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 tight 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 derive 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. We develop and apply a stochastic variant of the Frank-Wolfe (1956) algorithm to find a stationary point for any given service-time distribution. We also 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: November 29, 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 steady-state 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, but 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 that 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.

There are also variants of the classic extremal problem above when one of the two underlying distributions is given. An appealing simple story is developed for higher moments of both the transient and steady-state $GI/GI/1$ 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 $W_n(F, G)$ be the waiting time of customer (arrival) n starting empty with interarrival-time cdf F and service-time cdf G. Let F_u and G_0 be two-point distributions 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, $1 \leq n \leq \infty$ ($n = \infty$ means steady-state) and $k \geq 2$:

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

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, the nice story in (1) breaks down for $k = 1$, i.e., for the transient and steady-state mean. For $k = 1$, counterexamples to cases (a) and (b) in the first two lines of (1) for $n \leq \infty$ and for case (c) in 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). (The paper Chen and Whitt (2021b) studies optimization over two-point and over three-point distributions.) Partial positive results for cases (a) and (b) with $k=1$ and $n=\infty$ (for the steady-state mean) are contained in Theorem 2 of Chen and Whitt (2021a). In particular, case (a) was established for $k=1$ and $n=\infty$ when G can be represented as a mixture of exponential distributions.

In this paper, we contribute by applying classical optimization theory to develop new mathematical tools to study the extremal theory for the transient mean waiting time. To the best of our knowledge, the present paper is the first to focus on tight bounds for the transient mean. Since the transient mean increases to the steady-state mean as n increases (see (4) below), we also provide new ways to study tight bounds for the steady-state mean.

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 (case(a) in (1) for $k=1$ and $n<\infty$). We show that this problem can be represented as a non-convex nonlinear program. (In an online supplement Chen and Whitt (2021c) we obtain related results for maximizing $E[W_n(F,G)]$ over candidate service-time distributions G, for given interrarrival-time distribution F, corresponding to case (b) of (1). We also consider the associated minimization problem there.)

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. The first step is to derive the gradient of the transient mean with respect to F, which we do for F having finite support in §3.

We next show in §4 that we can test whether or not F_0 (or any other candidate F) is a stationary point of the optimization by solving a linear program. We specify the objective function and show that it easily can be accurately 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 optimizations over F given G_u and over G given F_0 for the steady-state mean in numerical examples, thus providing evidence to support the main outstanding conjecture about case (c) mentioned in the opening paragraph.

Given the gradient of the transient mean, we also show in §5 that we can apply the conditional-gradient or Frank and Wolfe (1956) (FW) algorithm as in §3.2 of Bertsekas (2016) to calculate a stationary point of the optimization in numerical examples. Because we estimate the objective function by simulation, we use a stochastic variant of FW as in Reddi et al. (2016). Our version of the algorithm typically converged very rapidly, in 2-5 steps, and only rarely in up to 15 steps, thus providing a practical way to find stationary points.

Finally, in §6 we also provide numerical methods to determine structural properties, i.e., whether the extremal distribution is a two-point or three-point distribution. In §6.1 we develop an abstraction of our optimization problem, so that the results can be applied to other related stochastic models. This involves a moment problem over product measures. In §6.2 we establish structural properties (monotonicity and convexity) of the objective function in our queueing problem. In §6.3 we then state positive structural results in the general setting from the structure established in Lemma 1. The following §7 is devoted to the proofs. In §8 we give simulation examples related to §6. in §9 we draw conclusions. Additional supporting material appears in Chen and Whitt (2021c), a supplement to this paper available from the authors' web pages.

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

In this section we review 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 empty, so that $W_0 = 0$. 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. 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 $W_n \stackrel{\mathrm{d}}{=} \max\{S_k : 0 \leq k \leq n\}$ for $n \leq \infty$, $S_0 \equiv 0$, $S_k \equiv X_0 + \cdots + X_{k-1}$ and $X_k \equiv V_k - U_k$, $k \geq 1$. Moreover, it is 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.

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. Let $\mathcal{P}(\mu, c^2, M)$ be the set of probability measures on $[0, M]$ with finite mean μ and scv c^2, i.e., with second moment $\mu^2(1 + c^2)$.

With this notation, our primary goal is to establish results for the optimization problem

$$\sup\{E[W_n(F, G) : F \in \mathcal{P}(1, c_a^2, M)\}, \tag{5}$$

for fixed cdf G with $E[V]=\rho<1$ and scv $c_s^2<\infty$. The objective function is given by (4), but finding the global optimal solution of (5) 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, under the additional assumption that F has finite support.

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. Analogs of the following results can be established for cdf's with densities; see §3.1 in Chen and Whitt (2021c). These results supplement the literature on continuity of queues, e.g., §X.6 of Asmussen (2003).

Let the finite support set in $[0,M]$ be $\mathcal{F}$. Let the elements of $\mathcal{F}$ be $0=u_1<u_2<\ldots<u_m=M$ with $m\equiv|\mathcal{F}|\geq 3$. Let $\mathcal{P}(\mathcal{F})$ be the subset of $\mathcal{P}(1,c_a^2,M)$ with support set $\mathcal{F}$. 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) $p\equiv(p_1,\ldots,p_m)$. Let $w_n(p)\equiv E[W_n(p,G)]\equiv E[W_n(F,G)]$.

With finite support and this new notation, the optimization problem in (5) becomes

$$\begin{aligned}&\max\{w_n(p)\equiv E[W_n(p)]:p\in\mathcal{P}(\mathcal{F})\}\\ &\text{such that}\quad \sum_{i=1}^m p_i=1,\quad \sum_{i=1}^m u_ip_i=1,\quad \sum_{i=1}^m u_i^2p_i=(1+c_a^2)\quad\text{and}\quad p_i\geq 0,\end{aligned}\tag{6}$$

where $0=u_1<u_2<\ldots<u_m=M$ are the support points in $\mathcal{F}\subseteq[0,M]$. There is no loss of generality in going from the optimization problem in (5) to the optimization problem in (6) with finite support, provided that the optimal solution to (5) has support in $\mathcal{F}$. Thus, we always require that $\mathcal{F}$ contains the support of the natural candidate F_0, which has mass $1/(1+c_a^2)$ in $1+c_a^2$ and the rest at 0. Support for conclusions can be gained by considering successively larger finite support sets.

We now show that the function $w_n(p)$ in (6) 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 $\|x\|$ be the l_1 norm in $\mathbb{R}^m$, i.e., for $x \in \mathbb{R}^m$,

$$\|x\| \equiv \sum_{i=1}^{m} |x_i|. \tag{7}$$

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

$$\lim_{\|p-\hat{p}\|\to 0} \frac{\left|w_n(p) - w_n(\hat{p}) - \nabla w_n(\hat{p})^t \cdot (p - \hat{p})\right|}{\|p - \hat{p}\|} = 0, \tag{8}$$

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, i.e.,

$$\nabla w_n(\hat{p}) \equiv \left(\left(\frac{\partial w_n}{\partial p_1}(\hat{p})\right), \ldots, \left(\frac{\partial w_n}{\partial p_m}(\hat{p})\right)\right)^t \tag{9}$$

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

$$w_n(p) \approx w_n(\hat{p}) + \nabla w_n(\hat{p})^t \cdot (p - \hat{p}). \tag{10}$$

We now show that the transient mean waiting time in this finite support setting is Frechet differentiable with respect to the interarrival-time pmf p and derive the gradient and Hessian. We write $V(G)$ to indicate that V has cdf G; similarly, we write $U(\hat{p})$ to indicate that U has pmf $\hat{p}$.

THEOREM 1. (*Frechet derivative*) *For the $GI/GI/1$ queue in the finite support setting above, the function $w_n(p)$ in (6) is Frechet differentiable with respect to p at $\hat{p}$ in $\mathcal{P}(\mathcal{F})$ with partial derivatives with respect to p at $\hat{p}$ given by*

$$\frac{\partial w_n}{\partial p_i}(\hat{p}) = \sum_{j=1}^{n} E[(\sum_{k=1}^{j} V_{k-1}(G) - \sum_{k=1}^{j-1} U_{k-1}(\hat{p}) - u_i)^+], \tag{11}$$

so that

$$\nabla w_n(\hat{p})^t \cdot (p - \hat{p}) = \sum_{i=1}^{m} \frac{\partial w_n}{\partial p_i}(\hat{p})(p_i - \hat{p}_i). \tag{12}$$

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

$$H(l,k) \equiv \frac{\partial^{(2)} w_n}{\partial p_l \partial p_k}(\hat{p}) = \sum_{j=1}^{n} (j-1) E[(\sum_{k=1}^{j} V_{k-1}(G) - \sum_{k=1}^{j-2} U_{k-1}(\hat{p}) - u_l - u_k)^+]. \tag{13}$$

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 exist constants c_1 and c_2 such that $0 < c_1 < c_2 < \infty$ and $c_1 g(x) \le |f(x)| \le c_2 g(x)$ for all x. Then, adding and subtracting by $\hat{p}_i$ and $\hat{p}_j$ inside the expression for $w_2(p)$ from (4), we get

$$\begin{aligned}
w_2(p) &= \sum_i E[(V_0 - u_i)^+]p_i + \frac{1}{2}\sum_{i,j} E[(V_0 + V_1 - u_i - u_j)^+]p_i p_j \\
&= \sum_i E[(V_0 - u_i)^+](p_i - \hat{p}_i + \hat{p}_i) + \frac{1}{2}\sum_{i,j} E[(V_0 + V_1 - u_i - u_j)^+](p_i - \hat{p}_i + \hat{p}_i)(p_j - \hat{p}_j + \hat{p}_j) \\
&= \sum_i E[(V_0 - u_i)^+]\hat{p}_i + \frac{1}{2}\sum_{i,j} E[(V_0 + V_1 - u_i - u_j)^+]\hat{p}_i \hat{p}_j \\
&\quad + \sum_i E[(V_0 - u_i)^+](p_i - \hat{p}_i) + \sum_i E[(V_0 + V_1 - U_0(\hat{F}) - u_i)^+(p_i - \hat{p}_i) + \Theta(\|p - \hat{p}\|^2) \\
&= w_2(\hat{p}) + \sum_i \frac{\partial w_2}{\partial p_i}(\hat{p})(p_i - \hat{p}_i) + \Theta(\|p - \hat{p}\|^2),
\end{aligned} \tag{14}$$

where

$$\frac{\partial w_2}{\partial p_i}(\hat{p}) = \sum_{j=1}^{2} E[(\sum_{k=1}^{j} V_{k-1}(\hat{G}) - \sum_{k=1}^{j-1} U_{k-1}(F) - u_i)^+]. \tag{15}$$

To justify the conclusion in (14), we observe that there exists a constant C such that $E[(V_0 + V_1 - u_i - u_j)^+] \le C < \infty$ for all i and j. Consequently, the second term in the second line of (14) associated with the second order of $(p_i - \hat{p}_i)$ can be bounded by the square of the norm, in particular,

$$\begin{aligned}
&\Big|\frac{1}{2}\sum_{i,j} E[(V_0 + V_1 - u_i - u_j)^+](p_i - \hat{p}_i)(p_j - \hat{p}_j)\Big| \le C \sum_{i,j} \big|(p_i - \hat{p}_i)(p_j - \hat{p}_j)\big| \\
&\le C \sum_{i,j} \big|(p_i - \hat{p}_i)\big|\big|(p_j - \hat{p}_j)\big| = C\|p - \hat{p}\|^2.
\end{aligned}$$

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

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

Hence, we have shown that $w_2(p)$ is Frechet differentiable. We can extend to general n by observing that the argument above implies that for $n = 2$ the relation (10) extends to

$$w_n(p) \approx w_n(\hat{p}) + \nabla w_n(\hat{p})^t \cdot (p - \hat{p}) + O(\|p - \hat{p}\|^2) \quad \text{as} \quad \|p - \hat{p}\| \to 0. \tag{16}$$

It is not difficult to prove that (16) holds for all $n \geq 2$ by mathematical induction.

Given (11), we can continue to take the derivative with respect to p at $\hat{p}$, so that

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

Therefore, we directly obtain (13) for the case $n=2$. We can continue the argument to obtain (13). ∎

4. The Linear Program for a Stationary Point

We now show how to exploit the smoothness established in Theorem 1 to establish partial results for the optimization problem formulated in (5) and (6). First, we observe that there exists a global maximum because we are maximizing a continuous function over a compact subset of R^m.

Recall that a point $\hat{p}$ is a local maximum for (6) if there exists $\delta > 0$ such that

$$w_n(p) \leq w_n(\hat{p}) \quad \text{for all} \quad p \quad \text{such that} \quad \|p - \hat{p}\| < \delta. \tag{18}$$

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

Proposition 1. (*necessary condition for a local maximum, Proposition 3.1.1 of Bertsekas (2016)*) *If $\hat{p} \in \mathcal{P}(\mathcal{F})$ is a local maximum of $w_n(p)$ in (6), then*

$$\nabla w_n(\hat{p})^t \cdot (p - \hat{p}) \leq 0 \quad \textit{for all} \quad p \in \mathcal{P}(\mathcal{F}). \tag{19}$$

If there exists $\hat{p}$ satisfying (19), then $\hat{p}$ is called a stationary point (of the optimization).

It will be convenient to look at the partial derivatives in (11) as a function of the support point u with $\hat{p} \in \mathcal{P}(\mathcal{F})$ given. Hence, we define

$$\phi_a(u) \equiv \frac{\partial w_n}{\partial p_i}(\hat{p})(u) \equiv \sum_{i=1}^{n} E[(\sum_{k=1}^{i} V_{k-1}(G) - \sum_{k=1}^{i-1} U_{k-1}(\hat{p}) - u)^+], \quad u \geq 0. \tag{20}$$

Corollary 1. (*the key linear program*) *A pmf $\hat{p}$ in $\mathcal{P}(\mathcal{F})$ is a stationary point of the optimization in* (6), *satisfying* (19), *if and only if $\hat{p}$ is the solution of the linear program (LP)*

$$\sup\{\nabla w_n(\hat{p})^t \cdot p \equiv \sum_{i=1}^{m} \frac{\partial w_n}{\partial p_i}(\hat{p}) p_i \equiv \sum_{i=1}^{m} \phi_a(u_i) p_i : p \in \mathcal{P}(\mathcal{F})\}; \tag{21}$$

for $\phi_a(u)$ in (20), *i.e., if and and only if*

$$\sup\{\sum_{i=1}^{m} \phi_a(u_i) p_i : p \in \mathcal{P}(\mathcal{F})\} = \sum_{i=1}^{m} \phi_a(u_i)\hat{p}_i. \tag{22}$$

For the steady-state mean, the two-point cdf F_0 provides the tight upper bound for $E[W(F,G)]$ for many G, but that is not true for G_0, as shown in §8 of Wolff and Wang (2003). Hence, we now apply Corollary 1 to study the special two-point interarrival-time distribution F_0 for the case $G \equiv G_0$.

Example 1. (application of Corollary 1 to an established counterexample)

We now assume that the service-cdf is the two-point cdf G_0. We consider two cases, one designed to approximately represent steady state and one to be genuinely transient. The nearly-steady-state example has $n = 40, \rho = 0.1, c_a^2 = c_s^2 = 0.5, M = 10$. The support contains $m = 401$ points in $[0, 10]$ (including the endpoints) so that, F_0 is in the support, while the transient example has $n = 4, \rho = 0.7, c_a^2 = c_s^2 = 0.5, M = 10$. (The cdf F_0 as mass $1/(1 + c_a^2)$ on $1 + c_a^2 = 1.50$.)

In both cases we apply simulation to estimate the objective function in (20) when $G = G_0$ and $F = F_0$ and then solve the linear program in (21). We perform 5 independent replications, so that we can estimate 95% confidence intervals. 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.

When we apply this procedure for most standard service-time distributions, we find that F_0 is 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 when the service-time cdf is G_0. 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. For G_0, we find that the stationary point of the optimization with respect to F can depend on ρ.

5. A Version of Frank-Wolfe (1956) to Find a Stationary Point

The availability of the gradient of the transient mean allows us to apply the conditional-gradient or Frank and Wolfe (1956) algorithm as in §3.2 of Bertsekas (2016), Lacoste-Julien (2016) and references there to compute a stationary point starting from any initial feasible F, provided we can calculate the objective function (20) in the LP in Corollary 1. As in stochastic variants of the Frank and Wolfe (1956) algorithm, such as in Reddi et al. (2016), we estimate the objective function of the LP by applying simulation.

We can exploit the first-order linear approximation in (10). By Proposition 1, if $\hat{p} \in \mathcal{P}(\mathcal{F})$ is not a stationary point of the optimization in (6), then we can find a $p \in \mathcal{P}(\mathcal{F})$ such that $\nabla w_n(\hat{p})^t \cdot (p - \hat{p}) > 0$. We thus apply line search to find a p that improves the objective function. The FW algorithm computes a succession of improvements until a stationary point is found.

Let the successive cdf's F be indexed by $j \geq 1$. (These successive F_j play the role of $\hat{p}$ in Corollary 1.) The first step is to use Monte-Carlo simulation to estimate the objective value in (20) via

$$\phi_a(u; F_j) \equiv \sum_{i=1}^{n} E[(\sum_{k=1}^{i} V_{k-1}(G) - \sum_{k=1}^{i-1} U_{k-1}(F_j) - u)^+] \tag{23}$$

$$\approx \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{n} (\sum_{k=1}^{i} V_{k-1}^{(b)}(G) - \sum_{k=1}^{i-1} U_{k-1}^{(b)}(F_j) - u)^+, u \in \mathcal{F}. \tag{24}$$

where we sample B i.i.d. copies of $\{(V_k, U_k) : 0 \leq k \leq n-1\}$ for each j. In each iteration we solve a linear program in the optimization step. In the following practical algorithm, we have made an additional simplifying approximation, letting the step size be $\varepsilon_j = 2/(j+2)$, $j \geq 1$. We found that this approximation was effective in all our numerical examples. See the supplement Chen and Whitt (2021c) for a more complicated step size algorithm following Lacoste-Julien (2016). There we prove that the sequence of cdf's $\{F_j : j \geq 1\}$ converges to a stationary point as $j \to \infty$, assuming accuracy in the objective function, by applying Lacoste-Julien (2016). Here we give a practical algorithm that we have found to be effective in identifying a stationary point in only a few iterations.

To state the practical algorithm, let $E_F[\cdot]$ denote the expectation with respect to candidate the cdf F of U.

Algorithm 1: Practical Stochastic Frank-Wolfe Algorithm

Initialization: A distribution F_1 in the feasible region $\mathcal{P}(\mathcal{F})$.

Input: Step size $\varepsilon_j \equiv 2/(2+j)$ for each step $j = 1, 2, \ldots$ and a stopping threshold $\delta > 0$

Procedure: For each iteration $j = 1, 2, \ldots,$ given a distribution F_j:

1 Compute the estimate of $\phi_a(u)$ in (23) by

$$\hat{\phi}_a(u; F_j) \equiv \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{n} (\sum_{k=1}^{i} V_{k-1}^{(b)}(G) - \sum_{k=1}^{i-1} U_{k-1}^{b}(F_j) - u)^+, u \in \mathcal{F} \tag{25}$$

2 Apply the LP in Corollary 1 to solve $Q_j = \arg\max_{F \in \mathcal{P}(\mathcal{F})} E_F[\hat{\phi}_a(U(F_{j-1}); F)]$ and let the FW gap at iteration j be

$$\bar{g}_j \equiv E_{Q_j}[\hat{\phi}_a(U; F_j)] - E_{F_j}[\hat{\phi}_a(U; F_j)] \tag{26}$$

3 Update $F_{j+1} = (1 - \varepsilon_j) F_j + \varepsilon_j Q_j$.

Repeat until $\bar{g}_j \leq \delta$ or Q_j is not changed for two consecutive iterations. If Q_j has not changed for two consecutive iterations, test whether Q_j itself is a stationary point. If so, stop; otherwise, continue iterating.

In all our numerical experiments, we found that, for given service-time cdf G, the stochastic FW algorithm converged to the same stationary point whatever initial cdf F is used. (The traffic intensity ρ is the mean of G, so it is fixed given G.) In addition, we observed that the sequence of $\{Q_j\}$ does not change after the initial few steps and $F_j \to Q_\infty$ as $j \to \infty$. Algorithm (1) always terminated within at most 15 steps.

With regard to the extremal cdf's, here is a summary of our findings: For the case (a), we determine stationary points for $F/G_0/1$ and found the F_0 is not always stationary point. We also found examples of cdf's G having a density for which F_0 is not optimal; see §8. We have not yet found an example of a completely monotone G for which F_0 is not optimal. Thus, we conjecture that the transient analog of Theorem 2 of Chen and Whitt (2021a) is valid.

For the case (b), we found that G_0 was the only stationary point for $E[W_n(M,G)]$ with $n<\infty$. The uniqueness for $n<\infty$ is in contrast to the insensitivity property of the steady-state mean. For the case (c), we confirm the conjectured solution (F_0, G_u) and $(F_0, G_{u,n})$ are stationary points for $E[W_\infty(F,G)]$ and $E[W_n(F,G)]$. (Recall that two-point distributions $G_{u,n}$ is a two-point distribution, where the upper mass point converges to M_s as $n\to\infty$.)

To illustrate, we describe two experiments, one for the transient mean and one for the (approximate) steady-state mean. For the transient mean, we let $n=4, \rho=0.5$, $B=1\times 10^7$ and support consisting of $m=401$ points uniformly distributed in the interval $[0,10]$. (Since F_0 has mass on $1+c_a^2=1.50$, F_0 is in the support.) For steady-state waiting time, we let $\rho=0.1$ and $n=40$. In the simulation studies, we consider different initial distributions. In all experiments, the optimization step in the algorithm (1) is numerically solved via the Gurobi solver in CVX.

5.1. The Transient Mean Waiting Time

We first consider the transient mean $E[W_4]$ for the four service-time distributions: exponential (M), Erlang (E_{10}) and the special two-point distributions G_u with one mass point on 10 and G_0 with one mass point on $\rho^2(1+c_a^2)=3/8=0.375$. For G being M, E_{10} and G_u, the algorithm converged to F_0 in two steps for all initial F considered, leading to $Q_j=F_0$ for all subsequent j.

For G_0, the story was different. Table 1 gives the numerical calues of Q_j for G_0 and $F_1=F_u$ when the support contains $m=401$ points uniformly distributed in $[0,10]$.

Table 1 **The successive optimal distributions Q_j for $E[W_4(F,G_0)]$ with $c_a^2=c_s^2=0.5, \rho=0.5$ when the initial distribution is $F=F_u$**

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.3333	0.6667		0.0000	1.500	
2	0.3795	0.1538	0.4667	0.1000	1.400	1.600
3	0.4190	0.5810		0.1500	1.600	
4	0.3816	0.4828	0.1356	0.1000	1.550	1.575
5	0.3816	0.4828	0.1356	0.1000	1.550	1.575

The solution in Table 1 is a three-point distribution, but it has two adjacent points in its support (1.5500 and 1.575), suggesting that it might change if we refined the support. Indeed, when we increase m to 801 from 401, we find the right to mass points change to 1.5500 and 1.5625. Continuing to increase the support set in this way, our numerical estimate of the extremal distribution for (5) is actually a two-point distribution with masses 0.3816 and 0.6184 on 0.1000 and 1.5556.

5.2. Steady-State Mean Waiting Time

We repeat the above experiments for the approximate steady-state mean waiting time $E[W_{40}(F,G)]$ under the same four models. The story is not changed for $F=M,E_{10},F_u$, again yielding $F=F_0$ as the stationary point. For $F/G_0/1$, we obtain F_0 being approximate stationary point under $\rho=0.1$.

But when we set $\rho=0.5$, we obtain a different stationary points F with three masses on $\{0.3295,0.3232,0.3472\}$ on support $\{0.000,1.375,1.600\}$. Table 2 shows the successive Q_j.

Table 2 The sequence of optimal distribution Q_j for $GI/G_0/1$ for $E[W_{60}(F,G)]$ during each iterations when initial distribution is $F=M$ with $\rho=0.5$

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.3333	0.6667		0.000	1.500	
2	0.4170	0.1413	0.4417	0.000	0.175	1.325
3	0.3326	0.3448	0.3226	0.000	1.450	1.550
4	0.3317	0.3509	0.3175	0.000	1.425	1.575
5	0.3304	0.3571	0.3125	0.000	1.400	1.600
6	0.3304	0.3571	0.3125	0.000	1.400	1.600
7	0.3287	0.3636	0.3077	0.000	1.375	1.625
8	0.3287	0.3636	0.3077	0.000	1.375	1.625
9	0.3295	0.3232	0.3472	0.000	1.375	1.600
10	0.3295	0.3232	0.3472	0.000	1.375	1.600

Therefore, we obtain different stationary points under different ρ. We see that F_0 is not the stationary point for $F/G_0/1$ for all ρ. It seems likely that a three-point distribution will be the extremal distribution for $E[W_\infty(F,G_0)]$ for some ρ.

6. Sufficient Conditions for Structured Extremal Distributions

In this section we establish sufficient conditions for the extremal cdf to have special structure, e.g., to be a three-point distribution or a two-point distribution. However, we first 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 function $\phi_a(u)$ in (20). We will then identify structure needed for these functions in order to characterize the solutions of the optimization problems. To provide guidance, in §6.2 we will establish structure of the objective function arising with the transient mean waiting time.

6.1. An Abstraction to a Multi-Dimensional Moment Problem

Our abstraction extends the classical moment problem, 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

$$\max_F \{E[\hat{g}(X)] \equiv \int_0^M \hat{g}(x)dF(x) \quad \text{subject to} \quad F \in \mathcal{P}(\mu, c_a^2, M)\}, \tag{27}$$

where $\hat{g}$ is a real-valued continuous function defined on $[0,M]$ and X is a random variable distributed as F where F lies in the domain $\mathcal{P}(\mu, c_a^2, M)$ with fixed first two moments μ and $\mu^2(1+c_a^2)$ and bounded support $[0,M]$, 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^* with all mass on at most three points.

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 $(X_1, \dots, X_n)$, where X_i are independent random variables with a common marginal cdf's F. The new formulation is

$$\max_F \{E[\hat{g}(X_1, \dots, X_n)] \equiv \int_0^M \hat{g}(x_1, \dots, x_n)dF(x_1)\dots dF(x_n) \quad \text{subject to} \quad F \in \mathcal{P}(\mu, c_a^2, M)\}, \tag{28}$$

where $\hat{g}(x_1, \dots, x_n)$ is a nonnegative continuous real-valued function defined on the product space $[0,M]^n$ with $M \geq 1+c^2$ (to have a feasible solution). In (28) the common marginal distribution has specified first two moments. The program formulation in (28) has many applications such as robust estimation in tail analysis and rare-event simulation; e.g., Lam and Mottet (2015, 2017) propose the reformulation in (28) with $\hat{g}$ being an indicator function. In that case, for some positive b, we are interested in solving

$$\max_F \{P(X_1 + \ldots + X_n \geq b) = \int_0^M 1_{\{x_1+\ldots+x_n \geq b\}} dF(x_1)\ldots dF(x_n) \quad \text{subject to} \quad F \in \mathcal{P}(\mu, c_a^2, M)\},$$

where all X_i are independent and distributed as the same cdf F, where F lies in an uncertainty set with unspecified tail.

As in §3, we restrict attention to probability distributions with finite support. We assume that all $F \in \mathcal{P}(\mu, c_a^2, M)$ have the common finite support $\mathcal{F}$ with elements $0 = u_1 < \ldots < u_m = M$ with sufficient large m. So that we have the following alternative formulation for (28),

$$\begin{aligned} &\max_p \hat{w}_n(p) \equiv \sum_{i_1,\ldots,i_n} \hat{g}(u_{i_1}, \ldots, u_{i_n}) p_{i_1} \ldots p_{i_n} \\ &\text{subject to} \quad \sum_{i=1}^m p_i = 1, \sum_{i=1}^m u_i p_i = 1, \sum_{i=1}^m u_i^2 p_i = 1 + c^2 \quad \text{and} \quad p_i \geq 0, \end{aligned} \tag{29}$$

where $g : \mathcal{F} \to R$ and the probability mass function p belongs to $\mathcal{P}(\mathcal{F})$, which is a compact and convex subset of R^m.

6.2. Structural Properties of the Objective Function in (20)

To provide guidance about what possible conditions to assume for our general objective function, We next establish structural properties of the objective function in (20) and (21) regarded as a function of u over the interval $[0, M]$.

LEMMA 1. (*structure of the objective function in* (20)) *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 $[-(i-1)M, \infty)$ which has a positive pdf γ_i over $[0, \infty)$ for each i, $1 \leq i \leq m$. Hence, for $x > 0$, the cdf of Y_i can be expressed by*

$$\Gamma_i(x) = \Gamma_i(0) + \int_0^x \gamma_i(y)\, dy \quad \textit{for} \quad x \geq 0, \tag{30}$$

so that the function ϕ_a in (20) can be expressed as

$$\phi_a(u) \equiv \frac{\partial w_n}{\partial p}(\hat{p}) = \sum_{i=1}^{n} \int_0^{\infty} (x-u)^+ \gamma_i(x)\, dx > 0, \quad u \geq 0. \tag{31}$$

Hence, $\phi_a(u) > 0$ and the first two derivatives of ϕ_a in (20) with respect to u exist for $u > 0$ and satisfy

$$\dot{\phi}_a(u) = \sum_{i=1}^{n} (\Gamma_i(u) - 1) < 0, \quad \ddot{\phi}_a(u) = \sum_{i=1}^{n} \gamma_i(u) > 0, \quad u \geq 0. \tag{32}$$

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

Proof. We directly calculate the derivative of $\phi_a(u)$ in (20) 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

$$\dot{\phi}_a(u) = -\sum_{i=1}^{n} \int_u^{\infty} \gamma_i(x)\, dx = \sum_{i=1}^{n} (\Gamma_i(u) - 1). \tag{33}$$

The rest follows directly. ∎

Going forward, we will see that the extremal distributions will depend on the structure of $\ddot{\phi}_a(u)$ in (32), which is the second derivative of $\phi_a(u)$ in (20). where γ_i is the pdf of $Y_i \equiv \sum_{k=1}^{i} V_k - \sum_{k=1}^{i-1} U_k$. We will establish concrete results in the next section.

6.3. 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 (20) replaced by a new function

$$\psi(u) \equiv \frac{\partial \hat{w}_n}{\partial p_i}(\hat{p})(u) \tag{34}$$

for $\hat{w}_n$ in (29). It suffices to check the optimality for

$$\max\{\sum_{i=1}^{m} \psi(u_i) p_i \equiv \nabla g(\hat{p})^t p,\ p \in \mathcal{P}(\mathcal{F})\} = \nabla g(\hat{p})^t \hat{p}. \tag{35}$$

As regularity conditions we require the properties deduced for ϕ_a in Lemma 1, but we also an extra condition on the second derivative $\ddot{\psi}$.

We apply duality theory for the LP in (35). From basic LP duality theory as in Ch. 4 of Bertsimas and Tsitsiklis (1997), the dual problem associated with the LP in (35) is to find the vector $\lambda^* \equiv (\lambda_0^*, \lambda_1^*, \lambda_2^*)$ that attains the minimum

$$\begin{aligned} &\min \{\lambda_0 + \lambda_1 + \lambda_2(1+c^2)\} \\ \text{such that} \quad & r(u_i) \equiv \lambda_0 + \lambda_1 u_i + \lambda_2 u_i^2 \geq \psi(u_i) \quad \text{for all} \quad i, \quad 1 \leq i \leq m. \end{aligned} \tag{36}$$

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 this purpose, we now introduce additional notation. Let $\mathcal{P}_n(\mathcal{F})$ denote the subset of $\mathcal{P}(\mathcal{F})$ with support on at most n points.

For the queueing problem, Lemma 1 shows that the first three 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* (29)-(35)*:*

(*i*) *The objective function $\hat{w}_n(p)$ in* (29) *is Frechet differentiable at all $p \in \mathcal{P}(\mathcal{F})$.*

(*ii*) *$\psi(u)$ in* (34) *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 (29)*, the LP given $\hat{p}$ in* (35) *has a unique optimal solution, which is thus an extreme point, and thus an element of $\mathcal{P}_3(\mathcal{F})$, if and only if the quadratic function $r(u)$ in* (36) *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 $f:[0,M]\to\mathbb{R}$ 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 (35). 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 (35). Hence, if this condition is satisfied for $F=F_u$, then F_u. must be a stationary point.

Corollary 2. (*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 $F\in\mathcal{P}(\mathcal{F})$, then the identified stationary points in Theorem 3 provide the unique global optimal solution.*

We can also extend to other functional forms of $\ddot{\psi}$ 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. (*implication of multiple peaks*) *Under the setting of Theorem 3. If $\ddot{\psi}(u;F)$ has at most n $(1\le n<\infty)$ peaks over $[0,M]$ for any candidate $F\in\mathcal{P}(\mathcal{F})$, then all stationary points of the optimization in (35) must lie in $\mathcal{P}_{n+1}(\mathcal{F})$.*

7. Proofs for Theorems 2-4 in Section 6.3

We now prove the results above.

7.1. 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 points with $\psi(u;\hat{p})$, the main goal is to show such (35) 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 2. (*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 2 from Corollary 1 to Theorem 4 in Tijssen and Sierksma (1998), we express the dual (36) in standard form by introducing slack variables and dividing the three variables λ_i into their positive and negative parts as

$$\begin{aligned} &\min\{(\lambda_0^+ - \lambda_0^-) + (\lambda_1^+ - \lambda_1^-) + (\lambda_2^+ - \lambda_2^-)(1+c^2)\} \\ &\text{such that}\quad (\lambda_0^+ - \lambda_0^-) + (\lambda_1^+ - \lambda_1^-)u_i + (\lambda_2^+ - \lambda_2^-)u_i^2 + s_i = \psi(u_i) \quad \text{for all}\quad i, \quad 1 \le i \le m, \\ &\text{and}\quad \lambda_j^+ \ge 0, \lambda_j^- \ge 0, 1 \le j \le 3; \quad s_i \ge 0, 1 \le i \le m. \end{aligned} \tag{37}$$

In the setting of (37), 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 (36) are not equal to zero. We will achieve the goal by establishing Lemma 3 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 (35) and $\hat{p}$ must be in $\mathcal{P}_3(\mathcal{F})$, i.e., must be a three-point distribution.

LEMMA 3. (*non-degeneracy for the dual*) *For the dual formulation (36), 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 (36), 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 (35), under the assumed conditions. (The first three are established in Lemma 1.) Under those conditions, $\psi(u)$ 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 (36) 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 (36) and the primal LP (35) have feasible solutions and the feasible region of the primal LP (35) is compact, thus they both have at least one optimal solution. We will show that the primal LP (35) has a unique solution by applying Lemma 2 and showing that no optimal solution of the dual (36) can be degenerate.

Hence, we will show (i) that we cannot have the optimal λ_i^* be 0 for any i in the (36).

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 (35) which must be $\in \mathcal{P}_3(\mathcal{F})$, 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 $\hat{p}$.

7.2. Proof of Theorem 3

We now consider the LP (35) 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 (35) has a unique optimal solution and the unique optimal solutions will be the specified special two-point distribution.

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 (36), we also have

$$\begin{aligned} \ddot{r}(u) &= 2\lambda_2^* = \ddot{\psi}(u), \\ \dot{r}(u) &= \lambda_1^* + 2\lambda_2^* u = \dot{\psi}(u), \\ r(u) &= \lambda_0^* + \lambda_1^* u + \lambda_2^* u^2 = \psi(u). \end{aligned} \tag{38}$$

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,

$$\begin{aligned} &\ddot{r}(x) = 2\lambda_2^* = \ddot{\psi}(x), \dot{r}(x) = \dot{\psi}(x), r(u) = \psi(x), \\ &\ddot{r}(y) = 2\lambda_2^* = \ddot{\psi}(y), \dot{r}(y) = \dot{\psi}(y), r(y) = \psi(y). \end{aligned} \tag{39}$$

Looking at the differences of these derivatives, we obtain

$$2\lambda_2^* = \frac{\dot{\psi}(y) - \dot{\psi}(x)}{y - x} = \ddot{\psi}(x) = \ddot{\psi}(y). \tag{40}$$

Therefore, by Mean Value Theorem, there exists $\tilde{u} \in (x, y)$ such that $\dddot{\psi}(\tilde{u}) = 2\lambda_2^*$. That leads to a contradiction because such $\dddot{\psi}(u)$ 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

$$2\lambda_2^* = \ddot{\psi}(y), \dot{r}(y) = \dot{\psi}(y), r(y) = \psi(y). \tag{41}$$

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 $\dddot{\psi}(u)$, the point y must be the final intersection point for $\ddot{\psi}(u)$ and $2\lambda^*$. For $u > y$, since $2\lambda^* > \ddot{\psi}(u)$ ($\dddot{\psi}$ 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 the 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. ∎

7.3. Proof of Theorem 4

Paralleling with lines before (40) 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 (41), 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 $\mathcal{P}_{n+1}(\mathcal{F})$.

8. Numerical Examples Exploiting Theorem 3 (a)

In this section we apply simulation to examine if the conditions in Theorem 3 (a) for F_0 and other F to be a stationary point of the optimization are satisfied for various $GI/GI/1$ examples, in the context of Corollary 1 and Lemma 1. 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 (32) for $\phi_a(u)$ in (20) 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.

First, Figure 1 shows simulation estimates of $\ddot{\phi}(u)$ in (32) and Lemma 1 for $F_0/M/1$ (LHS) and $M/M/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.7, n = 4, M = 10$ with $m = 501$ equally spaced points in the support. Both plots show that $\ddot{\phi}_a(u)$ is monotonically decreasing over $[0, M]$, implying that F_0 is the optimal solution in the LP in (21) or (35)) in both cases. That in turn implies that, when the service-time distribution is M, F_0 is a stationary point for the optimization, while M is not. As shown in the supplement Chen and Whitt (2021c), corresponding plots for the models $F_u/M/1$ and $E_2/M/1$ look very similar to the LHS and RHS of Figure 1, respectively, again implying that F_0 is the optimal solution in the LP in (21) or (35)) in both cases. Hence, neither F_u nor E_2 is a stationary point when the service-time distribution is M.

Next, Figure 2 shows simulation estimates of $\ddot{\phi}(u)$ in (32) and Lemma 1 again in the case $c_a^2 = 0.5, \rho = 0.7, n = 4, M = 10$ for two cases involving Erlang distributions. First, the LHS shows the simulation estimates for the $F_0/E_2/1$ model. In this case we do not see monotonicity, but instead we see the single-peak property over $[0, M]$. Thus, the LHS shows that F_0 is again a solution of the LP in (35), because of the single-peak property, so that F_0 is a stationary point. The model with interarrival-time cdf F_u looks very similar, again showing the single-peak property, but that implies F_u is not a stationary point.

The RHS in Figure 2 considers a more complex service-time cdf. Let $E_k(\mu)$ denote an E_k distribution with mean μ, i.e., the distribution of the sum of k i.i.d. exponential random variables, each

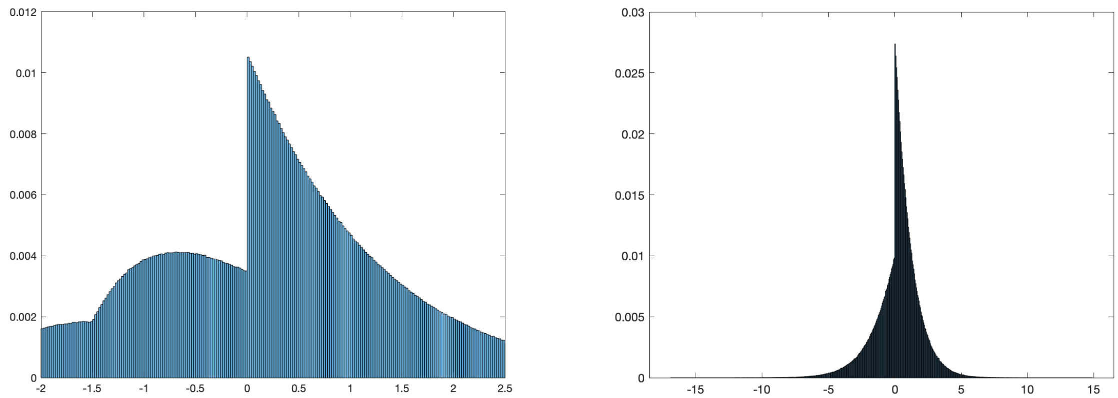

Figure 1 **Simulation estimates of $\ddot{\phi}(u)$ in (32) and Lemma 1 for $F_0/M/1$ (LHS) and $M/M/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.7, n = 4, M = 10$. These plots show that F_0 is a solution of the LP in (21) or (35) in both cases, so that F_0 is a stationary point, while M is not.**

with mean μ/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 $pm_1 + (1-p)m_2$. The specific G is $mix(E_{20}(0.4), E_{20}(1.6), 0.5)$. The RHS shows that that the condition of Theorem 3 (a) is not satisfied for this more complicated service-time distribution. Nevertheless, even though the condition of Theorem 3 (a) is not satisfied for this more complicated service-time distribution, an application of the FW algorithm shows that F_0 is a stationary point.

9. 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, assuming that the the interarrival-time distribution has finite support. We established mathematical properties justifying three different numerical algorithms, and illustrated each in §§4, 5, 8 and the supplement, Chen and Whitt (2021c).

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

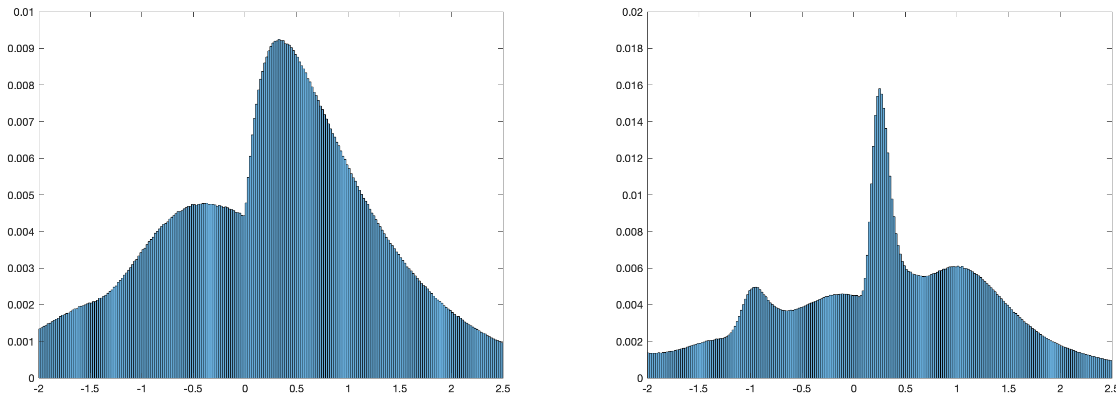

Figure 2 **Simulation estimates of $\ddot{\phi}_a(u)$ in (32) for $F_0/E_2/1$ (LHS) and $F_u/GI/1$ (RHS) where the service-time distribution is a mixture of two Erlang distributions, specifically $mix(E_{20}(0.4), E_{20}(1.6), 0.5)$, as defined above. in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M = 10$. The LHS plots show that F_0 is a solution of the LP in (35) in both cases because of the single-peak property, so that F_0 is a stationary point. In contrast, the RHS shows that the condition of Theorem 3 (a) is not satisfied**

optimization as solutions of a linear program. This provides an efficient way to construct counterexamples, as we illustrate in Example 1. The gradient also provides a basis for the Frank and Wolfe (1956) or conditional-gradient algorithm to find a stationary point, as we discuss in §5.

In §6 we develop new structural theorems. In §6.1 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 1 in §6.2. 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 §7 we prove Theorems 2 and 3. We prove Theorem 2 by applying Lemma 2 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 functional forms. We present numerical examples illustrating Theorem 3 (a) in §8. More examples appear in the supplement Chen and Whitt (2021c).

There is much yet to be done. It remains to prove or disprove that there is a unique stationary point of the maximization of $E[W_n(F, G)]$ over $F \in \mathcal{P}(\mathcal{F})$ for given G. It remains to be seen if the numerical examples can be extended to theorems, e.g., by efficiently calculated the functions that here have been estimated by simulation in our numerical examples. It remains to derive expressions for the tight upper bound of the mean $E[W_n(F_0, G)]$ as a function of the service-time cdf G, extending the results for $E[W_n(F_0, G_u)]$ in Chen and Whitt (2020a). It remains to obtain corresponding results for other stochastic models. Hopefully the results here can be helpful for that purpose.

Acknowledgments

This research was supported by NSF CMMI 1634133.

References

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

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

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

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: Three-point and two-point distributions. Working paper, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

Chen, Y., W. Whitt. 2021c. Supplement to applying optimization to study extremal $GI/GI/1$ transient mean waiting times. Working paper, 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.

Frank, M., P. Wolfe. 1956. Algorithm for quadratic programming. *Naval Research Logistics Quarterly* **3** 95–110.

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.

Lacoste-Julien, S. 2016. Convergence rate of Frank-Wolfe for non-convex objectives. Arxiv:1607:00345.

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.

Reddi, S. J., S. Sra, B. Poczos, A. Smola. 2016. Stochastic Frank-Wolfe methods for non-convex optimization. *2016* 54^{th} *Annual Allerton Conference on Communicatiom, Control and Computing*. IEEE, 1244–1251.

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.

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.

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

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.