Hidden Convexity in Queueing Models

Xin Chen
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Tech, xin.chen@isye.gatech.edu

Linwei Xin
School of Operations Research and Information Engineering, Cornell University, lx267@cornell.edu

Minda Zhao
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Tech, mindazhao@gatech.edu

Abstract. We study the joint control of arrival and service rates in queueing systems with the objective of minimizing long-run expected cost minus revenue. Although the objective function is non-convex, first-order methods have been empirically observed to converge to globally optimal solutions. This paper provides a theoretical foundation for this empirical phenomenon by characterizing the optimization landscape and identifying a hidden convexity: the problem admits a convex reformulation after an appropriate change of variables. Leveraging this hidden convexity, we establish the Polyak-Łojasiewicz-Kurdyka (PŁK) condition for the original control problem, which excludes spurious local minima and supports global convergence guarantees for first-order methods. Our analysis applies to a broad class of $GI/GI/1$ queueing models, including those with Gamma-distributed interarrival and service times, as well as $GI/M/1$ queues with log-concave interarrival times. As a key ingredient in the proof, we establish a new convexity property of the expected queue length under a square-root transformation of the traffic intensity.

Key words: Control of arrival and service rates, Queueing model, Hidden convex, Polyak-Łojasiewicz-Kurdyka (PŁK) condition

1. Introduction

This work studies a queueing model in which a service provider manages the system by jointly controlling the arrival and service rates, a topic that has been well studied in the literature over the past decades (e.g., Mendelson 1985, Maglaras and Zeevi 2003, Lee and Ward 2014, Chen et al. 2024b). We follow the standard setting where the service provider seeks the optimal arrival rate λ^* and service rate μ^* that minimize the long-run expected cost minus revenue in a $GI/GI/1$ queue: $\mathbb{E}[Q_\infty(\lambda,\mu)] + c(\mu) - r(\lambda)$, where $Q_\infty(\lambda,\mu)$ denotes the steady-state queue length under arrival rate λ and service rate μ; $c(\mu)$ is the cost of providing service rate μ; and $r(\lambda)$ is the revenue corresponding to the arrival rate λ.

Solving this joint control problem over arrival and service rates is challenging because the objective function is non-convex (Harel 1990). Existing studies often restrict attention to tractable settings such as $M/M/1$ queues for exact analysis (Ata and Shneorson 2006) or asymptotic heavy-traffic regimes (Lee and Ward 2014, 2019). More recently, Chen et al. (2024b) propose a gradient-based method to numerically solve the problem. Surprisingly, despite the non-convex nature of the underlying problem, both Chen et al. (2024b) and Hu et al. (2024), who control the arrival rate through pricing, observe that first-order methods empirically converge to global optimal solutions. This phenomenon reveals a striking gap between theory and practical performance. Understanding the mechanisms behind the success of first-order methods in these non-convex queueing models is essential for developing more principled online learning and offline optimization algorithms.

In this paper, we provide a theoretical foundation for this empirical phenomenon by identifying a hidden convexity. The key step is a reparameterization of the optimization problem through a change of variables. Under mild structural assumptions, we show that the original non-convex objective can be reformulated as a convex function over the new variables, subject to convex constraints. This convex reformulation applies to a broad class of $GI/GI/1$ queues, including $\Gamma_k/\Gamma_m/1$ queues whose interarrival and service times follow Gamma distributions with shape parameters k and m no less than 1, as well as $GI/M/1$ queues with log-concave interarrival times. It also extends to Jackson networks of $M/M/1$ queues, $M/GI/1$ queues, and well-known approximate formulations of $GI/GI/s$ queues. Finally, leveraging this reformulation, we establish that the original problem satisfies the Polyak-Łojasiewicz-Kurdyka (PŁK) condition (Polyak et al. 1963, Łojasiewicz 1963, Kurdyka 1998), which ensures the global convergence of first-order methods and helps explain the observed empirical success under standard assumptions.

A key ingredient in our analysis is a new convexity result. To construct the reformulation, we require the expected queue length to be convex after a square-root transformation of the traffic intensity ρ. Specifically, let $L_q(\rho)$ denote the expected queue length. We prove that $\hat{L}_q(\tau) := L_q(\sqrt{\tau})$ is convex for $\Gamma_k/\Gamma_m/1$ queues with $k, m \geq 1$ and $GI/M/1$ queues with log-concave interarrival times. Interestingly, using Spitzer's identity, $\hat{L}_q(\tau)$ can be represented as the sum of a finite collection of potentially non-convex terms and infinitely many, rapidly vanishing convex terms. Because of this mixture, the global convexity of $\hat{L}_q(\tau)$ is far from trivial, which may be of independent interest.

1.1. Literature Review

In recent years, learning and optimization in queueing problems have attracted considerable attention (see, for example, Chen and Hong 2023, Chen et al. 2025, Kanoria and Qian 2024, Zhong et al. 2024). One powerful tool for solving these problems is convexity, which has been studied for decades in queueing systems. Much of the literature establishes univariate convexity of performance measures while holding other parameters fixed, e.g., convexity in the service rate μ with fixed arrival rate λ (Weber 1983), and convexity in λ with fixed μ (Fridgeirsdottir and Chiu 2005). Related results also show convexity with respect to the number of servers (Weber 1980). By contrast, joint convexity in arrival and service rates generally fails. Harel (1990) demonstrates that "neither the expected number of customers in the queue nor in the system is jointly convex for the queues $M/M/x$, $M/G/1$."

Despite the loss of joint convexity, the optimization problem remains tractable when the decision-maker's objective is solely to minimize queue performance metrics, such as the expected queue length. Existing work establishes convexity of various performance measures in traffic intensity ρ for $M/M/c$ queues (see details in Harel and Zipkin 1987). The difficulty arises when service cost and revenue are involved, as the convexity fails due to the non-convex constraint $\rho = \lambda/\mu$. In addition, Agrawal and Boyd (2020) apply a log-log convex programming approach to minimize the weighted sum of the service loads in $M/M/c$ queues. However, their formulation does not contain the service cost minus revenue term, which is generally not log-log convex. Chen et al. (2024b) impose assumptions ensuring the joint strong convexity of the long-run expected cost. The assumptions are limited to $M/GI/1$ queues, which allow for a closed-form expression of the expected system length as a function of traffic intensity. Departing significantly from previous approaches, our work is the first to reveal the hidden convexity in the joint control of arrival and service rates for $GI/GI/1$ queues. Under mild conditions, we construct convex reformulations and establish the PŁK condition of the original problem. We contribute to the optimization literature by identifying a class of queueing control problems that satisfy the PŁK condition, thereby complementing recent advances on tractable non-convex optimization problems in operations and control models (Bhandari and Russo 2024, Chen et al. 2024a).

Our use of square-root transformations is conceptually related to, but technically distinct from, the classical line of work in queueing and stochastic networks. In capacity-allocation problems for Jackson networks, square-root allocation formulas have a long history, dating back to Kleinrock's work (Kleinrock 1964) and subsequent developments for product-form and generalized Jackson networks (Wein 1989, Dieker et al. 2017). These formulas typically arise from applying Lagrangian methods to approximately solve capacity-allocation problems. In other models, such as multiclass queueing networks, entropy-type formulations arise in the large-deviation objective (Pittel 1979) and in the proportional-fairness interpretation of such networks (Walton 2009). Our contribution is different from these classical analyses: we use square-root transformations as reparameterizations of the original decision variables to reveal hidden convexity and establish a PŁK landscape for the underlying nonconvex queueing-control problem.

Our paper is also related to the rich literature on joint pricing and capacity sizing in queueing systems, where both the arrival and service rates are controlled via a transformation between price and arrival rate. This transformation is commonly used in the revenue management literature (Talluri and Van Ryzin 2005) to address the non-concavity of revenue with respect to price. Several existing studies focus on static pricing control, in which the price is independent of the queue length. Since a closed-form expression for the expected queue length is generally unavailable in $GI/GI/1$ queues, some studies adopt heavy-traffic approximations to derive asymptotically optimal arrival and service rates (Lee and Ward 2014, 2019). Differently, Chen et al. (2023) and Chen et al. (2024b) apply online first-order methods to solve the problem. In the dynamic setting, Ata and Shneorson (2006) investigate the optimal dynamic pricing and service-rate control policy in $M/M/1$ queues, while Kim and Randhawa (2018) use an asymptotic large-system approach.

2. Problem Setting

Consider a single-server $GI/GI/1$ queueing system with customer arrivals according to a stochastic process. The interarrival times $\{T_n\}_{n=1}^{\infty}$ are independent and identically distributed (i.i.d.) with a general distribution, corresponding to the first “GI." Similarly, the service times $\{S_n\}_{n=1}^{\infty}$ are also i.i.d. with a general distribution, representing the second “GI." We model the queueing system by allowing control over both the arrival rate $\lambda > 0$ and service rate $\mu > 0$, such that the interarrival and service times are given by T_n/λ and S_n/μ, respectively; namely, each interarrival (service) time is expressed as the product of the inverse deterministic arrival (service) rate and a random variable. This decomposition is a standard approach in the queueing literature on rate control (e.g., Lee and Ward 2014, Whitt 2015, 2018, Chen et al. 2024b). A natural interpretation is that S_n may represent the size of a message to be transmitted in a communication network, while μ corresponds to the processing rate of messages. Without loss of generality, we assume $\mathbb{E}[T_n] = \mathbb{E}[S_n] = 1$, which can be achieved by appropriately rescaling the arrival and service rates.

By controlling the arrival and service rates, the system receives revenue $r(\lambda)$ and incurs a non-decreasing capacity cost $c(\mu)$ per time unit. The decision-maker searches for the optimal arrival rate λ^* and service rate μ^* to minimize the long-run expected cost over a bounded domain:

$$\begin{aligned} \min_{\lambda,\mu} \quad & \mathbb{E}\left[Q_\infty(\lambda,\mu)\right] + c(\mu) - r(\lambda) \\ \text{s.t.} \quad & \underline{\lambda} \le \lambda \le \bar{\lambda}, \quad \underline{\mu} \le \mu \le \bar{\mu}, \end{aligned} \tag{1}$$

with $\underline{\lambda} > 0$. To ensure stability of the queueing system and to state the landscape results in a differentiable form, we impose the following standing assumptions.

Assumption 1. *All of the following conditions hold:*

1. *The revenue $r(\lambda)$ and service cost $c(\mu)$ are continuously differentiable.*
2. *The bounds $\bar{\lambda}$ and $\underline{\mu}$ satisfy $\bar{\lambda} < \underline{\mu}$, implying that the service system is uniformly stable.*
3. *The steady-state expected queue length is finite and differentiable on the feasible region.*

The first two conditions are adopted from Chen et al. (2024b). Under uniform stability, finite second moments are standard sufficient conditions for the finiteness of the steady-state expected queue length. We impose the differentiability condition because the PŁK condition is formulated in terms of the gradient of the objective. This differentiability condition can be verified directly for the distributional classes analyzed later, including the exponential and Gamma cases.

In general, (1) is non-convex (Harel 1990). To address this issue, we introduce two different approaches to construct convex reformulations of (1) in the following section. The analysis applies to a broad class of $GI/GI/1$ queueing systems, including $\Gamma_k/\Gamma_m/1$ queues with $k, m \ge 1$ and $GI/M/1$ queues with log-concave interarrival times.

3. Convex Reformulation

We begin by analyzing the expected queue length $\mathbb{E}[Q_\infty(\lambda,\mu)]$, whose closed-form expression is generally unavailable for $GI/GI/1$ queues. Nevertheless, we can relate it to the expected waiting time via Little's law (Little 1961) by $\mathbb{E}[Q_\infty(\lambda,\mu)] = \lambda\mathbb{E}[W_\infty(\lambda,\mu)]$. Here, $W_\infty(\lambda,\mu)$ denotes the stationary waiting time, i.e., the limit of waiting times $W_n(\lambda,\mu)$ as $n \to \infty$ and the sequence $\{W_n(\lambda,\mu)\}_n$ satisfies Lindley's equation (Lindley 1952):

$$W_{n+1}(\lambda,\mu) = \left(W_n(\lambda,\mu) + \frac{S_n}{\mu} - \frac{T_n}{\lambda}\right)^+.$$

Under Assumption 1, the increments have negative drift with $\mathbb{E}[S_i/\mu - T_i/\lambda] < 0$. This ensures the stability and allows us to apply the celebrated Spitzer's identity (e.g., Kingman 1962):

$$\mathbb{E}\left[W_\infty(\lambda,\mu)\right] = \sum_{n=1}^{\infty} \frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n} \frac{S_i}{\mu} - \sum_{i=1}^{n} \frac{T_i}{\lambda}\right)^+\right].$$

Define

$$l_n(\rho) := \frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n} \rho S_i - \sum_{i=1}^{n} T_i\right)^+\right] \quad \text{and} \quad L_q(\rho) := \sum_{n=1}^{\infty} l_n(\rho).$$

For general cases, $L_q(\rho)$ is defined on the stability interval $\rho < \mathbb{E}[T_1]/\mathbb{E}[S_1]$. In the unit-mean normalization, the stability interval reduces to $\rho \in (0,1)$. By Little's law, it follows that $\mathbb{E}[Q_\infty(\lambda,\mu)] = L_q(\lambda/\mu)$. For any realization of $\{S_i, T_i\}$, the function $(\sum_{i=1}^n \rho S_i - \sum_{i=1}^n T_i)^+$ is a maximum of two linear functions and is therefore convex. Since convexity is preserved under expectation and countable summation, $L_q(\rho)$ is convex. However, the composition term λ/μ destroys the joint convexity. We present two approaches for constructing a convex reformulation: the first involves a change of variables with respect to λ, and the second involves a change of variables with respect to μ.

3.1. Reformulation I

Define $\hat{\lambda} := \sqrt{\lambda}$, let $\pi(x,y) := x^2/y$ for $y > 0$, and define $\hat{r}(\hat{\lambda}) := r(\hat{\lambda}^2)$. Then, (1) can be reformulated as

$$\begin{aligned} \min_{\hat{\lambda},\mu} \quad & L_q\left(\pi(\hat{\lambda},\mu)\right) + c(\mu) - \hat{r}(\hat{\lambda}) \\ \text{s.t.} \quad & \sqrt{\underline{\lambda}} \le \hat{\lambda} \le \sqrt{\bar{\lambda}}, \quad \underline{\mu} \le \mu \le \bar{\mu}. \end{aligned} \tag{R1}$$

To ensure the convexity (or strong convexity) of (R1), we need to make the following assumption.

ASSUMPTION 2. *We impose one of the following two conditions:*

1. $\hat{r}(\hat{\lambda})$ *is concave in* $\hat{\lambda}$ *and* $c(\mu)$ *is convex in* μ.
2. $\hat{r}(\hat{\lambda})$ *is* κ_1*-strongly concave in* $\hat{\lambda}$ *and* $c(\mu)$ *is* κ_1*-strongly convex in* μ.

THEOREM 1. *Suppose Assumption 1 holds. Under Assumption 2.1, the optimization problem (R1) is convex. If, moreover, Assumption 2.2 holds, then problem (R1) is* κ_1*-strongly convex.*

Proof of Theorem 1. For any fixed realization of $\{S_i, T_i\}$, the map $\rho \mapsto \left(\sum_{i=1}^n \rho S_i - \sum_{i=1}^n T_i\right)^+$ is convex and nondecreasing in ρ. Therefore, $L_q(\rho)$ is convex and nondecreasing on $(0,1)$, since both properties are preserved under expectation and countable summation. Next, $\pi(\hat{\lambda},\mu)$ is jointly convex in $(\hat{\lambda},\mu)$, because it is the perspective of the convex quadratic function: $\pi(\hat{\lambda},\mu) = \mu(\hat{\lambda}/\mu)^2$. By Assumption 1, $\pi(\hat{\lambda},\mu) \in (0,1)$.

Since L_q is convex and nondecreasing on $(0,1)$, the composition rule for convex functions implies that $L_q(\pi(\hat{\lambda},\mu))$ is jointly convex in $(\hat{\lambda},\mu)$.

By Assumption 2.1, $c(\mu)-\hat{r}(\hat{\lambda})$ is jointly convex in $(\hat{\lambda},\mu)$. Hence, problem (R1) is convex. Moreover, by Assumption 2.2, $c(\mu)-\hat{r}(\hat{\lambda})$ is jointly κ_1-strongly convex in $(\hat{\lambda},\mu)$. Adding the convex term $L_q(\pi(\hat{\lambda},\mu))$ preserves strong convexity. Therefore, problem (R1) is κ_1-strongly convex. □

REMARK 1. Besides convexity, (R1) is second-order cone (SOC) representable for $M/M/1$ queue if $c(\mu)$ and $-\hat{r}(\hat{\lambda})$ are SOC representable. When $S_n, T_n \sim \exp(1)$, $L_q(\rho)=\rho^2/(1-\rho)$ and (R1) is equivalent to

$$\begin{aligned}
\min_{\hat{\lambda},\mu} \quad & t_q+t_c+t_r \\
\text{s.t.} \quad & \rho^2/(1-\rho)\le t_q, \quad c(\mu)\le t_c, \quad -\hat{r}(\hat{\lambda})\le t_r, \quad \rho\ge \hat{\lambda}^2/\mu, \\
& \sqrt{\underline{\lambda}}\le\hat{\lambda}\le\sqrt{\bar{\lambda}}, \quad \underline{\mu}\le\mu\le\bar{\mu}.
\end{aligned}$$

The constraint $\rho^2/(1-\rho)\le t_q$ is equivalent to $2t_q(1-\rho)\ge\|\sqrt{2}\rho\|_2^2$ and $\rho\ge\hat{\lambda}^2/\mu$ is equivalent to $2\rho\mu\ge\|\sqrt{2}\hat{\lambda}\|_2^2$. All constraints are SOC representable. Thus, the optimization problem is a second-order cone program (SOCP), which can be efficiently solved by solvers such as CPLEX and Gurobi.

To interpret Assumption 2, we adopt the standard revenue model in which the price p and the arrival rate λ are linked by an invertible mapping. Under this specification, revenue can be written equivalently as $r(\lambda)=\lambda p(\lambda)$ or $r(p)=p\lambda(p)$. When r is twice differentiable, $\hat{r}(\hat{\lambda})=r(\hat{\lambda}^2)$ is concave if and only if $2r'(\lambda)+4\lambda r''(\lambda)\le 0$, and it is κ_1-strongly concave if $2r'(\lambda)+4\lambda r''(\lambda)\le-\kappa_1$ throughout the feasible region. As an example, consider the demand function $\lambda(p)=p^{-\gamma}$ with $\gamma\in(1,2]$. In this case, $\hat{r}(\hat{\lambda})=\hat{\lambda}^{2-2/\gamma}$, which is concave for all $\gamma\in(1,2]$, and is strongly concave on any compact feasible interval when $\gamma\in(1,2)$.

However, for many demand functions studied in revenue management (Talluri and Van Ryzin 2005, Besbes and Zeevi 2015), Assumption 2 may not hold globally.

- For linear demand $\lambda(p)=a-bp$ with $a,b>0$, we have $r(\lambda)=(a\lambda-\lambda^2)/b$, so that $2r'(\lambda)+4\lambda r''(\lambda)=(2a-12\lambda)/b$ and hence the condition holds whenever $\lambda\ge a/6$.
- For exponential demand $\lambda(p)=ae^{-bp}$ with $a,b>0$, we have $r(\lambda)=\frac{\lambda}{b}\log\left(\frac{a}{\lambda}\right)$. Thus, $2r'(\lambda)+4\lambda r''(\lambda)=\frac{2}{b}\left(\log\left(\frac{a}{\lambda}\right)-3\right)$, which is nonpositive whenever $\lambda\ge ae^{-3}$.
- For logit demand $\lambda(p):=\frac{\exp(a-p)}{1+\exp(a-p)}$, we have $\hat{r}(\hat{\lambda})=\hat{\lambda}^2\ln(1-\hat{\lambda}^2)-\hat{\lambda}^2\ln(\hat{\lambda}^2)+a\hat{\lambda}^2$, which is generally non-concave, although one can still characterize local regions on which concavity holds. For example, when $a=1$, $\hat{r}(\hat{\lambda})$ is concave for $\hat{\lambda}>\hat{\lambda}^*/2$, where $\hat{\lambda}^*$ denotes the global maximizer of $\hat{r}(\hat{\lambda})$.

Hence, to fully accommodate these demand functions, we propose an alternative approach to constructing the convex reformulation.

3.2. Reformulation II

Define $\hat{\mu} := \mu^2$ and $\hat{c}(\hat{\mu}) := c(\sqrt{\hat{\mu}})$. Then, (1) can be reformulated as

$$\begin{aligned} \min_{\lambda,\hat{\mu}} \quad & L_q\left(\sqrt{\pi(\lambda,\hat{\mu})}\right) + \hat{c}(\hat{\mu}) - r(\lambda) \\ \text{s.t.} \quad & \underline{\lambda} \le \lambda \le \bar{\lambda}, \quad \underline{\mu}^2 \le \hat{\mu} \le \bar{\mu}^2. \end{aligned} \tag{R2}$$

Similar to (R1), we require the following assumption to ensure the convexity (or strong convexity) of (R2).

ASSUMPTION 3. *We impose one of the following two conditions:*

1. *$\hat{c}(\hat{\mu})$ is convex in $\hat{\mu}$ and $r(\lambda)$ is concave in λ.*
2. *$\hat{c}(\hat{\mu})$ is κ_2-strongly convex in $\hat{\mu}$ and $r(\lambda)$ is κ_2-strongly concave in λ.*

When c is twice differentiable, $\hat{c}(\cdot)$ is convex if and only if $\mu c''(\mu) - c'(\mu) \ge 0$. For example, $c(\mu) = \alpha\mu^\gamma$ with $\gamma \ge 2$ and $\alpha \ge 0$ is a simple class that satisfies Assumption 3. Setting $\gamma = 2$ recovers the quadratic cost function commonly used in the literature (Ata and Shneorson 2006, Chen et al. 2024b). Moreover, the strong concavity requirement of r in Assumption 3.2 holds true for a wide range of demand functions. For example, under linear demand $\lambda(p) = a - bp$ with $a, b > 0$, the revenue function $r(\lambda) = (a\lambda - \lambda^2)/b$ is a $2/b$-strongly concave quadratic. Under the logit demand model, $r(\lambda) = \lambda\ln(1-\lambda) - \lambda\ln(\lambda) + a\lambda$, which is $27/4$-strongly concave on $(0,1)$. Hence, (R2) is broadly applicable as long as the service cost remains convex under the square root transformation, which typically holds under mild conditions.

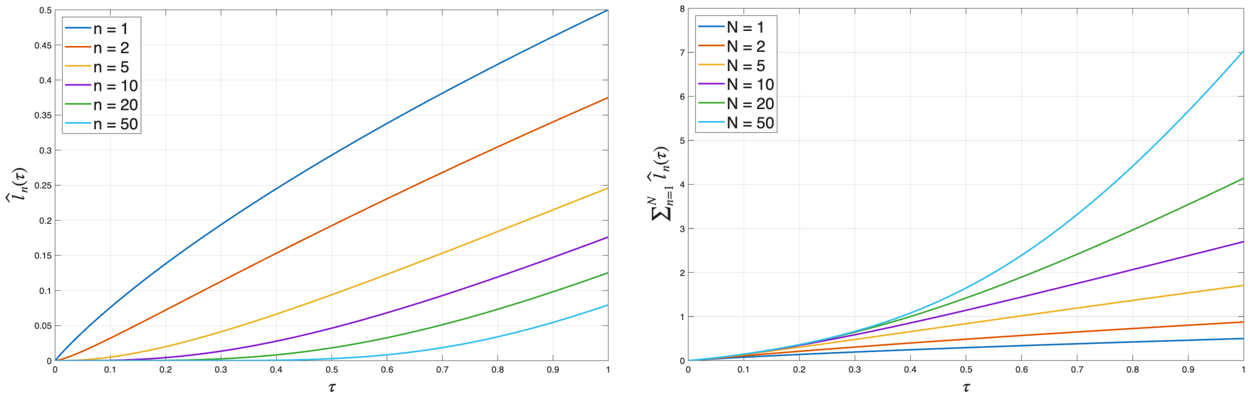
By Assumption 3, $\hat{c}(\hat{\mu}) - r(\lambda)$ is jointly convex in $(\lambda, \hat{\mu})$. For $\tau \in (0,1)$, define

$$\hat{l}_n(\tau) := \frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^n \sqrt{\tau} S_i - \sum_{i=1}^n T_i\right)^+\right] \quad \text{and} \quad \hat{L}_q(\tau) := \sum_{n=1}^\infty \hat{l}_n(\tau).$$

Then, $\hat{L}_q(\pi(\lambda,\hat{\mu})) = L_q(\sqrt{\pi(\lambda,\hat{\mu})})$. To establish the convexity of (R2), it is sufficient to prove that $\hat{L}_q(\tau)$ is convex because $\hat{L}_q(\tau)$ is non-decreasing and $\pi(\lambda,\hat{\mu})$ is jointly convex. However, this is far from trivial, because $\hat{l}_n(\tau)$ can be non-convex (or even concave) for small n. Although $\hat{l}_n(\tau)$ eventually becomes convex as n grows large, its magnitude seems to decay rapidly (see Figure 1a for an illustration). Hence, it remains unclear whether the curvature of $\hat{L}_q(\tau)$ is primarily driven by the finitely many non-convex initial terms or by the infinitely many rapidly vanishing convex tails. For the $M/M/1$ queue, the summation $\sum_{n=1}^N \hat{l}_n(\tau)$ appears to be dominated by the convex terms $\hat{l}_n(\tau)$ corresponding to large n as $N \to \infty$ (see Figure 1b). In what follows, we rigorously establish the convexity result not only for the $M/M/1$ queue but also for $\Gamma_k/\Gamma_m/1$ queues with $k, m \ge 1$ and $GI/M/1$ queues with log-concave interarrival times.

Special Case: $M/M/1$ Queue To build intuition, we start with the $M/M/1$ queue where $S_n, T_n \sim \exp(1)$. Then, the interarrival time T_n/λ and the service time S_n/μ follow $\exp(\lambda)$ and $\exp(\mu)$, respectively. As a special case of the $GI/GI/1$ queue, the $M/M/1$ queue admits a closed-form expression of the expected queue length, e.g., $\mathbb{E}[Q_\infty(\lambda,\mu)] = \lambda^2/[\mu(\mu-\lambda)]$. Hence, $\hat{L}_q(\tau)$ simplifies to $\tau/(1-\sqrt{\tau})$.

Figure 1 Plots of $\hat{l}_n(\tau)$ and $\sum_{n=1}^{N}\hat{l}_n(\tau)$ when $S_n, T_n \sim \exp(1)$.



(a) Plot of $\hat{l}_n(\tau)$ for different n. (b) Plot of $\sum_{n=1}^{N}\hat{l}_n(\tau)$ for different N.

THEOREM 2. *Suppose Assumption 1 holds. In the case $T_n, S_n \overset{i.i.d.}{\sim} \exp(1)$, the function $\hat{L}_q(\tau)$ is convex on $(0,1)$. Under Assumption 3.1, the reformulated problem (R2) is convex. If, moreover, Assumption 3.2 is satisfied, then problem (R2) is κ_2-strongly convex.*

Proof of Theorem 2. For any $\tau \in (0,1)$, the geometric series expansion gives $\hat{L}_q(\tau) = \tau/(1-\sqrt{\tau}) = \sum_{n=0}^{\infty}\tau^{1+n/2}$. Since each term $\tau^{1+n/2}$ is convex and increasing on $(0,1)$, it follows that $\hat{L}_q$ is convex and increasing on $(0,1)$. By Assumption 1, $\pi(\lambda,\hat{\mu}) \in (0,1)$. Since $\pi(\lambda,\hat{\mu})$ is convex and $\hat{L}_q$ is convex and nondecreasing, the composition rule for convex functions implies that $\hat{L}_q(\pi(\lambda,\hat{\mu}))$ is jointly convex in $(\lambda,\hat{\mu})$.

Under Assumption 3.1, the function $\hat{c}(\hat{\mu}) - r(\lambda)$ is jointly convex in $(\lambda,\hat{\mu})$. Hence, problem (R2) is convex. Moreover, by Assumption 3.2, $\hat{c}(\hat{\mu}) - r(\lambda)$ is jointly κ_2-strongly convex in $(\lambda,\hat{\mu})$. Since adding a convex function preserves strong convexity, problem (R2) is κ_2-strongly convex. □

Although Theorem 2 is subsumed by more general Gamma distribution results presented in the following subsection, we state the theorem explicitly because the convex reformulation for $M/M/1$ queues extends naturally to other settings, including Jackson networks of $M/M/1$ queues, $M/GI/1$ queues, and approximate formulations of $GI/GI/s$ queues.

Extension: Jackson Network Consider a Jackson queueing network with N stations. Each station $i \in [N]$ has an external Poisson arrival process with rate Λ_i. Revenue is collected upon admission of exogenous arrivals to the network. The service time at station i follows an exponential distribution with rate μ_i. A customer completing service at station i will either move to some new station j with probability P_{ij} or leave the system with probability $1-\sum_{j=1}^{N}P_{ij}$. We assume the network is open, i.e., every customer can only visit finitely many stations before leaving the network with probability one. For example, the network is open if the strict inequality $\sum_{j=1}^{N}P_{ij} < 1$ holds for any station $i \in [N]$. Then, the actual arrival rate λ_i at station i satisfies the balance equation

$$\lambda_i = \Lambda_i + \sum_{j=1}^{N} P_{ji}\lambda_j, \quad \forall i \in [N].$$

Applying Jackson's Theorem (Jackson 1957, 1963), the expected queue length at station i is $\lambda_i^2/[\mu_i(\mu_i - \lambda_i)]$. Then, we aim to solve the following optimization problem:

$$
\begin{aligned}
\min_{\lambda,\mu,\Lambda} \quad & \sum_{i=1}^{N} \left[\frac{\lambda_i^2}{\mu_i(\mu_i - \lambda_i)} + c(\mu_i) - r(\Lambda_i) \right] \\
\text{s.t.} \quad & \lambda_i = \Lambda_i + \sum_{j=1}^{N} P_{ji}\lambda_j, \qquad\qquad \forall i \in [N], \\
& \underline{\lambda} \le \lambda_i \le \bar{\lambda}, \quad \underline{\mu} \le \mu_i \le \bar{\mu}, \quad 0 \le \Lambda_i \le \bar{\Lambda}, \qquad \forall i \in [N].
\end{aligned} \tag{2}
$$

Define $\hat{\mu}_i := \mu_i^2$. Then, (2) can be reformulated as

$$
\begin{aligned}
\min_{\lambda,\hat{\mu},\Lambda} \quad & \sum_{i=1}^{N} \left[\frac{\lambda_i^2}{\sqrt{\hat{\mu}_i}\left(\sqrt{\hat{\mu}_i} - \lambda_i\right)} + \hat{c}(\hat{\mu}_i) - r(\Lambda_i) \right] \\
\text{s.t.} \quad & \lambda_i = \Lambda_i + \sum_{j=1}^{N} P_{ji}\lambda_j, \qquad\qquad \forall i \in [N], \\
& \underline{\lambda} \le \lambda_i \le \bar{\lambda}, \quad \underline{\mu}^2 \le \hat{\mu}_i \le \bar{\mu}^2, \quad 0 \le \Lambda_i \le \bar{\Lambda}, \qquad \forall i \in [N].
\end{aligned} \tag{3}
$$

Under Assumption 3.1, each transformed expected queueing length is convex in $(\lambda_i, \hat{\mu}_i)$, each transformed service-cost term is convex in $\hat{\mu}_i$, and each revenue term $r(\Lambda_i)$ is concave in Λ_i. Since the balance equations and box constraints are affine, the reformulated Jackson-network problem (3) is convex. Crucially, this exact network extension relies on the product-form stationary distribution of Jackson networks, which makes the steady-state queueing cost separable across stations. Without such separability, e.g., a generalized Jackson network, the single-station convexity argument does not directly lift to the network level.

Extension: $M/GI/1$ Queue Consider the setting where the interarrival times are exponential random variables, and the service times follow general distributions with finite variance $\mathrm{Var}(S) < +\infty$. Applying the Pollaczek-Khinchine formula, (R2) reduces to

$$
\begin{aligned}
\min_{\lambda,\hat{\mu}} \quad & \frac{1 + \mathrm{Var}(S)}{2} \cdot \frac{\lambda^2}{\sqrt{\hat{\mu}}\left(\sqrt{\hat{\mu}} - \lambda\right)} + \hat{c}(\hat{\mu}) - r(\lambda) \\
\text{s.t.} \quad & \underline{\lambda} \le \lambda \le \bar{\lambda}, \quad \underline{\mu}^2 \le \hat{\mu} \le \bar{\mu}^2.
\end{aligned} \tag{4}
$$

The only difference between the optimization of $M/GI/1$ and $M/M/1$ queues lies in the coefficient in the objective. Therefore, we can similarly prove the convexity of (4).

Extension: $GI/GI/s$ Approximation Consider the setting where interarrival and service times follow general distributions with finite variance $\mathrm{Var}(T)$ and $\mathrm{Var}(S)$, respectively. Based on the approximation of $GI/GI/s$ system in Shanthikumar et al. (2007), (R2) can be approximated by

$$
\begin{aligned}
\min_{\lambda,\hat{\mu}} \quad & \frac{\mathrm{Var}(T) + \mathrm{Var}(S)}{2} \cdot \frac{\left(\lambda/\sqrt{\hat{\mu}}\right)^{\sqrt{2(s+1)}}}{1 - \lambda/\sqrt{\hat{\mu}}} + \hat{c}(\hat{\mu}) - r(\lambda) \\
\text{s.t.} \quad & \underline{\lambda} \le \lambda \le \bar{\lambda}, \quad \underline{\mu}^2 \le \hat{\mu} \le \bar{\mu}^2.
\end{aligned} \tag{5}
$$

Similar to Theorem 2, the first term of the objective is a composition of two functions $\hat{L}_q(\pi(\lambda,\hat{\mu}))$ with $\hat{L}_q(\tau)=\tau^{\sqrt{2(s+1)}/2}/(1-\sqrt{\tau})$. For $s\geq 1$, $\hat{L}_q$ is monotone non-decreasing. Then, it is sufficient to check the convexity of $\hat{L}_q(\tau)$ by decomposing it via the Taylor's expansion for $\tau\in(0,1)$:

$$\hat{L}_q(\tau)=\frac{\tau^{\frac{\sqrt{2(s+1)}}{2}}}{1-\sqrt{\tau}}=\tau^{\frac{\sqrt{2(s+1)}}{2}}\sum_{n=0}^{\infty}\left(\sqrt{\tau}\right)^n=\sum_{n=0}^{\infty}\tau^{\frac{\sqrt{2(s+1)}+n}{2}}.$$

For each $n\geq 0$, the summand is convex, which implies that $\hat{L}_q(\tau)$ is convex. By the preservation under the composition of convex functions, we can establish the convexity of (5).

$\Gamma_k/\Gamma_m/1$ **Queue** We turn to the more general setting of $\Gamma_k/\Gamma_m/1$ queues. To enforce $\mathbb{E}[T_n]=\mathbb{E}[S_n]=1$, we set $T_n\sim\Gamma(k,k)$ and $S_n\sim\Gamma(m,m)$ under the (shape, rate) parameterization. Leveraging the structural properties of Gamma distributions, we establish the main result of this paper: $\hat{L}_q(\tau)$ is convex, despite being composed of finitely many non-convex components and infinitely many rapidly diminishing convex tails.

THEOREM 3. *Suppose that Assumption 1 holds. In the case $T_n\sim\Gamma(k,k), S_n\sim\Gamma(m,m)$ with $k,m\geq 1$, the function $\hat{L}_q(\tau)$ is convex on $(0,1)$. Under Assumption 3.1, the reformulated optimization problem (R2) is convex. If moreover, Assumption 3.2 is satisfied, then problem (R2) is κ_2-strongly convex.*

Note that Theorem 3 holds for any $k,m\geq 1$. In the special case where both k and m are integers, we recover the Erlang queueing model $E_k/E_m/1$, a cornerstone in queueing theory.

$GI/M/1$ **Queue with Log-concave Interarrival Times** In addition to the Gamma family with shape parameters no less than 1, the convexity results also apply to a broad class of $GI/M/1$ queues. The next theorem shows that when service times are exponential and the interarrival-time density is log-concave on its support, the transformed expected queue length $\hat{L}_q(\tau)$ remains convex.

THEOREM 4. *Suppose that Assumption 1 holds. If T_n has a log-concave density on its support, where the support is an interval contained in $[0,\infty)$, and $S_n\sim\exp(1)$, the function $\hat{L}_q(\tau)$ is convex on $(0,1)$. Under Assumption 3.1, the reformulated optimization problem (R2) is convex. If moreover, Assumption 3.2 is satisfied, then problem (R2) is κ_2-strongly convex.*

Theorem 4 complements Theorem 3 by establishing the convexity of $\hat{L}_q(\tau)$ beyond the Gamma family to log-concave interarrival-time distributions. This result further shows that the convexity result does not rely solely on distribution-specific expressions but also follows from structural properties of the distribution.

REMARK 2. The square-root transformations used in Reformulations (R1) and (R2) are not the only transformations that can yield convex reformulations. For any $\alpha,\beta>0$, define

$$\hat{\lambda}_\alpha=\lambda^{1/\alpha},\qquad \hat{\mu}_\beta=\mu^\beta,\qquad \hat{r}_\alpha(\hat{\lambda}_\alpha)=r(\hat{\lambda}_\alpha^\alpha),\qquad \hat{c}_\beta(\hat{\mu}_\beta)=c(\hat{\mu}_\beta^{1/\beta}).$$

Under this joint transformation, we have

$$L_q(\lambda/\mu) = L_q\left(\frac{\hat{\lambda}_\alpha^\alpha}{\hat{\mu}_\beta^{1/\beta}}\right) = \hat{L}_q\left(\frac{\hat{\lambda}_\alpha^{2\alpha}}{\hat{\mu}_\beta^{2/\beta}}\right), \qquad \hat{L}_q(\tau) = L_q(\sqrt{\tau}).$$

For the $M/M/1$ queue, the expected queue-length has the closed form $L_q(\rho) = \rho^2/(1-\rho)$. A direct Hessian calculation shows that the transformed expected queue length is jointly convex in $(\hat{\lambda}_\alpha, \hat{\mu}_\beta)$ on the stable domain if and only if $2\alpha \geq 1 + 2/\beta$. More generally, consider the following two settings:

- If L_q is convex and nondecreasing, as in the setting of Theorem 1, then the transformed expected queue-length component remains convex in $(\hat{\lambda}_\alpha, \hat{\mu}_\beta)$ whenever $\alpha \geq 1 + 1/\beta$. This follows from the fact that $(x, y) \mapsto x^a/y^b$ is jointly convex on the positive orthant whenever $b \geq 0$ and $a \geq b + 1$. In particular, this condition is satisfied by $\alpha = 2, \beta = 1$, which reduces to Reformulation (R1).
- If $\hat{L}_q$ is convex and nondecreasing, as in the settings of Theorems 3 and 4, then the transformed expected queue-length component remains convex in $(\hat{\lambda}_\alpha, \hat{\mu}_\beta)$ whenever $2\alpha \geq 1 + 2/\beta$. In particular, this condition is satisfied by $\alpha = 1, \beta = 2$, which reduces to Reformulation (R2). Within the family $\alpha = 1, \beta \geq 2$, the choice $\beta = 2$ imposes the least restrictive condition on the service-cost function, because convexity of $\hat{c}_\beta$ with $\beta \geq 2$ implies convexity of $\hat{c}(\hat{\mu}) = c(\sqrt{\hat{\mu}})$ when c is nondecreasing.

Thus, power transformations in this family can yield convex reformulations, subject to the corresponding concavity and convexity requirements on the transformed revenue and service-cost components.

Counterexample: Mixture of Exponential Distributions Given the positive results for Gamma distributions and log-concave distributions, one may naturally wonder whether the same property holds for general phase-type distributions, i.e., whether $\hat{L}_q(\tau)$ remains convex when S_n and T_n are phase-type random variables. However, this is not the case, as stated in the following theorem.

THEOREM 5. *$\hat{L}_q(\tau)$ can be non-convex on $(0, 1)$ when S_n and T_n are general phase-type random variables.*

The proof of Theorem 5 follows a constructive approach. Specifically, we present a counterexample in which S_n and T_n are mixtures of exponential distributions, demonstrating that $\hat{L}_q(\tau)$ is not convex. By definition, it is sufficient to prove that there exists $0 < \tau_1 < \tau_2 < 1$, such that $\sum_{n=1}^\infty \hat{l}_n'(\tau_1) > \sum_{n=1}^\infty \hat{l}_n'(\tau_2)$.

4. Polyak-Łojasiewicz-Kurdyka (PŁK) Condition

Building on the convex reformulations, we analyze the landscape of the original non-convex problem (1). Our analysis relies on a specific form of the PŁK condition (also referred to as the KŁ condition in Karimi et al. 2016; see their Appendix G). We refer interested readers to Bento et al. (2025, Definition 1) for the general formulation. Following Bento et al. (2025), we adopt the PŁK terminology, rather than the more common PŁ or KŁ condition, to acknowledge the foundational contributions of Polyak et al. (1963), Łojasiewicz (1963), and Kurdyka (1998).

DEFINITION 1 (PŁK CONDITION). Consider a convex and compact set $\mathcal{X} \subseteq \mathbb{R}^n$ and a differentiable function f. Denote f^* as the optimal function value with $f^* := \min_{x\in\mathcal{X}} f(x)$. The function f satisfies the PŁK condition on $\mathcal{X}$ if there exists $\nu > 0$ such that

$$f(x) - f^* \leq \frac{1}{2\nu} \min_{g\in\partial\delta_{\mathcal{X}}(x)} \|\nabla f(x) + g\|_2^{\alpha} \quad \forall x \in \mathcal{X},$$

where $\nu > 0$ denotes the PŁK constant and α is a real-valued PŁK exponent satisfying $\alpha \geq 1$.

In Definition 1, $\partial\delta_{\mathcal{X}}(x)$ denotes the subdifferential of $\delta_{\mathcal{X}}(x)$ and is the normal cone of $\mathcal{X}$ at x. Despite variations such as whether the domain is bounded or not, or whether the objective function is differentiable or not, an essential property of the PŁK condition is that optimization problems satisfying it exclude all suboptimal stationary or local optimal points (Karimi et al. 2016). Hence, the PŁK condition serves as a perfect candidate to explain the empirical success of first-order methods.

THEOREM 6. *Suppose Assumption 1 holds. Consider problem (1) in one of the following settings:*

1. *If Assumption 2.1 holds, then (1) satisfies the PŁK condition with exponent $\alpha = 1$. Moreover, if Assumption 2.2 is satisfied, then (1) satisfies the PŁK condition with exponent $\alpha = 2$.*
2. *Suppose that $T_n \sim \Gamma(k,k)$ and $S_n \sim \Gamma(m,m)$ with $k,m \geq 1$. If Assumption 3.1 holds, then (1) satisfies the PŁK condition with exponent $\alpha = 1$. Moreover, if Assumption 3.2 is satisfied, then (1) satisfies the PŁK condition with exponent $\alpha = 2$.*

The intuition behind Theorem 6 is as follows. The original problem (1) can be viewed as the reformulation composed with a transformation map of decision variables. For simplicity, let us express (1) by:

$$\min_{x\in\mathcal{X}} F(x) = H(\varphi(x)), \tag{6}$$

where F and $\mathcal{X}$ are the objective function and feasible region of (1), respectively, x denotes the decision variables (λ, μ), $\varphi(x)$ is the transformation map used to construct the reformulation, and H is the objective function of the reformulation. Li and Pong (2018, Theorem 3.2) and Fatkhullin et al. (2025, Proposition 2) prove that the PŁK condition can be established through such a composition when the transformation map is uniformly inverse-Lipschitz on the feasible region. Specifically, if the reformulated feasible region $\mathcal{U} = \varphi(\mathcal{X})$ is closed and convex, and there exists $\sigma_\varphi > 0$ such that

$$\|\varphi(x) - \varphi(y)\|_2 \geq \sigma_\varphi \|x - y\|_2, \qquad \forall x, y \in \mathcal{X},$$

then convexity of H on $\mathcal{U}$ implies that $F = H \circ \varphi$ satisfies the PŁK condition with exponent $\alpha = 1$ on $\mathcal{X}$. Moreover, if H is strongly convex on $\mathcal{U}$, then F satisfies the PŁK condition with exponent $\alpha = 2$.

In our case, we have established the convexity and strong convexity results for reformulations (R1) and (R2) in Theorem 1 and Theorem 3, respectively. In the proof of Theorem 6, we verify the required inverse-Lipschitz condition directly for the two transformation maps $\varphi_1(\lambda, \mu) = (\sqrt{\lambda}, \mu)$ and $\varphi_2(\lambda, \mu) = (\lambda, \mu^2)$.

REMARK 3. For the Jackson networks, the $M/GI/1$ queue, the $GI/GI/s$ approximation, and the $GI/M/1$ queue with log-concave interarrival times discussed in Section 3.2, the proof extends directly to establish the PŁK condition with exponent $\alpha = 1$ under Assumption 3.1, and with exponent $\alpha = 2$ under Assumption 3.2.

Theorem 6 shows that the original optimization problem satisfies the PŁK condition, providing a strong characterization of the non-convex landscape. This insight also extends to formulations in which the decision variables are price p and service rate μ, rather than arrival rate λ and service rate μ. Specifically, with the demand model $\lambda = \lambda(p)$, where $\lambda(\cdot)$ is continuously differentiable and $\inf_{p\in\mathcal{P}} |\lambda'(p)| > 0$ on a compact price interval $\mathcal{P}$, the mapping $(p, \mu) \mapsto (\lambda(p), \mu)$ is uniformly inverse-Lipschitz. Hence, by the same argument used in the proof of Theorem 6, the PŁK condition is preserved under this change of variables, up to a modification of the constant. Therefore, the formulation with decision variables (p, μ) also satisfies a PŁK condition. This shows that our theory directly applies to the practically relevant setting in which the decision maker controls price rather than arrival rate, and thus provides a theoretical explanation for the success of first-order methods in the (p, μ) formulations studied by Chen et al. (2024b) and Hu et al. (2024).

5. Technical Proofs

5.1. Proof of Theorem 3

Proof of Theorem 3. For the proof, it is convenient to rescale the unit-mean Gamma variables. Let $\widetilde{T}_n := kT_n \sim \Gamma(k,1)$ and $\widetilde{S}_n := mS_n \sim \Gamma(m,1)$, and let $\widetilde{L}_q$ denote the corresponding transformed queue-length function. Then, $\hat{L}_q(\tau) = \frac{1}{k}\widetilde{L}_q(\tau k^2/m^2)$. Therefore, it suffices to prove that $\widetilde{L}_q$ is convex on $(0, k^2/m^2)$. For notational simplicity, we write $\hat{L}_q$ for $\widetilde{L}_q$ and suppress the tildes on S_n and T_n throughout the proof.

We first establish a monotonicity property needed later. For $n \geq 1$, define

$$s_n := \frac{\Gamma((k+m)n)}{\Gamma(mn)\Gamma(kn)} \left[\frac{m^m k^k}{(m+k)^{m+k}}\right]^n.$$

We claim that s_n is non-decreasing in n. To see this, define, for $x > 0$,

$$h(x) := \ln\Gamma((k+m)x) - \ln\Gamma(kx) - \ln\Gamma(mx) - x\left[k\ln\left(\frac{k+m}{k}\right) + m\ln\left(\frac{k+m}{m}\right)\right].$$

Therefore, $s_n = e^{h(n)}$. Let $\psi(x) := \Gamma'(x)/\Gamma(x)$ be the digamma function. Using its integral representation (Abramowitz and Stegun 1965, Equation (6.3.22) on Page 259), we obtain

$$\begin{aligned} h'(x) = k\int_0^\infty \frac{e^{-kxt} - e^{-(k+m)xt}}{1-e^{-t}}dt + m\int_0^\infty \frac{e^{-mxt} - e^{-(k+m)xt}}{1-e^{-t}}dt \\ - k\ln\left(\frac{k+m}{k}\right) - m\ln\left(\frac{k+m}{m}\right). \end{aligned}$$

Since $1 - e^{-t} \leq t$ and the numerators in the two integrals are non-negative, Frullani's integral (Gradshteyn and Ryzhik 2014, Equation (3.434.2) on Page 363) gives

$$\begin{aligned} h'(x) \geq k\int_0^\infty \frac{e^{-kxt} - e^{-(k+m)xt}}{t}dt + m\int_0^\infty \frac{e^{-mxt} - e^{-(k+m)xt}}{t}dt \\ - k\ln\left(\frac{k+m}{k}\right) - m\ln\left(\frac{k+m}{m}\right) = 0. \end{aligned}$$

Thus, s_n is non-decreasing in n.

Now define

$$U_n := \sum_{i=1}^{n}(S_i + T_i), \qquad V_n := \frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n}(S_i + T_i)}.$$

By the standard beta-gamma decomposition, $U_n \sim \Gamma((k+m)n, 1)$, $V_n \sim \text{Beta}(mn, kn)$, and U_n is independent of V_n. Let p_n denote the density of V_n, i.e.,

$$p_n(v) = \frac{\Gamma((k+m)n)}{\Gamma(mn)\Gamma(kn)} v^{mn-1}(1-v)^{kn-1}.$$

Hence,

$$\hat{l}_n(\tau) = \frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n}\sqrt{\tau}S_i - \sum_{i=1}^{n}T_i\right)^+\right] = \frac{1}{n}\mathbb{E}\left[U_n\left((\sqrt{\tau}+1)V_n - 1\right)^+\right] = (k+m)\mathbb{E}\left[\left((\sqrt{\tau}+1)V_n - 1\right)^+\right].$$

It follows that

$$\hat{L}_q(\tau) = (k+m)\sum_{n=1}^{\infty}\int_a^1 \left[(1+\sqrt{\tau})v - 1\right] p_n(v)dv, \qquad a := \frac{1}{1+\sqrt{\tau}}. \tag{7}$$

Since $\tau \in (0, k^2/m^2)$, we have $a \in (m/(m+k), 1)$.

The termwise differentiations below are justified by uniform convergence on compact subintervals of $(0, k^2/m^2)$. This follows because, on such intervals, the differentiated terms are bounded by a summable geometrically decaying sequence. Differentiating (7) twice and using $a = (1+\sqrt{\tau})^{-1}$, we obtain

$$\hat{L}_q''(\tau) = \frac{(k+m)a^3}{4(1-a)^3}D(a), \qquad D(a) := \sum_{n=1}^{\infty}\left[a^2(1-a)p_n(a) - \int_a^1 vp_n(v)dv\right]. \tag{8}$$

Hence, it remains to prove that $D(a) \geq 0$ for all $a \in (m/(m+k), 1)$.

Observe that $D(1) = 0$. Thus, it is sufficient to prove that $D'(a) \leq 0$ on $(m/(m+k), 1)$. Since

$$p_n'(a) = \frac{mn - (k+m)an - (1-2a)}{a(1-a)}p_n(a),$$

direct differentiation gives

$$D'(a) = (2a - a^2)\sum_{n=1}^{\infty}p_n(a) - a\left[(k+m)a - m\right]\sum_{n=1}^{\infty}np_n(a). \tag{9}$$

Define

$$r(a) := \frac{(k+m)^{k+m}}{k^k m^m}a^m(1-a)^k.$$

For $a \in (m/(m+k), 1)$, we have $0 < r(a) < 1$, since $a^m(1-a)^k$ attains its maximum at $m/(m+k)$. Moreover,

$$p_n(a) = \frac{(k+m)^{k+m}}{k^k m^m}a^{m-1}(1-a)^{k-1}s_n\left[r(a)\right]^{n-1}.$$

Since s_n is non-decreasing in n and $0 < r(a) < 1$,

$$\sum_{n=1}^{\infty}np_n(a) \geq \frac{1}{1-r(a)}\sum_{n=1}^{\infty}p_n(a). \tag{10}$$

Indeed, after removing the common positive factor in $p_n(a)$, this follows from

$$\sum_{n=1}^{\infty} n s_n r^{n-1} = \sum_{q=1}^{\infty}\sum_{n=q}^{\infty} s_n r^{n-1} \geq \sum_{q=1}^{\infty}\sum_{n=q}^{\infty} s_q r^{n-1} = \frac{1}{1-r}\sum_{q=1}^{\infty} s_q r^{q-1},$$

where $r = r(a)$.

Combining (9) and (10), we obtain

$$D'(a) \leq a\left(\sum_{n=1}^{\infty} p_n(a)\right)\left[2 - a - \frac{(k+m)a - m}{1 - r(a)}\right]. \tag{11}$$

Therefore, it remains only to prove the scalar inequality

$$(k+m)a - m \geq (2-a)(1 - r(a)), \qquad a \in \left(\frac{m}{m+k}, 1\right). \tag{12}$$

If $a \geq (m+2)/(m+k+1)$, then $(k+m)a - m \geq 2 - a \geq (2-a)(1-r(a))$, so (12) holds. It remains to consider $m/(m+k) < a < (m+2)/(m+k+1)$. Define

$$\eta := \frac{(k+m)a - m}{2 - a} \in (0,1), \qquad a = \frac{m + 2\eta}{k + m + \eta}.$$

Then (12) is equivalent to $r(a) \geq 1 - \eta$. Substituting the above expression for a into $r(a)$ gives

$$r(a) = \left(\frac{k+m}{k+m+\eta}\right)^{k+m}\left(1 + \frac{2\eta}{m}\right)^{m}\left(1 - \frac{\eta}{k}\right)^{k}.$$

Let $s := k + m$. Since $0 < 2\eta/m < 2$ and $\ln(1+z) \geq z/2$ for $0 \leq z \leq 2$, we have $(1 + 2\eta/m)^m \geq e^{\eta}$. Also, $(1 + \eta/s)^s \leq e^{\eta}$. Therefore, $(1 + 2\eta/m)^m \geq (1 + \eta/s)^s$. In addition, Bernoulli's inequality gives $(1 - \eta/k)^k \geq 1 - \eta$. Hence,

$$r(a) \geq \left(\frac{s}{s+\eta}\right)^{s}\left(1 + \frac{\eta}{s}\right)^{s}(1-\eta) = 1 - \eta.$$

This proves (12). By (11), $D'(a) \leq 0$ on $(m/(m+k), 1)$. Since $D(1) = 0$, we have $D(a) \geq 0$ on $(m/(m+k), 1)$. Therefore, (8) implies that $\hat{L}_q$ is convex on $(0, k^2/m^2)$.

The preceding argument proves that $\widetilde{L}_q$ is convex and nondecreasing on $(0, k^2/m^2)$. Therefore, for the original unit-mean Gamma variables, $\hat{L}_q(\tau) = \frac{1}{k}\widetilde{L}_q(\tau k^2/m^2)$ is convex and nondecreasing on $(0,1)$. Under Assumption 1, $\pi(\lambda, \hat{\mu}) \in (0,1)$ for all feasible $(\lambda, \hat{\mu})$. Since $\hat{L}_q$ is convex and nondecreasing on $(0,1)$, and $\pi(\lambda, \hat{\mu})$ is convex, the composition rule implies that $\hat{L}_q(\pi(\lambda, \hat{\mu}))$ is jointly convex in $(\lambda, \hat{\mu})$. Under Assumption 3.1, the function $\hat{c}(\hat{\mu}) - r(\lambda)$ is jointly convex in $(\lambda, \hat{\mu})$. Therefore, problem (R2) is convex. Moreover, under Assumption 3.2, the function $\hat{c}(\hat{\mu}) - r(\lambda)$ is jointly κ_2-strongly convex in $(\lambda, \hat{\mu})$. Because the sum of a strongly convex function and a convex function remains strongly convex with the same modulus, problem (R2) is κ_2-strongly convex. $\square$

5.2. Proof of Theorem 4

Proof of Theorem 4. Let $\phi(u) := \mathbb{E}[e^{-uT_1}]$ with $u > 0$ be the Laplace transform of the interarrival-time distribution. For the $GI/M/1$ queue with traffic intensity $\rho \in (0,1)$, let $\sigma \in (0,1)$ be the non-trivial solution to $\sigma = \phi((1-\sigma)/\rho)$. The stationary waiting time in queue satisfies $\mathbb{P}(W_\infty > t) = \sigma e^{-\mu(1-\sigma)t}$ for $t \geq 0$. Hence, by Little's law, $L_q(\rho) = \lambda \mathbb{E}[W_\infty] = \rho\sigma/(1-\sigma)$. Define $u := (1-\sigma)/\rho$. Then,

$$\sigma = \phi(u), \qquad \rho(u) = \frac{1-\phi(u)}{u}, \qquad \hat{L}_q(\tau(u)) = \frac{\phi(u)}{u} = \frac{1}{u} - \rho(u), \qquad \tau(u) = \rho(u)^2.$$

Let $\bar{F}(t) := \mathbb{P}(T_1 > t)$. Since $\mathbb{E}[T_1] = 1$, integration by parts gives

$$\rho(u) = \frac{1-\phi(u)}{u} = \int_0^\infty e^{-ut}\bar{F}(t)dt. \tag{13}$$

Hence,

$$\rho'(u) = -\int_0^\infty te^{-ut}\bar{F}(t)dt < 0, \qquad u > 0.$$

Moreover, $\lim_{u\to 0^+}\rho(u) = \mathbb{E}[T_1] = 1$ and $\lim_{u\to\infty}\rho(u) = 0$. Thus, $u \mapsto \rho(u)$ is a bijection from $(0,\infty)$ to $(0,1)$, and $u \mapsto \tau(u)$ is a bijection from $(0,\infty)$ to $(0,1)$.

Define $C(u) := -\rho'(u)/\rho(u)$. By direct differentiation of the parametric representation $\hat{L}_q(\tau(u)) = 1/u - \rho(u)$ and $\tau(u) = \rho(u)^2$, we obtain

$$\hat{L}_q''(\tau(u)) = \frac{2C(u) - 2uC(u)^2 + uC'(u) + u^3\rho(u)C(u)^3}{4u^3\rho(u)^4C(u)^3}. \tag{14}$$

Since the denominator in (14) is positive, it remains to prove that the numerator is positive.

We next establish two bounds for $C(u)$. Let f be the density of T_1, extended by zero outside its support. Since f is log-concave on an interval, and log-concavity is preserved under integration, the survival function $\bar{F}(t)$ is log-concave on $[0,\infty)$. Therefore, for each fixed $u > 0$, the density

$$g_u(t) := \frac{e^{-ut}\bar{F}(t)}{\rho(u)}, \qquad t \geq 0,$$

is log-concave. Let Z_u denote a random variable with density g_u. By (13),

$$\mathbb{E}[Z_u] = C(u), \qquad C'(u) = -\mathrm{Var}(Z_u).$$

Let $\bar{G}_u$ and r_u denote the survival function and hazard rate of Z_u, respectively. Since g_u is log-concave, $\bar{G}_u$ is log-concave, and hence r_u is increasing. Moreover, $\bar{F}(0) = 1$, so $r_u(0) = g_u(0) = 1/\rho(u)$. Hence,

$$\bar{G}_u(t) = \exp\left\{-\int_0^t r_u(s)ds\right\} \leq e^{-t/\rho(u)}, \qquad \forall\, t \geq 0.$$

Integrating both sides gives

$$C(u) = \mathbb{E}[Z_u] = \int_0^\infty \bar{G}_u(t)dt \leq \rho(u). \tag{15}$$

We also need a lower bound on $C'(u)$. Since r_u is increasing, for any $s,t \geq 0$,

$$\frac{\bar{G}_u(t+s)}{\bar{G}_u(t)} = \exp\left\{-\int_t^{t+s} r_u(v)dv\right\} \leq \exp\left\{-\int_0^s r_u(v)dv\right\} = \bar{G}_u(s).$$

Thus,

$$\int_t^\infty \bar{G}_u(s)ds = \bar{G}_u(t)\int_0^\infty \frac{\bar{G}_u(t+s)}{\bar{G}_u(t)}ds \leq C(u)\bar{G}_u(t).$$

It follows that

$$\mathbb{E}[Z_u^2] = 2\int_0^\infty t\bar{G}_u(t)dt = 2\int_0^\infty\int_t^\infty \bar{G}_u(s)dsdt \leq 2C(u)\int_0^\infty \bar{G}_u(t)dt = 2C(u)^2.$$

Hence,

$$C'(u) = -\mathrm{Var}(Z_u) \geq -C(u)^2. \tag{16}$$

Combining (14) with (16), it is sufficient to show that $2C(u) - 3uC(u)^2 + u^3\rho(u)C(u)^3 > 0$. Set

$$x := uC(u), \qquad y := u\rho(u) = 1 - \phi(u).$$

Then, $x > 0$, $0 < y < 1$, and $x \leq y$ by (15). Moreover,

$$2C(u) - 3uC(u)^2 + u^3\rho(u)C(u)^3 = C(u)\left(2 - 3x + yx^2\right).$$

For any fixed $y \in (0,1)$, the function $2 - 3x + yx^2$ is decreasing in x on $(0,y]$, because $2yx - 3 \leq 2y^2 - 3 < 0$. Hence, $2 - 3x + yx^2 \geq y^3 - 3y + 2 = (1-y)^2(y+2) > 0$. Therefore, $\hat{L}_q''(\tau(u)) > 0$ for every $u > 0$. Since $u \mapsto \tau(u)$ is a bijection from $(0,\infty)$ to $(0,1)$, $\hat{L}_q(\tau)$ is convex on $(0,1)$.

Moreover, $\hat{L}_q$ is nondecreasing on $(0,1)$ because $L_q(\rho)$ is nondecreasing in ρ. Under Assumption 1, $\pi(\lambda,\hat{\mu}) \in (0,1)$ for all feasible $(\lambda,\hat{\mu})$. Since $\hat{L}_q$ is convex and nondecreasing on $(0,1)$, and $\pi(\lambda,\hat{\mu})$ is convex, the composition rule implies that $\hat{L}_q(\pi(\lambda,\hat{\mu}))$ is jointly convex in $(\lambda,\hat{\mu})$. Under Assumption 3.1, the function $\hat{c}(\hat{\mu}) - r(\lambda)$ is jointly convex in $(\lambda,\hat{\mu})$. Therefore, problem (R2) is convex. Moreover, under Assumption 3.2, the function $\hat{c}(\hat{\mu}) - r(\lambda)$ is jointly κ_2-strongly convex in $(\lambda,\hat{\mu})$. Because the sum of a strongly convex function and a convex function remains strongly convex, problem (R2) is κ_2-strongly convex. □

5.3. Proof of Theorem 5

Proof of Theorem 5. Consider a setting when S_n and T_n are mixtures of exponential random variables:

$$S_n, T_n \overset{\text{i.i.d.}}{\sim} \begin{cases} \mathrm{Exp}(\lambda_1 = 0.1), & \text{w.p. } \frac{1}{2}; \\ \mathrm{Exp}(\lambda_2 = 10), & \text{w.p. } \frac{1}{2}. \end{cases} \tag{17}$$

To enforce $\mathbb{E}[S_n] = \mathbb{E}[T_n] = 1$, we should use $S_n/\mathbb{E}[S_n]$ and $T_n/\mathbb{E}[T_n]$. In our example, $\mathbb{E}[S_n] = \mathbb{E}[T_n]$, so both sequences are rescaled by the same positive constant, which does not affect the non-convexity result. For simplicity, we keep using S_n and T_n as in (17) in the following analysis.

Let us define $X_n := \sum_{i=1}^n S_i$ and $Y_n := \sum_{i=1}^n T_i$. By definition,

$$\hat{l}_n(\tau) = \frac{1}{n}\mathbb{E}\left[\left(\sqrt{\tau}X_n - Y_n\right)^+\right] = \frac{1}{n}\int_0^\infty \left(\int_0^{\sqrt{\tau}x} (\sqrt{\tau}x - y) p_{X_n}(x) p_{Y_n}(y) dy\right) dx.$$

Applying the Leibniz rule:

$$\hat{l}_n'(\tau) = \frac{1}{2n\sqrt{\tau}}\int_0^\infty \left(\int_0^{\sqrt{\tau}x} x p_{X_n}(x) p_{Y_n}(y) dy\right) dx.$$

Set $\tau_1 = 10^{-6}$ and $\tau_2 = 0.01$. It is sufficient to prove that $\sum_{n=1}^\infty \hat{l}_n'(\tau_1) > \sum_{n=1}^\infty \hat{l}_n'(\tau_2)$.

Step 1: Lower bound of $\sum_{n=1}^\infty \hat{l}_n'(\tau)$: Since $\hat{l}_n'(\tau)$ is non-negative, we have:

$$\sum_{n=1}^\infty \hat{l}_n'(\tau) \geq \hat{l}_1'(\tau), \quad \forall \tau \in (0,1).$$

When $n = 1$, $X_1 = S_1$ and $Y_1 = T_1$ with PDFs:

$$p_{S_1}(s) = \frac{\lambda_1}{2}e^{-\lambda_1 s} + \frac{\lambda_2}{2}e^{-\lambda_2 s}, \quad p_{T_1}(t) = \frac{\lambda_1}{2}e^{-\lambda_1 t} + \frac{\lambda_2}{2}e^{-\lambda_2 t}.$$

By definition, we have:

$$\begin{aligned}
\hat{l}_1'(\tau) &= \frac{1}{2\sqrt{\tau}}\int_0^\infty s p_{S_1}(s)\left(\int_0^{\sqrt{\tau}s} p_{T_1}(t) dt\right) ds \\
&= \frac{1}{8\sqrt{\tau}}\sum_{i,j=1}^2 \lambda_i \int_0^\infty s e^{-\lambda_i s}\left(1 - e^{-\lambda_j\sqrt{\tau}s}\right) ds \qquad (18)\\
&= \frac{1}{8\sqrt{\tau}}\sum_{i,j=1}^2 \left[\frac{1}{\lambda_i} - \frac{\lambda_i}{(\lambda_i + \lambda_j\sqrt{\tau})^2}\right].
\end{aligned}$$

Plugging $\lambda_1 = 0.1$, $\lambda_2 = 10$, and $\tau_1 = 10^{-6}$ into the formulation, we have:

$$\sum_{n=1}^\infty \hat{l}_n'(\tau_1) \geq \hat{l}_1'(\tau_1) > 219.$$

Step 2: Upper bound of $\sum_{n=1}^\infty \hat{l}_n'(\tau)$: We split the summation into two parts: $n = 1$ and $n \geq 2$. When $n = 1$, from (18), we already have

$$\hat{l}_1'(\tau) = \frac{1}{8\sqrt{\tau}}\sum_{i=1}^2\sum_{j=1}^2 \left[\frac{1}{\lambda_i} - \frac{\lambda_i}{(\lambda_i + \lambda_j\sqrt{\tau})^2}\right].$$

Plugging $\lambda_1 = 0.1$, $\lambda_2 = 10$, and $\tau_2 = 0.01$ into the above formulation, we have $\hat{l}_1'(\tau_2) < 15$. When $n \geq 2$,

$$\hat{l}_n'(\tau) = \frac{1}{2n\sqrt{\tau}}\int_0^\infty x p_{X_n}(x)\left(\int_0^{\sqrt{\tau}x} p_{Y_n}(y) dy\right) dx = \frac{1}{2n\sqrt{\tau}}\mathbb{E}\left[X_n \mathbf{1}_{(Y_n \leq \sqrt{\tau}X_n)}\right].$$

For any $\alpha > 0$, we have $\mathbf{1}_{(Y_n \le \sqrt{\tau}X_n)} \le e^{\alpha(\sqrt{\tau}X_n - Y_n)}$. Thus, we have:

$$
\begin{aligned}
\hat{l}'_n(\tau) &\le \frac{1}{2n\sqrt{\tau}}\mathbb{E}\left[X_n e^{\alpha(\sqrt{\tau}X_n - Y_n)}\right] \\
&= \frac{1}{2n\sqrt{\tau}}\mathbb{E}\left[\left(\sum_{i=1}^{n} S_i\right)\prod_{i=1}^{n} e^{\alpha(\sqrt{\tau}S_i - T_i)}\right] \\
&\le \frac{1}{2n\sqrt{\tau}}\mathbb{E}\left[\sum_{i=1}^{n}\left(S_i e^{\alpha\sqrt{\tau}S_i}\prod_{j\ne i} e^{\alpha(\sqrt{\tau}S_j - T_j)}\right)\right],
\end{aligned}
$$

where the last inequality holds because $e^{-\alpha T_i} \le 1$. Since $\{S_i\}_{i=1}^n$ are i.i.d. and independent of $\{T_i\}_{i=1}^n$, we have

$$
\hat{l}'_n(\tau) \le \frac{1}{2\sqrt{\tau}}\mathbb{E}\left[S_1 e^{\alpha\sqrt{\tau}S_1}\right]\left(M_S(\alpha\sqrt{\tau})M_T(-\alpha)\right)^{n-1},
$$

where M_S and M_T are the moment-generating functions of S_i and T_i. Set $\alpha = 0.2$ and $\tau_2 = 0.01$, we can calculate each term one by one:

$$
\begin{aligned}
\mathbb{E}\left[S_1 e^{\alpha\sqrt{\tau_2}S_1}\right] &= \frac{1}{2}\sum_{i=1}^{2}\int_0^\infty s e^{\alpha\sqrt{\tau_2}s}\lambda_i e^{-\lambda_i s}ds = \frac{1}{2}\sum_{i=1}^{2}\frac{\lambda_i}{(\lambda_i - \alpha\sqrt{\tau_2})^2} < 8, \\
M_S(\alpha\sqrt{\tau_2}) &= \frac{1}{2}\sum_{i=1}^{2}\frac{\lambda_i}{\lambda_i - \alpha\sqrt{\tau_2}} < 1.13, \\
M_T(-\alpha) &= \frac{1}{2}\sum_{i=1}^{2}\frac{\lambda_i}{\lambda_i + \alpha} < 0.66.
\end{aligned}
$$

Combining all the results, we have $\hat{l}'_n(\tau_2) < 40 \times (0.75)^{n-1}$. Therefore, we have

$$
\sum_{n=1}^{\infty}\hat{l}'_n(\tau_2) = \hat{l}'_1(\tau_2) + \sum_{n=2}^{\infty}\hat{l}'_n(\tau_2) < 15 + 40\sum_{n=2}^{\infty}(0.75)^{n-1} = 135.
$$

As a conclusion, we have $\sum_{n=1}^{\infty}\hat{l}'_n(\tau_1) > 219 > 135 > \sum_{n=1}^{\infty}\hat{l}'_n(\tau_2)$ with $\tau_1 = 10^{-6} < 0.01 = \tau_2$, which violates the convexity of $\hat{L}_q(\tau)$. □

5.4. Proof of Theorem 6

LEMMA 1 (**Li and Pong 2018**, **Fatkhullin et al. 2025**). *Consider the composition problem (6). Let $\mathcal{X}$ be closed and convex, and let $\mathcal{U} := \varphi(\mathcal{X})$ denote the feasible region of the reformulated problem. Suppose that $\mathcal{U}$ is also closed and convex. Assume further that there exists a constant $\sigma_\varphi > 0$ such that*

$$
\|\varphi(x) - \varphi(y)\|_2 \ge \sigma_\varphi\|x - y\|_2, \qquad \forall x, y \in \mathcal{X}.
$$

Then, the following statements hold.

1. *If H is convex on $\mathcal{U}$ and $\mathcal{U}$ is bounded with diameter $D_{\mathcal{U}}$, then F satisfies the PŁK condition on $\mathcal{X}$ with exponent $\alpha = 1$ and constant $\sigma_\varphi/(2D_{\mathcal{U}})$.*

2. *If H is α_H-strongly convex on $\mathcal{U}$, then F satisfies the PŁK condition on $\mathcal{X}$ with exponent $\alpha = 2$ and constant $\alpha_H \sigma_\varphi^2$.*

Proof of Theorem 6 First consider reformulation (R1), which uses the change of variables $\hat{\lambda} = \sqrt{\lambda}$. Set $\varphi_1(\lambda, \mu) = (\sqrt{\lambda}, \mu)$. Then, $\mathcal{U}_1 = \varphi_1(\mathcal{X}) = \{(\hat{\lambda}, \mu) : \sqrt{\underline{\lambda}} \le \hat{\lambda} \le \sqrt{\bar{\lambda}},\ \underline{\mu} \le \mu \le \bar{\mu}\}$, which is closed and convex. Moreover, $\mathcal{U}_1$ is bounded with diameter $D_{\mathcal{U}_1} = \sqrt{(\sqrt{\bar{\lambda}} - \sqrt{\underline{\lambda}})^2 + (\bar{\mu} - \underline{\mu})^2}$. For any $(\lambda_1, \mu_1), (\lambda_2, \mu_2) \in \mathcal{X}$,

$$\|\varphi_1(\lambda_1, \mu_1) - \varphi_1(\lambda_2, \mu_2)\|_2^2 = (\sqrt{\lambda_1} - \sqrt{\lambda_2})^2 + (\mu_1 - \mu_2)^2.$$

Since

$$|\sqrt{\lambda_1} - \sqrt{\lambda_2}| = \frac{|\lambda_1 - \lambda_2|}{\sqrt{\lambda_1} + \sqrt{\lambda_2}} \ge \frac{|\lambda_1 - \lambda_2|}{2\sqrt{\bar{\lambda}}},$$

we obtain

$$\|\varphi_1(\lambda_1, \mu_1) - \varphi_1(\lambda_2, \mu_2)\|_2^2 \ge \frac{1}{4\bar{\lambda}}(\lambda_1 - \lambda_2)^2 + (\mu_1 - \mu_2)^2 \ge \min\left\{\frac{1}{4\bar{\lambda}}, 1\right\} \|(\lambda_1, \mu_1) - (\lambda_2, \mu_2)\|_2^2.$$

By Theorem 1, the objective of reformulation (R1) is convex under Assumption 2.1, and κ_1-strongly convex under Assumption 2.2. Therefore, Lemma 1 implies that the original problem (1) satisfies the PŁK condition with exponent $\alpha = 1$ and constant

$$\frac{\min\{1/(2\sqrt{\bar{\lambda}}), 1\}}{2\sqrt{(\sqrt{\bar{\lambda}} - \sqrt{\underline{\lambda}})^2 + (\bar{\mu} - \underline{\mu})^2}}$$

under Assumption 2.1, and satisfies the PŁK condition with exponent $\alpha = 2$ and constant $\kappa_1 \min\{1/(4\bar{\lambda}), 1\}$ under Assumption 2.2.

Next consider reformulation (R2), which uses the change of variables $\hat{\mu} = \mu^2$. Set $\varphi_2(\lambda, \mu) = (\lambda, \mu^2)$. Then, $\mathcal{U}_2 = \varphi_2(\mathcal{X}) = \{(\lambda, \hat{\mu}) : \underline{\lambda} \le \lambda \le \bar{\lambda},\ \underline{\mu}^2 \le \hat{\mu} \le \bar{\mu}^2\}$, which is closed and convex. Moreover, $\mathcal{U}_2$ is bounded with diameter $D_{\mathcal{U}_2} = \sqrt{(\bar{\lambda} - \underline{\lambda})^2 + (\bar{\mu}^2 - \underline{\mu}^2)^2}$. For any $(\lambda_1, \mu_1), (\lambda_2, \mu_2) \in \mathcal{X}$,

$$\|\varphi_2(\lambda_1, \mu_1) - \varphi_2(\lambda_2, \mu_2)\|_2^2 = (\lambda_1 - \lambda_2)^2 + (\mu_1^2 - \mu_2^2)^2.$$

Because

$$|\mu_1^2 - \mu_2^2| = |\mu_1 - \mu_2|\,|\mu_1 + \mu_2| \ge 2\underline{\mu}\,|\mu_1 - \mu_2|,$$

it follows that

$$\|\varphi_2(\lambda_1, \mu_1) - \varphi_2(\lambda_2, \mu_2)\|_2^2 \ge (\lambda_1 - \lambda_2)^2 + 4\underline{\mu}^2(\mu_1 - \mu_2)^2 \ge \min\{1, 4\underline{\mu}^2\}\|(\lambda_1, \mu_1) - (\lambda_2, \mu_2)\|_2^2.$$

By Theorem 3, the objective of reformulation (R2) is convex under Assumption 3.1, and κ_2-strongly convex under Assumption 3.2. Applying Lemma 1, we conclude that the original problem (1) satisfies the PŁK condition with exponent $\alpha = 1$ and constant

$$\frac{\min\{1, 2\underline{\mu}\}}{2\sqrt{(\bar{\lambda} - \underline{\lambda})^2 + (\bar{\mu}^2 - \underline{\mu}^2)^2}}$$

under Assumption 3.1, and satisfies the PŁK condition with exponent $\alpha = 2$ and constant $\kappa_2 \min\{1, 4\underline{\mu}^2\}$ under Assumption 3.2. This completes the proof. $\square$

6. Conclusion

We study the joint control of arrival and service rates in queueing systems, with the objective of minimizing the long-run expected cost minus revenue. Despite the non-convexity of the objective function in the original decision variables, an appropriate reparameterization yields a convex reformulation. Leveraging this hidden convexity, we show that the original problem satisfies the PŁK condition, which helps explain the observed empirical success of first-order methods under standard assumptions. Our results apply to a broad class of $GI/GI/1$ models, including $\Gamma_k/\Gamma_m/1$ queues with $k,m \geq 1$ and $GI/M/1$ queues with log-concave interarrival times. A key ingredient in our analysis is a new convexity property showing that the expected queue length remains convex after a square-root transformation of the traffic intensity.

Our work opens several directions for future research. A fundamental question is to identify the general characteristics of the interarrival and service time distributions that guarantee hidden convexity. Moving beyond the Gamma family or $GI/M/1$ queues to characterize this property for more general systems would represent a major theoretical advancement. In addition, our findings suggest two paths forward for learning and control. For online learning of static policies, it would be interesting to design algorithms explicitly leveraging the PŁK condition. A more challenging frontier is dynamic pricing, where decisions are state-dependent. It remains an open question whether a similar benign optimization landscape exists in this setting.

Acknowledgments

We thank Jie Wang for bringing to our attention the empirical observation of global convergence for first-order methods.

References

Abramowitz M, Stegun IA (1965) *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables* (New York: Dover Publications).

Agrawal A, Boyd S (2020) Differentiating through log-log convex programs. *arXiv preprint arXiv:2004.12553* .

Ata B, Shneorson S (2006) Dynamic control of an M/M/1 service system with adjustable arrival and service rates. *Management Science* 52(11):1778–1791.

Bento G, Mordukhovich B, Mota T, Nesterov Y (2025) Convergence of descent optimization algorithms under polyak-łojasiewicz-kurdyka conditions. *Journal of Optimization Theory and Applications* 207(3):41.

Besbes O, Zeevi A (2015) On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science* 61(4):723–739.

Bhandari J, Russo D (2024) Global optimality guarantees for policy gradient methods. *Operations Research* 72(5):1906–1927.

Chen C, Chen Y, Qian P (2025) Incentivizing resource pooling. *Management Science* Forthcoming.

Chen X, Hong G (2023) Steady-state analysis and online learning for queues with hawkes arrivals. *arXiv preprint arXiv:2311.02577* .

Chen X, Hu Y, Zhao M (2024a) Landscape of policy optimization for finite horizon mdps with general state and action. *arXiv preprint arXiv:2409.17138* .

Chen X, Liu Y, Hong G (2023) Online learning and optimization for queues with unknown demand curve and service distribution. *arXiv preprint arXiv:2303.03399* .

Chen X, Liu Y, Hong G (2024b) An online learning approach to dynamic pricing and capacity sizing in service systems. *Operations Research* 72(6):2677–2697.

Dieker A, Ghosh S, Squillante MS (2017) Optimal resource capacity management for stochastic networks. *Operations Research* 65(1):221–241.

Fatkhullin I, He N, Hu Y (2025) Stochastic optimization under hidden convexity. *SIAM Journal on Optimization* 35(4):2544–2571.

Fridgeirsdottir K, Chiu S (2005) A note on convexity of the expected delay cost in single-server queues. *Operations Research* 53(3):568–570.

Gradshteyn IS, Ryzhik IM (2014) *Table of Integrals, Series, and Products* (Academic Press), eighth edition.

Harel A (1990) Convexity properties of the Erlang loss formula. *Operations Research* 38(3):499–505.

Harel A, Zipkin P (1987) Strong convexity results for queueing systems. *Operations Research* 35(3):405–418.

Hu Y, Wang J, Chen X, He N (2024) Multi-level Monte-Carlo gradient methods for stochastic optimization with biased oracles. *arXiv preprint arXiv:2408.11084* .

Jackson JR (1957) Networks of waiting lines. *Operations Research* 5(4):518–521.

Jackson JR (1963) Jobshop-like queueing systems. *Management Science* 10(1):131–142.

Kanoria Y, Qian P (2024) Blind dynamic resource allocation in closed networks via mirror backpressure. *Management Science* 70(8):5445–5462.

Karimi H, Nutini J, Schmidt M (2016) Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. *Joint European conference on machine learning and knowledge discovery in databases*, 795–811 (Springer).

Kim J, Randhawa RS (2018) The value of dynamic pricing in large queueing systems. *Operations Research* 66(2):409–425.

Kingman J (1962) Some inequalities for the queue GI/G/1. *Biometrika* 49(3/4):315–324.

Kleinrock L (1964) *Communication Nets: Stochastic Message Flow and Delay* (New York: McGraw-Hill).

Kurdyka K (1998) On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier*, volume 48, 769–783.

Lee C, Ward AR (2014) Optimal pricing and capacity sizing for the GI/GI/1 queue. *Operations Research Letters* 42(8):527–531.

Lee C, Ward AR (2019) Pricing and capacity sizing of a service facility: Customer abandonment effects. *Production and Operations Management* 28(8):2031–2043.

Li G, Pong TK (2018) Calculus of the exponent of kurdyka–łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of Computational Mathematics* 18(5):1199–1232.

Lindley DV (1952) The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society* 48(2):277–289.

Little JD (1961) A proof for the queuing formula: L= λ W. *Operations Research* 9(3):383–387.

Maglaras C, Zeevi A (2003) Pricing and capacity sizing for systems with shared resources: Approximate solutions and scaling relations. *Management Science* 49(8):1018–1038.

Mendelson H (1985) Pricing computer services: Queueing effects. *Communications of the ACM* 28(3):312–321.

Pittel B (1979) Closed exponential networks of queues with saturation: The jackson-type stationary distribution and its asymptotic analysis. *Mathematics of Operations Research* 4(4):357–378.

Polyak BT, et al. (1963) Gradient methods for minimizing functionals. *Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki* 3(4):643–653.

Shanthikumar JG, Ding S, Zhang MT (2007) Queueing theory for semiconductor manufacturing systems: A survey and open problems. *IEEE Transactions on Automation Science and Engineering* 4(4):513–522.

Talluri KT, Van Ryzin GJ (2005) *The Theory and Practice of Revenue Management* (Springer New York, NY).

Walton NS (2009) Proportional fairness and its relationship with multi-class queueing networks. *The Annals of Applied Probability* 19(6):2301–2333.

Weber RR (1980) Note—on the marginal benefit of adding servers to G/GI/m queues. *Management Science* 26(9):946–951.

Weber RR (1983) A note on waiting times in single server queues. *Operations Research* 31(5):950–951.

Wein LM (1989) Capacity allocation in generalized jackson networks. *Operations Research Letters* 8(3):143–146.

Whitt W (2015) Stabilizing performance in a single-server queue with time-varying arrival rate. *Queueing Systems* 81:341–378.

Whitt W (2018) Time-varying queues. *Queueing Models and Service Management* 1(2).

Zhong Y, Birge JR, Ward AR (2024) Learning to schedule in multiclass many-server queues with abandonment. *Operations Research* 73(6):3085–3103.

Łojasiewicz S (1963) A topological property of real analytic subsets. *Coll. du CNRS, Les équations aux dérivées partielles* 117(87-89):2.