Applying Optimization Theory to Study Extremal GI/GI/1 Transient Mean Waiting Times

Yan Chen
Industrial Engineering and Operations Research, Columbia University, yc3107@columbia.edu

Ward Whitt
Industrial Engineering and Operations Research, Columbia University, ww2040@columbia.edu

We study the upper bound of the transient mean waiting time in the classical $GI/GI/1$ queue over candidate interarrival-time distributions with finite support, given the first two moments of the interarrival time and the full service-time distribution. We formulate the problem as a non-convex nonlinear program. We calculate the gradient of the transient mean waiting time and then show that a stationary point of the optimization can be characterized by a linear program. In this way, we are able to construct counterexamples to candidate optima and establish necessary conditions and sufficient conditions for stationary points to be three-point distributions or special two-point distributions. We illustrate by applying simulation together with optimization to analyze several examples.

Key words: GI/GI/1 queue, tight bounds, extremal queues, bounds for the transient mean waiting time, moment problem

History: April 4, 2021

1. Introduction

It is often helpful to have a bound on the possible performance in a stochastic performance model given only partial information, which can serve as a useful approximation. A classic example is the mean waiting time in the $GI/GI/1$ queueing model, given the first two moments of the underlying interarrival-time and service-time distributions. For that problem, the Kingman (1962) bound has often been applied; see (5) in §2.

However, that bound is not tight. A long-standing open problem is to determine the tight upper bound of the steady-state mean waiting time and the distributions that attain it, exactly or asymptotically; see Daley et al. (1992), especially §10, Wolff and Wang (2003) and references therein.

Progress on this problem is reviewed in Chen and Whitt (2020a), where algorithms are developed to compute the conjectured upper bound, which is attained asymptotically by two point distributions, where the interarrival-time distribution, denoted by F_0, has one mass at 0, while the service-time distribution, denoted by G_u, has one mass at the upper limit of support M_s, and then M_s is allowed to increase to infinity. A convenient formula is also developed in Theorem 3.2 of Chen and Whitt (2020a) for an upper bound to the conjectured tight upper bound, which provides a good approximation overall, but the main conjecture remains unresolved.

An appealing simple story is developed for higher moments of the $GI/GI/1$ waiting time in Chen and Whitt (2021a) by applying the theory of Tchebycheff systems from Karlin and Studden (1966) and stochastic comparison theory from Rolski (1976) and Denuit et al. (1998). To state them, let F_u and G_0 be defined the same as G_u and F_0 above. For these performance measures and for interarrival-time and service-time distributions with bounded support, Theorems 1 and 3 of Chen and Whitt (2021a) show that the following order relations hold for all n, $1 \le n \le \infty$ ($n = \infty$ means steady-state) and $k \ge 2$:

$$
\begin{aligned}
E[W_n(F_u, G)^k] &\le E[W_n(F, G)^k] \le E[W_n(F_0, G)^k] \quad \text{for all} \quad G, \\
E[W_n(F, G_0)^k] &\le E[W_n(F, G)^k] \le E[W_n(F, G_u)^k] \quad \text{for all} \quad F, \\
E[W_n(F_u, G_0)^k] &\le E[W_n(F, G)^k] \le E[W_n(F_0, G_u)^k] \quad \text{for all} \quad F \quad \text{and} \quad G.
\end{aligned} \tag{1}
$$

Corresponding simple comparison results for the asymptotic decay rate of the steady-state waiting time appear in Chen and Whitt (2020b). (These results require assumptions to avoid heavy tails.)

Unfortunately, this nice story in (1) breaks down for $k = 1$, i.e., for the transient and steady-state mean. Counterexamples to the first two lines of (1) for $n \le \infty$ and the final line for $n < \infty$ were constructed by considering the special case of two point distributions in Chen and Whitt (2021b), extending previous results in §V of Whitt (1984) and §8 of Wolff and Wang (2003).

Partial positive results for the first two lines of (1) with $n = \infty$ (for the steady-state mean) are contained in Theorem 2 of Chen and Whitt (2021a). In particular, the first line of (1) was established for $n = \infty$ when G is completely monotone, i.e., can be represented as a mixture of exponential distributions.

In this paper, we contribute by applying classical optimization theory. In particular, we study the upper bound of the transient mean $E[W_n(F,G)]$ over candidate interarrival-time distributions F assumed to have finite support and specified first two moments, for any given service-time distribution G assumed to have finite second moment. We show that this problem can be represented as a non-convex nonlinear program.

In order to establish counterexamples and to obtain partial positive results, we focus on stationary points of the optimization, as in Proposition 3.1.1 of Bertsekas (2016) (see §4 below). It is well known that any local optimum must be a stationary point. We show that we can test whether or not F_0 (or any other candidate) is a stationary point of the optimization by solving a linear program with an explicit objective function that can easily be estimated by stochastic simulation. In that way, we can construct counterexamples and develop candidates for the optimal distribution. By combining simulation and optimization, in this paper we show that the pair (F_0, G_u) is a stationary point of the optimization for the steady-state mean in numerical examples.

A key step in carrying out this program is calculating the gradient of the transient mean waiting time with respect to the interarrival-time distribution. We do that in §3, after formulating our problem in §2. In §4 we show how this smoothness can be exploited in the optimization. In Lemma 2 there we establish important structure of the objective function. In §5 we develop an abstraction of our optimization problem, so that the results can be applied to related stochastic models. This involves a moment problem over product measures. We then state positive results following from the structure established in Lemma 2. The following §6 is devoted to the proofs. In §7 we study the associated minimization problem, for achieving lower bounds, again when the service-time distribution has a positive pdf. In §8 we show how results for finite support can be extended to other distributions by taking limits. In §9 we give simulation examples. In §10 we draw conclusions.

Before proceeding, we mention other related work. The use of optimization to study the bounding problem for queues seems to have begun with Klincewicz and Whitt (1984) and Johnson and Taaffe

(1990). Due to intractability(e.g., lack of convexity), new approaches have been proposed to simplify the problem, e.g, reformulating the problem into tractable relaxed convex programs, imposing extra conditions and limitations; see Bertsimas and Natarajan (2007) and Gupta and Osogami (2011)). Optimal solutions are not difficult to obtain, but it is difficult to assess the approximation error.

In addition, several researchers have studied bounds for the more complex many-server queue. Bertsimas and Natarajan (2007), Gupta et al. (2010) and Gupta and Osogami (2011) investigate the bounds and approximations of the $M/GI/c$ queue. Gupta et al. (2010) explains why two-moment information is insufficient for good accuracy of steady-state approximations of $M/GI/c$. Gupta and Osogami (2011) establishes a tight bound for the $M/GI/K$ in light traffic. Osogami and Raymond (2013) bounds the transient tail probability of $GI/GI/1$ by a semi-definite program. Li and Goldberg (2017) establishes bounds for $GI/GI/c$ intended for the many-server heavy-traffic regime. van Eekelen et al. (2019) address the classical extremal queueing problem by measuring dispersion in terms of Mean Absolute Deviation (MAD) instead of variance. Finally, we mention that optimization also plays a critical role in recent work on robust queueing, as in Bandi et al. (2015) and Whitt and You (2018, 2019).

2. Formulation

We now formulate our problem. We review the $GI/GI/1$ model in §2.1 and the notation for the spaces of probability measures we consider in §2.2.

2.1. The $GI/GI/1$ Model and the Optimization Problem

The $GI/GI/1$ single-server queue has unlimited waiting space and the first-come first-served service discipline. There is a sequence of independent and identically distributed (i.i.d.) service times $\{V_n : n \geq 0\}$, each distributed as V with cumulative distribution function (cdf) G, which is independent of a sequence of i.i.d. interarrival times $\{U_n : n \geq 0\}$ each distributed as U with cdf F. With the understanding that a 0^{th} customer arrives at time 0, V_n is the service time of customer n, while U_n is the interarrival time between customers n and $n+1$.

Let U have mean $E[U] \equiv \lambda^{-1} \equiv 1$ and squared coefficient of variation (scv, variance divided by the square of the mean) $c_a^2 < \infty$; let a service time V have mean $E[V] \equiv \tau \equiv \rho$ and scv $c_s^2 < \infty$, where $\rho \equiv \lambda\tau < 1$, so that the model is stable. (Let $\equiv$ denote equality by definition.)

Let W_n be the waiting time of customer n, i.e., the time from arrival until starting service, assuming that the system starts with an initial workload W_0 having cdf H_0 with a finite mean. The sequence $\{W_n : n \geq 0\}$ is well known to satisfy the Lindley recursion

$$W_n = [W_{n-1} + V_{n-1} - U_{n-1}]^+, \quad n \geq 1, \tag{2}$$

where $x^+ \equiv \max\{x, 0\}$. Let W be the steady-state waiting time, satisfying $W_n \Rightarrow W$ as $n \to \infty$, where $\Rightarrow$ denotes convergence in distribution for any proper cdf H_0. It is well known that the cdf H of W is the unique cdf satisfying the stochastic fixed point equation

$$W \stackrel{\mathrm{d}}{=} (W + V - U)^+, \tag{3}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. It is also well known that, if $P(W_0 = 0) = 1$, then $W_n \stackrel{\mathrm{d}}{=} \max\{S_k : 0 \leq k \leq n\}$ for $n \leq \infty$, $S_0 \equiv 0$, $S_k \equiv X_0 + \cdots + X_{k-1}$ and $X_k \equiv V_k - U_k$, $k \geq 1$; e.g., It is also known that, under the specified finite moment conditions, for $1 \leq n \leq \infty$, W_n is a proper random variable with finite mean, given by

$$E[W_n] \equiv E[W_n | W_0 = 0] = \sum_{k=1}^{n} \frac{E[S_k^+]}{k} < \infty, \quad 1 \leq n < \infty, \quad \text{and} \quad E[W] = \sum_{k=1}^{\infty} \frac{E[S_k^+]}{k} < \infty; \tag{4}$$

see §§X.1-X.2 of Asmussen (2003) or (13) in §8.5 of Chung (2001). We will exploit the formula for the transient mean in (4) in our analysis. For reference, the Kingman (1962) upper bound is

$$E[W] \leq \frac{\rho^2([(2-\rho)c_a^2/\rho] + c_s^2)}{2(1-\rho)}. \tag{5}$$

We consider the mean waiting time $E[W_n]$ for $1 \leq n \leq \infty$ expressed as a mapping of the underlying distributions; i.e., let

$$w_n(F, G) \equiv E[W_n(F, G)], \quad 1 \leq n \leq \infty, \tag{6}$$

in the $GI/GI/1$ queue with interarrival-time cdf F and service-time cdf G, as given explicitly in (4). The goal is to identify the distribution that yields a tight upper bound over F, given a specification of the cdf G and the first two moments of F. In this paper we assume that the distribution F has bounded support.

In particular, our primary goal is to establish results for the optimization problem

$$\begin{aligned} &\sup \{w_n(F,G) \quad \text{for} \quad \text{for fixed cdf } G \text{ with } E[V]=\rho<1 \\ &\text{such that} \quad \int_0^M u\,dF(u)=1, \quad \text{and} \quad \int_0^M u^2\,dF(u)=(1+c_a^2), \end{aligned} \tag{7}$$

where F is a cdf with support over the bounded interval $[0,M]$. We can use (4) to explicitly write the objective function. However, finding the global optimal solution of (7) is challenging because it is a non-convex nonlinear program with affine constraints. Thus we focus on local optimal solutions, which must be stationary points of the optimization; see §4.

2.2. Notation

Let $\mathcal{P}_n$ be the set of all probability measures on a subset of $\mathbb{R}$ with specified first n moments. We use the scv to parameterize, so let $\mathcal{P}_2 \equiv \mathcal{P}_2(m,c^2)$ be the set of all cdf's with mean m and second moment $m^2(c^2+1)$ where $c^2<\infty$. Let $\mathcal{P}_2(M) \equiv \mathcal{P}_2(m,c^2,M)$ be the subset of all cdf's in $\mathcal{P}_2$ with support in the closed interval $[0,mM]$ having mean m and second moment $m^2(c^2+1)$ where $c^2+1<M<\infty$. (The last property ensures that the set $\mathcal{P}_2(M)$ is non-empty.) We let $\mathcal{P}_2(S) \equiv P_2(1,c^2,S)$ be the set of probability measures for inter-arrival time distribution F on $[0,\infty)$ with two moments specified, as determined by the parameter pair $(1,c^2)$, and support in the set S. For example, if $S=[0,M]$, then we write $P_2(M) \equiv P_2(1,c^2,[0,M])$. If $S=\mathcal{F}$ where $\mathcal{F}$ is a finite set including ending points in $\{0,M\}$, then we write $\mathcal{P}_2(\mathcal{F}) \equiv \mathcal{P}_2(1,c^2,\mathcal{F})$. If S is omitted, i.e., if we write $\mathcal{P}_2 \equiv \mathcal{P}_2(1,c^2)$, then the support is understood to be $[0,\infty)$. Let $\mathcal{P}_{2,k}(S)$ and $\mathcal{P}_{2,k}(S)$ denote the subset with support on at most k points within S for various S as above. To guarantee the $P_2(M)$ being feasible, require $M \geq 1+c^2$. Finally, we introduce notations for specific two-point distributions. Let F_0 denote the two-point cdf with $c^2/(1+c^2)$ on 0 and $1/(1+c^2)$ on $1+c^2$. Let F_u denote another two-point cdf with $(M-1)^2/(c^2+(M-1)^2)$ on $1-c^2/(M-1)$ and $c^2/(c^2+(M-1)^2)$ on M; Typically, we use $P_{a,2}$ and $P_{s,2}$ with notations a and s to denote the sets of probability measures for inter-arrival time and service time.

3. The Gradient of the Transient Mean Waiting Time

In this section we establish smoothness properties of the transient mean waiting time $E[W_n]$ in the $GI/GI/1$ queue as a function of the underlying interarrival-time cdf F for given service-time cdf G. For this purpose, we consider interarrival-time distributions with finite support, but analogs of the following results can be established for cdf's with densities; see Remark 1. The smoothness results here supplement the large literature on continuity of queues, e.g., Whitt (1974) and §X.6 of Asmussen (2003).

For $n \geq 2$, we consider finite support $\mathcal{F}$ in $\mathcal{P}_{a,2}(M_a)$, i.e., $\mathcal{P}_{a,2}(\mathcal{F})$. Let the elements of $\mathcal{F}$ be $0 = u_1 < u_2 < \ldots < u_m = M_a$ with $m \equiv |\mathcal{F}| \geq 3$. With this assumption, we will simplify the notation. In particular, we will suppress the fixed service-time cdf G and we will replace F by its pmf (probability mass function) $p \equiv (p_1, \ldots, p_m)$.

With this new notation, the optimization problem in (7) becomes

$$\max\{w_n(p) \equiv w_n(F,G) \equiv E[W_n(F,G)] : F \in \mathcal{P}_{a,2}(\mathcal{F})\}$$
$$\text{such that} \quad \sum_{i=1}^{m} p_i = 1, \quad \sum_{i=1}^{m} u_i p_i = 1, \quad \sum_{i=1}^{m} u_i^2 p_i = (1+c_a^2) \quad \text{and} \quad p_i \geq 0, \tag{8}$$

where $0 = u_1 < u_2 < \ldots < u_m = M_a$ are the support points in $[0, M_a]$.

We now show that the function $w_n(p)$ in (8) is a smooth function of $p \equiv (p_1, \ldots, p_m)$. In particular, we show that the gradient is well defined. We do that by showing that the Frechet derivative is well defined. For that purpose, let $\|p\|$ be the l_1 norm in $\mathbb{R}^m$, i.e.,

$$\|p\| \equiv \sum_{i=1}^{m} |p_i|. \tag{9}$$

The function $w_n(p)$ is said to be Frechet differentiable if it is Frechet differentiable at each $\hat{p}$. The function $w_n(p)$ is Frechet differentiable at $\hat{p}$ if the following limit as $p \to \hat{p}$ is well defined:

$$\lim_{\|p-\hat{p}\| \to 0} \frac{\left| w_n(p) - w_n(\hat{p}) - \nabla w_n(\hat{p})^t \cdot (p - \hat{p}) \right|}{\|p - \hat{p}\|} = 0, \tag{10}$$

where $\nabla w_n(\hat{p})$ is the gradient of w_n at $\hat{p}$, which we regard as an $m \times 1$ column vector (function of $(u_1, \ldots, u_m)$ in the support of $\hat{p}$),

$$\nabla w_n(\hat{p}) \equiv \left(\left(\frac{\partial w_n}{\partial p_1}(\hat{p}) \right), \ldots, \left(\frac{\partial w_n}{\partial p_m}(\hat{p}) \right) \right)^t \tag{11}$$

with t denoting the transpose of vector in $\mathbb{R}^m$. The gradient is associated with the local linear approximation of $w_n(p)$ at some $\hat{p} \in \mathbb{R}^m$, using the dot product, as

$$w_n(p) \approx w_n(\hat{p}) + \nabla w_n(\hat{p})^t \cdot (p - \hat{p}).$$

REMARK 1. (extension) The Frechet derivative can be generalized to Banach spaces using the total variation metric, which in our setting is just $d_{TV}(p, \hat{p}) = (1/2)\|p - \hat{p}\|$; see Ch. 6 of Serfling (1980) and Wang (1993). For example, the following result also holds if the cdf F has a pdf f over $\mathbb{R}$ instead of having finite support. Then $d_{TV}(F_1, F_2) \equiv \int_0^\infty |f_1(x) - f_2(x)| dx$. However, convergence in the total variation metric is not implied by the usual weak convergence, as in Billingsley (1999).

We now show that the transient mean waiting time in this finite support setting is a smooth function of the interarrival-time pmf p. We show that it is Frechet differentiable and exhibit the gradient and Hessian.

THEOREM 1. (*Frechet derivative*) *For the $GI/GI/1$ queue in the finite support setting above, the function $w_n(p)$ in (8) is Frechet differentiable with partial derivatives at $\hat{p}$ given by*

$$\frac{\partial w_n}{\partial p_i}(\hat{p}) = \sum_{j=1}^{n} E[(\sum_{k=1}^{j} V_k(G) - \sum_{k=1}^{j-1} U_k(\hat{p}) - u_i)^+], \tag{12}$$

so that

$$\nabla w_n(\hat{p})^t \cdot (p - \hat{p}) = \sum_{i=1}^{m} \frac{\partial w_n}{\partial p_i}(\hat{p})(p_i - \hat{p}_i). \tag{13}$$

Higher-order derivatives hold as well. The Hessian matrix H of $w_n(p)$ at $\hat{p}$ given by

$$H(l,k) \equiv \frac{\partial^{(2)} w_n}{\partial p_l \partial p_k}(\hat{p}) = \sum_{j=1}^{n} (j-1) E[(\sum_{k=1}^{j} V_k(G) - \sum_{k=1}^{j-2} U_k(\hat{p}) - u_l - u_k)^+]. \tag{14}$$

Proof. We do the proof of the gradient for $n=2$; the argument for higher n and higher-order differentiation is analogous. For any real-valued functions $f(x)$ and $g(x)$, let $f(x)=\Theta(g(x))$ denote that there exists $m, M>0$ such that $mg(x)\leq |f(x)|\leq Mg(x)$ for all x. Then, adding and subtracting by $\hat{p}_i$ and $\hat{p}_j$ inside the expression for $w_2(p)$, we get

$$\begin{aligned}
w_2(p) &= \sum_i E[(V_1-u_i)^+]p_i + \frac{1}{2}\sum_{i,j} E[(V_1+V_2-u_i-u_j)^+]p_ip_j \\
&= \sum_i E[(V_1-u_i)^+](p_i-\hat{p}_i+\hat{p}_i) + \frac{1}{2}\sum_{i,j} E[(V_1+V_2-u_i-u_j)^+](p_i-\hat{p}_i+\hat{p}_i)(p_j-\hat{p}_j+\hat{p}_j) \\
&= \sum_i E[(V_1-u_i)^+]\hat{p}_i + \frac{1}{2}\sum_{i,j} E[(V_1+V_2-u_i-u_j)^+]\hat{p}_i\hat{p}_j \\
&\quad + \sum_i E[(V_1-u_i)^+](p_i-\hat{p}_i) + \sum_i E[(V_1+V_2-U_1(\hat{F})-u)^+(p_i-\hat{p}_i) + \Theta(\|p-\hat{p}\|^2) \\
&= w_2(\hat{p}) + \sum_i \frac{\partial w_2}{\partial p_i}(\hat{p})(p_i-\hat{p}_i) + \Theta(\|p-\hat{p}\|^2), \qquad (15)
\end{aligned}$$

where

$$\frac{\partial w_2}{\partial p_i}(\hat{p}) = \sum_{j=1}^{2} E[(\sum_{k=1}^{j} V_k(\hat{G}) - \sum_{k=1}^{j-1} U_k(F) - u_i)^+]. \qquad (16)$$

To justify the conclusion in (15), we observe that there exists a constant C such that $E[(V_1+V_2-u_i-u_j)^+]\leq C<\infty$ for all i and j. Consequently, the second term in the second line of (15) associated with the second order of $(p_i-\hat{p}_i)$ can be bounded by the square of the norm, in particular,

$$\begin{aligned}
&\Big|\frac{1}{2}\sum_{i,j} E[(V_1+V_2-u_i-u_j)^+](p_i-\hat{p}_i)(p_j-\hat{p}_j)\Big| \leq C\sum_{i,j}\big|(p_i-\hat{p}_i)(p_j-\hat{p}_j)\big| \\
&\leq C\sum_{i,j}\big|(p_i-\hat{p}_i)\big|\big|(p_j-\hat{p}_j)\big| = C\|p-\hat{p}\|^2.
\end{aligned}$$

Therefore, as $\|p-\hat{p}\|\to 0$,

$$\frac{\big|w_2(p)-w_2(\hat{p})-\sum_i \frac{\partial w_2}{\partial p_i}(\hat{p})(p_i-\hat{p}_i)\big|}{\|p-\hat{p}\|} \leq C\frac{\|p-\hat{p}\|^2}{\|p-\hat{p}\|} = C\|p-\hat{p}\| \to 0.$$

Hence, we have shown that $w_n(p)$ is Frechet differentiable.

Given (12), we continue to take the derivative with respect p_j, so that

$$\begin{aligned}
\frac{\partial^2 w_2}{\partial p_l \partial p_k}(\hat{p}) &= \sum_{j=1}^{2} E[(\sum_{k=1}^{j} V_k(\hat{G}) - \sum_{k=1}^{j-2} U_k(F) - u_l - u_k)^+] \\
&= E[(\sum_{k=1}^{2} V_k(\hat{G}) - u_l - u_k)^+] \qquad (17)
\end{aligned}$$

Therefore,

$$H(l,k) = \sum_{j=1}^{n} (j-1) E[(\sum_{k=1}^{j} V_k(G) - \sum_{k=1}^{j-2} U_k(\hat{p}) - u_l - u_k)^+], \quad (18)$$

which supports (14). ■

4. Exploiting the Smoothness for Optimization

We now show how to exploit the smoothness established in Theorem 1 in order to establish partial results for the optimization problem formulated in (7) and (8). First, we observe that there exists a global optimum because we are maximizing a continuous function over a compact subset of R^m.

4.1. Necessary Condition for a Local Optimum: a Stationary Point

Recall that a point $\hat{p}$ is a local optimum for (8) if there exists $\delta > 0$ such that

$$w_n(p) \le w_n(\hat{p}) \quad \text{for all} \quad p \quad \text{such that} \quad \|p - \hat{p}\| < \delta. \quad (19)$$

Clearly, there exists at least one local optimum because the global optimum is necessarily a local optimum. We apply the following necessary condition for a local optimum from Proposition 3.1.1 of Bertsekas (2016).

PROPOSITION 1. (*necessary condition for a local optimum, Proposition 3.1.1 of Bertsekas (2016)*) *If $\hat{p}$ is a local optimum of $w_n(p)$, then*

$$\nabla w_n(\hat{p})^t \cdot (p - \hat{p}) \le 0 \quad \textit{for all} \quad p \in \mathcal{P}_{a,2}(\mathcal{F}). \quad (20)$$

If there exists $\hat{p}$ satisfying (20), then $\hat{p}$ is called a stationary point (of the optimization).

It will be convenient to look at the partial derivatives in (12) as a function of the support point u. Hence, we define

$$\phi_a(u) \equiv \phi_a(u;\hat{p}) \equiv \frac{\partial w_n}{\partial p_i}(\hat{p})(u) \equiv \sum_{i=1}^{n} E[(\sum_{k=1}^{i} V_k(G) - \sum_{k=1}^{i-1} U_k(\hat{p}) - u)^+], \quad u \ge 0. \quad (21)$$

COROLLARY 1. (*the key linear program*) *The pmf $\hat{p}$ is a stationary point, satisfying (20), if and only if $\hat{p}$ is the solution of the linear program (LP)*

$$\sup\{\nabla w_n(\hat{F})^t \cdot p \equiv \sum_{i=1}^{m} \frac{\partial w_n}{\partial p_i}(\hat{p}) p_i \equiv \sum_{i=1}^{m} \phi_a(u_i) p_i : p \in \mathcal{P}_{a,2}(\mathcal{F})\}. \quad (22)$$

i.e., if and and only if

$$\sup\{\sum_{i=1}^{m}\phi_a(u_i)p_i : p\in\mathcal{P}_{a,2}(\mathcal{F})\}=\sum_{i=1}^{m}\phi_a(u_i)\hat{p}_i. \tag{23}$$

4.2. Applications of Corollary 1 to F_0 given G_0

We now apply Corollary 1 to study the special two-point interarrival-time distribution F_0, for the case $G\equiv G_0$, which is the counterexample for the steady-state mean from §8 of Wolff and Wang (2003). We consider two cases, one designed to approximately represent steady state and one to be genuinely transient. The nearly-steady-state example has $n=40,\rho=0.1,c_a^2=c_s^2=0.5,M_a=10$. The support contains 501 points in $[0,10]$ (including the endpoints) so that, F_0 is in the support, while the transient example has $n=4,\rho=0.7,c_a^2=c_s^2=0.5,M_a=10$.

In both cases we apply simulation to estimate the objective function in (21) when $G=G_0$ and $F=F_0$ and then solve the linear program in (22). We perform 5 independent replications, so that we can estimate 95% confidence intervals (denoted by CIL). In each replication, use a large sample size such as 10^6, so that the randomness in the objective function can be ignored. When we do the optimization, we always find that the solution has support on at most three points, so that there is little ambiguity. Figure 1 shows the estimates of the objective function $\phi_a(u)$ in (21) for the two experiments with (F_0,G_0).

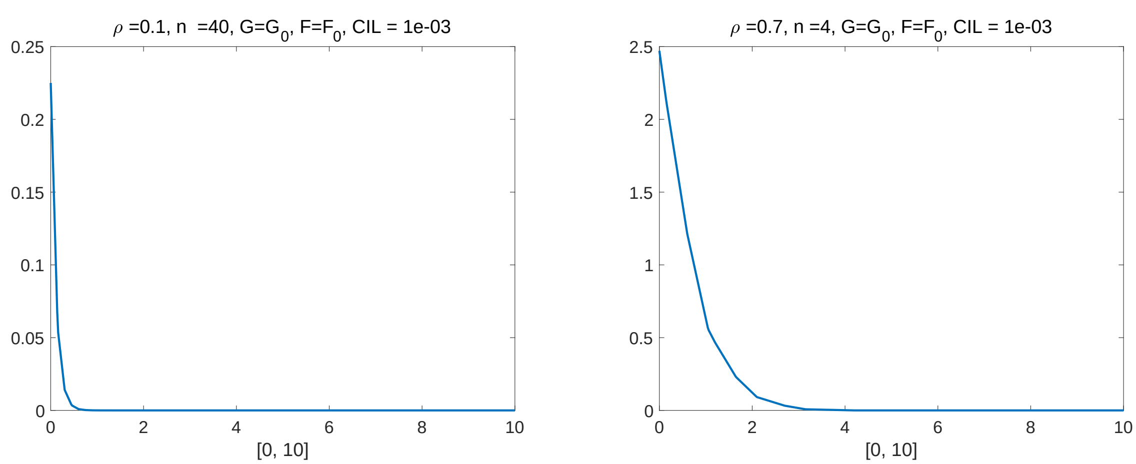

Figure 1 **Simulation estimates of the objective function $\phi_a(u)$ in (21) over $[0,10]$ for the $F_0/G_0/1$ model with $n=40,\rho=0.1,c_a^2=c_s^2=0.5,M_a=10$ (left) and with $n=4,\rho=0.7,c_a^2=c_s^2=0.5,M_a=10$ (right), based on 5 replications of 10^6 arrivals.**

When we carry out this simulation+optimization program for the other service-time distributions considered in the examples of §9, we find that F_0 is always a stationary point. However, for G_0, for the example with $n=4$, we find that F_0 is not the solution of the linear program. In particular, the solution F^* of the linear program has masses $0.3423, 0.3242, 0.3333$ on $0.020, 1.500, 1.520$, respectively. Hence, F_0 is not a stationary point. As a consequence, F_0 is not locally optimal, and thus not optimal. On the other hand, for the nearly-steady-state example with $n=40$, we find that F_0 is a stationary point, even though we know that it is not optimal. (A stationary point need not be locally optimal.) This shows that there may be more than one stationary point.

We also considered our associated numerical study over two-point distributions in Chen and Whitt (2021b). Tables 2 and 3 there display the mean waiting times $E[W]$ and $E[W_{20}]$ for two-point distributions F and G. These tables confirm the counterexample in §8 of Wolff and Wang (2003) for the case $c_a^2 = c_s^2 = 4.0, \rho = 5$ for $n=20$ and steady-state. We first applied Corollary 1 to (F_0, G_0). With a spacing of 0.25 between points in the support of F, we found that the optimal solution of the linear program had masses of $0.8767, 0.0833, 0.0400$ on $0.25, 6.25, 6.50$, respectively. Hence, F_0 is not a stationary point. Moreover, starting from the optimal solution F^{2*} among the two-point distributions shown in Table 3, which has one mass on 5.25, we find that it too is not a stationary point. We found that the optimal solution of that linear program had masses of $0.4525, 0.3810, 0.1645$ on $0.00, 0.25, 5.50$, respectively. Thus, we conclude that neither F_0 nor the optimal two-point cdf F^{2*} is a stationary point, and thus neither is optimal overall.

4.3. Extending the Class of Counterexamples

We next show that a variant of the counterexample in §4.2 holds for service-time cdf's with a positive pdf, as will be assumed in Lemma 2 below. We first establish the following basic property.

Lemma 1. *The objective function $\phi_a(u;\hat{p},G)$ (21) is uniformly bounded and continuous as a function of candidate G, $\hat{p}$ and u.*

Proof. Note that

$$0 \le \phi_a(u) \le \sum_{i=1}^{n} iE[V] \le n(n+1)\rho/2. \blacksquare \tag{24}$$

Corollary 2. *(extension for $G_n \Rightarrow G$) Suppose that $G_n \Rightarrow G$ as $n \to \infty$ and $\hat{p}_n$ is a stationary point of the optimization for $\phi_a(u; G_n)$ for $n \geq 1$. Then there exists a convergent subsequence of $\{\hat{p}_n : n \geq 1\}$ and the limit of any such convergent subsequence is a stationary point of the optimization for $\phi_a(u; G)$.*

Equivalently, if $\hat{p}$ is not a stationary point for $\phi_a(u; G)$ and if $G_n \Rightarrow G$ as $n \to \infty$, then, for all sufficiently large n, $\hat{p}$ is not a stationary point of $\phi_a(u; G_n)$.

Proof. Since the space $\mathcal{P}_{a,2}(\mathcal{F})$ is a compact metric space, there exists a convergent subsequence of $\{\hat{p}_n : n \geq 1\}$. Suppose that the limit is $\hat{p}$. By continuity, $\hat{p}$ must be a stationary point of the optimization for $\phi_a(u; G)$.■

4.4. Stronger Conclusions about Optimality from the Hessian

Stronger conclusions about global optimality can be obtained from the Hessian. Even though we do not exploit the Hessian in this paper, we state the result for future reference. See Appendix A.4 on p. 760, §1.1.2 on p. 15 and §3.1.11 on p. 252 of Bertsekas (2016) for background.

Proposition 2. (*sufficient condition for local and global optimality*) *Consider the Hessian matrix H from Theorem 1 for $GI/GI/1$ queue with the specified $G \in \mathcal{P}_{s,2}$.*

(a) If $-H$ is a positive semi-definite matrix for all $F(p) \in \mathcal{P}_{a,2}(\mathcal{F})$, then the program (8) is a convex program, so that there exists a unique global optimal distribution which is also the stationary point.

(b) If $-H$ is positive semi-definite matrix for some specific $F(\hat{P}) \in \mathcal{P}_{a,2}(\mathcal{F})$ and the $\hat{p}$ satisfies (20), then the $\hat{p}$ will be a local optimal distribution.

4.5. Structural Properties of the Objective Function

We next establish structural properties of the objective function in (21) and (22) regarded as a function of u over the interval $[0, M_a]$.

Lemma 2. (*structure of the objective function in* (21)) *If the fixed cdf G of V has a positive pdf g over $[0, \infty)$, then the random variable $Y_i \equiv \sum_{k=1}^{i} V_k - \sum_{k=1}^{i-1} U_k$ has a cdf Γ_i with support in*

$[-(i-1)M_a,\infty)$ which has a positive pdf γ_i over $[0,\infty)$ for each i, $1\le i\le m$. Hence, for $x>0$, the cdf of Y_i can be expressed by

$$\Gamma_i(x)=\Gamma_i(0)+\int_0^x \gamma_i(y)\,dy \quad \textit{for} \quad x\ge 0, \tag{25}$$

so that the function ϕ_a in (21) can be expressed as

$$\phi_a(u)\equiv \frac{\partial w_n}{\partial p}(\hat{p})=\sum_{i=1}^{n}\int_0^\infty (x-u)^+\gamma_i(x)\,dx>0, \quad u\ge 0. \tag{26}$$

Hence, $\phi_a(u)>0$ and the first two derivatives of ϕ_a in (21) exist for $u>0$ and satisfy

$$\dot{\phi}_a(u)=\sum_{i=1}^{n}(\Gamma_i(u)-1)<0, \quad \ddot{\phi}_a=\sum_{i=1}^{n}\gamma_i(u)>0, \quad u\ge 0. \tag{27}$$

Thus, ϕ_a is continuous, strictly decreasing and strictly convex on $[0,M_a]$.

Proof. We directly calculate the derivative of $\phi_a(u)$ in (21) term by term. Since the random variable V with cdf G has a positive pdf, so does Y_i for each i; see §V.4 of Feller (1971). To calculate the derivative of each term in the sum, we apply the Leibniz integral rule for differentiation of integrals of integrable functions that are differentiable almost everywhere. Each term involves the positive part function $(x)^+\equiv\max\{x,0\}$. Observe that the derivative of $(x-u)^+\gamma_i(x)$ with respect to u is $-\gamma_i(x)$ for $u<x$. That implies that

$$\dot{\phi}_a(u)=-\sum_{i=1}^{n}\int_u^\infty \gamma_i(x)\,dx=\sum_{i=1}^{n}(\Gamma_i(u)-1). \tag{28}$$

The rest follows directly. ∎

Going forward, we will see that the extremal distributions will depend on the structure

$$\ddot{\phi}_a=\sum_{i=1}^{n}\gamma_i(u) \tag{29}$$

where γ_i is the pdf of Y_i and we define

$$Y_i\equiv\sum_{k=1}^{i}V_k-\sum_{k=1}^{i-1}U_k. \tag{30}$$

We will establish concrete results in the next section.

4.6. Maximizing over G for Fixed F

It is evident that we obtain comparable results when we maximize over G for fixed F. First, we observe that an analog of Lemma 2 arises if we consider the dual problem of optimizing over the cdf G given fixed F, assuming that we impose corresponding regularity conditions. Even though the optimization problem (8) and its gradient vectors are changed if inter-arrival time distribution F is given. But we can exploit a reverse-time representation for the service time G to yield the same structure. For that purpose, let

$$\phi_s(\tilde{v}) \equiv \sum_{i=1}^{n} E[(\sum_{k=1}^{i-1} V_k(\hat{q}) - \sum_{k=1}^{i} U_k(F) + \rho M_s - \tilde{v})^+]. \tag{31}$$

LEMMA 3. (*structure of the objective function in* (31)) *If the fixed cdf F of U has a positive pdf f over $[0,\infty)$, then $Z_i \equiv \sum_{k=1}^{i-1} V_k(\hat{q}) - \sum_{k=1}^{i} U_k(F) + \rho M_s$ has support in $(-\infty, \rho M_s + (i-1)a]$, where $a > 0$ is the upper limit of the support of V. Thus Z_i has a positive pdf θ_i over $(-\infty, \rho M_s]$ for each i, $1 \le i \le m$. Hence,*

$$\begin{aligned}\phi_s(\tilde{v}) &= \sum_{i=1}^{n} E[(Z_i - \tilde{v})^+] = \sum_{i=1}^{n} \left(E[Z_i - \tilde{v}] - E[(Z_i - \tilde{v})^-]\right), \\ &= \sum_{i=1}^{n} \left(E[Z_i - \tilde{v}] - \int_{-\infty}^{\rho M_s} (x - \tilde{v})^- \, d\Theta_i(x)\right),\end{aligned} \tag{32}$$

where

$$\Theta_i(x) = \int_{-\infty}^{x} \theta_i(y)\, dy \quad for \quad x \le \rho M_s, \tag{33}$$

so that, paralleling Lemma 2, the first two derivatives of $\phi_s(\tilde{v})$ are

$$\dot{\phi}_s(\tilde{v}) = \sum_{i=1}^{n} (\Theta_i(\tilde{v}) - 1)) < 0 \quad and \quad \ddot{\phi}_s(\tilde{v}) = \sum_{i=1}^{n} \theta_i(\tilde{v}) > 0, \quad \tilde{v} \in [0, \rho M_s]. \tag{34}$$

Thus $\phi_s(\tilde{v})$ in (31) *is a continuous, strictly positive, strictly decreasing and strictly convex function on $[0, \rho M_s]$.*

Proof. Just as in Lemma 2, we differentiate the integral to go from (32) to (34). For each term in the sum for $\dot{\phi}_s(\tilde{v})$, we get -1 from the first term in (32) and $\Theta_i(\tilde{v})$ from the second. ∎

With Lemma 3, the rest of the proof for optimization over G for fixed F can use the same detailed argument used for optimization over F for fixed G However, we must recall the change of variables made in (31). For example, 0 appears in the extremal cdf for F if and only if ρM_s appears in the extremal cdf for G.

After exploiting the reverse-time representation, the shape of $\phi_s(.)$ and $\phi_a(.)$ are identical, so that the all results holding for these two cases will be the same.

4.7. Applications of Corollary 1 to G_u given F_0

We now apply Corollary 1 to study the special two-point service-time distribution G_u, for the case $F \equiv F_0$, which is a natural candidate upper bound overall. From §V of Whitt (1984) and Theorem 2 of Chen and Whitt (2021a), we know that the extremal G for given F, is more complicated, depending critically on the shape of F. However, the accumulated evidence indicates that G_u is optimal given F_0 for the steady-state mean. For example, in §3.3 of Chen and Whitt (2021b), we found that G_u was optimal within two-point distributions for steady state, but not for the transient mean. For the transient mean, we found that the optimal was obtained at distributions $G_{u,n}$, where the upper mass point converges to M_s as $n \to \infty$.

Consistent with that numerical experience, we find that G_u is a stationary point for the nearly-steady-state example with $n = 40, \rho = 0.1, c_a^2 = c_s^2 = 0.5, \rho M_s = 10$, while it is not in the transient example with $n = 4, \rho = 0.7, c_a^2 = c_s^2 = 0.5, \rho M_s = 10$.

Considering the joint optimality over (F, G), from this numerical analysis we find that (F_0, G_u) is a stationary point of the optimization in the nearly nearly-steady-state example with $n = 40, \rho = 0.1, c_a^2 = c_s^2 = 0.5, M_a = 10, \rho M_s = 10$, whereas it is not for the transient example with $n = 4, \rho = 0.7, c_a^2 = c_s^2 = 0.5, M_a = 10, \rho M_s = 10$.

5. An Abstraction to a Multi-Dimensional Moment Problem

We now abstract the queueing problem we have considered so far to provide a framework that can be used for other stochastic models in addition to the $GI/GI/1$ transient mean waiting time. We show that our problem can be regarded as a special case of a multi-dimensional moment problem. That generalization leads to extensions of the functions $\phi_a(u)$ in (21) and $\phi_s(\tilde{v})$ in (31). We will then identify structure needed for these functions, in addition to the structure established in Lemmas 2 and 3, is needed in order to characterize the solutions of the optimization problems.

Our abstraction extends the classical moment problem, which was established many years ago, as reviewed in Birge and Dula (1991), Smith (1995) and other references therein. A version (special case) of the classical moment problem is the optimization

$$\max\ E[\hat{g}(X_1)] \equiv \int_0^M \hat{g}(x_1)dF_1(x_1)$$
$$\text{subject to } F_1 \in \mathcal{P}_2(M) \tag{35}$$

where $\hat{g}$ is a real-valued continuous function defined on $[0, M] \to R$ and X_1 is a random variable distributed as F_1 where F_1 lies in the domain $\mathcal{P}_2(M)$ with fixed first two moments and bounded support M, which is thus convex and compact. The classical moment problem in our setting is a convex program over a compact domain and it has been shown that there always exists an optimal distribution F_1^* in $\mathcal{P}_{2,3}(M)$; i.e., with all mass on at most three points.

5.1. A Moment Problem Over Product Measures

In this paper we consider a similar moment problem for a continuous objective function $\hat{g}$ over independent random variables with a specified common marginal distribution; i.e., over random vectors $(X_1, \ldots, X_n)$, where X_i are independent random variables with a common marginal cdf's F. The new formulation is

$$\max\ E[\hat{g}(X_1, \ldots, X_n)] \equiv \int_0^M \hat{g}(x_1, \ldots, x_n)dF_1(x_1)\ldots dF_n(x_n) \tag{36}$$
$$\text{subject to } F_1 = F_2 = \ldots = F_n \in \mathcal{P}_2(M)$$

where $\hat{g}(x_1, \ldots, x_n)$ is a nonnegative continuous real-valued function defined on the product space $[0, M]^n$ with $M \geq 1 + c^2$. In (36) the common marginal distribution has specified first two moments. The program formulation in (36) has many applications such as robust estimations in tail analysis and rare-event simulation problems. Lam and Mottet (2015) and Lam and Mottet (2017) propose the reformulation as (36) when setting g to be indicator function. That implies that we consider some positive b, one is interested in solving

$$\max P_F(X_1 + \ldots + X_n \geq b) = \int_0^M 1_{\{x_1+\ldots+x_n \geq b\}} dF_1(x_1)\ldots dF_n(x_n)$$
$$\text{subject to } F_1 = F_2 = \ldots = F_n \in \mathcal{P}_2(M)$$

where all X_i are independent and are distributed as a same unknown distribution F and F lies in an uncertain set with unspecified tail. Also, we might extend to any utility function $g(x_1,\ldots,x_n)$ over the product measure space.

As in §3, we restrict attention to probability distributions with finite support. We assume that all $F\in\mathcal{P}_2(M)$ have the common finite support $\mathcal{F}$ with elements $0=u_1\leq\ldots\leq u_m=M$ with sufficient large m. So that we have the following alternative formulation for (36),

$$\begin{aligned}&\max\ g(p)\ \equiv \sum_{i_1,\ldots,i_n}\hat{g}(u_{i_1},\ldots,u_{i_n})p_{i_1}\ldots p_{i_n}\\&\text{subject to}\sum_{i=1}^m p_i=1,\sum_{i=1}^m u_ip_i=1,\sum_{i=1}^m u_i^2p_i=1+c^2\quad\text{and}\quad p_i\geq 0,\end{aligned}\tag{37}$$

where $g:\mathcal{P}_2(\mathcal{F})\rightarrow R$ and $\mathcal{P}_2(\mathcal{F})$ is a compact and convex subset of R^m.

5.2. Sufficient Conditions to be a Stationary Point

We clearly have a generalization of the linear program in Corollary 1 with the objective function $\phi_a(u)$ in (21) replaced by a new function

$$\psi(u)\equiv\frac{\partial g}{\partial p_i}(\hat{p})(u)\tag{38}$$

It suffices to check the optimality for

$$\max\{\sum_{i=1}^m\psi(u_i)p_i\equiv\nabla g(\hat{p})^tp,\ p\in\mathcal{P}_2(\mathcal{F})\}=\nabla g(\hat{p})^t\hat{p}.\tag{39}$$

As regularity conditions we require the properties deduced for ϕ_a in Lemma 2, but we also an extra condition on the second derivative $\ddot{\psi}$.

We apply duality theory for the LP in (39). From basic LP duality theory as in Ch. 4 of Bertsimas and Tsitsiklis (1997), the dual problem associated with the LP in (39) is to find the vector $\lambda^*\equiv(\lambda_0^*,\lambda_1^*,\lambda_2^*)$ that attains the minimum

$$\begin{aligned}&\min\{\lambda_0+\lambda_1+\lambda_2(1+c^2)\}\\&\text{such that}\quad r(u_i)\equiv\lambda_0+\lambda_1u_i+\lambda_2u_i^2\geq\psi(u_i)\quad\text{for all}\quad i,\quad 1\leq i\leq m.\end{aligned}\tag{40}$$

We are now ready to state the results obtained in this paper. Our first theorem establishes sufficient conditions for any specific stationary point to be a three-point distribution. For the queueing problem, Lemma 2 shows that these conditions are satisfied if the fixed service-time cdf G has a positive pdf.

THEOREM 2. (*sufficient condition for a stationary point* $\hat{p}$ *to be a three-point distribution*) *We make the following initial three assumptions for the optimization problem in* (37)-(39)*:*

(*i*) *The objective function* $g(p)$ *in* (37) *is Frechet differentiable at all* $p \in \mathcal{P}_2(\mathcal{F})$*.*

(*ii*) $\psi(u)$ *in* (38) *is a strictly convex, strictly positive and strictly decreasing function over* $[0, M]$.

(*iii*) $\psi(u)$ *is twice differentiable and the second derivative* $\ddot{\psi}(u)$ *is a smooth function over* $[0, M]$.

For any stationary point $\hat{p}$ *of* (37)*, the LP given* $\hat{p}$ *in* (39) *has a unique optimal solution, which is thus an extreme point, and is thus a three-point distribution, if and only if the quadratic function* $r(u)$ *in* (40) *has at most three intersection with* $\psi(u) \equiv \psi(u;\hat{p})$ *over* $[0, M]$*.*

Our next theorem establishes sufficient conditions for one of the special two-point distributions F_0 or F_u to be a stationary point of the optimization. For the shape of $\ddot{\psi}(u)$, we introduce the following strong from of unimodality.

DEFINITION 1. (single peak) A nonnegative continuous function $f:[0,M] \rightarrow \mathbb{R}$ is said to have a single peak if its maximum value is achieved uniquely at an interior point $\hat{u}$ and if f is monotone increasing over $[0, \hat{u}]$ and monotone decreasing over $[\hat{u}, M]$.

THEOREM 3. (*sufficient conditions for* F_0 *or* F_u *to be a stationary point*) *Under the same initial three assumptions as Theorem 2,*

(a) For any candidate cdf F*, if* $\ddot{\psi}(u;F)$ *is strictly decreasing or has a single peak over* $[0, M]$*, then* F_0 *must be a solution of the LP in* (39)*. Hence, if this condition is satisfied for* $F = F_0$*, then* F_0*. must be a stationary point.*

(b) Similarly, for any candidate cdf F*, if* $\ddot{\psi}(u;F)$ *is strictly increasing over* $[0, M]$*, then* F_u *must be a solution of the LP in* (39)*. Hence, if this condition is satisfied for* $F = F_u$*, then* F_u*. must be a stationary point.*

Corollary 3. (*sufficient conditions for F_0 or F_u to be a global optimum*) *Under the same initial three assumptions as Theorem 2, if $\ddot{\psi}(u;F)$ satisfies the specified conditions for all $F \in \mathcal{P}_2(\mathcal{F})$, then the identified stationary points in Theorem 3 provide the unique global optimal solution.*

We can also extend to other shapes for $\ddot{\psi}$ using the following generalization of Definition 1.

Definition 2. (multiple peaks) A nonnegative continuous function $f : [0, M] \to \mathbb{R}$ is said to have n peaks if it has n unique interior local maximum points and it is monotone increasing before the first maximum point and then thereafter the function is first monotone decreasing and then monotone increasing between each adjacent two peaks before the final maximum point. Then the function is monotone decreasing after the final maximum point.

Theorem 4. (*more structures*) *Under the setting of Theorem 3. If $\ddot{\psi}(u;F)$ has at most n $(1 \leq n < \infty)$ peaks over $[0, M]$ for any candidate $F \in \mathcal{P}_2(\mathcal{F})$, then all stationary points of the optimization in (39) must lie in $\mathcal{P}_{2,n+1}(\mathcal{F})$.*

6. Proofs

We now prove the results above.

6.1. Proof of Theorem 2

We first show the necessary condition, and then the sufficient condition.

Necessary Condition: Starting with $\hat{p}$ being a stationary point satisfying the condition that $r(u)$ has at most three intersection point with $\psi(u;\hat{p})$, the main goal is to show such (39) has a unique solution, so that the $\hat{p}$ must be an extremal point. For that purpose, we apply the following lemma, which is Corollary 1 to Theorem 4 in Tijssen and Sierksma (1998).

Lemma 4. (*non-degeneracy and uniqueness in LP*) *A standard LP has a unique optimal solution if and only if its dual has a non-degenerate optimal solution.*

To apply Lemma 4 from Corollary 1 to Theorem 4 in Tijssen and Sierksma (1998), we express the dual (40) in standard form by introducing slack variables and dividing the three variables λ_i into their positive and negative parts as

$$\min \{(\lambda_0^+ - \lambda_0^-) + (\lambda_1^+ - \lambda_1^-) + (\lambda_2^+ - \lambda_2^-)(1 + c^2)\}$$

$$\text{such that}\quad (\lambda_0^+ - \lambda_0^-) + (\lambda_1^+ - \lambda_1^-)u_i + (\lambda_2^+ - \lambda_2^-)u_i^2 + s_i = \psi(u_i) \quad \text{for all} \quad i, \quad 1 \le i \le m,$$

$$\text{and} \quad \lambda_j^+ \ge 0, \lambda_j^- \ge 0, 1 \le j \le 3; \quad s_i \ge 0, 1 \le i \le m. \tag{41}$$

In the setting of (41), we have $m+6$ variables and m equality constraints. To show that there exists a non-degenerate optimal solution, will show that at least one among $(\lambda_i^+, \lambda_i^-)$ for $i = 0,1,2$ are not equal to be zero, e.g., $\lambda_0^+ > 0, \lambda_1^- > 0$ and $\lambda_2^+ > 0$, while $\lambda_0^- = 0$, $\lambda_1^+ = 0$ and $\lambda_2^- = 0$. That is equivalent to show all λ_i^* in (40) are not equal to zero. We will achieve the goal by establishing Lemma 5 below.

Hence, when at most three of the slack variables s_i are 0 (at most three intersection points), the dual problem has a non-degenerate solution solution, thus the $\hat{p}$ will be the unique solution in (39) and $\hat{p}$ must be in $\mathcal{P}_{2,3}(\mathcal{F})$.

Lemma 5. (*non-degeneracy for the dual*) *Consider the dual formulation (40), for any optimal dual solution* $(\lambda_0^*, \lambda_1^*, \lambda_2^*)$ *associated with* $\hat{p}$, λ_i^* *for* $i = 0,1,2$ *can not be zero.*

From (40), we see that the constraints produce the quadratic function $r(u)$ that is required to dominate $r(u)$ for all $u \in \mathcal{F}$. We exploit the structure of the function $\psi(u)$ in (39) from regulations and assumptions. Under the condition, $\psi(u)$ is continuous, strictly positive, strictly decreasing and strictly convex. Recall that we are working with standard LP's, where the cdf F has finite support set $\mathcal{F}$, but the support set $\mathcal{F}$ always contains the two endpoints 0 and M.

The inequality constraints in (40) are only required to hold at the finitely many point in the support set $\mathcal{F}$. Even though we exploit the structure of continuous functions, the following argument applies to any finite support set.

If $M = 1 + c^2$, the second moment, which is the lower limit of the support, then the primal has the unique feasible, and thus optimal, two-point feasible distribution with masses on 0 and $1 + c^2$. So henceforth assume that $M > 1 + c^2$ as well.

We start knowing that both the dual LP (40) and the primal LP (39) have feasible solutions and the feasible region of the primal LP (39) is compact, thus they both have at least one optimal solution. We will show that the primal LP (39) has a unique solution by applying Lemma 4 and showing that no optimal solution of the dual (40) can be degenerate.

Hence, we will show (i) that we cannot have the optimal λ_i^* be 0 for any i in the (40).

We start with the λ_i^*. First, we must have $\lambda_0 \geq \psi(0) > 0$, so we cannot have $\lambda_0^* = 0$. Next, suppose that $\lambda_1 = 0$. In this setting, with $\lambda_0^* > 0$ and $\lambda_1^* = 0$, if $\lambda_2^* \geq 0$, then r can intersect $\psi(u)$ only at 0, which cannot correspond to a feasible solution of the primal. (We exploit complementary slackness here and in the following.) On the other hand, if $\lambda_2^* < 0$, then $\psi(u)$ can only intersect ψ at the two endpoints, without violating the conditions at the endpoints, but that does not correspond to a feasible solution of the primal, assuming that $M > 1 + c^2$. Hence, we cannot have a degenerate optimal solution with $\lambda_1^* = 0$. Finally, suppose that $\lambda_2^* = 0$, which makes ψ linear. If $\lambda_0 = \psi(0) > 0$, then again ψ can only meet $\psi(u)$ at the two endpoints without violating the conditions at the endpoints, but that does not correspond to a feasible solution of the primal, assuming that $M > 1 + c^2$. Otherwise, r can only have one intersection point with $\psi(u)$ (as we have done).

Sufficient Condition: To prove the sufficient condition, if $\hat{p}$ is the unique optimal solution for (39) which must be $\in \mathcal{P}_{2,3}(\mathcal{F})$, by Strict Complimentary Slackness Condition in LP, the optimal distribution can be identified from the solution to the LP, so that such ψ and r has at most three intersection points over $[0, M]$ which corresponds to the same points having positive masses in $\hat{p}$.

6.2. Proof of Theorem 3

We now consider the LP (39) based on an objective function determined by a cdf F under the conditions of Theorem 3. In each case we will show that the LP (39) has a unique optimal solution and the unique optimal solutions will be the specified special two-point distributio.

We first do the proof for (a) and then (b). For (a), we first establish the claim for only one unique interior intersection point and then the claim for F_0.

The argument for the single peak case is essentially same as that for the strictly monotone decreasing case. So we do the proof for the both two cases together.

We first show that at most one of the internal inequality constraints for $0 = u_1 < u_i < u_m = M$ can be satisfied as equalities if ψ is strict monotone (strictly decreasing or strictly increasing) or has a single peak. For any interior intersection point u where $r(u) = \psi(u)$, according to (40), we also have

$$\begin{aligned}
\ddot{r}(u) &= 2\lambda_2^* = \ddot{\psi}(u),\\
\dot{r}(u) &= \lambda_1^* + 2\lambda_2^* u = \dot{\psi}(u),\\
r(u) &= \lambda_0^* + \lambda_1^* u + \lambda_2^* u^2 = \psi(u). \qquad (42)
\end{aligned}$$

We first assume that equalities are obtained at the two interior points x, y, where $0 < x < y < M$ and show that produces a contradiction. Since x, y are interior intersection points,

$$\begin{aligned}
\ddot{r}(x) &= 2\lambda_2^* = \ddot{\psi}(x), \dot{r}(x) = \dot{\psi}(x), r(u) = \psi(x),\\
\ddot{r}(y) &= 2\lambda_2^* = \ddot{\psi}(y), \dot{r}(y) = \dot{\psi}(y), r(y) = \psi(y). \qquad (43)
\end{aligned}$$

Looking at the differences of these derivatives, we obtain

$$2\lambda_2^* = \frac{\dot{\psi}(y) - \dot{\psi}(x)}{y - x} = \ddot{\psi}(x) = \ddot{\psi}(y). \qquad (44)$$

Therefore, by Mean Value Theorem, there exists $\tilde{u} \in (x, y)$ such that $\ddot{\psi}(\tilde{u}) = 2\lambda_2^*$. That leads to a contradiction because such $\ddot{\psi}(u)$ can only have at most two intersection points with $2\lambda_2^*$.

Assume the only one interior intersection point is y, we next show the $\psi(u)$ and $r(u)$ can not intersect at $u = M$.

Recall at the point y, we must have

$$2\lambda_2^* = \ddot{\psi}(y), \dot{r}(y) = \dot{\psi}(y), r(y) = \psi(y). \qquad (45)$$

Since $r(u) > \psi(u)$ for $u \in (y, M)$, then $2\lambda_2^* > \ddot{\psi}(u)$ for $u \in (y, y+\delta)$ for some small $\delta > 0$. Therefore, given the shape of $\ddot{\psi}(u)$, the point y must be the final intersection point for $\ddot{\psi}(u)$ and $2\lambda^*$. After $u > y$, since $2\lambda^* > \ddot{\psi}(u)$ ($\ddot{\psi}$ has a single peak or is strictly monotone decreasing), that implies the $\psi(u) < r(u)$ for all u so that they can not intersect again at $u = M$.

The only remaining possible case is that the ψ and r will intersect at 0 and an interior point $b \in (0, M)$. By Strict Complementary Slackness Condition in LP, the optimal distribution can be identified from the solution to the LP. So that the optimal distribution only has the positive mass on 0 and b. A two-point distribution which has one mass at 0 must be F_0.

Essentially the same argument applies in part (b), but now the two-point distribution must have one inner point and mass at the upper end point M, which corresponds to the claimed F_u. ■

6.3. Proof of Theorem 4

Paralleling with lines before (44) in the proof of Theorem 3, given the number of peaks equal to $n \geq 2$, we can first show the number of interior intersection points between ψ and r is at most n. Then paralleling the arguments after (45), since the first intersection point of ψ and r must be the second intersection point between $\ddot{\psi}$ and $\ddot{r}$, the ψ and r will not intersect at M. With at most n interior intersection points and possible additional one intersection point at 0, the total number intersection points between ψ and r is at most $n+1$. Therefore, the optimal distribution in $\mathcal{P}_{n,n+1}(M)$.

7. The Associated Minimization Problem

We now consider the associated minimization problem, which corresponds to the supremum in (7), (22), (23) being replaced by an infimum, and then the associated maximum in (8), (35), (36), (37) and (39) being replaced by a minimum. Then the inequality must be reversed in the inequality (20) which expresses the definition of a stationary point. Accordingly the associated dual problem in (40) becomes finding the vector $\lambda^* \equiv (\lambda_0^*, \lambda_1^*, \lambda_2^*)$ that attains the maximum

$$\begin{aligned} &\max\{\lambda_0 + \lambda_1 + \lambda_2(1+c^2)\} \\ &\text{such that} \quad r(u_i) \equiv \lambda_0 + \lambda_1 u_i + \lambda_2 u_i^2 \leq \psi(u_i) \quad \text{for all} \quad i, \quad 1 \leq i \leq m. \end{aligned} \tag{46}$$

If we simply replace λ_i via $-\lambda_i$ such that the max in (46) can be replaced by min, we found the dual problem in (46) is not equivalent to that in (40) because the $-\psi(u)$ is not a strictly monotone decreasing and strictly positive convex function satisfying Lemma 2. That leads to failure of completely replicating proof of Theorem 3, thus leading to weaker conclusions for the minimization problem.

Proposition 3. (*sufficient conditions for F_0 or F_u to be a stationary point*) *Under the same initial three assumptions as Theorem 2, (a) if $\ddot{\psi}(u;F)$ is strictly decreasing over $[0,M]$ for any candidate cdf F, then F_u must be a solution of the LP in (39). Hence, if this condition is satisfied for $F = F_u$, then F_u. must be a stationary point.*

(b) Similarly, if $\ddot{\psi}(u;F)$ is strictly increasing over $[0,M]$ for any candidate cdf F, then F_0 must be a solution of the LP in (39). Hence, if this condition is satisfied for $F = F_0$, then F_0. must be a stationary point.

Proof. Replace λ_i by $-\lambda_i$ and make $\psi(u)$ become $-\psi(u)$ in (46). Thus we shall solve

$$\begin{aligned} &\min\{\lambda_0+\lambda_1+\lambda_2(1+c^2)\} \\ &\text{such that} \quad r(u_i)\equiv\lambda_0+\lambda_1 u_i+\lambda_2 u_i^2 \geq -\psi(u_i) \quad \text{for all} \quad i, \quad 1\leq i\leq m. \end{aligned} \tag{47}$$

It is the dual problem of maximization problem with $-\psi$ as in the objective function. That implies if ψ is strictly monotone decreasing, then $-\psi$ is strictly monotone increase. Paralleling with proof of Theorem 3, we can conclude the opposite conclusions. ∎

Proposition 4. (*a single peak for minimization problem*) *For the minimization problem, if $\ddot{\psi}(u;F)$ has a single peak over $[0,M]$ for all $F\in\mathcal{P}_2(\mathcal{F})$, then all local optimizers must be in $\mathcal{P}_{2,3}(\mathcal{F})$, in particularly, the optimizers must be one of the $\{F_0,F_u,F_b\}$ where F_b is a three-point distribution with only one interior point $b\in(0,M)$. So that one of these three will be the global optimal solution.*

Proof. Given (47), if ψ has one peak, then the $-\psi$ is not a one peak function. We apply the same argument before (44) to show there is at most one interior intersection point. However, due to $-\psi$ not being one peak function, there is no other cases which can be ruled out, thus it is possible to have F_0, F_u as well as F_b being optimizers where F_b has two ending points 0 and M and it also has one interior point $b\in(0,M)$. ∎

Combine the result of Proposition 4 and parallel with Theorem 4, we establish the following result for minimization problem.

Theorem 5. (*more structures for minimization*) *Under the setting of Theorem 3. If $\ddot{\psi}(u;F)$ has at most n $(1\leq n<\infty)$ peaks over $[0,M]$ for any candidate $F\in\mathcal{P}_2(\mathcal{F})$, then all stationary points of the optimization in (39) must lie in $\mathcal{P}_{2,n+2}(\mathcal{F})$.*

Proof. We apply the same argument before (44) in the proof of Theorem 3 to show there is at most n interior intersection point, then following via the proof of Proposition 4, the proof is thus complete. ∎

8. Asymptotic Analysis

We now show how to relax the finite support condition introduced in Section 3. For that purpose, we will consider a sequence of nested finite support sets. We say that a sequence of finite support sets $\{\mathcal{F}_k : k \geq 1\}$ is nested if $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ for all $n \geq 1$. We say that $\mathcal{F}_k \to [0, M]$ as $k \to \infty$ for a nested sequence of support sets if each $x \in [0, M]$ can be expressed as

$$x = \lim_{k \to \infty} \{x_k : x_k \in \mathcal{F}_k\}. \tag{48}$$

We have the following well known approximation lemma.

Lemma 6. (*approximation lemma*) *If $\mathcal{F}_k \to [0, M]$ as $k \to \infty$ for a nested sequence of support sets, then Any cdf $F \in \mathcal{P}_2(M)$ can be expressed as the limit of cdf's $F_k \in \mathcal{P}_2(\mathcal{F}_k)$.*

Proof. We perform a direct construction. Let F_k be the right-continuous piecewise-constant function satisfying Let $F_k(x_k) = F(x_k)$. Then F is the limit of this constructed F_k. In particular, for all x that are continuity points of F, $F_k(x) \to F(x)$ as $k \to \infty$.∎

Lemma 7. (*extremal cdf for support* $[0, M]$) *Assume that $\mathcal{F}_k \to [0, M]$ as $k \to \infty$ for a nested sequence of support sets. If $F_k^* \in \mathcal{P}_{2,3}(\mathcal{F}_k)$ is the optimal cdf for support set $\mathcal{F}_k$, then there exists a convergent subsequence of $\{F_k^* : k \geq 1\}$ with limiting cdf $F^* \in \mathcal{P}_{2,3}([0, M])$ and the cdf F^* is an optimal cdf in $\mathcal{P}_{2,3}(M)$.*

Proof. The key fact is that $\mathcal{P}_{2,3}(M)$ is a compact subset of $\mathcal{P}_2(M)$. That implies the existence of the convergent subsequence with a limit in the same space. Then the continuity implies the extremal property in the limit. ∎

9. Simulation Verification for Some $GI/GI/1$ Examples

In this section we apply simulation to examine if the conditions in Theorem 3 for F_0 or F_u to be a stationary point of the optimization are satisfied for various $GI/GI/1$ examples, in the context of Corollary 1 and Lemma 2. That is, we consider the maximization over interarrival-time cdf's F with specified first two moments for given service-time cdf G. For that purpose, we will look at $\ddot{\phi}_a(u)$ in (27) for $\phi_a(u)$ in (21), which is the summation of pdf functions of Y_i in (30). We obtain supporting positive results for the exponential (M) and Erlang (E_2) service-time distributions and negative results for a mixture of two Erlang service-time distributions.

9.1. An Exponential Service Time Distribution

We first show simulation results for $GI/M/1$ models, with fixed exponential service-time distribution.

9.1.1. $F_0/M/1$ and $F_u/M/1$ Models. We start by considering the special two-point distributions F_0 with one mass on 0 and F_u with one mass on the upper limit of support M_a. If U is distributed as F_0 with mass at $\{0, 1+c_a^2\}$, then

$$\ddot{\phi}_a(u) = \sum_{i=1}^{n} \gamma_i(u) = \sum_{i=1}^{n} e_i(u + z_k^{(i)}) p_k^{(i)} \tag{49}$$

where $e_i(.)$ is the pdf of $\sum_{k=1}^{i} V_i(M)$ and $z_k^{(i)}$ is the point from convolution $\sum_{k=1}^{i-1} U_k$. For example, if $n = 2$, then

$$\ddot{\phi}_a(u) = e_2(u) \frac{c_a^2}{1+c_a^2} + e_2(u + 1 + c_a^2) \frac{1}{1+c_a^2} + e_1(u). \tag{50}$$

Direct calculation shows that $\ddot{\phi}_a(u) < 0$, implying that $\ddot{\phi}_a(u)$ is a strictly monotone decreasing function over $[0, M_a]$.

For large n, we can verify the monotonicity property in any instance by applying stochastic simulation. As before, we used Monte-Carlo simulation to create 5 replications of 10^6 random samples in order to estimate the summation of pdf functions $\ddot{\phi}_a$ for $\sum_{i=1}^{n} Y_i$ in $GI/M/1$ with $F = F_0$. To illustrate, Figure 2 shows the simulation estimates of the second derivative $\ddot{\phi}_a(u)$ in (27) of the objective function in (21) for $F_0/M/1$ (LHS) and $F_u/M/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.7, n =$

$4, M_a = 10$. These plots show that $\ddot{\phi}_a(u)$ is monotonically decreasing over $[0, M_a]$ in both cases. Hence, F_0 is the optimal solution in the LP in (22) or (39)) in both cases. Thus, we conclude that F_0 is a stationary point of the optimization, whereas F_u is not. These conclusions were confirmed by applying Corollary 1. In particular, F_0 was found to be the solution to the LP in both the nearly-steady-state example with $(\rho = 0.1, n = 40)$ and the transient example with $(\rho = 0.7, n = 4)$.

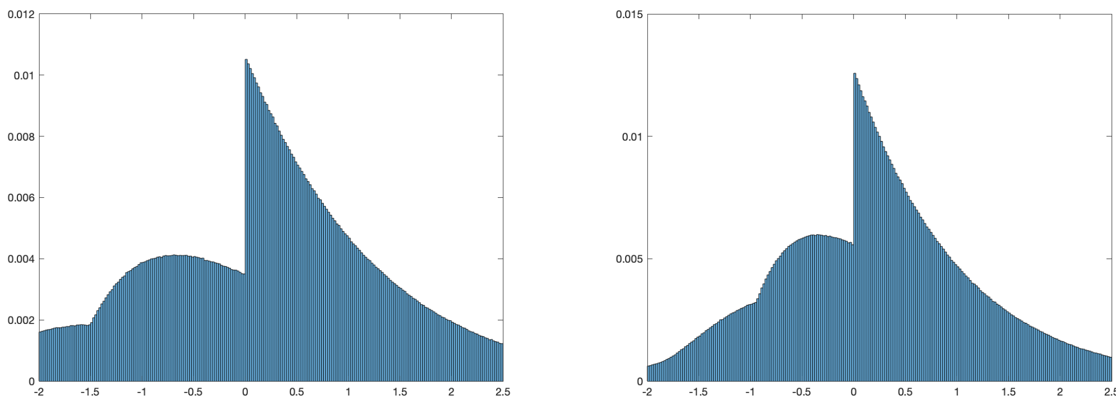

Figure 2 **Simulation estimates of $\ddot{\phi}(u)$ in (27) and Lemma 2 for $F_0/M/1$ (LHS) and $F_u/M/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.7, n = 4, M_a = 10$. These plots show that F_0 is a solution of the LP in (22) or (39) in both cases, so that F_0 is a stationary point, while F_u is not.**

9.1.2. Beyond Two-point Distributions. In order to better understand Theorem 3, we present the simulation results $GI/M/1$ models when the inter-arrival time distributions are not the special two-point distributions considered in Figure 2. Figure 3 displays the simulation estimates of $\ddot{\phi}(u)$ in (21) and Lemma 2 for $M/M/1$ (LHS) and $E_2/M/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.7, n = 4, M_a = 10$. These plots show that $\ddot{\phi}(u)$ is monotonically decreasing over $[0, M_a]$ in both cases. That implies that F_0 is a solution of the LP in both cases, so that these M and E_2 interarrival-time distributions are not stationary points of the optimization. As in the previous example, these conclusions were confirmed by applying Corollary 1. As before, F_0 was found to be the solution to the LP in both the nearly-steady-state example with $(\rho = 0.1, n = 40)$ and the transient example with $(\rho = 0.7, n = 4)$.

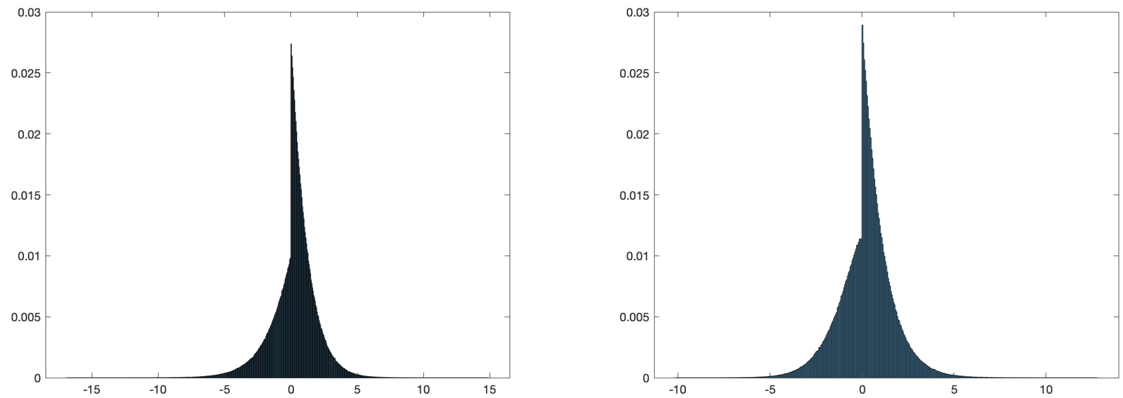

Figure 3 **Simulation estimates of $\ddot{\phi}(u)$ in (27) for $M/M/1$ (LHS) and $E_2/M/1$ (RHS) in the case $\rho = 0.7, n = 4, M_a = 10$. These plots show that F_0 is a solution of the LP in (22) in both cases, so that neither of these interarrival-time cdf's is a stationary point of the optimization.**

9.1.3. $GI/E_2/1$ Models. We now consider a fixed Erlang E_2 service-time distribution. The Erlang E_k service-time distributions are appealing because they are strongly unimodal, i.e., the convolution of the an Erlang distribution with any other unimodal distribution is again unimodal.

We now discuss $GI/E_2/1$. Figure 4 displays simulation estimates of $\ddot{\phi}(u)$ in (21) and Lemma 2 for $F_0/E_2/1$ (LHS) and $F_u/E_2/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$. In this case we do not see monotonicity, but instead we see the single-peak property over $[0, M_a]$. Thus, these plots also show that F_0 is a solution of the LP in (39) in both cases, because of the single-peak property, so that F_0 is a stationary point, while F_u is not. As in the previous example, these conclusions were confirmed by applying Corollary 1. As before, F_0 was found to be the solution to the LP in both the nearly-steady-state example with $(\rho = 0.1, n = 40)$ and the transient example with $(\rho = 0.7, n = 4)$.

9.1.4. Examples for a More Complex Service-Time Distribution. We now show that the sufficient condition in Theorem 3 involving a single peak is not always satisfied. For that purpose, we let the service-time distribution be the mixture of two Erlang distributions. Let $E_k(m)$ denote an E_k distribution with mean m, i.e., the distribution of the sum of k i.i.d. exponential random variables, each with mean m/k. Let $mix(E_{k_1}(m_1), E_{k_2}(m_2), p)$ denote the mixture of an Erlang $E_{k_1}(m_1)$ distribution with probability p and an $E_{k_2}(m_2)$ distribution with probability $1 - p$, which necessarily has mean $pm_1 + (1 - p)m_2$.

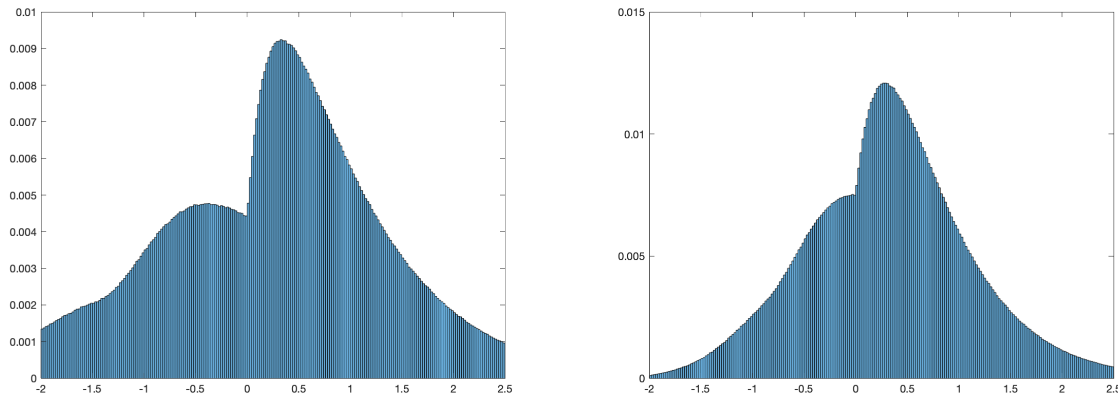

Figure 4 **Simulation estimates of $\ddot{\phi}_a(u)$ in (27) for $F_0/E_2/1$ (LHS) and $F_u/E_2/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$. These plots show that F_0 is a solution of the LP in (39) in both cases because of the single-peak property, so that F_0 is a stationary point, while F_u is not.**

Figure 5 presents simulation estimates of $\ddot{\phi}(u)$ in (21) and Lemma 2 for $F_0/GI/1$ (LHS) and $F_u/GI/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$, where the service-time distribution is chosen to be $G = mix(E_{20}(0.4), E_{20}(1.6), 0.5)$, which has mean $0.5(0.4) + 0.5(1.6) = 1.0$. Figure 5 shows that the condition of Theorem 3 is not satisfied in either case.

Unlike the previous three examples, the conclusions from applying Corollary 1 are more complicated. As before, F_0 was found to be the solution to the LP in the nearly-steady-state example with $(\rho = 0.1, n = 40)$, but it was not in the transient example with $(\rho = 0.7, n = 4)$.

9.2. Maximization Over G Given F

We next show simulation results for the associated maximization problem over candidate service-time distributions G, given a specified inter-arrival time distribution F.

From Lemma 3, we know that we can apply a reverse-time representation to reduce this problem to the case previously considered. That implies that, in the reverse-time representation, we should look at the shape of $\ddot{\phi}_s(\tilde{v})$ in the range $[0, \rho M_s]$. However, that is equivalent to looking at the corresponding shape of $\ddot{\phi}_s(v)$ over $[-\rho M_s, 0]$ without time-reverse representation. That means if the original shape is strictly monotone increasing over $[-\rho M_s, 0]$, that is equivalent to the time-reverse shape is strictly monotone increasing over $[0, \rho M_s]$ because 0 under the time-reverse representation corresponds to ρM_s under no time-reverse representation.

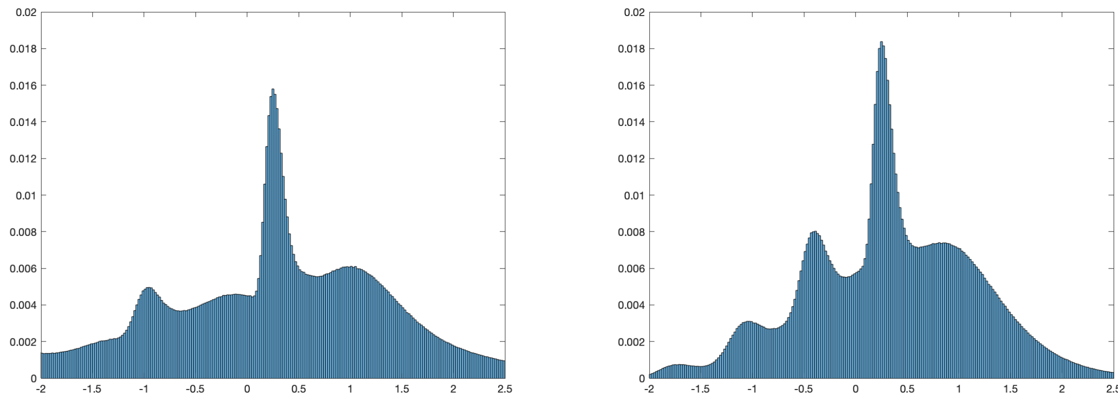

Figure 5 **Simulation estimates of $\ddot{\phi}(u)$ in (21) and Lemma 2 for $F_0/GI/1$ (LHS) and $F_u/GI/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$, where the service-time distribution in both cases is a mixture of two Erlang distributions, specifically $mix(E_{20}(0.4), E_{20}(1.6), 0.5)$, as defined above.**

We next show the summation of simulated pdf function $\sum_{i=1}^{n} Y_i$ for $M/GI/1$ with $G = G_0$ and $M/GI/1$ with $G = G_u$ with $n = 4, M_a = M_s = 10$.

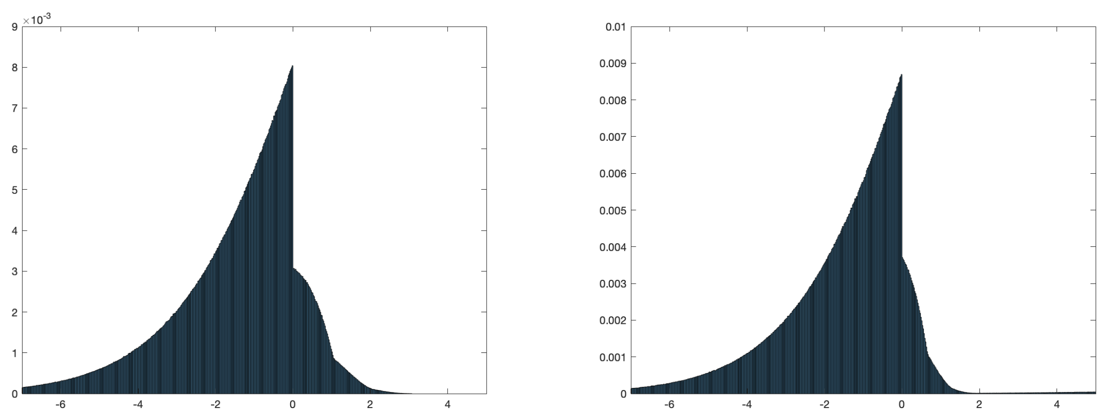

Figure 6 **Simulation estimates of $\ddot{\phi}_s(\tilde{v})$ in (34) associated with Lemma 3 for $M/G_0/1$ (LHS) and for $M/G_u/1$ (RHS) in the case: $c_a^2 = 01.0, c_s^2 = 0.5, \rho = 0.7, n = 4, M_s = 10$**

From $-\rho M_s$ to 0, we observe the strictly monotone increasing shape, thus the $\ddot{\phi}_s$ is strictly monotone increasing over $[0, \rho M_s]$ such that $\tilde{G}_u$ is the optimal solution under time reverse representation from Theorem 3. That implies that G_0 is a stationary point of the optimization for $M/GI/1$, whereas G_u is not.

10. Conclusions

We applied the theory of non-convex nonlinear programs together with the explicit expression for the transient mean $E[W_n]$ in (4) to study the interarrival-time distribution that maximizes the transient mean waiting time in the $GI/GI/1$ queue, given a specified service-time distribution and the first two moments of the interarrival time. We assume that the the interarrival-time distribution has finite support.

Theorem 1 first establishes the gradient of transient mean waiting time $E[W_n]$ with respect to the interarrival-time distribution F under finite support. Then Corollary 1 applies well-known first-order optimality conditions stated in Proposition 1 to characterize the stationary points of the optimization as solutions of a linear program. This provides an efficient way to construct counterexamples, as we illustrate in §4.2 and §4.7.

In §5 we develop an abstraction of the $GI/GI/1$ queueing problem that applies to other models in addition to the $GI/GI/1$ queue, provided that the objective function inherits the structure established for the $GI/GI/1$ model in Lemma 2. In that context, Theorem 2 establishes sufficient conditions for a stationary point to be a three-point distribution, while Theorem 3 establishes the sufficient conditions for the special two-point distributions F_0 and F_u to be stationary points of the optimization.

In §6 we prove Theorems 2 and 3. We prove Theorem 2 by applying Lemma 4 which establishes that an LP has a unique solution if and only if its dual has a nondegenerate optimal solution. We extend the proof of Theorem 3 to establish Theorem 4 for more complicated shapes.

In §7 we observe that minimization is not symmetric with maximization. In fact, the minimization problem is harder. We establish the corresponding results in Proposition 4 and Theorem 5. In particular, our results suggest that even when $\ddot{\phi}_a(u)$ is a one-peak function, it is possible that a specific three-point distribution will be a stationary point.

Finally, in §9 we apply simulation to provide some concrete numerical examples. We report results of simulation experiments showing that the sufficient conditions of Theorem 3 and Theorem 4 are satisfied for some concrete $GI/GI/1$ models.

There is much yet to be done; e.g., to verify or refute Conjectures 1 and 2 of Chen and Whitt (2021b).

Acknowledgments

This research was supported by NSF CMMI 1634133.

References

Asmussen, S. 2003. *Applied Probability and Queues*. 2nd ed. Springer, New York.

Bandi, C., D. Bertsimas, N. Youssef. 2015. Robust queueing theory. *Operations Research* **63**(3) 676–700.

Bertsekas, D.P. 2016. *Nonlinear Programming*. 3rd ed. Athena Scientific.

Bertsimas, D., K. Natarajan. 2007. A semidefinite optimization approach to the steady-state analysis of queueing systems. *Queueing Systems* **56** 27–39.

Bertsimas, D., J. N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Athena, Belmont, MA.

Billingsley, P. 1999. *Convergence of Probability Measures*. 2nd ed. Wiley, New York.

Birge, J. R., J. H. Dula. 1991. Bounding separable recourse functions with limited distribution information. *Annals of Operations Research* **30**(1) 277–298.

Chen, Y., W. Whitt. 2020a. Algorithms for the upper bound mean waiting time in the $GI/GI/1$ queue. *Queueing Systems* **94** 327–356.

Chen, Y., W. Whitt. 2020b. Extremal models for the $GI/GI/K$ waiting-time tail-probability decay rate. *Operations Research Letters* **48** 770–776.

Chen, Y., W. Whitt. 2021a. Extremal $GI/GI/1$ queues given two moments: Exploiting Tchebycheff systems. *Queueing Systems* **97** 101–124.

Chen, Y., W. Whitt. 2021b. Extremal $GI/GI/1$ queues given two moments: The case of two-point distributions. Submitted for publication, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

Chung, K. L. 2001. *A Course in Probability Theory*. 3rd ed. Academic Press, New York.

Daley, D. J., A. Ya. Kreinin, C.D. Trengove. 1992. Inequalities concerning the waiting-time in single-server queues: a survey. U. N. Bhat, I. V. Basawa, eds., *Queueing and Related Models*. Clarendon Press, 177–223.

Denuit, M., C. Lefevre, M. Shaked. 1998. The s-convex orders among real random variables with applications. *Mathematical Inequalitities and Applications* **1**(4) 585–613.

Feller, W. 1971. *An Introduction to Probability Theory and its Applications*. Second edition ed. John Wiley, New York.

Gupta, V., J. Dai, M. Harchol-Balter, B. Zwart. 2010. On the inapproximability of $M/G/K$: why two moments of job size distribution are not enough. *Queueing Systems* **64** 5–48.

Gupta, V., T. Osogami. 2011. On Markov-Krein characterization of the mean waiting time in $M/G/K$ and other queueing systems. *Queueing Systems* **68** 339–352.

Johnson, M. A., M.R. Taaffe. 1990. Matching moments to phase distributions: nonlinear programming approaches. *Stochastic Models* **6**(2) 259–281.

Karlin, S., W. J. Studden. 1966. *Tchebycheff Systems; With Applications in Analysis and Statistics*, vol. 137. Wiley, New York.

Kingman, J. F. C. 1962. Inequalities for the queue $GI/G/1$. *Biometrika* **49**(3/4) 315–324.

Klincewicz, J. G., W. Whitt. 1984. On approximations for queues, II: Shape constraints. *AT&T Bell Laboratories Technical Journal* **63**(1) 139–161.

Lam, H., C. Mottet. 2015. Simulating tail events with unspecified tail models. *Proceedings of the 2015 Winter Simulation Conference* **2015** 392–402.

Lam, H., C. Mottet. 2017. Tail analysis without parametric models: A worst-case perspective. *Operations Research* **65**(6) 1696–1711.

Li, Y., D. A. Goldberg. 2017. Simple and explicit bounds for multi-server queues with universal $1/(1-\rho)$ and better scaling. ArXiv:1706.04628v1.

Osogami, T., R. Raymond. 2013. Analysis of transient queues with semidefinite optimization. *Queueing Systems* **73** 195–234.

Rolski, T. 1976. Order relations in the set of probability distribution functions and their applications in queueing theory. *Dissertationes Mathematicae, Polish Academy of Sciences* **132**(1) 3–47.

Serfling, R.J. 1980. *Approximations Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Smith, J. 1995. Generalized Chebychev inequalities: Theory and application in decision analysis. *Operations Research* **43** 807–825.

Tijssen, G. A., G. Sierksma. 1998. Balinski-Tucker simplex tableaus: Dimensions, degeneracy degrees, and interior points of optimal faces. *Mathematical Programming* **81** 349–372.

van Eekelen, W., D. J. den Hartog, J. S. H. van Leeuwaarden. 2019. Mad dispersion measure makes extremal queue anlysis simple. Working paper.

Wang, T. 1993. L_p-Frechet differentiable preference and local utility analysis. *Journal of Economic Theory* **61** 139–159.

Whitt, W. 1974. The continuity of queues. *Advances in Applied Probability* **6** 175–183.

Whitt, W. 1984. On approximations for queues, I: Extremal distributions. *AT&T Bell Laboratories Technical Journal* **63**(1) 115–137.

Whitt, W., W. You. 2018. Using robust queueing to expose the impact of dependence in single-server queues. *Operations Research* **66**(1) 100–120.

Whitt, W., W. You. 2019. Time-varying robust queueing. *Operations Research* **67**(6) 1766–1782.

Wolff, R. W., C. Wang. 2003. Idle period approximations and bounds for the $GI/G/1$ queue. *Advances in Applied Probability* **35**(3) 773–792.