

Extremal $GI/GI/1$ Queues Given Two Moments: Exploiting Tchebycheff Systems

Yan Chen

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

Ward Whitt

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

This paper studies tight upper bounds for the mean and higher moments of the steady-state waiting time in the $GI/GI/1$ queue given the first two moments of the interarrival-time and service-time distributions. We apply the theory of Tchebycheff systems to obtain conditions for classical two-point distributions to yield the extreme values. These distributions are determined by having one mass at 0 or at the upper limit of support.

Key words: $GI/GI/1$ queue, tight bounds, extremal queues, bounds for the mean steady-state mean waiting time, moment problem

History: February 5, 2020

1. Introduction

In this paper we apply the theory of Tchebycheff (T) systems from [Karlin and Studden \(1966\)](#) to identify the extremal interarrival-time and service-time distributions with given first two moments for the mean and higher moments of the steady-state waiting time in the $GI/GI/1$ queue. Thus, this paper contributes to a long-standing open problem for the classical $GI/GI/1$ queueing model: determining a tight upper bound (UB) for the mean steady-state waiting time, and the distributions that attain them, given the first two moments of the interarrival-time and service-time distributions; see [Daley et al. \(1992\)](#), [Wolff and Wang \(2003\)](#) and references therein. This paper is a sequel to [Chen and Whitt \(2019\)](#) in which we applied T systems to determine interarrival-time and service-time distributions with given moments and other properties that maximize or minimize the asymptotic decay rate of the steady-state waiting time in the $GI/GI/k$ queue. The theory of

T systems was previously used for queueing problems in [Rolski \(1972\)](#), [Holtzman \(1973\)](#), [Eckberg \(1977\)](#), [Whitt \(1984a,b\)](#), [Johnson and Taaffe \(1991, 1993\)](#), [Gupta and Osogami \(2011\)](#).

1.1. The $GI/GI/1$ Model

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 $\{V_n : n \geq 1\}$, each distributed as V with cumulative distribution function (cdf) G , which is independent of a sequence of i.i.d. interarrival times $\{U_n : n \geq 1\}$ each distributed as U with cdf F . With the understanding that the first customer (customer 1) 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 $E[U] \equiv 1$ and squared coefficient of variation (scv, variance divided by the square of the mean) c_a^2 ; let a service time V have mean $E[V] \equiv \tau \equiv \rho$ and scv c_s^2 , where $\rho < 1$, 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

$$W_n = [W_{n-1} + V_n - U_n]^+, \quad n \geq 1, \quad (1)$$

where $x^+ \equiv \max\{x, 0\}$. Let H_n be the cdf of W_n , which is determined by (1). Let $W \equiv W_\infty$ (both used) be the steady-state waiting time, satisfying $W_n \Rightarrow W_\infty$ as $n \rightarrow \infty$, where \Rightarrow denotes convergence in distribution; see §§X.1-X.2 of [Asmussen \(2003\)](#). The cdf H_∞ of $W \equiv W_\infty$ is the unique cdf satisfying the stochastic fixed point equation

$$W_\infty \stackrel{d}{=} (W_\infty + V - U)^+, \quad (2)$$

where $\stackrel{d}{=}$ denotes equality in distribution. If $P(W_0 = 0) = 1$, then $W_n \stackrel{d}{=} \max\{S_k : 0 \leq k \leq n\}$ for $n \leq \infty$, $S_0 \equiv 0$, $S_k \equiv X_1 + \cdots + X_k$ and $X_k \equiv V_k - U_k$, $k \geq 1$. Under the specified finite moment conditions, for $1 \leq n \leq \infty$, W_n is a proper random variable with finite mean, given by

$$E[W_n | W_0 = 0] = \sum_{k=1}^n \frac{E[S_k^+]}{k} < \infty, \quad 1 \leq n < \infty, \quad \text{and} \quad E[W_\infty] = \sum_{k=1}^{\infty} \frac{E[S_k^+]}{k} < \infty. \quad (3)$$

1.2. Classical Steady-State Results: Exact, Approximate and Bounds

For the $M/GI/1$ special case, when the interarrival time has an exponential distribution, we have the classical Pollaczek-Khintchine formula

$$E[W] = \frac{\tau\rho(1+c_s^2)}{2(1-\rho)} = \frac{\rho^2(1+c_s^2)}{2(1-\rho)}. \quad (4)$$

A natural commonly used approximation for the $GI/GI/1$ model, inspired by (4), which we call the heavy-traffic approximation, because it is motivated by the early heavy-traffic limit in [Kingman \(1961\)](#), is

$$E[W] \equiv E[W(\rho, c_a^2, c_s^2)] \approx \frac{\rho^2(c_a^2 + c_s^2)}{2(1-\rho)}. \quad (5)$$

The heavy traffic limit for the mean states that $(1-\rho)E[W(\rho, c_a^2, c_s^2)] \rightarrow (c_a^2 + c_s^2)/2$ as $\rho \uparrow 1$.

The most familiar UB on $E[W]$ is the [Kingman \(1962\)](#) bound,

$$E[W] \leq \frac{\rho^2([c_a^2/\rho^2] + c_s^2)}{2(1-\rho)}, \quad (6)$$

which also satisfies the same heavy traffic limit.

A better UB depending on these same parameters was obtained by [Daley \(1977\)](#). In particular, the [Daley \(1977\)](#) UB replaces the term c_a^2/ρ^2 by $(2-\rho)c_a^2/\rho$, i.e.,

$$E[W] \leq \frac{\rho^2([(2-\rho)c_a^2/\rho] + c_s^2)}{2(1-\rho)}. \quad (7)$$

Note that $(2-\rho)/\rho < 1/\rho^2$ because $\rho(2-\rho) < 1$ for all ρ , $0 < \rho < 1$.

In contrast to the tight UB that we study, the tight lower bound (LB) for the steady-state mean has been known for a long time; see §5.4 of [Stoyan \(1983\)](#), §V of [Whitt \(1984a\)](#), Theorem 3.1 of [Daley et al. \(1992\)](#) and references there. The LB is

$$E[W] \geq \frac{\rho((1+c_s^2)\rho - 1)^+}{2(1-\rho)}. \quad (8)$$

The LB is attained asymptotically at a deterministic interarrival time with the specified mean and at any three-point service-time distribution that has all mass on nonnegative-integer multiples of the deterministic interarrival time. The service part follows from [Ott \(1987\)](#). (All service-time distributions satisfying these requirements yield the same mean.)

2. The Main Results

In this section we state our main results. These results will be proved in following sections.

2.1. Sets of Probability Distributions with Specified Moments

Let \mathcal{P}_n be the set of all probability measures on a subset of the positive real line $[0, \infty)$ with specified first n moments. The set \mathcal{P}_n is a convex set, because the convex combination of two probability measures is just the mixture; i.e., for all p , $0 \leq p \leq 1$, $pP_1 + (1-p)P_2 \in \mathcal{P}_n$ if $P_1 \in \mathcal{P}_n$ and $P_2 \in \mathcal{P}_n$, because the n^{th} moment of the mixture is the mixture of the n^{th} moments, which is just the common value of the components. Let $\mathcal{P}_{n,k}$ be the subset of probability measures in \mathcal{P}_n that have support on at most k points.

Let $\mathcal{P}_2(m, c^2)$ be the subset of all cdf's in \mathcal{P}_2 with support in the interval $[0, \infty)$ having mean m and second moment $m^2(c^2 + 1)$. Let $\mathcal{P}_2(m, c^2, M)$ be subset of $\mathcal{P}_2(m, c^2)$ denoting all cdf's with support in the close interval $[0, M]$, where $1 + c^2 < M < \infty$ (The last property ensures that the set is non-empty.). Let subscripts a and s denote sets for the inter-arrival and service times, respectively. Therefore, $\mathcal{P}_{a,2}(1, c_a^2, M_a)$ is the set of all interarrival-time cdf's F with mean 1, scv c_a^2 and compact support within $[0, M_a]$, while $\mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ is the set of all service-time cdf's G with mean ρ , scv c_s^2 and compact support within $[0, \rho M_s]$.

A special role is played by two-point distributions, which necessarily have finite support. Let $\mathcal{P}_{2,2}(m_1, c^2, M)$ be the set of all two-point distributions with mean m_1 and second moment $m_2 = m_1^2(c^2 + 1)$ with support in $[0, m_1 M]$. The set $\mathcal{P}_{2,2}(m_1, c^2, M)$ is a one-dimensional parametric family. Any element is determined by specifying one mass point. Let $F_b^{(2)}$ be the cdf that has probability mass $c^2/(c^2 + (b-1)^2)$ on $m_1 b$, and mass $(b-1)^2/(c^2 + (b-1)^2)$ on $m_1(1 - c^2/(b-1))$ for $1 + c^2 \leq b \leq M$. The cases $b = 1 + c^2$ and $b = M$ constitute the two extremal distributions.

Since we are only interested in the extremal cdf's here, we will use different notation. We let $F_0 \equiv F_{1+c^2}^{(2)}$, because it is the unique element that has lower mass point 0 and we let $F_u \equiv F_M^{(2)}$, because it is the unique element that has upper mass point $m_1 M$. We use this definition for both the cdf's we consider: F of U and G of V , but recall that our parameter specification with $\mathbb{E}[U] = 1$ makes the support of F_u be $[0, M_a]$, while the support of G_u is $[0, \rho M_s]$. Therefore, with $M_a \geq 1 + c_a^2$ for F and $M_s \geq 1 + c_s^2$ for G , we have:

- F_0 : $c_a^2/(1+c_a^2)$ on 0 and $1/(1+c_a^2)$ on $1+c_a^2$;
- F_u : $(M_a-1)^2/(c_a^2+(M_a-1)^2)$ on $1-c_a^2/(M_a-1)$ and $c_a^2/(c_a^2+(M_a-1)^2)$ on M_a ;
- G_0 : $c_s^2/(1+c_s^2)$ on 0 and $1/(1+c_s^2)$ on $\rho(1+c_s^2)$;
- G_u : $(M_s-1)^2/(c_s^2+(M_s-1)^2)$ on $\rho(1-c_s^2/(M_s-1))$ and $c_s^2/(c_s^2+(M_s-1)^2)$ on ρM_s .

2.2. Extremal Distributions for Higher Moments of W

Ever since [Bergmann et al. \(1979\)](#) (see p. 97 of [Stoyan \(1983\)](#)), it is known that the extremal theory is quite orderly for higher moments (and cumulants) even though it is challenging for the mean. Thus, we start by applying the T system theory to the higher moments. To treat higher moments, we require that the service time V has a finite moment generating function (mgf); that implies the same is true for the transient and steady-state waiting time; see §3.1 of [Chen and Whitt \(2019\)](#) and references there. For a nonnegative random variable Z , we say that it has a finite mgf if there exists $t^* > 0$ such that

$$E[e^{tZ}] < \infty \quad \text{for } t < t^*. \quad (9)$$

That implies that all moments of Z are finite. We remark that condition (9) can be relaxed. In order for $E[W^k]$ to be finite for $k \geq 1$, it suffices to have $E[V^{(k+1)}] < \infty$, e.g., see §10.2 of [Asmussen \(2003\)](#).

THEOREM 1. (*higher steady-state moments*) *Consider the GI/GI/1 model where $F \in \mathcal{P}_{a,2}(1, c_a^2)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2)$.*

(a) *Let the service-time cdf G be fixed satisfying (9). Then*

$$E[W(F_u, G)^k] \leq E[W(F, G)^k] \leq E[W(F_0, G)^k] \quad (10)$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ and $k \geq 2$. For each and $k \geq 2$, these extrema are unique.

(b) *Let the interarrival-time cdf F be fixed. Then*

$$E[W(F, G_0)^k] \leq E[W(F, G)^k] \leq E[W(F, G_u)^k] \quad (11)$$

for all $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ and $k \geq 2$. For each and $k \geq 2$, these extrema are unique.

(c) Suppose that neither F nor G is fixed. Then

$$E[W(F_u, G_0)^k] \leq E[W(F, G)^k] \leq E[W(F_0, G_u)^k] \quad \text{for all } k \geq 2 \quad (12)$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ with $M_s < \infty$. For each and $k \geq 2$, these extrema are unique.

We prove Theorem 1 in §4 by establishing results for the transient mean in a more general model. Specifically, for part (a) we first consider the extremal F for the more general $GI_{(n)}/GI/1$ model, which allows the interarrival-time cdf to vary from period to period, provided that they all belong in $\mathcal{P}_{a,2}(1, c_a^2, M_a)$.

Given Theorem 1, it is natural to expect that corresponding results also hold for the steady mean $E[W]$. However, counterexamples to the natural analogs of Theorem 1 (a) and (b) above (without additional conditions) were provided, respectively, in §8 of [Wolff and Wang \(2003\)](#) and in §V of [Whitt \(1984a\)](#). Indeed, counterexamples for (b) are provided by Theorem 2 (b) in §2.3 below. However, we conjecture that the analog of Theorem 1 (c) is valid. Accordingly, we directly studied the distribution of $W(F_0, G_u)$ and its limiting behavior as $M_s \rightarrow \infty$ in [Chen and Whitt \(2018a\)](#). Theorem 2 there provides a tractable bound for the limit of $E[W(F_0, G_u)]$ as $M_s \rightarrow \infty$, which serves as an excellent approximation of the conjectured tight upper bound.

2.3. Extremal Distributions for the Steady-State Mean

We now turn to the more challenging problem of the mean $E[W]$. To obtain corresponding comparison results for the steady-state mean approaching Theorem 1, we will exploit stochastic-order properties for cdf's of nonnegative random variables; e.g., see §1 of [Muller and Stoyan \(2002\)](#) and Ch. 1 of [Szekli \(1995\)](#). Recall that a hyperexponential (H_k , mixtures of k exponentials) distribution is completely monotone (CM), which in turn has strictly decreasing failure rate (DFR), which has a strictly decreasing pdf, which has a strictly concave cdf; i.e., we have the implications

$$H_k \mapsto CM \mapsto DFR \mapsto \text{strictly concave cdf.} \quad (13)$$

To show the dependence of random variables on the cdf assigned to them, we will include the cdf in parentheses, so we write $U(F)$ ($V(G)$) for an interarrival time U with cdf F (service time V with cdf G). Let $W(F, G)$ denote the steady-state waiting time when the pair (U, V) have the pair of cdf's (F, G) . When we write sums of random variables as occurs in the Lindley recursion (1), we assume that the random variables are independent.

We prove part (a) of the following theorem in §5 and then apply the same methods to prove parts (b) and (c) in later sections.

THEOREM 2. (*extremal distributions for the steady-state mean*) Consider the class of GI/GI/1 queues with $F \in \mathcal{P}_{a,2}(1, c_a^2)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2)$, $0 < \rho < 1$, where $\mathcal{P}_{a,2}$ and $\mathcal{P}_{s,2}$ are nonempty.

(a) If the service-time cdf $G \in \mathcal{P}_{s,2}$ is completely monotone and $1 + c_a^2 \leq M_a \leq \infty$, then

$$W(F_u, G) \leq_{icx} W(F, G) \leq_{icx} W(F_0, G) \quad \text{for all } F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \quad (14)$$

so that

$$E[W(F_u, G)] \leq E[W(F, G)] \leq E[W(F_0, G)] \quad \text{for all } F \in \mathcal{P}_{a,2}(1, c_a^2, M_a). \quad (15)$$

The extrema in (14) and (15) are uniquely attained.

(b) If the interarrival-time cdf $F \in \mathcal{P}_{a,2}$ is strictly concave and $1 + c_s^2 \leq M_s < \infty$, then

$$W(F, G_u) \leq_{icx} W(F, G) \leq_{icx} W(F, G_0) \quad \text{for all } F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \quad (16)$$

so that

$$E[W(F, G_u)] \leq E[W(F, G)] \leq E[W(F, G_0)] \quad \text{for all } F \in \mathcal{P}_{a,2}(1, c_a^2, M_a). \quad (17)$$

If the cdf F has support in $[0, M_a]$ and is strictly convex, then (16) and (17) hold with the roles of G_0 and G_u switched.

(c) If $M_s < \infty$ and

$$E[(W(F_0, G_u) + V(G) - U(F) - t)^+] \leq E[(W(F_0, G_u) + V(G_u) - U(F_0) - t)^+] \quad \text{for all } t, \quad (18)$$

then $W(F, G) \leq_{icx} W(F_0, G_u)$ and $E[W(F, G)] \leq E[W(F_0, G_u)]$. If (18) holds for all $F \in \mathcal{P}_{a,2}(1, c_a^2)$ and $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$, then

$$\sup \{E[(F, G)] : (F, G) \in \mathcal{P}_{a,2}(1, c_a^2) \times \mathcal{P}_{s,2}(\rho, c_s^2, M_s)\} = E[W(F_0, G_u)]. \quad (19)$$

It is worthwhile to mention the $E[W(F_u, G_0)]$ is not a lower bound (see numerical study in [Chen and Whitt \(2018b\)](#)). We regard Theorem 2 (c) as a promising tool to do further analysis together with the algorithms and properties of $W(F_0, G_u)$ developed in [Chen and Whitt \(2018a\)](#). From those results, it suggests a tractable way to justify the optimum by solving an approximate version of the stochastic optimization in (18); i.e., we solve for all $t \geq 0$,

$$\sup \left\{ \frac{1}{n} \sum_{i=1}^n (W_i(F_0, G_u) + V_i(G) - U_i(F) - t)^+ : F \in \mathcal{F}, G \in \mathcal{G} \right\} = E[W(F_0, G_u)] \quad (20)$$

where \mathcal{F} is a proper finite support over $[0, M_a]$ and \mathcal{G} is also a proper finite support over $[0, \rho M_s]$.

3. Connecting to Basic T System Theory

As indicated above, we apply the theory of T systems, as reviewed in §2 of [Chen and Whitt \(2019\)](#), which draws on [Karlin and Studden \(1966\)](#). In particular, we apply Lemma 1 in §2.3 [Chen and Whitt \(2019\)](#), which is a consequence of the tractable Wronskian condition for a T system.

DEFINITION 1. (*T System*) Consider a set of $n + 1$ continuous real-valued functions $\{u_i(t) : 0 \leq i \leq n\}$ on the closed interval $[a, b]$. This set of functions constitutes a T system if the $(n + 1)^{\text{st}}$ -order determinant of the $(n + 1) \times (n + 1)$ matrix formed by $u_i(t_j)$, $0 \leq i \leq n$ and $0 \leq j \leq n$, is strictly positive for all $a \leq t_0 < t_1 < \dots < t_n \leq b$.

Equivalently, except for an appropriate choice of sign, we could instead require that every non-trivial real linear combination $\sum_{i=0}^n a_i u_i(t)$ of the $n + 1$ functions (called a u -polynomial; see §I.4 of [Karlin and Studden \(1966\)](#)) possesses at most n distinct zeros in $[a, b]$. (Nontrivial means that $\sum_{i=0}^n a_i^2 > 0$.)

We next state a consequence of Lemma 1 in §2.2.3 of [Chen and Whitt \(2019\)](#). Let $\phi^{(n)}$ denote the n^{th} derivative of the function ϕ .

LEMMA 1. (from the $(n + 1)^{\text{st}}$ derivative to a T system) Consider the real-valued functions $u_i(t) \equiv t^i$, $0 \leq i \leq n$, and ϕ on the interval $[a, b]$ for $0 \leq a < b < \infty$. Suppose that ϕ has $n + 1$ continuous derivatives. If $\phi^{(n+1)}(t) > 0$ for $a \leq t \leq b$, then $\{u_0(t), u_1(t), \dots, u_n(t), \phi(t)\}$ is a T system of functions on $[a, b]$. If $(-1)^{n+1}\phi^{(n+1)}(t) > 0$ for $a \leq t \leq b$, then $\{u_0(t), u_1(t), \dots, u_n(t), -\phi(t)\}$ is a T system of functions on $[a, b]$.

As reviewed in §2.2.1 of [Chen and Whitt \(2019\)](#), Lemma 1 applies to our setting when $n = 2$. For Theorem 2 (a), we want the UB and LB of the integral

$$\int_0^{M_a} \phi(u) dF(u), \quad (21)$$

so that we will be applying Lemma 1 over the interval $[0, M_a]$. In part (a) of our queueing extremal problem we work with the integral form in (21) with integrand

$$\phi(u) \equiv \int_0^\infty h((y - u)^+) d\Gamma(y) = h(0)\Gamma(u) + \int_{u+}^\infty h(y - u) d\Gamma(y), \quad 0 \leq u \leq M_a, \quad (22)$$

where Γ is a cdf of a nonnegative real-valued random variable Y with a finite moment generating function (mgf); i.e., satisfying (9).

LEMMA 2. If the condition of Lemma 1 is satisfied with $n = 2$ and $(-1)^3\phi^{(3)}(u) > 0$ for $0 \leq u \leq M_a$, then

$$\sup \left\{ \int_0^{M_a} \phi(u) dF(u) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \right\} = \int_0^{M_a} \phi(u) dF_0(u) \quad (23)$$

and

$$\inf \left\{ \int_0^{M_a} \phi(u) dF(u) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \right\} = \int_0^{M_a} \phi(u) dF_u(u). \quad (24)$$

If the condition of Lemma 1 is satisfied with $n = 2$ and $(-1)^3\phi^{(3)}(u) < 0$ for $0 \leq u \leq M_a$, then the roles of F_0 and F_u are switched in (23) and (24).

We now give sufficient conditions on h and the cdf Γ in (22) for the system $\{1, u, u^2, -\phi(u)\}$ to be a T system on $[0, M_a]$. For a real-valued function h of a real variable that has at least k continuous derivatives, let $h^{(k)}$ denote its k^{th} derivative; let $h^{(0)} \equiv h$. Let 1_A be the indicator function of the set A , which equals 1 on A and 0 on its complement. For part of this result, we will be assuming that the cdf Γ has a smooth pdf γ , but we will relax that assumption in §6.1.

LEMMA 3. (condition for the third derivative to be negative) Consider a nonnegative real-valued random variable Y with a finite mgf (satisfying (9)) and the cdf Γ with support in $[a, b]$ or $[a, b)$ such that

$$0 \leq a < M_a \leq b \leq \infty. \quad (25)$$

For ϕ in (22), in order to have

$$(-1)^3 \phi^{(3)}(u) > 0 \quad \text{for } 0 \leq u \leq M_a, \quad (26)$$

so that for $\{1, u, u^2, -\phi(u)\}$ to be a T system on $[0, M_a]$, implying that F_0 attains the UB in (23), while F_u attains the LB (24), each of the following is a sufficient condition:

(i) $h(x) \equiv x$ and Γ has a positive pdf γ that is differentiable with $\gamma^{(1)}(x) < 0$ for $a \leq x \leq M_a$,

(ii) $h(x) \equiv x^2$ and Γ has a positive pdf γ for $a \leq x \leq M_a$,

(iii) $h(x) \equiv h(x; p) \equiv x^p$ for $p \geq 3$,

(iv) $h(x) \equiv h(x; t) \equiv e^{tx} - tx - \frac{(tx)^2}{2} - \frac{(tx)^3}{6} = 1 + \sum_{k=4}^{\infty} \frac{(tx)^k}{k!}$ for $t > 0$,

(v) $h^{(k)}(x) > 0$, $a < x \leq M_a$, $0 \leq k \leq 3$ and $h^{(k)}(a) = 0$, $1 \leq k \leq 2$.

For the function $h(x) \equiv x$ in condition (i), the condition on γ is necessary as well as sufficient, given that γ has a continuous positive derivative. In condition (i), if instead $\gamma^{(1)}(x) > 0$ for $0 \leq x \leq M_a$, then the roles of F_0 and F_u are switched in (23) and (24).

Proof. First, observe that condition (9) implies that all integrals are finite. Next, we consider what happens if $0 \leq u \leq a$ with $a > 0$ or $u > b$. First, if $u \leq a$, then $\phi(u) = E[h(Y - u)]$, $0 \leq u \leq a$, so that the desired property of ϕ holds over $[0, a]$. In particular,

$$\phi^{(3)}(u) = - \int_a^{\infty} h^{(3)}(y - u) d\Gamma(y) < 0, \quad 0 \leq u \leq a. \quad (27)$$

On the other hand, if $u \geq b$, then $\phi(u) = h(0)$, so that the desired property cannot hold for $u > b$. However, we have ruled that case out by assuming that $M_a \leq b$. It suffices for Γ to have a pdf over $[a, M_a]$.

In each case we can apply Lemmas 1 and 2 with (22). To do so, we apply the Leibniz rule for differentiation of an integral with (22). Using that condition with $a \leq u \leq M_a$, we have

$$\begin{aligned}\phi(u) &= \int_a^\infty h((y-u)^+) d\Gamma(y) = \int_u^\infty h(y-u) d\Gamma(y) + h(0)\Gamma(u) \quad \text{and} \\ \phi^{(1)}(u) &= - \int_u^\infty h^{(1)}(y-u) d\Gamma(y) - h(0)\gamma(u) + h(0)\gamma(u) = - \int_u^\infty h^{(1)}(y-u) d\Gamma(y).\end{aligned}\quad (28)$$

For $h(x) \equiv x$ in condition (i), we have $h^{(1)}(x) = 1$ for all x , so that

$$\phi^{(1)}(u) = - \int_u^\infty h^{(1)}(y-u) d\Gamma(y) = - \int_u^\infty d\Gamma(y) = -(1 - \Gamma(u)),\quad (29)$$

so that, by the condition on Γ ,

$$\phi^{(2)}(u) = \gamma(u) > 0 \quad \text{and} \quad \phi^{(3)}(u) = \gamma^{(1)}(u) < 0 \quad \text{for} \quad u \geq a.\quad (30)$$

From the form of $\phi^{(3)}(u)$ in (30), we see that the condition on γ is necessary as well as sufficient.

We also see that the UB and LB are switched if instead $\gamma^{(1)}(u) > 0$.

Turning to $h(x) = x^2$ in condition (ii), we use $h^{(1)}(0) = 0$ and $h^{(2)}(x) = 2$ for all x with the second line of (28) to get

$$\phi^{(2)}(u) = \int_u^\infty h^{(2)}(y-u) d\Gamma(y) = 2 \int_u^\infty d\Gamma(y) = 2(1 - \Gamma(u)) > 0,\quad (31)$$

so that $\phi^{(3)}(u) = -2\gamma(u) < 0$ for $a \leq u \leq M_a$.

Conditions (iii) and (iv) are both special cases of condition (v), which implies that

$$\phi^{(3)}(u) = - \int_u^\infty h^{(3)}(y-u) d\Gamma(y) < 0. \quad \blacksquare\quad (32)$$

4. Proof of Theorem 1

We prove Theorem 1 by establishing results for the transient higher moments. in a more general model.

4.1. Proof of Theorem 1 (a)

We start by considering part (a). To do so, we first consider the extremal F for the more general $GI_{(n)}/GI/1$ model, which allows the interarrival-time cdf to vary from period to period, provided that they all belong in $\mathcal{P}_{a,2}(1, c_a^2, M_a)$.

LEMMA 4. (*higher transient moments for time-varying arrivals*) Consider the $GI_{(n)}/GI/1$ model where the fixed service-time cdf G satisfies (9) and is in $\mathcal{P}_{s,2}(\rho, c_s^2)$. Let the initial waiting time $W_0 \equiv W_0(F, G)$ also satisfy (9) and be independent of future interarrival times and service times. Then, regardless of the choice of interarrival-time cdf's in previous periods, and thus independent of W_{n-1} ,

$$E[W_n(F_u, G)^k | W_{n-1}] \leq E[W_n(F, G)^k | W_{n-1}] \leq E[W_n(F_0, G)^k | W_{n-1}] \quad w.p.1 \quad (33)$$

for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$, $1 + c_a^2 \leq M_a < \infty$, $n \geq 1$ and $k \geq 2$. For each $n \geq 1$ and $k \geq 2$, these extrema are unique.

Proof. We first directly treat the case of $M_a < \infty$, but since the extremal cdf is independent of M_a , that constraint can be dropped. It suffices to apply Lemma 3 under conditions (ii) and (iii) with $M_a < \infty$ and the Lindley recursion in (1). To directly apply Lemma 3 under conditions (ii), we need to assume that the service time cdf G has a positive pdf. That approximation step is justified in §6.1. We exploit the fact that W_n and Y_n satisfy (9) when V does. Uniqueness follows from the T-system theory. ■

4.2. Proof of Theorem 1 (b)

By a minor modification of the same reasoning (see §6.2), we obtain the following analog for the extremal G in the $GI/GI_{(n)}/1$ model, which allows the service-time cdf to vary from period to period, provided that they all belong in $\mathcal{P}_{s,2}(\rho, c_s^2, M_s)$. We do not need to assume condition (9), because it automatically holds if G has bounded support.

COROLLARY 1. (*higher transient moments for time-varying service*) Consider the $GI/GI_{(n)}/1$ model with fixed interarrival-time cdf F in $\mathcal{P}_{a,2}(1, c_a^2)$. Let the initial waiting time W_0 satisfy (9) and be independent of future interarrival times and service times. Then, regardless of the choice of service-time cdf's in previous periods, and thus independent of W_{n-1} ,

$$E[W_n(F, G_0)^k | W_{n-1}] \leq E[W_n(F, G)^k | W_{n-1}] \leq E[W_n(F, G_u)^k | W_{n-1}] \quad w.p.1 \quad (34)$$

for all $G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$, $1 + c_s^2 \leq M_s < \infty$, $n \geq 1$ and $k \geq 2$. For each $n \geq 1$ and $k \geq 2$, these extrema are unique.

4.3. The $GI/GI/1$ Model

As an immediate consequence of Lemma 4 and Corollary 1, we get the corresponding result for the $GI/GI/1$ model, because the extremal cdf's for the $GI_{(n)}/GI/1$ and $GI/GI_{(n)}/1$ models are the same in every period, and so reduce to the special $GI/GI/1$ case. These results for all transient models then in turn imply corresponding results for the steady-state waiting time, as we show in §6.4. Uniqueness below follows from the uniqueness in the one-stage problem, starting with the steady-state waiting time with an extremal cdf.

5. Proof of Theorem 2 (a)

We next prove Theorem 2 (a): finding the extremal interarrival-time cdf F on $[0, M_a]$ for the mean $E[W(F, G)]$ for any given service-time cdf G . We again apply the theory of T systems. After treating part (a), we apply the same methods to treat Theorem 2 (b) and (c).

5.1. Optimality for the $GI/GI/1$ Queue

For Theorem 2 (a), we apply Lemma 3 (i) with the random variable Y being the sojourn time, i.e., the time spent by the arrival in the system, also called the response time. It is the sum of two independent nonnegative random variables, one being a service time V and the other steady-state waiting time W or the transient waiting time W_n for $n \geq 0$. Let $Y \equiv W + V$ and let $Y_n \equiv W_n + V$. Let H and H_n be the cdf of W and W_n , respectively. Let $Y_n(H_0, F, G)$ and $Y(F, G)$ denote the dependence of Y_n and Y on the underlying cdf triple (H_0, F, G) of (W_0, U, V) and similarly for other random variables. (Recall that the steady-state distributions are independent of the initial conditions, assuming a finite mean $E[W_0]$). We can apply the previous results to deduce the following two theorems.

For the first theorem, we exploit increasing convex stochastic order, denoted by \leq_{icx} , and state results in that form. For real-valued random variables, $Z_1 \leq_{icx} Z_2$ if $E[f(Z_1)] \leq E[f(Z_2)]$ for all nondecreasing convex functions f for which the expectations are well defined; e.g., see §1.5 of Muller and Stoyan (2002).

THEOREM 3. (*a one-step condition for ordered steady-state means*) Consider the $GI/GI/1$ model with given service-time cdf G satisfying (9). Let $W_0 \stackrel{d}{=} W(F_1, G)$ and $Y_0 \stackrel{d}{=} Y(F_1, G)$, where $W(F, G)$

and $Y(F, G)$ are the steady-state waiting time and sojourn time using $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$. For any $F_2 \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$, if

$$\int_0^{M_a} E[(Y_0(F_1, G) - t - u)^+] dF_2(u) \leq \int_0^{M_a} E[Y_0(F_1, G) - t - u]^+ dF_1(u) \quad \text{for all } t \geq -M_a, \quad (35)$$

then

$$W_n(F_2, G) \leq_{icx} W_{n-1}(F_2, G) \quad \text{and} \quad Y_n(F_2, G) \leq_{icx} Y_{n-1}(F_2, G) \quad \text{for } n \geq 1 \quad (36)$$

for

$$\begin{aligned} Y_n(F_i, G) &\equiv (W_n(F_i, G) + V(G)) \quad \text{for } n \geq 2, \quad Y_1(F_i, G) \equiv W_0 + V(G) = W(F_1, G) - U(F_i)^+, \\ W_n(F_i, G) &\equiv (Y_{n-1}(F_i, G) - U(F_i))^+ \quad \text{for } n \geq 2 \quad \text{and} \quad W_1(F_i, G) \equiv (Y_0 - U(F_i))^+. \end{aligned} \quad (37)$$

Hence,

$$W(F_1, G) \geq_{icx} W(F_2, G) \quad \text{and} \quad Y(F_1, G) \geq_{icx} Y(F_2, G) \quad (38)$$

and thus

$$E[W(F_1, G)] \geq E[W(F_2, G)] \quad \text{and} \quad E[Y(F_1, G)] \geq E[Y(F_2, G)]. \quad (39)$$

Proof. We start by observing that the increasing convex stochastic ordering

$$Y(F_1, G) - U(F_1) \geq_{icx} Y(F_1, G) - U(F_2), \quad (40)$$

where the random variables Y and U are independent, is equivalent to the expectation orderings

$$E[(Y(F_1, G) - U(F_1) - t)^+] \geq E[(Y(F_1, G) - U(F_2) - t)^+] \quad \text{for all } t, \quad (41)$$

by virtue of Theorem 1.5.7 of [Muller and Stoyan \(2002\)](#). We can then rewrite (41) equivalently as

$$\int_0^{M_a} E[(Y(F_1, G) - u - t)^+] dF_1(u) \geq \int_0^{M_a} E[(Y(F_1, G) - u - t)^+] dF_2(u) \quad \text{for all } t. \quad (42)$$

Since U has support in $[0, M_a]$, we only need to consider $t \geq -M_a$. Thus, the condition in (35) is equivalent to each of the expressions in (40)-(42).

Now, given (40), because $(x)^+$ is a nondecreasing convex function, we have

$$W_0 \equiv W(F_1, G) \stackrel{d}{=} (Y(F_1, G) - U(F_1))^+ \geq_{icx} (Y(F_1, G) - U(F_2))^+ \equiv W_1. \quad (43)$$

where $W_1 \equiv W_1(F_1, F_2, G) \equiv [W(F_1, G) + V(G) - U(F_2)]^+$. Then, by Theorem 1.5.5 (b) of [Muller and Stoyan \(2002\)](#), we see that the order is maintained if we add the same independent random variable from both sides. That gives

$$Y_0 \equiv W_0 + V(G) \geq_{icx} W_1 + V(G) \equiv Y_1, \quad (44)$$

where independence is assumed in the sums, as usual. Then, by Theorem 1.5.5 (b) of [Muller and Stoyan \(2002\)](#) again, we see that the order is maintained if we subtract the same independent random variable from both sides. Hence, from (44) we deduce that

$$Y_0 - U(F_2) \geq_{icx} Y_1 - U(F_2). \quad (45)$$

Then, because $(x)^+$ is a nondecreasing convex functions of x , we have

$$W_1 = (Y_0 - U(F_2))^+ \geq_{icx} (Y_1 - U(F_2))^+ \equiv W_2. \quad (46)$$

By the same reasoning, we deduce recursively, and using mathematical induction, that

$$W_{n-1} \geq_{icx} W_n \quad \text{and} \quad Y_{n-1} \geq_{icx} Y_n \quad \text{for all } n \geq 1. \quad (47)$$

But then observe that $(W_n, Y_n) \Rightarrow (W(F_2, G), Y(F_2, G))$ as $n \rightarrow \infty$, so that we can apply Theorem 1.5.9 of [Muller and Stoyan \(2002\)](#) to deduce (38), which of course implies (39). ■

The following theorem provides a sufficient condition for the condition in Theorem 3 in terms of the steady-state sojourn time $Y(F_0, G)$. We remark that it is known that the steady-state waiting-time cdf is always new worse than used (NWU), is concave if the service-time cdf is has decreasing failure rate (is DFR), and is completely monotone if the service-time cdf is completely monotone; see §§1.7-1.9 of [Szekli \(1995\)](#). However, these properties are not preserved under convolution in general.

THEOREM 4. (*concavity condition for F_0*) Consider the GI/GI/1 model with given service-time V having cdf G with support in $[a, \infty)$ for $0 \leq a < \rho = E[V]$ and a finite mgf (satisfying (9)). If the

sojourn-time cdf $\Gamma(x) \equiv P(Y(F_0, G) \leq x)$ is strictly concave in x over $[a, \infty)$, then condition (35) in Theorem 3 is satisfied for $F_1 = F_0$ and for all $F_2 \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$, so that

$$Y(F, G) \leq_{icx} Y(F_0, G) \quad \text{for all } F \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \quad (48)$$

and

$$\sup \{E[W(F, G)] : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)\} = E[W(F_0, G)]. \quad (49)$$

If the sojourn-time cdf $\Gamma(x) \equiv P(Y(F, G) \leq x)$ is strictly concave in x over $[a, \infty)$ for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$, then condition (35) in Theorem 3 is satisfied for all $F_2 = F_u$ and $F_1 = F$ for all F in $\mathcal{P}_{a,2}(1, c_a^2, M_a)$, so that

$$Y(F_u, G) \leq_{icx} Y(F, G) \quad \text{for all } F \in \mathcal{P}_{a,2}(1, c_a^2, M_a), \quad (50)$$

and

$$\inf \{E[W(F, G)] : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)\} = E[W(F_u, G)]. \quad (51)$$

Proof. The condition on the cdf Γ in Theorem 4 implies condition (35) in Theorem 3. That implication follows by combining Lemma 3 when the service-time pdf under condition (i) when the service-time strictly decreasing pdf, which we have not yet assumed. However, it is possible to treat the more general case by an additional asymptotic argument, as we indicate in §6.1 below. In particular, we apply Lemmas 5 and 6. The concavity of the cdf Γ over the entire interval $[a, \infty)$ is important for covering the subtraction by t in condition (35). ■

We now provide a sufficient condition for the conditions in Theorem 4. Recall that a cdf G on $[0, \infty)$ is completely monotone if it is a mixture of exponential cdf's, i.e., if

$$G(x) = \int_0^\infty (1 - e^{-\lambda x}) dP(\lambda)$$

for some probability measure P .

THEOREM 5. (*hyperexponential sojourn-time distribution*) *In the GI/GI/1 queue, if the service-time distribution is H_k , then so is the sojourn-time distribution. Hence, the concavity conditions on the sojourn-time cdf Γ in Theorem 4 are satisfied for all $F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ if the service-time cdf G is hyperexponential or completely monotone.*

Proof. For the $GI/PH/1$ queue with any interarrival-time cdf, it is known that the waiting time and sojourn time cdf's inherit the phase-type (PH) matrix structure of the service-time cdf; see pp. 46 and 151 of Neuts (1981) or Corollary 2.2 of Asmussen (1992). From the diagonal structure of the substochastic matrix in the PH representation when the PH distribution is H_k , we see that the distributions of W and Y are also H_k when the cdf G of V is H_k . We continue to provide a detailed proof. To do so, we use the known fact that the waiting-time cdf is H_k with an atom at the origin; see §II.5.10 of Cohen (1982) or Theorem 1.9 D (i) of Szekli (1995). Thus, from equation (4.1.41) of Neuts (1981), we deduce that the matrix function $\Theta(x)$ that appears there in the cdf of W , and is characterized in previous equations, is a diagonal matrix with exponential functions $e^{-\mu_i x}$ appearing on the diagonal. (The constant matrices appearing there are all strictly positive, so that there is no cancellation.) Hence, this must also be true for the matrix $\Psi(x)$ in equation (4.1.40) of Neuts (1981). Thus, the ODE for the function $\Psi_1(x)$ involves diagonal matrices S and $\Psi(x)$. Hence, $\Psi_1(x)$ has a diagonal matrix function solution. Thus, the distribution of Y given in equation (4.1.44) is a mixture of exponentials. Given that H_k structure, the cdf Γ has a strictly decreasing pdf, so that the concavity condition is satisfied. The general completely monotone case can be represented as the limit of H_k cdf's, using §6.1. The asymptotic argument was used for the $GI/GI/k$ waiting time by Szekli (1987) ■

We now observe that the concavity condition for the sojourn-time cdf cannot be satisfied for any $GI/E_k/1$ queue for $k \geq 2$ or any queue where the service time is a finite mixture of E_k distributions all of which have $k \geq 2$.

PROPOSITION 1. (*negative results for Erlang service*) *For any service-time cdf G that has a pdf g that is differentiable and strictly increasing over $[0, x]$ for some $x > 0$, the pdf γ of the associated steady-state sojourn-time $Y \equiv W + V$ must be strictly increasing over $[0, x]$, so that the cdf Γ cannot be strictly concave.*

Proof. Let H be the cdf of W , which has an atom at 0. Given that V has a continuous pdf, $Y \equiv V + W$ has a pdf γ , where

$$\gamma(t) = H(0)g(t) + \int_{0+}^t g(t-u) dH(u), \quad t \geq 0. \quad (52)$$

Consequently, the derivative satisfies

$$\gamma^{(1)}(t) = H(0)g^{(1)}(t) + \int_{0+}^t g^{(1)}(t-u) dH(u), \quad (53)$$

which is strictly positive for $0 \leq t \leq x$ for sufficiently small x by the assumption on g . ■

6. Remaining Proofs

We now provide the remaining proofs.

6.1. Relaxing the PDF Condition in Lemma 3

We now relax the pdf condition on Γ in Lemma 3 under conditions (i) and (ii) above. Recall that convergence in distribution can be expressed in terms of cdf's; i.e., corresponds to pointwise convergence at all points x that are continuity points of the limiting cdf. Let \Rightarrow denote convergence in distribution.

LEMMA 5. (*preservation of optimality*) Suppose that $\{Y_n : n \geq 1\}$ is a sequence of real-valued random variables such that the conditions of Lemma 3 are satisfied for each $n \geq 1$ and $Y_n \Rightarrow Y$ as $n \rightarrow \infty$. If F_0 (F_u) yields the UB for (23) and F_u (F_0) yields the LB in (24) for all $n \geq 1$, then the same is true for the limit Y .

Proof. We directly compare F_0 to any alternative cdf F for the UB. First, by the continuous mapping theorem, we obtain

$$\phi_n(u) \rightarrow \phi(u) \quad \text{as } n \rightarrow \infty \quad (54)$$

for each u from (22). Then, by the dominated convergence theorem,

$$\int_0^{M_a} \phi(u) dF(u) = \lim_{n \rightarrow \infty} \int_0^{M_a} \phi_n(u) dF(u) \leq \lim_{n \rightarrow \infty} \int_0^{M_a} \phi_n(u) dF_0(u) = \int_0^{M_a} \phi(u) dF_0(u). \quad (55)$$

Hence, F_0 remains optimal for the limit. Essentially the same argument applies to the lower bound. ■

LEMMA 6. (*preservation of optimality for the first two moments*) In order to apply Lemma 5 to condition (i) in Lemma 3 with $\gamma^{(1)}(x) < 0$, $a \leq x \leq M_a$, it suffices to have, in addition to $Y_n \Rightarrow Y$, the cdf's Γ of Y be strictly concave over $[a, M_a]$, i.e., satisfy $\Gamma(x+\delta) - \Gamma(x)$ be strictly decreasing in x over $[a, M_a]$ for all $\delta > 0$. For condition (ii) in Lemma 3, it suffices to have $\Gamma(x)$ be strictly decreasing in x over $[a, M_a]$.

Proof. If Γ has the stated property, then Γ can be made the limit of cdf's Γ_n with the properties stated in Lemma 3. ■

6.2. Proof of Theorems 1 (b) and 2 (b)

The proofs for part (b) can be short, because we can apply variants of the proof for part (a). For part (b), we are concerned with

$$\sup \left\{ \int_0^\infty E[(W(F, G) + v - U)^+]^k dG(v) : G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s) \right\}. \quad (56)$$

It is convenient to use a reverse-time formulation and work with the cdf \tilde{G} of $\tilde{V} \equiv \rho M_s - V$, and adjusting the moments consistently. Then we can focus on

$$\sup \left\{ \int_0^\infty E[(W(F, G) - U + \rho M_s - v)^+]^k d\tilde{G}(v) : \tilde{G} \in \mathcal{P}_{s,2}(\tilde{\rho}, \tilde{c}_s^2, M_s) \right\}. \quad (57)$$

Then we can apply Lemma 3 as in the proof of Theorem 1 (a) before.

For the proof of Theorem 2 (b) with $k = 1$, let $\phi_s(v) \equiv E[(W(F, G) + \rho M_s - v - U)^+]$ be the integrand in the integral in (57) with $k = 1$. We make structural assumptions about the cdf's F of U and H of W , which can be relaxed by the asymptotic methods of §6.1.

LEMMA 7. *If (i) the cdf F is differentiable with a strictly positive pdf f that can be expressed as*

$$f(u) = \int_0^u f^{(1)}(x) dx, \quad u \geq 0, \quad (58)$$

where $f^{(1)}$ is integrable, and (ii) W has a cdf H with $H(0) > 0$ and

$$H(x) = H(0) + \int_0^x h(w) dw \quad x \geq 0, \quad (59)$$

where h is strictly positive and integrable over the halfline, then the integrand ϕ_s in (57) with $k = 1$ can be expressed as

$$\phi_s(v) = H(0)E[(\rho M_s - v - U)^+] + \int_0^\infty h(w)E[(w + \rho M_s - v - U)^+] dw > 0, \quad (60)$$

so that the first three derivatives of ϕ_s exist for $v > 0$ and, with satisfy

$$\begin{aligned} \phi_s^{(1)}(v) &= -P(U - W \leq \rho M_s - v) = -H(0)F(\rho M_s - v) - \int_0^\infty h(w)F(w + \rho M_s - v) dw < 0, \\ \phi_s^{(2)}(v) &= \theta(v) = H(0)f(\rho M_s - v) + \int_0^\infty h(w)f(w + \rho M_s - v) dw > 0, \\ \phi_s^{(3)}(v) &= \dot{\theta}(v) = -H(0)f^{(1)}(\rho M_s - v) - \int_0^\infty h(w)f^{(1)}(w + \rho M_s - v) dw > 0, \quad v \geq 0, \end{aligned} \quad (61)$$

where $\theta(v)$ is the pdf of $U - W$ over $[0, \rho M_s]$ because $v \in [0, \rho M_s]$. So that ϕ_s is strictly positive, strictly decreasing and strictly convex on $[0, \rho M_s]$. Moreover, from (61) we see that if $f^{(1)} > 0$, then $\phi_s^{(3)}(v) < 0$ as well.

Thus, the previous proof applies until we come to Theorem 5, but we now need to replace the steady-state sojourn time $Y = W + V$ by $U - W$, where W is the steady-state waiting time. Fortunately, the analog of Theorem 5 is already covered by Lemma 7.

6.3. Proof of Theorem 2 (c)

The proof here is essentially the same as the proof of Theorem 3. As before, we establish increasing convex stochastic order as we move from one steady-state distribution to another through a sequence of transient distributions, based on the Lindley recursion (1),

6.4. Completing the Proof of Theorem 1

For the higher moments we can apply Lemma 3 without having to worry about the pdf γ appearing in condition (i). However, we have yet to prove that we can obtain the results for the steady-state waiting time from the corresponding result for all transient waiting times. We complete that step here. We will state and prove a result for the steady-state mean, but a corresponding result holds for higher moments under the extra condition (9) by the same reasoning.

LEMMA 8. (*reduction to the transient mean*) Consider the GI/GI/1 queue. Let the initial waiting time W_0 have fixed cdf H_0 with $E[W_0] < \infty$.

(a) For any specified $G \in \mathcal{P}_{s,2}(\rho, c_s^2)$, if there exists $F_n^* \in \mathcal{P}_{a,2}(1, c_a^2, M_a)$ such that

$$E[W_n(F_n^*, G)] = \sup \{E[W_n(F, G) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)]\} \quad \text{for all } n \geq 1, \quad (62)$$

then the sequence $\{F_n^* : n \geq 1\}$ is tight, so that there exists a convergent subsequence. Moreover, if F_∞^* is the limit of any convergent subsequence, then F_∞^* is in $\mathcal{P}_{a,2}(M_a)(1, c_a^2, M_a)$ and F_∞^* is optimal for the steady-state mean $E[W(F, G)]$, i.e.,

$$E[W(F_\infty^*, G)] = \sup \{E[W(F, G) : F \in \mathcal{P}_{a,2}(1, c_a^2, M_a)]\}. \quad (63)$$

(b) For any specified $F \in \mathcal{P}_{a,2}(1, c_a^2)$, if there exists $G_n^* \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ such that

$$E[W_n(F, G_n^*)] = \sup \{E[W_n(F, G) : G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)]\} \quad \text{for all } n \geq 1, \quad (64)$$

then the sequence $\{G_n^* : n \geq 1\}$ is tight, so that there exists a convergent subsequence. Moreover, if G_∞^* is the limit of any convergent subsequence, then G_∞^* is in $\mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ and G_∞^* is optimal for $E[W(F, G)]$, i.e.,

$$E[W(F, G_\infty^*)] = \sup \{E[W(F, G) : G \in \mathcal{P}_{s,2}(\rho, c_s^2, M_s)]\}. \quad (65)$$

(c) If there exists (F_n^*, G_n^*) in $\mathcal{P}_{a,2}(1, c_a^2, M_a) \times \mathcal{P}_{s,2}(1, c_s^2, M_s)$ such that

$$E[W_n((F_n^*, G_n^*))] = \sup \{E[W_n(F, G) : (F, G) \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \times \mathcal{P}_{s,2}(1, c_s^2, M_s)]\} \quad (66)$$

for all $n \geq 1$, then the sequence $\{(F_n^*, G_n^*) : n \geq 1\}$ is tight, so that there exists a convergent subsequence. Moreover, if (F_∞^*, G_∞^*) is the limit of any convergent subsequence, then (F_∞^*, G_∞^*) is in $\mathcal{P}_{a,2}(1, c_a^2, M_a) \times \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ and the pair (F_∞^*, G_∞^*) is optimal for $E[W(F, G)]$, i.e.,

$$E[W((F_\infty^*, G_\infty^*))] = \sup \{E[W(F, G) : (F, G) \in \mathcal{P}_{a,2}(1, c_a^2, M_a) \times \mathcal{P}_{s,2}(1, c_s^2, M_s)]\}. \quad (67)$$

Proof of Lemma 8. We only prove (c), because the others are proved in the same way. As observed before, because the support sets $[0, M_a]$ and $[0, \rho M_s]$ are compact intervals, the spaces $\mathcal{P}_{a,2}(1, c_a^2, M_a)$, $\mathcal{P}_{s,2}(\rho, c_s^2, M_s)$ and their product are compact metric spaces. Hence the tightness follows, which implies that there exists a convergent subsequence by Prohorov's theorem in §11.6 of Whitt (2002) and the limit (F^*, G^*) of any such subsequence $\{(F_{n_k}^*, G_{n_k}^*) : k \geq 1\}$ must remain in the space $\mathcal{P}_{a,2}(1, c_a^2, M_a) \times \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$. Moreover, the associated sequence of steady-state waiting times $\{W_\infty(F_{n_k}^*, G_{n_k}^*) : k \geq 1\}$ converges in distribution to $W_\infty(F^*, G^*)$ and the means converge as well, by the continuity results for $GI/GI/1$ in §X.6 of Asmussen (2003).

We conclude by demonstrating optimality. Suppose that (F', G') is another candidate pair of cdf's in $\mathcal{P}_{a,2}(1, c_a^2, M_a) \times \mathcal{P}_{s,2}(\rho, c_s^2, M_s)$. By the assumed optimality, we must have $w_{n_k}(F_{n_k}^*, G_{n_k}^*) \geq w_{n_k}(F', G')$ for all k . Then, by continuity, using §X.6 of Asmussen (2003), we conclude that $w_\infty^\uparrow = w(F^*, G^*)$ for the steady-state mean. ■

7. Supporting Simulation Results

In this section we present results of simulation experiments illustrating the the shape conditions for parts (a) and (b) of Theorem 2.

7.1. Supporting Simulation Results for F

We now present supporting simulation results for the shape of the steady-state sojourn time $Y \equiv W + V$ asserted in Theorem 5 and thus for the extremal results in Theorems 3 and 4. We consider the supremum and infimum for F , e.g.,

$$E[W(F_0, H_k)] = \sup\{E[W(F, H_k)] : F \in \mathcal{P}_{a,2}(M_a)\}, k \geq 2. \quad (68)$$

Figure 1 shows that for the examples $F_0/H_2/1$ and $F_u/H_2/1$ ($M_a = 10$) with balanced means under traffic level $\rho = 0.5$, the pdf γ of Y under the same simulation settings in both cases has a monotone density. In these cases, the estimated mean steady-state values are $E[W(F_u, H_2)] = 2.00$ and $E[W(F_0, H_2)] = 3.28$.

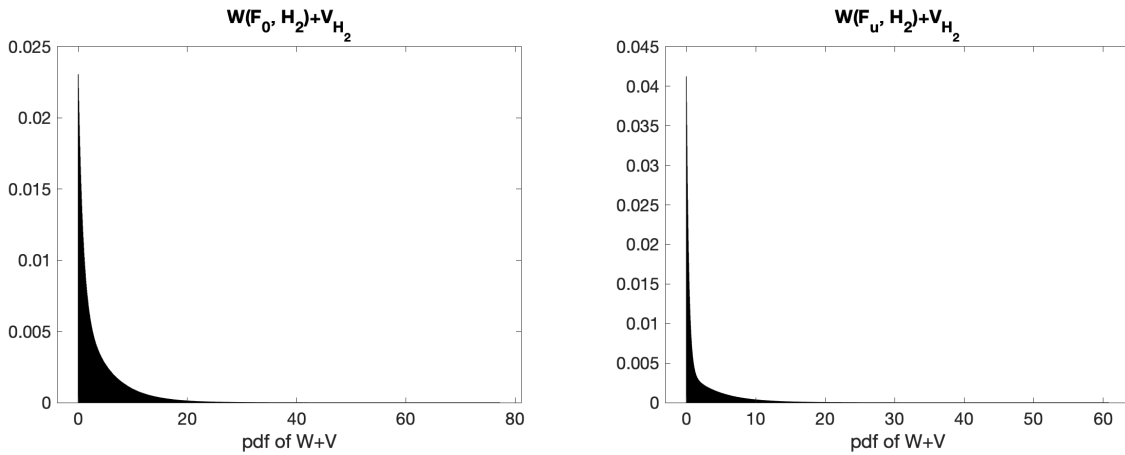


Figure 1 The shape of the pdf of $W(F_0, H_2) + V_{H_2}$ and $W(F_u, H_2) + V_{H_2}$ for H_2 service with $M_a = 10$, $\rho = 0.5$ and $c_a^2 = c_s^2 = 4$

In addition, Figure 2 shows supporting simulation results for $F_0/H_5/1$ and $F_u/H_5/1$ ($M_a = 10$) where H_5 has the service rates $[1.0, 1.5, 2.0, 2.5, 3.0]$ with the respective probabilities $[0.1, 0.15, 0.2, 0.25, 0.3]$. In these cases, $E[W(F_u, H_5)] = 1.03$ and $E[W(F_0, H_5)] = 2.75$.

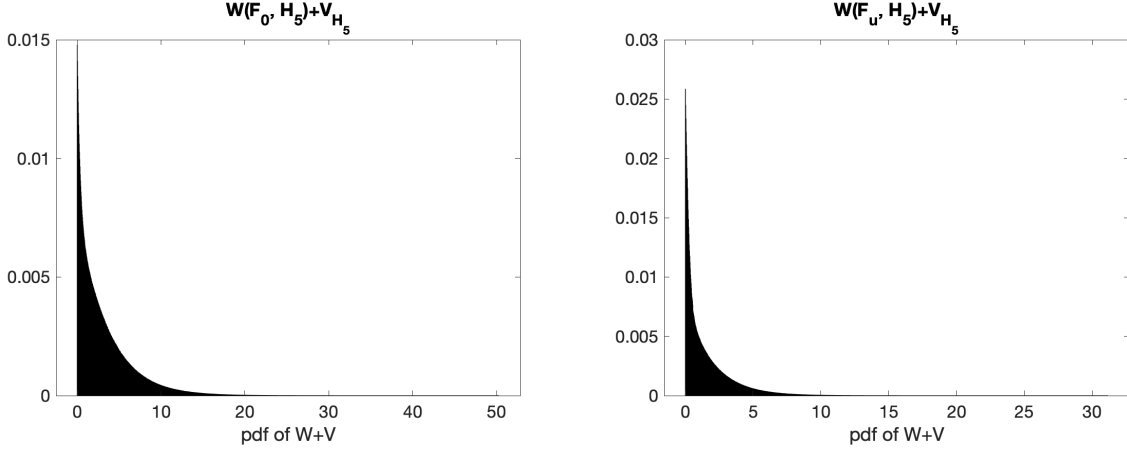


Figure 2 The shape of the pdf of $W(F_0, H_5) + V_{H_5}$ and $W(F_u, H_5) + V_{H_5}$ for H_5 service (as specified in the text) with $M_a = 10$, $\rho = 0.5$ and $c_a^2 = c_s^2 = 4$

Figure 3 shows corresponding results for E_2 service. Figure 3 shows that the pdf of $W + V$ is nearly monotone, but a careful examination shows that the pdf is increasing over a short interval $[0, x]$. In these cases, $E[W(F_u, E_2)] = 0.528$ and $E[W(F_0, E_2)] = 2.55$.

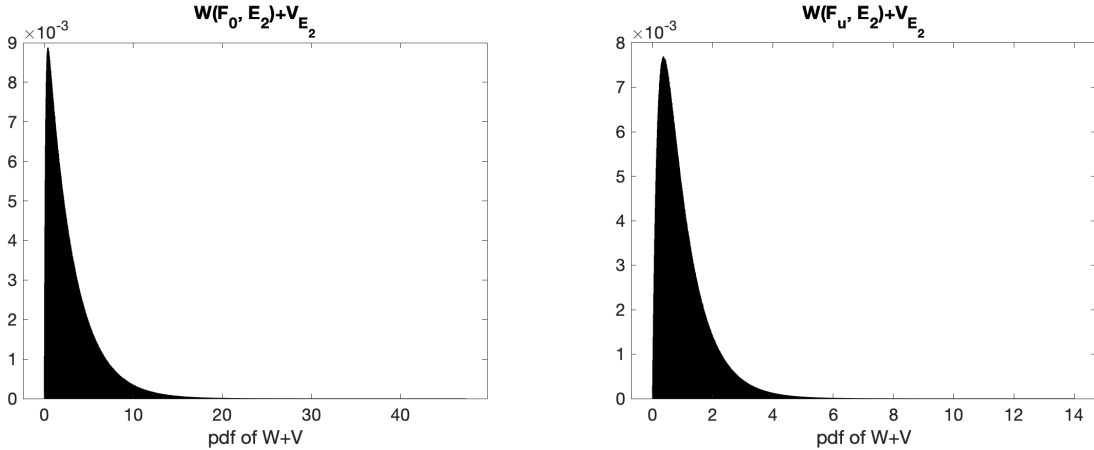


Figure 3 The shape of the pdf of $W(F_0, E_2) + V_{E_2}$ and $W(F_u, E_2) + V_{E_2}$ for Erlang E_2 service with $M_a = 10$, $\rho = 0.5$, $c_a^2 = 4$ and $c_s^2 = 0.5$

7.2. Supporting Simulation Results for G

Figure 4 experimentally confirms the conclusion of Lemma 7 in §6.2. that the condition of Theorem 2 (b) can be satisfied by presenting simulation estimates of the pdf of $W(H_2, G_0) - U_{H_2}$

and $W(H_2, G_u) - U_{H_2} \leq 0$ with $M_s = 10$ and $\rho = 0.5$ for the case $c_a^2 = c_s^2 = 4$. In these cases, $E[W(H_2, G_0)] = 2.17$ and $E[W(H_2, G_u)] = 2.03$.

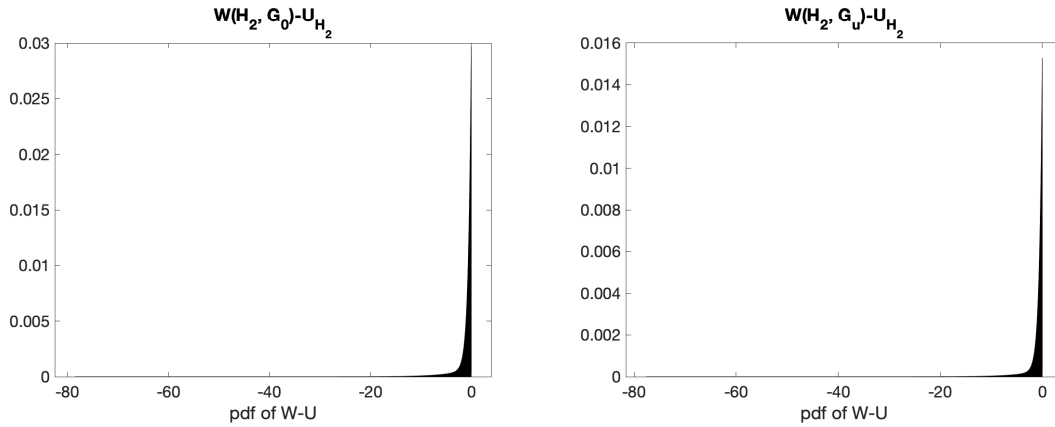


Figure 4 The shape of the pdf of $W(H_2, G_0) - U_{H_2}$ and $W(H_2, G_u) - U_{H_2} \leq 0$ with H_2 interarrival-time cdf, $M_s = 10$, $\rho = 0.5$ and $c_a^2 = c_s^2 = 4$

Figure 5 presents corresponding simulation results when F has an E_2 distribution. We see that monotonicity fails, just as in Figure 3. In these cases, $E[W(E_2, G_0)] = 1.01$ and $E[W(E_2, G_u)] = 1.06$. Notice that there is a switch of order of G_0 and G_u going from H_2 arrivals to E_2 arrivals.

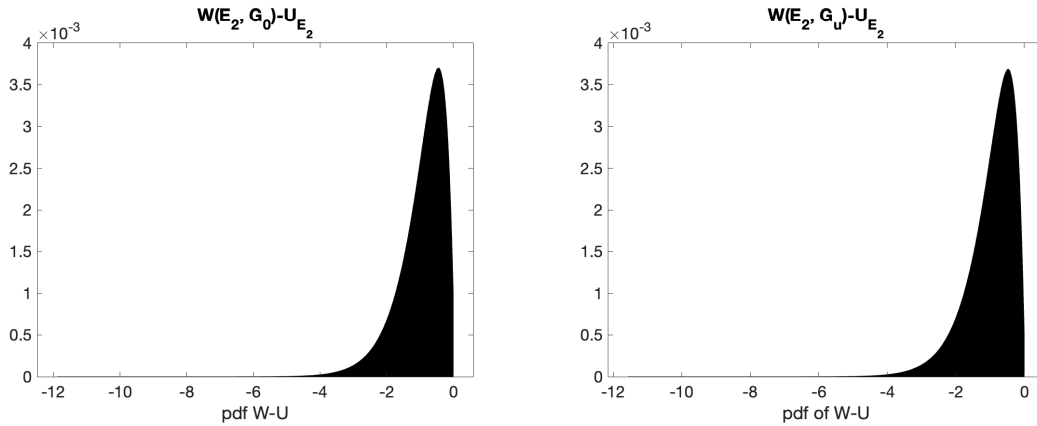


Figure 5 The shape of the pdf of $W(E_2, G_0) - U_{E_2}$ and $W(E_2, G_u) - U_{E_2} \leq 0$ with E_2 interarrival-time cdf, $M_s = 10$, $\rho = 0.5$ and $c_a^2 = 0.5, c_s^2 = 4$

8. Conclusions

This paper applies the theory of Tchebycheff (T) systems to identify interarrival-time and service-time cdf's that maximize or minimize the transient and steady state moments of the waiting time

in the $GI/GI/1$ queue, given the first two moments of the underlying interarrival-time and service-time cdf's. In all cases the extremal cdf's are classical two-point distributions that are determined by either having one mass point at 0 or the upper limit of the support. The proofs rely on the tractable characterization of T systems in §3 in terms of Wronskians, which was previously used for the asymptotic decay rate in [Chen and Whitt \(2019\)](#).

Theorem 1 establishes that higher moments of the steady-state waiting times are maximized (minimized) over all by interarrival-time cdf's F with given first two moments in the $GI/GI/1$ model by the classical extremal two-point cdf F_0 with one mass on 0 (F_u with one mass on the upper limit of support). Corresponding results are also obtained for the service times and for the two cdf's jointly. We prove Theorem 1 in §4 by establishing corresponding results for the transient means $GI_{(n)}/GI/1$ where the interarrival-time cdf can vary with the time period.

Theorem 2 establishes sufficient conditions for corresponding results to hold for the steady state mean $E[W]$. For given service-time cdf G , F_0 yields the upper bound if G is completely monotone. For optimizing over F , a key supporting result was Theorem 4, establishing the concavity of the sojourn-time cdf when the service-time cdf is completely monotone. For given interarrival-time cdf F , G_0 yields the upper bound if the cdf F is strictly concave, as occurs when G has a strictly decreasing pdf. Increasing convex order is used to give a sufficient condition for the overall upper bound to be $E[W(F_0, G_u)]$, as widely conjectured.

Acknowledgments

This research was supported by NSF CMMI 1634133.

References

- Asmussen, S. 1992. Phase-type representations in random walk and queueing problems. *Ann. Probability* **20**(2) 772–789.
- Asmussen, S. 2003. *Applied Probability and Queues*. 2nd ed. Springer, New York.
- Bergmann, R., D. J. Daley, T. Rolski, D. Stoyan. 1979. Bounds for cumulants of waiting times in $GI/GI/1$ queues. *Mathematische Operationsforschung und Statistik. Series Optimization* **10**(2) 257–263.

- Chen, Y., W. Whitt. 2018a. Algorithms for the upper bound mean waiting time in the $GI/GI/1$ extremal queue. Submitted for publication, Columbia University, <http://www.columbia.edu/~ww2040/allpapers.html>.
- Chen, Y., W. Whitt. 2018b. Extremal $GI/GI/1$ queues given two moments. Submitted for publication, Columbia University, <http://www.columbia.edu/~ww2040/allpapers.html>.
- Chen, Y., W. Whitt. 2019. Set-valued queueing approximations given partial information. Submitted for publication, Columbia University, <http://www.columbia.edu/~ww2040/allpapers.html>.
- Cohen, J. W. 1982. *The Single Server Queue*. 2nd ed. North-Holland, Amsterdam.
- Daley, D. J. 1977. Inequalities for moments of tails of random variables, with queueing applications. *Zeitschrift für Wahrscheinlichkeitstheorie Verw. Gebiete* **41** 139–143.
- 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.
- Eckberg, A. E. 1977. Sharp bounds on Laplace-Stieltjes transforms, with applications to various queueing problems. *Mathematics of Operations Research* **2**(2) 135–142.
- 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.
- Holtzman, J. M. 1973. The accuracy of the equivalent random method with renewal inouts. *Bell System Technical Journal* **52**(9) 1673–1679.
- Johnson, M. A., M. R. Taaffe. 1991. An investigation of phase-distribution moment-matching algorithms for use in queueing models. *Queueing Systems* **8**(1-2) 129–148.
- Johnson, M. A., M. R. Taaffe. 1993. Tchebycheff systems for probability analysis. *American Journal of Mathematical and Management Sciences* **13**(1-2) 83–111.
- Karlin, S., W. J. Studden. 1966. *Tchebycheff Systems; With Applications in Analysis and Statistics*, vol. 137. Wiley, New York.
- Kingman, J. F. C. 1961. The single server queue in heavy traffic. *Proc. Camb. Phil. Soc.* **77** 902–904.

-
- Kingman, J. F. C. 1962. Inequalities for the queue $GI/G/1$. *Biometrika* **49**(3/4) 315–324.
- Muller, A., D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. John Wiley and Sons, New York.
- Neuts, M. F. 1981. *Matrix-Geometric Solutions of Stochastic Models*. The Johns Hopkins University Press, Baltimore, MD.
- Ott, T. J. 1987. Simple inequalities for the $D/G/1$ queue. *Operations Research* **35**(4) 589–597.
- Rolski, T. 1972. Some inequalities for $GI/M/n$ queues. *Zast. Mat.* **13**(1) 43–47.
- Stoyan, D. 1983. *Comparison Methods for Queues and Other Stochastic Models*. John Wiley and Sons, New York. Translated and edited from 1977 German Edition by D. J. Daley.
- Szekli, R. 1987. On the complete monotonicity of the waiting time density in some $GI/G/k$ systems. *Queueing Systems* **1** 401–406.
- Szekli, R. 1995. *Stochastic Ordering and Dependence in Applied Probability*. Springer.
- Whitt, W. 1984a. On approximations for queues, I: Extremal distributions. *AT&T Bell Laboratories Technical Journal* **63**(1) 115–137.
- Whitt, W. 1984b. On approximations for queues, III: Mixtures of exponential distributions. *AT&T Bell Laboratories Technical Journal* **63**(1) 163–175.
- Whitt, W. 2002. *Stochastic-Process Limits*. Springer, New York.
- 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.