



# The M/M/1+M queue with a utility-maximizing server

Dongyuan Zhan<sup>a,\*</sup>, Amy R. Ward<sup>b</sup>

<sup>a</sup> School of Management, University College London, London E14 5AB, United Kingdom

<sup>b</sup> Booth School of Business, University of Chicago, Chicago, IL 60637, United States

## ARTICLE INFO

### Article history:

Received 9 May 2018

Received in revised form 26 July 2018

Accepted 23 August 2018

Available online 28 August 2018

### Keywords:

M/M/1+M

Utilization

Server utility

Endogenous service rate

## ABSTRACT

We consider an M/M/1+M queue with a human server, who is influenced by incentives. Specifically, the server chooses his service rate by maximizing his utility function. Our objective is to guarantee the existence of a unique maximum. The complication is that most sensible utility functions depend on the server utilization, a non-simple expression. We derive a property of the utilization that guarantees quasiconcavity of any utility function that multiplies the server's concave (including linear) "value" of his service rate by the server utilization.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Traditional single-server queueing models assume the service rate is exogenous. This is sensible when the server is a machine. However, when the server is human, the balance between payment and required effort determines the resulting service rate. We model this by assuming the server has a utility function, and that the service rate emerges when the server maximizes his utility function.

The relevant question is: When does there exist a unique service rate? We answer this question in the context of a single-server queue with abandonment. Customers arrive to the system according to a Poisson process at rate  $\lambda$ , and the maximum amount of time each customer is willing to wait for service is an independent random sample from an exponential distribution with mean  $1/\theta$ . Any customer whose wait exceeds that maximum abandons the system without receiving service. Service times are independent and exponential, and, for a given utility function  $U$ , have rate

$$\hat{\mu} := \operatorname{argmax}_{\mu \in [\underline{\mu}, \bar{\mu}]} U(\mu) \quad (1)$$

when the maximum is unique, where  $U$  is the utility function, and  $0 < \underline{\mu} \leq \bar{\mu} < \infty$  are the lower and upper bounds on the possible service rate. In summary, we consider an M/M/1+M queue except that the service rate is determined as in (1).

The utility function is assumed to have the form

$$U(\mu) := V(\mu) \times B(\mu) \geq 0,$$

\* Corresponding author.

E-mail addresses: [d.zhan@ucl.ac.uk](mailto:d.zhan@ucl.ac.uk) (D. Zhan), [amy.ward@chicagobooth.edu](mailto:amy.ward@chicagobooth.edu) (A.R. Ward).

where the non-negative function  $V$  is the server's value of completing  $\mu$  tasks per time unit and

$$B(\mu) := \frac{\sum_{i=1}^{\infty} a_i(\mu)}{1 + \sum_{i=1}^{\infty} a_i(\mu)} \text{ for } a_i(\mu) := \prod_{k=0}^{i-1} \frac{\lambda}{\mu + k\theta}$$

is the utilization function, which can be computed in a straightforward manner because the M/M/1+M queue is a birth and death process. The function  $V$  could model a piece-rate payment to the server, as in [11], in which case  $V(\mu) = p\mu$ , where  $p \geq 0$  is the payment per finished service. Or, the function  $V$  could represent a piece-rate payment that additionally accounts for quality, as in [12]. As a final example, the function  $V$  could trade off monetary payment and effort. Then,  $V(\mu) = M(\mu) - c(\mu)$ , where  $M$  is some kind of piece-rate payment function and  $c$  represents the per-unit-busy-time cost. The function  $c$  would model the decrease in utility caused by increased effort, as is common in labor economics (see [3]).

In this paper, we prove the following result.

**Theorem 1.** *If  $V$  is a concave (including linear) function in  $\mu$ , then there exists a unique utility-maximizing service rate  $\hat{\mu}$ .*

The key to the proof of Theorem 1 is the following property of the utilization function in an M/M/1+M queue, that may be helpful in other endogenous service rate queueing models.

### Lemma 1.

$$2(B'(\mu))^2 - B(\mu)B''(\mu) > 0.$$

The remainder of this paper is organized as follows. We end this section with a brief literature review. Section 2 proves

**Theorem 1** under the assumption that **Lemma 1** holds. Section 3 proves **Lemma 1**.

## Literature review

The service rate that maximizes the server's utility is solution to a queueing game. There is a large literature on the queueing games, well surveyed by [9] and [8]. Much of that literature assumes a fixed service rate (or rates in the case of a many-server model), and focuses on how utility-maximizing customer decisions (such as whether or not to queue, and behind which server to queue) affect system performance. Some exceptions (that is, papers that allow the service rates to be the solution to a game) are [10,7,1,2,5,6]. A main challenge in all these papers is to guarantee the existence and uniqueness of the equilibrium service rate(s), for which the assumptions on the server utility function are critical. However, none of the aforementioned papers allow for customer abandonment.

## 2. The utility-maximizing service rate

The server wants to maximize his utility

$$\max_{\mu \in [\underline{\mu}, \bar{\mu}]} U(\mu) = V(\mu)B(\mu).$$

The issue is to know when the above maximization leads to a unique service rate  $\mu$ . If there is non-uniqueness, then  $\hat{\mu}$  in (1) is not well-defined. As a consequence, system performance cannot be predicted through the use of M/M/1+M queueing formulae. **Theorem 1** provides a sufficient condition for the maximum in (1) to be unique, implying that the service rate in (1) is well-defined.

**Proof of Theorem 1.** The first order condition of the utility maximization is

$$U'(\mu) = V'(\mu)B(\mu) + V(\mu)B'(\mu) = 0.$$

The second order derivative is

$$\begin{aligned} U''(\mu) &= V''(\mu)B(\mu) + 2V'(\mu)B'(\mu) + V(\mu)B''(\mu) \\ &= V''(\mu)B(\mu) + \frac{V(\mu)}{B(\mu)} (B(\mu)B''(\mu) - 2(B'(\mu))^2) \\ &\quad + 2U'(\mu) \frac{B'(\mu)}{B(\mu)}. \end{aligned}$$

From the concavity of  $V$  we know  $V'' \leq 0$  so the first term is non-positive. From **Lemma 1**,  $2(B'(\mu))^2 - B(\mu)B''(\mu) > 0$ , and the second term is negative. At a stationary point satisfying  $U'(\mu) = 0$ , the third term is zero, and so  $U''(\mu) < 0$ . Hence, any stationary point is a local maximum. Since any stationary point must be a local maximum, there can exist at most one stationary point. In summary,  $U'(\mu)$  has at most one zero point, which is a local maximum, implying quasiconcavity. ■

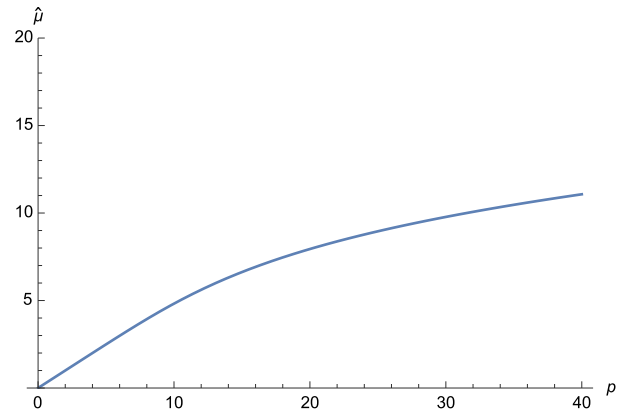
We end this section with an example in which there are piece-rate payment and a convex effort cost; specifically, assume

$$V(\mu) = p\mu - c\mu^r \text{ for } \mu \geq 0, c > 0, r > 1.$$

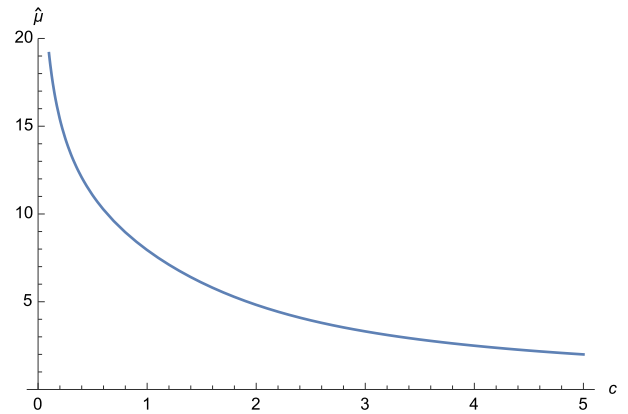
Since  $V$  is concave, **Theorem 1** guarantees

$$\hat{\mu} = \operatorname{argmax}_{\mu \in [0, \infty]} (p\mu - c\mu^r)B(\mu)$$

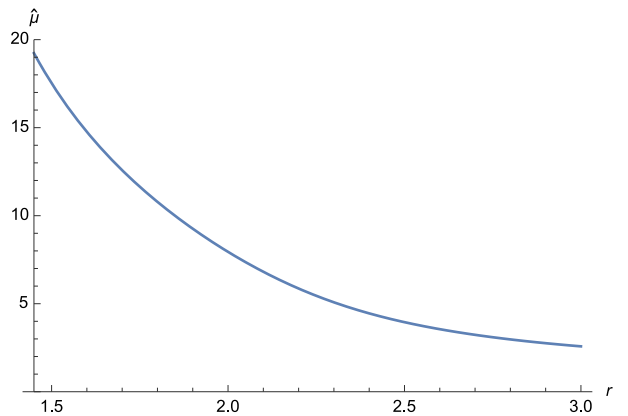
is unique. Assuming  $\lambda = 10$ ,  $\theta = 1$ , **Fig. 1** solves for  $\hat{\mu}$  numerically by changing one parameter among  $p$ ,  $c$ ,  $r$  and fixing the other two. Consistent with intuition, the service rate increases as the piece-rate incentive  $p$  increases, and decreases as the effort cost coefficient  $c$  or  $r$  increases.



(a)  $c = 1, r = 2$



(b)  $p = 20, r = 2$



(c)  $p = 20, c = 1$

**Fig. 1.** The service rate as functions of the piece-rate incentive and effort cost.

## 3. Proof of Lemma 1

We prove **Lemma 1** in three steps:

- Step 1: We show  $B''(\mu)B(\mu) - 2(B'(\mu))^2 < 0$  is equivalent to

$$2 \left( \sum_{i=1}^{\infty} a_i b_i \right)^2 - \sum_{i=1}^{\infty} a_i \sum_{i=1}^{\infty} a_i (b_i^2 + c_i) > 0, \quad (2)$$

for  $a_i$  as defined earlier and

$$b_i = \sum_{k=0}^{i-1} \frac{1}{\mu + k\theta}, \quad c_i = \sum_{k=0}^{i-1} \frac{1}{(\mu + k\theta)^2}, \quad i \in \mathcal{N}^+,$$

where  $\mathcal{N}^+ := \{1, 2, \dots\}$ .

- Step 2: We rewrite (2) as

$$\sum_{m=2}^{\infty} \sum_{j=1}^{m-1} a_{m-j} a_j (2b_{m-j} b_j - b_j^2 - c_j).$$

Then, for each  $m \geq 2$ , we show that

$$\begin{aligned} & \sum_{j=1}^{m-1} a_{m-j} a_j (2b_{m-j} b_j - b_j^2 - c_j) \\ & \geq a_{m-J(m)} a_{J(m)} \sum_{j=1}^{m-1} 2b_{m-j} b_j - b_j^2 - c_j, \end{aligned}$$

where  $J(m) \in \{1, \dots, \lfloor m/2 \rfloor\}$  will be defined in that step.

- Step 3: We prove that for any  $m \geq 2$ ,

$$\sum_{j=1}^{m-1} 2b_{m-j} b_j - b_j^2 - c_j > 0,$$

which completes the proof.

### Step 1:

We regard  $a_i$  and  $b_i$  as functions of  $\mu$ , and write  $a'_i(\mu)$  and  $b'_i(\mu)$  for their derivatives. For any  $i \in \mathcal{N}^+$ , we have

$$a'_i(\mu) = -a_i(\mu) \sum_{k=0}^{i-1} \frac{1}{\mu + k\theta} = -a_i(\mu) b_i(\mu),$$

$$b'_i(\mu) = -\sum_{k=0}^{i-1} \frac{1}{(\mu + k\theta)^2} = -c_i(\mu).$$

Assuming the interchange of summation and derivative,

$$\begin{aligned} B'(\mu) &= \left(1 - \frac{1}{1 + \sum_{i=1}^{\infty} a_i(\mu)}\right)' = \frac{\sum_{i=1}^{\infty} a'_i(\mu)}{(1 + \sum_{i=1}^{\infty} a_i(\mu))^2} \\ &= -(1 - B(\mu))^2 \sum_{i=1}^{\infty} a_i(\mu) b_i(\mu), \end{aligned}$$

and, using the above expression,

$$\begin{aligned} B''(\mu) &= 2(1 - B(\mu)) B'(\mu) \sum_{i=1}^{\infty} a_i(\mu) b_i(\mu) \\ &\quad - (1 - B(\mu))^2 \left( \sum_{i=1}^{\infty} a_i(\mu) b_i(\mu) \right)' \\ &= -2(1 - B(\mu))^3 \left( \sum_{i=1}^{\infty} a_i(\mu) b_i(\mu) \right)^2 \\ &\quad - (1 - B(\mu))^2 \sum_{i=1}^{\infty} (a'_i(\mu) b_i(\mu) + a_i(\mu) b'_i(\mu)) \\ &= -2(1 - B(\mu))^3 \left( \sum_{i=1}^{\infty} a_i(\mu) b_i(\mu) \right)^2 \\ &\quad + (1 - B(\mu))^2 \sum_{i=1}^{\infty} a_i(\mu) (b_i^2(\mu) + c_i(\mu)). \end{aligned}$$

Substituting for the above two expressions shows

$$\begin{aligned} B''(\mu) B(\mu) - 2(B'(\mu))^2 &= -(1 - B(\mu))^3 \times \\ &\quad \left( 2 \left( \sum_{i=1}^{\infty} a_i(\mu) b_i(\mu) \right)^2 - \sum_{i=1}^{\infty} a_i(\mu) \sum_{i=1}^{\infty} a_i(\mu) \right. \\ &\quad \times (b_i(\mu)^2 + c_i(\mu)) \left. \right). \end{aligned}$$

Since  $1 - B(\mu) > 0$  for all  $\mu \in [\underline{\mu}, \bar{\mu}]$ , showing  $B''(\mu) B(\mu) - 2(B'(\mu))^2 < 0$  is equivalent to showing (2).

Finally, the two interchanges of summation and derivative we must justify are:

$$(i) \left( \sum_{i=1}^{\infty} a_i(\mu) \right)' = \sum_{i=1}^{\infty} a'_i(\mu),$$

$$(ii) \left( \sum_{i=1}^{\infty} a_i(\mu) b_i(\mu) \right)' = \sum_{i=1}^{\infty} a'_i(\mu) b_i(\mu) + a_i(\mu) b'_i(\mu).$$

To show (i), first observe it is straightforward to see  $\sum_{i=1}^{\infty} a_i(\mu) < \infty$  (by, for example, the ratio test), and that  $a_i(\mu)$  is continuous in  $\mu$  for each  $i \in \mathcal{N}^+$ . Then, to establish (i) (see, for example, Theorem 2.4.3 in [4]), we must show  $\sum_{i=1}^{\infty} a'_i(\mu)$  converges uniformly on every closed bounded subinterval of  $[\underline{\mu}, \bar{\mu}]$ . Since  $|a'_i(\mu)| = a_i(\mu) b_i(\mu)$  and  $b_i < i/\mu$  for every  $i \in \mathcal{N}^+$ , if we define  $m_i := m \left( \frac{\lambda}{\theta} \right)^{i-1} \frac{i}{(i-1)!}$  for any finite  $m > \frac{\lambda}{\mu^2}$ , then

$$|a'_i(\mu)| < \frac{i}{\mu} \prod_{k=0}^{i-1} \frac{\lambda}{\mu + k\theta} = \frac{i\lambda}{\mu^2} \prod_{k=1}^{i-1} \frac{\lambda}{\mu + k\theta} < m_i,$$

for all  $\mu \in [\underline{\mu}, \bar{\mu}]$ .

The Weierstrass M-test implies the uniform convergence holds if  $\sum_{i=1}^{\infty} m_i < \infty$ , which is true by the ratio test

$$\lim_{i \rightarrow \infty} \left| \frac{m_{i+1}}{m_i} \right| = \lim_{i \rightarrow \infty} \frac{\lambda}{\theta} \frac{i+1}{i^2} = 0.$$

Similarly, if we re-define  $m_i := \frac{(i^2+i)\lambda}{(i-1)!\mu^3} \left( \frac{\lambda}{\theta} \right)^{i-1}$  for any  $i \in \mathcal{N}^+$ , then

$$\begin{aligned} |(a_i(\mu) b_i(\mu))'| &= |a'_i(\mu) b_i(\mu) + a_i(\mu) b'_i(\mu)| = a_i(\mu) (b_i^2(\mu) + c_i(\mu)) \\ &< \prod_{k=0}^{i-1} \frac{\lambda}{\mu + k\theta} \left( \frac{i^2}{\mu^2} + \frac{i}{\mu^2} \right) \\ &= \frac{(i^2+i)\lambda}{\mu^3} \prod_{k=1}^{i-1} \frac{\lambda}{\mu + k\theta} < m_i. \end{aligned}$$

The Weierstrass M-test implies  $\sum_{i=1}^{\infty} (a_i(\mu) b_i(\mu))'$  converges uniformly on every closed bounded subinterval of  $[\underline{\mu}, \bar{\mu}]$  if  $\sum_{i=1}^{\infty} m_i < \infty$ , which is true by the ratio test

$$\lim_{i \rightarrow \infty} \left| \frac{m_{i+1}}{m_i} \right| = \lim_{i \rightarrow \infty} \frac{\lambda}{\theta} \frac{i+2}{i^2} = 0.$$

### Step 2:

We drop the argument  $\mu$  in  $a_i, b_i, c_i$  for  $i \in \{1, 2, \dots\}$  for notational convenience. Then from algebra and re-arranging the order of summation terms,

$$\begin{aligned} & 2 \left( \sum_{i=1}^{\infty} a_i b_i \right)^2 - \sum_{i=1}^{\infty} a_i \sum_{i=1}^{\infty} a_i (b_i^2 + c_i) \\ &= 2 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i b_i a_j b_j - \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i a_j (b_j^2 + c_j) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i a_j (2b_i b_j - b_j^2 - c_j) \\
&= \sum_{m=2}^{\infty} \sum_{j=1}^{m-1} a_{m-j} a_j (2b_{m-j} b_j - b_j^2 - c_j).
\end{aligned}$$

To identify a lower bound for the inner summation term, it is helpful to observe the following:

(i) First, for  $\Delta(m, j) := 4b_{m-j}b_j - b_j^2 - b_{m-j}^2 - c_j - c_{m-j}$ ,

$$\begin{aligned}
&\sum_{j=1}^{m-1} a_{m-j} a_j (2b_{m-j} b_j - b_j^2 - c_j) \\
&= \sum_{j=1}^{\lfloor (m-1)/2 \rfloor} a_{m-j} a_j \Delta(m, j) + \frac{\Delta(m, m/2)}{2} \mathbf{1}\{m \text{ is even}\};
\end{aligned}$$

(ii) Second,  $a_j a_{m-j}$  is increasing for  $j \in \{1, \dots, \lfloor m/2 \rfloor\}$  because

$$\begin{aligned}
\frac{a_{m-j-1} a_{j+1}}{a_{m-j} a_j} &= \frac{\frac{\lambda}{\mu+j\theta}}{\frac{\lambda}{\mu+(m-j-1)\theta}} \\
&= \frac{\mu + (m-j-1)\theta}{\mu + j\theta} > 1 \text{ for } j \in \{1, \dots, \lfloor m/2 \rfloor - 1\}.
\end{aligned}$$

Suppose we can show that for every positive integer  $m \geq 2$ , there exists an integer  $J(m)$  such that  $\Delta(m, j) < 0$  for  $j \in \{1, \dots, J(m) - 1\}$  and  $\Delta(m, j) \geq 0$  for  $j \in \{J(m), \dots, \lfloor m/2 \rfloor\}$ . Then, also using (i) and (ii) above,

$$a_{m-j} a_j \Delta(m, j) \geq a_{m-J(m)} a_{J(m)} \Delta(m, j) \text{ for all } j \in \{1, \dots, \lfloor m/2 \rfloor\}.$$

so that

$$\begin{aligned}
&\sum_{j=1}^{m-1} a_{m-j} a_j (2b_{m-j} b_j - b_j^2 - c_j) \\
&\geq a_{m-J(m)} a_{J(m)} \left( \sum_{j=1}^{\lfloor (m-1)/2 \rfloor} \Delta(m, j) + \frac{\Delta(m, m/2)}{2} \mathbf{1}\{m \text{ is even}\} \right) \\
&= a_{m-J(m)} a_{J(m)} \sum_{j=1}^{m-1} (2b_{m-j} b_j - b_j^2 - c_j).
\end{aligned}$$

Note that the above equality is still valid if  $\Delta(m, j) > 0$  for all  $j \in \{1, \dots, \lfloor m/2 \rfloor\}$  ( $J(m) = 1$ ) or if  $\Delta(m, j) \leq 0$  for all  $j \in \{1, \dots, \lfloor m/2 \rfloor\}$  ( $J(m) = \lfloor m/2 \rfloor$ ).

Now we justify the existence of  $J(m)$  for any  $m \geq 2$  by showing the monotonicity of  $\Delta(m, j)$  in  $j$  when  $j \in \{1, \dots, \lfloor m/2 \rfloor\}$ . Specifically, for  $j \in \{1, \dots, \lfloor m/2 \rfloor - 1\}$ ,

$$\begin{aligned}
&\Delta(m, j) - \Delta(m, j+1) \\
&= 4(b_{m-j}b_j - b_{m-j-1}b_{j+1}) + (b_{j+1}^2 - b_j^2) - (b_{m-j}^2 - b_{m-j-1}^2) \\
&\quad + (c_{j+1} - c_j) - (c_{m-j} - c_{m-j-1}) \\
&= 4 \left( -\frac{b_{m-j-1}}{\mu + j\theta} + \frac{b_j}{\mu + (m-j-1)\theta} \right) + \left( \frac{2b_{j+1}}{\mu + j\theta} - \frac{1}{(\mu + j\theta)^2} \right) \\
&\quad - \left( \frac{2b_{m-j}}{\mu + (m-j-1)\theta} - \frac{1}{(\mu + (m-j-1)\theta)^2} \right) \\
&\quad + \frac{1}{(\mu + j\theta)^2} - \frac{1}{(\mu + (m-j-1)\theta)^2} \\
&= 2 \left( -\frac{b_{m-j-1}}{\mu + j\theta} + \frac{b_j}{\mu + (m-j-1)\theta} \right) \\
&\quad - \frac{2(b_{m-j-1} - b_{j+1})}{\mu + j\theta} - \frac{2(b_{m-j} - b_j)}{\mu + (m-j-1)\theta}.
\end{aligned}$$

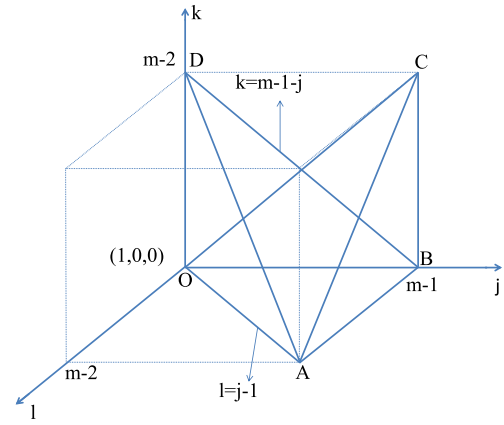


Fig. 2. The equivalence of two summations.

Since  $j < \lfloor m/2 \rfloor$ , we have  $j+1 \leq m-j-1$ . Therefore,  $\frac{1}{\mu+j\theta} > \frac{1}{\mu+(m-j-1)\theta}$ ,  $b_{j+1} \leq b_{m-j-1}$  and  $b_j < b_{m-j-1} < b_{m-j}$ . The above is negative.

**Step 3:**

From the definition of  $b_j, j \in \mathcal{N}^+$ ,

$$\begin{aligned}
\sum_{j=1}^{m-1} b_{m-j} b_j &= \sum_{j=1}^{m-1} \sum_{k=0}^{m-j-1} \frac{1}{\mu + k\theta} \sum_{l=0}^{j-1} \frac{1}{\mu + l\theta} \\
&= \sum_{j=1}^{m-1} \sum_{k+l < j} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta}.
\end{aligned}$$

The second equality, which is equivalent to

$$\sum_{j=1}^{m-1} \sum_{l=0}^{j-1} \sum_{k=0}^{m-j-1} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} = \sum_{j=1}^{m-1} \sum_{l=0}^{j-1} \sum_{k=0}^{j-1-l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta},$$

can be shown through a geometric argument. Define  $f(k, l, j) := \frac{1}{\mu+k\theta} \frac{1}{\mu+l\theta}$ . Then  $\sum_{j=1}^{m-1} \sum_{l=0}^{j-1} \sum_{k=0}^{m-j-1} f(k, l, j)$  corresponds to triangular pyramid OABD in Fig. 2;  $\sum_{j=1}^{m-1} \sum_{l=0}^{j-1} \sum_{k=0}^{j-1-l} f(k, l, j)$  corresponds to triangular pyramid OABC. OABD is equivalent to OABC if we map the points in the direction of  $j$  axis since the value of  $f(k, l, j)$  is independent of  $j$ .

Next, algebra gives

$$\begin{aligned}
&\sum_{j=1}^{m-1} 2b_{m-j} b_j - b_j^2 - c_j \\
&= \sum_{j=1}^{m-1} \left( 2 \sum_{k+l < j} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} - \left( \sum_{k=0}^{j-1} \frac{1}{\mu + k\theta} \right)^2 \right. \\
&\quad \left. - \sum_{k=0}^{j-1} \frac{1}{(\mu + k\theta)^2} \right).
\end{aligned}$$

To complete the proof, it is sufficient to show each term in the above summation is positive. This follows from the following sequence of equalities

$$\begin{aligned}
&2 \sum_{k+l < j} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} - \left( \sum_{k=0}^{j-1} \frac{1}{\mu + k\theta} \right)^2 - \sum_{k=0}^{j-1} \frac{1}{(\mu + k\theta)^2} \\
&= 2 \sum_{k+l < j} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} - \sum_{k=0}^{j-1} \sum_{l=0}^{j-1} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta}
\end{aligned}$$

$$\begin{aligned}
& - \sum_{k=0}^{j-1} \frac{1}{(\mu + k\theta)^2} \\
& = \sum_{k+l < j} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} - \sum_{k < j, l < j, k+l \geq j} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} \\
& \quad - \sum_{k=0}^{j-1} \frac{1}{(\mu + k\theta)^2} \\
& = \sum_{k+l \leq j-1, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} + \sum_{k < j/2} \frac{1}{(\mu + k\theta)^2} - \sum_{j/2 \leq k < j} \frac{1}{(\mu + k\theta)^2} \\
& \quad - \sum_{k < j, l < j, k+l \geq j, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} - \sum_{k=0}^{j-1} \frac{1}{(\mu + k\theta)^2} \\
& = \sum_{k+l \leq j-2, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} + \sum_{k+l=j-1, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} \\
& \quad - \sum_{k < j, l < j, k+l \geq j, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} - 2 \sum_{j/2 \leq k < j} \frac{1}{(\mu + k\theta)^2}.
\end{aligned}$$

For any pair  $(k, l)$  in the set  $\{(k, l) | 0 \leq k \leq j-1, 0 \leq l \leq j-1, k+l \leq j-2, k \neq l\}$ ,  $(j-1-k, j-1-l)$  is in the set  $\{(k, l) | 0 \leq k \leq j-1, 0 \leq l \leq j-1, k+l \geq j, k \neq l\}$ , and vice versa. Therefore,

$$\begin{aligned}
& \sum_{k+l \leq j-2, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} - \sum_{k < j, l < j, k+l \geq j, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} \\
& = \sum_{k+l \leq j-2, k \neq l} \left( \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} \right. \\
& \quad \left. - \frac{1}{\mu + (j-1-l)\theta} \frac{1}{\mu + (j-1-k)\theta} \right).
\end{aligned}$$

Note that  $j-1-k > l$  and  $j-1-l > k$ , hence the above is positive. Meanwhile,

$$\begin{aligned}
\sum_{k+l=j-1, k \neq l} \frac{1}{\mu + k\theta} \frac{1}{\mu + l\theta} & = \sum_{j/2 \leq k < j} \frac{2}{\mu + k\theta} \frac{1}{\mu + (j-1-k)\theta} \\
& > \sum_{j/2 \leq k < j} \frac{2}{(\mu + k\theta)^2}. \blacksquare
\end{aligned}$$

## References

- [1] G.P. Cachon, P.T. Harker, Competition and outsourcing with scale economies, *Manage. Sci.* 48 (10) (2002) 1314–1333.
- [2] G.P. Cachon, F. Zhang, Obtaining fast service in a queueing system via performance-based allocation of demand, *Manage. Sci.* 53 (3) (2007) 408–420.
- [3] P. Cahuc, A. Zylberberg, *Labor Economics*, MIT Press, 2004.
- [4] G. Casella, R.L. Berger, *Statistical Inference*, second ed., Duxbury Press, 2002.
- [5] L.G. Debo, L.B. Toktay, L.N. Van Wassenhove, Queuing for expert services, *Manage. Sci.* 54 (8) (2008) 1497–1512.
- [6] X. Geng, W.T. Huh, M. Nagarajan, Fairness among servers when capacity decisions are endogenous, *Prod. Oper. Manage.* 24 (6) (2015) 961–974.
- [7] S.M. Gilbert, Z.K. Weng, Incentive effects favor nonconsolidating queues in a service system: The principal–agent perspective, *Manage. Sci.* 44 (12) (1998) 1662–1669.
- [8] R. Hassin, *Rational Queueing*, CRC Press, 2016.
- [9] R. Hassin, M. Haviv, *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, Kluwer Academic Publishers, 2003.
- [10] E. Kalai, M.I. Kamien, M. Rubinfeld, Optimal service speeds in a competitive environment, *Manage. Sci.* 38 (8) (1992) 1154–1163.
- [11] Z.J. Ren, Y.P. Zhou, Call center outsourcing: Coordinating staffing level and service quality, *Manage. Sci.* 54 (2) (2008) 369–383.
- [12] D. Zhan, A.R. Ward, Incentive based service system design: Staffing and compensation to trade off speed and quality, 2017 (working paper).