# Rearrangement, Majorization and Stochastic Scheduling

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

David D. Yao
Department of Industrial Engineering and Operations Research
Columbia University, New York, NY 10027-6699

(October, 1990; revision: April, 1992)

**Abstract**

Rearrangement inequalities, such as the classical Hardy-Littlewood-Polya inequality and the more general Day's inequality, and related majorization results are often useful in solving scheduling problems. Among other things, they are essential for pairwise interchange arguments. Motivated by solving stochastic scheduling problems, we develop stochastic versions of Day's inequality, over both unrestricted and restricted (specifically, one-cycle) permutations. These lead to a general and unified approach, which we apply to solve the stochastic versions of several classical deterministic scheduling problems. In most cases, the approach leads to new or stronger results; in other cases it recovers known results with new insight. The approach is built upon recent development on stochastic majorization and multivariate characterization of stochastic order relations.

Keywords: rearrangement inequalities, stochastic majorization, stochastic scheduling

# 1 Introduction

That $a \geq b$ and $x \geq y$ ($a, b, x, y$ are all real values) implies $ax + by \geq ay + bx$ is obvious. A generalization of this simple result is the classical Hardy-Littlewood-Polya (HLP) [11] inequality: Let $x$ and $y$ be two $n$-vectors, $\pi$ and $\sigma$ be any two permutations (of $n$ integers). Let $x^\pi$ denote the rearrangement of $x$ according to the permutation $\pi$; in particular $x^\uparrow$ denotes the arrangement of $x$ in increasing order. Denote similarly $y^\sigma$, $y^\uparrow$ and $y^\downarrow$ (decreasing arrangement). Then the inner products satisfy the following inequalities:

$$x^\uparrow \cdot y^\downarrow \leq x^\pi \cdot y^\sigma \leq x^\uparrow \cdot y^\uparrow.$$

A generalization of the above inequality is due to Day [7]. Denote

$$f \odot \phi(x, y) = f(\phi(x_1, y_1), ..., \phi(x_n, y_n)).$$

Then,

$$f \odot \phi(x^\uparrow, y^\downarrow) \leq f \odot \phi(x^\pi, y^\sigma) \leq f \odot \phi(x^\uparrow, y^\uparrow),$$

for any monotone supermodular function $\phi$ and any increasing Schur convex function $f$. Since Schur convex functions form a defining class for majorization ordering, the above result can be recast into majorization inequalities. In fact there is an intimate relationship between rearrangement and majorization; refer to Marshall and Olkin [17], Chapter 6.

That rearrangement inequalities play a central role in scheduling is no surprise. After all, scheduling problems are typically concerned with various rearrangement of jobs (processing sequence). Many such problems are solved through a pairwise interchange argument (i.e., interchanging the processing order of a pair of jobs leads to an improved objective value), which relies heavily on rearrangement inequalities. For instance, the optimality of the classical SPT rule (shortest processing time first) in terms of minimizing the total delay ("flow time") of all jobs follows directly from the HLP inequality. Although Day's inequality has yet to become part of the scheduler's tool kit, many deterministic scheduling problems in the literature can indeed be viewed as a simple application of Day's results.

Therefore it becomes evident that in order to solve *stochastic* scheduling problems (i.e., stochastic counterparts of deterministic scheduling problems), it would be helpful to first develop a stochastic version of Day's results. One asks, what happens to Day's inequality when $x$ and $y$ become *random* vectors (denoted $X$ and $Y$ below)?

Here, the first issue one encounters is the ordering of the components of the random vectors $X$ and $Y$. (In the deterministic setting, the components of a real vector are always naturally ordered.) Conceivably, different orderings will lead to different stochastic versions of rearrangement inequalities. The situation is typical in stochastic scheduling: a stronger ordering of the processing times, for instance, will lead to a stronger result; weakening the ordering weakens the result.

We shall focus on those commonly used stochastic order relations: likelihood ratio ordering, hazard rate ordering, stochastic ordering and increasing convex (concave) ordering. Refer to Ross [25] (Chapter 8) and Stoyan [32]. Based on these order relations, we develop several stochastic versions of Day's inequality. Our approach is based upon a fruitful merging of the results in two recent pieces of work: the multivariate characterization of various stochastic order relations of Shanthikumar and Yao [31], and a new class of functions characterizing stochastic majorization developed by Chang [4].

Motivated by certain stochastic scheduling applications, we also develop a stochastic version of Day's inequality with the permutations restricted to the so-called "one-cycle" permutations. Results in this direction appear to have other applications as well, for instance, the stochastic traveling salesman problem.

Applying the stochastic rearrangement inequalities to scheduling problems, we present solutions to the stochastic counterpart of many classical deterministic scheduling problems in the literature (e.g., Conway, Maxwell and Miller [6]). In most cases, we are able to generate new or stronger results. In a few other cases, we recover known results but provide these with new insight through our unified approach.

Briefly, the rest of the paper is organized as follows. In §2 we present preliminaries on majorization and stochastic order relations, focusing on those results in [4], [17] and [31] that will be used later in the paper. In §3 we develop the stochastic versions of Day's inequality, first allowing all permutations and then restricting to one-cycle permutations. Applications in stochastic scheduling are elaborated in §4, which consists of seven subsections, each focusing on one model. There, we present solutions to the stochastic version of many classical deterministic scheduling models, such as the SPT rule, Jackson's rule, Johnson's rule, flow shops with infinite buffer or without buffer, and parallel machines. Finally, in §5 our main results in stochastic scheduling are summarized in three tables.

## 2 Preliminaries

Unless otherwise specified, throughout we shall use lower case letters such as $x$ and $y$ to denote (deterministic) vectors or scalars (the distinction is often transparent from the context, or is otherwise of no importance). In the vector case, a subscript will index the components of a vector, e.g., $x_i$; on the other hand, a superscript will index a set of vectors, e.g., $x^1$, $x^2$, etc.

Also use $x_{[i]}$ and $x_{(i)}$ to denote, respectively, the $i$ th largest and the $i$ th smallest components of $x$. Upper case letters will denote random vectors or random variables, such as $X$ and $Y$, with subscripts and superscripts following the same conventions as in the deterministic case.

Throughout, "increasing/decreasing" and "convex/concave" are used in the non-strict sense. LHS and RHS denote, respectively "left hand side" and "right hand side".

### 2.1 Majorization and Stochastic Majorization

Most of the preliminaries in this subsection can be found in Marshall and Olkin [17]. Exceptions are the results associated with the class of functions $\mathbf{C}_5$ and the induced stochastic majorization. For these, refer to Chang [4].

Recall an $n$-vector $x$ is majorized by $y$, denoted $x \prec y$, if (i) $\sum_{i=1}^{j} x_{[i]} \le \sum_{i=1}^{j} y_{[i]}$, $j = 1, \dots, n-1$ and (ii) $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$. The vector $x$ is weakly submajorized by $y$, denoted $x \prec_w y$, if $\sum_{i=1}^{j} x_{[i]} \le \sum_{i=1}^{j} y_{[i]}$, $j = 1, \dots, n$. The vector $x$ is weakly supermajorized by $y$, denoted $x \prec^w y$, if $\sum_{i=j}^{n} x_{[i]} \ge \sum_{i=j}^{n} y_{[i]}$, $j = 1, \dots, n$.

A function that preserves the majorization order is called a Schur convex function. Specifically, $f : \Re^n \to \Re$ is termed Schur convex if $f(x) \le f(y)$ for all $x \prec y$. The function $f$ is termed Schur concave if $-f$ is Schur convex.

The following four classes of functions are considered in [17], Chapter 11:

$\mathbf{C}_1 = \{f : f \text{ is Schur convex }\}$.

$\mathbf{C}_2 = \{f : f \text{ is convex and symmetric }\}$.

$\mathbf{C}_3 = \{f : f(x) = \sum_{i=1}^{n} g(x_i) \text{ for some convex function } g \text{ on } \Re\}$.

$\mathbf{C}_4 = \{f : f(x) = \sum_{i=1}^{j} x_{[i]} \text{ or } f(x) = -\sum_{i=j}^{n} x_{[i]} \text{ for some } j = 1, \dots, n\}$.

Chang [4] added to the above list a fifth class:

$\mathbf{C}_5 = \{f : f \text{ is symmetric, submodular and convex in each variable }\}$.

Recall a function $f$ is submodular (supermodular), if for $\epsilon_1, \epsilon_2 \ge 0$ and $i \ne j$, $f(x + \epsilon_1 e^i + \epsilon_2 e^j) + f(x) \le (\ge) f(x + \epsilon_1 e^i) + f(x + \epsilon_2 e^j)$, where $e^i$ and $e^j$ denote the $i$ th and the $j$ th unit vector. (Note that in [17], submodular and supermodular functions are referred to as L-subadditive and L-superadditive functions.)

In addition, for $i = 1,2,3,4,5$, let $\mathbf{C}_i^{\uparrow} = \{f : f \in \mathbf{C}_i$ and $f$ is increasing $\}$ and $\mathbf{C}_i^{\downarrow} = \{f : f \in \mathbf{C}_i$ and $f$ is decreasing $\}$.

The following relations among the five classes of functions are known: $\mathbf{C}_3 \subset \mathbf{C}_5 \subset \mathbf{C}_1$, $\mathbf{C}_3 \subset \mathbf{C}_2 \subset \mathbf{C}_1$, $\mathbf{C}_4 \subset \mathbf{C}_5$, and $\mathbf{C}_4 \subset \mathbf{C}_2$. Inclusions not listed here are in general false. In particular, Chang [4] presents examples to show that $\mathbf{C}_5 \not\subset \mathbf{C}_2$ and $\mathbf{C}_2 \not\subset \mathbf{C}_5$. However, all five classe are equally effective in characterizing majorization ([17, 4]), i.e., for $i = 1,2,3,4,5$, $x \prec (\prec_w, \prec^w) y$ if and only if $f(x) \leq f(y)$ for all $f \in \mathbf{C}_i$ $(\mathbf{C}_i^{\uparrow}, \mathbf{C}_i^{\downarrow})$.

Based on the five classes of functions one can define different versions of *stochastic* majorization. Versions that are based on the first four classes were elaborated in [17]. Chang [4] proposed a new version based on $\mathbf{C}_5$ and illustrated its various applications. For the stochastic scheduling applications in this paper, we will primarily use stochastic majorizations that are based on $\mathbf{C}_1$ and $\mathbf{C}_5$, and occasionally $\mathbf{C}_3$. Other forms of stochastic majorization and their applications can be found in Shanthikumar [29] and Liyanage and Shanthikumar [16].

It is easy to verify that in general the class $\mathbf{C}_5$ is not closed under increasing convex functional composition. That is, given $f \in \mathbf{C}_5$ and $g$ an increasing and convex function, it is not true in general that $g \circ f \in \mathbf{C}_5$. For our applications below it is convenient to modify the class $\mathbf{C}_5$ to make it closed under composition of increasing and convex functions. The resulting class is $\mathbf{C}_{5-icx}$ defined below:

$$\mathbf{C}_{5-icx} = \{f = g \circ h \,:\, h \in \mathbf{C}_5,\, g \text{ is increasing and convex}\}.$$

Define $\mathbf{C}_{5-icx}^{\uparrow}$ and $\mathbf{C}_{5-icx}^{\downarrow}$ accordingly if $h \in \mathbf{C}_5^{\uparrow}$ or $h \in \mathbf{C}_5^{\downarrow}$ in the above definition.

### 2.2. Stochastic Order Relations

We start with briefly recalling several commonly used stochastic order relations. Refer to Stoyan [32] and Ross ([25], Chapter 8) for details.

Let $X$ and $Y$ be two random variables. They are ordered in, respectively, the stochastic ordering, increasing convex ordering and increasing concave ordering, denoted $X \geq_{st} Y$, $X \geq_{icx} Y$, and $X \geq_{icv} Y$), if $Ef(X) \geq Ef(Y)$ for all $f : \Re \mapsto \Re$ that is, respectively, increasing, increasing and convex, and increasing and concave.

Clearly, stochastic ordering implies both increasing convex and increasing concave orderings. On the other hand, the order relations stronger than stochastic ordering are: the likelihood-ratio and hazard-rate orderings, denoted $X \geq_{lr} Y$ and $X \geq_{hr} Y$, and defined as respectively, $f(x)/f(x+a) \leq g(x)/g(x+a)$ for all $x$ and all $a > 0$, and $f(x)/[1-F(x)] \leq g(x)/[1-G(x)]$, for all $x$; where $f$ and $g$ denote the density functions, while $F$ and $G$ denote the distribution functions of $X$ and $Y$. When $X$ and $Y$ have discrete distributions, likelihood ratio and hazard rate orderings are similarly defined: replace the density functions by the probability mass functions. It is known that likelihood ratio ordering implies hazard rate ordering, which in turn implies stochastic ordering.

Likelihood ratio and hazard rate orderings do not admit univariate functional characterizations as the other weaker orderings (stochastic, convex and concave orderings). However, Shanthikumar and Yao [31] developed the following classes of *bivariate* functions to characterize these orderings.

$\mathbf{C}_{lr} = \{f : f(x,y) - f(y,x) \geq 0, \forall x \geq y\}$ ($\mathcal{G}_{lr}$ in [31]).

$\mathbf{C}_{hr} = \{f : f(x,y) - f(y,x)$ is increasing in $x$, $\forall x \geq y\}$ ($\mathcal{G}_{hr}$ in [31]).

$\mathbf{C}^*_{hr} = \{f : f(x,y)$ is increasing in $x$, $\forall x \geq y$ and; $f \in \mathbf{C}_{hr}\}$ ($\mathcal{G}^1_{hr}$ in [31]).

In addition, stochastic ordering also admits a bivariate characterization through the following class:

$\mathbf{C}_{st} = \{f : f(x,y) - f(y,x)$ is increasing in $x$, $\forall x\}$ ($\mathcal{G}_{st}$ in [31]).

**Theorem 2.1** *([31]) $X \geq_{lr} Y$ (resp. $X \geq_{hr} Y$, $X \geq_{st} Y$) if and only if there exists two independent random variables $\hat{X}$ and $\hat{Y}$, equal in distribution to $X$ and $Y$ respectively, such that $Ef(\hat{X},\hat{Y}) \geq Ef(\hat{Y},\hat{X})$ for all $f \in \mathbf{C}_{lr}$ (resp. $\mathbf{C}^*_{hr}, \mathbf{C}_{st}$).*

It is easy to verify that $\mathbf{C}_{lr}$ is preserved under increasing functional composition, while $\mathbf{C}^*_{hr}$ is preserved under increasing and convex functional composition. Hence, the results in the above theorem can also be rephrased as follows: $f(\hat{X},\hat{Y}) \geq_{st} f(\hat{Y},\hat{X})$ for all $f \in \mathbf{C}_{lr}$ and $f(\hat{X},\hat{Y}) \geq_{icx} f(\hat{Y},\hat{X})$ for all $f \in \mathbf{C}^*_{hr}$, respectively, for the $lr$ and the $hr$ cases.

# 3 Stochastic Majorization via Rearrangement

### 3.1 Rearrangement without Restrictions

In the deterministic setting, there is a well known result, due to Day [7] (also see [17], Chapter 6), that relates rearrangement to majorization. Here we develop the stochastic counterpart of Day's result. For ease of exposition, we focus our presentation on two-dimensional vectors, and only list the general results in higher dimension at the end of the section.

**Theorem 3.1** *(Day) Given two vectors $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Suppose $x_1 \geq x_2$ and $y_1 \geq y_2$. Then, for any $\phi(\cdot, \cdot)$ that is monotone and supermodular (resp. submodular),*

$$(\phi(x_1, y_1), \phi(x_2, y_2)) \succ_w \text{ (resp. } \succ^w)(\phi(x_1, y_2), \phi(x_2, y_1)).$$

*In other words,*

$$f(\phi(x_1, y_1), \phi(x_2, y_2)) \geq f(\phi(x_1, y_2), \phi(x_2, y_1)) \tag{1}$$

*for all $f \in \mathbf{C}_1^{\uparrow}$ (resp. $f \in \mathbf{C}_1^{\downarrow}$).*

(Note that Day's original result also includes a converse statement, for instance, if the weak majorization holds then $\phi$ must be monotone and supermodular. This converse part, however, is not relevant to our development below.)

We now present the stochastic counterpart of the above theorem.

**Theorem 3.2** *$X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ are two independent random vectors, both have independent components. Suppose $X_1 \geq_a X_2$ and $Y_1 \geq_b Y_2$. (The orderings $\geq_a$ and $\geq_b$ will be specified below.) Suppose $\phi(\cdot, \cdot)$ is a monotone and supermodular function. Then,*

$$Ef(\phi(X_1, Y_1), \phi(X_2, Y_2)) \geq Ef(\phi(X_1, Y_2), \phi(X_2, Y_1)),$$

*where (i) $f \in \mathbf{C}_1^{\uparrow}$ if $(a, b) = (lr, lr)$, (ii) $f \in \mathbf{C}_5^{\uparrow}$ if $(a, b) = (st, st)$, (iii) $f \in \mathbf{C}_{5-icx}^{\uparrow}$ if $\phi$ is increasing and $(a, b) = (lr, hr)$ or $(a, b) = (hr, lr)$.*

*Moreover, when $\phi$ is a monotone and submodular function, the above results (i) and (ii) hold with $\mathbf{C}_i^{\uparrow}$ replaced by $\mathbf{C}_i^{\downarrow}$ for i=1 and 5, respectively. If $\phi$ is a decreasing and submodular function, then (iii) holds with $\mathbf{C}_{5-icx}^{\uparrow}$ replaced by $\mathbf{C}_{5-icx}^{\downarrow}$.*

In later applications, we find two forms of $\phi$ particularly useful: $\phi(x, y) = x + y$ and $\phi(x, y) = \max(x, y)$. Note that in the first case $(+)$, $\phi$ is both supermodular and submodular, hence Day's result holds as a majorization [i.e., $f \in \mathbf{C}_1$ in (1)]. The stochastic counterparts in Theorem 3.2 hold for $f \in \mathbf{C}_i$ for $i$=1 and 5, respectively.

Below, we prove part (i) of Theorem 3.2 first.

**Proof** [of Theorem 3.2(i)]. For $x_1 \geq x_2$ and $y_1 \geq y_2$, we have

$$f(\phi(x_1, y_1), \phi(x_2, y_2)) \geq f(\phi(x_1, y_2), \phi(x_2, y_1)) \tag{2}$$

for any $f \in \mathbf{C}_1^{\uparrow}$, following Theorem 3.1. Now, fix $x_1$ and $x_2$ ($x_1 \geq x_2$), write the LHS of (2) as $g(y_1, y_2)$, then the RHS is $g(y_2, y_1)$; hence, $g \in \mathbf{C}_{lr}$, and

$$Ef(\phi(x_1, Y_1), \phi(x_2, Y_2)) \geq Ef(\phi(x_1, Y_2), \phi(x_2, Y_1)).$$

(Refer to Theorem 2.7.) Next, write the LHS above as $h(x_1, x_2)$ and repeat the argument. The case $f \in \mathbf{C}_1^{\downarrow}$ is similarly proved. ∎

To prove part (ii) of Theorem 3.2, let's first have a closer examination of the class of functions in $\mathbf{C}_5$. The following is easily verified. (Note in Lemmas 3.3 and 3.4 below, the symmetry of $f$ is *not* needed.)

**Lemma 3.3** *$f \in \mathbf{C}_5$ (resp. $f \in \mathbf{C}_5^{\uparrow}$, $f \in \mathbf{C}_5^{\downarrow}$) if and only if*

$$f(z^1) + f(z^4) \geq f(z^2) + f(z^3)$$

*for any four vectors $z^i = (x_i, y_i)$, $i = 1, 2, 3, 4$, satisfying:*

$$x_1 \geq \max\{x_2, x_3\}, \quad y_4 \geq \max\{y_2, y_3\};$$

$$x_1 + x_4 = \text{ (resp. } \geq, \leq)\, x_2 + x_3, \quad y_1 + y_4 = \text{ (resp. } \geq, \leq)\, y_2 + y_3.$$

Note the four points (vectors) $z^i$ above have the following relation: $z^1$ and $z^4$ are, respectively, the southeast and the northwest corners of the rectangle they form, $z^2$ and $z^3$ are then two points which fall inside (or at most at the boundary) of the rectangle. The (componentwise) convexity of $f$ follows from collapsing the rectangle into a single horizontal or vertical line. Submodularity corresponds to moving $z^2$ and $z^3$ to the other two (i.e., the northeast and the southwest) corners of the rectangle. We also remark that the same four points play an essential role in the construction of sample-path based stochastic majorization in [4].

**Lemma 3.4** *Suppose $f \in \mathbf{C}_5^{\uparrow}$ (resp. $f \in \mathbf{C}_5^{\downarrow}$) and $\phi$ is monotone and supermodular (resp. submodular). Given $y_1 \geq y_2$, then $f(\phi(x_1, y_1), \phi(x_2, y_2)) - f(\phi(x_1, y_2), \phi(x_2, y_1))$ is increasing in $x_1$. In other words, given $y_1 \geq y_2$, $g(x_1, x_2) = f(\phi(x_1, y_1), \phi(x_2, y_2)) \in \mathbf{C}_{st}$ (requiring symmetry of $f$).*

**Proof.** Suppose $\phi$ is increasing. Let $a > 0$ be any given constant. Define the four points $z^i$ as follows:

$$z^1 = (\phi(x_1 + a, y_1), \phi(x_2, y_2)), \quad z^4 = (\phi(x_1, y_2), \phi(x_2, y_1));$$

$$z^2 = (\phi(x_1 + a, y_2), \phi(x_2, y_1)), \quad z^3 = (\phi(x_1, y_1), \phi(x_2, y_2)).$$

Then clearly $z^i$ satisfy the requirements in Lemma 3.3. In particular,

$$\phi(x_1 + a, y_1) + \phi(x_1, y_2) \geq \phi(x_1 + a, y_2) + \phi(x_1, y_1)$$

follows from the supermodularity of $\phi$. When $f \in \mathbf{C}_5^{\uparrow}$, we have $f(z^1) - f(z^2) \geq f(z^3) - f(z^4)$, and hence the desired increasingness. When $\phi$ is decreasing, interchange the roles played by $z^1$ and $z^4$. The case $f \in \mathbf{C}_5^{\downarrow}$ is similarly proved. The last statement in this theorem is obvious, given the established increasingness in $x_1$ and the symmetry of $f$. ∎

Note that if in addition to the conditions in Lemma 3.4, we also have $x_1 \geq x_2$, then (requiring the symmetry of $f$)

$$\begin{aligned} & f(\phi(x_1, y_1), \phi(x_2, y_2)) - f(\phi(x_1, y_2), \phi(x_2, y_1)) \\ \geq\ (\leq)\ & f(\phi(x_2, y_1), \phi(x_2, y_2)) - f(\phi(x_2, y_2), \phi(x_2, y_1)) = 0, \end{aligned}$$

for $f \in \mathbf{C}_5^{\uparrow}$ ($\mathbf{C}_5^{\downarrow}$). Since $\mathbf{C}_5^{\uparrow}$ ($\mathbf{C}_5^{\downarrow}$) is equivalently effective in characterizing weak submajorization (supermajorization), we recover Day's result [cf. Theorem 3.1].

**Proof** [of Theorem 3.2 (ii)]. Consider $f \in \mathbf{C}_5^{\uparrow}$ and $\phi$ increasing supermodular. (The other case is similarly proved.) We want to show, under the given conditions, that

$$Ef(\phi(x_1, Y_1), \phi(x_2, Y_2)) - Ef(\phi(x_1, Y_2), \phi(x_2, Y_1)) \tag{3}$$

is increasing in $x_1$. Since $Y_1 \geq_{st} Y_2$, there exists a pair of random variables, $\hat{Y}_1 \geq \hat{Y}_2$ (a.s.), with $\hat{Y}_i$ equal in distribution to $Y_i$ for $i$=1,2. (Refer to [32], for instance.) Generate *independently* another pair, $\hat{Y}_1' \geq \hat{Y}_2'$ (a.s.), that are also equal in distribution to $Y_1$ and $Y_2$, respectively. Then, for any given $a > 0$, it is easy to verify that

$$\begin{aligned} & f(\phi(x_1 + a, \hat{Y}_1), \phi(x_2, \hat{Y}_2')) - f(\phi(x_1 + a, \hat{Y}_2), \phi(x_2, \hat{Y}_1')) \\ & \geq f(\phi(x_1, \hat{Y}_1), \phi(x_2, \hat{Y}_2')) - f(\phi(x_1, \hat{Y}_2), \phi(x_2, \hat{Y}_1')), \end{aligned} \tag{4}$$

since $f \in \mathbf{C}_5^{\uparrow}$ and the four points above correspond to $z^1$, $z^2$, $z^3$ and $z^4$, respectively. Taking expectations on both sides of (4), we obtain the relation in (3). The desired result then follows by replacing $x_1$ and $x_2$ by $X_1$ and $X_2$, taking into account Theorem 2.1. ∎

Finally, we prove part (iii) of Theorem 3.2.

**Proof** [of Theorem 3.2(iii)]. Suppose $(a, b) = (lr, hr)$. Again, consider the case $f \in \mathbf{C}_{5-icx}^{\uparrow}$ and $\phi$ is increasing and supermodular, the other cases being similar. Fix $x_1 \geq x_2$, and write $f$ as

$$f(y_1, y_2) = g \circ h(\phi(x_1, y_1), \phi(x_2, y_2)),$$

where $g$ is an increasing and convex function and $h \in \mathbf{C}_5^{\uparrow}$. Write the $h$ part above as $h(y_1, y_2)$, then $h(y_1, y_2) \in \mathbf{C}_{st}$ following Lemma 3.4, and hence $h(y_1, y_2) \in \mathbf{C}_{hr}$. Since both $\phi$ and $h$ are increasing, $h(y_1, y_2) \in \mathbf{C}_{hr}^*$. Since the class $\mathbf{C}_{hr}^*$ is closed under increasing and convex functional composition, it follows that $g \circ h \in \mathbf{C}_{hr}^*$. Therefore [cf. Theorem 2.1],

$$Eg \circ h(\phi(x_1, Y_1), \phi(x_2, Y_2)) \geq Eg \circ h(\phi(x_1, Y_2), \phi(x_2, Y_1)),$$

since $Y_1 \geq_{hr} Y_2$. Now, write the LHS above as $\tilde{f}(x_1, x_2)$, then the RHS is $\tilde{f}(x_2, x_1)$, taking into account the symmetry of $h$. Since $X_1 \geq_{lr} X_2$, we have following Theorem 2.1,

$$Eg \circ h(\phi(X_1, Y_1), \phi(X_2, Y_2)) \geq Eg \circ h(\phi(X_1, Y_2), \phi(X_2, Y_1)),$$

which is the desired conclusion. ∎

To conclude this subsection, we present the generalization of Day's result (Theorem 3.1) and its stochastic counterpart Theorem 3.2. We need the following notations. Let $x$ and $y$ be two $n$-vectors. Let $\pi$ and $\sigma$ be two permutations over $n$ integers $\{1, 2, ..., n\}$. Denote

$$f \odot \phi(x^{\pi}, y^{\sigma}) = f(\phi(x_{\pi(1)}, y_{\sigma(1)}), ..., \phi(x_{\pi(n)}, y_{\sigma(n)}));$$

$$f \odot \phi(x^{\uparrow}, y^{\uparrow}) = f(\phi(x_{(1)}, y_{(1)}), ..., \phi(x_{(n)}, y_{(n)}));$$

$$f \odot \phi(x^{\uparrow}, y^{\downarrow}) = f(\phi(x_{(1)}, y_{[1]}), ..., \phi(x_{(n)}, y_{[n]})).$$

**Theorem 3.5** *(Day [7]) For any two $n$-vectors $x$ and $y$, and any two permutations $\pi$ and $\sigma$, the following inequalities*

$$f \odot \phi(x^{\uparrow}, y^{\downarrow}) \leq f \odot \phi(x^{\pi}, y^{\sigma}) \leq f \odot \phi(x^{\uparrow}, y^{\uparrow}),$$

*hold for any $\phi$ that is monotone and supermodular (resp. submodular) and any $f \in \mathbf{C}_1^{\uparrow}$ (resp. $f \in \mathbf{C}_1^{\downarrow}$).*

**Theorem 3.6** *$X$ and $Y$ are two independent random $n$-vectors, each with independent components. Suppose the components of $X$ ($Y$) are ordered in the sense of $\geq_a$ ($\geq_b$) to be specified below. Suppose $\phi(\cdot, \cdot)$ is a monotone and supermodular function. Then, for any permutations $\pi$ and $\sigma$ of $\{1, 2, ..., n\}$, we have*

$$Ef \odot \phi(X^{\uparrow}, Y^{\downarrow}) \leq Ef \odot \phi(X^{\pi}, Y^{\sigma}) \leq Ef \odot \phi(X^{\uparrow}, Y^{\uparrow}),$$

*where, (i) $f \in \mathbf{C}_1^{\uparrow}$ if $(a, b) = (lr, lr)$, (ii) $f \in \mathbf{C}_5^{\uparrow}$ if $(a, b) = (st, st)$, (iii) $f \in \mathbf{C}_{5-icx}^{\uparrow}$ if $\phi$ is increasing and $(a, b) = (lr, hr)$ or $(a, b) = (hr, lr)$.*

*Moreover, when $\phi$ is a monotone and submodular function, the above results (i) and (ii) hold with $\mathbf{C}_i^{\uparrow}$ replaced by $\mathbf{C}_i^{\downarrow}$ for $i = 1$ and 5, respectively; while (iii) holds with $\mathbf{C}_{5-icx}^{\uparrow}$ replaced by $\mathbf{C}_{5-icx}^{\downarrow}$ if $\phi$ is decreasing and submodular. (In particular, when $\phi(x,y) = x+y$, the results hold with $f \in \mathbf{C}_i$, for $i = 1$ and 5, respectively.)*

### 3.2 Rearrangement under Restrictions

Let $\Pi$ denote the set of $n!$ permutations on the set of integers $\mathbf{N} = \{1, 2, ..., n\}$. In §3.1 the main result, Theorem 3.6, is concerned with all permutations $\pi, \sigma \in \Pi$. Here in contrast, we restrict $\Pi$ to a special subset, that of the so-called "one-cycle" permutations. Denote this subset as $\Pi^1$. Below we briefly review this concept.

Recall, a permutation can be expressed in the "circular form". For instance, (1,4)(2,3,5,6) denotes the following permutation:

$$1 \mapsto 4 \mapsto 1, \quad 2 \mapsto 3 \mapsto 5 \mapsto 6 \mapsto 2;$$

where $1 \mapsto 4$, e.g., reads as $\pi(1) = 4$ (i.e., 1 is mapped to 4), and so forth. Clearly, this particular permutation has two "cycles". A "one-cycle" permutation is one that has a single cycle, such as

$$1 = i_0 \mapsto i_1 \mapsto i_2 \mapsto ... \mapsto i_{n-1} \mapsto i_n = 1,$$

where, except $i_0 = i_n = 1$, all the other $i_k$'s are all distinct members of $\{2, ..., n\}$. Denote the set of one-cycle permutations as $\Pi^1$. Clearly, $\Pi^1$ is a subset of $\Pi$, with cardinality $(n-1)!$. (Refer to, for instance, Riordan [24], Chapter 4, p.68.)

One can view a permutation $\pi$ as a mapping, and define

$$\pi^0(i) = i; \quad \pi^k(i) = \pi(\pi^{k-1}(i)), \quad k = 1, ..., n.$$

Then, clearly

$\pi \in \Pi^1$ if and only if $\{\pi^k(i),\, k = 1, ..., n\} = \mathbf{N}$ for any $i \in \mathbf{N}$.

In words, applying a one-cycle permutation to any $i \in \mathbf{N}$ successively for $n$ times yields the $n$ distinct members of $\mathbf{N}$, and hence the set $\mathbf{N}$ itself. Based on this fact, without loss of generality, whenever we refer to a generic one-cycle permutation $\pi$ below, we will let $i = 1$, and write $\pi^k(i)$ simply as $\pi^k$.

Conceptually it helps to relate a one-cycle permutation to a TSP ("traveling salesman problem") tour among the $n$ "points" in $\mathbf{N}$. For instance, if $\pi \in \Pi^1$ is such a permutation, expressed as follows:

$$1 = \pi^0 \mapsto \pi^1 \mapsto ... \mapsto \pi^{n-1} \mapsto \pi^n = 1,$$

then the corresponding TSP tour is to start from 1 and travel to $\pi^1$, ..., to $\pi^{n-1}$ and then return to 1. Indeed, with this in mind, the lemma below is immediate:

**Lemma 3.7** *If $\pi \in \Pi^1$ is defined as follows:*

$$1 \mapsto \pi^1 \mapsto ... \mapsto \pi^i \mapsto ... \mapsto \pi^j \mapsto ... \mapsto \pi^{n-1} \mapsto 1,$$

*where $i < j$, let $\tilde{\pi}$ be another permutation obtained from reversing the segment from $\pi^i$ to $\pi^j$ in $\pi$:*

$$1 \mapsto \pi^1 \mapsto ... \mapsto \pi^{i-1} \mapsto \pi^j \mapsto \pi^{j-1} \mapsto ... \mapsto \pi^i \mapsto \pi^{j+1} \mapsto ... \mapsto 1.$$

*Then, $\tilde{\pi} \in \Pi^1$.*

Note that $\tilde{\pi}$ in the above lemma follows the tour of $\pi$ up to $i-1$ and then visits the points $i$ through $j$ in reverse order; after that it follows $\pi$'s itinerary again. Obviously, $\tilde{\pi}$ is also a TSP tour. In words, the above lemma says that reversing any segment of a one-cyle permutation results in another (possibly the same) one-cycle permutation.

Below, we want to develop the parallel result to Day's rearrangement theorem [cf. Theorem 3.5] but with the permutations restricted to $\Pi^1$.

**Definition 3.8** *The following one-cycle permutations are termed, respectively, "odd-even" (denoted $\pi^{oe}$) and "small-large" (denoted $\pi^{s\ell}$): for $n = 2k+1$,*

$$\pi^{oe} = \quad 1 \mapsto 3 \mapsto 5 \mapsto ... \mapsto 2k-1 \mapsto 2k+1 \mapsto 2k \mapsto 2k-2 \mapsto ... \mapsto 4 \mapsto 2 \mapsto 1,$$

$$\pi^{s\ell} = \quad 1 \mapsto 2k+1 \mapsto 2 \mapsto 2k-1 \mapsto 4 \mapsto ... \mapsto k \mapsto ... \mapsto 5 \mapsto 2k-2 \mapsto 3 \mapsto 2k \mapsto 1;$$

*and for $n = 2k$,*

$$\pi^{oe} = \quad 1 \mapsto 3 \mapsto 5 \mapsto ... \mapsto 2k-3 \mapsto 2k-1 \mapsto 2k \mapsto 2k-2 \mapsto ... \mapsto 4 \mapsto 2 \mapsto 1,$$

$$\pi^{s\ell} = \quad 1 \mapsto 2k \mapsto 2 \mapsto 2k-2 \mapsto 4 \mapsto ... \mapsto k \mapsto ... \mapsto 5 \mapsto 2k-3 \mapsto 3 \mapsto 2k-1 \mapsto 1.$$

In words, an odd-even permutation corresponds to a tour that first visits the odd points in increasing order and then visits the even points in decreasing order. A small-large permutation is formulated by assigning the $n$ integers, one pair at a time and alternating between small and large pairs, to positions from the left and the right ends towards the center. Suppose $n = 2k+1$. Start by assigning two 1's to the left and the right ends; followed by assigning the pair $(2k+1, 2k)$: $2k+1$ to the left and $2k$ to the right; followed by assigning (2,3) in the same fashion, followed by assigning $(2k-1, 2k-2)$, and so forth. (The case $n = 2k$ is similar.) As an example, consider $n = 10$. Then,

$$\pi^{oe} = \quad 1 \mapsto 3 \mapsto 5 \mapsto 7 \mapsto 9 \mapsto 10 \mapsto 8 \mapsto 6 \mapsto 4 \mapsto 2 \mapsto 1;$$

$$\pi^{s\ell} = \quad 1 \mapsto 10 \mapsto 2 \mapsto 8 \mapsto 4 \mapsto 6 \mapsto 5 \mapsto 7 \mapsto 3 \mapsto 9 \mapsto 1.$$

Suppose $\pi \in \Pi^1$, denote

$$f \odot \phi(\pi) = f(\phi(\pi^0, \pi^1), \phi(\pi^1, \pi^2), ...\phi(\pi^{n-1}, \pi^n)),$$

and similarly denote $f \odot \phi(\pi^{oe})$ and $f \odot \phi(\pi^{s\ell})$.

**Lemma 3.9** *Suppose $\phi$ is a monotone, symmetric, and supermodular (resp. submodular) function. Then for any $\pi \in \Pi^1$, the following inequalities hold:*

$$f \odot \phi(\pi^{s\ell}) \leq f \odot \phi(\pi) \leq f \odot \phi(\pi^{oe}) \tag{5}$$

*for any $f \in \mathbf{C}_1^{\uparrow}$ (resp. $f \in \mathbf{C}_1^{\downarrow}$).*

**Proof.** For any given $\pi \in \Pi^1$, suppose it takes the following form:

$$1 \mapsto \pi^1 \mapsto ... \mapsto \pi^{i-1} \mapsto 2 \mapsto \pi^{i+1} \mapsto ... \mapsto \pi^{n-1} \mapsto 1.$$

We can bring 2 next to 1 on the RHS by reversing the segment 2 through $\pi^{n-1}$, with the resulting permutation, denoted $\tilde{\pi}$, still in $\Pi^1$:

$$1 \mapsto \pi^1 ... \mapsto \pi^{i-1} \mapsto \pi^{n-1} \mapsto \pi^{n-2} \mapsto ... \mapsto 2 \mapsto 1.$$

In terms of the values of the two functions $f \odot \phi(\pi)$ and $f \odot \phi(\tilde{\pi})$, they only differ in the following pairs (taking into account the symmetry of $\phi$):

$$(\phi(\pi^{i-1}, 2), \phi(\pi^{n-1}, 1)) \text{ for } f \odot \phi(\pi), \text{ versus } (\phi(\pi^{i-1}, \pi^{n-1}), \phi(2, 1)) \text{ for } f \odot \phi(\tilde{\pi}).$$

Since $\pi^{i-1} > 1$ and $\pi^{n-1} > 2$, we know the value of the function $f \odot \phi$ is increased from $\pi$ to $\tilde{\pi}$, following Theorem 3.1.

Next, we continue with bringing 3 next to 1 on the LHS. This will further increase the value of the function. Continue in this fashion until $\pi^{oe}$ is reached. At that point clearly no further improvement can be made, hence the upper bound in (5).

The approach to reach the lower bound is similar. Suppose $n$ is odd, and $\pi$ takes the following form:

$$1 \mapsto \pi^1 \mapsto ... \mapsto \pi^{i-1} \mapsto n \mapsto \pi^{i+1} \mapsto ... \mapsto 1.$$

Then, move $n$ next to 1 on the LHS by reversing the segment $\pi^1$ through $n$. Since $\pi^{i+1} > 1$ and $n > \pi^1$, we know that the value of the function will be reduced. Next, bring $n-1$ next to 1 on the RHS. Continue in the fashion until $\pi^{s\ell}$ is reached. ∎

For any given one-cycle permutation $\pi \in \Pi^1$, denote

$$f \odot \phi(x^\pi) = f(\phi(x_{\pi^0}, x_{\pi^1}), \phi(x_{\pi^1}, x_{\pi^2})...\phi(x_{\pi^{n-1}}, x_{\pi^n})),$$

and denote $f \odot \phi(x^{\pi^{oe}})$ and $f \odot \phi(x^{\pi^{s\ell}})$ similarly. The above lemma extends readily to

**Theorem 3.10** *Suppose $\phi$ is a monotone, symmetric, and supermodular (resp. submodular) function; $x$ is an $n$-vector with $x_1 \leq x_2 \leq ... \leq x_n$. Then for any $\pi \in \Pi^1$, the following inequalities hold:*

$$f \odot \phi(x^{\pi^{s\ell}}) \leq f \odot \phi(x^\pi) \leq f \odot \phi(x^{\pi^{oe}})$$

*for any $f \in \mathbf{C}_1^\uparrow$ (resp. $f \in \mathbf{C}_1^\downarrow$).*

Following the same spirit as in §3.1, the stochastic counterpart of the above is

**Theorem 3.11** *Suppose $X$ is a random $n$-vector with independent components, ordered as $X_1 \leq_a X_2 \leq_a ... \leq_a X_n$, where $\leq_a$ is to be specified below. Suppose $Y$ is an independent replica of $X$ (i.e., $Y =_{st} X$). Suppose $\phi$ is a monotone, symmetric, and supermodular (resp. submodular) function. Then, for any $\pi \in \Pi^1$, the following inequalities hold:*

$$Ef \odot \phi(X, Y^{\pi^{s\ell}}) \leq Ef \odot \phi(X, Y^\pi) \leq Ef \odot \phi(X, Y^{\pi^{oe}}),$$

*where (i) $f \in \mathbf{C}_1^\uparrow$ (resp. $f \in \mathbf{C}_1^\downarrow$) if $a = lr$, and (ii) $f \in \mathbf{C}_5^\uparrow$ (resp. $f \in \mathbf{C}_5^\downarrow$) if $a = st$.*

*Furthermore, (iii) if $Y = X$ (a.s.), then the above inequalities hold for $f \in \mathbf{C}_3^\uparrow$ (resp. $f \in \mathbf{C}_3^\downarrow$), provided $a = st$.*

The proof of the above theorem mimics that of Lemma 3.9 through conditioning. Note that part (iii) of Theorem 3.11 is the actual counterpart of the deterministic result in Theorem 3.10. The setting there is obviously more restrictive. The results in (i) and (ii), on the other hand, are tailored towards typical applications in stochastic scheduling. Also note that due to the required symmetry, there is no parallel results for $f \in \mathbf{C}_{5-icx}$ here.

To conclude this subsection, we present an application of the above theorem [specifically, part (iii)] to the stochastic TSP on the line.

**Corollary 3.12** *Consider the TSP on the set of points* $\mathbf{N}$*. Suppose the distance between* $i$ *and* $j$ *is* $d(X_i, X_j)$*, where* $d$ *is monotone, symmetric and submodular,* $X_1 \leq_{st} ... \leq_{st} X_n$ *are a set of independent random variables. Then the odd-even tour and the small-large tour [cf. Definition 3.8], respectively, minimizes and maximize the objective function* $E\sum_{i=1}^{n} d(X_i, X_{\pi(i)})$ *(over all* $\pi \in \Pi^1$*).*

(In the Corollary above, when the submodularity of $d$ is replaced by supermodularity, the odd-even tour and the small-large tour, respectively, maximizes and minimizes the objective function.)

As an example, suppose $d(x,y) = |x-y|$, i.e., the usual distance on a real line, write $d(x,y) = \max(x,y) - \min(x,y)$, then the objective function breaks into two parts. As the reader will verify, the odd-even tour will minimize the max part (max being submodular) and maximize the min part (min being supermodular), and hence minimize the total distance. A similar result applies to the small-large tour.

## 4 Stochastic Scheduling

In this section, we consider stochastic versions of some classical deterministic scheduling problems. Wherever possible we will follow the notations in Conway, Maxwell and Miller [6]. Jobs are indexed by the integers $1, 2, \ldots, n$, and machines are indexed similarly by the integers $1, 2, \ldots, m$. For each job $j$, let $P_j$, $R_j$ and $D_j$ be, respectively, its processing time, ready (or release/arrival) time and due date. We assume these are *independent* random variables. For each job $j$, the flow time $F_j$ is the total time it spends in the system, the completion time is $C_j = F_j + R_j$ (note that when the job is ready at time zero the flow time coincides with the completion time), the lateness is defined as $L_j = C_j - D_j$, while the tardiness, $T_j = \max\{0, L_j\}$. Let $C = (C_1, \ldots, C_n)$, $D = (D_1, \ldots, D_n)$, $L = (L_1, \ldots, L_n)$ and $T = (T_1, \ldots T_n)$. In particular, $C^{\uparrow} = (C_{(1)}, \ldots C_{(n)})$ denotes the *departure* sequence of the $n$ jobs. [Note: $C_{(j)}$ is denoted $C_{[n]}$ in [6].]

In addition to the classes of functions defined earlier, we shall also need the following classes (increasing, increasing and supermodular, increasing and separable):

$\mathbf{C}^{\uparrow} = \{f : f \text{ is increasing}\}$.

$\mathbf{C}^{\uparrow}_{sup} = \{f : f \text{ is increasing and supermodular}\}$.

$\mathbf{C}^{\uparrow}_{sep} = \{f : f(x) = \sum_{i=1}^{n} g_i(x_i) \text{ for some } g_i, i = 1, 2, \ldots, n, \text{ increasing}\}$.

Clearly, $\mathbf{C}^{\uparrow}_{sep} \subset \mathbf{C}^{\uparrow}_{sup} \subset \mathbf{C}^{\uparrow}$.

### 4.1 Single Machine, No Preemption: without Arrivals

Consider the scheduling problem of a single machine with $n$ jobs, all present at time zero, i.e., $R_j = 0$, $j = 1, \ldots, n$. We consider the class of *permutation schedules* in which preemptions and insertion of idle times are not allowed. The schedules are hence completely specified by the sequence in which the jobs are processed. The problem is a classical one in the deterministic setting (e.g., [6], Chapter 3). Here we establish results in the stochastic setting. We start with results concerning the departure sequence.

**Theorem 4.1** *Suppose* $P_1 \leq_a P_2 \leq_a \ldots \leq_a P_n$. *Let* $C^s_j$ *and* $C^\ell_j$, $j = 1, 2, \ldots, n$, *be the completion time of job* $j$ *under the sequence* $1 \mapsto 2 \mapsto \ldots \mapsto n$ *and the sequence* $n \mapsto (n-1) \mapsto \ldots \mapsto 1$, *respectively. Then*

$$Ef(C^s_{(1)}, C^s_{(2)}, \ldots, C^s_{(n)}) \leq Ef(C_{(1)}, C_{(2)}, \ldots, C_{(n)}) \leq Ef(C^\ell_{(1)}, C^\ell_{(2)}, \ldots, C^\ell_{(n)}),$$

*where (i)* $f \in \mathbf{C}^{\uparrow}$, *if* $a = lr$, *(ii)* $f \in \mathbf{C}^{\uparrow}_{sup}$, *if* $a = hr$ *and (iii)* $f \in \mathbf{C}^{\uparrow}_{sep}$, *if* $a = st$.

[Note that the result in Theorem 4.1 (i) (resp. (iii)) is the stochastic ordering for the *joint* (resp. *marginal*) distribution of the departure sequence. The result in Theorem 4.1 (ii) can be viewed as another form of stochastic ordering which is weaker than the stochastic ordering for joint distributions but stronger than the stochastic ordering for marginal distributions.]

**Proof**. We prove (iii) first. Based on the pairwise interchange argument, it suffices to compare two sequences which differ only in the execution of two adjacent jobs. Let sequence 1 be $\pi(1) \mapsto \pi(2) \mapsto \ldots \mapsto \pi(j) \mapsto \pi(j+1) \mapsto \ldots \mapsto \pi(n)$ and sequence 2 be $\pi(1) \mapsto \pi(2) \mapsto \ldots \mapsto \pi(j+1) \mapsto \pi(j) \mapsto \ldots \mapsto \pi(n)$ with $\pi(j) < \pi(j+1)$. For $k = 1, 2$, let $C^k = (C^k_{(1)}, C^k_{(2)}, \ldots, C^k_{(n)})$ be the departure times under sequences 1 and 2, respectively. Observe that $C^1_{(i)} = C^2_{(i)} = \sum_{m=1}^{i} P_{\pi(m)}$ for $i \neq j$ and that $C^1_{(j)} = C^1_{(j-1)} + P_{\pi(j)}$, $C^2_{(j)} = C^1_{(j-1)} + P_{\pi(j+1)}$. Since $\pi(j) < \pi(j+1)$ and $P_1 \leq_{st} P_2 \leq_{st} \ldots \leq_{st} P_n$, we have $P_{\pi(j)} \leq_{st} P_{\pi(j+1)}$ and hence $Ef(C^1_{(j)}) \leq Ef(C^2_{(j)})$ for all $f$ increasing. This completes the proof for (iii).

Now, write $f(C^{\uparrow 1})$ explicitly as a function of the processing times:

$$f(C^{\uparrow 1}) = h(P_{\pi(1)}, P_{\pi(2)}, \ldots, P_{\pi(j)}, P_{\pi(j+1)}, \ldots, P_{\pi(n)}),$$

and accordingly, write

$$f(C^{\uparrow 2}) = h(P_{\pi(1)}, P_{\pi(2)}, \ldots, P_{\pi(j+1)}, P_{\pi(j)}, \ldots, P_{\pi(n)}).$$

From the earlier analysis we know $C^1_{(j)} \leq C^2_{(j)}$ if $P_{\pi(j)} \leq P_{\pi(j+1)}$, and hence $f(C^{\uparrow 1}) \leq f(C^{\uparrow 2})$ for all $f$ increasing. Thus, $h \in \mathbf{C}_{lr}$ when $f$ is increasing. Applying Theorem 2.1 then completes the proof of (i). In a similar fashion, (ii) follows if $f$ increasing and supermodular implies $h \in \mathbf{C}^*_{hr}$, which we prove in the next lemma. ∎

**Lemma 4.2** *Consider a function $h$ that takes the following form:*

$$h(x_1, x_2, \ldots, x_n) = f(g_1(x_1), g_2(x_1, x_2), \ldots, g_n(x_1, x_2, \ldots, x_n)),$$

*where $f$ is increasing and supermodular, and $g_j$ increasing and symmetric in $(x_1, x_2, \ldots, x_j)$, for all $j = 1, \ldots, n$. Then, $h \in \mathbf{C}^*_{hr}$. Specifically,*

$$h(x_1, \ldots, x_i, x_{i+1}, \ldots, x_n) - h(x_1, \ldots, x_{i+1}, x_i, \ldots, x_n)$$

*is increasing in $x_i$ for all $x_i \geq x_{i+1}$, $i = 1, \ldots, n-1$.*

[Note: to complete the proof of Theorem 4.1, take $g_j(x_1, \ldots, x_j) = \sum_{\ell=1}^{j} x_\ell$].

**Proof**. It suffices to show that for $\tilde{x}_i \geq x_i \geq x_{i+1}$,

$$\begin{aligned}
& f\Big(\ldots,\ g_i(x_1, \ldots, x_{i-1}, \tilde{x}_i),\ g_{i+1}(x_1, \ldots, x_{i-1}, \tilde{x}_i, x_{i+1}),\ \ldots\Big) \\
& -f\Big(\ldots,\ g_i(x_1, \ldots, x_{i-1}, x_{i+1}),\ g_{i+1}(x_1, \ldots, x_{i-1}, x_{i+1}, \tilde{x}_i),\ \ldots\Big) \\
& \geq f\Big(\ldots,\ g_i(x_1, \ldots, x_{i-1}, x_i),\ g_{i+1}(x_1, \ldots, x_{i-1}, x_i, x_{i+1}),\ \ldots\Big) \\
& -f\Big(\ldots,\ g_i(x_1, \ldots, x_{i-1}, x_{i+1}),\ g_{i+1}(x_1, \ldots, x_{i-1}, x_{i+1}, x_i),\ \ldots\Big)
\end{aligned}$$

Since $g_i$ is increasing and $\tilde{x}_i \geq x_i$,

$$g_i(x_1, \ldots, x_{i-1}, \tilde{x}_i) \geq g_i(x_1, \ldots, x_{i-1}, x_i).$$

Since $f$ is increasing, it then follows that the first term above is not less than and hence can be replaced by

$$f\Big(\ldots,\ g_i(x_1, \ldots, x_{i-1}, x_i),\ g_{i+1}(x_1, \ldots, x_{i-1}, \tilde{x}_i, x_{i+1}),\ \ldots\Big).$$

Since $x_i \geq x_{i+1}$ and $g_i$ is increasing and symmetric, the above inequality follows from the supermodularity of $f$. ∎

It is well known ([6], Theorem 3.2) that in the deterministic setting the SPT rule (shortest processing time first) minimizes $\frac{1}{n}\sum_{j=1}^{n} C_j^\alpha$, $\alpha > 0$, where $C_j^\alpha$ is the $\alpha$-th power of the completion time of job $j$, while the LPT rule (longest processing time first) maximizes the same objective function. The sequences $C^s$ and $C^\ell$ in Theorem 4.1 correspond to SPT and LPT, respectively. Following standard terminologies in stochastic scheduling, we shall refer to these as SEPT and LEPT rules, respectively ("E" for "expected").

**Corollary 4.3** *Suppose the processing times follow the order in Theorem 4.1: $P_1 \le_a \ldots \le_a P_n$. Then SEPT minimizes $\frac{1}{n}\sum_{j=1}^{n} C_j^\alpha$, $\alpha > 0$ in the sense of (i) stochastic ordering, if $a = lr$; (ii) increasing convex ordering, if $a = hr$; (iii) expectation, if $a = st$. LEPT maximizes the same objective function similarly.*

**Proof**. Observe that the objective function remains the same if $C_j$ is replaced by $C_{(j)}$ for all $j$. Hence the results follow directly from Theorem 4.1. In particular, for (ii), note that the objective function belongs to $\mathbf{C}_{sup}^{\uparrow}$, which is closed under the composition of increasing convex functions ([17], Proposition 6.D.2). ∎

Another well-known result in the deterministic setting is that processing jobs in increasing order of their due dates, smallest due date first, minimizes the maximum lateness and the maximum tardiness. (This is known as Jackson's rule; see Jackson [8]; [6], Theorem 3.3). In the stochastic setting, however, this result will not hold in general. Consider the following example with two jobs. The processing time of job 1 is one and the processing time of job 2 is zero. Assume that $D_1 \le_{st} D_2$. In order to minimize the expected maximum lateness and tardiness, Jackson's rule suggest that we should schedule job 1 first. The corresponding maximum lateness is $\max(1 - D_1, 1 - D_2)$. However, if job 2 is processed first, the maximum lateness is $\max(-D_2, 1 - D_1)$, which is stochastically smaller than that following Jackson's rule. The lesson learned here is that the processing times and the due dates should be ordered in the same direction, i.e., if $D_1 \le D_2$, then we also want $P_1 \le P_2$. We shall refer to this kind of conditions as "agreeability" conditions.

**Theorem 4.4** *Suppose the processing times and due dates are ordered as follows: $P_1 \le_a P_2 \le_a \ldots \le_a P_n$ and $D_1 \le_b D_2 \le_b \ldots \le_b D_n$. Let $L^s$ ($T^s$) and $L^\ell$ ($T^\ell$) be the lateness (tardiness) vectors incurred by the sequence $1 \mapsto 2 \mapsto \ldots \mapsto n$ and the sequence $n \mapsto (n-1) \mapsto \ldots \mapsto 1$, respectively. Then*

$$Ef(L^s) \le Ef(L) \le Ef(L^\ell) \;\; \text{and} \;\; Ef(T^s) \le Ef(T) \le Ef(T^\ell),$$

*where (i) $f \in \mathbf{C}_1^{\uparrow}$ (resp. $\mathbf{C}_{5-icx}^{\uparrow}, \mathbf{C}_5^{\uparrow}$) if $(a,b) = (lr, lr)$ [resp. $(lr, hr), (lr, st)$], (ii) $f \in \mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_1$*

*(resp.* $\mathbf{C}^{\uparrow}_{sup} \cap \mathbf{C}_{5-icx}, \mathbf{C}^{\uparrow}_{3}$*) if* $(a,b) = (hr, lr)$ *[resp.* $(hr,hr),(hr,st)$*], and (iii)* $f \in \mathbf{C}^{\uparrow}_{3}$ *if* $(a,b) = (st, st)$*.*

**Proof**. Since the proofs for the lower and upper bounds are similar, we only show that for the lower bound. Under the sequence $1 \mapsto 2 \mapsto \ldots \mapsto n$, $C^s_j = C^s_{(j)}$ and thus

$$Ef(L^s) = Ef(C^s_1 - D_1, \ldots, C^s_n - D_n) = Ef(C^s_{(1)} - D_1, \ldots, C^s_{(n)} - D_n).$$

Now consider specifically the cases (i) $f \in \mathbf{C}^{\uparrow}_1$, (ii) $f \in \mathbf{C}^{\uparrow}_{sup} \cap \mathbf{C}_1$, and (iii) $f \in \mathbf{C}^{\uparrow}_3$. Note that $f$ here also belongs to the three classes in Theorem 4.1 accordingly. Hence, applying the results there (with respect to the departure sequence), we have

$$Ef(C^s_{(1)} - D_1, \ldots, C^s_{(n)} - D_n) \leq Ef(C_{(1)} - D_1, \ldots, C_{(n)} - D_n),$$

provided the processing times follow the ordering $a = lr$, $a = hr$ and $a = st$, respectively for the three cases. Next, since $C_{(1)} \leq C_{(2)} \leq \ldots \leq C_{(n)}$ *(almost surely)*, applying Theorem 3.6 yields

$$Ef(C_{(1)} - D_1, \ldots, C_{(n)} - D_n) \leq Ef(C_1 - D_1, \ldots, C_n - D_n) = Ef(L),$$

provided the due dates follow the ordering $b = lr$, $b = hr$ and $b = st$, respectively for the three cases. [Here it would be helpful to note that in order to apply both Theorem 4.1 and Theorem 3.6, it is necessary to restrict $f$ to the intersection of the respective function classes in the two theorems. For instance, $\mathbf{C}^{\uparrow}_3 = \mathbf{C}^{\uparrow}_{sep} \cap \mathbf{C}^{\uparrow}_5$, hence $f \in \mathbf{C}^{\uparrow}_3$ in case (iii) here.]

The other cases under "resp." are similarly verified. This concludes the proof for lateness. To establish the results for tardiness, note that the composite function $f(g(x_1), g(x_2), \ldots, g(x_n))$ belongs to $\mathbf{C}^{\uparrow}$ (resp. $\mathbf{C}^{\uparrow}_{sup}$, $\mathbf{C}^{\uparrow}_{sep}$) if $g$ is increasing and $f$ belongs to $\mathbf{C}^{\uparrow}$ (resp. $\mathbf{C}^{\uparrow}_{sup}$, $\mathbf{C}^{\uparrow}_{sep}$); similarly, it belongs to $\mathbf{C}^{\uparrow}_1$ (resp. $\mathbf{C}^{\uparrow}_3$, $\mathbf{C}^{\uparrow}_5$) if $g$ is increasing convex and $f$ belongs to $\mathbf{C}^{\uparrow}_1$ (resp. $\mathbf{C}^{\uparrow}_3$, $\mathbf{C}^{\uparrow}_5$). Now, since for any job $j$, $T_j = \max\{L_j, 0\}$ is an increasing and convex function of $L_j$, the results for tardiness readily follow. ∎

**Remark 4.5** From the proof of Theorem 4.4, it is clear that the results there hold for any performance measure in the form $\phi(C_j, W_j)$ for job $j$, where $\phi$ is the function specified in Theorem 3.6, and $W_j$'s are independent random variables appropriately ordered. For instance, $\phi(C_j, W_j) = W_j C_j$, the "weighted" completion time, is one such example.

## 4.2 Single Machine, No Preemption: with Arrivals

Consider the same scheduling problem as in §4.1, but now jobs are not all present at time zero, they arrive at random time epochs: $R_j$ for job $j$. In this case, we do not have results similar to Theorem 4.1 for the departure sequence. However, we may still have some positive results, which are elaborated below.

We start with considering a special case of Theorem 4.4; the results (Theorem 4.6 below) apply to systems with or without arrivals. When the processing times are iid (independent and identically distributed) random variables, in the proof of Theorem 4.4 we only need to apply Theorem 3.6; Theorem 4.1 is not needed, since any processing order yields the same departure times (i.e., equal in joint distribution). Hence the results in Theorem 4.4 can be strengthened by enlarging the classes of functions (for $f$) to those in Theorem 3.6 (since the $\phi$ function in the *lateness* vectors is the simple summation here, the increasingness can also be dropped in the cases of $\mathbf{C}_1$ and $\mathbf{C}_5$). In particular, if the objective function is maximum lateness/tardiness, then $f$ is simply the max function, which belongs to $\mathbf{C}_5^{\uparrow}$ but is also preserved under composition of increasing convex functions, i.e., $g(\max(x_1, \ldots, x_n)) \in \mathbf{C}_5^{\uparrow}$ for all $g$ that is increasing and convex. With these in mind, we have the following theorem, where SEDD (smallest expected due date) and LEDD (largest expected due date) refer to the processing sequences indexed by, respectively, $s$ and $\ell$ in Theorem 4.4.

**Theorem 4.6** *If the processing times are iid and $D_1 \leq_b D_2 \leq_b \ldots \leq_b D_n$, then with or without arrivals,*

$$Ef(L^s) \leq Ef(L) \leq Ef(L^\ell),$$

*where (i) $f \in \mathbf{C}_1$, if $b = lr$; (ii) $f \in \mathbf{C}_{5-icx}^{\uparrow}$, if $b = hr$; and (iii) $f \in \mathbf{C}_5$, if $b = st$.*

*In particular, SEDD minimizes (LEDD maximizes) the maximum lateness and maximum tardiness in the sense of*

*(a) stochastic ordering, if the due dates are likelihood ratio ordered ($b = lr$);*

*(b) increasing convex ordering, if the due dates are stochastically ordered ($b = st$).*

We note that Baccelli, Liu and Towsley [1] showed Theorem 4.6 (i) under a stronger ordering (a.s.) for the due dates, $D_1 \leq D_2 \leq \ldots \leq D_n$. Based on an argument in [1], suppose $R_j$'s are known, deterministic values, letting $R_j = D_j$ for all $j$, SEDD (LEDD) becomes FCFS, first come first served (LCFS, last come first served). Also, flow time then becomes equal to lateness: $F_j = C_j - R_j = C_j - D_j = L_j$. Hence from Theorem 4.6(i), we immediately have that

FCFS (LCFS) minimizes (maximizes) $E\sum_{j=1}^n f(F_j)$ for all $f$ convex. This recovers a previously known result (Kingman [14], Shanthikumar and Sumita [30]).

Another positive result is a restricted version of Theorem 4.1 to two job classes. This is also analogous to Hirayama, Kijima and Nishimura [13]. The proof is similar to the pairwise interchange argument in the proof of Theorem 4.1, and hence omitted.

**Theorem 4.7** *Suppose there are two independent classes of job arrivals. Within each class, the processing times are iid. Let $G_1$ and $G_2$ be the generic processing time of the two classes. Suppose $G_1 \leq_a G_2$. Let $C^s$ and $C^\ell$ follow the definition in Theorem 4.1. Then*

$$Ef(C^s_{(1)}, C^s_{(2)}, \ldots, C^s_{(n)}) \leq Ef(C_{(1)}, C_{(2)}, \ldots, C_{(n)}) \leq Ef(C^\ell_{(1)}, C^\ell_{(2)}, \ldots, C^\ell_{(n)}),$$

*where (i) $f \in \mathbf{C}^\uparrow$, if $a = lr$; (ii) $f \in \mathbf{C}^\uparrow_{sup}$, if $a = hr$; and (iii) $f \in \mathbf{C}^\uparrow_{sep}$, if $a = st$.*

Note the average flow time equals $\frac{1}{n}\sum_{j=1}^n F_j = \frac{1}{n}\sum_{j=1}^n C_{(j)} - \frac{1}{n}\sum_{j=1}^n R_j$. Applying Theorem 4.7 yields the following corollary.

**Corollary 4.8** *SEPT (LEPT) minimizes (maximizes) the average flow time in the sense of (i) stochastic ordering, (ii) increasing convex ordering, and (iii) expectation, respectively under the three cases in Theorem 4.7.*

**Remark 4.9** In general, the result in Theorem 4.7 is not true for more than two classes of jobs. Consider the following example. Suppose there are three jobs: short, medium and long; refer to these as $s$, $m$ and $\ell$, and use the same to denote their processing times. Suppose $m$ and $\ell$ are present at time zero, $s$ will arrive at a time that is smaller than $\ell$ but greater than $m$. If $m$ is processed first, then after its completion, the server has to continue with the long job $\ell$, since $s$ has not arrived yet and idling is not allowed. Even when $s$ arrives, nothing can be done, since preemption is not allowed. On the other hand, if $\ell$ is processed first, then after its completion, the server can process the short job $s$. Hence, processing shorter jobs first does not necessarily speed up completions. The reader can verify this kind of "trap" will not happen if there are only two job classes.

## 4.3 Single Machine, with Preemtions: with or without Arrivals

Here we allow preemption, everything else remains the same as in §4.1 and §4.2. We treat the case with arrivals in detail, and summarize the result for the no-arrival case at the end of this subsection. Suppose $G(t)$ is a distribution function, and denote $\bar{G}(t) = 1 - G(t)$. Recall the *hazard rate* function (associated with $G(t)$) is defined as $\gamma(t) = \frac{d}{dt}[-\log \bar{G}(t)]$; or equivalently, $\bar{G}(t) = \exp[-\int_0^t \gamma(s)ds]$. Now, let $\gamma_j(t)$ be the hazard rate of the processing time of job $j$. Similar to [34], we assume that there is an monotone function $\gamma(t)$ such that $\gamma_j(t) = \gamma(t + \tau_j)$, for some $\tau_j \geq 0$. Since $\gamma(t)$ is monotone, the processing times are hazard rate ordered, through the associated constants $\tau_j$. Let HHR (LHR) denote the schedule that processes jobs according to highest (lowest) hazard rate first. Note that with $\gamma(t)$ being increasing, when HHR is followed, once a job is started, it can only be preempted by a new arrival; it can never be preempted by any job that is present in the system. This is because with increasing hazard rate, the more a job is processed the higher its hazard rate. (This fact plays a crucial role to the validity of the results below.) The same argument applies to LHR with $\gamma(t)$ being decreasing. The main result here is the following theorem.

**Theorem 4.10** *Let $Q_t$ be the number of jobs in the system at time $t$. If $\gamma(t)$ is increasing (decreasing), then, for any $f \in \mathbf{C}^{\uparrow}_{sup}$, HHR (LHR) minimizes (maximizes)*

**(i)** *$Ef(Q_{t_1}, Q_{t_2}, \ldots, Q_{t_k})$, for all $k$ and all $(t_1, t_2, \ldots, t_k)$;*

**(ii)** *$Ef(C_{(1)}, C_{(2)}, \ldots, C_{(n)})$.*

We note that Righter and Shanthikumar [23] have recently obtained a result parallel to Theorem 4.10 under a stronger assumption: if all the jobs have a common log concave density, then the policy that schedules jobs according to the highest likelihood ratio first minimizes the corresponding objective functions for all $f \in \mathbf{C}^{\uparrow}$.

We need the following lemma to prove Theorem 4.10.

**Lemma 4.11** *Consider two systems indexed by $i = 1, 2$, each with $k$ jobs at time zero. Suppose both systems have the same job arrival process, and the jobs have the same processing time distributions, characterized by the hazard rate functions specified earlier. The only difference is: for those $k$ jobs present at time zero, $\tau_j^1 \geq \tau_j^2$, for all $j = 1, \ldots, k$. (The $\gamma$ function is common for both systems.) Let $\{Q_t^i\}$, $i = 1, 2$, be the queue length processes of the two systems (i.e., total number of jobs in either system) and $\{C_{(n)}^i, n \geq 1\}$, $i = 1, 2$ be the $n^{th}$ departures from the two systems. If $\gamma(t)$ is increasing and both systems follow HHR, then $\{Q_t^1\} \leq_{st} \{Q_t^2\}$ and $\{C_{(n)}^1, n \geq 1\} \leq_{st} \{C_{(n)}^2, n \geq 1\}$.*

**Proof**. On a common probability space, generate the processing times of the $k$ initial jobs in either system. Since these are hazard rate ordered, they are stochastically ordered, and hence the generated processing times for system 1 are smaller than those for system 2 (refer to, e.g., Stoyan [32]). On the same probability space, generate all future arrivals and their processing times, common for both systems. Since both systems operate under HHR, it is clear that every job (including the initial $k$ jobs and the new arrivals) is completed in system 1 *no later than* its counterpart in system 2. (It would be helpful at this point to take into account the assumption on the hazard rate function and its implication alluded to before Theorem 4.10.) The stochastic ordering of the two queue length processes then follows. ∎

**Remark 4.12** Similar results also hold under LHR with $\gamma(t)$ being decreasing, as the reader will verify. In addition, the results remain true for parallel machine systems, which will be used in §4.7.

**Proof** [of Theorem 4.10]. We prove for the case HHR only, the LHR case is similarly proved. Consider the case under discrete time. Results for the original continuous time problem will follow from taking limits. Index time by $t = 0, 1, 2, \ldots$; and use induction on $t$. Denote the system state at time $t$, $s(t)$, as the vector of hazard rates of jobs in the system at time $t$ arranged in the decreasing order, i.e., $s(t) = (\gamma_{[1]}, \gamma_{[2]}, \ldots)$. Let $|s(t)|$ denote the number of jobs in the system in state $s(t)$. Clearly, HHR minimizes $Ef(Q_1|s(0))$ for all $f$ increasing and any initial state $s(0)$. Now assume that HHR minimizes $Ef(Q_1, Q_2, \ldots, Q_m|s(0))$ for all $f \in \mathbf{C}^{\uparrow}_{sup}$ and any initial state $s(0)$. In order to show that HHR is the best policy for $m+1$, we argue that any policy that does not follow HHR at time 0 cannot be optimal. (This is the argument used in Buyukkoc, Varaiya and Walrand [35].) Let $p$ be the highest hazard rate in $s(0)$ and $q$ be the hazard rate of some job with $q < p$ in $s(0)$. Without loss of generality, name the job with hazard rate $p$ (resp. $q$) as job 1 (resp. 2). Consider the policy, $\sigma_{21}$, that schedules job 2 at time 0. We have

$$\begin{aligned}
&Ef(Q_1, Q_2, \ldots, Q_{m+1}|s(0); \sigma_{21}) \\
&= qEf(|s(0)| - 1, Q_2, \ldots, Q_{m+1}|s'(1)) + (1-q)Ef(|s(0)|, Q_2, \ldots, Q_{m+1}|s''(1)),
\end{aligned}$$

where $s'(1)$ and $s''(1)$ are the new states (at the next time unit). It then follows from the induction hypothesis for $m$ time units that the policy $\sigma_{21}$ should schedule jobs according to HHR from time 1 onwards. Due to the increasing hazard rate, the hazard rate of job 2 cannot be higher than that of job 1 if job 2 is not completed in one time unit. If there is no arrival at time 0 or the jobs which arrive at time 0 have lower hazard rate than $p$, then job 1 will be scheduled for service at time 1. If there are arrivals at time 0 with hazard rates higher than $p$, then job 1 will be delayed. However, job 1 will always be the first job scheduled for service among the jobs that are present at time 0. Let $T_1$ be the time that job 1 is scheduled for service.

We now illustrate how the policy $\sigma_{21}$ can be improved. Denote by $\sigma_{12}$ the policy that schedules job 1 at time 0, follows the same schedule as $\sigma_{21}$ before $T_1$, schedules job 2 at time $T_1$, and then follows HHR onwards. In order to compute the difference

$$Ef(Q_1, Q_2, \ldots, Q_{m+1}|s(0); \sigma_{12}) - Ef(Q_1, Q_2, \ldots, Q_{m+1}|s(0); \sigma_{21}),$$

there are four cases to be considered:

**(i)** with probability $(1-p)q$, job 2 is completed in one time unit but job 1 is not;

**(ii)** with probability $p(1-q)$, job 1 is completed in one time unit but job 2 is not;

**(iii)** with probability $pq$, both job 1 and job 2 are completed in one time unit;

**(iv)** with probability $(1-p)(1-q)$, neither job 1 and job 2 is completed in one time unit.

Observe that in cases (iii) and (iv) the queue length processes under $\sigma_{21}$ and $\sigma_{12}$ are the same for all t. Let $Q = (Q_1, \ldots, Q_{m+1})$, $Q^c = (Q_1, \ldots, Q_{\min(T_1, m+1)})$ and $Q^d = (Q_{T_1+1}, \ldots, Q_{m+1})$. In the case that $T_1 + 1 > m+1$, $Q^d$ is a null vector (vanishes) by convention. Thus, $Q = (Q^c, Q^d)$. Hence, we have

$$\begin{aligned}
&Ef(Q|\sigma_{21}) - Ef(Q|\sigma_{12}) \\
&= q(1-p)\Big(Ef(Q|\sigma_{21}; \text{(i)}) - Ef(Q|\sigma_{12}; \text{(i)})\Big) + p(1-q)\Big(Ef(Q|\sigma_{21}; \text{(ii)}) - Ef(Q|\sigma_{12}; \text{(ii)})\Big) \\
&= q(1-p)\Big(Ef(Q|\sigma_{21}; \text{(i)}) - Ef(Q|\sigma_{12}; \text{(i)}) + Ef(Q|\sigma_{21}; \text{(ii)}) - Ef(Q|\sigma_{12}; \text{(ii)})\Big) \qquad (6) \\
&+(p-q)\Big(Ef(Q|\sigma_{21}; \text{(ii)}) - Ef(Q|\sigma_{12}; \text{(ii)})\Big), \qquad (7)
\end{aligned}$$

where the intitial state $s(0)$ is omitted from the conditional expectations to simplify notation. Since $(Q_t|\sigma_{21}; \text{(ii)}) = (Q_t|\sigma_{12}; \text{(ii)}) + 1$ for $t \le T_1$ and $(Q_t|\sigma_{21}; \text{(ii)}) = (Q_t|\sigma_{12}; \text{(ii)})$ for $t > T_1$, the expression in (7) is non-negative. Moreover,

$$(Q^c|\sigma_{21}; \text{(ii)}) = (Q^c|\sigma_{12}; \text{(i)}) \ge (Q^c|\sigma_{12}; \text{(ii)}) = (Q^c|\sigma_{21}; \text{(i)}). \qquad (8)$$

Due to the increasing hazard rate, the hazard rate of the uncompleted job 1 in (i) is higher than that of the uncompleted job 2 in (ii). Thus, the states in (i) and (ii) at time $T_1 + 1$ satisfy the condition in Lemma 4.11 and we can construct two realizations such that the queue length in (i) is smaller than that in (ii) from $T_1 + 1$ onwards. Hence,

$$(Q^d|\sigma_{21}; \text{(ii)}) = (Q^d|\sigma_{12}; \text{(ii)}) \ge (Q^d|\sigma_{21}; \text{(i)}) = (Q^d|\sigma_{12}; \text{(i)}). \qquad (9)$$

In conjunction with (8), we have from the supermodularity of $f$ that the expression in (6) is non-negative. Thus, $Ef(Q|\sigma_{21}) \ge Ef(Q|\sigma_{12})$ and this completes the proof for part (i) of the theorem.

To show the result for the departure processes, use the idea of truncating the completion time in Glazebrook [10]. Let $C_j^k = \min(k, C_j)$, $j = 1, 2, \ldots, n$. Then follow the same arguments for the queue length process to show that HHR minimizes $Ef(C_{(1)}^k, C_{(2)}^k, \ldots, C_{(n)}^k)$ by induction on $k$. ∎

**Remark 4.13** The results in [34] are for parallel machines without arrivals. We will discuss more about this model in §4.7. Also note that the *single* dimensional version of the results in Theorem 4.10 was previously established, respectively, in Righter and Shanthikumar [22] (Theorems 2.2 and 2.4), and in Hirayama and Kijima [12] (Lemma 1). The following corollary presents a generalization of Theorem 1 in [12].

**Corollary 4.14** *If the processing times are iid, with an IFR (DFR) distribution, then FCFS minimizes (maximizes)* $Ef(F_1, F_2, \ldots, F_n)$ *for all* $f \in \mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_1$.

**Proof**. We prove the IFR case, the DFR case being similar. Notice that for each job $j$, $F_j = C_j - R_j$. Since all jobs have iid processing times, we can assume they are numbered such that $R_1 \leq R_2 \leq \ldots \leq R_n$. The desired conclusion then follows from mimicking the proof of Theorem 4.4 *mutatis mutandis* [specifically, part (ii) with $(a, b) = (hr, lr)$], with $R$ replacing $D$, and Theorem 4.10(ii) replacing Theorem 4.1. ∎

Finally, consider the no-arrival case, i.e., all jobs are present at time zero, and there is no future arrivals. As a special case of Theorem 4.10 (ii), we know that HHR (LHR) minimizes (maximizes) $Ef(C^{\uparrow})$ for all $f \in \mathbf{C}_{sup}^{\uparrow}$ *without preemptions* if the hazard rate function, $\gamma(t)$, is increasing (decreasing). Therefore, we can mimic the proof of Theorem 4.4 to establish the following theorem. (A word of caution: the results below are different from part (ii) of Theorem 4.4 in the conditions on processing times; refer to the first paragraph of this subsection.)

**Theorem 4.15** *Let* $L^s$ *and* $L^\ell$ *be the lateness vectors defined in Theorem 4.4. Suppose* $\gamma(t)$ *is increasing; suppose* $P_1 \leq_{hr} P_2 \leq_{hr} \ldots \leq_{hr} P_n$ *and* $D_1 \leq_b D_2 \leq_b \ldots \leq_b D_n$. *Then*

$$Ef(L^s) \leq Ef(L),$$

*for all* $f \in \mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_1$ *(resp.* $\mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_{5-icx}$, $\mathbf{C}_3^{\uparrow}$*) if* $b = lr$ *(resp.* $b = hr$, $b = st$*). When* $\gamma(t)$ *is decreasing, the results hold with the inequality above replaced by* $Ef(L) \leq Ef(L^\ell)$. *Similar results also hold for the tardiness vector* $T$.

## 4.4 Two-Machine Flow-Shop with Infinite Buffer

This is yet another classical problem in deterministic scheduling; see Johnson [9]; also, [6], Chapter 5. All $n$ jobs are present at time zero; each job has to be processed by the first machine and then by the second machine. There is no buffer limit between the two machines, so there is no blocking. For each job $j$, let its processing time at the first (second) machine be $A_j$ ($B_j$). Again, consider permutation schedules only (no preemption and no insertion of idle time). The objective is to minimize the makespan (i.e., time until completion of all $n$ jobs on both machines). What's known as "Johnson's rule" says that job $i$ should precede job $j$ if and only if $\min(A_i, B_j) \leq \min(A_j, B_i)$.

In the stochastic setting, assume the processing times (at the two machines) are independent random variables. Talwar [33] conjectured that in the case of *exponential* processing times job $i$ should precede job $j$ if and only if $(EA_i)^{-1} - (EB_i)^{-1} \geq (EA_j)^{-1} - (EB_j)^{-1}$. This conjecture was later proved by Ku and Niu [15] and other authors (see [15]). Under the agreeability conditions $A_i \leq_{lr} A_j$ and $B_i \geq_{lr} B_j$, it is shown in [15] that job $i$ should precede job $j$ in order to minimize the makespan in the sense of stochastic ordering. Here we complement this result with other stochastic order relations.

**Theorem 4.16** *Job $i$ should precede job $j$ in order to minimize the makespan in the sense of*

**(i)** *stochastic ordering, if $A_i \leq_{lr} A_j$ and $B_i \geq_{lr} B_j$,*

**(ii)** *increasing convex ordering, if $A_i \leq_{hr} A_j$ and $B_i \geq_{hr} B_j$,*

**(iii)** *expectation, if $A_i \leq_{st} A_j$ and $B_i \geq_{st} B_j$.*

**Proof**. Without loss of generality, we consider two sequences which differ from each other only in the order of the execution of jobs $i$ and $j$. Let sequence 1 be $1 \mapsto 2 \mapsto \ldots \mapsto i \mapsto \ldots \mapsto j \mapsto \ldots \mapsto n$ and sequence 2 be $1 \mapsto 2 \mapsto \ldots \mapsto j \mapsto \ldots \mapsto i \mapsto \ldots \mapsto n$. Let $C^1_{max}$ and $C^2_{max}$ be the makespan under the two sequences. We want to show $Ef(C^1_{max}) \leq Ef(C^2_{max})$ for all $f$ (i) increasing, (ii) increasing convex, and (iii) linear, respectively for the three cases listed above.

From [6], pp. 91, we have

$$C^1_{max} = \sum_{k=1}^{n} B_k + \max(Y^1_1, \ldots, Y^1_n), \text{ where } Y^1_k = \sum_{\ell=1}^{k} A_\ell - \sum_{\ell=1}^{k-1} B_\ell.$$

Similarly,

$$C^2_{max} = \sum_{k=1}^n B_k + \max(Y^2_1, \ldots, Y^2_n), \text{ where } Y^2_k = \sum_{\ell=1}^k A_{\pi(\ell)} - \sum_{\ell=1}^{k-1} B_{\pi(\ell)}$$

and $\pi$ is the permutation schedule defined by sequence 2, i.e., $\pi(k) = k$ for $k \neq i, j$, $\pi(i) = j$ and $\pi(j) = i$. Now we introduce a fictitious makespan,

$$C^3_{max} = \sum_{k=1}^n B_k + \max(Y^3_1, \ldots, Y^3_n), \text{ where } Y^3_k = \sum_{\ell=1}^k A_{\pi(\ell)} - \sum_{\ell=1}^{k-1} B_\ell.$$

We want to show $Ef(C^1_{max}) \leq Ef(C^3_{max})$ and $Ef(C^3_{max}) \leq Ef(C^2_{max})$. Observe that $C^1_{max}$ can be obtained from $C^3_{max}$ by interchanging $A_i$ and $A_j$. Thus, if we write $C^1_{max} = h(A_i, A_j)$, and accordingly $C^3_{max} = h(A_j, A_i)$, it then follows from Theorem 2.1 that $Ef(C^1_{max}) \leq Ef(C^3_{max})$, for all $f$ (i) increasing, (ii) increasing convex, and (iii) linear, provided, respectively, (i) $h(A_j, A_i) \in \mathbf{C}_{lr}$, (ii) $h(A_j, A_i) \in \mathbf{C}^*_{hr}$, and (iii) $h(A_j, A_i) \in \mathbf{C}_{st}$. Taking into account the relationship between the three classes (cf. §2.2), we only need to show $h(A_j, A_i) \in \mathbf{C}_{st}$ and that it is increasing in $A_j$.

Observe that

$$C^3_{max} - C^1_{max} = \max(Y^3_1, \ldots, Y^3_n) - \max(Y^1_1, \ldots, Y^1_n),$$

and that $Y^3_k = Y^1_k$ for $k = 1, \ldots, i-1$, and $k = j, \ldots, n$, and $Y^3_k = Y^1_k - A_i + A_j$ for $k = i, \ldots, j-1$. Moreover, $Y^1_k$, $k \leq j-1$ do *not* involve $A_j$. Thus, if $\max(Y^1_1, \ldots, Y^1_n) = Y^1_k$ for some $k \leq j-1$, then $C^3_{max} - C^1_{max} = C^3_{max} - Y^1_k$ is increasing in $A_j$. If $\max(Y^1_1, \ldots, Y^1_n) = Y^1_k$ for some $k > j-1$, then $\max(Y^3_1, \ldots, Y^3_n) = Y^3_\ell$ for some $\ell \geq i$ due to the fact that $Y^3_k = Y^1_k$ for $k = 1, \ldots, i-1$ and $k = j, \ldots, n$. Since $Y^3_\ell$, $\ell \geq i$ and $Y^1_k$, $k \geq j$ are *linear* in $A_j$, $Y^3_\ell - Y^1_k$, $\ell \geq i$, $k \geq j$ are constant (and hence increasing) in $A_j$. Since max is a continuous function, $C^3_{max} - C^1_{max}$ is increasing in $A_j$, and hence $h(A_j, A_i)$ is in $\mathbf{C}_{st}$.

It is also clear that $C_{max}$ is increasing in $A_j$ and $B_j$ for all $j$ and hence $h(A_j, A_i)$ is increasing in $A_j$. This concludes the proof for $Ef(C^1_{max}) \leq Ef(C^3_{max})$. That $Ef(C^3_{max}) \leq Ef(C^2_{max})$ follows similar arguments. ∎

In the case that the processing times at the first machine is *smaller* (*larger*) that those at the second machine, we have the following theorem.

**Theorem 4.17** *Suppose* $A_k \leq B_k$ *(resp.* $A_k \geq B_k$*),* $k = 1, 2, \ldots, n$*; and* $A_1 \leq_a A_2 \leq_a \ldots \leq_a A_n$ *(resp.* $B_n \leq_a B_{n-1} \leq_a \ldots \leq_a B_1$*). Then the sequence* $1 \mapsto 2 \mapsto \ldots \mapsto n$ minimizes *the makespan in the sense of (i) stochastic ordering, if* $a = lr$*; (ii) increasing convex ordering, if* $a = hr$*; and (iii) expectation, if* $a = st$.

**Proof**. Consider the case $A_k \leq B_k$; the other case is similarly proved. Let $C^1_{max}$, $C^2_{max}$ and $C^3_{max}$ follow the definition in the proof of Theorem 4.16 with jobs $i$ and $j$ being two adjacent jobs, i.e., $j = i+1$. Similar to the proof of Theorem 4.16, we can establish $Ef(C^1_{max}) \leq Ef(C^3_{max})$. Next, we prove $C^3_{max} \leq C^2_{max}$ under the assumption that $A_k \leq B_k$, $k = 1, 2, \ldots, n$. Since $j = i+1$, we have $Y^2_k = Y^3_k$, for all $k \neq i+1$. Note that $Y^3_{i+1} = Y^3_i + A_i - B_i \leq Y^3_i$ since $A_i \leq B_i$. Thus,

$$\max(Y^3_1, \ldots, Y^3_i, Y^3_{i+1}, Y^3_{i+2}, \ldots, Y^3_n) = \max(Y^3_1, \ldots, Y^3_i, Y^3_{i+2}, \ldots, Y^3_n)$$
$$= \max(Y^2_1, \ldots, Y^2_i, Y^2_{i+2}, \ldots, Y^2_n) \leq \max(Y^2_1, \ldots, Y^2_i, Y^2_{i+1}, Y^2_{i+2}, \ldots, Y^2_n),$$

which completes the proof. ■

## 4.5 Two-Machine Flow Shop without Buffer

Pinedo [18] considers a scheduling problem similar to the one in §4.4 except there is no buffer in front of the second machine. Thus, a job completed by the first machine is blocked if the second machine is busy. It is further assumed that the processing times at both machines are i.i.d. random variables, i.e., $A_j =_{st} B_j$, $j = 1, 2, \ldots, n$. Among other results, it is shown that if $A_1 \leq_{st} A_2 \leq_{st} \ldots \leq_{st} A_n$, then the odd-even sequence (in [18] this is referred to as the SEPT-LEPT sequence) minimizes the *expected* makespan.

Let's have a closer look at this scheduling problem. Note that the makespan can be expressed as $\sum_{i=1}^{n+1} \max(A_{\pi(i)}, B_{\pi(i-1)})$, where $\pi \in \Pi^1$ (one-cycle permutation), and $A_{\pi(n+1)} = B_{\pi(0)} = 0$ by definition. Then clearly, the makespan problem belongs to the setting of Theorem 3.11. Hence, more general results readily follow:

**Theorem 4.18** *Suppose the processing times at the two machines are two independent vectors, $A =_{st} B$, both with independent components. Suppose $A_1 \leq_a A_2 \leq_a \ldots \leq_a A_n$. Suppose the objective function is*

$$f\{\max(A_{\pi(i)}, B_{\pi(i-1)}),\ i = 1, \ldots, n+1\},$$

*where, $A_{n+1} = B_0 = 0$. Then, among all $\pi \in \Pi^1$ (one-cycle permutations),*

*(i) the odd-even (small-large) sequence minimizes (maximizes) the objective function, $\forall f \in \mathbf{C}^{\uparrow}_1$, in the sense of* stochastic ordering, *provided $a = lr$;*

*(ii) the result in (i) holds for $f \in \mathbf{C}^{\uparrow}_5$ in* expectation *if $a = st$;*

*(iii) if $A = B$ (a.s.) (instead of $A =_{st} B$), then the result in (ii) holds for $f \in \mathbf{C}^{\uparrow}_3$.*

In particular, when $f$ takes the form of a simple summation, e.g., in the case of makespan, then $f \in \mathbf{C}_3^{\uparrow} \subset \mathbf{C}_5^{\uparrow} \subset \mathbf{C}_1^{\uparrow}$. Hence, all three parts in the above theorem apply. In particular, the minimization case of part (ii) recovers Pinedo's original result.

An easy special case for part (i) is that of exponential processing times. (Recall, exponential random variables can be ordered under likelihood ratio ordering according to their means.)

### 4.6 $m$-Machine Flow-Shop with Infinite Buffer

Pinedo and Weber [20] study the permutation schedules for processing $n$ jobs on $m$ machines in tandem, with infinite buffer in front of each machine. All $n$ jobs are present at time zero. For any given job $i$, the processing times on all $m$ machines are equal, denoted $P_i$ (a random variable). Among other things, they show that the makespan, $C_{max}$, can be expressed as follows:

$$C_{max} = \max(C_1, \ldots, C_n) = \sum_{i=1}^{n} P_i + (m-1)\max(P_1, \ldots, P_n). \tag{10}$$

Based on this expression, Pinedo and Wie [21], Theorem 1, compare the makespan for two systems (indexed by superscripts 1 and 2 below) with different processing times. In particular, they assume that $P_i$'s are independent, and follow the distribution functions with "proportional hazard rates": $G_i(t) = 1 - \exp\left(-\theta_i \int_0^t \gamma(s)ds\right)$ for some constant $\theta_i > 0$ and some hazard rate function $\gamma(\cdot)$ that is common for all $i$. Thus, the processing times $P = (P_i)_{i=1}^n$ are parameterized by the vector $\theta = (\theta_1, \ldots, \theta_n)$. They show that if $\theta^1 \prec \theta^2$, then $C^1_{max} \leq_{st} C^2_{max}$.

Below, we show that similar results hold under assumptions that are weaker than proportional hazard rates. To this end, we first introduce a hierarchy of conditions, in increasing order of generality, all of which include proportional hazard rates as a special case.

**Definition 4.19** *(Shaked and Shanthikumar [26, 27, 28])* $\{X(\theta)\}$ *is a family of random variables parameterized by* $\theta$*. Let* $g(t;\theta)$ *and* $G(t;\theta)$ *be the density and the distribution of* $X(\theta)$*.* $\{X(\theta)\}$ *is called*

**(i)** SICX(hr) *if the hazard function,* $\gamma(t;\theta) = g(t;\theta)/[1 - G(t;\theta)]$ *is decreasing and concave in* $\theta$ *for each fixed* $t \in \Re_+$*;*

**(ii)** SICX(st) *if the survival function,* $1 - G(t;\theta)$ *is increasing and convex in* $\theta$ *for each fixed* $t \in \Re_+$*;*

**(iii)** SICX(sp) *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 (a)* $\hat{X}_i$ *has the distribution* $G(t;\theta_i)$, *(b)* $\hat{X}_4 \geq \max(\hat{X}_2,\hat{X}_3)$ *and (c)* $\hat{X}_1+\hat{X}_4 \geq \hat{X}_2+\hat{X}_3$.

It is known (see Liyanage and Shanthikumar [16], as well as the above references of Shaked and Shanthikumar) that proportionate hazard rates $\Rightarrow$ SICX(hr) $\Rightarrow$ SICX(st) $\Rightarrow$ SICX(sp).

**Theorem 4.20** *Let* $\{X(\theta)\}$ *be a family of* independent *random variables parameterized by* $\theta$. *Write* $\theta=(\theta_i)_{i=1}^n$. *Then,* $\theta^1 \prec \theta^2$ *implies*

$$Ef(X(\theta_1^1),...,X(\theta_n^1)) \leq Ef(X(\theta_1^2),...,X(\theta_n^2))$$

*for all* $f$ *that are*

*(i) symmetric and increasing, if* $\{X(\theta)\}$ *is SICX(hr);*

*(ii) symmetric, increasing and submodular, if* $\{X(\theta)\}$ *is SICX(st);*

*(iii) symmetric, increasing, submodular and convex (in each variable), i.e.,* $f \in \mathbf{C}_5^{\uparrow}$, *if* $\{X(\theta)\}$ *is SICX(sp).*

Part (i) and part (ii) above are due to Liyanage and Shanthikumar [16], while part (iii) is due to Chang [4]. In fact, in the parametric case, the three parts constitute a hierarchy of stochastic majorization, based on different notions of stochastic convexity.

Now, coming back to the scheduling problem introduced at the beginning of this subsection. Clearly, the makespan function in [21] satisfies all three cases in Theorem 4.20. Also notice that the class of functions in (i) are closed under composition with increasing functions, and the class of functions in (ii) are closed under composition with increasing concave functions. Hence, the following results are immediate:

**Theorem 4.21** *Suppose the processing times are independent and parameterized following the specifications in Theorem 4.20. Then,* $\theta^1 \prec \theta^2$ *implies (i)* $C^1_{max} \leq_{st} C^2_{max}$, *(ii)* $C^1_{max} \leq_{icv} C^2_{max}$, *(iii)* $EC^1_{max} \leq EC^2_{max}$, *corresponding to the three cases in 4.20.*

We note that the increasingness in SICX can be replaced by decreasingness without affecting the conclusions in the above theorem. Hence, a similar extension of Theorem 2 of [21] is also possible.

**Remark 4.22** Equation (10) is also the completion time of the $n$-th departure, i.e., $C_{(n)}$. Thus, for any permutation schedule, $\pi$, the completion time of the $j$-th departure, $C_{(j)}$, is $\sum_{i=1}^{j} P_{\pi(i)} + (m-1)\max(P_{\pi(1)}, \ldots, P_{\pi(j)})$. Since $C_{(j)}$ is *symmetric* and *increasing* in $(P_{\pi(1)}, \ldots, P_{\pi(j)})$, Lemma 4.2 can be applied. The results in Theorems 4.1 and 4.4 can be extended similarly to $m$-machine flow-shop with infinite buffer and identical processing times in front of each machine.

### 4.7 Parallel Machines

Weber, Varaiya and Walrand [35] consider scheduling problems on $m$ parallel machines. They assume that the processing times of the $n$ jobs are *stochastically* ordered, i.e., $P_1 \leq_{st} P_2 \leq_{st} \ldots \leq_{st} P_n$ and machine $i$ becomes available at time $\tau_i \geq 0$. They show that among all *nonpreemptive* policies, the SEPT policy maximizes the objective function, $E\sum_{j=1}^{n} f(C_j)$, where $f$ is decreasing and convex and $C_j$ is the completion time of job $j$. In other words, $f(C_1, C_2, \ldots, C_n)$ is maximized in the expectation for the class of functions in $\mathbf{C}_3^{\downarrow}$.

It is perhaps interesting to note that the steps that lead to the result in [35] are quite similar to the steps that we used to reach many of the conclusions in this paper. For instance, it was first shown in [35] that the objective function, with respect to $\tau = (\tau_1, \ldots, \tau_n)$, is in $\mathbf{C}_5^{\downarrow}$ (which we will use later). Next, it was argued that, with respect to a pairwise interchange of two jobs, the objective function belongs to $\mathbf{C}_{st}$. The main result then follows. These are almost parallel to our development from Lemma 3.4 to Theorem 3.6. However, we must note that the objective function in [35] does not take the form of $f \odot \phi$, hence it is still necessary to go through the specific proofs there in order to reach the desired conclusion.

Here we consider *preemptive* policies. We shall assume that there are $n$ jobs, all present at time zero. As in §4.3, we further assume that the processing times are iid, with a common hazard rate function $\gamma(t)$ and that each job may have already received a distinct amount of service. Under these conditions, Weber [34] showed that HHR minimizes the expected sum of completion times among the class of preemptive policies when $\gamma(t)$ is monotone. In the following, we extend his result to the class of functions in $\mathbf{C}_3^{\downarrow}$ when $\gamma(t)$ is increasing.

**Theorem 4.23** *If $\gamma(t)$ is increasing, then HHR maximizes $Ef(C_1, C_2, \ldots, C_n)$ for all $f \in \mathbf{C}_3^{\downarrow}$.*

A special case of the above assumption on the processing times is when they are exponentially distributed with different means (see [34]). In this case, HHR is nothing but SEPT. This recovers the previously known result: SEPT minimizes $E\sum_{j=1}^{n} g(C_j)$ for all $g$ increasing concave when the processing times are exponentially distributed. (See Pinedo [19], Theorem 6.)

We note that Chang, Hordijk and Righter [5] have recently obtained a result parallel to Theorem 4.23 under a stronger assumption: if all the jobs have a common log concave density, then the policy that schedules jobs according to the highest likelihood ratio first maximizes $Ef(C_1, C_2, \ldots, C_n)$ for all $f \in \mathbf{C}_1^{\downarrow}$.

**Proof**. The proof here is an adaptation of the proof in Glazebrook [10]. Consider the discrete-time problem. Let $C_j^k = \min(k, C_j)$ be the truncated flowtime of job $j$, $j = 1, 2, \ldots, n$. Let $m(t)$ be the number of machines functioning (i.e., processing jobs) at time $t$. We first show that HHR minimizes $Ef(C^k)$ for all $k$ and all $f$ that is increasing, separable and Schur concave if $m(t)$ is increasing in $t$. The proof is by induction on $m(0)$ and on $k$. Let $S(m^*, k^*)$ denote the statement, "the theorem is true for $m(0) = m^*$ and $k = k^*$." Then, to prove the theorem, we will have to show the following:

**(i)** $S(m, 1)$, $(1 \leq m \leq n)$,

**(ii)** $S(m, k^*)$, $(1 \leq m \leq n) \Rightarrow S(1, k^* + 1)$,

**(iii)** $S(m, k^*)$, $(1 \leq m \leq n)$ and $S(m, k^*{+}1)$, $(1 \leq m \leq m^*{-}1)) \Rightarrow S(m^*, k^*{+}1)$, $(2 \leq m^* \leq n)$.

It is easy to verify (i). To prove (ii), let job 1 be the one with the highest hazard rate and job 2 be any other job with a smaller hazard rate. We want to show that starting job 2 at $t = 0$ cannot be optimal.

First note that if for some reason job 2 is not processed at $t = 1$ (e.g., it was completed, or replaced by another job), in other words, if the machine that was processing job 2 at $t = 0$ is processing another job at $t = 1$, then applying to this particular machine and the time slots $t = 0, 1$ the pairwise interchange argument similar to that in the proof of Theorem 4.1, we immediately reach the conclusion that starting job 2 at $t = 0$ cannot be optimal. Thus, we can assume that both jobs 1 and 2 are processed at $t = 1$, provided job 2 is not completed in $t = 0$. Moreover, since $\gamma(t)$ is increasing, we know that jobs 1 and 2 will be processed until they are completed.

Denote the policy that starts from job 2 at $t = 0$ and thereafter follows HHR by $S^2$, and by $S^1$ HHR itself. Let $F(S, k) = Ef(C^k|S)$ be the objective value under the policy $S$. Note that under $S^1$, $C_1 = X_1$, $C_2 = X_2 + 1$, and $C_3, \ldots, C_n$ will be the departure sequence (under SEPT) for a problem in which machines 1 and 2 are available at $X_1$ and $X_2 + 1$. Similarly, under $S^2$, $C_1 = X_1 + 1$, $C_2 = X_2$, and $C_3, \ldots, C_n$ will be the departure sequence (under SEPT) for a problem in which machines 1 and 2 are available at $X_1 + 1$ and $X_2$. Thus, if we denote $F(S^1, k^* + 1)$ by $Eh(X_1, X_2 + 1)$ for some function $h(x_1, x_2)$ with $x_1$ and $x_2$ being the available times for machines 1 and 2. Then $F(S^2, k^* + 1) = Eh(X_2, X_1 + 1)$. Since we assume that $f$

is increasing, separable and Schur concave, it follows from Lemmas 1 and 2(a) in [35] that $-h$ is in $\mathbf{C}_5^{\downarrow}$. Since $X_1 \leq_{hr} X_2$ and $0 < 1$, we have from Theorem 3.2(ii) (with $\phi(x,y) = x + y$) that $F(S^1, k^* + 1) \leq F(S^2, k^* + 1)$. Thus, $S^2$ cannot be optimal. HHR is therefore optimal in $S(1, k^* + 1)$.

It remains to prove (iii). This is based on induction on $m^*$. We argue that (iii) holds for $m^* = 2$ first. The case $m^* \geq 3$ will follow a similar argument. Suppose we schedule a job on machine 1 at $t = 0$, then the problem reduces to $S(1, k^*+1)$. Then it follows from the induction hypothesis that we should also schedule the highest hazard rate job, which is job 1, at $t = 0$. However, this policy cannot be better than the policy that schedules job 1 and the second highest hazard rate job at $t = 0$ according to the induction hypothesis for $S(1, k^* + 1)$. Thus, HHR is optimal. ∎

## 5 Summary and Concluding Remarks

We summarize our main results in stochastic scheduling in the following three tables. Table 1 lists the results for the departure processes. For example, the first row shows that for single-machine systems without arrivals and without preemption, the SEPT policy minimizes $Ef(C_1, C_2, \ldots, C_n)$, where $C_j$ is the completion time of job $j$, over the class of increasing and symmetric functions if the processing times are likelihood-ratio ordered. Details are listed in Theorem 4.1. (The functions involved in the Table are defined at the beginning of §4, prior to §4.1.)

Table 1. Departure Processes

| Machine | Arrival | Preemption | Processing times | $f(C_1, C_2, \ldots, C_n)$ | Ref |
|---|---|---|---|---|---|
| 1 | No | No | lr | symmetric, $\mathbf{C}^{\uparrow}$ | Thm 4.1 |
| 1 | No | No | hr | symmetric, $\mathbf{C}^{\uparrow}_{sup}$ | Thm 4.1 |
| 1 | No | No | st | symmetric, $\mathbf{C}^{\uparrow}_{sep}$ | Thm 4.1 |
| 1 | two classes | No | lr | symmetric, $\mathbf{C}^{\uparrow}$ | Thm 4.7 |
| 1 | two classes | No | hr | symmetric, $\mathbf{C}^{\uparrow}_{sup}$ | Thm 4.7 |
| 1 | two classes | No | st | symmetric, $\mathbf{C}^{\uparrow}_{sep}$ | Thm 4.7 |
| 1 | Yes | Yes | hr (increasing) | symmetric, $\mathbf{C}^{\uparrow}_{sup}$ | Thm 4.10 |
| 1 | Yes | Yes | lr (increasing) | symmetric, $\mathbf{C}^{\uparrow}$ | [23] |
| m | No | Yes | hr (increasing) | Schur concave, $\mathbf{C}^{\uparrow}_{sep}$ | Thm 4.23 |
| m | No | Yes | lr (increasing) | Schur concave, $\mathbf{C}^{\uparrow}$ | [5] |
| m | No | No | st | Schur concave, $\mathbf{C}^{\uparrow}_{sep}$ | [35] |

Table 2 summarizes the results for the lateness vector in single-machine systems. For example, the first row shows that for single-machine systems without arrivals and without preemptions, the sequence $1 \mapsto 2 \mapsto \ldots \mapsto n$, which corresponds to the SEDD policy, minimizes $Ef(L_1, L_2, \ldots, L_n)$, where $L_j$ is the lateness of job $j$, over the class of increasing and Schur convex function if the processing times are likelihood ratio ordered, i.e., $P_1 \le_{lr} P_2 \le_{lr} \ldots \le_{lr} P_n$, and the due dates are also likelihood-ratio ordered, i.e., $D_1 \le_{lr} D_2 \le_{lr} \ldots \le_{lr} D_n$. This result is listed in Theorem 4.4. (The functions involved in the Table are defined in §2.1.)

Table 2. Lateness in Single-Machine Systems

| Arrival | Preemption | $P_j$ | $D_j$ | $f(L_1, L_2, \ldots, L_n)$ | Ref |
|---|---|---|---|---|---|
| No | No | $\le_{lr}$ | $\le_{lr}$ | $\mathbf{C}_1^{\uparrow}$ | Thm 4.4 |
| No | No | $\le_{lr}$ | $\le_{hr}$ | $\mathbf{C}_{5-icx}^{\uparrow}$ | Thm 4.4 |
| No | No | $\le_{lr}$ | $\le_{st}$ | $\mathbf{C}_5^{\uparrow}$ | Thm 4.4 |
| No | No | $\le_{hr}$ | $\le_{lr}$ | $\mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_1$ | Thm 4.4 |
| No | No | $\le_{hr}$ | $\le_{hr}$ | $\mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_{5-icx}$ | Thm 4.4 |
| No | No | $\le_{st}$ | $\le_{st}$ | $\mathbf{C}_3^{\uparrow}$ | Thm 4.4 |
| Yes | No | $=_{st}$ | $\le_{lr}$ | $\mathbf{C}_1$ | Thm 4.6 |
| Yes | No | $=_{st}$ | $\le_{hr}$ | $\mathbf{C}_{5-icx}^{\uparrow}$ | Thm 4.6 |
| Yes | No | $=_{st}$ | $\le_{st}$ | $\mathbf{C}_5$ | Thm 4.6 |
| No | Yes | $\le_{hr}$ (IFR) | $\le_{lr}$ | $\mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_1$ | Thm 4.15 |
| No | Yes | $\le_{hr}$ (IFR) | $\le_{hr}$ | $\mathbf{C}_{sup}^{\uparrow} \cap \mathbf{C}_{5-icx}$ | Thm 4.15 |
| No | Yes | $\le_{hr}$ (IFR) | $\le_{st}$ | $\mathbf{C}_3^{\uparrow}$ | Thm 4.15 |

Table 3 lists results for minimizing makespan in two-machine flow shops. For example, the first row shows that for two-machine flow shops with infinite buffer between the two machines, under the assumptions for the processing times of job $j$ at the first machine, $A_j$, and the second machine, $B_j$, the sequence $1 \mapsto 2 \mapsto \ldots \mapsto n$ minimizes the makespan, $C_{max}$, in the sense of stochastic ordering. This result is listed in [15].

Table 3. Makespan in Two-Machine Flow Shops

| Buffer | Assumption | Sequence | $C_{max}$ | Ref |
|---|---|---|---|---|
| Infinite | $A_1 \le_{lr} \ldots \le_{lr} A_n$, $B_1 \ge_{lr} \ldots \ge_{lr} B_n$ | $1 \mapsto \ldots \mapsto n$ | st | [15] |
| Infinite | $A_1 \le_{hr} \ldots \le_{hr} A_n$, $B_1 \ge_{hr} \ldots \ge_{hr} B_n$ | $1 \mapsto \ldots \mapsto n$ | icx | Thm 4.16 |
| Infinite | $A_1 \le_{st} \ldots \le_{st} A_n$, $B_1 \ge_{st} \ldots \ge_{st} B_n$ | $1 \mapsto \ldots \mapsto n$ | mean | Thm 4.16 |
| Infinite | $A_1 \le_{lr} \ldots \le_{lr} A_n$, $A_k \le B_k$, $\forall k$, | $1 \mapsto \ldots \mapsto n$ | st | Thm 4.17 |
| Infinite | $A_1 \le_{hr} \ldots \le_{hr} A_n$, $A_k \le B_k$, $\forall k$, | $1 \mapsto \ldots \mapsto n$ | icx | Thm 4.17 |
| Infinite | $A_1 \le_{st} \ldots \le_{st} A_n$, $A_k \le B_k$, $\forall k$, | $1 \mapsto \ldots \mapsto n$ | mean | Thm 4.17 |
| Infinite | $B_1 \ge_{lr} \ldots \ge_{lr} B_n$, $A_k \ge B_k$, $\forall k$, | $1 \mapsto \ldots \mapsto n$ | st | Thm 4.17 |
| Infinite | $B_1 \ge_{hr} \ldots \ge_{hr} B_n$, $A_k \ge B_k$, $\forall k$, | $1 \mapsto \ldots \mapsto n$ | icx | Thm 4.17 |
| Infinite | $B_1 \ge_{st} \ldots \ge_{st} B_n$, $A_k \ge B_k$, $\forall k$, | $1 \mapsto \ldots \mapsto n$ | mean | Thm 4.17 |
| zero | $A_1 \le_{lr} \ldots \le_{lr} A_n$, $A_k =_{st} B_k$, $\forall k$, | odd-even | st | Thm 4.18 |
| zero | $A_1 \le_{st} \ldots \le_{st} A_n$, $A_k =_{st} B_k$, $\forall k$, | odd-even | mean | [18] |
| zero | $A_1 \le_{st} \ldots \le_{st} A_n$, $A_k = B_k$, $\forall k$, | odd-even | mean | Thm 4.18 |

We have presented in this paper a unified approach to solve stochastic scheduling problems through stochastic rearrangement inequalities. Extensions are possible in several directions. For instance, one can consider the multivariate *joint* order relations among processing times and due dates (cf. §2.2 and [31]), instead of assuming independence as we did throughout §4. One can also apply the approach here to study some of the scheduling problems in more complex systems, such as fork-join queues. Some results in this direction will be reported in another paper.

**Acknowledgements:**

Insightful discussions with Randy Nelson and with Michael Pinedo at various stages are gratefully acknowledged. This work was undertaken while David Yao was an academic visitor at IBM T.J. Watson Research Center. He thanks Peter Franaszek, Steve Lavenberg and Randy Nelson for their hospitality. His research is also supported in part by NSF under Grants ECS-89-96201 and DDM-91-08540.

## References

[1] F. Baccelli, Z. Liu and D. Towsley, "Extremal Scheduling of Parallel Processing with and without Real-Time Constraints," preprint, 1990.

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

[3] C. Buyukkoc, P. Varaiya, and J. Walrand, "The $c\mu$ Rule Revisited," *Adv. Appl. Prob.*, Vol. 17, pp. 237-238, 1985.

[4] C.S. Chang, "A New Ordering for Stochastic Majorization: Theory and Applications," to appear in *Adv. Appl. Prob.*, Sep., 1992.

[5] C.S. Chang, A. Hordijk and R. Righter, "The Stochastic Optimality of SEPT in Parallel Machine Scheduling," IBM RC-17243, 1991.

[6] R.W. Conway, W.L. Maxwell and L.W. Miller, *Theory of Scheduling.* Addison-Welsey Publishing Company, Reading, M.A., 1967.

[7] P.W. Day, "Rearrangement Inequalities," *Can. J. Math.*, Vol. 24, pp. 930-943, 1972.

[8] J. R. Jackson, "Scheduling a Production Line to Minimize Maximum Tardiness," Research Report 43, *Management Science Research project*, UCLA, January 1955.

[9] S.M. Johnson, "Optimal Two- and Three-Stage Production Schedules with Setup Times Included," *Nav. Res. Log. Quart.*, Vol. 1, pp. 61-68, 1954.

[10] K.D. Glazebrook, "Scheduling Tasks with Exponential Service Times on Parallel Processors," *J. Appl. Prob.*, Vol. 16, 685-689, 1979.

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

[12] T. Hirayama, and M. Kijima, "An Extremal Property of FIFO Displine in G/IFR/1 Queues," *Adv. Appl. Prob.*, Vol. 21, pp. 481-484, 1989.

[13] T. Hirayama, M. Kijima and S. Nishimura, "Further Results for Dynamic Scheduling of Multiclass G/G/1 Queues," *J. Appl. Prob.*, Vol. 26, 595-603, 1989.

[14] J.F.C. Kingman, "The Effect of Queue Displine On Waiting Time Variance," *Proc. Camb. Phil. Soc.*, Vol. 58, pp. 163-164, 1962.

[15] P.-S. Ku and S.-C. Niu, "On Johnson's Two-Machine Flow Shop With Random Processing Times," *Oper. Res.*, Vol. 34, pp. 130-136, 1986.

[16] L. Liyanage and J.G. Shanthikumar, "Allocation Through Stochastic Schur Convexity and Stochastic Arrangement Increasingness," preprint, 1990.

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

[18] M. Pinedo, "Minimizing the Expected Makespan in Stochastic Flow Shops," *Oper. Res.*, Vol. 30, pp. 148-162, 1982.

[19] M. Pinedo, "Stochastic Scheduling with Release Dates and Due Dates," *Oper. Res.*, Vol. 31, pp. 559-572, 1983.

[20] M. Pinedo and R.R. Weber, "Inequalities and Bounds in Stochastic Shop Scheduling," *SIAM J. Appl. Math.*, Vol. 44, pp. 869-879, 1984.

[21] M. Pinedo and S.H. Wie, "Inequalities for Stochastic Flow Shops and Job Shops," *Applied Stochastic Models and Data Analysis*, Vol. 2, pp. 61-69, 1986.

[22] R. Righter and J.G. Shanthikumar, "Scheduling Multiclass Single Server Queueing Systems to Stochastically Maximize the Number of Successful Departures," *Prob. Eng. Inf. Sciences*, Vol. 3, pp. 323-334, 1989.

[23] R. Righter and J.G. Shanthikumar, "Extremal Properties of the FIFO Displine in Queueing Networks," to appear in *J. Appl. Prob.*, 1991.

[24] J. Riordan, *An Introduction to Combinatorial Analysis.* Princeton University Press, 1978.

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

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

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

[28] M. Shaked and J.G. Shanthikumar, "Convexity of a Set of Stochastically Ordered Random Variables," *Adv. Appl. Prob.*, Vol. 22, pp. 160-177, 1990.

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

[30] J.G. Shanthikumar and U. Sumita, "Convex Ordering of Sojourn Times in Single-Server Queues: Extremal Properties of FIFO and LIFO Service Displines," *J. Appl. Prob.*, Vol. 24, pp. 737-748, 1987.

[31] J.G. Shanthikumar and D.D. Yao, "Bivariate Characterization of Some Stochastic Order Relations," *Adv. Appl. Prob.*, Vol. 23, 642-659, 1991.

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

[33] P.P. Talwar, "A Note on Sequencing Problems with Uncertain Job Times," *J. Opns. Res. Soc. Jpn.*, Vol. 9, pp. 93-97, 1967.

[34] R.R. Weber, "Scheduling Jobs with Stochastic Processing Requirements on Parallel Machines to Minimize Makespan or Flowtime," *J. Appl. Prob.*, Vol. 19, pp. 167-182, 1982.

[35] R.R. Weber, P. Varaiya and J. Walrand, "Scheduling Jobs with Stochastically Ordered Processing Times On Parallel Machines to Minimize Expected Flowtime," *J. Appl. Prob.*, Vol. 23, pp. 841-847, 1987.