Adv. Appl. Prob. **22,** 764–767 (1990)
Printed in N. Ireland
© *Applied Probability Trust* 1990

CONCAVITY OF THE THROUGHPUT OF TANDEM QUEUEING SYSTEMS WITH FINITE BUFFER STORAGE SPACE

LUDOLF E. MEESTER* AND
J. GEORGE SHANTHIKUMAR,** *University of California, Berkeley*

Abstract

We consider a tandem queueing system with m stages and finite intermediate buffer storage spaces. Each stage has a single server and the service times are independent and exponentially distributed. There is an unlimited supply of customers in front of the first stage. For this system we show that the number of customers departing from each of the m stages during the time interval $[0, t]$ for any $t \geqq 0$ is strongly stochastically increasing and concave in the buffer storage capacities. Consequently the throughput of this tandem queueing system is an increasing and concave function of the buffer storage capacities. We establish this result using a sample path recursion for the departure processes from the m stages of the tandem queueing system, that may be of independent interest. The concavity of the throughput is used along with the reversibility property of tandem queues to obtain the optimal buffer space allocation that maximizes the throughput for a three-stage tandem queue.

BLOCKING; SAMPLE PATH CONSTRUCTION; STOCHASTIC CONCAVITY; BUFFER ALLOCATION

1. Introduction and summary

Consider a tandem queueing system with m stages numbered $1, 2, \cdots, m$. Stage j has a single server and the service times are independent and exponentially distributed with mean $1/\mu_j$, $j = 1, 2, \cdots, m$. The buffer storage capacity between stages j and $j + 1$ is $b_j < +\infty$, $j = 1, 2, \cdots, m - 1$. There is an unlimited number of customers in front of the first stage (that is, the first stage is never starved) and the output buffer storage for stage m has an unlimited capacity (that is, $b_m = +\infty$ and stage m is never blocked). Customers require service from all stages in the order $1, 2, \cdots, m$ and the service to a customer at stage j is initiated only if the number of customers at stage $j + 1$ (that is the number of customers in the buffer storage j including the one at the server at stage $j + 1$, if any) is less than b_j, $j = 1, 2, \cdots, m - 1$. Service to a customer at any stage, once initiated, is completed without interruptions. Let $D_j(t, \boldsymbol{b})$ be the number of customers departing from stage j during the time interval $[0, t]$ and $D(t, \boldsymbol{b}) = (D_j(t, \boldsymbol{b}), j = 1, \cdots, m)$. Then we show the following (see Section 2).

(1.1) *Theorem.* $D_j(t, \boldsymbol{b})$ is strongly stochastically increasing and concave in $\boldsymbol{b}$ almost everywhere.

(1.2) *Definition.* A collection $\{X(\theta), \theta \in \Theta\}$ of random variables with a convex parameter set $\Theta \subset \mathbb{R}^m$ is said to be strongly stochastically increasing and concave in θ almost everywhere if there exists a collection $\{X'(\theta), \theta \in \Theta\}$ of random variables such that $X'(\theta)$ has the same distribution as $X(\theta)$ and it is increasing and concave in θ almost surely.

Received 16 October 1989; revision received 20 April 1990.

* Postal address: Department of Statistics, University of California, Berkeley, CA 94720, USA.

** Postal address: Walter A. Haas School of Business, University of California, Berkeley, CA 94720, USA.

https://doi.org/10.2307/1427472 Published online by Cambridge University Press

This definition of strong stochastic concavity is stronger than the stochastic concavity definition of Shaked and Shanthikumar (1988) and Shanthikumar and Yao (1991). Now let $TH(\boldsymbol{b})$ be the steady state throughput of this tandem queueing system. Then

$$TH(\boldsymbol{b}) = \lim_{t\to\infty} \frac{D_j(t, \boldsymbol{b})}{t}. \tag{1.3}$$

From Theorem (1.1) and (1.3) we have the following corollary.

(1.4) *Corollary.* $TH(\boldsymbol{b})$ is an increasing and concave function of $\boldsymbol{b}$.

Tandem queueing systems with finite buffer spaces serve as models for several manufacturing and communication systems. Consequently, considerable attention has been focused on the analysis of such systems (e.g. see the review paper by Perros (1986)). One of the main design issues addressed in these studies is the optimal allocation of buffer spaces (e.g. see Hillier et al. (1986)). The concavity of the throughput has remained a conjecture and often used in good faith in the empirical studies of optimal buffer space allocation problems. Corollary (1.4) now validates most of these empirical studies and allows the development of efficient optimization techniques to obtain solutions for the buffer space allocation problems. We illustrate this through a simple example. It has been conjectured that in a three-stage (that is $m = 3$) tandem queue with $\mu_1 = \mu_3$, the optimal buffer space allocation (b_1', b_2') that maximizes the throughput subject to $b_1 + b_2 = B$ should satisfy $|b_1' - b_2'| \leqq 1$. The following result provides an affirmative answer to this conjecture.

(1.5) *Theorem.* For $m = 3$ and $\mu_1 = \mu_3$, $TH(\boldsymbol{b})$ is Schur-concave in $\boldsymbol{b}$.

(1.6) *Remark.* Recall that a function $f: \mathbb{Z}^2 \to \mathbb{R}$ is Schur-concave if for any $\boldsymbol{n} \in \mathbb{Z}^2$, $n_1 > [<] n_2$ implies $f(n_1 - 1, n_2 + 1) \geqq [\leqq] f(n_1, n_2)$. It is then immediate that $\max \{f(\boldsymbol{n}): \boldsymbol{n} \in \mathbb{Z}_+^2,\ n_1 + n_2 = N\} = f(n')$ where $|n_1' - n_2'| \leqq 1$.

Proof of Theorem (1.5). From the duality results of Yamazaki and Sakasegawa (1975) it follows that $TH(\boldsymbol{b})$ is symmetric in $\boldsymbol{b}$ (that is, $TH(b_1, b_2) = TH(b_2, b_1)$). Then from Corollary (1.4) and Proposition 3.C.2 of Marshall and Olkin (1979) for discrete functions, the Schur-concavity of $TH(\boldsymbol{b})$ follows.

2. Concavity of the departure processes

In this section we prove Theorem (1.1). We do this by first constructing a process $\{D'(t, \boldsymbol{b}), t \geqq 0\}$, where $D'(t, \boldsymbol{b}) = (D_j'(t, \boldsymbol{b}),\ j = 1, \cdots, m)$, such that $\{D'(t, \boldsymbol{b}),\ t \geqq 0\} \overset{\text{st}}{=} \{D(t, \boldsymbol{b}),\ t \geqq 0\}$ and $D_j'(t, \boldsymbol{b})$ is increasing and concave in $\boldsymbol{b}$ for each $j = 1, \cdots, m$.

Suppose the number of customers at stage j at time 0 is r_j, $j = 1, \cdots, m$. By our earlier assumption, $r_1 = +\infty$. Then the number of customers at stage j at time t is $r_j + D_{j-1}(t) - D_j(t)$, $j = 2, \cdots, m$. Note that the number of customers at stage 1 is always equal to $+\infty$. Therefore it is not hard to see that $\{D(t, \boldsymbol{b}), t \geqq 0\}$ is a Markov process on the state space $S = \{\boldsymbol{d} \in \mathbb{Z}_+^m: -r_j \leqq d_{j-1} - d_j \leqq b_{j-1} - r_j,\ j = 2, \cdots, m\}$. The transition rate from state $\boldsymbol{d}$ to state $\boldsymbol{d} + \boldsymbol{e}_j$ is $\mu_j \cdot I\{d_{j-1} - d_j > -r_j\} \cdot I\{d_j - d_{j+1} < b_j - r_{j+1}\}$. Here $\boldsymbol{e}_j$ is the jth unit vector and $I\{\cdot\}$ is the indicator function. Next we construct the process $\{D'(t, \boldsymbol{b}),\ t \geqq 0\}$ on the state space S such that it is Markov with the same transition rates as that of $\{D(t, \boldsymbol{b}), t \geqq 0\}$.

Let $\{T_n, n = 1, 2, \cdots\}$ be the sequence of arrival epochs of a Poisson process with rate $\eta = \sum_{j=1}^m \mu_j$, $T_0 = 0$ and $\{U_n, n = 1, 2, \cdots\}$ be a sequence of i.i.d. uniform random variables on $(0, \eta)$ independent of the Poisson process. Define

$$Z_{j:n} = I\left\{\sum_{k=1}^{j-1} \mu_k \leqq U_n < \sum_{k=1}^{j} \mu_k\right\}, \qquad j = 1, \cdots, m;\ n = 1, 2, \cdots. \tag{2.1}$$

In the constructed process a service completion is allowed to take place only at time points $\{T_n, n = 1, 2, \cdots\}$. In particular, a service completion at stage j takes place at time T_n if $Z_{j:n} = 1$, there is at least one customer at that stage and the number of customers at stage

https://doi.org/10.2307/1427472 Published online by Cambridge University Press

$j+1$ is less than b_j. If $D'_{j:n}(\boldsymbol{b})$ is the number of departures from stage j for the constructed process during the time interval $[0, T_n]$, then we set

$$D'_{j:0}(\boldsymbol{b})=0, \qquad j=1,\cdots,m, \tag{2.2}$$

and

$$\begin{aligned} D'_{j:n+1}(\boldsymbol{b}) &= D'_{j:n}(\boldsymbol{b})+Z_{j:n}\cdot I\{D'_{j-1:n}(\boldsymbol{b})-D'_{j:n}(\boldsymbol{b})>-r_j\}\cdot I\{D'_{j:n}(\boldsymbol{b})-D'_{j+1:n}(\boldsymbol{b})<b_j-r_{j+1}\}, \\ & j=1,\cdots,m; \qquad n=0,1,2,\cdots, \end{aligned} \tag{2.3}$$

where $D'_{0:n}(\boldsymbol{b})=+\infty$, $D'_{m+1:n}(\boldsymbol{b})=+\infty$, $n=0,1,2,\cdots,b_m=+\infty$ and $r_{m+1}=0$ so that $I\{D'_{0:n}(\boldsymbol{b})-D'_{1:n}(\boldsymbol{b})>-r_1\}=1$ always (that is, stage 1 is never starved) and $I\{D'_{m:n}(\boldsymbol{b})-D'_{m+1:n}(\boldsymbol{b})<b_m-r_{m+1}\}=1$ always (that is, stage m is never blocked). Since no state change is supposed to take place during any time interval (T_n, T_{n+1}), $n=1,2,\cdots$, we set

$$D'_j(t,\boldsymbol{b})=D'_{j:n}(\boldsymbol{b}),\ T_n\leqq t<T_{n+1},\ j=1,\cdots,m;\ n=0,1,2,\cdots. \tag{2.4}$$

Clearly $\{D'(t,\boldsymbol{b}),\ t\geqq 0\}$ is a Markov process on the state space S. The transition rate of this process from state $\boldsymbol{d}$ to $\boldsymbol{d}+\boldsymbol{e}_j$ is $\eta\cdot P\{Z_{j:n+1}=1\}\cdot I\{d_{j-1}-d_j>-r_j\}\cdot I\{d_j-d_{j+1}<b_j-r_{j+1}\}=\mu_j\cdot I\{d_{j-1}-d_j>-r_j\}\cdot I\{d_j-d_{j+1}<b_j-r_{j+1}\}$, same as that for $\{D(t,\boldsymbol{b}),t\geqq 0\}$. Since the initial states of these processes are the same (that is, $d'(0,\boldsymbol{b})=D(0,\boldsymbol{b})=0$) it is clear that $\{D'(t,\boldsymbol{b}),t\geqq 0\}\overset{\text{st}}{=}\{D(t,\boldsymbol{b}),t\geqq 0\}$. We use the following result to prove Theorem (1.1).

(2.5) *Lemma.* For $n=0,1,2,\cdots$,

$$D'_{j:n+1}(\boldsymbol{b})=\min\{D'_{j-1:n}(\boldsymbol{b})+r_j,\ D'_{j:n}(\boldsymbol{b})+Z_{j:n+1},\ D'_{j+1:n}(\boldsymbol{b})+b_j-r_{j+1}\}, \qquad j=1,\cdots,m. \tag{2.6}$$

Proof. Since $D'_{j-1:n}(\boldsymbol{b})+r_j\geqq D'_{j:n}(\boldsymbol{b})$, $j=1,\cdots,m$; $n=0,1,2,\cdots$, one sees that $I\{D'_{j-1:n}(\boldsymbol{b})-D'_{j:n}(\boldsymbol{b})>-r_j\}=0\ [1]\Leftrightarrow D'_{j-1:n}(\boldsymbol{b})+r_j=[>]D'_{j:n}(\boldsymbol{b})$. Similarly since $D'_{j:n}(\boldsymbol{b})-D'_{j+1:n}(\boldsymbol{b})\leqq b_j-r_{j+1}$, $j=1,\cdots,m$; $n=0,1,2,\cdots$, one has $I\{D'_{j:n}(\boldsymbol{b})-D'_{j+1:n}(\boldsymbol{b})<b_j-r_{j+1}\}=0\ [1]\Leftrightarrow D'_{j:n}(\boldsymbol{b})=[<]D'_{j+1:n}(\boldsymbol{b})+b_j-r_{j+1}$. Then from (2.3) one sees that $D'_{j:n+1}(\boldsymbol{b})=D'_{j:n}(\boldsymbol{b})$ if either $D'_{j-1:n}(\boldsymbol{b})+r_j=D'_{j:n}(\boldsymbol{b})$ and/or $D'_{j:n}(\boldsymbol{b})=D'_{j+1:n}(\boldsymbol{b})+b_j-r_{j+1}$; otherwise $D'_{j:n+1}(\boldsymbol{b})=D'_{j:n}(\boldsymbol{b})+Z_{j:n+1}$. It is now easily seen that the values obtained from (2.6) for all these different cases are the same as above.

Proof of Theorem (1.1). From (2.6) it is easily seen that if $D'_{j:n}(\boldsymbol{b})$ is increasing and concave in $\boldsymbol{b}$ for every $j=1,\cdots,m$, then $D'_{j:n+1}(\boldsymbol{b})$ is also increasing and concave in $\boldsymbol{b}$ for all $j=1,\cdots,m$. Then by induction and the initial condition $D'_{j:0}(\boldsymbol{b})=0$, $j=1,\cdots,m$, it is immediate that $D'_{j:n}(\boldsymbol{b})$ is increasing and concave in $\boldsymbol{b}$ for every $j=1,\cdots,m$. The required result now follows from (2.4).

(2.7) *Remark.* The blocking mechanism we considered in this paper is such that the service is initiated only if there is room in the downstream buffer. This is called *communication* blocking. In an alternate blocking mechanism, called *production* blocking, a service is initiated even if the downstream buffer is full. In such a case a server is blocked only if the customer it served cannot be advanced to the next stage. The above sample path construction and the conclusions easily extend to this case and to several other blocking mechanisms.

(2.8) *Remark.* After this paper had been submitted for publication it was brought to our attention that Anantharam and Tsoucas (1990) had independently derived the componentwise concavity of the throughput of the exponential tandem queue with respect to the individual buffer size. They use a sample path construction similar to ours, but we have derived an explicit representation of the dynamics of the system which allowed us to obtain the stronger joint concavity result. In particular we have obtained the joint stochastic concavity of the number departures and the throughput with respect to the buffer capacities. We have therefore also answered the open question raised in Section 3 of Anantharam and Tsoucas (1990).

https://doi.org/10.2307/1427472 Published online by Cambridge University Press

References

ANANTHARAM, V. AND TSOUCAS, P. (1990) Stochastic concavity of throughput in series of queues with finite buffers *Adv. Appl. Prob.* **22,** 761–763.

HILLIER, F. S., BOLING, R. W. and SO, K. C. (1986) Toward characterizing the optimal allocation of storage space in production line systems with variable operation times. Technical report, Department of Operations Research, Stanford University, Stanford, CA.

MARSHALL, A. W. AND OLKIN, I. (1979) *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

PERROS, H. G. (1986) A survey of queueing networks with blocking, Technical report, Department of Computer Science, North Carolina State University, Raleigh, NC 27695.

SHAKED, M. AND SHANTHIKUMAR, J. G. (1988) Stochastic convexity and its applications. *Adv. Appl. Prob.* **20,** 427–446.

SHANTHIKUMAR, J. G. AND YAO, D. D. (1991) Strong stochastic convexity: Closure properties and applications. *J. Appl. Prob.* **28,** to appear.

YAMAZAKI, G. AND SAKASEGAWA, H. (1975) Properties of duality in tandem queueing systems. *Ann. Inst. Statist. Math.* **27,** 201–212.

https://doi.org/10.2307/1427472 Published online by Cambridge University Press