J. Appl. Prob. **43,** 363–376 (2006)
Printed in Israel

ON THE VALUE FUNCTION OF THE M/Cox(r)/1 QUEUE

SANDJAI BHULAI,* *Vrije Universiteit Amsterdam*

Abstract

We consider a single-server queueing system at which customers arrive according to a Poisson process. The service times of the customers are independent and follow a Coxian distribution of order r. The system is subject to costs per unit time for holding a customer in the system. We give a closed-form expression for the average cost and the corresponding value function. The result can be used to derive nearly optimal policies in controlled queueing systems in which the service times are not necessarily Markovian, by performing a single step of policy iteration. We illustrate this in the model where a controller has to route to several single-server queues. Numerical experiments show that the improved policy has a close-to-optimal value.

Keywords: Coxian distribution; M/Cox(r)/1 queue; one-step policy improvement; optimal routeing; Poisson equation; value function

2000 Mathematics Subject Classification: Primary 93E25
Secondary 60K25

1. Introduction

The application of Markov decision theory to the control of queueing systems is to a large extent hindered by two things. First, in most practical situations the state space of the Markov decision process is enormous, and renders it almost impossible to derive optimal policies with standard techniques and algorithms. This phenomenon is known as 'the curse of dimensionality'. Second, handling non-Markovian systems within the framework of Markov decision theory is not straightforward. Consequently, there is clear motivation to develop approximation methods to avoid these problems.

Two approximation methods that deal with the first aspect of the problem are one-step policy improvement and reinforcement learning. One-step policy improvement dates back to Norman [13]. The method requires an explicit expression of the value function, which can be obtained by solving the dynamic programming optimality equations for a fixed policy. The result is then used in one step of the policy iteration algorithm from Markov decision theory, to obtain an improved policy. The method has been successfully applied to derive nearly optimal state-dependent policies in, e.g. [14], [16], and [20]. In reinforcement learning the value function is approximated based on a certain functional form. However, choosing the initial functional form such that the approximations are good is difficult. Successful applications of this method to queueing systems include [10], [12], and [19]. It is clear that both one-step policy improvement and reinforcement learning require insight into the structure of the value function. Systematic studies of the structure of value functions for specific queueing models can be found in [2] and [9].

Received 3 November 2003; revision received 27 February 2006.
* Postal address: Vrije Universiteit Amsterdam, Faculty of Sciences, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands. Email address: sbhulai@few.vu.nl
Supported by the Netherlands Organization for Scientific Research (NWO).

The second aspect of the problem can be dealt with by using Markovian approximations to non-Markovian systems. A useful tool in this respect is the Coxian distribution. Coxian distributions have the important feature that they are dense in the class of all nonnegative distributions (see, e.g. [17]). This property makes them very useful in approximating general nonnegative distributions. Note that non-Markovian systems can be formulated in a Markov decision-theoretic framework by finding an embedded Markov chain (see, e.g. [8]). This formulation makes the dynamic programming optimality equations more difficult to solve, however, since the cost structure has to be altered accordingly. The Coxian distribution does not change the cost structure, but does add supplementary variables to the state space, increasing the dimensionality.

In this paper we study the M/Cox(r)/1 queue as a first step toward the application of the above-mentioned ideas. The M/Cox(r)/1 queue can be described as a single-server queue at which customers arrive according to a Poisson process, and where the service durations of the customers are independent and follow a Coxian distribution of order r. Additionally, the system is subject to costs per unit time for holding a customer in the system. The main result of this paper is the explicit solution to the dynamic programming optimality equations, known as the Poisson equations. This result is derived in Section 2 and some important special cases of it are listed in Section 3.

The main contribution of our result with respect to reinforcement learning is the structure of the value function. It provides theoretical insight into approximation choices for the functional equation. Furthermore, the result can be directly used in the one-step policy improvement method. This will be illustrated in Section 4, by means of a routeing problem to parallel queues. We numerically show that the improved policy has a close-to-optimal value.

2. The M/Cox(r)/1 queue

Consider a single-server queueing system at which customers arrive according to a Poisson process with parameter λ. The service times are independent, identically distributed, and follow a Coxian distribution of order r. Thus, the service of a customer can last for up to r exponential phases. The mean duration of phase i is μ_i, for $i = 1, \dots, r$. The service starts at phase 1. After phase i the service ends with probability $1 - p_i$ or enters phase $i+1$ with probability p_i, for $i = 1, \dots, r-1$. The service is completed with certainty after phase r, if not completed at an earlier phase. We assume that $p_i > 0$ for $i = 1, \dots, r-1$, to avoid trivial situations. Let

$$\mathcal{S} = \{(0,0)\} \cup (\mathbb{N} \times \{0, \dots, r-1\})$$

denote the state space, where, for $(x, y) \in \mathcal{S}$, the component x represents the number of customers in the system and y the number of completed phases of the service process.

Assume that the system is subject to costs for holding a customer in the system. Without loss of generality we assume that unit costs are incurred per unit time for holding a customer in the system. Let $u_t(x)$ denote the total expected cost up to time t when the system starts in state x. Let $\gamma(i) = \prod_{k=1}^{i} p_k$ for $i = 0, \dots, r-1$, with the convention that $\gamma(0) = 1$. Note that the Markov chain satisfies the unichain condition. Assume that the stability condition

$$\sum_{k=1}^{r} \gamma(k-1) \frac{\lambda}{\mu_k} = \sum_{k=1}^{r} \prod_{l=1}^{k-1} p_l \frac{\lambda}{\mu_k} < 1 \tag{1}$$

holds, such that, consequently, the average cost,

$$g = \lim_{t \to \infty} u_t(x)/t,$$

is independent of the initial state x (due to Proposition 8.2.1 of [15, p. 337]). Therefore, the dynamic programming optimality equations for the M/Cox(r)/1 queue are given by

$$\begin{aligned} g + \lambda V(0,0) &= \lambda V(1,0), \\ g + (\lambda + \mu_i) V(x, i-1) &= \lambda V(x+1, i-1) + p_i \mu_i V(x, i) \\ &\quad + (1-p_i)\mu_i V(x-1, 0) + x, \qquad i = 1, \dots, r-1, \\ g + (\lambda + \mu_r) V(x, r-1) &= \lambda V(x+1, r-1) + \mu_r V(x-1, 0) + x. \end{aligned}$$

In this set of equations, known as the Poisson equations, the function $V(x, y)$ is called the (relative) value function. This function has the interpretation of the asymptotic difference in total costs that results from starting the process in state (x, y) instead of some reference state. Without loss of generality we take the reference state to be $(0, 0)$. The value function will be the main subject of interest, since it plays a central role in deriving (nearly) optimal policies in control problems. In Section 4 we illustrate this for the problem of routeing to parallel queues.

When the state space is not finite, as is the case in our model, it is known that there are many pairs of g and V that satisfy the Poisson equations (see, e.g. [3]). There is only one pair that is the correct solution, however, and we refer to this as the unique solution. Finding this pair involves constructing a weighted norm such that the Markov chain is geometrically recurrent with respect to that norm. This weighted norm imposes extra conditions on the solution to the Poisson equations, such that their unique solution can be obtained. Solving the Poisson equations therefore requires two steps: first we have to find an expression that satisfies the Poisson equations (existence), and then we have to show that it is the unique solution (uniqueness). We separate these two steps in the proof of the following theorem.

Theorem 1. *Let $\gamma(i) = \prod_{k=1}^{i} p_k$ for $i = 0, \dots, r-1$, with the convention that $\gamma(0) = 1$. Define*

$$\alpha = \frac{\sum_{k=1}^{r} \gamma(k-1)/\mu_k}{1 - \sum_{k=1}^{r} \gamma(k-1)\lambda/\mu_k}, \qquad a_0 = \sum_{k=1}^{r} \frac{1-\gamma(k-1)}{\mu_k}\lambda\alpha - \sum_{k=1}^{r}\sum_{l=1}^{k-1} \frac{\gamma(l-1)}{\mu_k \mu_l}\lambda(1+\lambda\alpha).$$

The solution to the Poisson equations is given by the average cost $g = \lambda(\alpha + a_0)$ and the corresponding value function

$$\begin{aligned} V(x,y) &= \alpha\frac{x(x+1)}{2} + \left[a_0 + \left[\frac{1}{\gamma(y)} - 1\right]\alpha - \sum_{k=1}^{y} \frac{\gamma(k-1)}{\gamma(y)}\frac{1+\lambda\alpha}{\mu_k}\right]x \\ &\quad - \left[a_0 + \sum_{k=y+1}^{r} \frac{\gamma(k-1)}{\gamma(y)}\frac{\lambda}{\mu_k}\left(\sum_{l=1}^{k-1}\frac{\gamma(l-1)}{\gamma(k-1)}\frac{1+\lambda\alpha}{\mu_l} - \left[\frac{1}{\gamma(k-1)} - 1\right]\alpha\right)\right], \qquad (2) \end{aligned}$$

for $(x, y) \in \mathcal{S}$.

Before we prove Theorem 1, we first provide some insight into the structure of $V(x, y)$. The following argument shows that we could have guessed $V(x, y)$ to be quadratic in the number of customers, x, in the queue. Assume, without loss of generality, that the service discipline is last-come–first-served. Let κ_y denote the waiting cost incurred by the first customer until he completes service (and also by all customers who arrive while he is being served), and τ_y the busy period duration of the other customers when starting in state (x, y). Note that κ_y and τ_y are not equal in general, since the former takes the residual service time of the first customer into account. Since the cost in state (x, y) accumulates at rate x, the total cost of serving one customer is $x\tau_y$ if $x > 1$ and κ_y if $x = 1$. Hence, the total cost incurred until the system is empty is given by $V(x, y) = \sum_{i=1}^{x-1} i\tau_y + \kappa_y$, showing that the structure of $V(x, y)$ is indeed of the form claimed in Theorem 1.

Note that the argument above implies that the value function for the M/G/1 queue is likely to be quadratic as well. Moreover, suppose that the holding costs are given by a polynomial in x of degree k. Then the argument shows that the value function $V(x, y)$ will be a polynomial in x of degree $k + 1$, with coefficients that, except for the coefficient of x^{k+1}, depend on y.

To prove Theorem 1 we first develop recursive equations for the terms appearing in $V(x, y)$. These equations will be useful in showing that $V(x, y)$ satisfies the Poisson equations. Additionally, the recursive equations might be preferred in practical applications, from a computational viewpoint.

Lemma 1. *Let the constants* α *and* a_0 *and the value function* $V(x, y)$ *be defined as in Theorem 1. The value function can then be written as*

$$V(x, y) = \alpha \frac{x(x+1)}{2} + a_y x + b_y$$

for $(x, y) \in \mathcal{S}$. *Moreover, the terms* a_y *and* b_y *satisfy the recursions*

$$a_y = \frac{a_{y-1}}{p_y} + \frac{(1 - p_y)\mu_y(\alpha - a_0) - (1 + \lambda\alpha)}{p_y \mu_y},$$

$$b_y = \frac{b_{y-1}}{p_y} + \frac{\lambda(a_0 - a_{y-1}) + (1 - p_y)\mu_y a_0}{p_y \mu_y},$$

for $y = 1, \dots, r - 1$, *with* $b_0 = 0$.

Proof. Let $y \in \{0, \dots, r - 1\}$ and consider a_y. From (2) it follows that

$$a_y = a_0 + \left[\frac{1}{\gamma(y)} - 1\right]\alpha - \sum_{k=1}^{y} \frac{\gamma(k-1)}{\gamma(y)} \frac{1 + \lambda\alpha}{\mu_k} \tag{3}$$

$$= \frac{a_0}{\gamma(y)} + \sum_{k=1}^{y} \frac{\gamma(k)}{\gamma(y)} \frac{(1 - p_k)\mu_k(\alpha - a_0) - (1 + \lambda\alpha)}{p_k \mu_k}. \tag{4}$$

The second line follows from rearranging terms in the first line and using the fact that $\gamma(k) = \gamma(k-1)p_k$. From (4), the recursive formula for a_y immediately follows. Now, let $y \in \{0, \dots, r - 1\}$ and consider b_y. From (2) it follows that

$$b_y = -\left[a_0 + \sum_{k=y+1}^{r} \frac{\gamma(k-1)}{\gamma(y)} \frac{\lambda}{\mu_k} \left(\sum_{l=1}^{k-1} \frac{\gamma(l-1)}{\gamma(k-1)} \frac{1 + \lambda\alpha}{\mu_l} - \left[\frac{1}{\gamma(k-1)} - 1\right]\alpha \right)\right]$$

$$= \sum_{k=1}^{y} \frac{\gamma(k-1)}{\gamma(y)} \frac{\lambda}{\mu_k} \left(\sum_{l=1}^{k-1} \frac{\gamma(l-1)}{\gamma(k-1)} \frac{1 + \lambda\alpha}{\mu_l} - \left[\frac{1}{\gamma(k-1)} - 1\right]\alpha \right) + \left[\frac{1}{\gamma(y)} - 1\right]a_0$$

$$= \sum_{k=1}^{y} \frac{\gamma(k)}{\gamma(y)} \frac{\lambda(a_0 - a_{k-1})}{p_k \mu_k} + \left[\frac{1}{\gamma(y)} - 1\right]a_0 \tag{5}$$

$$= \sum_{k=1}^{y} \frac{\gamma(k)}{\gamma(k)} \frac{\lambda(a_0 - a_{k-1}) + (1 - p_k)\mu_k a_0}{p_k \mu_k}. \tag{6}$$

The second line follows from the first by adding and subtracting $a_0/\gamma(y)$, using the definition of a_0 in Theorem 1. The third line follows from observing that the quantity in parentheses equals $a_0 - a_{k-1}$, and by using the fact that $\gamma(k) = \gamma(k-1)p_k$. The last line follows from rearranging terms in the third line. From (6), both $b_0 = 0$ and the recursive formula for b_y immediately follow.

Recall that $V(x, y)$ has the interpretation of the asymptotic difference in total costs that results from starting the process in state (x, y) instead of some reference state. Since we assumed the reference state to be (0, 0) it follows that $V(0, 0) = b_0$ should equal 0. Lemma 1 shows that this requirement is indeed met.

We will next show that $V(x, y)$ satisfies the Poisson equations.

Proof of Theorem 1. (*Existence.*) Consider the optimality equation for state (0, 0). We have

$$g + \lambda V(0,0) - \lambda V(1,0) = \lambda(\alpha + a_0) + \lambda b_0 - \lambda(\alpha + a_0 + b_0) = 0.$$

Next consider state $(x, y) \in \mathbb{N} \times \{1, \ldots, r-1\}$. For this we have

$$\begin{aligned}
g &+ (\lambda + \mu_y)V(x, y-1) - \lambda V(x+1, y-1) - p_y\mu_y V(x, y) - (1-p_y)\mu_y V(x-1, 0) - x \\
&= g - \lambda[V(x+1, y-1) - V(x, y-1)] - p_y\mu_y[V(x, y) - V(x, y-1)] \\
&\quad + (1-p_y)\mu_y[V(x, y-1) - V(x-1, 0)] - x \\
&= \lambda(\alpha + a_0) - \lambda[\alpha(x+1) + a_{y-1}] - p_y\mu_y[(a_y - a_{y-1})x + (b_y - b_{y-1})] \\
&\quad + (1-p_y)\mu_y[\alpha x + (a_{y-1} - a_0)x + a_0 + b_{y-1}] - x \\
&= [\mu_y a_{y-1} + (1-p_y)\mu_y(\alpha - a_0) - (1 + \lambda\alpha)]x - p_y\mu_y a_y x \\
&\quad + [\mu_y b_{y-1} + \lambda(a_0 - a_{y-1}) + (1-p_y)\mu_y a_0] - p_y\mu_y b_y.
\end{aligned}$$

By substituting the recursion relations from Lemma 1 for a_y and b_y, it follows that the last quantity equals 0. Finally consider the state $(x, r-1)$, with $x \in \mathbb{N}$. We have

$$\begin{aligned}
g &+ (\lambda + \mu_r)V(x, r-1) - \lambda V(x+1, r-1) - \mu_r V(x-1, 0) - x \\
&= g - \lambda[V(x+1, r-1) - V(x, r-1)] + \mu_r[V(x, r-1) - V(x-1, 0)] - x \\
&= \lambda(\alpha + a_0) - \lambda[\alpha(x+1) + a_{r-1}] + \mu_r[\alpha x + (a_{r-1} - a_0)x + a_0 + b_{r-1}] - x \\
&= [(\mu_r - \lambda)\alpha + \mu_r(a_{r-1} - a_0) - 1]x + [(\lambda + \mu_r)a_0 - \lambda a_{r-1} + \mu_r b_{r-1}].
\end{aligned}$$

Let us study the coefficient of x and the constant separately. The coefficient of x is

$$\begin{aligned}
(\mu_r &- \lambda)\alpha + \mu_r(a_{r-1} - a_0) - 1 \\
&= (\mu_r - \lambda)\alpha + \mu_r\left[\frac{1}{\gamma(r-1)} - 1\right]\alpha - \mu_r\sum_{k=1}^{r-1}\frac{\gamma(k-1)}{\gamma(r-1)}\frac{1+\lambda\alpha}{\mu_k} - 1 \\
&= \frac{\mu_r}{\gamma(r-1)}\left[1 - \sum_{k=1}^{r}\frac{\gamma(k-1)}{\mu_k}\lambda\right]\alpha - \sum_{k=1}^{r}\frac{\gamma(k-1)}{\gamma(r-1)}\frac{\mu_r}{\mu_k} = 0.
\end{aligned}$$

The second line follows from (3) and the third line follows from rearranging the terms in the second line. The constant in the optimality equation for state $(x, r-1)$ is

$$
\begin{aligned}
&(\lambda + \mu_r)a_0 - \lambda a_{r-1} + \mu_r b_{r-1} \\
&\quad = \mu_r a_0 + \lambda(a_0 - a_{r-1}) + \sum_{k=1}^{r-1} \frac{\gamma(k)}{\gamma(r-1)} \frac{\mu_r}{\mu_k} \frac{\lambda}{p_k}(a_0 - a_{k-1}) + \mu_r \left[\frac{1}{\gamma(r-1)} - 1\right] a_0 \\
&\quad = \frac{\mu_r}{\gamma(r-1)} \left[a_0 + \sum_{k=1}^{r} \frac{\gamma(k-1)}{\mu_k} \lambda(a_0 - a_{k-1})\right] \\
&\quad = \frac{\mu_r}{\gamma(r-1)} \left[a_0 - \sum_{k=1}^{r} \frac{\gamma(k-1)}{\mu_k} \lambda\left(\left[\frac{1}{\gamma(k-1)} - 1\right]\alpha + \sum_{l=1}^{k-1} \frac{\gamma(l-1)}{\gamma(k-1)} \frac{(1+\lambda\alpha)}{\mu_l}\right)\right] \\
&\quad = \frac{\mu_r}{\gamma(r-1)} \left[a_0 - \sum_{k=1}^{r} \frac{1-\gamma(k-1)}{\mu_k} \lambda\alpha + \sum_{k=1}^{r} \sum_{l=1}^{k-1} \frac{\gamma(l-1)}{\mu_k \mu_l} \lambda(1+\lambda\alpha)\right] = 0.
\end{aligned}
$$

The second line follows from (5), and by rearranging terms we obtain the third line. The fourth line follows from (3) and, after simplifying, yields the last line.

As mentioned before, there are many pairs of g and V that satisfy the Poisson equations when the state space is not finite. Hence, we need to show that the solution proposed in Theorem 1 is unique. Consider a function $w\colon \mathcal{S} \to [1, \infty)$, which we shall refer to as a weight function. Define the weighted w-norm of a real-valued function u defined on $\mathcal{S}$ by

$$
\|u\|_w = \sup_{(x,y)\in\mathcal{S}} \frac{|u(x,y)|}{w(x,y)}.
$$

The approach we shall adopt in proving uniqueness is summarized in the next theorem.

Theorem 2. (Lemma 1 and Theorem 4 of [3].) *Consider a stable and aperiodic Markov cost chain. Let $M \subset \mathcal{S}$ be finite and let w be a weight function such that the Markov chain is w-geometrically recurrent, i.e.*

$$
\sum_{(x',y')\notin M} \frac{P_{(x,y)(x',y')} w(x',y')}{w(x,y)} < 1 \tag{7}
$$

for all $(x, y) \in \mathcal{S}$, where P is the transition matrix of the Markov chain. Assume that the cost function satisfies $\|c\|_w < \infty$. A pair (g, V) satisfying the Poisson equations is the unique solution when $\|g\|_w < \infty$ and $\|V\|_w < \infty$.

Note that the geometric recurrence condition is somewhat stricter than necessary. Dekker [4] required the more general geometric ergodicity condition, and Hernández-Lerma and Lasserre [7, pp. 11–17] generalized this even further. In the sequel we shall use the geometric recurrence condition, however, since in practice it is hard to verify that the other two conditions are satisfied.

Proof of Theorem 1. (*Uniqueness.*) Note that, from (1), the Markov chain of the M/Cox(r)/1 queue is stable. Furthermore, (1) also ensures that $g < \infty$ and, hence, that $\|g\|_w < \infty$ for any weight function w. Assume, without loss of generality, that $\lambda + \mu_i < 1$ for $i = 1, \ldots, r$;

this can always be obtained through scaling. The M/Cox(r)/1 queue then has the following transition rate matrix for $(x, y) \in \mathcal{S}$:

$$\begin{aligned} P_{(x,y)(x+1,y)} &= \lambda, \\ P_{(x,y)(x,y+1)} &= p_{y+1}\mu_{y+1}\,\mathbf{1}(x > 0, y < r-1), \\ P_{(x,y)(x-1,0)} &= [1 - p_{y+1}\,\mathbf{1}(y < r-1)]\mu_{y+1}\,\mathbf{1}(x > 0), \\ P_{(x,y)(x,y)} &= 1 - P_{(x,y)(x+1,y)} - P_{(x,y)(x,y+1)} - P_{(x,y)(x-1,0)}. \end{aligned}$$

Observe that the Markov chain is aperiodic, since $P_{(x,y)(x,y)} > 0$ for $(x, y) \in \mathcal{S}$. Let $M = \{(0, 0)\}$ and assume that $w(x, y) = (1 + k)^x$ for some $k > 0$. Now consider (7), the left-hand side of which is given by

$$\begin{cases} \lambda(1+k), & x = 0, y = 0, \\ \lambda(1+k) + p_{y+1}\mu_{y+1} + (1 - \lambda - \mu_{y+1}), & x = 1, y = 0, \dots, r-2, \\ \lambda(1+k) + (1 - \lambda - \mu_r), & x = 1, y = r-1, \\ \lambda(1+k) + p_{y+1}\mu_{y+1} + \dfrac{(1 - p_{y+1})\mu_{y+1}}{1+k} + (1 - \lambda - \mu_{y+1}), & x > 1, y = 0, \dots, r-2, \\ \lambda(1+k) + \dfrac{\mu_r}{1+k} + (1 - \lambda - \mu_r), & x > 0, y = r-1. \end{cases}$$

We need to choose k such that these expressions are all strictly less than 1. Note that if the fourth expression is less than 1, then the second expression is less than 1, and that if the fifth expression is less than 1, then the third expression is less than 1. The first expression immediately gives $k < (1-\lambda)/\lambda$. The fourth expression shows that $0 < k < [(1-p_i)\mu_i - \lambda]/\lambda$ for $i = 1, \dots, r-1$. The last expression yields $0 < k < (\mu_r - \lambda)/\lambda$. Define

$$k^* = \frac{\min\{1, (1 - p_1)\mu_1, \dots, (1 - p_{r-1})\mu_{r-1}, \mu_r\} - \lambda}{\lambda}.$$

Observe that $k^* > 0$, and that for any k, $0 < k < k^*$, the Markov chain is geometrically recurrent. Furthermore, the cost function $c(x) = x$ satisfies $\|c\|_w < \infty$. Moreover, the value function defined in Theorem 1, which is a quadratic polynomial in x, satisfies $\|V\|_w < \infty$. Hence, by Theorem 2 the value function defined in Theorem 1 is the unique solution to the Poisson equations.

3. Special cases

In this section we consider important special cases of the M/Cox(r)/1 queue. This includes queues with the hyper-exponential (H_r), the hypo-exponential (Hypo_r), the Erlang (E_r), and the exponential (M) distributions.

3.1. The M/H_r/1 queue

The single-server queue with hyper-exponential-distributed service times of order r is obtained by letting the service times consist of only one exponential phase, with parameter μ_i with probability q_i for $i = 1, \dots, r$. Note that the hyper-exponential distribution has the property that the coefficient of variation is greater than or equal to 1. Unfortunately, the hyper-exponential distribution is not directly obtained from the Coxian distribution through interpretation, but rather from showing that the Laplace transforms of the distribution functions are equal for specific parameter choices. For $r = 2$ this result follows from, e.g. Appendix B of [18]. The general case is obtained in the following theorem.

Theorem 3. *Under the assumption that $\mu_1 > \cdots > \mu_r$, a Coxian distribution with parameters $(p_1, \dots, p_{r-1}, \mu_1, \dots, \mu_r)$ is equivalent to a hyper-exponential distribution with parameters $(q_1, \dots, q_r, \mu_1, \dots, \mu_r)$ when the probabilities p_i are defined by*

$$p_i = \frac{\sum_{j=i+1}^{r} q_j \prod_{k=1}^{i}(\mu_k - \mu_j)}{\mu_i \sum_{j=i}^{r} q_j \prod_{k=1}^{i-1}(\mu_k - \mu_j)} \tag{8}$$

for $i = 1, \dots, r-1$.

Proof. Note that the probability p_i can be written as $p_i = h(i)/[\mu_i h(i-1)]$, where $h(i) = \sum_{j=i+1}^{r} q_j \prod_{k=1}^{i}(\mu_k - \mu_j)$ with $h(0) = 1$. Also observe that the Coxian distribution has exactly i exponential phases with probability $\prod_{j=1}^{i-1} p_j(1-p_i)$ for $i = 1, \dots, r$, with $p_r = 0$. Therefore, the Laplace transform of the Coxian distribution is given by

$$\begin{aligned} F^*(s) &= \sum_{i=1}^{r} \prod_{j=1}^{i-1} p_j(1-p_i) \prod_{l=1}^{i} \frac{\mu_l}{\mu_l + s} \\ &= \sum_{i=1}^{r} \frac{h(i-1)}{\mu_1 \cdots \mu_{i-1}}(1-p_i) \prod_{l=1}^{i} \frac{\mu_l}{\mu_l + s} \\ &= \sum_{i=1}^{r} \frac{\mu_i h(i-1) - h(i)}{\mu_1 \cdots \mu_i} \prod_{l=1}^{i} \frac{\mu_l}{\mu_l + s}. \end{aligned}$$

The final equality is obtained by substituting for p_i using (8). Note that

$$\mu_i h(i-1) - h(i) = \sum_{j=i}^{r} q_j \mu_j \prod_{m=1}^{i-1}(\mu_m - \mu_j) = \sum_{j=i}^{r} q_j \mu_j (-1)^{i-1} \prod_{m=1}^{i-1}(\mu_j - \mu_m).$$

Expanding the last product in $F^*(s)$ into partial fractions (see Section XI.4 of [5]) yields

$$\prod_{l=1}^{i} \frac{\mu_l}{\mu_l + s} = (-1)^{i+1} \sum_{k=1}^{i} \frac{\mu_1 \cdots \mu_i}{\prod_{m=1,\, m \neq k}^{i}(\mu_k - \mu_m)(\mu_k + s)}.$$

By combining the two expressions we obtain

$$\begin{aligned} F^*(s) &= \sum_{i=1}^{r} \sum_{j=i}^{r} q_j \mu_j \sum_{k=1}^{i} \frac{\prod_{m=1}^{i-1}(\mu_j - \mu_m)}{\prod_{m=1,\, m \neq k}^{i}(\mu_k - \mu_m)} \frac{1}{\mu_k + s} \\ &= \sum_{k=1}^{r} \sum_{j=k}^{r} q_j \mu_j \sum_{i=k}^{j} \frac{\prod_{m=1}^{i-1}(\mu_j - \mu_m)}{\prod_{m=1,\, m \neq k}^{i}(\mu_k - \mu_m)} \frac{1}{\mu_k + s}. \end{aligned}$$

Let $z(i, j)$ denote the term within the brackets. It can easily be checked using induction that $\sum_{i=k}^{j} z(i, j)$ equals 1 for $j = k$ and 0 for $j > k$. The latter follows by using the fact that $z(i+1, j) = [(\mu_j - \mu_i)/(\mu_k - \mu_{i+1})]z(i, j)$, whence the partial sums are given by $\sum_{i=l}^{j} z(i, j) = [(\mu_k - \mu_l)/(\mu_k - \mu_j)]z(l, j)$. Substituting this into $F^*(s)$ yields the Laplace transform of the hyper-exponential distribution,

$$F^*(s) = \sum_{k=1}^{r} q_k \frac{\mu_k}{\mu_k + s}.$$

Hence, from Theorem 1 of Section XIII.1 of [6], it follows that the Coxian distribution with parameters as defined in (8) is equivalent to a hyper-exponential distribution.

3.2. The M/Hypo$_r$/1 queue

The single-server queue with hypo-exponential-distributed service times of order r is obtained by letting the service be the sum of r independent random variables that are exponentially distributed with parameter μ_i in phase i, for $i = 1, \dots, r$. Thus, it can be obtained from the M/Cox(r)/1 queue by letting $p_1 = \cdots = p_{r-1} = 1$. The optimality equations are given by

$$\begin{aligned} g + \lambda V(0,0) &= \lambda V(1,0), \\ g + (\lambda + \mu_i) V(x, i-1) &= \lambda V(x+1, i-1) + \mu_i V(x,i) + x, \qquad i = 1, \dots, r-1, \\ g + (\lambda + \mu_r) V(x, r-1) &= \lambda V(x+1, r-1) + \mu_r V(x-1, 0) + x. \end{aligned}$$

Define $\beta(i) = \sum_{k=1}^{i} (1/\mu_k)$. The average cost is then given by

$$g = \frac{\lambda \beta(r)}{1 - \lambda\beta(r)} - \frac{\lambda^2}{1 - \lambda\beta(r)} \sum_{k=1}^{r} \frac{\beta(k-1)}{\mu_k}$$

and, under the assumption that $\lambda\beta(r) < 1$, the value function becomes

$$V(x,y) = \frac{\beta(r) x(x+1)}{2(1-\lambda\beta(r))} - \frac{x}{1-\lambda\beta(r)} \left[\lambda \sum_{k=1}^{r} \frac{\beta(k-1)}{\mu_k} + \beta(y) \right] + \frac{\lambda}{1-\lambda\beta(r)} \sum_{k=1}^{y} \frac{\beta(k-1)}{\mu_k}.$$

3.3. The M/E_r/1 queue

The single-server queue with Erlang-distributed service times of order r is obtained by letting the service be the sum of r independent random variables having a common exponential distribution. Thus, it can be obtained from the M/Cox(r)/1 queue by letting $p_1 = \cdots = p_{r-1} = 1$ and $\mu = \mu_1 = \cdots = \mu_r$. Note that the Erlang distribution can also be seen as a special case of the hypo-exponential distribution, and has coefficient of variation $1/r \le 1$. The optimality equations are given by

$$\begin{aligned} g + \lambda V(0,0) &= \lambda V(1,0), \\ g + (\lambda + \mu) V(x, i-1) &= \lambda V(x+1, i-1) + \mu V(x,i) + x, \qquad i = 1, \dots, r-1, \\ g + (\lambda + \mu) V(x, r-1) &= \lambda V(x+1, r-1) + \mu V(x-1, 0) + x. \end{aligned}$$

The average cost is given by

$$g = \frac{\lambda r}{\mu - \lambda r} - \frac{\lambda^2 r(r-1)}{2\mu(\mu - \lambda r)}$$

and, under the assumption that $\lambda r/\mu < 1$, the value function becomes

$$V(x,y) = \frac{r x(x+1)}{2(\mu - \lambda r)} - \frac{x}{\mu - \lambda r} \left[\frac{\lambda r(r-1)}{2\mu} + y \right] + \frac{\lambda y(y-1)}{2\mu(\mu - \lambda r)}.$$

3.4. The M/M/1 queue

The standard single-server queue with exponentially distributed service times is obtained by having only one phase in the M/Cox(r)/1 queue, i.e. $r = 1$. Let $\mu = \mu_1$. The optimality equations are then equivalent to

$$g + (\lambda + \mu)V(x) = \lambda V(x+1) + \mu V([x-1]^+) + x,$$

where $[x]^+ = \max\{x, 0\}$. The average cost is given by $g = \lambda/(\mu - \lambda)$ and, under the assumption that $\lambda/\mu < 1$, the value function becomes

$$V(x) = \frac{x(x+1)}{2(\mu - \lambda)}.$$

4. Application to routeing problems

In this section we illustrate how the value function can be used to obtain nearly optimal policies in controlled queueing systems. We show this by studying a routeing problem to parallel single-server queues. The general idea is to start with a policy such that each queue behaves as an independent single-server queue. By doing so, the average cost and the value function can be readily determined from the results of the previous sections. Then one step of policy iteration is performed, to obtain an improved policy without having to compute the policy iteratively.

The distribution of the service times does not necessarily need to be Coxian. Since the Coxian distributions are dense in the set of nonnegative distribution functions, we shall restrict ourselves to general nonnegative distributions. The approach to handling such distributions is to approximate these distributions with a Coxian distribution by using, e.g. the expectation-maximization (EM) algorithm (see [1]). The EM algorithm is an iterative scheme that minimizes the information divergence given by the Kullback–Leibler information for a fixed order r. For the cases we have considered, a Coxian distribution of order $r = 5$ was adequate to describe the original distribution.

Given that the service time distribution is adequately described by a Coxian distribution, the control problem we study can be formalized as follows. Consider two parallel, infinite-buffer single-server queues. The service times of server i are Cox-distributed with order r_i, with parameters $(p_1^i, \dots, p_{r_i-1}^i, \mu_1^i, \dots, \mu_{r_i}^i)$ for $i = 1, 2$. Furthermore, queue i has holding cost h_i for $i = 1, 2$. An arriving customer can be sent to either queue one or queue two. The objective is to minimize the average cost. Let x_i be the number of customers in queue i and y_i the number of completed phases of the customer in service, if there is one, at queue i, for $i = 1, 2$. The optimality equation for this system is then given by

$$\begin{aligned} g &+ (\lambda + \mathbf{1}(x_1 > 0)\mu_{y_1+1}^1 + \mathbf{1}(x_2 > 0)\mu_{y_2+1}^2)V(x_1, y_1, x_2, y_2) \\ &= h_1 x_1 + h_2 x_2 + \lambda \min\{V(x_1+1, x_2, y_1, y_2), V(x_1, x_2+1, y_1, y_2)\} \\ &\quad + \mathbf{1}(x_1 > 0)\mu_{y_1+1}^1 [p_{y_1+1}^1 V(x_1, x_2, y_1+1, y_2) + \bar{p}_{y_1+1}^1 V(x_1-1, x_2, 0, y_2)] \\ &\quad + \mathbf{1}(x_2 > 0)\mu_{y_2+1}^2 [p_{y_2+1}^2 V(x_1, x_2, y_1, y_2+1) + \bar{p}_{y_2+1}^2 V(x_1, x_2-1, y_1, 0)] \end{aligned}$$

for $(x_i, y_i) \in \{(0,0)\} \cup (\mathbb{N} \times \{0, \dots, r_i - 1\})$, with $p_{r_i}^i = 0$ and $\bar{p}_y^i = 1 - p_y^i$ for $i = 1, 2$.

Take as initial control policy the Bernoulli policy with parameter $\eta \in [0, 1]$, i.e. the policy that splits the arrivals into two streams such that arrivals occur with rate $\eta\lambda$ at queue one and

TABLE 1: Numerical results for $r = 2$ with $p^1 = 1$ and $\mu^1 = (2, 2)$.

Parameters for queue two	g_B	g'	g^*
$p^2 = \frac{2}{3}, \mu^2 = (2, \frac{4}{3})$	5.147 786	3.208 688	3.208 588
$p^2 = \frac{1}{2}, \mu^2 = (2, 1)$	5.405 949	3.332 179	3.332 038
$p^2 = \frac{2}{5}, \mu^2 = (2, \frac{4}{5})$	5.652 162	3.445 815	3.445 787

TABLE 2: Numerical results for $r = 5$ with $p^1 = (1, 1, 1, 1)$ and $\mu^1 = (2, 3, 2, 3, 4)$.

Parameters for queue two	g_B	g'	g^*
$p^2 = (\frac{9}{10}, \frac{4}{5}, \frac{7}{10}, \frac{3}{5}),\ \mu^2 = (2, 3, 2, 3, 4)$	6.175 842	3.787 954	3.783 727
$p^2 = (\frac{3}{5}, \frac{7}{10}, \frac{4}{5}, \frac{9}{10}),\ \mu^2 = (2, 3, 2, 3, 4)$	3.729 859	2.493 349	2.480 818
$p^2 = (\frac{2}{5}, \frac{1}{5}, \frac{4}{5}, \frac{1}{2}),\ \mu^2 = (3, 2, 4, 2, 3)$	1.399 628	1.169 286	1.132 408

rate $(1 - \eta)\lambda$ at queue two. Under the Bernoulli policy, the optimality equation is obtained by replacing the term

$$\lambda \min\{V(x_1 + 1, x_2, y_1, y_2), V(x_1, x_2 + 1, y_1, y_2)\}$$

with $\eta\lambda V(x_1 + 1, x_2, y_1, y_2) + (1 - \eta)\lambda V(x_1, x_2 + 1, y_1, y_2)$. Hence, it follows that the two queues behave as independent single-server queues for which the average cost g_i and the value function V_i, for $i = 1, 2$, can be determined from Theorem 1. The average cost, g', and the value function, V', for the whole system are then given by the sum of the individual average costs, $g = g_1 + g_2$, and the sum of the individual value functions, $V' = V_1 + V_2$, respectively. Note that, for a specified set of parameters, the optimal Bernoulli policy obtained by minimizing with respect to η is straightforward. We shall therefore use the optimal Bernoulli policy in the numerical examples. The policy improvement step now follows from the minimizing action in

$$\min\{V'(x_1 + 1, x_2, y_1, y_2), V'(x_1, x_2 + 1, y_1, y_2)\}.$$

The Coxian distribution allows for many interesting numerical experiments. Therefore, we restrict ourselves to four representative examples that display the main ideas. We shall use the notation g_B, g', and g^* for the average costs obtained under the optimal Bernoulli policy, the one-step improved policy, and the optimal policy, respectively. Moreover, we set $h_1 = h_2 = 1$, with $\lambda = \frac{3}{2}$ for the first example and $\lambda = 1$ for the other three examples.

We start with two queues having a Cox(2)-distribution. Queue one has an Erlang E_2-distribution with parameter $\mu = 2$, such that the mean and the variance of the service time equal 1 and 2, respectively. The parameters for queue two are chosen such that the mean remains 1 but the variance increases to 3, 4, and 5, in the respective examples. In Table 1 we summarize the results and show that the one-step improved policy has a close-to-optimal value. The proportional extra cost, $(g' - g^*)/g^*$, is practically 0 in all cases.

Next, we show the results of a similar experiment with $r = 5$. The service distribution at queue one is a fixed hypo-exponential Hypo_5-distribution with parameter $\mu = (2, 3, 2, 3, 4)$. The one-step improved policy again performs quite well. Table 2 shows that the greatest proportional extra cost is given by 0.03 (the third experiment).

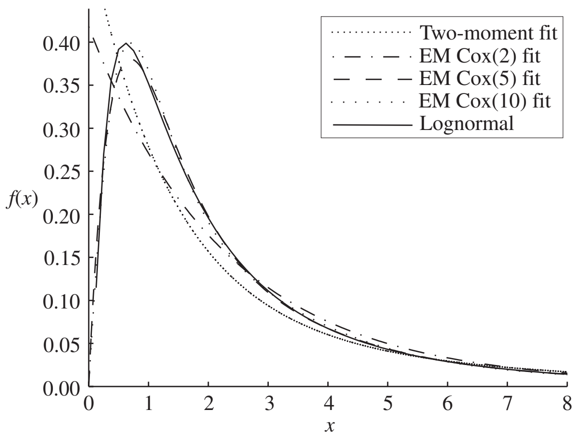

FIGURE 1: Lognormal ($\mu = 0.5, \sigma = 1$) probability density.

TABLE 3: Lognormal distribution: numerical results with $p^1 = 1$ and $\mu^1 = (2, 2)$.

Approximation for queue two	g_B	g'	g^*
Two-moment fit	4.617 707	3.021 571	2.976 950
EM algorithm with $r = 2$	4.554 838	2.982 955	2.933 345
EM algorithm with $r = 5$	4.526 013	2.965 625	2.919 847
EM algorithm with $r = 10$	4.527 392	2.963 318	2.917 011
EM algorithm with $r = 20$	4.527 040	2.963 311	2.917 169

In the third example we take an Erlang E_2-distribution with parameter $\mu = 2$ at queue one, and a lognormal distribution with parameters $\mu = 0.5$ and $\sigma = 1$ at queue two. Recall that the probability density function, f, of the lognormal distribution is given by

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right]$$

for $x > 0$. We variously approximate the lognormal distribution with Cox(r)-distributions of order $r = 2, 5, 10, 20$ using the EM algorithm. We also compare this with a two-moment fit of the Coxian distribution. Let X be a random variable having a coefficient of variation $c_X \geq \frac{1}{2}\sqrt{2}$. The following parameters were suggested in [11]: $\mu_1 = 2/\operatorname{E}(X)$, $p_1 = 0.5/c_X^2$, and $\mu_2 = p_1\mu_1$.

The results of the EM algorithm and the two-moment fit are displayed in Figure 1. The fit with the Cox(20)-distribution is omitted, since it could not be distinguished from the lognormal probability density. Therefore, the optimal value when using the Cox(20)-distribution can be considered representative of the optimal value when using the lognormal distribution. Note that the EM approximation with the Cox(2)-distribution captures more characteristics of the lognormal distribution than does the two-moment fit. This result is reflected in Table 3, since the value of the policy for the Cox(20)-distribution is closer to g^* than is the value of the policy for the Cox(2)-distribution. The greatest proportional extra cost for the EM approximations is given by 0.02 (EM fit with $r = 2$).

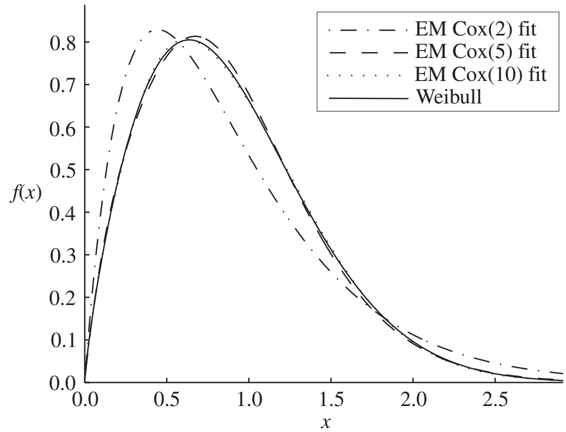

FIGURE 2: Weibull ($a = 1.8$, $b = 1$) probability density.

TABLE 4: Weibull distribution: numerical results with $p^1 = 1$ and $\mu^1 = (2, 2)$.

Approximation for queue two	g_B	g'	g^*
EM algorithm with $r = 2$	1.563 547	1.167 233	1.167 232
EM algorithm with $r = 5$	1.524 032	1.148 638	1.148 511
EM algorithm with $r = 10$	1.522 566	1.148 570	1.148 100
EM algorithm with $r = 20$	1.522 387	1.148 826	1.148 552

In the final example we take an Erlang E_2-distribution with parameter $\mu = 2$ at queue one, and a Weibull distribution with parameters $a = 0.3$ and $b = 0.107\,985$ at queue two. Recall that the probability density function, f, of the Weibull distribution is given by

$$f(x) = ax^{a-1}\frac{\mathrm{e}^{-(x/b)^a}}{b^a}$$

for $x > 0$. Note that the parameters a and b are chosen such that the Weibull distribution has mean 1. We variously approximate the Weibull distribution with Cox(r)-distributions of order $r = 2, 5, 10, 20$ using the EM algorithm. The results of the EM algorithm are depicted in Figure 2. We again omit the fit of the Cox(20)-distribution, since it could not be distinguished from the Weibull probability density. Moreover, since the coefficient of variation is less than $\frac{1}{2}\sqrt{2}$, we have also omitted the two-moment fit. The results displayed in Table 4 again indicate that the one-step improved policy has a close-to-optimal value, since the proportional extra cost is practically 0.

The previous examples show that the one-step policy improvement method yields nearly optimal policies, even when non-Markovian service distributions are approximated by a Coxian distribution. For the lognormal and the Weibull distributions we studied, a Coxian distribution of order $r = 5$ was already sufficient for an adequate approximation. Note that the one-step improved policy can be easily obtained for more than two queues. In this section we have restricted our attention to two queues, since the numerical computation of the value of the optimal policy, g^*, becomes rapidly intractable for more than two queues. Observe that the computational complexity is exponential in the number of queues, in contrast to a single step of policy iteration, which has linear complexity.

References

[1] ASMUSSEN, S., NERMAN, O. AND OLSSON, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scand. J. Statist.* **23,** 419–441.

[2] BHULAI, S. AND KOOLE, G. M. (2003). On the structure of value functions for threshold policies in queueing models. *J. Appl. Prob.* **40,** 613–622.

[3] BHULAI, S. AND SPIEKSMA, F. M. (2003). On the uniqueness of solutions to the Poisson equations for average cost Markov chains with unbounded cost functions. *Math. Meth. Operat. Res.* **58,** 221–236.

[4] DEKKER, R. (1985). Denumerable Markov decision chains: optimal policies for small interest rates. Doctoral Thesis, Leiden University.

[5] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1. John Wiley, New York.

[6] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2. John Wiley, New York.

[7] HERNÁNDEZ-LERMA, O. AND LASSERRE, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.

[8] KENDALL, D. G. (1951). Some problems in the theory of queues. *J. R. Statist. Soc.* **13,** 151–173.

[9] KOOLE, G. M. AND NAIN, P. (2004). An explicit solution for the value function of a priority queue. *Queueing Systems* **47,** 251–282.

[10] MARBACH, P., MIHATSCH, O. AND TSITSIKLIS, J. N. (2000). Call admission control and routing in integrated services networks using neuro-dynamic programming. *IEEE J. Sel. Areas Commun.* **18,** 197–208.

[11] MARIE, R. (1980). Calculating equilibrium probabilities for $\lambda(m)/C_k/1/N$ queues. In *Joint International Conference on Measurement and Modeling of Computer Systems* (Proc. 1980 Internat. Symp., Toronto), Association for Computing Machinery, New York, pp. 117–125.

[12] NORDSTROM, E. AND CARLSTROM, J. (1995). A reinforcement learning scheme for adaptive link allocation in ATM networks. In *Proc. Internat. Workshop Appl. Neural Networks Telecommun.*, eds J. Alspector *et al.*, Lawrence Erlbaum, Mahwah, NJ, pp. 88–95.

[13] NORMAN, J. M. (1972). *Heuristic Procedures in Dynamic Programming*. Manchester University Press.

[14] OTT, T. J. AND KRISHNAN, K. R. (1992). Separable routing: a scheme for state-dependent routing of circuit switched telephone traffic. *Ann. Operat. Res.* **35,** 43–68.

[15] PUTERMAN, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.

[16] SASSEN, S. A. E., TIJMS, H. C. AND NOBEL, R. D. (1997). A heuristic rule for routing customers to parallel servers. *Statist. Neerlandica* **51,** 107–121.

[17] SCHASSBERGER, R. (1973). *Warteschlangen*. Springer, Vienna.

[18] TIJMS, H. C. (1994). *Stochastic Models: An Algorithmic Approach*. John Wiley, New York.

[19] TONG, H. AND BROWN, T. X. (2000). Adaptive call admission control under quality of service constraints: a reinforcement learning solution. *IEEE J. Sel. Areas Commun.* **18,** 209–221.

[20] TOWSLEY, D., HWANG, R. H. AND KUROSE, J. F. (2000). MDP routing for multi-rate loss networks. *Computer Networks* **34,** 241–261.