leq z(j, j-1) \leq b(j)$, (C5) implies that

$$
\begin{aligned}
\tilde{c}(i,a) - \tilde{c}(i-1,a) &\geq \sum_{j=i+1}^{\infty} (\tilde{g}_{i-1,j-1}(a) - \tilde{g}_{ij}(a))' z(j,j-1) \\
&\geq \sum_{j=i+1}^{\infty} (\tilde{g}_{i-1,j-1}(a) - \tilde{g}_{ij}(a)) z(j,j-1).
\end{aligned}
$$

for all $a \in A$. Setting $a = a(i)$ and using the fact that $z(j, j-1) \geq z(j-1, j-2)$ for all $j \geq i+1$, it follows from (13.4) that

$$
\begin{aligned}
z(i,i-1) &= \tilde{c}(i,a) + \sum_{j=i+1}^{\infty} \tilde{g}_{ij}(a)) z(j,j-1) \\
&\geq \tilde{c}(i-1,a) + \sum_{j=i+1}^{\infty} \tilde{g}_{i-1,j-1}(a) z(j,j-1) \\
&\geq \tilde{c}(i-1,a) + \sum_{j=i+1}^{\infty} \tilde{g}_{i-1,j-1}(a) z(j-1,j-2) \\
&= \tilde{c}(i-1,a) + \sum_{j=i}^{\infty} \tilde{g}_{i-1,j}(a) z(j,j-1) \\
&\geq z(i-1,i-2).
\end{aligned}
$$

thus establishing the desired result. ■

To complete the inductive proof, we need to accomplish two tasks:

1. find a function $b(\cdot)$ satisfying (C5) such that $b(j) \geq z(j, j-1)$, $j \geq 1$; and
2. find a state $i = k$ at which to start the downward induction.

To be useful in applications, the upper-bounding function $b(j)$ must be easy to calculate a priori and be as tight as possible. Similarly, the starting state k for the downward induction must have the property that we can establish a priori that $z(j+1, j) \geq z(j, j-1)$ for all $j \geq k$. To a large extent, the methods for accomplishing these tasks tend to be specific to the application. We shall give illustrative examples for problems with certain types of specific structure, focusing on processes with *asymptotically additive* transitions:

(C6) There exists a function $q_m(a)$, $m \geq 0$, and a positive integer k such that, for all $i \geq 1$, $a \in A$, $p_{ij}(a) = q_{j-i+1}(a)$, for all $j \geq \max\{i-1, k\}$.

Note that under (C6), $p_{i,i+m}(a)$, $m = -1, 0, 1, 2, \ldots$, is independent of i, for $i \geq k$. Note also that (C6) implies that $g_{ij}(a) = \Sigma_{m=j-i+1}^{\infty} q_m(a)$, for $j \geq \max\{i-1, k\}$.

Using the same coupling argument as in the proof of Theorem 13.1, we can prove the following lemma.

Lemma 13.3. Assume (C4) and (C6). Then $z(j+1, j) \geq z(j, j-1)$ for $j \geq k$.

Thus, when (C4) and (C6) hold, we can start the downward induction at $i = k$. It remains to find a function $b(j)$ such that $z(j, j-1) \leq b(j)$, $j \leq k$. For polynomially bounded one-step cost functions, this can be done utilizing the asymptotically additive transition structure. To keep the exposition simple, we shall present this approach in detail for the case of a linearly bounded $c(i, a)$, and then sketch the extension to polynomially bounded $c(i, a)$.

(C7) There exist nonnegative constants c and h such that $c(i, a) \leq c + h \cdot i$, for all $i \geq 1$, $a \in A$.

First note that $z(k, k-1) \leq z^{\bar{\pi}}(k, k-1)$, when $\bar{\pi}$ is a policy that uses the maximal action $\bar{a}$ in every state $j \geq k$. Let

$$q_m := q_m(\bar{a}),\ m \geq 0, \quad \alpha := \sum_{m=0}^{\infty} m q_m, \quad \beta := \sum_{m=0}^{\infty} m^2 q_m, \tag{13.5}$$

$$d := \alpha + \frac{\beta - \alpha}{2(1-\alpha)}. \tag{13.6}$$

Define the function $b(j)$, $j \geq k$, by

$$b(j) := \frac{c + h \cdot (d + j - 1)}{1 - \alpha}. \tag{13.7}$$

Lemma 13.4. Assume (C6) and (C7). Suppose $\alpha < 1$ and $\beta < \infty$. Then $z(j, j-1) \leq b(j) < \infty$, for all $j \geq k$.

Proof. Since $z(j, j-1) \leq z^{\bar{\pi}}(j, j-1)$, it suffices to show that $z^{\bar{\pi}}(j, j-1) \leq b(j)$ for all $j \geq k$. To this end, we treat $j - 1$ as a costless absorbing state and note that (C6) implies that, starting in state $j \geq k$ and following policy $\bar{\pi}$, the state of the system evolves as a left-skip-free *DTMC*, $\{X_n, n \geq 0\}$, with transition probabilities

$$p_{il} = q_{l+1-i},\ l \geq i-1, \tag{13.8}$$

as long as $i \geq j$. It follows from (13.8) that

$$X_{n+1} = X_n - 1 + Y_n, \tag{13.9}$$

where Y_n is independent of X_n, with $P\{Y_n = m\} = q_m$, as long as $X_n \geq j$. This recursion is identical to that for an M/GI/1 queue, observed at service-completion epochs, in which Y_n represents the number of arrivals during the nth service time.

Now suppose we incur a one-stage cost $c + h \cdot i$ whenever $\{X_n, n \geq 0\}$ visits state $i \geq j$. For obvious reasons, we shall refer to h as the per-unit *holding cost*. Let $c(j, j-1)$ denote the total cost incurred until X_n enters the absorbing state $j - 1$, given that it starts in state j. Then (C7) implies that $z^{\bar{\pi}}(j, j-1) \leq c(j, j-1)$. Now

$$c(j, j-1) = c \cdot t(j, j-1) + h(j, j-1),$$

where $t(j, j-1)$ and $h(j, j-1)$ are, respectively, the expected number of transitions and the expected total holding cost incurred until the first entrance into the absorbing state $j-1$, starting from state j. It follows from classical results for this system (cf., e.g., [4]) that

$$t(j, j-1) = (1-\alpha)^{-1}, \; h(j, j-1) = (1-\alpha)^{-1} h \cdot (d + j - 1),$$

which implies that $c(j, j-1) = b(j)$. This completes the proof. ■

We are now ready to state our main result. For convenient reference we restate (C1)–(C7) here.

(C1) $c(i, a)$ is submodular in (i, a).
(C2) $g_{i,i+m}(a)$ is submodular in (i, a), for each $m \geq 0$.
(C3) $g_{i,i+m}(a)$ is nonincreasing in $a \in A$ for each $i \geq 1$, $m \geq 0$.
(C4) $c(i, a)$ is nondecreasing in $i \geq 1$, for all $a \in A$.
(C5) There exists a nonnegative function $b(j)$, $j \geq 1$, such that

$$\tilde{c}(i, a) - \tilde{c}(i-1, a) \geq \sum_{j=i+1}^{\infty} (\tilde{g}_{i-1,j-1}(a) - \tilde{g}_{ij}(a))^+ b(j), \quad \text{for all } i \geq 2, a \in A.$$

(C6) There exists a function $q_m(a)$, $m \geq 0$, and a positive integer k such that, for all $i \geq 1$, $a \in A$, $p_{ij}(a) = q_{j-i+1}(a)$, for all $j \geq \max\{i-1, k\}$.
(C7) There exist nonnegative constants c and h such that $c(i, a) \leq c + h \cdot i$, for all $i \geq 1$, $a \in A$.

Theorem 13.2. Assume (C1)–(C7), with $\alpha < 1$, $\beta < \infty$, and $b(j)$ defined by (13.7) for $j \geq k$, and $b(j) := b(k)$ for $j \leq k$. Then a monotonic policy is optimal, i.e., $a(i)$ is nondecreasing in $i \geq 1$.

Proof. It follows from Lemma 13.3 that $z(j+1, j) \geq z(j, j-1)$ for all $j \geq k$. Lemma 13.4 implies that

$$z(k, k-1) \leq b(k) = \frac{c + h \cdot (d + k - 1)}{1 - \alpha}.$$

Now let $i \leq k$ and suppose (as an induction hypothesis) that $z(j, j-1) \leq b(j)$ and $z(j+1, j) \geq z(j, j-1)$ for all $j \geq i$. Then Lemma 13.2 implies that $z(i, i-1) \geq z(i-1, i-2)$, which in turn implies that $z(i-1, i-2) \leq b(i) = b(k) = b(i-1)$, thus completing the inductive step. It follows by downward induction on i that $z(i+1, i) \geq z(i, i-1)$ for all $i \geq 1$. Monotonicity of an optimal policy is then a consequence of (C1)–(C3) and Lemma 13.1. ■

Now suppose the one-stage cost is polynomially bounded:

(C7′) There exist nonnegative constants c and h and a positive integer n such that $c(i, a) \leq c + h \cdot i^n$, for all $i \geq 1$, $a \in A$.

The proof of Lemma 13.4 can be modified to derive an appropriate bounding function $b(j)$ in this case, again using the M/GI/1-type recursion (13.9). Of course, we still have $t(j, j-1) = (1-\alpha)^{-1}$, but now the expression for $h(j, j-1)$ is

an nth-order polynomial in j, which involves the first n moments of the steady-state distribution of $\{X_n, n \geq 0\}$, and hence the first $n+1$ moments of $\{q_m, m \geq 0\}$, all of which must therefore be assumed finite. We leave the details to the reader.

13.3 Special Cases and Applications

In this section we show how (C1)–(C7) can be verified in special cases and give some applications to control of queues.

13.3.1 Control of Downward Jumps

Suppose $p_{i,i-1}(a) = a$. That is, the control variable in each state is itself the probability of a downward jump. Then (C5) simplifies considerably. First note that in this case the inequality in (C5) holds if and only if

$$c(i, a) - c(i-1, a) \geq \sum_{j=i+1}^{\infty} (g_{i-1,j-1}(a) - g_{ij}(a))^+ b(j).$$

Now suppose (C1) and (C2) hold, so that

$$c(i, a) - c(i-1, a) \geq c(i, \bar{a}) - c(i-1, \bar{a}),$$

$$g_{i-1,j-1}(a) - g_{ij}(a) \leq g_{i-1,j-1}(\bar{a}) - g_{ij}(\bar{a}),$$

for all $1 \leq i < j$, $a \in A$. Thus we can replace (C5) in Theorem 13.2 with the following simpler condition:

(C5′) There exists a nonnegative function $b(j)$, $(j) \geq 1$, such that

$$c(i, \bar{a}) - c(i-1, \bar{a}) \geq \sum_{j=i+1}^{\infty} (g_{i-1,j-1}(\bar{a}) - g_{ij}(\bar{a}))^+ b(j), \quad \text{for all } i \geq 2.$$

13.3.2 An Example with Additive Transitions

(This example is taken from Stidham and Weber [9].) Consider an M/GI/1 queue with batch arrivals and nonpreemptive discipline. Batches of jobs arrive according to a Poisson process; the sizes of successive batches are i.i.d. random variables. At each service completion, if the queue is not empty, a job is removed from the queue and placed in service. We choose the service-time distribution for this job from a family of distributions indexed by the service rate μ chosen from the compact set $A = [0, \bar{\mu}]$. Services cannot be interrupted. Service times of successive jobs, conditional on their indices, are independent. Let $S(\mu)$ denote a generic service time from the distribution indexed by μ. We assume that $\mu > \mu'$ implies that $S(\mu')$ is stochastically smaller than $S(\mu')$. There is a cost rate $c(\mu)$ incurred while service rate μ is in effect and a holding cost incurred at rate $h(i)$ while i customers are in the system, where $c(\cdot)$ is nonnegative, nondecreasing, and continuous on A, and $h(\cdot)$ is nonnegative and nondecreasing in $i \geq 1$.

We observe the system at the beginning of each service. The one-stage cost function is given by

$$c(i, \mu) = \frac{c(\mu)}{\mu} + f(i; \mu),\ i \geq 1,\ \mu \in A,$$

and the transition probabilities by

$$p_{ij}(\mu) = P\{i + K(\mu) - 1 = j\},\ 1 \leq i < j + 1,\ \mu \in A,$$

where $f(i; \mu)$ is the expected total holding cost incurred during the service time $S(\mu)$ that begins with i jobs in the system, and $K(\mu)$ is the number of arrivals during the service time $S(\mu)$. The objective is to minimize the expected total cost until the first entrance into state 0, from each starting state $i \geq 1$. One can readily verify that (C1), (C2′), (C3), and (C4) are satisfied, so that Theorem 13.1 applies and hence a monotonic policy is optimal.

13.3.3 An M/M/s Queue with Batch Arrivals and Controllable Service Rate

In this example, batches of customers arrive according to a Poisson process with rate λ. Successive batch sizes are i.i.d. and distributed as a nonnegative integer-valued random variable, M, with probability mass function $p(m) = P\{M = m\}$, $m \geq 1$. There are s memoryless servers, each with mean service rate a, fed by a single queue. The control variable (action) is the service rate a, to be chosen from the interval $A = [0, \bar{a}]$, where $\bar{a} < \infty$. There is a service cost $c(a)$ per unit time while the service rate is a. We assume that $c(\cdot)$ is nonnegative, nondecreasing, and continuous on A. There is a holding cost, which is incurred at rate $h(i)$ while there are i customers in the system. We assume that $h(i)$ is nonnegative and nondecreasing in $i \geq 1$. The objective is to minimize the expected total cost until the first entrance into state 0, from each starting state $i \geq 1$.

We use uniformization (cf. [5,8]) to formulate the problem as a Markov decision process. The system is observed at the events of a Poisson process with mean rate $\lambda + s\bar{a}$. The expected cost until the next observation point, given that the current state is i and the current control is a, is given by

$$c(i, a) = \frac{c(a) + h(i)}{\lambda + s\bar{a}}.$$

The imbedded transition probabilities are

$$p_{i,i-1}(a) = \frac{(i \wedge s)a}{\lambda + s\bar{a}},$$

$$p_{i,i+m}(a) = \frac{\lambda p(m)}{\lambda + s\bar{a}}, \quad m \geq 1,$$

$$p_{i,i}(a) = 1 - p_{i,i-1}(a) - \sum_{m=1}^{\infty} p_{i,i+m}(a) = \frac{s\bar{a} - (i \wedge s)a}{\lambda + s\bar{a}}.$$

To simplify these expressions, assume (without loss of generality) that the time unit has been chosen so that $\lambda + s\bar{a} = 1$. Then we have

$$c(i, a) = c(a) + h(i),$$

$$g_{i,i+m}(a) = \lambda \sum_{k=m}^{\infty} p(m), \quad m \geq 1,$$

$$g_{i,i}(a) = 1 - p_{i,i-1}(a) = 1 - (i \wedge s)a.$$

It is easy to see that (C1)–(C4) are satisfied. Condition (C6) is satisfied with $k = s$, and

$$q_0(a) = s \cdot a,$$

$$q_1(a) = s(\bar{a} - a),$$

$$q_m(a) = \lambda p(m - 1), \quad m \geq 2.$$

Now suppose (C7) holds. With q_m, $m \geq 0$, α, and β defined by (13.5), d defined by (13.6), and $b(j)$ defined by (13.7), first note that $\alpha < 1$ if and only if

$$\rho := \frac{\lambda E[M]}{s\bar{a}} < 1.$$

Now let us examine (C5). For $i > s$, (C5) reduces to

$$\frac{c(a) + h(i)}{sa} - \frac{c(a) + h(i-1)}{sa} \geq \sum_{j=i+1}^{\infty} \left(\frac{\lambda p(j-i)}{sa} - \frac{\lambda p(j-i)}{sa} \right)' b(j) = 0,$$

which holds since $h(i) \geq h(i - 1)$. For $i \leq s$, (C5) holds if and only if

$$h(i) - h(i-1) \geq \frac{c(\bar{a}) + h(i) + \lambda f(i)}{i},$$

where $f(i) := \Sigma_{m=1}^{\infty} p(m) b(i + m)$.

Acknowledgments

The research of S.S. was supported by the National Science Foundation under Grant No. NCR-9406823. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Bertsekas, D. *Dynamic Programming and Optimal Control*, Vol. II. Athena Scientific, Belmont, MA, 1995.

[2] Keilson, J. The use of Green's functions in the study of bounded random walks with applications to queuing theory. *J. Math. Phys.* **41**, 42–52, 1962.

[3] Keilson, J. *Green's Function Methods in Probability Theory*. Griffin, London, 1965.

[4] Kulkarni, V. G. *Modeling and Analysis of Stochastic Systems*. Chapman-Hall, London, 1995.

[5] Lippman, S. A. Applying a new device in the optimization of exponential queuing systems. *Oper. Res.* **23**, 687–710, 1975.

[6] Puterman, M. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York, 1994.

[7] Schäl, M. Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. *Z. Wahrscheinlichkeitstheorie verw. Gerb.* **32**, 179–196, 1975.

[8] Serfozo, R. An equivalence between continuous and discrete-time markov decision processes. *Oper. Res.* **27**, 616–620, 1979.

[9] Stidham, S. Jr., and Weber, R. Monotonic and insensitive optimal policies for control of queues with undiscounted costs. *Oper. Res.* **87**, 611–625, 1989.

[10] Topkis, D. Minimizing a submodular function on a lattice. *Oper. Res.* **26**, 305–321, 1978.

[11] Wijngaard, J., and Stidham, S. Jr. Forward recursion for Markov decision processes with skip-free-to-the-right transitions, Part i: Theory and algorithms. *Math. Oper. Res.* **11**, 295–308, 1986.

CHAPTER 14

OPTIMAL ROUTING CONTROL IN RETRIAL QUEUES

H. M. Liang and V. G. Kulkarni

Dedication. One of the earliest papers in retrial queues is by Keilson, Cozzolino, and Young [6]. Retrials queues has grown into an important area of research over the last decade, as evidenced by the survey papers by Yang and Templeton [11], Falin [3], and Kulkarni and Liang [7]. However, as far as the authors are aware, there are no results on the control of retrial queues. In this chapter, we try to fill this gap.

14.1 Introduction

A single server retrial queue consists of a primary queue, an orbit, and a server serving the primary queue. Customers can arrive at the primary queue either from outside the system or from the orbit. If an arriving customer is blocked from entering the primary queue, he joins the orbit and conducts a retrial later. Otherwise, he enters the primary queue, waits for service, and leaves the system after being served. The main motivation for this model arises from the phenomenon of retrials in telephone and telecommunication systems.

In this chapter, we study dynamic routing control of the retrial queue. A controlled retrial queueing system consists of a system controller, a primary service facility, and an orbit (see Figure 14.1, where the system controller is represented by a circle with a question mark). Customers can arrive at the system controller either from outside the system (according to a Poisson process) or from the orbit (according to a rate that depends upon the number of customers in the orbit). The system controller decides whether to route the customer to the primary queue or to the orbit, based on the state of the system. All customers are admitted to the system, and no customer can leave the system without receiving service in the primary queue. The capacities in the primary queue and in the orbit are both infinite. A holding cost $h(i, j)$ is incurred per unit time whenever there are i customers in the primary queue and j customers in the orbit. Our goal is to characterize the optimal routing policy that minimizes the expected total discounted cost over an infinite horizon.

(We would like to point out the fact that we study the socially optimal policies here, i.e., the cost to the system is minimized. The problem of computing individu-