A New Ordering for Stochastic Majorization: Theory and Applications

Cheng-Shang Chang
IBM Research Division
T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

August, 1990
revised April, 1991

Abstract

In this paper, we develop a unified approach for stochastic load balancing on various multiserver systems. We expand the four partial orderings defined in Marshall and Olkin, by defining a new ordering based on the set of functions that are symmetric, L-subadditive and convex in each variable. This new partial ordering is shown to be equivalent to the previous four orderings for comparing deterministic vectors but differ for random vectors. Sample path criteria and a probability enumeration method for the new stochastic ordering are established and the ordering is applied to various fork-join queues, routing and scheduling problems. Our results generalize previous work and can be extended to multivariate stochastic majorization which includes tandem queues and queues with finite buffers.
Keywords: load balancing, stochastic convexity, majorization ordering, routing, scheduling, fork-join queues

1 Introduction

Recently, stochastic comparison has received a great deal of attention. Three methods are widely used: convex ordering, association of random variables and stochastic convexity. The basic ideas for applying these three methods are similar. Usually a mathematical model for a system consists of three parts: system parameters, system dynamics and performance measures. System parameters include initial conditions, input processes, service disciplines. System dynamics are the equations that govern system behavior, and examples of performance measures include expected response time and blocking probability. A standard approach used to establish a stochastic comparison between two systems consists of establishing that the system parameters are initially ordered in one of the orderings. Then, according to the system dynamics, one shows that performance measures are functions of system parameters. The classes of functions here could be convex functions, or the functions considered in this paper. Finally, one shows that the ordering is preserved under such functions. Therefore, functions that preserve these orderings are of importance. For the properties of convex ordering, we refer to Stoyan [34] and Ross [27]. Recent applications of convex ordering can be found in [25, 1, 2, 5, 6], among many others. For the properties of association of random variables, we refer to [12] and [3]. Recent applications can be found in [23, 22, 2], among many others. The results of stochastic convexity can be found in [37, 38, 28, 29, 32, 33, 8, 9, 20].

In this paper, we provide formal definitions of stochastic load balancing for various multiserver systems. In the case of identical processors, it is intuitively clear that the best policy to balance load across processors is to assign each processor "almost" the same amount of work. However, the criterion by which one says a stochastic system, characterized by a random vector, is more "load balanced" than another is seldom precisely specified. Under certain circumstances, such a specified criterion can be made precisely. For example, majorization ordering (Hardy, Littlewood and Polya [15]) is commonly used to compare deterministic vectors. To compare random vectors, Marshall and Olkin [19], Chapter 11, define four stochastic majorization orderings E_i, $i = 1, 2, 3$ and 4, based on sets of functions ($\mathcal{C}_i$, $i = 1, 2, 3$ and 4, in this paper) that map a random vector to a random variable. The comparison of two random vectors is based on the comparison of the distributions or the means of the random variables mapped from the random vectors by the set of functions, $\mathcal{C}_i$, $i = 1, 2, 3$ and 4 (see Definition 2.9, below). However, these four orderings are either too strong to verify if they are satisfied or too weak to provide sufficient information about the joint distributions of random vectors being compared. In our work, we define a new ordering which is strong enough to provide useful information about the joint distributions and yet is easy to verify.

In Section 2, we expand the above definitions to include a new ordering E_5 ($E_5^{\uparrow}$) based on the set of functions $\mathcal{C}_5$ ($\mathcal{C}_5^{\uparrow}$) that are (increasing) symmetric, L-subadditive and convex in each variable. For two deterministic vectors, this new ordering is equivalent to the majorization ordering and the other four orderings. However, the orderings are *not* equivalent when comparing random vectors. We establish some relationships among these five stochastic majorization orderings and provide counterexamples to show that there is no implication between E_5 and E_2. Using the notion of *stochastic convexity*, we also establish sample path sufficient criteria for obtaining the new stochastic ordering $E_5^{\uparrow}$, as well as for the two orderings $E_1^{\uparrow}$ and $E_3^{\uparrow}$. Closure properties for these criteria are derived. In Section 3, we apply the sample path criteria to various fork-join queues and routing problems. We show that under the assumption of (i) identical task processing times, (ii) exchangeable task processing times and (iii) identically distributed task processing times, the work load processes are minimized in the sense of (i) $E_1^{\uparrow}$, (ii) $E_5^{\uparrow}$ and (iii) $E_3^{\uparrow}$, respectively, when equal processor power is assigned to each server. We also show that the work load processes subject to the Bernoulli routing mechanism are minimized in the sense of $E_5^{\uparrow}$ when equal load is assigned to each queue. The result for the systems subject to Bernoulli routing mechanisms is extended to the systems subject to Markovian routing mechanisms with nonnegative correlation coefficients (see Example 3.8, below). We then consider scheduling problems with exponential processing times on parallel processors to which sample path criteria cannot be easily verified; for these problems, we provide a probability enumeration method to verify the new ordering and show that the amount of work assigned to each queue is minimized in the sense of E_5 under the Longest Expected Processing Time (LEPT) first policy. These generalize and unify some of the results in the literature. In Section 4, we extend our results to multivariate stochastic majorization, which covers, among other systems, the cases of tandem queues and queues with finite buffer. Finally, our conclusions are given in Section 5.

A few words on the notation. The set of (K-dimensional, $L \times K$-dimensional) real numbers is denoted by $I\!R$ ($I\!R^K$, $I\!R^{LK}$). An underlined letter, e.g. $\underline{x}$, denotes a *vector*; the k-th component of $\underline{x}$ is denoted by x_k. A letter with the *tilde* beneath, e.g. $\utilde{x}$, denotes a *matrix*; the component of $\utilde{x}$ at the i-th row and the j-th column is denoted by x_{ij}. We use an *asterisk* on a vector or a matrix, e.g. $\underline{x}^*$ and $\utilde{x}^*$, to denote the *transpose* of the vector or the matrix. Random variables, random vectors and random matrices are denoted by capital letters, e.g. X, $\underline{X}$ and $\utilde{X}$, except the letters, K, L and J, which are constants. We also use a *caret* to denote *constructed* random variables, random vectors and random matrices, e.g. $\hat{X}$, $\underline{\hat{X}}$ and $\utilde{\hat{X}}$.

2 Majorization

2.1 Majorization for Deterministic Vectors

In this subsection, we briefly review the notion of majorization from [19]. We then establish the equivalence of the orderings defined in [19] and ours when comparing deterministic vectors.

Definition 2.1 *Let $\underline{x}$, $\underline{y}$ and $\underline{z}$ be vectors in $\mathbb{R}^K$. Consider a function $f : \mathbb{R}^K \to \mathbb{R}$.*

(i) *f is* **increasing (decreasing)** *if $\underline{x} \leq \underline{y}$ coordinatewise implies $f(\underline{x}) \leq (\geq) f(\underline{y})$. This condition is equivalent to $\frac{\partial f}{\partial x_i} \geq (\leq) 0$ for $i = 1, \ldots, K$ if f has first partial derivatives. (Here we mean nondecreasing (nonincreasing) when we say increasing (decreasing).)*

(ii) *f is* **convex (concave)** *if for any $\underline{x} = \alpha\underline{y} + (1-\alpha)\underline{z}$ for all $\alpha \in [0,1]$ implies*

$$f(\underline{x}) \leq (\geq) \alpha f(\underline{y}) + (1-\alpha) f(\underline{z}).$$

This condition is equivalent to the Hessian matrix $(\frac{\partial^2 f}{\partial x_i \partial x_j})_{ij}$ is positive (negative) semidefinite if f has second partial derivatives.

(iii) *f is* **convex (concave) in each variable** *if*

$$\begin{aligned} &f(x_1, \ldots, x_{i-1}, \alpha y_i + (1-\alpha) z_i, x_{i+1}, \ldots, x_K) \\ &\leq (\geq) \alpha f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_K) + (1-\alpha) f(x_1, \ldots, x_{i-1}, z_i, x_{i+1}, \ldots, x_K) \end{aligned}$$

for $i = 1, \ldots, K$. This condition is equivalent to $\frac{\partial^2 f}{\partial x_i^2} \geq (\leq) 0$ for $i = 1, \ldots, K$ if f has second partial derivatives.

(iv) *f is* **L-subadditive (L-superadditive)** *if for any $1 \leq i < j \leq K$ and any nonnegative ϵ_1, ϵ_2 the following inequality holds:*

$$\begin{aligned} &f(\ldots, x_i + \epsilon_1, \ldots, x_j + \epsilon_2, \ldots) + f(\ldots, x_i, \ldots, x_j, \ldots) \\ &\leq (\geq) f(\ldots, x_i + \epsilon_1, \ldots, x_j, \ldots) + f(\ldots, x_i, \ldots, x_j + \epsilon_2, \ldots). \end{aligned}$$

The condition is equivalent to $\frac{\partial^2 f}{\partial x_i \partial x_j} \leq (\geq) 0$ for all i, j $(i \neq j)$ if f has second partial derivatives ([19], Chapter 6).

(v) *f is* **symmetric** *if $f(\underline{x}) = f(\underline{y})$, where $\underline{y}$ is any permutation of $\underline{x}$.*

Definition 2.2 *A vector $\underline{x}$ is* **majorized** *by $\underline{y}$ ($\underline{x} \prec \underline{y}$) if (i) $\sum_{i=1}^{j} x_{[i]} \leq \sum_{i=1}^{j} y_{[i]}$, $j = 1, \ldots, K-1$ and (ii) $\sum_{i=1}^{K} x_{[i]} = \sum_{i=1}^{K} y_{[i]}$, where $x_{[i]}$ ($y_{[i]}$) is the i-th largest component in $\underline{x}$ ($\underline{y}$).*

A vector $\underline{x}$ is **weakly submajorized** *by $\underline{y}$ ($\underline{x} \prec_w \underline{y}$) if $\sum_{i=1}^{j} x_{[i]} \leq \sum_{i=1}^{j} y_{[i]}$, $j = 1, \ldots, K$.*

A vector $\underline{x}$ is **weakly supermajorized** *by $\underline{y}$ ($\underline{x} \prec^w \underline{y}$) if $\sum_{i=j}^{K} x_{[i]} \geq \sum_{i=j}^{K} y_{[i]}$, $j = 1, \ldots, K$.*

Definition 2.3 *A function $f : I\!\!R^K \to I\!\!R$ is* **Schur convex** *if $f(\underline{x}) \leq f(\underline{y})$ for all $\underline{x} \prec \underline{y}$.*

The following four classes of functions are considered in [19]:

$\mathcal{C}_1 \stackrel{\text{def}}{=} \{f : f \text{ is Schur convex }\}$.

$\mathcal{C}_2 \stackrel{\text{def}}{=} \{f : f \text{ is convex and symmetric }\}$.

$\mathcal{C}_3 \stackrel{\text{def}}{=} \{f : f(\underline{x}) = \sum_{i=1}^{K} g(x_i) \text{ for some convex function } g \text{ on } I\!\!R \}$.

$\mathcal{C}_4 \stackrel{\text{def}}{=} \{f : f(\underline{x}) = \sum_{i=1}^{j} x_{[i]} \text{ or } f(\underline{x}) = -\sum_{i=j}^{K} x_{[i]} \text{ for some } j = 1, \ldots, K\}$.

We consider the following new class of functions:

$\mathcal{C}_5 \stackrel{\text{def}}{=} \{f : f \text{ is symmetric, L-subadditive and convex in each variable }\}$.

Let $\mathcal{C}_i^{\uparrow} \stackrel{\text{def}}{=} \{f : f \in \mathcal{C}_i \text{ and } f \text{ is increasing }\}$ and $\mathcal{C}_i^{\downarrow} \stackrel{\text{def}}{=} \{f : f \in \mathcal{C}_i \text{ and } f \text{ is decreasing }\}$.

As commented in [30], one can consider any family of symmetric functions to define a stochastic majorization ordering. The advantage of choosing $\mathcal{C}_5$ is that the stochastic majorization ordering induced by $\mathcal{C}_5$ is closely related to stochastic convexity ([28][29]).

In Lemma 2.4, we show that $\mathcal{C}_5$ is a richer class of functions than $\mathcal{C}_3$ and $\mathcal{C}_4$. This inclusion relationship is then used in the proof of Theorem 2.7 to show the equivalence in comparing deterministic vectors.

Lemma 2.4 *(i) $\mathcal{C}_3 \subset \mathcal{C}_5$, (ii) $\mathcal{C}_4 \subset \mathcal{C}_5$.*

Proof: (i) Note that $f(\underline{x}) = \sum_{i=1}^{K} g(x_i)$ for any $f \in \mathcal{C}_3$. It is clear that $f \in \mathcal{C}_3$ is symmetric and convex in each variable since g is convex. Moreover, $\frac{\partial^2 f}{\partial x_i \partial x_j} = 0$ for $i \neq j$ implies L-subadditivity.

(ii) Let $h^j(\underline{x}) \stackrel{\text{def}}{=} \max_{1 \leq i_1 < \ldots < i_j \leq K} (\sum_{l=1}^{j} x_{i_l})$ for $j = 1, \ldots, K$. In words, $h^j(\underline{x})$ is the largest sum of j components choosing out of $\underline{x}$. Note that any $f \in \mathcal{C}_4$ has either the form $f(\underline{x}) = h^j(\underline{x})$ or $f(\underline{x}) = h^j(\underline{x}) - \sum_{i=1}^{K} x_i$, for some $j = 1, \ldots, K$. Thus, it suffices to show that $h^j \in \mathcal{C}_5$ for all j. To simplify the notation in the proof, we omit the superscript j on h. It is clear that h is symmetric and convex in each variable. Thus, we only need to show that h is L-subadditive. Since h is symmetric, it suffices to show that for $\epsilon_1, \epsilon_2 \geq 0$,

$$h + h_{12} \leq h_1 + h_2,$$

where $h_{12} \stackrel{\text{def}}{=} h(x_1 + \epsilon_1, x_2 + \epsilon_2, x_3, \ldots, x_K)$, $h_1 \stackrel{\text{def}}{=} h(x_1 + \epsilon_1, x_2, \ldots, x_K)$ and $h_2 \stackrel{\text{def}}{=} h(x_1, x_2 + \epsilon_2, x_3, \ldots, x_K)$. The proof is then to enumerate all the cases. Let $M(h)$ be the set of the components that results in the maximum in h, i.e., $\sum_{l \in M(h)} x_l = h(\underline{x})$. $M(h_{12})$, $M(h_1)$ and $M(h_2)$ are defined accordingly. By symmetry, we need to consider the following five cases:

Case 1. $\{1, 2\} \subset M(h)$

Since the maximum in h occurs at a sum that contains x_1 and x_2, we have $M(h_{12}){=}M(h_1){=}M(h_2){=}M(h)$ in this case. Thus,

$$h = x_1 + x_2 + \sum_{l \in M(h) - \{1,2\}} x_l,$$

$$h_{12} = x_1 + \epsilon_1 + x_2 + \epsilon_2 + \sum_{l \in M(h) - \{1,2\}} x_l,$$

$$h_1 = x_1 + \epsilon_1 + x_2 + \sum_{l \in M(h) - \{1,2\}} x_l \quad \text{and}$$

$$h_2 = x_1 + x_2 + \epsilon_2 + \sum_{l \in M(h) - \{1,2\}} x_l.$$

Clearly, $h + h_{12} = h_1 + h_2$.

Case 2. $\{1\} \subset M(h)$ and $\{2\} \not\subset M(h)$

In this case, $M(h_1) = M(h)$ and $h_1 = h + \epsilon_1$. Observe that $h_{12} \leq h_2 + \epsilon_1$ since ϵ_1 only appears with x_1 in h_{12}. Thus, $h + h_{12} \leq h_1 + h_2$.

Case 3. $\{1\} \not\subset M(h)$, $\{2\} \not\subset M(h)$, $\{1\} \not\subset M(h_{12})$ and $\{2\} \not\subset M(h_{12})$

The maximum in h_{12} occurs at a sum that contains neither x_1 nor x_2. In this case, $M(h) = M(h_{12}) = M(h_1) = M(h_2)$. Thus, $h = h_{12} = h_1 = h_2$.

Case 4. $\{1\} \not\subset M(h)$, $\{2\} \not\subset M(h)$, $\{1\} \subset M(h_{12})$ and $\{2\} \not\subset M(h_{12})$

In this case, we still have $M(h_{12}) = M(h_1)$ and $h_{12} = h_1$. Since $h \leq h_2$, we have that $h + h_{12} \leq h_1 + h_2$.

Case 5. $\{1\} \not\subset M(h)$, $\{2\} \not\subset M(h)$, $\{1\} \subset M(h_{12})$ and $\{2\} \subset M(h_{12})$

Let s be one of the two components that are in $M(h)$ but not in $M(h_{12})$. Since h_1 and h_2 are the largest sum of j components in $\{x_1 + \epsilon_1, x_2, \ldots, x_n\}$ and $\{x_1, x_2 + \epsilon_2, \ldots, x_n\}$, we have that

$$h_1 \geq x_1 + \epsilon_1 + x_s + \sum_{l \in M(h_{12}) - \{1,2\}} x_l$$

and that

$$h_2 \geq x_2 + \epsilon_2 + \sum_{l \in M(h) - \{s\}} x_l.$$

Therefore, we have $h + h_{12} \leq h_1 + h_2$. ■

Remark 2.5 As a corollary of Lemma 2.4 (ii), we have that $\max(x_1, x_2, \ldots, x_K)$ is in $\mathcal{C}_5$. Since the maximum function is increasing, it is also in $\mathcal{C}_5^{\uparrow}$. This result can be extended to include the class of functions

$$\{f : f(\underline{x}) = g(max(\underline{x})) \text{ for some increasing convex function } g \text{ } on \text{ } I\!R\},$$

as follows: It is clear that the class of functions above is increasing, symmetric and convex in each variable. To show that the class of functions is in $\mathcal{C}_5^{\uparrow}$, it suffices to show that it is L-subadditive. Observe that for ϵ_1, $\epsilon_2 \geq 0$,

$$\max(\max(x_1+\epsilon_1, x_2, x_3, \ldots, x_K), \max(x_1, x_2+\epsilon_2, x_3, \ldots, x_K)) = \max(x_1+\epsilon_1, x_2+\epsilon_2, x_3, \ldots, x_K)$$

and

$$\min(\max(x_1 + \epsilon_1, x_2, x_3, \ldots, x_K), \max(x_1, x_2 + \epsilon_2, x_3, \ldots, x_K)) \geq \max(x_1, x_2, x_3, \ldots, x_K).$$

Since g is increasing,

$$\begin{aligned} &g(\max(x_1 + \epsilon_1, x_2 + \epsilon_2, x_3, \ldots, x_K)) + g(\max(x_1, x_2, x_3, \ldots, x_K)) \\ &\leq g(\max(x_1 + \epsilon_1, x_2, x_3, \ldots, x_K)) + g(\max(x_1, x_2 + \epsilon_2, x_3, \ldots, x_K)). \end{aligned}$$

We need the following proposition in the proof of Theorem 2.7.

Proposition 2.6 *$\underline{x} \prec \underline{y}$ if and only if $\underline{x}$ can be derived from $\underline{y}$ by applications of a finite number of T-transformations ([19], p. 107), that is, transformations T having for some $1 \leq i < j \leq K$ and some $\alpha \in [0,1]$, the form*

$$T(\underline{y}) = (y_1, \ldots, y_{i-1}, \alpha y_i + (1-\alpha)y_j, y_{i+1}, \ldots, y_{j-1}, (1-\alpha)y_i + \alpha y_j, y_{j+1}, \ldots, y_K).$$

Theorem 2.7 *The following are equivalent*
(i) $\underline{x} \prec \underline{y}$;
(ii) $f(\underline{x}) \leq f(\underline{y})$ for all $f \in \mathcal{C}_1$;
(iii) $f(\underline{x}) \leq f(\underline{y})$ for all $f \in \mathcal{C}_2$;
(iv) $f(\underline{x}) \leq f(\underline{y})$ for all $f \in \mathcal{C}_3$;
(v) $f(\underline{x}) \leq f(\underline{y})$ for all $f \in \mathcal{C}_4$;
(vi) $f(\underline{x}) \leq f(\underline{y})$ for all $f \in \mathcal{C}_5$.

Proof: That (i),(ii),(iii),(iv) and (v) are equivalent is well known. (i) $\Rightarrow$ (ii) is by Definition 2.3. (ii) $\Rightarrow$ (iii) is by [19], Proposition 3.C.2. (iii) $\Rightarrow$ (iv) is from the fact that $\mathcal{C}_3 \subset \mathcal{C}_2$. (iv) $\Rightarrow$ (v) is by [19], Proposition 4.B.1 which is adapted from [15]. (v) $\Rightarrow$ (i) is trivial. (vi) $\Rightarrow$ (i) is a direct consequence of Lemma 2.4.

Thus, we only need to show (i) $\Rightarrow$ (vi). By Proposition 2.6, it suffices to show $f(T(\underline{y})) \leq f(\underline{y})$ for all $f \in \mathcal{C}_5$. To simplify the notation, we only consider the case $K = 2$ (without loss of generality). By convexity in each variable, we have

$$\begin{aligned} & f(T(\underline{y})) = f(\alpha y_1 + (1-\alpha)y_2, (1-\alpha)y_1 + \alpha y_2) \\ & \leq \alpha(1-\alpha)f(y_1,y_1) + \alpha^2 f(y_1,y_2) + (1-\alpha)^2 f(y_2,y_1) + \alpha(1-\alpha)f(y_2,y_2). \end{aligned}$$

From L-subadditivity, it follows that

$$f(y_1,y_1) + f(y_2,y_2) \leq f(y_1,y_2) + f(y_2,y_1)$$

and thus

$$f(T(\underline{y})) \leq \alpha f(y_1,y_2) + (1-\alpha)f(y_2,y_1).$$

Applying the symmetric property of f completes the proof. ∎

We note that Theorem 2.7 can be extended to weak majorizations by replacing $\mathcal{C}_i$ with $\mathcal{C}_i^{\uparrow}$ or $\mathcal{C}_i^{\downarrow}$, for $i = 1,2,3,4,5$. The proofs for these cases follow from the fact that $\underline{x} \prec_w (\prec^w)\underline{y}$ implies that $\underline{x}$ can be derived from $\underline{y}$ by a finite number of T-transformations, or transformations of the form

$$T(\underline{y}) = (y_1, \ldots, y_{i-1}, \alpha y_i, y_{i+1}, \ldots, y_K),$$

where $0 \leq \alpha < 1 (\alpha > 1)$ ([19], Proposition 4.A.2).

From Theorem 2.7 and Definition 2.3, it follows that all the functions in $\mathcal{C}_5$ are Schur convex. Using the facts in [19], pp. 311, Lemma 2.4 and Theorem 2.7, we have an exhaustive list of inclusions given by $\mathcal{C}_3 \subset \mathcal{C}_5 \subset \mathcal{C}_1$, $\mathcal{C}_3 \subset \mathcal{C}_2 \subset \mathcal{C}_1$, $\mathcal{C}_4 \subset \mathcal{C}_5$ and $\mathcal{C}_4 \subset \mathcal{C}_2$. Other inclusions are not possible. To see that $\mathcal{C}_5 \not\subset \mathcal{C}_2$, consider the function $f(x_1, x_2) = -x_1 x_2$. Clearly, it is symmetric. Moreover, $\frac{\partial^2 f}{\partial x_i^2} = 0$, $i = 1, 2$, and $\frac{\partial^2 f}{\partial x_1 \partial x_2} = -1 < 0$. Thus, $f \in \mathcal{C}_5$. However, the determinant of the Hessian matrix

$$det \begin{vmatrix} 0 & -1 \\ -1 & 0 \end{vmatrix} < 0$$

and thus $f \notin \mathcal{C}_2$. To see that $\mathcal{C}_2 \not\subset \mathcal{C}_5$, consider the function $f(x_1, x_2) = (x_1 + x_2)^2$. Clearly, $f \in \mathcal{C}_2$, but $\frac{\partial^2 f}{\partial x_1 \partial x_2} = 2 > 0$ and thus $f \notin \mathcal{C}_5$.

2.2 Stochastic Majorization

In order to define the notions of stochastic majorizations, we need the following partial orderings for random variables.

Definition 2.8 *Let X and Y be two random variables. X is smaller than Y in*

(i) stochastic ordering *($X \leq_{st} Y$) if $Ef(X) \leq Ef(Y)$ for all f increasing. Equivalently, $P(X > t) \leq P(Y > t)$ for all $t \in \mathbb{R}$.*

(ii) convex ordering *($X \leq_c Y$) if $Ef(X) \leq Ef(Y)$ for all f convex.*

(iii) increasing convex ordering *($X \leq_{ic} Y$) if $Ef(X) \leq Ef(Y)$ for all f increasing convex.*

Definition 2.9 *Let $\underline{X}$ and $\underline{Y}$ be K-dimensional random vectors. We define the following stochastic majorization orderings:*
For $i = 1, 2, 3, 4, 5$,
P_i: $\underline{X} \prec_{P_i} \underline{Y}$ if $f(\underline{X}) \leq_{st} f(\underline{Y})$ for all $f \in \mathcal{C}_i$;
E_i: $\underline{X} \prec_{E_i} \underline{Y}$ if $Ef(\underline{X}) \leq Ef(\underline{Y})$ for all $f \in \mathcal{C}_i$.
Similarly, for $i = 1, 2, 3, 4, 5$,
$P_i^{\uparrow}(P_i^{\downarrow})$: $\underline{X} \prec_{wP_i} (\prec^{w}{}_{P_i}) \underline{Y}$ if $f(\underline{X}) \leq_{st} f(\underline{Y})$ for all $f \in \mathcal{C}_i^{\uparrow}(\mathcal{C}_i^{\downarrow})$;
$E_i^{\uparrow}(E_i^{\downarrow})$: $\underline{X} \prec_{wE_i} (\prec^{w}{}_{E_i}) \underline{Y}$ if $Ef(\underline{X}) \leq Ef(\underline{Y})$ for all $f \in \mathcal{C}_i^{\uparrow}(\mathcal{C}_i^{\downarrow})$.
(The stochastic majorization orderings for $i = 1, 2, 3, 4$ are defined in [19], Chapter 11, pp. 282 and 311.)

In this paper, we focus on the new orderings induced by the functions in $\mathcal{C}_5$, $\mathcal{C}_5^{\uparrow}$ and $\mathcal{C}_5^{\downarrow}$. The following two examples provide insight into the new ordering.

Example 2.10 Suppose $X_i, i = 1, \ldots, K$, are independent random variables, $Y_i, i = 1, \ldots, K$ are independent random variables, and $X_i \leq_c (\leq_{ic}) Y_i$, $i = 1, \ldots, K$. Ross [27], Proposition 8.5.4, showed that $Ef(\underline{X}) \leq Ef(\underline{Y})$ for all f that is (increasing) convex in each variable (although Ross seems to implicitly assume convexity of f, his proof is valid under the weaker assumption of convexity in each variable). It then follows that $\underline{X} \prec_{E_5} (\prec_{wE_5}) \underline{Y}$.

Example 2.11 Suppose $X_i, i = 1, \ldots, K$, are identically distributed. Then it follows from Lorentz's inequality for L-superadditive functions (see [26] for details) that

$$(X_1, X_1, \ldots, X_1) \prec_{E_5} \underline{X}.$$

We should point out that there are many other partial ordering for random vectors (see [34]). Also, Chang, Chao, Pinedo and Shanthikumar [9] have studied the stochastic ordering for random vectors based on increasing and directionally convex (idcx) functions ($\frac{\partial f}{\partial x_i} \geq 0$ for all i and $\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0$ for all i, j including $i = j$). Examples 2.10 and 2.11 correspond to their variability condition and subsequence condition.

Though the conditions in Theorem 2.7 are equivalent when comparing deterministic vectors, they are *not* equivalent when comparing random vectors. The set of inclusions for the classes of functions shows that $P_1 \Rightarrow P_5$, $P_5 \Rightarrow P_3$, $P_5 \Rightarrow P_4$, $E_1 \Rightarrow E_5$, $E_5 \Rightarrow E_3$ and $E_5 \Rightarrow E_4$. It is also clear that $P_5 \Rightarrow E_5$. Two simple examples, however, show that $E_2 \not\Rightarrow E_5$ and $E_5 \not\Rightarrow E_2$. To establish $E_2 \not\Rightarrow E_5$, let $P(\underline{Y} = (1,0)) = P(\underline{Y} = (0,-1)) = \frac{1}{2}$ and $P(\underline{X} = (\frac{1}{2}, -\frac{1}{2}))$ $= 1$. By Jensen's inequality, $\underline{X} \prec_{E_2} \underline{Y}$. However, for the function $f(x_1, x_2) = -x_1 x_2$, we have $Ef(\underline{Y}) = 0 \leq Ef(\underline{X}) = \frac{1}{4}$ and thus $E_2 \not\Rightarrow E_5$. To show $E_5 \not\Rightarrow E_2$, let X_1 and X_2 be independent with $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$, $i = 1$ and 2. By Lorentz's inequality (Example 2.11), we have $(X_1, X_1) \prec_{E_5} (X_1, X_2)$. However, for the function $f(x_1, x_2) = (x_1 + x_2)^2$, we have $Ef(X_1, X_2) = 2 \leq Ef(X_1, X_1) = 4$. Thus, $E_5 \not\Rightarrow E_2$. An extended graph of [19], Fig. 11.1, which summarizes the relationships among these orderings, is shown in Fig. 1.

We now establish properties of P_5 and E_5. These properties have proofs that are analogous to those found in [19], pp. 314.

Theorem 2.12 (Closure under Convolutions) *Suppose $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ are independent, $\underline{Y}^{(1)}$ and $\underline{Y}^{(2)}$ are independent, and these random vectors have* exchangeable *components. If $\underline{X}^{(1)} \prec_{E_5} \underline{Y}^{(1)}$ and $\underline{X}^{(2)} \prec_{E_5} \underline{Y}^{(2)}$, then $\underline{X}^{(1)} + \underline{X}^{(2)} \prec_{E_5} \underline{Y}^{(1)} + \underline{Y}^{(2)}$.*

Theorem 2.13 (Preservation under Mixtures) *Suppose* $\underline{X}^{(\theta)} \prec_{P_5} (\prec_{E_5}) \underline{Y}^{(\theta)}$ *for all* θ, *then* $\underline{X}^{(\Theta)} \prec_{P_5} (\prec_{E_5}) \underline{Y}^{(\Theta)}$, *where* Θ *is a random variable that is independent of* $\underline{X}^{(\theta)}$ *and* $\underline{Y}^{(\theta)}$ *for all* θ.

Theorem 2.14 (Closure under Limits in Distribution) *Suppose* $\{\underline{X}^{(j)}\}$ *converges in distribution to* $\underline{X}$ *and* $\{\underline{Y}^{(j)}\}$ *converges in distribution to* $\underline{Y}$. *If* $\underline{X}^{(j)} \prec_{P_5} \underline{Y}^{(j)}$, $j = 1, 2, \ldots$, *then* $\underline{X} \prec_{P_5} \underline{Y}$. *If* $\underline{X}^{(j)} \prec_{E_5} \underline{Y}^{(j)}$, $j = 1, 2, \ldots$, *then* $\underline{X} \prec_{E_5} \underline{Y}$ *is preserved when* $\{f(\underline{X}^{(j)})\}$ *and* $\{f(\underline{Y}^{(j)})\}$ *are uniformly integrable (see [10]).*

Clearly, Theorems 2.12-2.14 also hold for the weak stochastic majorization. For the partial ordering $E_5^{\uparrow}$, we have the following theorem.

Theorem 2.15 (Closure property of Composition) *Let* $g : I\!R \rightarrow I\!R$ *and* $g \odot \underline{x} \stackrel{\text{def}}{=} (g(x_1), \ldots, g(x_K))$ *be a function from* $I\!R^K$ *to* $I\!R^K$. *If* $\underline{X} \prec_{wE_5} \underline{Y}$ *and* g *is increasing and convex, then* $g \odot \underline{X} \prec_{wE_5} g \odot \underline{Y}$.

Proof: Let $h(\underline{x}) \stackrel{\text{def}}{=} f(g \odot \underline{x})$. By calculus, it can be shown that $h \in \mathcal{C}_5^{\uparrow}$ if $f \in \mathcal{C}_5^{\uparrow}$ and g is increasing and convex. ∎

2.3 Sample Path Criteria

In the previous subsection, we established some closure properties for the new ordering E_5. However, the assumption of exchangeability in Theorem 2.12 is too strong for the problems to be considered in this paper. To avoid this difficulty, we establish relationships between the ordering and the sample paths (realizations) of random vectors. First, we define some sample path criteria and show that these criteria imply the new ordering. Under sample path constructions, it is easy to check if the criteria are preserved under convolutions and other operations. We then establish conditions for the closure properties of these criteria. To provide a clear comparison among these orderings, we not only develop sample path criteria for the partial ordering $E_5^{\uparrow}$, but also for $E_1^{\uparrow}$ and $E_3^{\uparrow}$.

Since two deterministic vectors ordered in majorization ordering can be derived from one to the other by applications of a finite number of T-transformations (Proposition 2.6), we define sample path criteria based on T-transformations. To use the tools for stochastic convexity (see [28] [29]), the sample path criteria for $E_1^{\uparrow}$ are defined in a form that is equivalent to T-transformations (see the proof of Theorem 2.17, below).

Definition 2.16 $E_{1(i,j)}^{\uparrow}$: $\underline{X} \prec_{wE_1(i,j)} \underline{Y}$ *if for some* $i \neq j$, *there exist two random vectors,* $\underline{\hat{X}}$ *and* $\underline{\hat{Y}}$, *on the same probability space such that*

(A1) $\underline{\hat{X}} =_{st} \underline{X}$ *and* $\underline{\hat{Y}} =_{st} \underline{Y}$,

(A2) $\hat{X}_k = \hat{Y}_k$ *a.s., for all* $k \neq i, j$,

(A3i) $\hat{Y}_i \geq max(\hat{X}_i, \hat{X}_j)$ *a.s., and*

(A3c) $\hat{X}_i + \hat{X}_j \leq \hat{Y}_i + \hat{Y}_j$ *a.s.*

Theorem 2.17 *If* $\underline{X} \prec_{wE_1(i,j)} \underline{Y}$, *then* $\underline{X} \prec_{wE_1} \underline{Y}$.

Proof: To simplify the notation, we only prove the theorem for the case $K = 2$ ($\underline{X} \prec_{wE_1(1,2)} \underline{Y}$). By Definition 2.16, we can construct $\underline{\hat{X}}$ and $\underline{\hat{Y}}$ satisfying (A1) and (A3). Since f is increasing, we need only consider the case $\hat{X}_1 + \hat{X}_2 = \hat{Y}_1 + \hat{Y}_2$. If $\hat{Y}_1 = \hat{Y}_2$, then it follows from (A3i) that $\hat{X}_1 = \hat{X}_2 = \hat{Y}_1 = \hat{Y}_2$. Thus, we may assume that $\hat{Y}_2 > \hat{Y}_1$ and write

$$\underline{\hat{X}} = (\alpha\hat{Y}_1 + (1-\alpha)\hat{Y}_2, (1-\alpha)\hat{Y}_1 + \alpha\hat{Y}_2),$$

where $\alpha = (\hat{X}_1 - \hat{Y}_2)/(\hat{Y}_1 - \hat{Y}_2)$. From (A3i), it follows that $\alpha \in [0,1]$. Thus, $\underline{\hat{X}}$ is a T-transformation of $\underline{\hat{Y}}$ which implies ([19], Proposition 4.A.1) $\underline{\hat{X}} \prec \underline{\hat{Y}}$ a.s. It then follows from (A1) that $Ef(\underline{X}) \leq Ef(\underline{Y})$ for all $f \in \mathcal{C}_1^{\uparrow}$. ∎

Theorem 2.18 (Preservation under Mixtures) *Suppose* $\underline{X}^{(\theta)} \prec_{wE_1(i,j)} \underline{Y}^{(\theta)}$ *for all* θ, *then* $\underline{X}^{(\Theta)} \prec_{wE_1(i,j)} \underline{Y}^{(\Theta)}$, *where* Θ *is a random variable that is independent of* $\underline{X}^{(\theta)}$ *and* $\underline{Y}^{(\theta)}$ *for all* θ .

Proof: Since Θ is independent of $\underline{X}^{(\theta)}$ and $\underline{Y}^{(\theta)}$, we may construct Θ first and then construct $\underline{X}^{(\Theta)}$ and $\underline{Y}^{(\Theta)}$.

Theorem 2.19 (Closure under Convolutions) *Suppose* $\underline{X}^{(1)}$ *and* $\underline{X}^{(2)}$ *are independent and* $\underline{Y}^{(1)}$ *and* $\underline{Y}^{(2)}$ *are independent. If* $\underline{X}^{(1)} \prec_{wE_1(i,j)} \underline{Y}^{(1)}$ *and* $\underline{X}^{(2)} \prec_{wE_1(i,j)} \underline{Y}^{(2)}$, *then* $\underline{X}^{(1)} + \underline{X}^{(2)} \prec_{wE_1(i,j)} \underline{Y}^{(1)} + \underline{Y}^{(2)}$.

Proof: It follows directly from the fact that if $max(x_2, x_3) \leq x_4$, $x_2 + x_3 \leq x_1 + x_4$ and $max(y_2, y_3) \leq y_4$, $y_2 + y_3 \leq y_1 + y_4$, then $max(x_2 + y_2, x_3 + y_3) \leq x_4 + y_4$ and $x_2 + y_2 + x_3 + y_3 \leq x_1 + y_1 + x_4 + y_4$. ∎

Theorem 2.20 (Closure property of Composition) *If* $\underline{X} \prec_{wE_1(i,j)} \underline{Y}$ *and* g *is increasing and convex, then* $g \odot \underline{X} \prec_{wE_1(i,j)} g \odot \underline{Y}$.

Proof: It follows directly from the fact that if $max(x_2, x_3) \leq x_4$, and $x_2 + x_3 \leq x_1 + x_4$, then $max(g(x_2), g(x_3)) \leq g(x_4)$ and $g(x_2) + g(x_3) \leq g(x_1) + g(x_4)$ for any g increasing convex. ■

Definition 2.16 follows intuitively from the operation of T-transformations; however, it is somewhat restrictive in that it requires the construction of two random vectors on the same probability space that satisfy inequalities (A3i) and (A3c) *simultaneously*. Since all the stochastic majorization orderings are based on symmetric functions, any interchange of two components in a random vector is equivalent to the original vector in all of the stochastic majorization orderings. To avoid construction difficulty, we construct two pairs of random vectors. In each pair, we construct a random vector that is stochastically equivalent to the original random vector and the other that is stochastically equivalent to an interchange of two components, say components i and j, in the original random vector. The criteria are that the four i-th components and the four j-th components in these constructed random vectors satisfy inequalities similar to (A3i) and (A3c). These considerations motivate Definition 2.21 below, which turns out to be a sample-path criterion for the partial ordering $E_5^{\uparrow}$.

Let $\underline{X}^{ij} \stackrel{\text{def}}{=} (X_1, \ldots, X_{i-1}, X_j, X_{i+1}, \ldots, X_{j-1}, X_i, X_{j+1}, \ldots, X_K)$ be the interchange of the i-th component and the j-th component of $\underline{X}$. $\underline{Y}^{ij}$ is defined accordingly (i is not necessarily smaller than j and $\underline{X}^{ij} = \underline{X}^{ji}$). The reader should be cautioned that $\hat{\underline{X}}^{ij}$ is not equivalent to $(\hat{\underline{X}})^{ij}$, which is an interchange of $\hat{\underline{X}}$. $\hat{\underline{X}}^{ij}$ and $(\hat{\underline{X}})^{ij}$ only have the same joint distributions as that of $\underline{X}^{ij}$. The reader may regard the caret and the superscript $\{ij\}$ as operators on a vector. The superscript $\{ij\}$ is executed before the caret in $\hat{X}^{ij}$. The order is reversed in $(\hat{\underline{X}})^{ij}$.

Definition 2.21 $E_5^{\uparrow}(i,j)$: $\underline{X} \prec_{wE_5(i,j)} \underline{Y}$ *if for some* $i \neq j$, *there exist four random vectors,* $\hat{\underline{X}}$, $\hat{\underline{X}}^{ij}$, $\hat{\underline{Y}}$ *and* $\hat{\underline{Y}}^{ij}$ *on the same probability space such that*

(B1) $\hat{\underline{X}} =_{st} \underline{X}$, $\hat{\underline{X}}^{ij} =_{st} \underline{X}^{ij}$, $\hat{\underline{Y}} =_{st} \underline{Y}$ *and* $\hat{\underline{Y}}^{ij} =_{st} \underline{Y}^{ij}$,

(B2) $\hat{X}_k = \hat{X}_k^{ij} = \hat{Y}_k = \hat{Y}_k^{ij}$ *a.s., for all* $k \neq i, j$,

(B3i) $\hat{Y}_i \geq max(\hat{X}_i, \hat{X}_i^{ij})$ *a.s.,*

(B3c) $\hat{X}_i + \hat{X}_i^{ij} \leq \hat{Y}_i + \hat{Y}_i^{ij}$ *a.s.,*

(B4i) $\hat{Y}_j^{ij} \geq max(\hat{X}_j, \hat{X}_j^{ij})$ *a.s., and*

(B4c) $\hat{X}_j + \hat{X}_j^{ij} \leq \hat{Y}_j + \hat{Y}_j^{ij}$ *a.s.*

Note that $\hat{X}_i^{ij}$ and $\hat{Y}_i^{ij}$ are stochastically the same as $\hat{X}_j$ and $\hat{Y}_j$ and that $\hat{X}_j^{ij}$ and $\hat{Y}_j^{ij}$ are stochastically the same as $\hat{X}_i$ and $\hat{Y}_i$. Thus, $E_1^{\uparrow}(i,j)$ implies $E_5^{\uparrow}(i,j)$ since we can construct $\underline{\hat{X}}^{ij}$ and $\underline{\hat{Y}}^{ij}$ as $(\underline{\hat{X}})^{ij}$ and $(\underline{\hat{Y}})^{ij}$.

In the following, we provide two examples for the new relation $E_5^{\uparrow}(i,j)$: one is a family of random vectors, parameterized by $\underline{\theta}$, in which components are independent and stochastic increasing and convex in sample-path sense ($SICX(sp)$) (see [28] [29]); the other is the family of Bernoulli random vectors.

Example 2.22 A family of distributions $\{F_\theta(\cdot)\}$ is $SICX(sp)$ ($SDCX(sp)$) in θ if for any $\theta_1 + \theta_4 = \theta_2 + \theta_3$, $\theta_4 \geq max(\theta_2, \theta_3)$, there exist four random variables, $\hat{X}_i$, $i = 1,2,3,4$, such that (i) $\hat{X}_i$ has the distribution $F_{\theta_i}(\cdot)$, (ii) $\hat{X}_4 \geq \max(\hat{X}_2, \hat{X}_3)$ ($\hat{X}_1 \geq \max(\hat{X}_2, \hat{X}_3)$) and (iii) $\hat{X}_1 + \hat{X}_4 \geq \hat{X}_2 + \hat{X}_3$. Now consider two vectors $\underline{p}$ and $\underline{q}$ with $\sum_{i=1}^K p_i = \sum_{i=1}^K q_i = 1$, $p_i = q_i$ for $i = 3, \ldots, K$ and $p_1 \geq max(p_2, q_1, q_2)$. Let $\{X_k(q_k)\}$ and $\{Y_k(p_k)\}$, $k = 1, \ldots, K$, be two sets of *independent* random variables with distributions $F_{q_k}(\cdot)$ and $F_{p_k}(\cdot)$, respectively. Let $\underline{X} \stackrel{\text{def}}{=} (X_1(q_1), \ldots, X_K(q_K))$ and $\underline{Y} \stackrel{\text{def}}{=} (Y_1(p_1), \ldots, Y_K(p_K))$. Note that $p_1 + p_2 = q_1 + q_2$ and that $\underline{X}^{12}$ and $\underline{Y}^{12}$ are random vectors parameterized by $(q_2, q_1, q_3, \ldots, q_K)$ and $(p_2, p_1, p_3, \ldots, p_K)$, respectively. It then follows from the independence assumption that $\underline{X} \prec_{wE_5(1,2)} \underline{Y}$ ($\underline{X} \prec_{wE_5(2,1)} \underline{Y}$) if $\{F_\theta(\cdot)\}$ is $SICX(sp)$ ($SDCX(sp)$) in θ.

Reference [19] (p. 102) shows that if $\{F_\theta(\cdot)\}$ satisfies some semigroup property and some total positivity property, then $\underline{X} \prec_{E_1} \underline{Y}$. Without the assumption of total positivity, [29] arrives at the conclusion that $\underline{X} \prec_{E_3} \underline{Y}$. Reference [29], Example 4.1 also shows that the semigroup property implies that $\{F_\theta(\cdot)\}$ is $SIL(sp)$ in θ which replaces the inequality (iii) in Example 2.22 by an equality. Using Example 2.22 and the proof of Theorem 2.24 (the next theorem), it can be shown that $\underline{X} \prec_{E_5} \underline{Y}$ under the independence assumption in Example 2.22 and the assumption of the semigroup property. This conclusion is weaker than that in [19] since we do not assume total positivity, but stronger than that in [29] since [29] does not assume the independence of components.

Example 2.22 shows that the relation $E_5^{\uparrow}(i,j)$ can be obtained if the components are independent and SICX(sp). In the following, we consider Bernoulli random vectors. We show the relation $E_5^{\uparrow}(i,j)$ still can be obtained even though the components are dependent.

Example 2.23 Let $\underline{e}_i$ be a K-dimensional row vector with a 1 in its i-th component and 0's in the others. Let $P(\underline{Y} = \underline{e}_i) = p_i$ and $P(\underline{X} = \underline{e}_i) = q_i$ with p_i and q_i described in Example 2.22. Then $\underline{X} \prec_{wE_5(1,2)} \underline{Y}$.

Proof: Our construction is similar to the constructive proof for showing that Bernoulli random variables with parameter p is SICX(sp) in p ([29], Example 4.4). The key observation is that not only the constructed random variables but also the "complementary random variables" satisfy the conditions for SICX(sp). To be precise, we define $r_j \stackrel{\text{def}}{=} \sum_{i=1}^{j} p_i = \sum_{i=1}^{j} q_i$ for $j = 2, \ldots, K$. Generate a random variable U uniformly distributed over [0,1]. Construct $\underline{\hat{X}}$, $\underline{\hat{X}}^{12}$, $\underline{\hat{Y}}$ and $\underline{\hat{Y}}^{12}$ as follows (see Fig. 2):

$$\underline{\hat{X}} = \underline{e}_1 \mathbf{I}[p_1 - q_1, p_1) + \underline{e}_2 \mathbf{I}[0, p_1 - q_1) \cup [p_1, r_2) + \sum_{i=3}^{K} \underline{e}_i \mathbf{I}[r_{i-1}, r_i),$$

$$\underline{\hat{X}}^{12} = \underline{e}_1 \mathbf{I}[0, q_2) + \underline{e}_2 \mathbf{I}[q_2, r_2) + \sum_{i=3}^{K} \underline{e}_i \mathbf{I}[r_{i-1}, r_i),$$

$$\underline{\hat{Y}} = \underline{e}_1 \mathbf{I}[0, p_1) + \underline{e}_2 \mathbf{I}[p_1, r_2) + \sum_{i=3}^{K} \underline{e}_i \mathbf{I}[r_{i-1}, r_i), \quad \text{and}$$

$$\underline{\hat{Y}}^{12} = \underline{e}_1 \mathbf{I}[p_1 - q_1, q_2) + \underline{e}_2 \mathbf{I}[0, p_1 - q_1) \cup [q_2, r_2) + \sum_{i=3}^{K} \underline{e}_i \mathbf{I}[r_{i-1}, r_i),$$

where $\mathbf{IA} = 1$ if U is in the set $\mathbf{A}$ and $\mathbf{IA} = 0$, otherwise. Clearly, $\underline{\hat{X}} =_{st} \underline{X}$, $\underline{\hat{X}}^{12} =_{st} \underline{X}^{12}$, $\underline{\hat{Y}} =_{st} \underline{Y}$ and $\underline{\hat{Y}}^{12} =_{st} \underline{Y}^{12}$. Moreover, (B2),(B3) and (B4) are satisfied ((B3c) and (B4c) are satisfied with equality). ∎

Theorem 2.24 *If* $\underline{X} \prec_{wE_5(i,j)} \underline{Y}$*, then* $\underline{X} \prec_{wE_5} \underline{Y}$.

Proof: To simplify the notation, we only prove the theorem for the case $K = 2$ ($\underline{X} \prec_{wE_5(1,2)} \underline{Y}$). By Definition 2.21, we can construct $\underline{\hat{X}}$, $\underline{\hat{X}}^{12}$, $\underline{\hat{Y}}$ and $\underline{\hat{Y}}^{12}$ satisfying (B1),(B3) and (B4). Since f is increasing, we need only consider the case $\underline{\hat{X}} + \underline{\hat{X}}^{12} = \underline{\hat{Y}} + \underline{\hat{Y}}^{12}$. Rewrite

$$\underline{\hat{X}} = (\alpha \hat{Y}_1 + (1-\alpha)\hat{Y}_1^{12}, (1-\beta)\hat{Y}_2 + \beta \hat{Y}_2^{12}),$$
$$\underline{\hat{X}}^{12} = ((1-\alpha)\hat{Y}_1 + \alpha \hat{Y}_1^{12}, \beta \hat{Y}_2 + (1-\beta)\hat{Y}_2^{12}),$$

where $\alpha = (\hat{X}_1 - \hat{Y}_1^{12})/(\hat{Y}_1 - \hat{Y}_1^{12})$ and $\beta = (\hat{X}_2^{12} - \hat{Y}_2)/(\hat{Y}_2^{12} - \hat{Y}_2)$. From (B3i) and (B4i), it follows that $\alpha, \beta \in [0,1]$. Using the property that f is convex in each variable yields

$$\begin{aligned}
&f(\underline{\hat{X}}) + f(\underline{\hat{X}}^{12}) \\
&\leq (\alpha(1-\beta) + (1-\alpha)\beta)(f(\underline{\hat{Y}}) + f(\underline{\hat{Y}}^{12})) \\
&+((1-\alpha)(1-\beta) + \alpha\beta)(f(\hat{Y}_1^{12}, \hat{Y}_2) + f(\hat{Y}_1, \hat{Y}_2^{12})).
\end{aligned}$$

Since $\hat{Y}_1 \geq \hat{Y}_1^{12}$ and $\hat{Y}_2^{12} \geq \hat{Y}_2$, applying the L-subadditivity yields

$$f(\hat{Y}_1^{12}, \hat{Y}_2) + f(\hat{Y}_1, \hat{Y}_2^{12}) \leq f(\underline{\hat{Y}}) + f(\underline{\hat{Y}}^{12}).$$

Thus, we have

$$f(\underline{\hat{X}}) + f(\underline{\hat{X}}^{12}) \leq f(\underline{\hat{Y}}) + f(\underline{\hat{Y}}^{12}).$$

It follows from the symmetry of f and (B1) that $Ef(\underline{\hat{X}}) = Ef(\underline{\hat{X}}^{12}) = Ef(\underline{X})$. Similarly, we have $Ef(\underline{\hat{Y}}) = Ef(\underline{\hat{Y}}^{12}) = Ef(\underline{Y})$. Therefore, $Ef(\underline{X}) \leq Ef(\underline{Y})$ for all $f \in \mathcal{C}_5^{\uparrow}$. ∎

Analogous to Theorems 2.18-2.20, we have the following theorems for the relation $E_{5}^{\uparrow}{}_{(i,j)}$.

Theorem 2.25 (Preservation under Mixtures) *Suppose* $\underline{X}^{(\theta)} \prec_{wE_5(i,j)} \underline{Y}^{(\theta)}$ *for all* θ*, then* $\underline{X}^{(\Theta)} \prec_{wE_5(i,j)} \underline{Y}^{(\Theta)}$*, where* Θ *is a random variable that is independent of* $\underline{X}^{(\theta)}$ *and* $\underline{Y}^{(\theta)}$ *for all* θ*.*

Theorem 2.26 (Closure under Convolutions) *Suppose* $\underline{X}^{(1)}$ *and* $\underline{X}^{(2)}$ *are independent and* $\underline{Y}^{(1)}$ *and* $\underline{Y}^{(2)}$ *are independent. If* $\underline{X}^{(1)} \prec_{wE_5(i,j)} \underline{Y}^{(1)}$ *and* $\underline{X}^{(2)} \prec_{wE_5(i,j)} \underline{Y}^{(2)}$*, then* $\underline{X}^{(1)} + \underline{X}^{(2)} \prec_{wE_5(i,j)} \underline{Y}^{(1)} + \underline{Y}^{(2)}$*.*

Theorem 2.27 (Closure property of Composition) *If* $\underline{X} \prec_{wE_5(i,j)} \underline{Y}$ *and* g *is increasing and convex, then* $g \odot \underline{X} \prec_{wE_5(i,j)} g \odot \underline{Y}$*.*

We note that the closure property of composition can be further extended to increasing and directional convex (idcx) functions. Let $h : I\!R^2 \to I\!R$ and $h \odot (\underline{x}, \underline{y}) \stackrel{\text{def}}{=} (h(x_1, y_1), \ldots, h(x_K, y_K))$ be a function from $I\!R^K \times I\!R^K \to I\!R^K$. Then it can be shown that under the conditions in Theorem 2.26, $h \odot (\underline{X}^{(1)}, \underline{X}^{(2)}) \prec_{wE_5(i,j)} h \odot (\underline{Y}^{(1)}, \underline{Y}^{(2)})$ for all h idcx. This is a generalization of Theorems 2.26 and 2.27 since the summation is idcx and an idcx function is increasing and convex in the first vector when all the components in the second vector are a fixed constant.

There are cases in which the constructions in Definition 2.21 are not possible. For instance, random vectors without any information about the joint distribution are hard to construct. However, if the marginal distributions are known, we can relax the construction to match marginal distributions only. Under the circumstance, the result is weaker than those in Definitions 2.16 and 2.21.

Definition 2.28 $E_3^{\uparrow}(i,j)$ *:* $\underline{X} \prec_{wE_3(i,j)} \underline{Y}$ *if for some* $i \neq j$*, there exist two random vectors,* $\hat{\underline{X}}$ *and* $\hat{\underline{Y}}$*, on the same probability space such that*

(C1) $\hat{X}_i =_{st} X_i$*, and* $\hat{Y}_i =_{st} Y_i$*, for* $i = 1, \ldots, K$*,*

(C2) $\hat{X}_k = \hat{Y}_k$*, for all* $k \neq i, j$ *a.s.,*

(C3i) $\hat{Y}_i \geq max(\hat{X}_i, \hat{X}_j)$ *a.s., and*

(C3c) $\hat{X}_i + \hat{X}_j \leq \hat{Y}_i + \hat{Y}_j$ *a.s.*

Note that Definition 2.28 differs from Definition 2.16 only in (C1) which only requires matching the marginal distributions.

Theorem 2.29 *If* $\underline{X} \prec_{wE_3(i,j)} \underline{Y}$*, then* $\underline{X} \prec_{wE_3} \underline{Y}$*.*

Proof: By Definition 2.28, we can construct $\hat{\underline{X}}$ and $\hat{\underline{Y}}$ satisfying (C1)(C2) and (C3). From (C3), it follows that $g(\hat{X}_i) + g(\hat{X}_j) \leq g(\hat{Y}_i) + g(\hat{Y}_j)$ for all g increasing convex. Using (C2), we have $\sum_{i=1}^{K} g(\hat{X}_i) \leq \sum_{i=1}^{K} g(\hat{Y}_i)$. Taking expectations on both sides and using (C1) yields $Ef(\underline{X}) \leq Ef(\underline{Y})$ for all $f \in \mathcal{C}_3^{\uparrow}$. ∎

Analogous to Theorems 2.18-2.20, we have the following theorems for the relation $E_3^{\uparrow}(i,j)$.

Theorem 2.30 (Preservation under Mixtures) *Suppose* $\underline{X}^{(\theta)} \prec_{wE_3(i,j)} \underline{Y}^{(\theta)}$ *for all* θ*, then* $\underline{X}^{(\Theta)} \prec_{wE_3(i,j)} \underline{Y}^{(\Theta)}$*, where* Θ *is a random variable that is independent of* $\underline{X}^{(\theta)}$ *and* $\underline{Y}^{(\theta)}$ *for all* θ*.*

Theorem 2.31 (Closure under Convolutions) *Suppose* $\underline{X}^{(1)}$ *and* $\underline{X}^{(2)}$ *are independent and* $\underline{Y}^{(1)}$ *and* $\underline{Y}^{(2)}$ *are independent. If* $\underline{X}^{(1)} \prec_{wE_3(i,j)} \underline{Y}^{(1)}$ *and* $\underline{X}^{(2)} \prec_{wE_3(i,j)} \underline{Y}^{(2)}$*, then* $\underline{X}^{(1)} + \underline{X}^{(2)} \prec_{wE_3(i,j)} \underline{Y}^{(1)} + \underline{Y}^{(2)}$*.*

Theorem 2.32 (Closure property of Composition) *If* $\underline{X} \prec_{wE_3(i,j)} \underline{Y}$ *and* g *is increasing and convex, then* $g \odot \underline{X} \prec_{wE_3(i,j)} g \odot \underline{Y}$.

One of the possible extensions is to define the sample path criteria for $E_2^{\uparrow}$. The idea is based on the notion of strong stochastic convexity [33] and Definition 2.21 for $E_5^{\uparrow}(i,j)$.

Definition 2.33 $E_2^{\uparrow}(i,j,\alpha)$*:* $\underline{X} \prec_{wE_2(i,j,\alpha)} \underline{Y}$ *if for some* $i \neq j$ *and* $\alpha \in [0,1]$, *there exist three random vectors,* $\underline{\hat{X}}$, $\underline{\hat{Y}}$ *and* $\underline{\hat{Y}}^{ij}$ *on the same probability space such that*

(E1) $\underline{\hat{X}} =_{st} \underline{X}$, $\underline{\hat{Y}} =_{st} \underline{Y}$ *and* $\underline{\hat{Y}}^{ij} =_{st} \underline{Y}^{ij}$,

(E2) $\hat{X}_k = \hat{Y}_k = \hat{Y}_k^{ij}$ *a.s., for all* $k \neq i, j$,

(E3) $\underline{\hat{X}} \leq \alpha \underline{\hat{Y}} + (1-\alpha)\underline{\hat{Y}}^{ij}$ *a.s.*

It is easy to see that if $\underline{X} \prec_{wE_2(i,j,\alpha)} \underline{Y}$, then $\underline{X} \prec_{wE_2} \underline{Y}$. Moreover, analogous to Theorems 2.18-2.20, the relation, $E_{2(i,j,\alpha)}^{\uparrow}$, enjoys the closure properties under mixtures, convolutions and compositions of increasing convex functions. We note that $\underline{X} \prec_{wE_2(i,j,\alpha)} \underline{Y}$ implies $\underline{X} \prec_{wE_5(i,j)} \underline{Y}$ since $\underline{\hat{X}}^{ij}$ can be constructed as $(\underline{\hat{X}})^{ij}$. Thus, the conditions of the sample path criteria for $E_2^{\uparrow}$ are stronger than those for $E_5^{\uparrow}$.

Since the class of functions $\mathcal{C}_4^{\uparrow}$ only consists of the k functions that define the weak majorization ordering, it would be difficult to define sample path criteria for $E_4^{\uparrow}$ with useful closure properties. Note that the evaluation of the expectations over the class of functions $\mathcal{C}_4^{\uparrow}$ require the knowledge of joint distributions. In order to establish the $E_4^{\uparrow}$ ordering, we suggest going up one level to establish the $E_5^{\uparrow}$ ordering instead. In fact, this is our approach for the problems considered in the next section.

3 Applications

In Sections 3.1 and 3.2, we apply the sample path criteria developed in the previous section to various fork-join queues and routing problems. We then establish stochastic majorization orderings $E_k^{\uparrow}$, $k = 1, 5$ and 3, for these systems under various assumptions. In Section 3.3, we consider scheduling problems to which sample path criteria cannot be easily applied. A probability enumeration method, which counts and cancels transition probabilities of the two systems being compared, is proposed to obtain the new ordering E_5.

3.1 Fork-Join Queues

In this subsection, we illustrate the sample path criteria for $E_i^{\uparrow}(i,j)$, $i = 1, 5$ and 3, by considering a simple example of fork-join queues.

Example 3.1 Consider an extended model for the fork/join queues in [22]. Assume jobs arrive at $\tau_n, n \geq 1$. Upon the arrival of the n-th job, it forks into K tasks. We denote the processing time of task k of the n-th job by $S_{n,k}$, $k = 1, 2, \ldots, K$, and assume that the random vectors $\{(S_{n,1}, \ldots, S_{n,K})\}_n$ are independent (and also independent of the system). Task k is assigned to the k-th queue which consists of a single server with speed p_k and an infinite buffer. We assume that $\sum_{k=1}^{K} p_k = 1$, and that the n-th job leaves the system as soon as all of its tasks complete service. The response time of the n-th job, T_n, is defined to be the time between its arrival epoch and its departure epoch. Let $\underline{V}_t = (V_{1,t}, \ldots, V_{K,t})$ and $\underline{V}_0 = (0, \ldots, 0) \stackrel{\text{def}}{=} \underline{0}$, where $V_{k,t}$ denotes the amount of work in the k-th queue at time t. Clearly, $T_n = \max_k V_{k,\tau_n^+}$, where τ_n^+ is the epoch after the n-th arrival.

Theorem 3.2 (Identical Task Processing Times) *Assume that the processing times of all the tasks in the same job are identical, i.e., $S_{n,k} = S_n$, $k = 1, \ldots, K$. Then (i) $Ef(\underline{V}_t)$ is Schur convex in $\underline{p}$ for all $f \in \mathcal{C}_1^{\uparrow}$. (ii) The response time of every job is minimized in the sense of stochastic ordering when $\underline{p} = (\frac{1}{K}, \ldots, \frac{1}{K})$.*

Theorem 3.3 (Exchangeable Task Processing Times) *Assume that the task processing times in the same job are exchangeable random variables. Then (i) $Ef(\underline{V}_t)$ is Schur convex in $\underline{p}$ for all $f \in \mathcal{C}_5^{\uparrow}$. (ii) The response time of every job is minimized in the sense of increasing convex ordering when $\underline{p} = (\frac{1}{K}, \ldots, \frac{1}{K})$.*

Theorem 3.4 (Identically Distributed Task Processing Times) *Assume that the task processing times in the same job only have the same distribution and no assumption about the joint distribution of task processing times is given. Then $Ef(\underline{V}_t)$ is Schur convex in $\underline{p}$ for all $f \in \mathcal{C}_3^{\uparrow}$.*

These results show that a strong assumption on task processing times, i.e. Theorem 3.2, leads to a strong stochastic majorization ordering. It is intuitive here that the assumption of identical processing times eases the construction on the same probability space. It is harder to construct the work load processes when the assumption of the task processing times is relaxed to be exchangeable, as in Theorem 3.3. Under such an assumption, we need to consider two pairs of systems and construct more processes than the case of identical processing times. Taking expectations on these two pairs of systems enables us to obtain some comparison results which provides some information about the joint distribution of queues. As expected, the result is weaker than the one under the assumption of identical task processing times in Theorem 3.2. If the assumption is further relaxed to identically distributed task processing times, as in Theorem 3.4, the result is even weaker since no assumption about the joint distribution is given. Our construction can only match the marginal distribution of each queue and no information of the joint distribution of queues can be derived.

The rest of this subsection is devoted to the proofs of Theorems 3.2, 3.3 and 3.4. The proofs of these theorems are based on the closure properties (convolution and composition) in the previous section. The governing equation for the fork-join queues is a K-dimensional Lindley's equation. At the instants of arrivals, tasks are added to the queues. By the closure property under convolution, we show that these relations are preserved. Between two successive arrivals, the governing equation is an increasing convex composition of each component. By the closure property under composition, we then show that the relations are also preserved.

Proof of Theorem 3.2: Consider two vectors $\underline{p}$ and $\underline{q}$ of processor speeds specified in Example 2.22. Let $\underline{V}_t(\underline{p})$ and $\underline{V}_t(\underline{q})$ be the corresponding work load processes. We would like to show that $\underline{V}_t(\underline{q}) \prec_{wE_1(2,1)} \underline{V}_t(\underline{p})$. We construct $\hat{\underline{V}}_t(\underline{p})$ and $\hat{\underline{V}}_t(\underline{q})$ as follows:

Case 1: At the instant of the n-th arrival, the K tasks are assigned to the K queues. We construct

$$\hat{\underline{V}}_{\tau_n^+}(\underline{p}) = \hat{\underline{V}}_{\tau_n}(\underline{p}) + S_n \underline{1} \text{ and}$$

$$\hat{\underline{V}}_{\tau_n^+}(\underline{q}) = \hat{\underline{V}}_{\tau_n}(\underline{q}) + S_n \underline{1},$$

where $\underline{1} \stackrel{\text{def}}{=} (1, \ldots, 1)$.

Case 2: Between two successive arrivals, the amount of work in the k-th queue is reduced linearly at the speed p_k or q_k respectively until the queue is empty. For $\tau_n < t \le \tau_{n+1}$ ($\tau_0 \stackrel{\text{def}}{=} 0$) and $1 \le k \le K$,

$$\hat{V}_{k,t}(\underline{p}) = (\hat{V}_{k,\tau_n^+}(\underline{p}) - p_k(t - \tau_n))^+ \text{ and}$$

$$\hat{V}_{k,t}(\underline{q}) = (\hat{V}_{k,\tau_n^+}(\underline{q}) - q_k(t - \tau_n))^+,$$

where $(x)^+ \stackrel{\text{def}}{=} max(0, x)$.

It is clear that (A1) is satisfied for all $t \geq 0$. We will show by induction that (A2) $\hat{V}_{i,t}(\underline{p}) = \hat{V}_{i,t}(\underline{q})$, $i = 3, \ldots, K$, (A3i) $\hat{V}_{2,t}(\underline{p}) \geq max(\hat{V}_{1,t}(\underline{q}), \hat{V}_{2,t}(\underline{q}))$ and (A3c) $\hat{V}_{1,t}(\underline{q}) + \hat{V}_{2,t}(\underline{q}) \leq \hat{V}_{1,t}(\underline{p}) + \hat{V}_{2,t}(\underline{p})$ are satisfied for all $t \geq 0$. Clearly, (A2) and (A3) are satisfied trivially at time 0 since $\underline{\hat{V}}_0(\underline{p}) = \underline{\hat{V}}_0(\underline{q}) = \underline{0}$. Let $\underline{D}_n(\underline{p}) \stackrel{\text{def}}{=} -(t-\tau_n)\underline{p}$ and $\underline{D}_n(\underline{q}) \stackrel{\text{def}}{=} -(t-\tau_n)\underline{q}$. Observe that $\underline{D}_n(\underline{q}) \prec_{wE_1(2,1)} \underline{D}_n(\underline{p})$ since $p_i = q_i$, $i = 3, \ldots, K$, $p_1 \geq max(p_2, q_1, q_2)$ and $p_1 + p_2 = q_1 + q_2$. Applying Theorem 2.19 yields $\underline{V}_{\tau_n^+}(\underline{q}) + \underline{D}_n(\underline{q}) \prec_{wE_1(2,1)} \underline{V}_{\tau_n^+}(\underline{p}) + \underline{D}_n(\underline{p})$. Since $(x)^+$ is increasing and convex, we have from Theorem 2.20 that for the case between two successive arrivals, $\underline{V}_t(\underline{q}) \prec_{wE_1(2,1)} \underline{V}_t(\underline{p})$. By Theorem 2.19, (A2) and (A3) are satisfied at the instants of arrivals. Thus, $\underline{V}_t(\underline{q}) \prec_{wE_1(2,1)} \underline{V}_t(\underline{p})$ for all $t \geq 0$. From Theorem 2.17, we have $\underline{V}_t(\underline{q}) \prec_{wE_1} \underline{V}_t(\underline{p})$. It then follows from Proposition 2.6 that (i) $Ef(\underline{V}_t)$ is Schur convex in $\underline{p}$ for all $f \in \mathcal{C}_1^{\uparrow}$. Since $g(\max(\underline{V}_t))$ is in $\mathcal{C}_1^{\uparrow}$ for all g increasing ([19], p. 61, Table 1), (ii) is a consequence of the fact that $(\frac{1}{K}, \ldots, \frac{1}{K}) \prec \underline{p}$. ■

Proof of Theorem 3.3: (i) The proof is similar to the one in Theorem 3.2. We need only to consider the case at the instants of arrivals since the case between two successive arrivals is concluded by the closure property in Theorem 2.27. At the instant of the n-th arrival, we construct

$$\underline{\hat{V}}_{\tau_n^+}(\underline{p}) = \underline{\hat{V}}_{\tau_n}(\underline{p}) + \underline{S}_n,$$

$$\underline{\hat{V}}^{12}_{\tau_n^+}(\underline{p}) = \underline{\hat{V}}^{12}_{\tau_n}(\underline{p}) + \underline{S}_n,$$

$$\underline{\hat{V}}_{\tau_n^+}(\underline{q}) = \underline{\hat{V}}_{\tau_n}(\underline{q}) + \underline{S}_n,$$

$$\underline{\hat{V}}^{12}_{\tau_n^+}(\underline{q}) = \underline{\hat{V}}^{12}_{\tau_n}(\underline{q}) + \underline{S}_n.$$

Since $\underline{S}_n$ consists of exchangeable random variables, (B1) is satisfied under this construction. It then follows from Theorem 2.26 that (B2) and (B3) are satisfied trivially at the instants of arrivals. (ii) follows from Remark 2.5 and the fact that $(\frac{1}{K}, \ldots, \frac{1}{K}) \prec \underline{p}$. ■

Proof of Theorem 3.4: The proof is similar to the one in Theorem 3.2 with the closure properties replaced by Theorems 2.31 and 2.32. ■

3.2 Routing

In this subsection, we consider various routing mechanisms used for load balancing.

Winston [41] (Shortest Queue Routing) considers a queueing system having K identical exponential servers. He shows that customers should be assigned to the shortest queue in order to maximize the number of departures by time t. The problem has been extensively studied in the literature (see [36], [11], [42], among others). Recently, by a sample path argument analogous to Definition 2.16, Menich [21] shows that the shortest queue policy minimizes $Ef(\underline{Q}_t)$ for all $f \in \mathcal{C}_1^{\uparrow}$, where $\underline{Q}_t$ is the numbers of customers in queues at time t. Thus, the shortest queue policy balances load in the sense of $E_1^{\uparrow}$ ordering. Ephremides, Varaiya and Walrand [11] (Round-Robin Routing) consider a queueing system having two identical exponential servers. They assume that the initial queue lengths are known and equal ($Q_{1,0} = Q_{2,0}$) and no further observations are made. The Round-Robin (RR) policy assigns the odd-numbered arrivals to queue 1 and the remainder to queue 2. They show that the RR policy minimizes $Ef(\underline{Q}_t)$ for all $f \in \mathcal{C}_3^{\uparrow}$, where $\underline{Q}_t$ is the number of customers in the queues at time t. Thus, the RR policy balances load in the sense of $E_3^{\uparrow}$ ordering.

In the following, we show that a Bernoulli routing mechanism with equal probability of splitting the arrivals balances load in the sense of $E_5^{\uparrow}$ among all Bernoulli routing mechanisms. Our result generalizes the result in [13] to general arrival processes.

Example 3.5 [Bernoulli Routing with Resequencing] Consider a queueing system consisting of K identical queues in parallel. Arrivals are characterized by a marked point process (τ_n, S_n), where τ_n and S_n denote the epoch and the service time of the n-th arrival, respectively. The marked point process is assumed to be independent of the queueing system. Arrivals are routed to one of the K queues by a sequence of i.i.d. Bernoulli random variables (r.v.'s) with parameter $\underline{p} \stackrel{\text{def}}{=} (p_1, p_2, \ldots, p_K)$. We assume that the buffer in each queue is infinite and thus that there is no blocking. We define $\underline{V}_t \stackrel{\text{def}}{=} (V_{1,t}, \ldots, V_{K,t})$ to be the amount of work in the queues at time t and assume that the system is empty at $t = 0$, i.e., $\underline{V}_0 = (0, \ldots, 0)$. The n-th customer departs from the system on the condition that all the customers $j, j \leq n$, have been served. The output process is thus in the same order as the input process. In the literature (see [13] and the listed references), this requirement is often called the *resequencing synchronization*. The response time of the n-th customer, T_n, is defined to be the elapsed time between arrival epoch and departure. Clearly, $T_n = \max_k V_{k,\tau_n^+}$, where τ_n^+ is the epoch that the n-th customer is routed to one of the K queues.

Theorem 3.6 *Let $\underline{p}$ and $\underline{q}$ be two routing vectors as defined in Example 2.22. Let $\underline{V}_t(\underline{p})$ and $\underline{V}_t(\underline{q})$ be the corresponding work load processes. Then $\underline{V}_t(\underline{q}) \prec_{wE_5(1,2)} \underline{V}_t(\underline{p})$.*

Corollary 3.7 *(i) $Ef(\underline{V}_t)$ is Schur convex in the routing vector $\underline{p}$ for all $f \in \mathcal{C}_5^{\uparrow}$. (ii) The expected total amount of work in the system at time t is minimized when $\underline{p} = (\frac{1}{K}, \dots, \frac{1}{K})$. (iii) The response time of every customer is minimized in the sense of increasing convex ordering when $\underline{p} = (\frac{1}{K}, \dots, \frac{1}{K})$.*

We note that if the arrival process is Poisson and the service times are exponentially distributed, then the system has a product form distribution for queues at steady state. In such a case, the $E_1^{\uparrow}$ ordering can be obtained as a special case of Shanthikumar [31] in which two random vectors consisting of independent random variables with proportional equilibrium rates are compared. In the case of Poisson arrivals and general service times, References [13][14] show that the waiting time in steady state is stochastically minimized when equal load is assigned to each queue. Tsoucas [35] shows that the departure process without resequencing is stochastically minimized under the assumption of general arrivals and exponential service times.

In the following, we establish the $E_5^{\uparrow}$ ordering for Markovian routing mechanisms, which is an extension of Bernoulli routing mechanisms.

Example 3.8 [Markovian Routing] Consider the Bernoulli routing problem in Example 3.5 with $K = 2$. Instead of assuming i.i.d. Bernoulli random vectors, the routing random vectors $\{\underline{X}_n, n \geq 1\}$ are assumed to be dependent on the outcome of the previous one. A customer is routed to the first queue with probability p_{11} (respectively p_{21}) if the previous one is routed to the first queue (respectively second) queue. Clearly, if $p_{11} = p_{21}$, the routing scheme is reduced to Bernoulli routing for two parallel queues. More precisely,

$$P(\underline{X}_{n+1} = \underline{e}_j | \underline{X}_n = \underline{e}_i) = p_{ij},$$

with $p_{11} + p_{12} = p_{21} + p_{22} = 1$. It is easy to show that the stationary distribution of this Markov chain is:

$$P(\underline{X}_n = \underline{e}_i) = 1 - \frac{1 - p_{ii}}{2 - (p_{11} + p_{22})} \stackrel{\text{def}}{=} p_i, \quad i = 1 \text{ and } 2.$$

Observe that p_i is the percentage of customers who are routed to the i-th queue. Define the correlation coefficient $\omega \stackrel{\text{def}}{=} p_{11} + p_{22} - 1$. Obviously, a stationary Markovian routing mechanisms is characterized by $\underline{p} \stackrel{\text{def}}{=} (p_1, p_2)$ and ω.

Theorem 3.9 *Let $(\underline{p}, \omega_1)$ and $(\underline{p}, \omega_2)$ be two stationary Markovian routing mechanisms defined in Example 3.8. Let $\underline{V}_t(\underline{p}, \omega_1)$ and $\underline{V}_t(\underline{p}, \omega_2)$ be the corresponding work load processes. If $\underline{p} = (1/2, 1/2)$ and $0 \leq \omega_1 \leq \omega_2 < 1$, then $Ef(\underline{V}_t(\underline{p}, \omega_1)) \leq Ef(\underline{V}_t(\underline{p}, \omega_2))$ for all $f \in \mathcal{C}_5^{\uparrow}$*

(in order to have uniform integrablity in our proof, we need the technical assumptions that $f(\underline{0}) > -\infty$ and $Ef(V_t^a \underline{1}) < \infty$, where V_t^a is the total amount of work arrived at the system by time t).

Corollary 3.10 *The stationary Markovian routing mechanism, $((1/2, 1/2), 0)$, minimizes $Ef(\underline{V}_t)$ for all $f \in \mathcal{C}_5^{\uparrow}$ among all nonnegative correlated ($\omega \geq 0$) Markovian routing mechanisms.*

To provide the intuition of the correlation coefficient ω, we note that the transition probabilities of the two-state Markov chain with $\underline{p} = (1/2, 1/2)$ can be rewirtten as $p_{ii} = \omega + (1-\omega)/2$, $i = 1$ and 2, and $p_{ij} = (1-\omega)/2$, $i \neq j$. This implies that if $\omega \geq 0$, the Markov chain stays in the same state with probability ω or chooses the next state with probability $1-\omega$ from the outcome of tossing a fair coin. Our proof for Theorem 3.9 is mainly based on this observation and the sample path criteria. For Markovian routing mechanisms with K states, similar results to those in Theorem 3.9 and Corollary 3.10 can be derived if there exists an $\omega \geq 0$ such that the transition probabilities are of the following form: $p_{ii} = \omega + (1-\omega)/K$, $i = 1, 2, \ldots, K$ and $p_{ij} = (1-\omega)/K$, $i \neq j$. The proof is similar to that for Theorem 3.9 and is omitted. The rest of this subsection is devoted to the proofs of Theorems 3.6 and 3.9 and Corollary 3.10.

Proof of Theorem 3.6: Analogous to the proof of Theorem 3.2, we construct $\underline{\hat{V}}_t(\underline{p})$, $\underline{\hat{V}}_t^{12}(\underline{p})$, $\underline{\hat{V}}_t(\underline{q})$ and $\underline{\hat{V}}_t^{12}(\underline{q})$ as follows:

Case 1: At the instant of the n-th arrival, we construct the routing vectors $\underline{\hat{X}}_n$, $\underline{\hat{X}}_n^{12}$, $\underline{\hat{Y}}_n$ and $\underline{\hat{Y}}_n^{12}$ as in Example 2.23 and

$$\underline{\hat{V}}_{\tau_n^+}(\underline{p}) = \underline{\hat{V}}_{\tau_n}(\underline{p}) + \underline{\hat{Y}}_n \cdot S_n,$$

$$\underline{\hat{V}}_{\tau_n^+}^{12}(\underline{p}) = \underline{\hat{V}}_{\tau_n}^{12}(\underline{p}) + \underline{\hat{Y}}_n^{12} \cdot S_n,$$

$$\underline{\hat{V}}_{\tau_n^+}(\underline{q}) = \underline{\hat{V}}_{\tau_n}(\underline{q}) + \underline{\hat{X}}_n \cdot S_n,$$

$$\underline{\hat{V}}_{\tau_n^+}^{12}(\underline{q}) = \underline{\hat{V}}_{\tau_n}^{12}(\underline{q}) + \underline{\hat{X}}_n^{12} \cdot S_n.$$

Case 2: For $\tau_n < t \leq \tau_{n+1}$, we construct

$$\hat{V}_{k,t}(\underline{p}) = (\hat{V}_{k,\tau_n^+}(\underline{p}) - (t - \tau_n))^+,$$

$$\hat{V}_{k,t}^{12}(\underline{p}) = (\hat{V}_{k,\tau_n^+}^{12}(\underline{p}) - (t - \tau_n))^+,$$

$$\hat{V}_{k,t}(\underline{q}) = (\hat{V}_{k,\tau_n^+}(\underline{q}) - (t - \tau_n))^+,$$

$$\hat{V}^{12}_{k,t}(\underline{q}) = (\hat{V}^{12}_{k,\tau_n^+}(\underline{q}) - (t - \tau_n))^+,$$

for $k = 1, \ldots, K$.

Since $(x)^+$ is increasing and convex and the arrivals are independent of the systems, applying Theorems 2.26 and 2.27 completes the proof. ∎

Proof of Theorem 3.9: From the proof of Theorem 3.6, it suffices to compare the arrival sequences. First, we show that the arrival sequences subject to the stationary Markovian routing mechanisms can be constructed by i.i.d. Bernoulli random vectors and random variables. Let $\{\underline{B}_n, n = 1, 2, \ldots\}$ be a sequence of i.i.d. Bernoulli routing vectors with $P(\underline{B}_n = \underline{e}_i) = \frac{1}{2}$, $i = 1$ and 2. Let $\{\delta_n, n = 1, 2, \ldots\}$ be a sequence of i.i.d. Bernoulli random variables, having the distribution, $P(\delta_n = 0) = \omega_1$ and $P(\delta_n = 1) = 1 - \omega_1$, and independent of $\{\underline{B}_n\}$. Construct a sequence of random vectors as follows:

$$\underline{X}_1 = \underline{B}_1,$$

$$\underline{X}_{n+1} = (1 - \delta_n)\underline{X}_n + \delta_n \underline{B}_{n+1}.$$

It was shown in [8] that the sequence of random vectors $\{\underline{X}_n, n = 1, 2 \ldots\}$ has the same joint distribution as that characterized by the stationary Markovian routing mechanism $((1/2, 1/2), \omega_1)$. Similarly, let $\{\zeta_n, n = 1, 2, \ldots\}$ be a sequence of i.i.d. Bernoulli random variables, having the distribution, $P(\zeta_n = 1) = \frac{1-\omega_2}{1-\omega_1}$ and $P(\zeta_n = 0) = 1 - \frac{1-\omega_2}{1-\omega_1}$, and independent of $\{\underline{B}_n, \delta_n\}$. Construct a sequence of random vectors as follows:

$$\underline{Y}_1 = \underline{B}_1,$$

$$\underline{Y}_{n+1} = (1 - \delta_n\zeta_n)\underline{Y}_n + \delta_n\zeta_n \underline{B}_{n+1}.$$

$\{\underline{Y}_n\}$ has the same joint distribution as that characterized by $((1/2, 1/2), \omega_2)$ since $\{\delta_n\zeta_n, n = 1, 2, \ldots\}$ is a sequence of i.i.d. Bernoulli random variables, having the distribution $P(\delta_n\zeta_n = 0) = \omega_2$ and $P(\delta_n\zeta_n = 1) = 1 - \omega_2$.

A direct comparison between $\{\underline{X}_n\}$ and $\{\underline{Y}_n\}$ is not obvious. We consider a family of auxiliary systems in which every adjacent pair differ only in a small time segment. Let $Z_0 \stackrel{\text{def}}{=} 0$ and $Z_{j+1} \stackrel{\text{def}}{=} inf\{n > Z_j : \delta_n = \zeta_n = 1\}$, $j \geq 0$. Note that $\{Z_j, j \geq 0\}$ is a sequence of regenerative points for both $\{\underline{X}_n\}$ and $\{\underline{Y}_n\}$. Let $\zeta_n(j) = \zeta_n$ if $n < Z_j$ and 1 otherwise. For each j, we construct a sequence of random vectors $\{\underline{X}_n(j)\}$ as follows:

$$\underline{X}_1(j) = \underline{B}_1,$$

$$\underline{X}_{n+1}(j) = (1 - \delta_n \zeta_n(j))\underline{X}_n(j) + \delta_n \zeta_n(j)\underline{B}_{n+1}.$$

Since $\zeta_n(j+1) = \zeta_n(j) = \zeta_n$ for $n < Z_j$ and $\zeta_n(j+1) = \zeta_n(j) = 1$ for $n \geq Z_{j+1}$, $\underline{X}_{n+1}(j) = \underline{X}_{n+1}(j+1)$ for $n < Z_j$ or $n \geq Z_{j+1}$. Moreover, we have $\{\underline{X}_n(0)\} =_{st} \{\underline{X}_n\}$ and $\lim_{j\to\infty} \{\underline{X}_n(j)\} =_{st} \{\underline{Y}_n\}$.

Now we show that either $\underline{X}_n(j) \prec_{wE_5(1,2)} \underline{X}_n(j+1)$ for all n or $\underline{X}_n(j) \prec_{wE_5(2,1)} \underline{X}_n(j+1)$ for all n is satisfied. Since $\underline{X}_{n+1}(j) = \underline{X}_{n+1}(j+1)$ for $n < Z_j$ or $n \geq Z_{j+1}$, we only have to consider the time segment in which they are different, i.e., $Z_j \leq n < Z_{j+1}$. Since $\{Z_j\}$ is a sequence of regenerative points, we can construct $\{\hat{\underline{X}}_n(j)\}$, $\{\hat{\underline{X}}_n^{12}(j)\}$, $\{\hat{\underline{Y}}_n(j)\}$ and $\{\hat{\underline{Y}}_n^{12}(j)\}$ such that they only differ when $Z_j \leq n < Z_{j+1}$. In this time segment, we use $\underline{B}_{n+1}^{12}$, which represent the two components in $\underline{B}_{n+1}$ interchanged, to construct $\{\hat{\underline{X}}_n^{12}(j)\}$ and $\{\hat{\underline{Y}}_n^{12}(j)\}$. The constructions are as follows:

$$\hat{\underline{X}}_1(j) = \hat{\underline{X}}_1^{12}(j) = \hat{\underline{Y}}_1(j) = \hat{\underline{Y}}_1^{12}(j) = \underline{B}_1,$$

$$\hat{\underline{X}}_{n+1}(j) = (1 - \delta_n \zeta_n(j))\hat{\underline{X}}_n(j) + \delta_n \zeta_n(j)\underline{B}_{n+1},$$

$$\begin{aligned} \hat{\underline{X}}_{n+1}^{12}(j) &= (1 - \delta_n \zeta_n(j))\hat{\underline{X}}_n^{12}(j) + \delta_n \zeta_n(j)\underline{B}_{n+1}, \text{ if } n < Z_j \text{ or } n \geq Z_{j+1}, \\ &= (1 - \delta_n \zeta_n(j))\hat{\underline{X}}_n^{12}(j) + \delta_n \zeta_n(j)\underline{B}_{n+1}^{12}, \text{ if } Z_j \leq n < Z_{j+1}, \end{aligned}$$

$$\hat{\underline{X}}_{n+1}(j+1) = (1 - \delta_n \zeta_n(j+1))\hat{\underline{X}}_n(j+1) + \delta_n \zeta_n(j+1)\underline{B}_{n+1},$$

$$\begin{aligned} \hat{\underline{X}}_{n+1}^{12}(j+1) &= (1 - \delta_n \zeta_n(j+1))\hat{\underline{X}}_n^{12}(j+1) + \delta_n \zeta_n(j+1)\underline{B}_{n+1}, \text{ if } n < Z_j \text{ or } n \geq Z_{j+1}, \\ &= (1 - \delta_n \zeta_n(j+1))\hat{\underline{X}}_n^{12}(j+1) + \delta_n \zeta_n(j+1)\underline{B}_{n+1}^{12}, \text{ if } Z_j \leq n < Z_{j+1}. \end{aligned}$$

By the property that $P(\underline{B}_{n+1} = \underline{e}_i) = \frac{1}{2}$, $i = 1$ and 2, we have $P(\underline{B}_{n+1}^{12} = \underline{e}_i) = \frac{1}{2}$, $i = 1$ and 2. Thus, (B1) is satisfied. It is also clear that $\hat{\underline{X}}_{n+1}(j) = \hat{\underline{X}}_{n+1}^{12}(j) = \hat{\underline{X}}_{n+1}(j+1) = \hat{\underline{X}}_{n+1}^{12}(j+1)$ for $n < Z_j$ or $n \geq Z_{j+1}$. Note that $\hat{\underline{X}}_{n+1}(j) + \hat{\underline{X}}_{n+1}^{12}(j) = \hat{\underline{X}}_{n+1}(j+1) + \hat{\underline{X}}_{n+1}^{12}(j+1) = (1,1)$ for $Z_j \leq n < Z_{n+1}$and thus (B3c) and (B4c) are satisfied. Observe that $\hat{\underline{X}}_{n+1}(j+1) = \underline{B}_{Z_j+1}$ and $\hat{\underline{X}}_{n+1}^{12}(j+1) = \underline{B}_{Z_j+1}^{12}$ for $Z_j \leq n < Z_{j+1}$ since $\zeta_n(j+1) = 1$ for $Z_j \leq n$. Thus, we have under the construction that $\underline{X}_n(j) \prec_{wE_5(1,2)} \underline{X}_n(j+1)$ for all n if $\underline{B}_{Z_j+1} = \underline{e}_1$ and $\underline{X}_n(j) \prec_{wE_5(2,1)} \underline{X}_n(j+1)$ for all n if $\underline{B}_{Z_j+1} = \underline{e}_2$.

Let $\underline{V}_t(j)$ be the work load process subject to the routing vectors $\{X_n(j)\}$. Analogous to the proof of Theorem 3.6, we have that either $\underline{V}_t(j) \prec_{wE_5(1,2)} \underline{V}_t(j+1)$ or $\underline{V}_t(j) \prec_{wE_5(2,1)} \underline{V}_t(j+1)$ is satisfied. It then follows from Theorem 2.24, $\underline{V}_t(j) \prec_{wE_5} \underline{V}_t(j+1)$. By the assumptions that $f(\underline{0}) > -\infty$ and $Ef(V_t^a \underline{1}) < \infty$, $\{f(\underline{V}_t(j))\}$ are uniformly integrable since $f(\underline{V}_t(j))$ is stochastically larger than $f(\underline{0})$ and stochastically smaller than $f(V_t^a \underline{1}))$. We have for all $f \in \mathcal{C}_5^{\uparrow}$

$$
\begin{aligned}
& Ef(\underline{V}_t((1/2,1/2),\omega_1)) = Ef(\underline{V}_t(0)) \leq \lim_{j\to\infty} Ef(\underline{V}_t(j)) \\
& = E \lim_{j\to\infty} f(\underline{V}_t(j)) = Ef(\underline{V}_t((1/2,1/2),\omega_2)). \quad \blacksquare
\end{aligned}
$$

Proof of Corollary 3.10: From Theorem 3.9, it suffices to show that $Ef(\underline{V}_t)$ under the routing mechanism $((1/2,1/2),\omega_1)$ is smaller than that under $((p_1,p_2),\omega_1)$, $p_1+p_2=1$, $p_1 \geq p_2$, for all f in $\mathcal{C}_5^{\uparrow}$. The construction of the routing vectors $\{\underline{X}_n\}$ for $((1/2,1/2),\omega_1)$ is the same as that in the proof of Theorem 3.9. Construct the routing vectors $\{\underline{Y}_n\}$ for $((p_1,p_2),\omega_1)$ as follows:

$$
\begin{aligned}
& \underline{Y}_1 = \underline{B}'_1, \\
& \underline{Y}_{n+1} = (1-\delta_n)\underline{Y}_n + \delta_n \underline{B}'_{n+1},
\end{aligned}
$$

where $\{\underline{B}'_n\}$ is a sequence of i.i.d. Bernoulli random vectors with $P(\underline{B}'_n = \underline{e}_1) = p_1$ and $P(\underline{B}'_n = \underline{e}_2) = p_2$. It then follows from Example 2.23 that $\underline{B}_n \prec_{wE_5(1,2)} \underline{B}'_n$. As a corollary of the closure properties in Theorems 2.26 and 2.27, we have $\underline{X}_n \prec_{wE_5(1,2)} \underline{Y}_n$ and $\underline{V}_t((1/2,1/2),\omega_1) \prec_{wE_5(1,2)} \underline{V}_t((p_1,p_2),\omega_1)$. $\blacksquare$

3.3 Probability Enumeration and Scheduling

In this subsection, we develop a probability enumeration method to establish the new partial ordering $E_5^{\uparrow}$. We consider scheduling problems with K identical processors in parallel and J jobs, with preemptions are allowed. We assume that all jobs are present at time 0 and there are no further arrivals. We also assume that all the jobs are independent and that the processing time of Job j is exponentially distributed with mean $\frac{1}{\mu_j}$. Let V_k be the amount of work assigned to Processor k. A policy is non-idling if (i) there are no idle processors when the number of uncompleted jobs is larger than or equal to the number of processors and (ii) all the uncompleted jobs are in service when the number of uncompleted jobs is less than the number of processors. It is a well known fact that the LEPT rule, which assigns priority in decreasing order of the expected processing time, is the best policy to balance the load of the system. In the following theorem, we provide a mathematical meaning for the fact.

Theorem 3.11 *Among all non-idling policies, the LEPT rule minimizes $Ef(V_1, \ldots, V_K)$ for all $f \in \mathcal{C}_5$. Preemptions are not needed.*

Corollary 3.12 *Let T_k be the last time that Processor k becomes idle. Clearly, $V_k = T_k$ under non-idling policies. Then the LEPT rule minimizes $Ef(T_1, \ldots, T_K)$ for all $f \in \mathcal{C}_5^{\uparrow}$ since inserting idle times cannot be better.*

The following examples are corollaries of Theorem 3.11 by choosing appropriate functions in $\mathcal{C}_5$.

Example 3.13 The expected sample variance,

$$Ef(V_1, \ldots, V_K) = E\frac{1}{K-1}\sum_{k=1}^{K}(V_k - \frac{1}{K}\sum_{k=1}^{K}V_k)^2$$

is in $\mathcal{C}_5$ since $\frac{\partial^2 f}{\partial V_k^2} = \frac{2}{K} > 0$ and $\frac{\partial^2 f}{\partial V_{k_1}\partial V_{k_2}} = -\frac{2}{K(K-1)} < 0$. To our best knowledge, this is a new result.

Example 3.14 The expected makespan [43] [4] [39],

$$Emax(V_1, \ldots, V_K),$$

is in $\mathcal{C}_5$ (Lemma 2.4(ii)). It should be noted that [43] and [39] allow arrivals.

Example 3.15 Let Z_t be the amount of work in the system at time t. The expected total amount of work accumulated up to infinity,

$$E\int_0^{\infty} Z_t dt = E\sum_{k=1}^{K}\frac{V_k^2}{2},$$

is in $\mathcal{C}_5$. This is a corollary of [16] [17] when $c_j = \frac{1}{\mu_j}$. This is also a corollary of [7] when there are no arrivals.

Example 3.16 The expected time to the first idle processor [40],

$$Emin(V_1, \ldots, V_K),$$

is symmetric, L-superadditive and concave in each variable. Thus, the LEPT rule maximized the expected time to the first idle processor.

In the literature, the LEPT rule has received a great deal of attention. Weiss and Pinedo [40] provides a unified criterion for optimal static policies in minimizing (maximizing) various costs (rewards) in scheduling exponential jobs in parallel processors. The objective functions they considered are formulated as $E\int_0^\infty g(Q_{1,t},\ldots,Q_{J,t})dt$, where $Q_{j,t}=1$ if Job j is still in the system and $Q_{j,t}=0$ otherwise. They show that the LEPT rule minimizes their objectives if $g(Q_{1,t},\ldots,Q_{J,t})=h(\sum_{j=1}^{J}Q_{j,t})$ and h is increasing and convex. Example 3.14 and 3.16 fall in their category; however, the other two cannot be formulated by their objectives (note that they also consider processors with different speeds).

Our proof of Theorem 3.11 consists of the following lemmas which are essentially based on the proof in [7] and the properties for functions $f\in\mathcal{C}_5$ (Lemma 3.20). Without loss of generality, we assume $\mu_1\le\mu_2\le\ldots\le\mu_J$. From the well known uniformization technique (see [18]), it follows that continuous-time optimization problems can be transformed into discrete-time ones. Generate a Poisson process with rate $\Delta\stackrel{\text{def}}{=}K\mu_J$. Let $\{\tau_n,n\ge1\}$ be the arrival epochs of the Poisson process and let $\tau_0\stackrel{\text{def}}{=}0$. In standard coupling, these epochs are possible epochs of job completions. By the memoryless property of exponential processing times, τ_n's cover all the decision epochs needed to be considered. That is, the states are the sets of uncompleted jobs at τ_n's and the optimal policy is only a sequence of decisions, $(\sigma_1,\sigma_2,\sigma_3,\ldots)$. At the i-th decision epoch, all the jobs being processed are preempted and a set of jobs, A^{σ_i}, determined by the decision rule σ_i according to the current state, are assigned into service (Jobs preempted may be reassigned into service and in practice there is no preemption for those jobs). Let π be the decision rule that assigns the K jobs with the longest expected processing times among the remaining jobs at a decision epoch. Clearly, the LEPT rule corresponds to the sequence of decision rules $(\pi,\pi,\pi,\ldots)$. Using a standard forward induction argument, the following lemma is obvious ([7]).

Lemma 3.17 *The LEPT rule is optimal if $(\pi,\pi,\pi,\ldots)$ is better than $(\sigma_1,\pi,\pi,\ldots)$ for any other non-idling policy σ_1 and any initial states ("better" means that $Ef(V_1,\ldots,V_K)$ is smaller for $f\in\mathcal{C}_5$).*

To show that $(\pi,\pi,\pi,\ldots)$ is better than $(\sigma_1,\pi,\pi,\ldots)$, we consider two policies $\underline{\pi}^\beta\stackrel{\text{def}}{=}(\pi^\beta,\pi,\pi,\ldots)$ and $\underline{\pi}^\alpha\stackrel{\text{def}}{=}(\pi^\alpha,\pi,\pi,\ldots)$ that differ from each other in only one job assignment at time 0, say Jobs β and α. Assume that Jobs β and α are assigned into service at time 0 under π^β and π^α, respectively, and that the expected processing time of Job β is longer than that of Job α, i.e., $\frac{1}{\mu_\beta}>\frac{1}{\mu_\alpha}$ $(\beta<\alpha)$. Using a pairwise interchange argument on jobs, we have the following lemma ([7]).

Lemma 3.18 *$(\pi, \pi, \pi, \ldots)$ is better than $(\sigma_1, \pi, \pi, \ldots)$ if $\underline{\pi}^\beta$ is better than $\underline{\pi}^\alpha$.*

Let U be the state at a decision epoch. Then U is the set of uncompleted jobs and $U \subseteq U^0 \stackrel{\text{def}}{=} \{1, 2, \ldots, J\}$. We denote the number of uncompleted jobs at the state U by $|U|$. Let $U_j \stackrel{\text{def}}{=} U - \{j\}$ and $U_{j,i} \stackrel{\text{def}}{=} U - \{j, i\}$. Let V_k^α and V_k^β be the amount of work assigned to processor k under $\underline{\pi}^\alpha$ and $\underline{\pi}^\beta$, respectively. Let $\underline{V}^\alpha \stackrel{\text{def}}{=} (V_1^\alpha, \ldots, V_K^\alpha)$. $\underline{V}^\beta$ is defined accordingly. Let A^α and A^β be the set of jobs that are assigned into service under the policies π^α and π^β at time 0, respectively. If $|U^0| = J \leq K$, then the results are the same for all non-idling policies since all the jobs are served at time 0. Thus, we only have to consider the case $|U^0| > K$. Let $V_k(U)$ be the amount of work assigned to processor k under the LEPT rule for a problem with the initial set of jobs U. Let $\underline{V}(U) \stackrel{\text{def}}{=} (V_1(U), \ldots, V_K(U))$. Analogous to Kolmogorov's backward equation, we have for any function f that

$$Ef(\underline{V}^\alpha) = \sum_{j \in A^\alpha} \frac{\mu_j}{\Delta} Ef(\underline{V}(U_j^0) + \tau_1 \underline{1}) + \frac{\Delta - \sum_{j \in A^\alpha} \mu_j}{\Delta} Ef(\underline{V}(U^0) + \tau_1 \underline{1}),$$

and

$$Ef(\underline{V}^\beta) = \sum_{j \in A^\beta} \frac{\mu_j}{\Delta} Ef(\underline{V}(U_j^0) + \tau_1 \underline{1}) + \frac{\Delta - \sum_{j \in A^\beta} \mu_j}{\Delta} Ef(\underline{V}(U^0) + \tau_1 \underline{1}),$$

where $\underline{1} \stackrel{\text{def}}{=} (1, 1, \ldots, 1)$.

Lemma 3.19 *$\underline{\pi}^\beta$ is better than $\underline{\pi}^\alpha$ if*

$$\mu_\alpha(Ef(\underline{V}(U_\alpha^0) + \tau_1 \underline{1}) - Ef(\underline{V}(U^0) + \tau_1 \underline{1})) - \mu_\beta(Ef(\underline{V}(U_\beta^0) + \tau_1 \underline{1}) - Ef(\underline{V}(U^0) + \tau_1 \underline{1})) \geq 0.$$

for all $f \in \mathcal{C}_5$.

Proof: Since A^α and A^β differ only in Jobs α and β,

$$\begin{aligned} Ef(\underline{V}^\alpha) - Ef(\underline{V}^\beta) = \frac{1}{\Delta}\Big(&\mu_\alpha(Ef(\underline{V}(U_\alpha^0) + \tau_1 \underline{1}) \\ &-Ef(\underline{V}(U^0) + \tau_1 \underline{1})) - \mu_\beta(Ef(\underline{V}(U_\beta^0) + \tau_1 \underline{1}) - Ef(\underline{V}(U^0) + \tau_1 \underline{1}))\Big). \blacksquare \end{aligned}$$

One observe that under the LEPT rule (see [7] for details)

$$\begin{aligned} &\mu_\alpha(Ef(\underline{V}(U_\alpha^0) + \tau_1 \underline{1}) - Ef(\underline{V}(U^0) + \tau_1 \underline{1})) \\ &-\mu_\beta(Ef(\underline{V}(U_\beta^0) + \tau_1 \underline{1}) - Ef(\underline{V}(U^0) + \tau_1 \underline{1})) \end{aligned}$$

can be written as a linear combination of

$$\mu_\alpha(Ef(\underline{V}(U^0_{\alpha,i}) + \tau_2\underline{1}) - Ef(\underline{V}(U^0_i) + \tau_2\underline{1}))$$
$$-\mu_\beta(Ef(\underline{V}(U^0_{\beta,i}) + \tau_2\underline{1}) - Ef(\underline{V}(U^0_i) + \tau_2\underline{1}))$$

with nonnegative coefficients summing up to one if $U^0 > K$. Further expansion will eventually reduce the number of uncompleted jobs to the number of processors, K. Thus, it suffices to show that for any state $|U| = K$ and all $n \geq 1$,

$$\mu_\alpha(Ef(\underline{V}(U_\alpha) + \tau_n\underline{1}) - Ef(\underline{V}(U) + \tau_n\underline{1}))$$
$$-\mu_\beta(Ef(\underline{V}(U_\beta) + \tau_n\underline{1}) - Ef(\underline{V}(U) + \tau_n\underline{1})) \geq 0.$$

In such a case, $\underline{V}(U)$ are the remaining processing times of the jobs in U, and so are $\underline{V}(U_\alpha)$ and $\underline{V}(U_\beta)$. Let X_j be the remaining processing time of Job j in U. By the memoryless property, X_j's are independent and exponentially distributed with mean $\frac{1}{\mu_j}$. Let $h(X_\alpha, X_\beta) \stackrel{\text{def}}{=} E(f(\underline{V}(U) + \tau_n\underline{1})|X_\alpha, X_\beta)$ be the conditional expectation on X_α and X_β. Clearly, we have $E(f(\underline{V}(U_\alpha) + \tau_n\underline{1})|X_\alpha, X_\beta) = h(0, X_\beta)$ and $E(f(\underline{V}(U_\beta) + \tau_n\underline{1})|X_\alpha, X_\beta) = h(X_\alpha, 0)$. Since $f \in \mathcal{C}_5$, we have $h \in \mathcal{C}_5$; thus, the proof for Theorem 3.11 is concluded if we can establish the following lemma.

Lemma 3.20 *If X_α and X_β are two independent exponential random variables and $EX_\alpha \leq EX_\beta$, then*

$$\frac{Eh(0, X_\beta) - Eh(X_\alpha, X_\beta)}{EX_\alpha} \geq \frac{Eh(X_\alpha, 0) - Eh(X_\alpha, X_\beta)}{EX_\beta}$$

for all $h \in \mathcal{C}_5$.

Proof: Let Z_1 and Z_2 be two independent and exponentially distributed random variables with mean 1. Let $g(a, b) \stackrel{\text{def}}{=} Eh(aZ_1, bZ_2)$, where $a, b \in \mathbb{R}$. It is easy to check that $g \in \mathcal{C}_5$ since $h \in \mathcal{C}_5$ and Z_1 and Z_2 are nonnegative i.i.d. random variables. Thus, it suffices to show that for $a \leq b$,

$$\frac{g(0, b) - g(a, b)}{a} \geq \frac{g(a, 0) - g(a, b)}{b}.$$

By L-subadditivity and symmetry,

$$g(a, 0) - g(a, b) \leq g(b, 0) - g(b, b) = g(0, b) - g(b, b).$$

By convexity in each variable,

$$\frac{g(0, b) - g(a, b)}{a} \geq \frac{g(0, b) - g(b, b)}{b}. \quad \blacksquare$$

4 Multivariate Stochastic Majorization

In Section 2, we defined the notion of stochastic majorization for random vectors. The notion is developed further for comparing random matrices. Our approach is based on rowwise majorization ([19]) and multidimensional stochastic convexity ([8, 9]). Several examples are given as applications of multivariate stochastic majorization which extend the results in Section 3 to tandem queues and queues with finite buffers.

Let $\underset{\sim}{x}$ and $\underset{\sim}{y}$ be two $\mathbb{R}^{LK}$ matrices. In the following, we review the definition of *rowwise majorization* ([19], Definition 11.A.6).

Definition 4.1 *The matrix $\underset{\sim}{x}$ is said to be* **rowwise majorized** *by $\underset{\sim}{y}$ ($\underset{\sim}{x} \prec_{:row} \underset{\sim}{y}$) if $\underline{e}_l \underset{\sim}{x} \prec \underline{e}_l \underset{\sim}{y}$ for $l = 1, \ldots, L$, where $\underline{e}_l$ is the row vector with the l-th component 1 and the others 0. In other words, every row of $\underset{\sim}{x}$ is majorized by the corresponding row of $\underset{\sim}{y}$. Let $\underset{\sim}{A}$ be a lower triangular $\mathbb{R}^{LL}$ matrix with $A_{ij} = 1$ if $i \geq j$ and $A_{ij} = 0$ if $i < j$. The matrix $\underset{\sim}{x}$ is said to be* **rowwise majorized** *for each* **partial sum** *by $\underset{\sim}{y}$ ($\underset{\sim}{x} \prec_{:ps} \underset{\sim}{y}$) if $\underset{\sim}{A}\underset{\sim}{x} \prec_{:row} \underset{\sim}{A}\underset{\sim}{y}$.*

Additional definitions for multivariate majorization can be found in [19]. In order to define multivariate stochastic majorization orderings, we consider the following class of functions ($f : \mathbb{R}^{LK} \to \mathbb{R}$):

$$\tilde{\mathcal{C}}_5 \ (\tilde{\mathcal{C}}_5^{\uparrow}) \stackrel{\text{def}}{=} \{f : f(\underset{\sim}{x}) = \textstyle\sum_{i=1}^{L} f_i(e_i \underset{\sim}{x}) \text{ and } f_i \text{ is in } \mathcal{C}_5 \ (\mathcal{C}_5^{\uparrow})\}.$$

Analogous to Definition 2.9, we define the following stochastic majorization orderings for random matrices based on rowwise majorization.

Definition 4.2 **(i)** *$E_{5:row}$ ($E_{5:row}^{\uparrow}$) : $\underset{\sim}{X} \prec_{E_{5:row}} (\prec_{wE_{5:row}}) \underset{\sim}{Y}$ if $Ef(\underset{\sim}{X}) \leq Ef(\underset{\sim}{Y})$ for all $f \in \tilde{\mathcal{C}}_5$ ($\tilde{\mathcal{C}}_5^{\uparrow}$).*

(ii) *$E_{5:ps}$ ($E_{5:ps}^{\uparrow}$) : $\underset{\sim}{X} \prec_{E_{5:ps}} (\prec_{wE_{5:ps}}) \underset{\sim}{Y}$ if $\underset{\sim}{A}\underset{\sim}{X} \prec_{E_{5:row}} (\prec_{wE_{5:row}}) \underset{\sim}{A}\,\underset{\sim}{Y}$.*

The other stochastic majorization orderings can be defined accordingly, but they will not be pursued further in this paper. In the following, we provide the sample path criteria for these orderings.

Definition 4.3 **(i)** $E_5^{\uparrow}{}_{(i,j):row}$: $\underset{\sim}{X} \prec_{wE_5(i,j):row} \underset{\sim}{Y}$ *if for some* $i \neq j$, *there exist four random matrices* $\hat{\underset{\sim}{X}}$, $\hat{\underset{\sim}{X}}^{ij}$, $\hat{\underset{\sim}{Y}}$ *and* $\hat{\underset{\sim}{Y}}^{ij}$, *where* $\underset{\sim}{X}^{ij}$ ($\underset{\sim}{Y}^{ij}$) *is the interchange of columns* i *and* j *of the matrix* $\underset{\sim}{X}$ ($\underset{\sim}{Y}$), *on the same probability space such that*

(B1) $\hat{\underset{\sim}{X}} =_{st} \underset{\sim}{X}$, $\hat{\underset{\sim}{X}}^{ij} =_{st} \underset{\sim}{X}^{ij}$, $\hat{\underset{\sim}{Y}} =_{st} \underset{\sim}{Y}$ *and* $\hat{\underset{\sim}{Y}}^{ij} =_{st} \underset{\sim}{Y}^{ij}$,

(B2) $\hat{\underset{\sim}{X}}\underline{e}_k^* = \hat{\underset{\sim}{X}}^{ij}\underline{e}_k^* = \hat{\underset{\sim}{Y}}\underline{e}_k^* = \hat{\underset{\sim}{Y}}^{ij}\underline{e}_k^*$ *a.s., for all* $k \neq i, j$,

(B3i) $\hat{\underset{\sim}{Y}}\underline{e}_i^* \geq max(\hat{\underset{\sim}{X}}\underline{e}_i^*, \hat{\underset{\sim}{X}}^{ij}\underline{e}_i^*)$ *a.s.*,

(B3c) $(\hat{\underset{\sim}{X}} + \hat{\underset{\sim}{X}}^{ij})\underline{e}_i^* \leq (\hat{\underset{\sim}{Y}} + \hat{\underset{\sim}{Y}}^{ij})\underline{e}_i^*$ *a.s.*,

(B4i) $\hat{\underset{\sim}{Y}}^{ij}\underline{e}_j^* \geq max(\hat{\underset{\sim}{X}}\underline{e}_j^*, \hat{\underset{\sim}{X}}^{ij}\underline{e}_j^*)$ *a.s., and*

(B4c) $(\hat{\underset{\sim}{X}} + \hat{\underset{\sim}{X}}^{ij})\underline{e}_j^* \leq (\hat{\underset{\sim}{Y}} + \hat{\underset{\sim}{Y}}^{ij})\underline{e}_j^*$ *a.s.*

(ii) $E_5^{\uparrow}{}_{(i,j):ps}$: $\underset{\sim}{X} \prec_{wE_5(i,j):ps} \underset{\sim}{Y}$ *if for some* $i \neq j$, *there exist four random matrices* $\hat{\underset{\sim}{X}}$, $\hat{\underset{\sim}{X}}^{ij}$, $\hat{\underset{\sim}{Y}}$ *and* $\hat{\underset{\sim}{Y}}^{ij}$ *on the same probability space such that (B1), (B2), (B3i) and (B4i) are satisfied and (B3c) and (B4c) are replaced by*

(B3c′) $\underset{\sim}{A}(\hat{\underset{\sim}{X}} + \hat{\underset{\sim}{X}}^{ij})\underline{e}_i^* \leq \underset{\sim}{A}(\hat{\underset{\sim}{Y}} + \hat{\underset{\sim}{Y}}^{ij})\underline{e}_i^*$ *a.s. and*

(B4c′) $\underset{\sim}{A}(\hat{\underset{\sim}{X}} + \hat{\underset{\sim}{X}}^{ij})\underline{e}_j^* \leq \underset{\sim}{A}(\hat{\underset{\sim}{Y}} + \hat{\underset{\sim}{Y}}^{ij})\underline{e}_j^*$ *a.s.*

Based on a standard pairwise comparison argument and the technique used in the proof of Theorem 2.24, it is easy to prove that these sample path criteria imply the corresponding orderings.

Theorem 4.4 *If* $\underset{\sim}{X} \prec_{wE_5(i,j):row} (\prec_{wE_5(i,j):ps}) \underset{\sim}{Y}$, *then* $\underset{\sim}{X} \prec_{wE_5:row} (\prec_{wE_5:ps}) \underset{\sim}{Y}$.

It is also clear that these sample path criteria still have the following closure properties (cf. Theorems 2.25-2.27).

Theorem 4.5 *The relations* $E_5^{\uparrow}{}_{(i,j):row}$ *and* $E_5^{\uparrow}{}_{(i,j):ps}$ *are preserved under mixtures and convolutions.*

Theorem 4.6 (Closure property of Composition) *Let* $(g \odot \underset{\sim}{x})_{ij} \stackrel{\text{def}}{=} g(x_{ij})$ *be a function from* $\mathbb{R}^{LK}$ *to* $\mathbb{R}^{LK}$*. If* $\underset{\sim}{X} \prec_{wE_5(i,j):row} (\prec_{wE_5(i,j):ps}) \underset{\sim}{Y}$ *and* g *is increasing and convex, then* $g \odot \underset{\sim}{X} \prec_{wE_5(i,j):row} g \odot \underset{\sim}{Y}$ $(\underset{\sim}{A}^{-1} g \odot (\underset{\sim}{A}\underset{\sim}{X}) \prec_{wE_5(i,j):ps} \underset{\sim}{A}^{-1} g \odot (\underset{\sim}{A}\underset{\sim}{Y}))$*, where* $\underset{\sim}{A}^{-1}$ *is the inverse of* $\underset{\sim}{A}$*.*

In the following, we provide examples that satisfy the sample path criteria.

Example 4.7 [Fork-Join Queues] Consider the fork-join queues in Example 3.1. Let $V_{n,k,t}$ be the amount of work, originated from the n-th job, in the k-th queue at time t. Let $\underset{\sim}{V}_t$ be the random matrix consisting of $V_{n,k,t}$. Let $\underline{p}$ and $\underline{q}$ be the vectors of processor speeds considered in the proof of Theorem 3.3. Let $\underset{\sim}{V}_t(\underline{p})$ and $\underset{\sim}{V}_t(\underline{q})$ be the corresponding "work load" processes. Using the proof of Theorem 3.3 with the closure properties replaced by Theorems 4.5 and 4.6, our results in Section 3 can be extended as follows:

Theorem 4.8 *With the assumption for task processing times in Theorem 3.3,* $\underset{\sim}{V}_t(\underline{p}) \prec_{wE_5(2,1):ps} \underset{\sim}{V}_t(\underline{q})$ *for all* $t \geq 0$*.*

Example 4.9 Consider a queueing system consisting of K parallel L-stage tandem queues. Assume the service times of customers at all stages are independent and exponentially distributed r.v.'s with mean 1. The speed (service rate) at the l-th stage is $\mu_l(n)$ when n customers are present. Assume that $\mu_l(n)$ is increasing concave and bounded. For instance, the speed of a queue with m servers is $\min(m, n)$. Arrivals are routed to one of the K queues by a Bernoulli routing mechanism (Markovian routing for $K = 2$). Let $Q_{l,k,t}$ be the number of customers in the l-th stage of the k-th tandem queue at time t and $\underset{\sim}{Q}_t$ be the matrix consisting of $Q_{l,k,t}$. Let $\underline{p}$ and $\underline{q}$ be two vectors considered in Example 2.22 and $\underset{\sim}{Q}_t(\underline{p})$ and $\underset{\sim}{Q}_t(\underline{q})$ be the corresponding queue length processes. Then $\underset{\sim}{Q}_t(\underline{q}) \prec_{wE_5(1,2):ps} \underset{\sim}{Q}_t(\underline{p})$. A similar result holds for Markovian routing mechanism.

Proof: By the construction in Example 2.23 and the closure property under convolution (Theorem 4.5), the relation $E_5^{\uparrow}(1,2){:}ps$ is preserved at the instants of arrivals. Between two successive arrivals, each tandem queue behaves *independently.* The relation is then preserved (cf. Example 2.22) since the tandem queue considered is $SICX - PS(sp)$ for all t ([8], Lemma 3.3). Recall that a stochastic vector process $\{\underline{X}_t, t \geq 0\}$ is $SICX - PS(sp)$ for all t if given four initial conditions $\underline{X}_0(i)$, $i = 1, 2, 3, 4$, satisfying

(D1) $\underline{X}_0^*(4) \geq max(\underline{X}_0^*(2), \underline{X}_0^*(3))$ a.s. and

(D2) $\underset{\sim}{A}(\underline{X}_0^*(2) + \underline{X}_0^*(3)) \leq \underset{\sim}{A}(\underline{X}_0^*(4) + \underline{X}_0^*(1))$ a.s.,

where $\underline{X}_0^*(i)$ is the transpose of $\underline{X}_0(i)$, then there exist $\hat{\underline{X}}_t(i)$, $i = 1, 2, 3, 4$ on the same probability space such that $\hat{\underline{X}}_t(i) =_{st} \underline{X}_t(i)$ and $\hat{\underline{X}}_t(i)$ satisfy (D1) and (D2). ■

We note that similar results can be obtained for queues with finite buffer. Consider the Bernoulli routing model described in Example 3.5 with these changes: in each queue the buffer size is finite and the speed depends on the number of customers present. Assume that the speed is increasing, concave and bounded and the service times are independent and exponentially distributed random variables, as in Example 4.9. An arrival routed to queue k is blocked if the buffer is full. Let $B_{k,t}$ be the number of customers routed to queue k and blocked by time t. Let $Q_{k,t}$ be the number of customers in queue k at time t. Let $\underset{\sim}{X}_t$ be the matrix process with $X_{1,k,t} = B_{k,t}$ and $X_{2,k,t} = Q_{k,t}$. Based on the proof in [8], Lemma 3.5 and the argument used in the previous example, it can be shown that $Ef(\underset{\sim}{A}\underset{\sim}{X}_t)$ is minimized for all $f \in \tilde{\mathcal{C}}_5^{\uparrow}$ when $\underline{p} = (\frac{1}{K}, \ldots, \frac{1}{K})$. As corollaries, the number of customers blocked by time t is minimized in expectation and the number of departures by time t are maximized in expectation when $\underline{p} = (\frac{1}{K}, \ldots, \frac{1}{K})$.

5 Conclusions

In this paper, we defined a new stochastic majorization ordering based on the class of functions that are symmetric, L-subadditive and convex in each variable and showed that this ordering is equivalent to the other four orderings in [19] when comparing deterministic vectors, but it is different when comparing random vectors. Two methods have been provided to establish this new ordering, sample path criteria and probability enumeration. Our framework is applied to various fork-join queues, routing and scheduling problems. We have also extended the notion to multivariate stochastic majorization.

It is interesting to note that it is usually not easy to establish the stochastic majorization ordering E_1 unless very strong assumptions are made. It is also not easy to establish the ordering E_4 because it lacks closure properties. The ordering E_3 is less interesting than the others since no information on the joint distribution is provided. Perhaps for these reasons, the ordering E_2 has been studied extensively in the literature [19] to bridge the gap. However, the sample path criteria in Definition 2.33 for E_2 may not be easy to establish, since the closure properties require not only matching i and j but also α. Thus, E_5 seems to be the appropriate choice for stochastic majorization ordering.

Acknowledgments: The author would like to thank Prof. X.L. Chao, Prof. A.M. Makowski, Dr. R.D. Nelson, Prof. M. Pinedo, and Prof. D.D. Yao for their valuable advice and discussion. The author would also like to thank the referee for his careful comments which enhanced the presentation of this work.

References

[1] F. Baccelli and A.M. Makowski, "Queueing Models for Systems with Synchronization Constraints," *Proc. of the IEEE*, vol. 77, pp. 138-161, 1989.

[2] F. Baccelli, A.M. Makowski and A. Shwartz, "The Fork-Join Queue and Related Systems with Synchronization Constraints: Stochastic Ordering and Computable Bounds," *Adv. Appl. Prob.* 21, pp. 629-660, 1989.

[3] R.E. Barlow and F. Proschan, *Stochastic Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, Reading, M.A., 1975.

[4] J. Bruno, P. Downey and G.N. Frederickson, "Sequencing Tasks with Exponential Service Times to Minimize the Expected Flow Time or Makespan," *J. Ass. Comp. Mach.*, vol. 28, pp. 100-113, 1981.

[5] C.S. Chang and M. Pinedo, "Bounds and Inequalities for Single Server Loss Systems," *Queueing Systems*, vol 6, pp. 425-436, 1990.

[6] C.S. Chang, X.L. Chao and M. Pinedo, "Integration of Discrete-Time Correlated Markov Processes in a TDM System: Structural Results," *Probability in the Engineering and Informational Sciences*, vol. 4, pp. 29-56, 1990.

[7] C.S. Chang, X.L. Chao and M. Pinedo, "On the Optimality of the LEPT Rule and the $c\mu$-Rule for Machines in Parallel," submitted to *J. Appl. Prob.*

[8] C.S. Chang, X.L. Chao and M. Pinedo, "Monotonicity Results for Queues with Doubly Stochastic Poisson Arrivals: Ross's Conjecture," *Adv. Appl. Prob*, vol. 23, pp. 210-228, 1991.

[9] C.S. Chang, X.L. Chao, M. Pinedo and J.G. Shanthikumar, "Stochastic Convexity for Multidimensional Processes and Its Applications," submitted for publications.

[10] Y.-S. Chow and H. Teicher, *Probability Theory, Independence, Interchangeability, Martingales*, New York: Springer-Verlag, 1988.

[11] A. Ephremides, P. Varaiya, and J. Walrand, "A Simple Dynamic Routing Problem," *IEEE Tran. Automat. Contr.*, vol. AC-25, pp. 690-693, 1980.

[12] J.D. Esary, F. Proschan and D.W. Walkup, "Association of Random Variables with Applications," *Ann. Math. Statist.*, vol. 38, pp. 1466-1474, 1967.

[13] L. Gun, "Performance evaluation and optimization of parallel systems with synchronization," Ph.D dissertation, University of Maryland, College Park, 1989.

[14] L. Gun, A. Jean-Marie, A.M. Makowski and Tedijanto, "Convexity Results for Queues With Bernoulli Mechanisms," preprint, 1990.

[15] G.H. Hardy, J.E. Littlewood and G. Polya, *Inequalities.* Cambridge Univ. Press, London and New York, 1952.

[16] T. Kampke, "On the Optimality of Static Priority Policies in Stochastic Scheduling on Parallel Machines," *J. Appl. Prob.*, vol. 24, pp. 430-448, 1987.

[17] T. Kampke, "Optimal Scheduling of Jobs with Exponential Service Times on Identical Parallel Processors," *Oper. Res.*, vol. 37, pp. 126-133, 1989.

[18] T. Keilson, *Markov Chain Models-Rarity and Exponentiality.* Springer-Verlag, New York, 1979.

[19] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press, 1979.

[20] L.E. Meester and J.G. Shanthikumar, "Stochastic Convexity on General Space," preprint, 1990.

[21] R. Menich, "Optimality of Shortest Queue Routing for Dependent Service Stations," *Proc. 26th Conference on Decision and Control*, pp. 1069-1072, 1987.

[22] R. Nelson and A.N. Tantawi, "Approximation analysis of fork/join synchronization in parallel queues," *IEEE Trans. Comput.*, vol. C-37, pp. 739-743, Jun. 1988.

[23] S.C. Niu, "Bounds for the Expected Delays in Some Tandem Queues," *J. Appl. Prob.*, vol. 17, pp. 831-838, 1980.

[24] Y. Rinott, "Multivariate Majorization and Rearrangement Inequalities with Some Applications to Probability and Statistics," *Israel J. Math.*, vol. 15, pp. 60-77, 1973.

[25] T. Rolski, "Queues with Non-stationary Input Stream: Ross's Conjecture," *Adv. Appl. Prob.*, vol. 13, pp. 603-618, 1981.

[26] T. Rolski, "Upper Bounds for Single Server Queues with Doubly Stochastic Poisson Arrivals," *Math. of Oper. Res.*, vol. 11, pp. 442-450, 1986.

[27] S.M. Ross, *Stochastic Processes.* New York: J. Wiley & Sons, 1983.

[28] M. Shaked and J.G. Shanthikumar, "Temporal Stochastic Convexity and Concavity," *Stochastic Processes and Their Applications*, vol. 27, pp. 1-20, 1988.

[29] M. Shaked and J.G. Shanthikumar, "Stochastic Convexity and Its Applications," *Adv. Appl. Prob.*, vol. 20, pp. 427-446, 1988.

[30] M. Shaked and Y.L. Tong, "Some Partial Orderings of Exchangeable Random Variables by Positive Dependence," *J. Multivar. Anal.*, vol. 17, pp. 333-349, 1985.

[31] J.G. Shanthikumar, "Stochastic Majorization of Random Variables with Proportional Equilibrium Rates," *Adv. Appl. Prob.*, Vol. 19, pp.854-872, 1987.

[32] J.G. Shanthikumar and D.D. Yao, "Spatiotemporal Convexity of Stochastic Processes and Applications," submitted for publication.

[33] J.G. Shanthikumar and D.D. Yao, "Strong Stochastic Convexity: Closure Properties and Application," *J. Appl. Prob.*, vol. 28, pp. 131-145, 1991.

[34] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, Berlin: J. Wiley & Sons, 1983.

[35] P. Tsoucas, "Stochastic monotonicity of the output process of parallel queues," preprint, 1989.

[36] R.R. Weber, "On the Optimal Assignment of Customers to Parallel Servers," *J. Appl. Prob.*, vol. 15, pp. 406-413, 1978.

[37] R.R. Weber, "On the marginal Benefit of Adding Servers to G/GI/m Queues," *Management Sci.*, vol. 25, pp. 946-951, 1979.

[38] R.R. Weber, "A Note on Waiting Time in Single Server Queues," *Operations Res.*, vol. 31, pp. 950-951, 1983.

[39] G. Weiss, "Multiserver Stochastic Scheduling," Deterministic and Stochastic Scheduling, Dempster, Lenstra & Rinnooy Kan, pp. 157-179, 1982.

[40] G. Weiss, and M. Pinedo, "Scheduling Tasks with Exponential Service Times On Non-Identical Processors to Minimize Various Cost Functions," *J. Appl. Prob.*, vol. 17, pp. 187-202, 1980.

[41] W. Winston, "Optimality of the Shortest Line Discipline," *J. Appl. Prob.*, vol. 14, pp. 181-189, 1977.

[42] W. Whitt, "Deciding Which Queue to Join: Some Conterexamples," *Operations Res.*, vol. 34, pp. 55-62, 1986.

[43] L. Van Der Heyden, "Scheduling Jobs with Exponential Processing and Arrival Times on Identical Processors so as to Minimize the Expected Makespan," *Math. of Oper. Res.*, vol. 6, pp. 305-312, 1981.