AT&T BELL LABORATORIES TECHNICAL JOURNAL
Vol. 63, No. 1, January 1984
Printed in U.S.A.

On Approximations for Queues, II: Shape Constraints

By J. G. KLINCEWICZ* and W. WHITT*

(Manuscript received April 22, 1983)

This paper continues the investigation begun in Part I of approximations for queues that are based on a few parameters partially characterizing the arrival process and the service-time distribution. Part I provides insight into approximations for intractable systems by considering the set of all possible values of the mean queue length in the GI/M/1 queue given the service rate and the first two moments of the interarrival-time distribution. The distributions yielding the maximum and minimum values of the mean queue length turn out to be quite unusual, e.g., two-point distributions. This paper shows that the range of possible values can be reduced dramatically by imposing realistic shape constraints on the interarrival-time distribution with given first two moments. We found extremal distributions in the presence of shape constraints by restricting our attention to discrete distributions with all mass on a fixed finite set of points and solving nonlinear programs. The results strongly support the use of two-moment approximations in general queueing systems when the interarrival-time and service-time distributions are not too irregular.

I. INTRODUCTION AND SUMMARY

This paper continues the investigation begun in Part I[1] of the set of possible values of the mean queue length L (number in system) in a GI/M/1 queue given the service rate, μ, and various parameters partially characterizing the interarrival time cdf F (e.g., the first two moments m_1 and m_2). As explained in Part I, we are not primarily

* AT&T Bell Laboratories.

interested in the GI/M/1 model itself; we wish to provide a basis for evaluating approximations for more complex queueing models such as the nodes in a non-Markov network of queues.[2] For such complex models, the arrival process may be approximated by a renewal process, partially characterized by the first two moments of the renewal interval. Then the GI/M/1 model arises as an approximation and there is no complete interarrival-time distribution for an exact solution. We examine the GI/M/1 queue because it is tractable and because we believe it is indicative of what happens more generally.

For the GI/M/1 queue,[3]

$$L = \rho/(1 - \sigma), \tag{1}$$

where ρ is the traffic intensity ($\rho = 1/\mu m_1$) and σ is the unique root in the open interval (0, 1) of the equation

$$\phi[\mu(1 - \sigma)] = \sigma, \tag{2}$$

with $\phi(s)$ the Laplace-Stieltjes transform of the interarrival-time cdf F:

$$\phi(s) = \int_0^\infty e^{-st} dF(t). \tag{3}$$

Unfortunately, given m_1, m_2, and μ, the range of possible values of L can be very wide. (See the example in Section I of Part I.) This wide range naturally raises doubts about the value of two-moment approximations, but the particular distributions yielding the extreme values of L suggest that the approximations may still be useful. As we indicated in Part I, these extremal distributions are discrete probability distributions with positive probability on just two points. These two-point distributions are obviously very unusual. We would hope that for typical (nice) distributions L would not vary much among interarrival-time distributions with the same moments. In this paper, we investigate how much the range is reduced by imposing regularity conditions on the interarrival-time distribution. The regularity conditions we consider are shape constraints such as unimodality and log-convexity (a natural smoothness condition; see Chapter 5 of Keilson[4] and Section II).

A major contribution here, we believe, is the method. To study the effect of the shape constraints, we restrict attention to discrete distributions with all mass on a fixed finite set of points. We then find the range of the mean queue length L by means of nonlinear programming.

Since typical interarrival-time distributions are smooth (have densities), some may distrust results based on discrete distributions. However, continuity theorems show that there is no loss of generality, at least in principle, in considering distributions concentrating on a fixed finite set of points; see Section 11 of Borovkov.[5] With enough points, such discrete distributions can be used to approximate an arbitrary interarrival-time distribution arbitrarily well (in the usual sense of convergence in distribution and convergence of moments). In turn, the queue-length distribution and the mean queue length L associated with finite-valued probability mass functions can be used to approximate the queue-length distribution and the mean queue length L associated with the arbitrary interarrival-time distribution.

The point is that we need not worry about the local behavior of the interarrival-time distribution. For sufficiently small positive ϵ, if we change an interarrival-time density, say $f(t)$, only on the interval $[t_0, t_0 + \epsilon]$, for example, by making

$$f_n(t) = \begin{cases} f(t) & t \notin [t_0, t_0 + \epsilon] \\ nf[t_0 + n(t - t_0)], & t_0 \le t \le t_0 + \epsilon/n \\ 0 & t_0 + \epsilon/n < t \le t_0 + \epsilon, \end{cases}$$

then the new density $f_n(t)$ will be very different from the density $f(t)$ on $[t_0, t_0 + \epsilon]$ for large n, but the associated cdf's will be close and the behavior of the associated queueing systems will be virtually indistinguishable.

While there is no loss in generality in restricting attention to discrete distributions, it is not clear how many points are enough and where they should be located. We have not made a systematic investigation of this question, but we believe that we have used enough points in our study. It is important to recognize that extra points are not free because the nonlinear programs typically become harder to solve.

Throughout this paper, we use 21 points on the integers $\{0, 1, 2, \cdots, 20\}$. By comparing the programming results without shape constraints here with the theoretical results based on the complete Tchebycheff systems in Part I, we can see the effect of the discreteness. This effect can be seen in Tables II and V. The upper bound 20 on the support of the distribution (which might not be regarded as an essential aspect of the discreteness) can have a significant impact, but otherwise the discreteness matters little.

Do the shape constraints help? For the GI/M/1 example, with $\rho = 2/3$ and $c^2 = 2$, assuming a log-convex probability mass function reduces the maximal possible error in L from 200 percent to 8 percent. If the third moment is fixed as well, the maximal possible error is less than 1 percent.

These results indicate that two-moment approximations can be very useful, provided the interarrival-time distribution is actually not unusually irregular. In this paper we only study the GI/M/1 queue, but we believe the results are indicative of what happens in GI/G/1 queues and more general systems.

On the other hand, even for the GI/M/1 queue, the results do not imply that the two-moment approximations will work well in all circumstances or that they should be used blindly. If it is known that the interarrival-time distribution has an unusual shape, then the approximation should probably be modified. If additional information is known that would permit working with a third parameter such as the third moment or the peakedness, then better results can be expected. As noted by Kuczura[6] in a related context, a third parameter seems to offer the possibility of significant improvement, but additional parameters are rarely worth the effort.

This paper is organized as follows. In Section II, we define prototype distributions, introduce the shape constraints, and formulate the mathematical programs. The prototype distributions are intended to be typical interarrival-time distributions, which we use to generate parameter values and the "exact" queue characteristics σ and L. In Section III, we discuss the computational results for shape constraints with the first two moments fixed. In Section IV, we discuss the results for shape constraints with other parameters fixed. In Section V, we compare our results to other bounds and approximations. Finally, in Section VI, we discuss mathematical programming issues. It turns out that solving the nonlinear programs was not routine. These queueing problems may be interesting test problems for nonlinear programming codes.

We conclude this introduction by mentioning an interesting outcome of our experiments. Unlike Part I, the approach here is primarily numerical, being based on nonlinear programs, but the extremal distributions yielding the minimum and maximum values of L obtained from the nonlinear programs exhibit regularity that suggests the possibility of an analytic treatment similar to Part I. The extremal distributions we obtain on the set $\{0, 1, \cdots, 20\}$ have special structure and evidently do not depend on the traffic intensity. Hence, it may be possible to obtain analytic characterizations; this is a promising direction of research. (In fact, an analytic approach to shape constraints is carried out in Part III,[7] but not for discrete distributions and not for the shape constraints considered here.) Also, the robustness of the extremal distributions suggests that, just as with the extremal distributions in Part I, they should be useful in other contexts, e.g., to study the quality of two-moment approximations for inventory and reliability models as well as other queues.

II. PROTOTYPE DISTRIBUTIONS, SHAPE CONSTRAINTS, AND NONLINEAR PROGRAMS

2.1 Prototype distributions

To compare alternate parameter specifications and shape constraints in a consistent and meaningful way, we introduce two "prototype" distributions. The specified parameter values, e.g., the moments, will be the parameter values of one of the prototype distributions. The specified shape constraints also will be satisfied by one of the prototype distributions. In this way, we guarantee that there is at least one reasonable probability mass function satisfying all the conditions.

Since mixtures and convolutions of two exponential distributions are frequently used in queueing, we use the discrete analogues: mixtures and convolutions of two geometric distributions. The mixture of two geometric distributions has probability mass function

$$p_k = \gamma(1-\alpha)\alpha^k + (1-\gamma)(1-\beta)\beta^k, \qquad k \geq 0, \tag{4}$$

for probabilities α, β, and γ. As is often done with mixtures of exponential distributions,[8] we assume balanced means; i.e., we assume that $\gamma\alpha/(1-\alpha) = (1-\gamma)\beta/(1-\beta)$. The convolution of two geometric distributions, on the other hand, has probability mass function

$$p_k = \sum_{j=0}^{k} (1-\alpha)\alpha^j(1-\beta)\beta^{k-j} \tag{5}$$

for probabilities α and β.

To have finite support, we truncate the distributions, and work with the conditional distribution given that the upper bound is not exceeded. We truncate at 20, so that the support is the set of 21 integers $\{0, 1, 2, \cdots, 20\}$. In each case the upper bound 20 is at least 5 standard deviations above the mean.

Mixtures of exponential and geometric distributions are relatively more variable with squared coefficient of variation $c^2 > 1$, while convolutions are relatively less variable with $c^2 < 1$. Hence, we consider one prototype distribution of each type. Prototype I is a truncated mixture of two geometric distributions, having $c^2 = 2.0$; Prototype II is a truncated convolution of two geometric distributions, having $c^2 = 0.8$.

To obtain the specific prototype distributions, we start with the first two moments. For Prototype I, $m_1 = 2.0$ and $m_2 = 12.0$ ($c^2 = 2.0$) and, for Prototype II, $m_1 = 4.0$ and $m_2 = 28.8$ ($c^2 = 0.8$). To obtain a Prototype I distribution with the chosen values of m_1 and m_2, we numerically solve a system of three nonlinear equations in the three unknowns α, β, and γ. Two of these equations are the formulas for the moments m_1 and m_2 of the truncated distribution; the third is the "balanced means" equation. To obtain a Prototype II distribution with the chosen values of m_1 and m_2, we solve the system of two nonlinear equations in α and β corresponding to the moments m_1 and m_2 of the truncated distribution. The two prototype distributions are displayed in Table I. Additional parameters (the third moment, m_3, and trans-

Table I—The two prototype distributions: probability mass functions with p_k on k

	Prototype I		Prototype II	
k	p_k	p_k/p_{k+1}	p_k	p_k/p_{k+1}
0	0.3572	1.58	0.1215	0.79
1	0.2262	1.58	0.1536	1.04
2	0.1435	1.57	0.1475	1.16
3	0.0912	1.57	0.1272	1.22
4	0.0583	1.56	0.1040	1.26
5	0.0374	1.54	0.0825	1.28
6	0.0243	1.52	0.0642	1.30
7	0.0160	1.49	0.0494	1.31
8	0.0107	1.45	0.0377	1.32
9	0.0074	1.40	0.0286	1.32
10	0.0053	1.34	0.0217	1.32
11	0.0040	1.27	0.0163	1.33
12	0.0031	1.21	0.0123	1.33
13	0.0026	1.15	0.0093	1.33
14	0.0022	1.11	0.0070	1.33
15	0.0020	1.08	0.0053	1.33
16	0.0019	1.05	0.0040	1.33
17	0.0018	1.04	0.0030	1.33
18	0.0017	1.03	0.0022	1.33
19	0.0017	1.02	0.0017	1.33
20	0.0016	—	0.0013	—
	mean m_1	2.00	mean m_1	4.00
	c^2	2.00	c^2	0.80

form values, e.g., evaluated at the service rate μ) are given in Tables IV and V of Part I.

2.2 Shape constraints

Mixtures and convolutions of exponential and geometric distributions have many nice properties; see Chapter 5 of Keilson.[4] Mixtures of exponential and geometric distributions are log-convex and thus are DFR, i.e., have decreasing failure rate. For discrete distributions with probability mass functions p_k on the nonnegative integers, *log-convexity* means

$$p_k^2 \leq p_{k-1}p_{k+1}, \qquad k \geq 1. \tag{6}$$

Since the ratios p_k/p_{k-1} are nondecreasing with log-convexity, the distribution changes smoothly. The *failure rate* is

$$r_k = p_k \bigg/ \sum_{j=k}^{\infty} p_j, \qquad k \geq 0. \tag{7}$$

Decreasing failure rate of course implies that the probability mass function is decreasing. For log-convex distributions, $c^2 \geq 1$ and $m_3 \geq (3/\sqrt{2})m_2^{3/2}$ (see p. 69 of Keilson[4]).

Convolutions of exponential and geometric distributions are *log-concave*, i.e., the inequality (6) is reversed. Log-concavity is equivalent to *strong unimodality*. A probability mass function p_k on the nonnegative integers is *unimodal* if there is an integer k_0 such that

$$p_k \geq p_{k-1} \quad \text{for} \quad k \leq k_0$$

and

$$p_k \geq p_{k+1} \quad \text{for} \quad k \geq k_0. \tag{8}$$

A probability mass function p_k is *strongly unimodal* if the convolution with any unimodal probability mass function remains unimodal. In addition to being strongly unimodal, log-concave distributions are IFR, i.e., have increasing failure rate. For log-concave distributions, $c^2 \leq 1$ and $m_3 \leq (3/\sqrt{2})m_2^{3/2}$ (see p. 69 of Keilson[4]).

Of course, truncation and conditioning alter some of these properties. For example, the failure rates are changed significantly. For Prototype I, the failure rate is decreasing for the first thirteen values but is increasing after that. The failure rate remains increasing for Prototype II. The mass function ratios p_k/p_{k+1} are unchanged by the truncation, however. Also the unimodality properties are unchanged: Prototype I is decreasing and Prototype II is unimodal with a mode at 1.

In our study, we focus on the shape constraints unaffected by the truncation, namely, log-convexity or log-concavity and unimodality. We also consider additional parameters such as the third moment, transform values, and constraints on the cdf F.

2.3 The nonlinear programs

From (1), we see that the mean queue length L depends only on the traffic intensity ρ and the root σ of (2). Since L is an increasing function of σ, the maximum and minimum values of L are attained by the maximum and minimum values of σ. For interarrival-time distributions with probability mass functions $\{p_k\}$ on the set $\{0, 1, 2, \cdots, 20\}$, (2) becomes

$$\sum_{k=0}^{20} e^{-\mu(1-\sigma)k} p_k = \sigma. \tag{9}$$

To find the maximum and minimum values of σ, we solve nonlinear programs. The variables are σ and the probability masses p_k, $k = 0, 1, \cdots, 20$. The constraints specify that $\{p_k\}$ is a proper probability distribution with the specified properties and that (9) holds.

Given the two interarrival-time moments m_1 and m_2, the upper bound $b = 20$ on the support of the interarrival-time distribution, the service rate μ, and no shape constraints, we have a nonlinear program (NLP) for the maximum of the form:

$$\text{(NLP)} \qquad \max \sigma, \tag{10a}$$

subject to:

$$\sum_{k=0}^{20} e^{-\mu(1-\sigma)k} p_k = \sigma, \tag{10b}$$

$$\sum_{k=0}^{20} p_k = 1, \tag{10c}$$

$$\sum_{k=0}^{20} k p_k = m_1, \tag{10d}$$

$$\sum_{k=0}^{20} k^2 p_k = m_2, \tag{10e}$$

$$p_k \geq 0 \quad \text{for all} \quad k, \tag{10f}$$

$$0 \leq \sigma \leq 1 - \epsilon, \tag{10g}$$

where $0 < \epsilon < 1$. For any probability mass function $\{p_k\}$, the queue is stable if and only if $\rho = 1/\mu m_1 < 1$, in which case σ is the unique solution to (10b) in the open interval (0, 1). Since $\sigma = 1$ is also a solution to (10b), we rule it out by bounding σ above in (10g).

Of course, we obtain a corresponding NLP for the minimum value of σ by changing (10a) from a maximum to a minimum. If there is a mode at k_0, then we add the constraints (8) to (10). In the nonlinear program for $c^2 = 2.0$, we assumed that the mode is at 0; in the nonlinear program for $c^2 = 0.8$, we assumed that the mode is at 1. This is consistent with the location of the modes in the prototype distributions. If we were to assume only unimodality without specifying where the mode is, then we would have to solve a program for each possible mode location, and then optimize over the solutions.

When log-convexity is assumed, we add the constraints (6) for $k = 1, \cdots, 19$, to (10). For log-concavity, we add the constraints (6) with the inequality reversed.

III. SHAPE CONSTRAINTS WITH TWO MOMENTS FIXED

In this section, we give the minimum and maximum values of the root σ, denoted by σ_ℓ and σ_u, respectively, and the interarrival-time distributions yielding these extreme values of σ. From (1), we obtain the extreme values of L, denoted by L_ℓ and L_u. We also give the maximum relative error (MRE) in L, which is computed as

$$\mathrm{MRE} = \frac{L_u - L_\ell}{L_\ell} = \frac{\sigma_u - \sigma_\ell}{1 - \sigma_u}. \tag{11}$$

Table II gives the extremal characteristics and the MRE for the two prototype distributions ($c^2 = 2.0$ and 0.8), two values of the traffic intensity ($\rho = 2/3$ and $9/10$), and five constraint cases:

1. Two moments fixed only
2. Plus an upper bound $b = 20$ on the support of the distribution
3. Plus discrete, all mass on $\{0, 1, \cdots, 20\}$
4. Plus unimodal
5. Plus log-convex ($c^2 = 2.0$) or log-concave ($c^2 = 0.8$).

The results in the first two cases, before discreteness is imposed, come from Tables IV and V of Part I. The last three cases are the solutions to the nonlinear programs described in Section 2.3. Tables III through V give the associated extremal probability mass functions. Notice that these extremal distributions are the same for both values of ρ.

Each successive case adds an additional constraint to the one before, so the subsets of feasible interarrival-time distributions are nested, and the extremal characteristics get closer to the values for the prototype distributions.

The main conclusion is that with fairly strong but reasonable shape constraints the maximum relative error given two moments is dramatically reduced, becoming small enough to justify two-moment approximations. In particular, with log-convexity or log-concavity the MRE is always less than 8 percent, with the average MRE over the four cases being 3.8 percent. Unimodality helps, but is not good in the case $c^2 = 2.0$ and $\rho = 2/3$, yielding a 33.7-percent MRE. However, from Tables III and IV it is apparent that the unimodal extremal distributions are still quite irregular.

As in Part I, we see that the MRE gets smaller as ρ increases and c^2 decreases. From Table II, it is evident that this property holds with shape constraints as well as without. We also see that the upper bound of 20 on the support of the interarrival-time distribution strongly affects the minimum characteristic σ_ℓ but does not change the maximum characteristic σ_u at all. The discreteness either has no effect (for σ_u when $c^2 = 2.0$) or only a very small effect.

As we indicated above, there is another significant conclusion. The extremal probability distributions on the set of integers $\{0, 1, 2, \cdots, 20\}$ obtained from the nonlinear programs evidently share an important property with the extremal distributions on $[0, 20]$ given fixed parameters, treated in Part I: The extremal distributions computed by the nonlinear programs are evidently independent of the traffic intensity ρ.

Consider the case of no shape constraints (Table V). The extremal distributions on the set $\{0, 1, \cdots, 20\}$ computed by the nonlinear

Table II—The extremal characteristics, σ_ℓ and σ_u, and maximum relative errors (MRE) for the GI/M/1 queue: the cases of Prototype Distributions I and II, traffic intensities 2/3 and 9/10, and different shape constraints

Constraints on the Inter-arrival-Time Distribution	Prototype Distribution I, $c^2 = 2.0$						Prototype Distribution II, $c^2 = 0.8$					
	$\rho = 2/3$			$\rho = 9/10$			$\rho = 2/3$			$\rho = 9/10$		
	σ_ℓ	σ_u	MRE	σ_ℓ	σ_u	MRE	σ_ℓ	σ_u	MRE	σ_ℓ	σ_u	MRE
Two moments m_1 and m_2	0.417	0.806	2.00	0.807	0.936	2.00	0.417	0.676	0.80	0.807	0.893	0.808
Plus upper bound at 20	0.645	0.806	0.83	0.926	0.936	0.16	0.571	0.676	0.32	0.885	0.893	0.075
Plus discrete, all mass on (0, 1, 2, . . . , 20)	0.660	0.806	0.75	0.9268	0.9356	0.14	0.5732	0.6760	0.32	0.8848	0.8926	0.073
Plus unimodal	0.730	0.798	0.34	0.9306	0.9350	0.07	0.6145	0.6608	0.14	0.8878	0.8915	0.034
Plus log-convex (I) or log-concave (II)	0.762	0.779	0.08	0.9320	0.9336	0.02	0.6387	0.6533	0.04	0.8897	0.8909	0.001
The prototype distribution	0.7676	0.7676	0.00	0.9324	0.9324	0.00	0.6429	0.6429	0.00	0.8901	0.8901	0.000

Table III—The distributions minimizing the GI/M/1 mean queue length L given two moments and the shape constraints

	Prototype Distribution I, $c^2 = 2.0$		Prototype Distribution II, $c^2 = 0.8$	
	Unimodal	Log-Convex	Unimodal	Log-Concave
	$\rho = 2/3$ and $9/10$	$\rho = 2/3$ and $9/10$	$\rho = 2/3$ and $9/10$	$\rho = 2/3$ and $9/10$
0	0.2460	0.3486	0.0000	0.0619
1	0.2460	0.2260	0.1810	0.2155
2	0.2460	0.1465	0.1810	0.1663
3	0.2090	0.0950	0.1810	0.1283
4	0.0031	0.0616	0.1810	0.0990
5	0.0031	0.0399	0.1748	0.0764
6	0.0031	0.0259	0.0068	0.0589
7	0.0031	0.0168	0.0068	0.0455
8	0.0031	0.0109	0.0068	0.0351
9	0.0031	0.0071	0.0068	0.0271
10	0.0031	0.0046	0.0068	0.0209
11	0.0031	0.0030	0.0068	0.0161
12	0.0031	0.0019	0.0068	0.0124
13	0.0031	0.0012	0.0068	0.0096
14	0.0031	0.0008	0.0068	0.0074
15	0.0031	0.0005	0.0068	0.0057
16	0.0031	0.0003	0.0068	0.0044
17	0.0031	0.0002	0.0068	0.0034
18	0.0031	0.0001	0.0068	0.0026
19	0.0031	0.0001	0.0068	0.0020
20	0.0031	0.0088	0.0068	0.0016

programs are related to the analytic extremal distributions on the interval [0, b], derived in Section I of Part I. In Part I, the distribution yielding the upper limit of L is a two-point distribution with positive probability mass on 0 and another point x_u. The extremal distribution yielding the lower bound is also a two-point distribution with mass on a point x_ℓ and on b. (The points x_u and x_ℓ are determined by the requirement that the distributions have moments m_1 and m_2.) Our results support the following conjecture:

Conjecture 1: The extremal distributions on the set of integers $\{0, 1, 2, \cdots b\}$ *given the same two moments* m_1 *and* m_2 *have as mass points the triples* $(0, \lfloor x_u \rfloor, \lceil x_u \rceil)$ *and* $(\lfloor x_\ell \rfloor, \lceil x_\ell \rceil, b)$, *respectively, where* $\lfloor x \rfloor$ *is the greatest integer less than* x *and* $\lceil x \rceil$ *is the least integer greater than* x. *If* x_u *or* x_ℓ *is an integer, then the three-point extremal distribution reduces to a two-point distribution.*

With no additional shape constraints, we can show that the solution to the NLP (10) has at most three nonzero values of p_k. To see this, consider the situation where an extreme value of σ in (10) is known for particular values of m_1 and m_2. Then, we can combine (10a) and (10b) to form a linear objective function in the remaining variables p_k. With this linear objective, the three linear constraints (10c), (10d),

Table IV—The distributions maximizing the GI/M/1 mean queue length L given two moments and the shape constraints

	Prototype Distribution I, $c^2 = 2.0$		Prototype Distribution II, $c^2 = 0.8$	
	Unimodal	Log-Convex	Unimodal	Log-Concave
	$\rho = 2/3$ and $9/10$	$\rho = 2/3$ and $9/10$	$\rho = 2/3$ and $9/10$	$\rho = 2/3$ and $9/10$
0	0.5778	0.4377	0.1985	0.1719
1	0.0500	0.1571	0.1985	0.1450
2	0.0500	0.1133	0.0629	0.1223
3	0.0500	0.0817	0.0629	0.1031
4	0.0500	0.0589	0.0629	0.0870
5	0.0500	0.0425	0.0629	0.0734
6	0.0500	0.0306	0.0629	0.0619
7	0.0500	0.0221	0.0629	0.0522
8	0.0500	0.0159	0.0629	0.0440
9	0.0222	0.0115	0.0629	0.0371
10	0.0000	0.0083	0.0629	0.0312
11	0.0000	0.0060	0.0367	0.0262
12	0.0000	0.0043	0.0000	0.0221
13	0.0000	0.0031	0.0000	0.0181
14	0.0000	0.0022	0.0000	0.0043
15	0.0000	0.0016	0.0000	0.0001
16	0.0000	0.0012	0.0000	0.0000
17	0.0000	0.0008	0.0000	0.0000
18	0.0000	0.0006	0.0000	0.0000
19	0.0000	0.0004	0.0000	0.0000
20	0.0000	0.0003	0.0000	0.0000

and (10e), and the bounding constraints (10f), we can determine the values for the p_k by solving a linear program for which only three variables will be in the basis. Hence, to establish Conjecture 1, it suffices to verify that the special three-point distributions are optimal among all feasible three-point distributions for all these objective functions, i.e., for all arguments of the transform. Of course, if the extremal mass points x_ℓ and x_u for the distributions on $[0, b]$ are integer, then these extremal distributions on $[0, b]$ are feasible for the smaller set $\{0, 1, \cdots, b\}$ and are thus still optimal. This happens here for the upper bound with $c^2 = 2.0$.

The following conjecture for the cases with shape constraints is also supported by our experiments (we solved the programs for traffic intensities ranging from 0.01 to 0.9):

Conjecture 2: For each kind of shape constraint considered, the extremal interarrival-time distributions on $\{0, 1, 2, \cdots, b\}$ for the GI/M/1 queue, given the first two moments of the interarrival-time distribution, are independent of the traffic intensity, ρ.

Moreover, there is an obvious regularity in the extremal unimodal

Table V—The extremal GI/M/1 interarrival-time distributions without shape constraints: the effect of discreteness and an upper bound on the support of the distribution

	Prototype Distribution I, $c^2 = 2.0$					
Upper Bounds, σ_u	p_1	x_1	p_2	x_2	p_3	x_3
Cases 1, 2, and 3	0.6667	0.000	0.3333	6.00	—	—
Lower Bounds, σ_ℓ	p_1	x_1	p_2	x_2	p_3	x_3
Case 3	0.4211	1.000	0.5555	2.00	0.0234	20.00
Case 2	0.9759	1.556	0.0241	20.00	—	—
Case 1	1.0000	2.000	—	—	—	—
	Prototype Distribution II, $c^2 = 0.8$					
Upper Bounds, σ_u	p_1	x_1	p_2	x_2	p_3	x_3
Cases 1 and 2	0.444	0.000	0.556	7.20	—	—
Case 3	0.4429	0.000	0.4571	7.00	0.1000	8.00
Lower Bounds, σ_ℓ	p_1	x_1	p_2	x_2	p_3	x_3
Case 3	0.7529	3.000	0.2000	4.00	0.0471	20.00
Case 2	0.9524	3.200	0.0476	20.00	—	—
Case 1	1.000	4.000	—	—	—	—

Note: The cases are described at the beginning of Section II of this paper. x_i is the ith point with positive probability mass; p_i is the probability mass at point x_i.

distributions: they have only a few points of mass change. This can be explained by making a change of variables. For unimodal distributions on $\{0, 1, \cdots, b\}$ with a mode at 0, we can make the change of variables

$$q_k = (k + 1)(p_k - p_{k+1}), \qquad k \geq 0, \tag{12}$$

with $p_{b+1} = 0$. Then $q_k \geq 0$ for all k, and the constraints for p_k become the following constraints for q_k:

$$\sum_{k=0}^{b} q_k = 1, \qquad \sum_{k=0}^{b} kq_k = 2m_1 \tag{13}$$

and

$$\sum_{k=0}^{b} k^2 q_k = 3m_2 - m_1/2. \tag{14}$$

Moreover, the linear objective function $\sum_{j=0}^{b} e^{-sj} p_j$ is transformed into the linear objective function

$$\sum_{k=0}^{b} q_k (k + 1)^{-1} \sum_{j=0}^{k} e^{-sj}.$$

Solving the transformed linear program yields three-point solutions. Hence, the extreme points for the decreasing distributions with unimodal constraints have at most three points of decrease after 0. For decreasing probability mass functions, we thus make the following conjecture.

Conjecture 3: *Let* $(0, x_u)$ *and* (x_ℓ, b) *be the pairs of mass points for the extremal distributions on* $[0, b]$ *given the first two moments* $2m_1$ *and* $3m_2 - m_1/2$, *obtained from Section II of Part I. Then, for the* $GI/M/1$ *queue characteristics, the extremal* decreasing *probability mass functions on the set of integers* $\{0, 1, 2, \cdots, b\}$ *given the first two moments* m_1 *and* m_2 *have as points of decrease the triples* $(0, \underline{x_u}, \overline{x_u})$ *and* $(\underline{x_\ell}, \overline{x_\ell}, b)$, *respectively.* (*This completely determines the extremal probability distributions.*)

For other modes, say k_0, we can do a similar change of variables, namely,

$$q_{k_0+j} = \begin{cases} \dfrac{(2j+1)}{2}(p_{k_0+j} - p_{k_0+j+1}), & 0 \le j \le b - k_0, \\[2ex] \dfrac{-(1+2j)}{2}(p_{k_0+j+1} - p_{k_0+j}), & -k_0 - 1 \le j \le -1, \end{cases} \tag{15}$$

with $p_{-1} = p_{b+1} = 0$, so that q_k is a probability mass function on $\{-1, 0, 1, \cdots, 20\}$ for which

$$p_j = \sum_{i=0}^{i=j} a_{ij} q_{i-1}, \qquad 0 \le j \le b, \tag{16}$$

where a_{ij} are appropriate constants determined by (15). As before, the two linear moment constraints for p_k become linear moment constraints for q_k. In addition, there is an extra linear constraint on the q_k when $k_0 > 0$ since the support of q_k has one more point, i.e., is $\{-1, 0, 1, \cdots, b\}$ instead of $\{0, 1, \cdots, b\}$. In particular, from (15) it is easy to see that

$$\sum_{j=0}^{b-k_0} \left(\frac{2}{2j+1}\right) q_{k_0+j} + \sum_{j=1}^{k_0+1} \left(\frac{2}{2j-1}\right) q_{k_0-j} = 0. \tag{17}$$

Therefore, solving the transformed linear program would require the inclusion of (17) as a fourth linear constraint. This results in four positive values among the q_k. Therefore, extremal unimodal distributions with mode $k_0 > 0$ must have at most four points of mass change, including any positive mass at 0 and 20. The unimodal extremal distributions for Prototype II with $k_0 = 1$ obtained from the nonlinear programs have this property; see Tables III and IV. For Prototype II, we also found that the extremal characteristics σ_ℓ and σ_u both decreased as the mode k_0 was changed from 0 to 1 to 2, e.g., σ_ℓ changed from 0.633 to 0.614 to 0.603.

The numerical results also show that the extremal distributions for the log-convex and log-concave constraints have special regularity that can be seen by looking at the successive ratios p_k/p_{k+1}.

Conjecture 4: *In the log-convex case, the upper-(lower-) bound distribution has constant ratios p_k/p_{k+1} for $k = 1, 2, \cdots, b$ ($k = 0, 1, \cdots, b-1$) with an extra mass at 0 (b). (This determines the interarrival-time distribution.)*

Conjecture 4 is supported by Tables II and IV. With log-convex constraints, the upper-(lower-) bound ratios are $p_1/p_2 \approx 1.387$ ($p_1/p_2 \approx 1.542$).

IV. SHAPE CONSTRAINTS WITH OTHER PARAMETER SPECIFICATIONS

In this section, we investigate different parameter specifications, both with and without shape constraints. However, attention is focused on the cases with shape constraints because alternate parameter specifications without shape constraints were considered in Sections III and IV of Part I. We consider only Prototype Distribution I with the traffic intensity $\rho = 2/3$. This is the difficult case in Section III, yielding the largest maximum relative errors.

Table VI contains the major results. It gives the maximum relative errors in L for various combinations of two and three parameters with no shape constraints and with log-convex shape constraints. Of course, we still consider the first two moments m_1 and m_2. The additional parameters that we consider are: the third moment, m_3, the Laplace-Stieltjes transform evaluated at the service rate, $\phi(\mu)$, and the interarrival time cdf evaluated at k, $F(k)$, i.e., $F(k) = p_0 + p_1 + \cdots + p_k$. These parameters are fixed at the values satisfied by Prototype I. In particular, we use $m_3 = 119.01$, $\phi(\mu) = 0.5098$, $F(0) = 0.35724$, $F(2) = 0.72692$, and $F(7) = 0.95409$. Combinations of two parameters are

Table VI—A comparison of alternate second- and third-parameter specifications: the maximum relative error (MRE) in the mean queue length L in a GI/M/1 queue, based on Prototype I with $\rho = 2/3$

	The Second Parameter in Addition to m_1				
	m_2	$\phi(\mu)$	$F(0)$	$F(2)$	$F(7)$
No shape constraint	0.75	1.53	3.10	5.09	3.81
Plus log-convexity	0.077	0.083	0.096	0.370	0.555
	The Third Parameter in Addition to m_1 and m_2				
	m_3	$\phi(\mu)$	$F(0)$	$F(2)$	$F(7)$
No shape constraint	0.069	0.175	0.331	0.604	0.609
Plus log-convexity	0.009	0.012	0.019	0.049	0.020

Note: m_k is the kth moment, $\phi(\mu)$ is the Laplace-Stieltjes transform evaluated at the service rate μ, and $F(k)$ is the cdf evaluated at k, i.e., $F(k) = p_0 + p_1 + \ldots + p_k$, of Prototype Distribution I. These values are $m_1 = 2.00$, $m_2 = 12.00$, $m_3 = 119.01$, $\phi(\mu) = 0.5098$, $F(0) = 0.35724$, $F(2) = 0.72692$, and $F(7) = 0.95409$. The distribution itself appears in Table I. All these results are obtained from the nonlinear programs.

formed by specifying each of the additional parameters together with the first moment m_1. Combinations of three parameters are formed by specifying each of the additional parameters together with the first two moments m_1 and m_2.

The first conclusion is that, with log-convexity, the third moment or almost any other third parameter in addition to the first two moments makes the maximum relative error negligible. For the third moment, the maximum relative error is less than one percent and for all but one of the other third parameters it is less than two percent. This suggests that with nice distributions three-moment approximations ought to work very well for more general models.

The second conclusion is that the next higher moment is the best additional parameter in all cases. However, the advantage of the moment over the transform value decreases dramatically with log-convexity. Although the cdf constraints certainly reduce the MRE, the next higher moment and the transform value perform better as additional parameters.

In order to have the maximum relative error small enough to justify approximations, say less than 10 percent, it appears that three constraints are enough. It suffices to specify either three moments without shape constraints (6.9-percent MRE) or specify two moments with log-convex shape constraints (7.7-percent MRE). We can think of log-convexity as being roughly equivalent to another moment parameter.

The values of σ, the GI/M/1 probability of delay, for the various parameter specifications and shape constraints are given in Tables VII and VIII. From Table VII we see an interesting reversal in form with log-convexity. Without shape constraints, the next higher moment is better than the transform value $\phi(\mu)$ as an additional parameter for the upper bound but not as the second parameter for the lower bound. With log-convexity, these orderings are reversed. From Table VIII, we see that $F(0)$ is significantly better than the other two cdf values as an additional parameter for the upper bound with log-convexity and for the lower bound with no shape constraints, but not in the other cases.

We also tabulated the extremal interarrival-time distributions for the different combinations of parameters and shape constraints, but they have been omitted to save space. As in Section III, these extremal distributions have important regularity properties. With k parameters and no shape constraints, the extremal distributions have at most $k + 2$ positive mass points; with k parameters and a decreasing mass function, the extremal distribution have at most $k + 2$ points of mass change after 0. As in Section III, there is also regularity in the extremal distributions in the log-convex case, which can be seen by looking at the successive ratios p_k/p_{k+1}. There appear to be only a few points

Table VII—The GI/M/1 extremal characteristics σ (the probability of delay) given $c^2 = 2.0$, $\rho = 2/3$, and the different shape constraints

		Parameters Specified in Addition to the Mean				
		m_2	$\phi(\mu)$	$\phi(\mu(1-\rho))$	m_2, m_3	$m_2, \phi(\mu)$
	No shape constraints	0.8057	0.8811	0.7994	0.7768	0.7844
σ_u	Unimodal	0.7983	0.8203	0.7777	0.7708	0.7777
	Log-convex	0.7790	0.7736	0.7690	0.7694	0.7680
	Prototype Distribution	0.7675	0.7675	0.7675	0.7675	0.7675
	Log-convex	0.7621	0.7542	0.7638	0.7673	0.7652
σ_ℓ	Unimodal	0.7374	0.7192	0.7574	0.7615	0.7606
	No shape constraints	0.6601	0.6986	0.7545	0.7548	0.7466

Table VIII—The extremal characteristics σ (the probability of delay) for the GI/M/1 queue given values of the cumulative distribution function F in addition to the first moment, m_1, or the first two moments, m_1 and m_2: the case of Prototype Distribution I with traffic intensity $\rho = 2/3$

		Additional Parameter with m_1			Additional Parameter with m_1 and m_2		
		$F(0)$	$F(2)$	$F(7)$	$F(0)$	$F(2)$	$F(7)$
σ_u	No shape constraint	0.9093	0.9135	0.9034	0.7927	0.8019	0.8042
	Unimodal	0.8523	0.8558	0.8408	0.7867	0.7940	0.7959
	Log-convex	0.7760	0.8278	0.8385	0.7684	0.7781	0.7706
σ_ℓ	Log-convex	0.7546	0.7641	0.7489	0.7639	0.7670	0.7659
	Unimodal	0.6849	0.6683	0.6859	0.7544	0.7388	0.7427
	No shape constraint	0.6280	0.4736	0.5351	0.7241	0.6823	0.6849

where these ratios change. Including the final mass point, for k parameters there appear to be k points where the ratios change. Given m_2 and m_3, the ratios p_{k-1}/p_k change for the lower bound at $k \in \{10, 11\}$ and for the upper bound at $k \in \{2, 20\}$. Given only m_2, the ratio changes for the lower bound at $k = 20$ and for the upper bound at $k = 2$.

Although we report results for only a single value of the traffic intensity ρ, it also appears that the extremal distributions do not change with ρ, i.e., the numerical solutions to the nonlinear programs were indistinguishable for a range of ρ values tested from 0.01 to 0.9. There are natural extensions for Conjectures 1 through 4 to other parameter specifications.

We also found the extreme values of the transform values $\phi(\mu)$, which are the blocking probabilities for the associated GI/M/1 loss system, for given moments and shape constraints. The numerical solutions for the extremal distributions appear to be the same as those in which σ is the objective. The extremal blocking probabilities are given in Table IX. As in Section V and Table X of Part I, the constraints pin down the delay probability σ better than the associated blocking probability $\phi(\mu)$.

V. OTHER BOUNDS AND APPROXIMATIONS FOR GI/G/1 QUEUES

Having obtained the extreme values of the GI/M/1 mean queue length L given two moments and various shape constraints, we note how these results compare with other bounds and approximations for the GI/G/1 queue that depend only on the first two moments of the interarrival times and service times. Several of these other bounds and approximations are defined and compared in Shanthikumar and Buzacott.[9] These bounds are stated for the mean waiting time, but they are easily translated into the mean queue length by Little's formula. Among the bounds and approximations treated there is the Kingman[10] upper bound and the Marchal[11] approximation based on it. Recently, Daley[12] obtained a better upper bound, (1.5) there, which can be used to produce an approximation by scaling to make the M/G/1 case exact, just as Marchal did for the Kingman bound. We call this new approximation Marchal (D) and the original Marchal approximation Marchal (K). Shanthikumar and Buzacott also discuss the Kraemer and Langenbach-Belz[13] approximation and a modification of Page's[14] approximation based on it, formula (8) there, which we call Modified-Page. They also discuss an approximation by Sakasegawa[15] and Yu,[16] which coincides with the monotone-failure-rate approximation in Whitt.[17] Another natural two-moment approximation is to fit a hyperexponential distribution with balanced means to the two moments, provided $c^2 \geq 1$ (see Section III of Whitt[7]) and solve the resulting $H_2^b/H_2^b/1$ queue via a vector-state Markov process. When a distribution is exponential, H_2^b becomes M, so for the setting of the GI/M/1 queue based on Prototype I we obtain the $H_2^b/M/1$ queue. Finally, the crudest

Table IX—The extremal blocking probabilities for the associated GI/M/1 loss system (the transform values $\phi(\mu)$) with given moments and shape constraints: case of Prototype I with $m_1 = 2$, $m_2 = 12$, $m_3 = 119$, and $\mu = 4/3$ ($\rho = 2/3$)

	The Moment Parameters	
The shape constraints	m_1, m_2	m_1, m_2, m_3
Max $\phi(\mu)$, no shape constraints	0.6641	0.5902
Max $\phi(\mu)$, unimodal	0.6225	0.5414
Max $\phi(\mu)$, log-convex	0.5499	0.5147
Prototype I	0.5098	0.5098
Min $\phi(\mu)$, log-convex	0.5026	0.5087
Min $\phi(\mu)$, unimodal	0.4395	0.4713
Min $\phi(\mu)$, no shape constraints	0.3279	0.4049

approximation is obtained by ignoring the second moments and using the M/M/1 formula $L = \rho/(1 - \rho)$. There is also a related collection of approximations arising from diffusion approximations that we will not consider here; see Whitt[18] and references there.

We also include bounds for GI/G/1 queue in which the interarrival-time distribution is IFR or DFR.[17] Marshall[19] obtained a lower bound for IFR/G/1 queues and an upper bound for DFR/G/1 queues. Stoyan and Stoyan[20] also obtained an upper bound for IFR/G/1 queues and a lower bound for DFR/G/1 queues, which is just the M/G/1 queue with the given arrival rate. (In fact, the interarrival-time distribution is only required to be NBUE or NWUE, i.e., new better or worse than used in expectation.) The DFR bounds, but not the IFR bounds, are tight.[17]

In Table X these various bounds and approximations are compared with the extreme values of L for $c^2 = 2.0$ and 0.8 (the two prototype distributions), and $\rho = 2/3$ and 9/10. When interpreting these results, note that none of the other bounds and approximations use the fact that the service-time distribution is exponential. Also, the DFR and the IFR bounds are based on interarrival-time distributions having densities with support on the entire positive half line, whereas the bounds obtained here in Section II are based on interarrival-time distributions with support $\{0, 1, \cdots, 20\}$. The upper bound $b = 20$ on the support of the interarrival-time distribution has a significant impact on the lower bound mean queue length, L_{ℓ}, when the interarrival-time distribution in DFR ($c^2 > 1$) and on the upper bound mean queue length, L_u, when the interarrival-time distribution is IFR ($c^2 < 1$).

The first conclusion is that all the approximations, with the exception of the M/M/1 approximation when $c^2 = 2$, appear to be within the range of reasonable values for actual GI/M/1 systems. However, for $c^2 = 2.0$ and $\rho = 2/3$, the Modified-Page approximation seems a bit high. The Kraemer and Langenbach-Belz approximations for $c^2 = 2.0$ seem low compared to the log-convex discrete lower bounds (Case 5), but note that the Kraemer and Langenbach-Belz approximations are close to the H_2^b/M/1 values.

The second conclusion is that the D/M/1 lower bound and the Kingman and Daley upper bounds are not close enough to be good approximations. Of course, the upper bounds are asymptotically tight in heavy traffic, so they are not too bad when $\rho = 0.9$.

We believe that the shape constraints play a very useful role. They narrow down the range of possible values for L, so it is reasonable to consider approximations based on two moments only. Instead of concluding that it is not possible to obtain a good approximation when $c^2 > 1$ (p. 765 of Shanthikumar and Buzacott[9]), we conclude that it is

Table X—A comparison of the GI/M/1 extreme values of the mean queue length, L, with other bounds and approximations for L in GI/G/1 queues that depend on the first two moments of the interarrival times and service times

	Prototype Distribution I, $c^2 = 2.0$		Prototype Distribution II, $c^2 = 0.8$	
GI/G/1 upper bounds	$\rho = 2/3$	$\rho = 9/10$	$\rho = 2/3$	$\rho = 9/10$
Kingman	4.33	14.95	2.53	8.95
Daley	4.00	14.85	2.40	8.91
DFR or IFR	3.00	13.50	2.00	9.00
GI/M/1 upper bounds				
Case 1, two moments only	3.44	14.06	2.06	8.41
Case 2, bound on support	3.44	14.06	2.06	8.41
Case 3, discrete	3.44	13.98	2.06	8.38
Case 4, unimodal	3.30	13.85	1.97	8.29
Case 5, log-concave or log-convex	3.02	13.55	1.92	8.25
Approximations				
Marchal (K)	2.92	13.48	1.82	8.11
Marchal (D)	2.89	13.46	1.82	8.11
Kraemer and L-B	2.56	12.85	1.86	8.18
Modified-Page	3.28	13.34	1.83	8.13
Sakasegawa and Yu	2.67	13.05	1.87	8.19
$H_2^b/M/1$	2.64	13.03	—	—
M/M/1	2.00	9.00	2.00	9.00
GI/M/1 lower bounds				
Case 5, log-concave or log-convex	2.80	13.24	1.85	8.16
Case 4, unimodal	2.47	12.97	1.73	8.02
Case 3, discrete	1.96	12.30	1.56	7.83
Case 2, bound on support	1.88	12.16	1.55	7.81
Case 1, two moments only	1.14	4.66	1.14	4.66
GI/G/1 lower bound				
DFR or IFR	2.00	9.00	1.80	8.00

Note: The actual values of L for the prototype distributions are 2.87, 13.31, 1.87, and 8.19, respectively.

possible to consider approximations based on two moments, with the caveat that the distributions should not be too irregular.

VI. MATHEMATICAL PROGRAMMING ISSUES

Solving the nonlinear programs turned out to be quite complicated, especially when the shape constraints were included. The programs involve 22 variables and up to 46 constraints. This is a reasonably large problem for most general-purpose nonlinear programming codes. In addition, when the nonlinear constraints (6) are present, the problems apparently become ill-conditioned and poorly scaled, causing numerical difficulties and, often, nonconvergence of standard nonlinear programming algorithms.

The numerical results reported in this paper were obtained using two nonlinear programming codes from the Harwell Subroutine Library, compiled by the Numerical Analysis Group at the United Kingdom Atomic Energy Authority. These codes were VF01AD, an augmented Lagrangian code described in Fletcher,[21] and VF02AD, a quadratic approximation code due to Powell.[22] They were run in double precision on an Amdahl 470/V6 computer operating under multiple virtual storage. Both codes are included in a recent performance comparison of available state-of-the-art computer codes compiled by Schittkowski.[23]

Although numerical problems were experienced with both codes, they were far more prevalent with the augmented Lagrangian code VF01AD. The augmented Lagrangian code solves a sequence of unconstrained optimization subproblems. Unfortunately, some of the problems (especially for $\rho = 0.9$ and Prototype II) resulted in ill-conditioned subproblems and, occasionally, subproblems with an unbounded optimum, in which case, the augmented Lagrangian code did not converge. Our experience bears out the experience of Schittkowski, who reported that the performance of this code deteriorates drastically for ill-conditioned problems and is highly sensitive to slight variations of the problem. Certain standard measures, however, were able to overcome the numerical difficulties in most instances. For example, in some runs, the default settings for certain penalty parameters were overridden according to rules of thumb suggested by Gill, Murray, and Wright (see pp. 295–6 of Ref. 24).

Fortunately, for those experiments for which code VF01AD did not obtain a solution, code VF02AD did. This supports Schittkowski's conclusion that code VF02AD is one of the most robust and reliable codes available. Even though several individual runs experienced numerical overflows and underflows and eventual nonconvergence, it was always possible eventually to obtain convergence with this quadratic approximation code using some starting point. In particular, problems with $\rho = 0.9$ and Prototype II were solved with less difficulty using VF02AD.

All runs were tried with a variety of starting points. These starting points included the prototype distributions, the uniform distribution, solutions obtained for other parameter settings, and an initial all-zero solution.

In summary, then, although this nonlinear programming method for analyzing the quality of queueing approximations provides considerable insight and potential for future applications, great care must be exercised in the solution of the nonlinear programs. Our experience indicates that the computer code, the parameter settings, the starting points, and the scaling of the variables must be chosen judiciously in order to obtain useful results.

VII. ACKNOWLEDGMENT

We thank Hanan Luss and Moshe Segal for many suggestions on the presentation.

REFERENCES

1. W. Whitt, "On Approximations for Queues, I: Extremal Distributions," AT&T Bell Lab. Tech. J., this issue.
2. W. Whitt, "The Queueing Network Analyzer," B.S.T.J., *62*, No. 9 (November 1983), pp. 2779–2815.
3. J. W. Cohen, *The Single Server Queue*, Amsterdam: North Holland, 1969.
4. J. Keilson, *Markov Chain Models–Rarity and Exponentiality*, New York: Springer-Verlag, 1979.
5. A. A. Borovkov, *Stochastic Processes in Queueing Theory*, New York: Springer-Verlag, 1976.
6. A. Kuczura, "The Interrupted Poisson Process as an Overflow Process," B.S.T.J., *53*, No. 3 (March 1973), pp. 437–48.
7. W. Whitt, "On Approximations for Queues, III: Mixtures of Exponential Distributions," AT&T Bell Lab. Tech. J., this issue.
8. W. Whitt, "Approximating a Point Process by a Renewal Process: Two Basic Methods," Oper. Res., *30*, No. 1 (January–February 1982), pp. 125–47.
9. J. G. Shanthikumar and J. A. Buzacott, "On the Approximations to the Single Server Queue," Int. J. Prod. Res., *18* (1980), pp. 761–73.
10. J. F. C. Kingman, "Some Inequalities for the Queue GI/G/1," Biometrika, *49*, No. 3 (1962), pp. 315–24.
11. W. G. Marchal, "An Approximate Formula for Waiting Time in Single Server Queues," AIIE Trans. (1976), pp. 473–4.
12. D. J. Daley, "Inequalities for Moments of Tails of Random Variables, with a Queueing Application," Z. Wahrscheinlichkeitstheorie verw. Gebiete, *41* (1977), pp. 139–43.
13. W. Kraemer and M. Langenbach-Belz, "Approximate Formulae for the Delay in the Queueing System GI/G/1," Eighth Int. Teletraffic Cong., Melbourne, 1976, pp. 235-1–8.
14. E. Page, *Queueing Theory in OR*, New York: Crane Russak and Co., 1972.
15. H. Sakasegawa, "An Approximation Formula $L_q \simeq \alpha\rho^{\beta}/(1-\rho)$," Ann. Inst. Statist. Math., *29A* (1977), pp. 67–75.
16. P. S. Yu, "On Accuracy Improvement and Applicability Conditions of Diffusion Approximation with Application to Modelling of Computer Systems," Digital Systems Laboratory, *TR-129*, Stanford University, 1977.
17. W. Whitt, "The Marshall and Stoyan Bounds for IMRL/G/1 Queues are Tight," Oper. Res. Lett., *1*, No. 6 (December 1982), pp. 209–13.
18. W. Whitt, "Refining Diffusion Approximations for Queues," Oper. Res. Lett., *1*, No. 5 (November 1982), pp. 165–9.
19. K. T. Marshall, "Some Inequalities in Queueing," Oper. Res., *16*, No. 3 (May–June 1968), pp. 651–65.
20. H. Stoyan and D. Stoyan, "Monotonicity Properties of Waiting Times in the GI/G/1 Queue," Zeit. angew Math. Mech., *49*, No. 12 (1969), pp. 729–34 (in German).
21. R. Fletcher, "An Ideal Penalty Function for Constrained Optimization," Nonlinear Programming 2, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., New York: Academic Press, 1975.
22. M. J. D. Powell, "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," *Proceedings of the 1977 Dundee Conference on Numerical Analysis*, Lecture Notes in Mathematics, New York: Springer-Verlag, 1978.
23. K. Schittkowski, *Nonlinear Programming Codes: Information, Tests, Performance*, Lecture Notes in Economics and Mathematical Systems, New York: Springer-Verlag, 1980.

24. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, New York: Academic Press, 1981.

AUTHORS

John G. Klincewicz, S.B. (Mathematics), 1975, Massachusetts Institute of Technology; M.A., 1977, and Ph.D. (Operations Research), 1979, Yale University; AT&T Bell Laboratories, 1979—. At AT&T Bell Laboratories, Mr. Klincewicz is a member of the Operations Research Department. His research interests include applications of mathematical programming and the development of algorithms for network flow problems and facility location problems. Member, Operations Research Society, Mathematical Programming Society.

Ward Whitt, A.B. (Mathematics), 1964, Dartmouth College; Ph.D. (Operations Research), 1968, Cornell University; Stanford University, 1968–1969; Yale University, 1969–1977; AT&T Bell Laboratories, 1977—. At Yale University, from 1973–1977, Mr. Whitt was Associate Professor in the departments of Administrative Sciences and Statistics. At AT&T Bell Laboratories he is in the Operations Research Department. His work focuses on stochastic processes and congestion models.