APPENDIX
to
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

(from the main paper) 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: August 21, 2021

1. Overview

This appendix provides additional material to supplement the main paper. In §2 we review related work. In §3 we make additional remarks about the gradient of the transient mean waiting time, supplementing Section 3 in the main paper. In §4 we study the maximization of the transient mean waiting time $E[W_n(F,G)]$ over the service-time cdf G given the interarrival-time cdf F.

The remainder of the appendix is primarily devoted to numerical examples. First, §5 supplements §4 of the main paper by giving examples applying the linear program in Corollary 1 of the main

paper to determine whether the special two-point cdf F_0 is a stationary point for the known difficult service-time distribution G_0. Second, §6 supplements §5 of the main paper by elaborating on the Frank and Wolfe (1956) (FW) algorithm. Additional examples are given in §6.2. Third, §7 elaborates on §§6-8 of the main paper by giving more examples applying the structural Theorem 3 (a) in the main paper to determine if the special two-point interarrival-time cdf F_0 is a stationary point of the optimization for various service-time cdf's G.

2. Related Work

The use of optimization to study the bounding problem for queues seems to have begun with Klincewicz and Whitt (1984) and Johnson and Taaffe (1990). Due to intractability(e.g., lack of convexity), new approaches have been proposed to simplify the problem, e.g, reformulating the problem into tractable relaxed convex programs, imposing extra conditions and limitations; see Bertsimas and Natarajan (2007) and Gupta and Osogami (2011)). Optimal solutions are not difficult to obtain, but it is difficult to assess the approximation error.

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

3. Complements and Remarks on the Gradient

In this section we make several remarks supplementing Section 3 of the main paper, which derives the gradient of the transient mean $E[W_n(F,G)]$ with respect to the interarrival-time cdf F using the Frechet differentiation, under the assumption that F has finite support.

3.1. Extending the Frechet Differentiation

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

3.2. Stronger Conclusions about Optimality from the Hessian

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

Recall the notation introduced in the main paper at the end of §2 and the beginning of §3.

PROPOSITION 1. (*sufficient condition for local and global optimality*) *Consider the Hessian matrix H from Theorem 1 of the main paper for $GI/GI/1$ queue with the specified G with finite second moment.*

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

(b) If $-H$ is positive semi-definite matrix for some specific $F(\hat{P}) \in \mathcal{P}(\mathcal{F})$ and the $\hat{p}$ satisfies (19) in Proposition 1 of the main paper, then the $\hat{p}$ will be a local optimal distribution.

3.3. Extending the Class of Counterexamples

We next show that a variant of the counterexample to F_0 being optimal for all G, based on the special service-time cdf G_0 in §5 holds for service-time cdf's with a positive pdf, as assumed in Lemma 1 in §6.2 of the main paper. We first observe that The objective function $\phi_a(u)$ in (20)-(22) of the main paper is uniformly bounded and continuous as a function of candidate G, $\hat{p}$ and u. Let $G_n \Rightarrow G$ denote convergence in distribution for a sequence of cdf's as $n \to \infty$.

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

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

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

3.4. Relaxing the Finite Support Condition

We now show how to relax the finite support condition used in Sections 3-5 of the main paper. For that purpose, we will consider a sequence of nested finite support sets. We say that a sequence of finite support sets $\{\mathcal{F}_k : k \geq 1\}$ is nested if $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ for all $n \geq 1$. We say that $\mathcal{F}_k \to [0, M]$ as $k \to \infty$ for a nested sequence of support sets if each $x \in [0, M]$ can be expressed as

$$x = \lim_{k \to \infty} \{x_k : x_k \in \mathcal{F}_k\}. \tag{1}$$

We have the following well known approximation lemma.

LEMMA 1. (*approximation lemma*) *If $\mathcal{F}_k \to [0, M]$ as $k \to \infty$ for a nested sequence of support sets, then Any cdf $F \in \mathcal{P}(1, c_a^2, M)$ can be expressed as the limit of cdf's $F_k \in \mathcal{P}(\mathcal{F}_k)$.*

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

Let $\mathcal{P}_3(1, c_a^2, M)$ be the set of probability distributions with support on three points within $[0, M]$ having specified first two moments. Let $\mathcal{P}_3(\mathcal{F})$ be the subset with support in $\mathcal{F}$.

LEMMA 2. (*extremal cdf for support $[0, M]$*) *Assume that $\mathcal{F}_k \to [0, M]$ as $k \to \infty$ for a nested sequence of support sets. If $F_k^* \in \mathcal{P}_3(\mathcal{F}_k)$ is the optimal cdf for support set $\mathcal{F}_k$, then there exists a convergent subsequence of $\{F_k^* : k \geq 1\}$ with limiting cdf $F^* \in \mathcal{P}_3(1, c_a^2, M)$ and the cdf F^* is an optimal cdf in $\mathcal{P}_3(1, c_a^2, M)$.*

Proof. The key fact is that $\mathcal{P}_3(1, c_a^2, M)$ is a compact subset of $\mathcal{P}(1, c_a^2, M)$. That implies the existence of the convergent subsequence with a limit in the same space. Then the continuity implies the extremal property in the limit. ∎

4. Maximizing Over G for Fixed F

In the main paper we focused on maximizing $E[W_n(F,G)]$ over F for given G. It is evident that we can obtain comparable results when we maximize over G for given F. First, we observe that an analog of Lemma 1 in §6.2 of the main paper arises if we consider the dual problem of optimizing over the cdf G given fixed F, assuming that we impose corresponding regularity conditions. This holds even though the optimization problem in (21) and (22) of the main paper and its gradient vectors are changed if inter-arrival time distribution F is given. But we can exploit a reverse-time representation for the service time G to yield the same structure. For that purpose, let

$$\phi_s(\tilde{v}) \equiv \sum_{i=1}^{n} E[(\sum_{k=1}^{i-1} V_k(\hat{q}) - \sum_{k=1}^{i} U_k(F) + \rho M_s - \tilde{v})^+]. \tag{2}$$

LEMMA 3. (*structure of the objective function in* (2)) *If the fixed cdf F of U has a positive pdf f over $[0,\infty)$, then $Z_i \equiv \sum_{k=1}^{i-1} V_k(\hat{q}) - \sum_{k=1}^{i} U_k(F) + \rho M_s$ has support in $(-\infty, \rho M_s + (i-1)a]$, where $a > 0$ is the upper limit of the support of V. Thus Z_i has a positive pdf θ_i over $(-\infty, \rho M_s]$ for each i, $1 \le i \le m$. Hence,*

$$\begin{aligned} \phi_s(\tilde{v}) &= \sum_{i=1}^{n} E[(Z_i - \tilde{v})^+] = \sum_{i=1}^{n} \left(E[Z_i - \tilde{v}] - E[(Z_i - \tilde{v})^-] \right), \\ &= \sum_{i=1}^{n} \left(E[Z_i - \tilde{v}] - \int_{-\infty}^{\rho M_s} (x - \tilde{v})^- \, d\Theta_i(x) \right), \end{aligned} \tag{3}$$

where

$$\Theta_i(x) = \int_{-\infty}^{x} \theta_i(y)\, dy \quad \textit{for} \quad x \le \rho M_s, \tag{4}$$

so that, paralleling Lemma 1 of the main paper, the first two derivatives of $\phi_s(\tilde{v})$ are

$$\dot{\phi}_s(\tilde{v}) = \sum_{i=1}^{n} (\Theta_i(\tilde{v}) - 1)) < 0 \quad \textit{and} \quad \ddot{\phi}_s(\tilde{v}) = \sum_{i=1}^{n} \theta_i(\tilde{v}) > 0, \quad \tilde{v} \in [0, \rho M_s]. \tag{5}$$

Thus $\phi_s(\tilde{v})$ in (2) *is a continuous, strictly positive, strictly decreasing and strictly convex function on $[0, \rho M_s]$.*

Lemma 3 is proved just like Lemma 1 of the main paper. With Lemma 3, the rest of the proof for optimization over G for fixed F can use the same detailed argument used for optimization over F for fixed G However, we must recall the change of variables made in (2). For example, 0 appears in the extremal cdf for F if and only if ρM_s appears in the extremal cdf for G.

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

5. Applying the LP in Corollary 1 of the Main Paper

We now supplement §4 of the main paper. Corollary 1 in §4 of the main paper provides a a simple verification algorithm. For any given G, we can verify if a candidate F is a stationary point of the optimization. When we solve the LP in Corollary 1 for the specified pair (F, G), if the optimal solution is F, then we have confirmed that F is a stationary point of the optimization. Otherwise, F will not be a stationary point. The verification algorithm is convenient because we already have ideas about the extremal cdf F. In particular, we think that it is often the special two-point cdf F_0.

We now give two numerical examples applying the LP in Corollary 1 of the main paper.

5.1. Determining if F_0 is Extremal Given G_0

For the steady-state mean, the cdf F_0 provides the tight upper bound for $E[W(F, G)]$ for many G, but that is not correct for G_0, as shown in §8 of Wolff and Wang (2003). Hence, we now apply Corollary 1 of the main paper to study the special two-point interarrival-time distribution F_0, for the case $G \equiv G_0$. We consider two cases, one designed to approximately represent steady state and one to be genuinely transient. The nearly-steady-state example has $n = 40, \rho = 0.1, c_a^2 = c_s^2 = 0.5, M_a = 10$. The support contains 401 points in $[0, 10]$ (including the endpoints) so that, F_0 is in the support, while the transient example has $n = 4, \rho = 0.7, c_a^2 = c_s^2 = 0.5, M_a = 10$.

In both cases we apply simulation to estimate the objective function in (20) of the main paper when $G = G_0$ and $F = F_0$ and then solve the linear program in (21) of the main paper. 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. Figure 1 shows the estimates of the objective function $\phi_a(u)$ for the two experiments with (F_0, G_0).

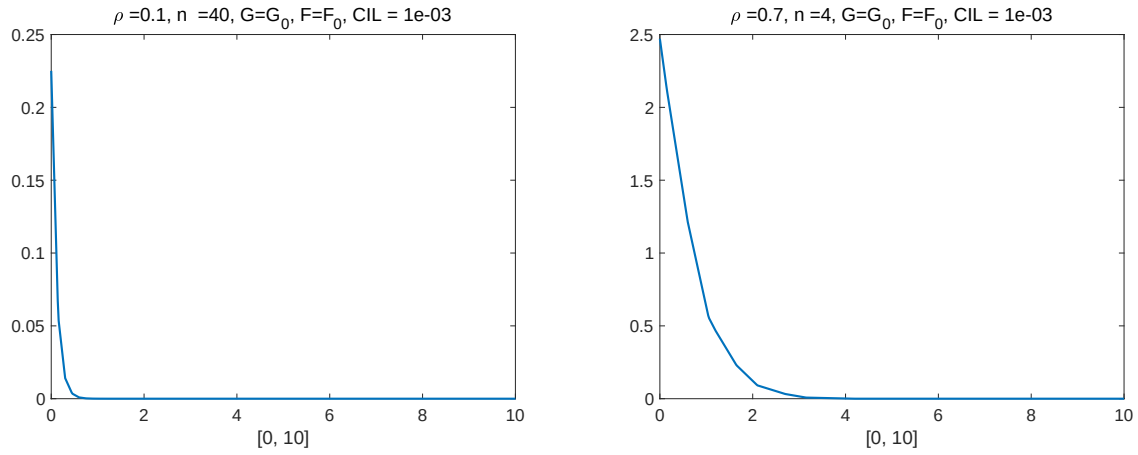

Figure 1 **Simulation estimates of the objective function $\phi_a(u)$ in (20) of the main paper over $[0,10]$ for the $F_0/G_0/1$ model with $n = 40, \rho = 0.1, c_a^2 = c_s^2 = 0.5, M_a = 10$ (left) and with $n = 4, \rho = 0.7, c_a^2 = c_s^2 = 0.5, M_a = 10$ (right), based on 5 replications of 10^6 arrivals.**

When we carry out this simulation+optimization program for the other service-time distributions considered in the examples of §8 of the main paper, we find that F_0 is always a stationary point. However, for G_0, for the example with $n = 4$, we find that F_0 is not the solution of the linear program. In particular, the solution F^* of the linear program has masses $0.3423, 0.3242, 0.3333$ on $0.020, 1.500, 1.520$, respectively. Hence, F_0 is not a stationary point. As a consequence, F_0 is not locally optimal, and thus not optimal. On the other hand, for the nearly-steady-state example with $n = 40$ and $\rho = 0.5$, we find that F_0 is a stationary point.

We also considered our associated numerical study over two-point distributions in Chen and Whitt (2021b). Tables 2 and 3 there display the mean waiting times $E[W]$ and $E[W_{20}]$ for two-point distributions F and G. These tables confirm the counterexample in §8 of Wolff and Wang (2003) for the case $c_a^2 = c_s^2 = 4.0, \rho = 5$ for $n = 20$ and steady-state. We first applied Corollary 1 of the main paper to (F_0, G_0). With a spacing of 0.25 between points in the support of F, we found that

the optimal solution of the linear program had masses of $0.8767, 0.0833, 0.0400$ on $0.25, 6.25, 6.50$, respectively. Hence, F_0 is not a stationary point. Moreover, starting from the optimal solution F^{2*} among the two-point distributions shown in Table 3, which has one mass on 5.25, we find that it too is not a stationary point. We found that the optimal solution of that linear program had masses of $0.4525, 0.3810, 0.1645$ on $0.00, 0.25, 5.50$, respectively. Thus, we conclude that neither F_0 nor the optimal two-point cdf F^{2*} is a stationary point, and thus neither is optimal overall.

5.2. Determining if G_u is Extremal Given F_0

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

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

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

6. More on the Frank and Wolfe (1956) Algorithm

We now elaborate on the Frank and Wolfe (1956) algorithm in §5 of the main paper. A basic reference is §3.2 of Bertsekas (2016). The algorithm is now widely used to solve convex and non-convex programs with convex constraints; see Lacoste-Julien (2016), Reddi et al. (2016) and references there.

6.1. A General FW Algorithm

We start by presenting a general FW algorithm, following Lacoste-Julien (2016) who establishes convergence results. As in the main paper, we denote the cdf F either directly or by its pmf p.

Recall from the first-order optimality condition in Corollary 1 of the main paper, the objective function given $\hat{p}$ can be expressed as

$$\phi_a(u) \equiv \sum_{i=1}^{n} E[(\sum_{k=1}^{i} V_{k-1}(G) - \sum_{k=1}^{i-1} U_{k-1}(\hat{p}) - u)^+], \quad u \in \mathcal{F}. \tag{6}$$

(Of course, $\phi_a(u)$ is also a function of G and $\hat{p}$.) Given the first-order expansion, we can establish the first order approximation for $E[W_n(p, G)]$:

$$E[W_n(p, G)] \approx E[W_n(\hat{p}, G)] + \phi_a(u)^t(p - \hat{p}).$$

Paralleling the setting of Frank-Wolfe algorithm 1 in Lacoste-Julien (2016), we define the *curvature constant* C_f of the continuous differentiable function $E[W_n(p, G)]$ of p, which thus maps R^m into R, as follows,

$$C_f \equiv \sup_{\gamma \in [0,1]} \frac{2}{\gamma^2} \Big(E[W_n(p, G)] - E[W_n(\hat{p}, G)] - \phi_a(u)^t(p - \hat{p}) \Big) \tag{7}$$

$$\text{subject to} \quad \hat{p}, p_2 \in \mathcal{P}(\mathcal{F}), p = \hat{p} + \gamma(p_2 - \hat{p}). \tag{8}$$

We now state the outline of the algorithm in our context. As in the main paper, let $E_F[\cdot]$ denote the expectation with respect to candidate the cdf F of U.

Algorithm 1: Frank-Wolfe algorithm

Initialization: A pmf p_1 in the feasible region $\mathcal{P}(\mathcal{F})$.

Input: Step size ε_j for each step $j = 1, 2, \ldots$ and a stopping threshold $\delta > 0$

Procedure: For each iteration $j = 1, 2, \ldots$, given a pmf p_j:

1 Compute

$$\phi_a(u) = \sum_{i=1}^{n} E[(\sum_{k=1}^{i} V_{k-1}(G) - \sum_{k=1}^{i-1} U_{k-1}(\hat{p})) - u)^+], \quad u \in \mathcal{F} \tag{9}$$

2 Solve $Q_j = \arg\max_{p \in \mathcal{P}(\mathcal{F})} E_p[\phi_a(U)] \equiv \arg\max_{p \in \mathcal{P}(\mathcal{F})} [\nabla w_n(\hat{p})^t \cdot p]$.

3 Update $F_{j+1} = (1 - \varepsilon_j) F_j + \varepsilon_j Q_j$ with two options for the choice of ε_j: (a)

$\varepsilon_j \in \text{argmax}_{\varepsilon \in [0,1]} E[W_n(F_{j+1}, G)]$ or (b) $\varepsilon_j = \min\{\bar{g}_j / C_f, 1\}$ for some constant $C_f > 0$,

where $\bar{g} \equiv E_{Q_j}[\phi_a(U)] - E_{F_j}[\phi_a(U;)]$ is the FW gap.

Repeat above procedure until $\bar{g}_j \leq \delta$.

In the setting of the algorithm above, the assumption of bounded C_f corresponds to the assumption on the gradient of $E[W_n(F, G)]$ being Lipschitz. We now show that condition holds for our queueing problem. We do that by showing that the Hessian is bounded.

Given that the domain $[0, M_a]$ has bounded support and $\mathcal{P}(\mathcal{F})$ is compact, the function $\phi_a(u)$ in (9) is a continuous function with respect to F over a compact domain. Therefore, $\phi_a(u) \equiv \phi_a(u, F)$ is uniformly bounded for all $F \in \mathcal{P}(\mathcal{F})$ and $u \in \mathcal{F}$. Since each term in $\phi_a(u; F)$ is non-negative, it is also uniformly bounded.

A similar argument applies to the Hessian. Recall that the Hessian matrix in this problem is

$$H(u_i, u_j) \equiv \sum_{i=1}^{n} (n-1) E[(\sum_{k=1}^{i} V_{k-1}(G) - \sum_{k=1}^{i-2} U_{k-1}(p(F)) - u_i - u_j)^+],$$

it is thus also uniformly bounded for all $u_i, u_j \in \mathcal{F}$ and all $F \in \mathcal{P}(\mathcal{F})$ such that

$$H(u_i, u_j) \equiv \sum_{i=1}^{n} (n-1) E[(\sum_{k=1}^{i} V_{k-1}(G) - \sum_{k=1}^{i-2} U_{k-1}(p(F)) - u_i - u_j)^+] < n(n-1) \max_{i,j} C_{i,j} < \infty.$$

This is equivalent to say the Hessian matrix is bounded $\|H\| \leq L$ for some constant L. Finally, for all $F(p_1), F(p_2) \in \mathcal{P}(\mathcal{F})$, we have $\|\phi_a(u; p_1) - \phi_a(u; p_2)\| \leq L \|p_1 - p_2\|$ for a uniform constant L with respect to the norm of Hessian matrix.

Given the Lipschitz condition of ϕ_a, the C_f in (7) is well defined, we thus apply the Theorem 1 in Lacoste-Julien (2016) to conclude the algorithm will converge to a stationary point, i.e., $p_j \to p^*$ as $j \to \infty$ and it will take $O(1/\delta^2)$ iterations to find an approximate stationary point with gap smaller than δ.

6.2. A Simple Practical FW Algorithm

In the simulation experiments, because there is no simple closed form expression of (9), we exploit the stochastic Frank-Wolfe algorithm. That means we use simulation to estimate the objective function (9).

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 of the main paper.) The first step is to use Monte-Carlo simulation to estimate the objective value via

$$\phi_a(u; F_j) \equiv \sum_{i=1}^{n} E[(\sum_{k=1}^{i} V_{k-1}(G) - \sum_{k=1}^{i-1} U_{k-1}(F_j) - u)^+] \tag{10}$$

$$\approx \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{n} (\sum_{k=1}^{i} V_{k-1}^{(b)}(G) - \sum_{k=1}^{i-1} U_{k-1}^{(b)}(F_j) - u)^+, u \in \mathcal{F}. \tag{11}$$

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.

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

Algorithm 2: 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 (10) by

$$\hat{\phi}_a(u; F_j) \equiv \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{n} (\sum_{k=1}^{i} V_{k-1}^{(b)}(G) - \sum_{k=1}^{i-1} U_{k-1}^{b}(F_j) - u)^+, u \in \mathcal{F} \tag{12}$$

2 Apply the LP in Corollary 1 of the main paper 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

$$\bar{g}_j \equiv E_{Q_j}[\hat{\phi}_a(U; F_j)] - E_{F_j}[\hat{\phi}_a(U; F_j)] \tag{13}$$

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.) Hence, our numerical studies support the conjecture that, for maximizing $E[W_n(F, G)]$ with respect to F in $\mathcal{P}(\mathcal{F})$ for $2 \leq n < \infty$ and fixed G, the stationary point is unique and so is the global optimal solution. 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 (2) 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 §7.3.

For the case (b), the stationary point for $M/GI/1$ under $E[W_n(M, G)]$ with $n < \infty$ is evidently unique, 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 (2) is numerically solved via the Gurobi solver in CVX.

6.2.1. Transient Mean Waiting Time We first consider the transient mean $E[W_4]$ for the classical four models $GI/M/1$, $GI/E_{10}/1$, $GI/G_u/1$ and $GI/G_0/1$.

The direct finding is that the algorithm will converge to F_0 within two steps for the three models $GI/M/1$, $GI/E_{10}/1$, $GI/G_u/1$ and two steps $Q_k = F_0$ for all k. For $GI/G_0/1$, the story is different. The sequence will converge to a specific two-point distributions within finite steps and Q_k is not changed after four steps. When we chance initial distribution into M and other distributions, the above conclusions are not changed.

The Q_k has some changes in the initial steps and keep the same for $GI/G_0/1$ after four iterations. We present numerical values for the sequence of Q_k. In this experiment we started with a support of $m = 401$ points evenly distributed in $[0, 10]$, but then refined thesupport to obtain a better approximation of the extremal distribution over all of $[0, 10]$. In that way we obtain a two point distribution was mass on 1.5556 instead of a three-point distribution (as shown in the main paper).

Table 1 **The sequence of optimal distribution Q_k for $GI/G_0/1$ during each iterations when 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	0	1.5	0
2	0.3795	0.1538	0.4667	0.1	1.4	1.6
3	0.4190	0.581	0	0.15	1.6	0
4	0.3816	0.6184	0	0.1000	1.5556	0
5	0.3816	0.6184	0	0.1000	1.5556	0

Table 2 **The sequence of optimal distribution Q_k for $GI/G_0/1$ during each iterations when initial distribution is $F = M$**

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.3333	0.6667	0	0	1.5	0
2	0.3795	0.1538	0.4667	0.1	1.4	1.6
3	0.4190	0.581	0	0.15	1.6	0
4	0.3816	0.6184	0	0.1000	1.5556	0
5	0.3816	0.6184	0	0.1000	1.5556	0

After 5 steps, the Q_k has not changed anymore. The optimization step in the algorithm (2) is numerically solved via Gurobi solver in CVX, in the final output we usually obtain a three-point distribution with two adjacent masses when the theoretical optimal solution is a two-point distribution. For example, in the above case, we obtain the final resulting solution provided by the solver is $\{0.3816, 0.4828, 0.1356\}$ on $\{0.1, 1.550, 1.575\}$ where errors result from the finite uniform discretization and numerical solver. In fact, it is a two-point distribution. According to the two-point distribution closed-form solution, we can determine the b via $1 - c_a^2/(b-1) = 0.1$ to obtain $b = 1.5556$ such that the resulting two-point distribution Q_∞ has mass on $\{0.1, 1.5556\}$.

We also look into the behaviors of the sequence $\{F_k\}$. After 20 iterations, the F_{20} has probability $\{0.043, 0.3717, 0.049, 0.0024, 0.2812, 0.3354\}$ on support $\{0, 0.1, 0.2, 1.4, 1.5, 1.6\}$. We have seen consistently decreasing masses on $0.0, 0.2, 1.4$ as $k \to \infty$. Thus the F_k will finally converge to Q_∞ as $k \to \infty$.

6.2.2. Steady-State Mean Waiting Time We repeat the above experiments for the approximate steady-state mean waiting time $E[W_{40}(F, G)]$ under such four models. The story is not changed for the three models with $F = F_0$ as stationary points. For $F/G_0/1$, we obtain F_0 being approximate stationary point under $\rho = 0.1$.

Table 3 **The sequence of optimal distribution Q_k for $GI/G_0/1$ for $E[W_{40}(F, G)]$ during each iterations when initial distribution is $F = M$ with $\rho = 0.1$**

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.3333	0.6667	0	0	1.5	0
2	0.3333	0.6667	0	0	1.5	0
3	0.3333	0.6667	0	0	1.5	0
4	0.3333	0.6667	0	0	1.5	0
5	0.3333	0.6667	0	0	1.5	0

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 $\{0, 1.375, 1.6\}$.

Table 4 **The sequence of optimal distribution Q_k 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	0	1.5	0
2	0.4170	0.1413	0.4417	0	0.1750	1.325
3	0.3326	0.3448	0.3226	0	1.45	1.55
4	0.3317	0.3509	0.3175	0	1.4250	1.575
5	0.3304	0.3571	0.3125	0	1.4	1.6
6	0.3304	0.3571	0.3125	0	1.4	1.6
7	0.3287	0.3636	0.3077	0	1.375	1.625
8	0.3287	0.3636	0.3077	0	1.375	1.625
9	0.3295	0.3232	0.3472	0	1.375	1.6
10	0.3295	0.3232	0.3472	0	1.375	1.6

Therefore, we obtain different stationary points under different ρ. F_0 is not always the stationary point for $F/G_0/1$ for all ρ. It is highly possible the three-point distribution will be the extremal distribution for $E[W_\infty(F,G_0)]$.

6.3. Numerical Examples for Case (b)

We set up the same settings as experiments in the case (a) for transient mean waiting time and approximate steady-state mean waiting time. The only difference is the range of b where $b\in[0,M_s]$ where $M_s=10/\rho$ due to mean one inter-arrival time.

6.3.1. Transient Mean Waiting Time We then consider the transient mean $E[W_4]$ for $M/GI/1$, $E_{10}/GI/1$, $F_0/GI/1$ and $F_u/GI/1$. The stationary points for $M/GI/1$ and $E_{10}/GI/1$ is G_u. For other models, the stationary points are specific two-point distributions with interior masses.

Table 5 The sequence of optimal distribution Q_k for $F_u/GI/1$ during each iterations when initial distribution is $G = M$

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.9592	0.0408	0	0.422	2.2	0
2	0.9792	0.0208	0	0.45	3	0
3	0.9792	0.0208	0	0.45	3	0
4	0.9792	0.0208	0	0.45	3	0
5	0.9792	0.0208	0	0.45	3	0

Table 6 The sequence of optimal distribution Q_k for $F_0/GI/1$ during each iterations when initial distribution is $G = M$

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.9822	0.0178	0	0.452	3.1	0
2	0.9774	0.0226	0	0.446	2.8	0
3	0.9774	0.0226	0	0.446	2.8	0
4	0.9774	0.0226	0	0.446	2.8	0
5	0.9774	0.0226	0	0.446	2.8	0

6.3.2. Transient Mean Waiting Time The approximate stationary points for the four models are F_u.

Table 7 The sequence of optimal distribution Q_k for $F_0/GI/1$ during each iterations when initial distribution is $G = M$

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.9939	0.0001	0	0.1	10	0
2	0.9939	0.0001	0	0.1	10	0
3	0.9939	0.0001	0	0.1	10	0
4	0.9939	0.0001	0	0.1	10	0
5	0.9939	0.0001	0	0.1	10	0

Table 8 **The sequence of optimal distribution Q_k for $F_u/GI/1$ during each iterations when initial distribution is $G = M$**

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.9939	0.0001	0	0.1	10	0
2	0.9939	0.0001	0	0.1	10	0
3	0.9939	0.0001	0	0.1	10	0
4	0.9939	0.0001	0	0.1	10	0
5	0.9939	0.0001	0	0.1	10	0

6.4. Numerical Examples for Case (c)

We combine the settings of the case (a) and the case (b).

6.4.1. Transient Mean Waiting Time We set up the same initial setting of experiments as before. We set $G = M$ to do maximization over F with initial $F = M$, we found the $F = F_0$ is the approximate stationary point. Then we given $F = F_0$ to do maximization over G under initial distribution $G = M$ to conclude $G_{u,n}$ in our setting is the approximate stationary point. (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 be specific, the $G_{u,n}$ has masses $\{0.9773, 0.0227\}$ on support $\{0.4461, 2.8199\}$. Finally given the $G_{u,n}$ we still are able to check the F_0 is also the stationary point. Therefore, the pair $(F_0, G_{u,n})$ is the fixed point solution for the LP program in Corollary 1 of the main paper for (a) and the analog ofr case (b) respectively, thus they are the approximate stationary points for the case (c).

We repeat the above experiment with $G = G_0$, we first optimize over F with initial $F = M$, the convergent solution is a two-point distribution $F_{u,n}$ which has masses on $\{0.381, 0.619\}$ on $\{0.099, 1.555\}$. Then we given $F = F_{u,n}$ to optimize over G with initial $G = M$, the approximate stationary point is the $G_{u,n}$ with $\{0.9808, 0.0192\}$ on support$\{0.45, 3\}$.

Table 9 **The sequence of optimal distribution Q_k for $F_{u,n}/GI/1$ during each iterations when initial distribution is $G = M$**

Iterations	p_1	p_2	p_3	x_1	x_2	x_3
1	0.9864	0.0154	0	0.4554	3.3	0
2	0.9808	0.0192	0	0.45	3	0
3	0.9808	0.0192	0	0.45	3	0
4	0.9808	0.0192	0	0.45	3	0
5	0.9808	0.0192	0	0.45	3	0

Given such above G, we continue to optimize over F with initial $F = M$, we obtain $F = F_0$ and finally obtain the same stationary pair $(F_0, G_{u,n})$ with $G_{u,n} = \{0.3810, 0.6190\}$ on support $\{0.099, 1.555\}$.

Therefore, we have confirmed that $(F_0, G_{u,n})$ is the stationary point pair for the case (c) maximization problem. Next, we study the performance of $G_{u,n}$ when increasing n.

Table 10 **The sequence of Q_{20} as a function of n for $E[W_n(F_0, G)]$ when initial distribution is $G = M$**

n	p_1	p_2	p_3	x_1	x_2	x_3
4	0.9774	0.0226	0	0.446	2.8	0
10	0.9916	0.0084	0	0.4671	4.3	0
15	0.9958	0.0042	0	0.4769	5.9	0
20	0.9975	0.0025	0	0.4821	7.5	0
40	0.9986	0.0014	0	0.4868	10	0

This table illustrates the $Q_{20} = G_{u,n}$ and the largest point in $G_{u,n}$ increases as n increases. When $n = 40$ such that the $E[W_n]$ approximates the $E[W_\infty]$ well, the $G_{u,n} = G_u$ with the largest point $b = \rho M_s = 10$.

6.4.2. The Steady-State Mean Waiting Time For the steady-state approximation, we consider more cases to make results convincing; We set $\rho = 0.1, 0.3, 0.5$ and $n = 40, 60, 80$ under the

same experiment setting with initial distributions $F, G = M$. We finally obtain consistent conclusion (F_0, G_u) is always the approximate stationary points. Given F_0 being inter-arrival time for $\rho = 0.1$, the G_u is approximate stationary point with mass on $\{0.988, 10\}$ and probability $\{0.9986, 0.0014\}$; Given such G_u the F_0 is the approximate stationary point. Therefore, (F_0, G_u) is the approximate stationary points for the case (c).

7. Applications of Theorem 3 (a) for Some $GI/GI/1$ Examples

This section is an elaboration of §8 of the main paper, where we give numerical examples illustrating the value of the structural theorems in §6 of the main paper. In particular, here we focus on the application of Theorem 3 (a) in the main paper.

In this section we apply simulation to examine if the conditions in Theorem 3 (a) of the main paper for F_0 or F_u to be a stationary point of the optimization are satisfied for various $GI/GI/1$ examples, in the context of Corollary 1 and Lemma 1 in the main paper. 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) of the main paper, which is the summation of pdf functions of Y_i in Lemma 1 of the main paper. 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.

7.1. An Exponential Service Time Distribution

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

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

$$\ddot{\phi}_a(u) = \sum_{i=1}^{n} \gamma_i(u) = \sum_{i=1}^{n} e_i(u + z_k^{(i)}) p_k^{(i)} \tag{14}$$

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

$$\ddot{\phi}_a(u) = e_2(u)\frac{c_a^2}{1+c_a^2} + e_2(u+1+c_a^2)\frac{1}{1+c_a^2} + e_1(u). \tag{15}$$

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

For large n, we can verify the monotonicity property in any instance by applying stochastic simulation. As before, we used Monte-Carlo simulation to create 5 replications of 10^6 random samples in order to estimate the summation of pdf functions $\ddot{\phi}_a$ for $\sum_{i=1}^{n} Y_i$ in $GI/M/1$ with $F = F_0$. To illustrate, Figure 2 shows the simulation estimates of the second derivative $\ddot{\phi}_a(u)$ of the objective function in $\phi_a(u)$ in (32) and (20) of the main paper for $F_0/M/1$ (LHS) and $F_u/M/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.7, n = 4, M_a = 10$. These plots show that $\ddot{\phi}_a(u)$ is monotonically decreasing over $[0, M_a]$ in both cases. Hence, F_0 is the optimal solution in the LP in Corollary 1 in both cases. Thus, we conclude that F_0 is a stationary point of the optimization, whereas F_u is not. These conclusions were confirmed by applying Corollary 1 of the main paper. In particular, F_0 was found to be the solution to the LP in both the nearly-steady-state example with $(\rho = 0.1, n = 40)$ and the transient example with $(\rho = 0.7, n = 4)$.

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

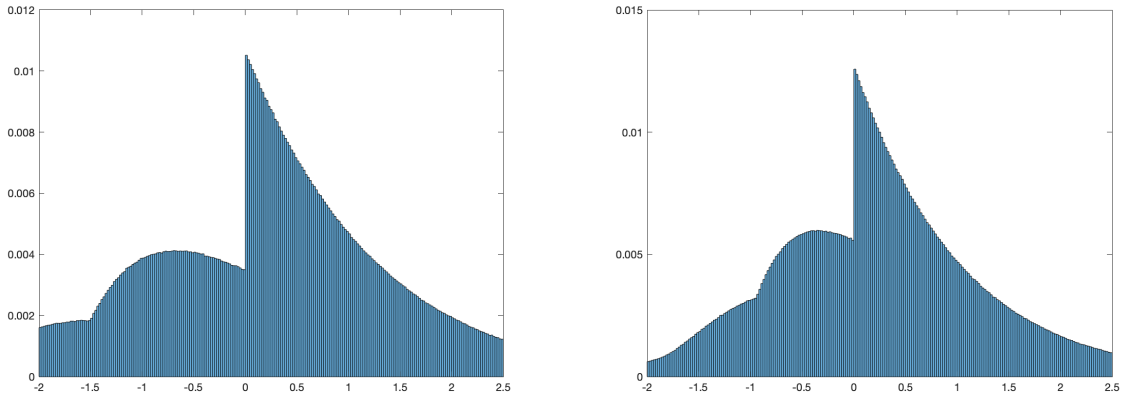

Figure 2 **Simulation estimates of $\ddot{\phi}(u)$ in Lemma 1 of the main paper for $F_0/M/1$ (LHS) and $F_u/M/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.7, n = 4, M_a = 10$. These plots show that F_0 is a solution of the LP in in both cases, so that F_0 is a stationary point, while F_u is not.**

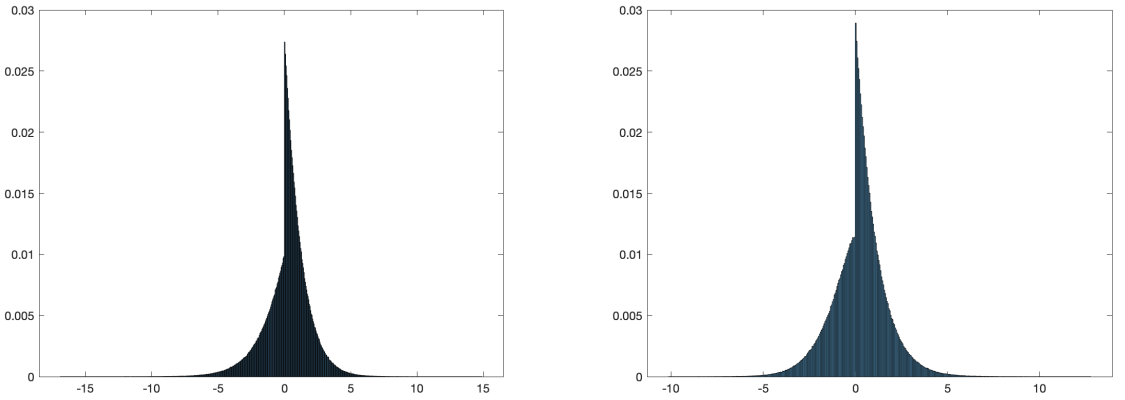

Figure 3 **Simulation estimates of $\ddot{\phi}(u)$ in Lemma 1 of the main paper for $M/M/1$ (LHS) and $E_2/M/1$ (RHS) in the case $\rho = 0.7, n = 4, M_a = 10$. These plots show that F_0 is a solution of the LP in Corollary 1 of the main paper in both cases, so that neither of these interarrival-time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

7.2. Erlang E_2 Service Distributions

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

We now discuss $GI/E_2/1$. Figure 4 displays simulation estimates of $\ddot{\phi}(u)$ in Lemma 1 of the main paper for $F_0/E_2/1$ (LHS) and $F_u/E_2/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$.

In this case we do not see monotonicity, but instead we see the single-peak property over $[0, M_a]$. Thus, these plots also show that F_0 is a solution of the LP in Corollary 1 of the main paper in both cases, because of the single-peak property, so that F_0 is a stationary point, while F_u is not. As in the previous example, these conclusions were confirmed by applying the LP in Corollary 1 of the main paper. As before, F_0 was found to be the solution to the LP in both the nearly-steady-state example with $(\rho = 0.1, n = 40)$ and the transient example with $(\rho = 0.7, n = 4)$.

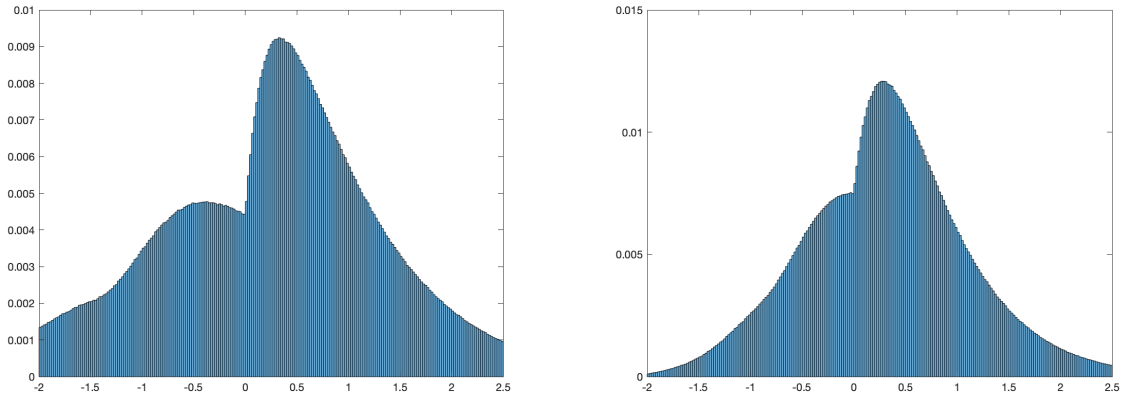

Figure 4 **Simulation estimates of $\ddot{\phi}_a(u)$ in Lemma 1 of the main paper for $F_0/E_2/1$ (LHS) and $F_u/E_2/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$. These plots show that F_0 is a solution of the LP in Corollary 1 of the main paper in both cases because of the single-peak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

7.3. A More Complex Service-Time Distribution

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

Figure 5 presents simulation estimates of $\ddot{\phi}(u)$ in (32) and Lemma 1 of the main paper for $F_0/GI/1$ (LHS) and $F_u/GI/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$, where the service-time distribution is chosen to be $G = mix(E_{20}(0.4), E_{20}(1.6), 0.5)$, which has mean $0.5(0.4) +$

$0.5(1.6) = 1.0$. Figure 5 shows that the condition of Theorem 3 (a) of the main paper is not satisfied in either case.

Unlike the previous three examples, the conclusions from applying Corollary 1 of the main paper are more complicated. As before, F_0 was found to be the solution to the LP in the nearly-steady-state example with $(\rho = 0.1, n = 40)$, but it was not in the transient example with $(\rho = 0.7, n = 4)$. Moreover, the FW algorithm confirmed these conclusions. Starting with M, the stationary point obtained has masses $(0.6667, 0.3333)$ on the points $(0.500, 2.000)$ in the transient example.

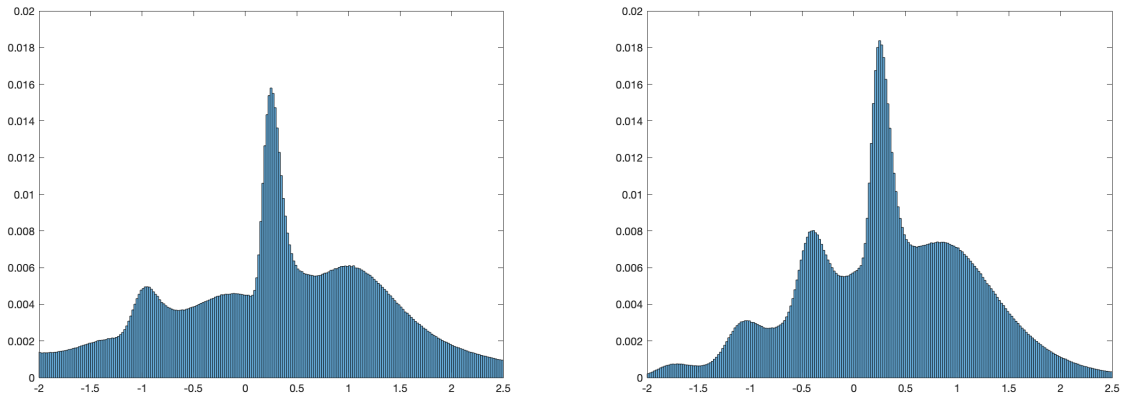

Figure 5 **Simulation estimates of $\ddot{\phi}(u)$ in Lemma 1 of the main paper for $F_0/GI/1$ (LHS) and $F_u/GI/1$ (RHS) in the case $c_a^2 = 0.5, \rho = 0.5, n = 4, M_a = 10$, where the service-time distribution in both cases is a mixture of two Erlang distributions, specifically $mix(E_{20}(0.4),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

7.4. Maximization Over G Given F

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

From Lemma 3, we know that we can apply a reverse-time representation to reduce this problem to the case previously considered. That implies that, in the reverse-time representation, we should look at the shape of $\ddot{\phi}_s(\tilde{v})$ in the range $[0, \rho M_s]$. However, that is equivalent to looking at the corresponding shape of $\ddot{\phi}_s(v)$ over $[-\rho M_s, 0]$ without time-reverse representation. That means if the original shape is strictly monotone increasing over $[-\rho M_s, 0]$, that is equivalent to the time-reverse

shape is strictly monotone increasing over $[0, \rho M_s]$ because 0 under the time-reverse representation corresponds to ρM_s under no time-reverse representation.

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

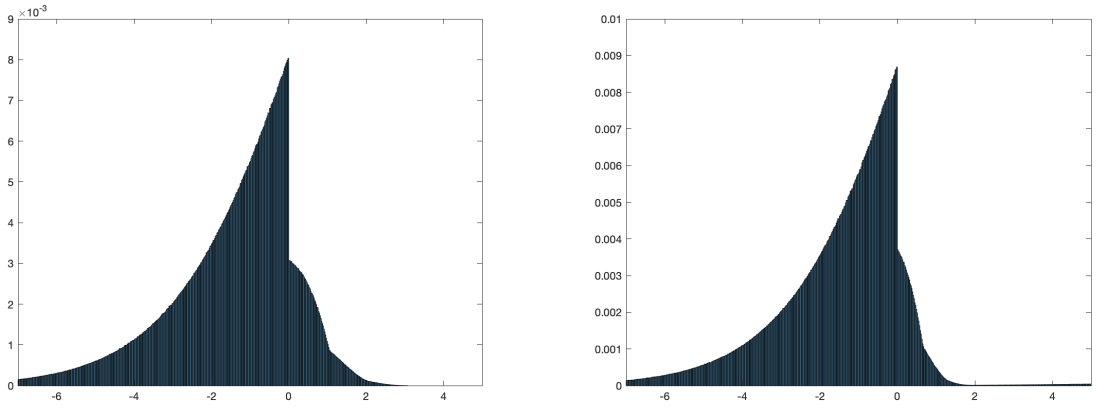

Figure 6 **Simulation estimates of $\ddot{\phi}_s(\tilde{v})$ in (5) associated with Lemma 3 for $M/G_0/1$ (LHS) and for $M/G_u/1$ (RHS) in the case: $c_a^2 = 01.0, c_s^2 = 0.5, \rho = 0.7, n = 4, M_s = 10$**

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

8. The Associated Minimization Problem

We now consider the associated minimization problem, which corresponds to the supremum in (5), (21) and (22) of the main paper being replaced by an infimum, and then the associated maximum in (6), (27), (28) and (29) being replaced by a minimum. Then the inequality must be reversed in the inequality (19) of the main paper which expresses the definition of a stationary point. Accordingly the associated dual problem in (35) of the main paper becomes finding the vector $\lambda^* \equiv (\lambda_0^*, \lambda_1^*, \lambda_2^*)$ that attains the maximum

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

$$\text{such that} \quad r(u_i) \equiv \lambda_0 + \lambda_1 u_i + \lambda_2 u_i^2 \le \psi(u_i) \quad \text{for all} \quad i, \quad 1 \le i \le m. \tag{16}$$

If we simply replace λ_i via $-\lambda_i$ such that the max in (16) can be replaced by min, we found the dual problem in (16) is not equivalent to that in (36) of the main paper because the $-\psi(u)$ is not a strictly monotone decreasing and strictly positive convex function satisfying Lemma 1 of the main paper. As a consequence, we are not able to completely replicate the proof of Theorem 3 of the main paper, thus leading to weaker conclusions for the minimization problem.

PROPOSITION 2. (*sufficient conditions for F_0 or F_u to be a stationary point*) *Under the same initial three assumptions as in Theorem 2 of the main paper, (a) if $\ddot{\psi}(u;F)$ is strictly decreasing over $[0,M]$ for any candidate cdf F, then F_u must be a solution of the LP in (34) of the main paper. Hence, if this condition is satisfied for $F=F_u$, then F_u. must be a stationary point.*

(b) Similarly, if $\ddot{\psi}(u;F)$ is strictly increasing over $[0,M]$ for any candidate cdf F, then F_0 must be a solution of the LP in (34) of the main paper. Hence, if this condition is satisfied for $F=F_0$, then F_0. must be a stationary point.

Proof. Replace λ_i by $-\lambda_i$ and make $\psi(u)$ become $-\psi(u)$ in (16). Thus we shall solve

$$\begin{aligned} &\min\{\lambda_0+\lambda_1+\lambda_2(1+c^2)\} \\ &\text{such that} \quad r(u_i)\equiv\lambda_0+\lambda_1u_i+\lambda_2u_i^2\geq-\psi(u_i) \quad \text{for all} \quad i, \quad 1\leq i\leq m. \end{aligned} \tag{17}$$

It is the dual problem of maximization problem with $-\psi$ as in the objective function. That implies if ψ is strictly monotone decreasing, then $-\psi$ is strictly monotone increase. Paralleling with proof of Theorem 2 in the main paper, we can conclude the opposite conclusions. ∎

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

Proof. Given (17), if ψ has one peak, then the $-\psi$ is not a one peak function. We apply the same argument before (39) in the main paper to show there is at most one interior intersection point. However, due to $-\psi$ not being one peak function, there is no other cases which can be ruled out, thus it is possible to have F_0, F_u as well as F_b being optimizers where F_b has two ending points 0 and M and it also has one interior point $b \in (0, M)$. ∎

Incorporating the result of Proposition 3 and paralleling Theorem 4 of the main paper, we establish the following result for minimization problem.

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

Proof. We apply the same argument before (39) of the main paper in the proof of Theorem 3 to show there is at most n interior intersection point, then following via the proof of Proposition 3, the proof is thus complete. ∎

Acknowledgments

This research was supported by NSF CMMI 1634133.

References

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

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

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

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

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. Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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