TECHNICAL RESEARCH REPORT

Simple Optimization Problems via Majorization Ordering

by Y.B. Kim, A.M. Makowski

T.R. 96-5

INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Simple Optimization Problems via Majorization Ordering

YOUNG B. KIM — dosa@src.umd.edu

ARMAND M. MAKOWSKI[†] — armand@src.umd.edu

Abstract

We introduce and explicitly solve a novel class of optimization problems which are motivated by load assignment issues in crossbar switches with output queueing. The optimization criterion is given in the majorization ordering sense. The solution to these problems indirectly provide solutions to a large class of convex optimization problems under a linear constraint.

1 Introduction

The notion of majorization (and its derivatives) provides a powerful tool to formalize statements concerning the relative size of components of two vectors, viz., the components $(x_1, \ldots, x_K)$ of the vector $\boldsymbol{x}$ are "less spread out" than the components $(y_1, \ldots, y_K)$ of the vector $\boldsymbol{y}$. As elegantly demonstrated in the monograph of Marshall and Olkin [7], these notions, in both deterministic and stochastic forms, have found widespread use in many diverse fields of mathematics and their applications.

Recently, several authors have made use of majorization ideas to identify optimal scheduling and load balancing strategies for various resource allocation problems [2, 3]. In this paper, we consider a novel class of optimization problems which are motivated by load assignment issues in crossbar switches with output queueing; this application is discussed in some detail in Sections 3 and 4. These optimization problems which we now define, are in the majorization sense, and have the following

[†]Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742. The work of these authors was supported partially through NSF Engineering Research Centers Program Grant NSFD CDR-88-03012 and through NASA Grant NAGW277S.

generic form: For every vector $\boldsymbol{p}$ in $\mathbb{R}_+^K$ and every constant $c > 0$, we define the subset $\mathcal{A}(\boldsymbol{p}, c)$ of $[0,1]^K$ by

$$\mathcal{A}(\boldsymbol{p}; c) \equiv \{\boldsymbol{\lambda} \in [0,1]^K : \sum_{k=1}^{K} p_k \lambda_k = c\}.$$

This set $\mathcal{A}(\boldsymbol{p}; c)$ is non-empty whenever c satisfies the condition

$$0 < c < \sum_{k=1}^{K} p_k. \tag{1.1}$$

With $\prec$ denoting the majorization ordering (a precise definition is given at the end of this section), we seek vectors $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ in $\mathcal{A}(\boldsymbol{p}; c)$ such that

$$\gamma(\boldsymbol{\lambda}^-, \boldsymbol{p}) \prec \gamma(\boldsymbol{\lambda}, \boldsymbol{p}) \prec \gamma(\boldsymbol{\lambda}^+, \boldsymbol{p}), \quad \boldsymbol{\lambda} \in \mathcal{A}(\boldsymbol{p}; c) \tag{1.2}$$

where we have used the notation

$$\gamma(\boldsymbol{\lambda}, \boldsymbol{p}) \equiv (\lambda_1 p_1, \ldots, \lambda_K p_K), \quad \boldsymbol{\lambda} \in [0,1]^K, \ \boldsymbol{p} \in \mathbb{R}_+^K.$$

The main results of this paper are contained in Theorems 2.1 and 2.2 of Section 2 where we provide explicit expressions for the optimizers $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ in terms of c and $\boldsymbol{p}$. We follow up in Sections 3 and 4 with an outline of how these optimization results can be used to determine the best and worst loading vectors for a non-blocking crossbar switch under the output queueing strategy.

In addition to providing answers to some natural questions raised by the results of [5], Theorems 2.1 and 2.2 have the following curious byproduct which is of independent interest, especially to readers interested in convex optimization: To simplify matters somewhat, we take $\boldsymbol{p} = (1, \ldots, 1) \equiv \boldsymbol{e}$ so that now $\gamma(\boldsymbol{\lambda}, \boldsymbol{p}) = \boldsymbol{\lambda}$ for every $\boldsymbol{\lambda}$ in $[0,1]^K$. The relation $\prec$ being a partial ordering on $\mathbb{R}^K$, it is natural to seek the mappings $\varphi : \mathbb{R}^K \to \mathbb{R}$ which are monotonic for the majorization ordering $\prec$, i.e., mappings with the property that $\varphi(\boldsymbol{x}) \le \varphi(\boldsymbol{y})$ whenever $\boldsymbol{x} \prec \boldsymbol{y}$. Such mappings are called Schur–convex mappings in honor of I. Schur who first studied them. The class of Schur–convex functions is very large, and includes convex and symmetric mappings [7, C.2, p. 67], among other things. With this definition, we conclude from (1.2) that

$$\varphi(\boldsymbol{\lambda}^-) \le \varphi(\boldsymbol{\lambda}) \le \varphi(\boldsymbol{\lambda}^+), \quad \boldsymbol{\lambda} \in \mathcal{A}(\boldsymbol{e}; c). \tag{1.3}$$

for *any* Schur–convex function $\varphi : \mathbb{R}^K \to \mathbb{R}$. Therefore, the vectors $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$, whose existence is established in Theorems 2.1 and 2.2 are solutions to the optimization problems

$$\text{Minimize } \varphi \text{ over the set } \mathcal{A}(\boldsymbol{e}; c) \tag{1.4}$$

and

$$\text{Maximize } \varphi \text{ over the set } \mathcal{A}(\boldsymbol{e}; c) \tag{1.5}$$

where $\varphi : \mathbb{R}^K \to \mathbb{R}$ is *any* Schur–convex function. In other words, the *same* vector $\boldsymbol{\lambda}^-$ (resp. $\boldsymbol{\lambda}^+$) can be chosen to simultaneously solve all the problems (1.4) (resp. (1.5)) with Schur–convex functions φ.

As should be apparent from the discussion above, the equivalence of (1.2) and (1.3) suggests that some information could be gleaned from viewing certain convex optimization problems as optimization problems in the majorization sense. In this vein, the reader will easily check that Theorems 2.1 and 2.2 yield answers to the following convex optimization problems which generalize (1.4) and (1.5): For every vector $\boldsymbol{p}$ in $\mathbb{R}_+^K$ with $\min_{k=1,\ldots,K} p_k > 0$ and a constant $c > 0$ such that $0 < c < \sum_{k=1}^K p_k$, consider the problem

$$\text{Minimize (resp. Maximize) } \varphi \text{ over the set } \mathcal{B}(\boldsymbol{p}; c)$$

where $\mathcal{B}(\boldsymbol{p}; c)$ is the subset of $\mathbb{R}^K$ defined by

$$\mathcal{B}(\boldsymbol{p}; c) \equiv \{\boldsymbol{x} \in \mathbb{R}^K : \ 0 \le x_k \le p_k, \ k = 1, \ldots, K; \sum_{k=1}^K x_k = c\}.$$

and $\varphi : \mathbb{R}^K \to \mathbb{R}$ is a Schur–convex function. Again, a *single* solution vector can be selected that will do the job for *all* Schur–convex functions!

A few words on the notation used in this paper: The k^{th} component of any element $\boldsymbol{x}$ in $\mathbb{R}^K$ is denoted either by x^k or by x_k, $k = 1, \ldots, K$, so that $\boldsymbol{x} \equiv (x^1, \ldots, x^K)$ or $(x_1, \ldots, x_K)$. A similar convention is used for $\mathbb{R}^K$–valued random variables (rvs). For any vector $\boldsymbol{x} = (x_1, \ldots, x_K)$ in $\mathbb{R}^K$, let $x_{(1)} \le x_{(2)} \le \cdots \le x_{(K)}$ denote the components of $\boldsymbol{x}$ arranged in increasing order. For vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^K$, we say that $\boldsymbol{x}$ is *majorized* by $\boldsymbol{y}$, and write $\boldsymbol{x} \prec \boldsymbol{y}$, whenever the conditions

$$\sum_{i=1}^k x_{(i)} \ge \sum_{i=1}^k y_{(i)}, \quad k = 1, 2, \ldots, K \tag{1.6}$$

and

$$\sum_{i=1}^K x_i = \sum_{i=1}^K y_i, \tag{1.7}$$

hold. If conditions (1.6) all hold without (1.7), then we say that $\boldsymbol{x}$ is *weakly super-majorized* by $\boldsymbol{y}$, and write $\boldsymbol{x} \prec^w \boldsymbol{y}$.

2 The Main Results

In this section we establish the existence of and expression for the vectors $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$ satisfying (1.2). Throughout the discussion we assume the vector $\boldsymbol{p}$ to be selected so that

$$0 < p_1 \le p_2 \le \cdots \le p_K, \tag{2.1}$$

and we leave it to the reader to check that there is no loss of generality in doing so.

Theorem 2.1 *Assume (1.1) and (2.1) to hold.*

1. *If $c \le p_K$, then the vector $\boldsymbol{\lambda}^+ \equiv (0, \ldots, 0, \frac{c}{p_K})$ is an element of $\mathcal{A}(\boldsymbol{p}; c)$ which satisfies*

$$\gamma(\boldsymbol{\lambda}, \boldsymbol{p}) \prec \gamma(\boldsymbol{\lambda}^+, \boldsymbol{p}), \quad \boldsymbol{\lambda} \in \mathcal{A}(\boldsymbol{p}; c). \tag{2.2}$$

2. *If $p_K < c$, there exists a unique integer $m = 2, \ldots, K$ such that*

$$\sum_{k=m}^{K} p_k \le c < \sum_{k=m-1}^{K} p_k, \tag{2.3}$$

and the vector $\boldsymbol{\lambda}^+$ defined by

$$\boldsymbol{\lambda}^+ \equiv (\underbrace{0, \ldots, 0}_{m-2}, \frac{c - \sum_{k=m}^{K} p_k}{p_{m-1}}, \underbrace{1, \ldots, 1}_{K-m+1}) \tag{2.4}$$

is an element of $\mathcal{A}(\boldsymbol{p}; c)$ which satisfies (2.2);

Before giving a proof of this result, we note that in both cases the vector $\boldsymbol{\lambda}^+$ can be expressed through the formula (2.4) provided we define m as

$$m \equiv \min\{i = 1, \ldots, K : \sum_{k=i}^{K} p_k \le c\} \tag{2.5}$$

with the natural convention that if the set of indices entering the definition (2.5) is empty, then $m = K + 1$ and $\sum_{k=K+1}^{K} p_k = 0$.

Proof. (Claim 1) If $c < p_K$, then the vector $(0, \ldots, 0, \frac{c}{p_K})$ is indeed an element of $\mathcal{A}(\boldsymbol{p}; c)$, and the validity of (2.2) is well known in that case [7].

(Claim 2) Set $Q_i \equiv \sum_{k=i}^{K} p_k$, $i = 1, \ldots, K$, and the statement on m is therefore equivalent to the monotone sequence $c - Q_i$, $i = 1, \ldots, K$, changing sign exactly once. Condition $p_K < c$ and (1.1) together imply $Q_K < c < Q_1$, so that a change of sign must occur. The uniqueness of the integer m satisfying (2.3) follows from the strict monotonicity of $c - Q_i$, $i = 1, \dots, K$, a fact implied by (2.1). We also conclude from (1.1) that $2 \leq m \leq K$, and $\boldsymbol{\lambda}^+$ is well defined. The definition (2.3) of m implies $0 \leq \lambda_{m-1}^+ < 1$, and $\boldsymbol{\lambda}^+$ is therefore an element of $\mathcal{A}(\boldsymbol{p}; c)$. It is also plain that $\lambda_1^+ p_1 \leq \lambda_2^+ p_2 \leq \dots \leq \lambda_K^+ p_K$.

To establish (2.2) for some $\boldsymbol{\lambda}$ in $\mathcal{A}(\boldsymbol{p}; c)$, it suffices to show that for any permutation σ of $\{1, \dots, K\}$, we have the inequalities

$$\sum_{i=k}^{K} \lambda_{\sigma(i)} p_{\sigma(i)} \leq \sum_{i=k}^{K} \lambda_i^+ p_i, \quad k = 1, \dots, K. \tag{2.6}$$

Using (2.4) and the definition of $\mathcal{A}(\boldsymbol{p}; c)$, we see that

$$\sum_{i=k}^{K} \lambda_{\sigma(i)} p_{\sigma(i)} \leq c = \sum_{i=k}^{K} \lambda_i^+ p_i, \quad k = 1, \dots, m-1$$

and (2.6) thus holds for $k = 1, \dots, m-1$. On the other hand, because $\lambda_{\sigma(i)} \leq \lambda_i^+ = 1$, $i = m, \dots, K$, we have from (2.1) that

$$\sum_{i=k}^{K} \lambda_{\sigma(i)} p_{\sigma(i)} \leq \sum_{i=k}^{K} \lambda_{\sigma(i)} p_i \leq \sum_{i=k}^{K} \lambda_i^+ p_i, \quad k = m, \dots, K$$

and (2.6) also holds for $k = m, \dots, K$. ■

The quantities

$$D_0 \equiv K p_1 - c, \quad D_s \equiv (K - s) p_{s+1} - (c - \sum_{i=1}^{s} p_i), \quad s = 1, \dots, K-1$$

will be useful for characterizing the vector $\boldsymbol{\lambda}^-$ which satisfies (1.2).

Theorem 2.2 *Assume (1.1) and (2.1) to hold.*

1. *If $c \leq K p_1$, then the vector $\boldsymbol{\lambda}^-$ given by*

$$\boldsymbol{\lambda}^- \equiv \left(\frac{c}{K p_1}, \dots, \frac{c}{K p_K}\right)$$

is an element in $\mathcal{A}(\boldsymbol{p}; c)$ which satisfies

$$\gamma(\boldsymbol{\lambda}^-, \boldsymbol{p}) \prec \gamma(\boldsymbol{\lambda}, \boldsymbol{p}), \quad \boldsymbol{\lambda} \in \mathcal{A}(\boldsymbol{p}; c). \tag{2.7}$$

2. *If $Kp_1 < c$, then there exists a unique integer $t = 1, 2, \ldots, K-1$ such that*

$$D_{t-1} < 0 \leq D_t \tag{2.8}$$

and the vector $\boldsymbol{\lambda}^-$ defined by

$$\boldsymbol{\lambda}^- \equiv (\underbrace{1, \ldots, 1}_{t}, \frac{c - \sum_{k=1}^{t} p_k}{p_{t+1}(K-t)}, \ldots, \frac{c - \sum_{k=1}^{t} p_k}{p_K(K-t)})$$

is an element of $\mathcal{A}(\boldsymbol{p}; c)$ which satisfies (2.7).

To clarify the proof of this result, and to see why we might expect it in the first place, we make the following change of variable: With $\mathcal{A}(\boldsymbol{p}; c)$ we associate the set $\mathcal{B}(\boldsymbol{p}; c)$ defined by

$$\begin{aligned} \mathcal{B}(\boldsymbol{p}; c) &\equiv \{\gamma(\boldsymbol{\lambda}, \boldsymbol{p}), \quad \boldsymbol{\lambda} \in \mathcal{A}(\boldsymbol{p}; c)\} \\ &\equiv \{\boldsymbol{x} \in \mathbb{R}^K : \ 0 \leq x_k \leq p_k, k = 1, \ldots, K; \sum_{k=1}^{K} x_k = c\}. \end{aligned}$$

Since $p_1 > 0$, the sets $\mathcal{A}(\boldsymbol{p}; c)$ and $\mathcal{B}(\boldsymbol{p}; c)$ are in one–to–one correspondence with each other through the transformations

$$\boldsymbol{x} = \gamma(\boldsymbol{\lambda}, \boldsymbol{p}) \text{ if and only if } \lambda_i = \frac{x_i}{p_i}, \quad i = 1, \ldots, K. \tag{2.9}$$

The original problem of finding $\boldsymbol{\lambda}^-$ in $\mathcal{A}(\boldsymbol{p}; c)$ satisfying (2.7) is clearly equivalent to that of finding an element $\boldsymbol{x}^\star$ in $\mathcal{B}(\boldsymbol{p}; c)$ such that

$$\boldsymbol{x}^\star \prec \boldsymbol{x}, \quad \boldsymbol{x} \in \mathcal{B}(\boldsymbol{p}; c) \tag{2.10}$$

with $\boldsymbol{x}^\star$ and $\boldsymbol{\lambda}^-$ related through (2.9).

We expect the minimizing element $\boldsymbol{x}^\star$ to be as "balanced" as possible given the constraints defining $\mathcal{B}(\boldsymbol{p}; c)$; in fact, in the absence of the component constraints, the minimizing element would be simply given by $\frac{c}{K}(1, \cdots, 1)$. In general this vector will not be the minimizing element since a priori it is possible for (1.1) to hold while $Kp_s < c$ for some $s = 1, \ldots, K-1$. This suggests that in constructing $\boldsymbol{x}^\star$ we should attempt to keep as many components identical as possible, while meeting the constraints on *all* the components. In view of (2.1) this construction would obviously start with $x_i^\star = p_i$ for the smallest indices, and would lead to guessing $\boldsymbol{x}^\star$ in the form

$$\boldsymbol{x}^\star = (p_1, p_2, \ldots, p_s, a, \ldots, a) \tag{2.11}$$

for some integer $s = 1, \ldots, K-1$ (the case $s = K$ is ruled out by the strict inequality in (1.1)) and scalar $a > 0$. Such a choice (2.11) should be a reasonable candidate for the most balanced vector in $\mathcal{B}(\boldsymbol{p}; c)$ provided additional constraints are met: First, given s, we must have $p_s < a$ for otherwise a more balanced vector in $\mathcal{B}(\boldsymbol{p}; c)$ could be constructed (by transfers [7, p. 134]) from $\boldsymbol{x}^\star$ given by (2.11). The fact that $\boldsymbol{x}^\star$ is an element of $\mathcal{B}(\boldsymbol{p}; c)$ further imposes $a \leq p_{s+1}$ and $p_1 + \ldots + p_s + (K-s)a = c$. Hence, a is uniquely determined and the index s must be selected such that

$$p_s < a \leq p_{s+1} \quad \text{with} \quad a \equiv \frac{c - \sum_{k=1}^{s} p_k}{K - s}. \tag{2.12}$$

Note that (2.12) for some $s = 1, \ldots, K-1$ is equivalent to $D_{s-1} < 0 \leq D_s$, thereby giving a clue for the need of condition (2.8). In Theorem 2.3 below we show that the guess (2.11)–(2.12) indeed satisfies (2.10).

Theorem 2.3 *Assume (1.1) and (2.1) to hold.*

1. *If $c \leq Kp_1$, then the vector $\frac{c}{K}(1, \ldots, 1)$ is an element in $\mathcal{B}(\boldsymbol{p}; c)$ which satisfies (2.10);*

2. *If $Kp_1 < c$, then there exists a unique integer $t = 1, 2, \ldots, K-1$ such that (2.8) holds and the vector $\boldsymbol{x}^\star$ defined by*

$$\boldsymbol{x}^\star \equiv \left(p_1, p_2, \ldots, p_t, \frac{c - \sum_{k=1}^{t} p_k}{K - t}, \ldots, \frac{c - \sum_{k=1}^{t} p_k}{K - t}\right) \tag{2.13}$$

is an element of $\mathcal{B}(\boldsymbol{p}; c)$ which satisfies (2.10).

In view of the transformation (2.9) Theorems 2.2 and 2.3 are clearly equivalent.

Proof. We have $D_0 \leq D_1 \leq D_2 \leq \ldots \leq D_{K-1}$ as we note that

$$D_s - D_{s-1} = (K - s)(p_{s+1} - p_s) \geq 0, \quad s = 1, \ldots, K. \tag{2.14}$$

(Claim 1) If $c \leq Kp_1$, then $\frac{c}{K} \leq p_k$, $k = 1, \ldots, K$, and the vector $\frac{c}{K}(1, \ldots, 1)$ is indeed an element of $\mathcal{B}(\boldsymbol{p}; c)$; that it satisfies (2.10) is well known [7, p. 7].

(Claim 2) The condition $Kp_1 < c$ is equivalent to $D_0 < 0$, and (1.1) yields $D_{K-1} > 0$. The existence and uniqueness of an integer t satisfying (2.8) is now immediate from (2.14), and $\boldsymbol{x}^\star$ given by (2.13) is thus well defined. As a consequence of (2.8) this vector is an element of $\mathcal{B}(\boldsymbol{p}; c)$, and its components are in *increasing* order. Thus, in order to establish its minimality within $\mathcal{B}(\boldsymbol{p};c)$, we need only show for any element $\boldsymbol{x}$ of $\mathcal{B}(\boldsymbol{p};c)$ that

$$\sum_{i=1}^{k} x_i^\star \geq \sum_{i=1}^{k} x_{(i)}, \quad k = 1, \ldots, K. \tag{2.15}$$

If $p_i < x_{(i)}$ for some $i = 1, \ldots, K$, then at most $(i-1)$ components of $\boldsymbol{x}$ do not exceed p_i, but this contradicts the fact that at least i components in $\boldsymbol{x}$ lie in the interval $[0, p_i]$. Hence, $x_{(i)} \leq p_i$ for all $i = 1, \ldots, K$, and (2.15) holds for $k = 1, \ldots, t$.

Next, suppose that n is the first index greater than t for which (2.15) fails, i.e.,

$$\sum_{i=1}^{n} x_i^\star < \sum_{i=1}^{n} x_{(i)} \quad \text{and} \quad \sum_{i=1}^{n-1} x_i^\star \geq \sum_{i=1}^{n-1} x_{(i)}. \tag{2.16}$$

From (2.16) we note that

$$\sum_{i=1}^{n-1} x_i^\star + x_n^\star = \sum_{i=1}^{n} x_i^\star < \sum_{i=1}^{n} x_{(i)} \leq \sum_{i=1}^{n-1} x_i^\star + x_{(n)} \tag{2.17}$$

so that $x_n^\star < x_{(n)}$. On the other hand, the first part of (2.16) being equivalent to $c - \sum_{i=1}^{n} x_i^\star > c - \sum_{i=1}^{n} x_{(i)}$, we get from (2.13) that

$$(K-n)\frac{c - \sum_{k=1}^{t} p_k}{K-t} > \sum_{i=n+1}^{K} x_{(i)} \geq (K-n)x_{(n)}. \tag{2.18}$$

The resulting inequality

$$x_{(n)} < \frac{c - \sum_{k=1}^{t} p_k}{K-t} = x_n^\star \tag{2.19}$$

is in clear contradiction with the conclusion $x_n^\star < x_{(n)}$ derived earlier from (2.17), and (2.15) must hold for all $k = t+1, \ldots, K$. ■

3 Non–Blocking Switches with Output Queueing

In this section we present the model used by the authors in [4, 5, 6] to discuss various stochastic comparison results for a class of non–blocking switches with output queueing. With K input ports and L output ports, this model is parameterized by a vector of rates $\boldsymbol{\lambda}$ (in $[0,1]^L$) and by probability vectors $\boldsymbol{r}_k = (r_{k1}, \ldots, r_{kL})$ (in

$\mathcal{S}_L \equiv \{\boldsymbol{r} = (r_1, \ldots, r_L) \in [0,1]^L : \sum_{\ell=1}^L r_\ell = 1\})$, $k = 1, \ldots, K$. We organize these K vectors into the $K \times L$ routing matrix $\mathbf{R}$ given by

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \\ \boldsymbol{r}_K \end{bmatrix} = \begin{bmatrix} r_{11} \ldots r_{1L} \\ \vdots \\ r_{K1} \ldots r_{KL} \end{bmatrix}.$$

With each set of such vectors, we associate $\{0,1\}$–valued rvs $\{A_{t+1}^k(\lambda_k),\ t = 0,1,\ldots\}$ and $\{1,\ldots,L\}$–valued rvs $\{\nu_t^k(\boldsymbol{r}_k),\ t = 0,1,\ldots\}$, $k = 1,\ldots,K$. These rvs are all defined on some probability triple $(\Omega, \mathcal{F}, \mathbf{P})$. During the discussion we make the following assumptions: (i) For each $k = 1,\ldots,K$, the rvs $\{A_{t+1}^k(\lambda_k),\ t = 0,1,\ldots\}$ are *i.i.d.* rvs with

$$\mathbf{P}\left[A_{t+1}^k(\lambda_k) = 1\right] = 1 - \mathbf{P}\left[A_{t+1}^k = 0\right] = \lambda_k$$

for all $t = 0,1,\ldots$; (ii) For each $k = 1,\ldots,K$, the rvs $\{\nu_t^k(\boldsymbol{r}_k),\ t = 0,1,\ldots\}$ are *i.i.d.* rvs with

$$\mathbf{P}\left[\nu_t^k(\boldsymbol{r}_k) = \ell\right] = r_{k\ell}, \quad \ell = 1,\ldots,L$$

for all $t = 0,1,\ldots$; and (iii) The $2K$ collections of rvs $\{A_{t+1}^k(\lambda_k),\ t = 0,1,\ldots\}$ and $\{\nu_t^k(\boldsymbol{r}_k),\ t = 0,1,\ldots\}$, $k = 1,\ldots,K$, are *mutually independent.*

These quantities are given the following interpretation [4, 5]: At the beginning of time slot $[t, t+1)$, new cells arrive into the system, with $A_{t+1}^k(\lambda_k)$ cell arriving at the k^{th} input port, $k = 1,\ldots,K$. The destination of a cell arriving at the k^{th} input port is encoded in the rv $\nu_t^k(\boldsymbol{r}_k)$, and is declared upon arrival. All cells which arrive during a time slot and which are destined for a given output port, are transported across the switch during that single time slot, and put into the corresponding output buffer in *random* order. With the notation

$$\xi_{t+1}^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \equiv \sum_{k=1}^K \mathbf{1}\left[\nu_t^k(\boldsymbol{r}_k) = \ell\right] A_{t+1}^k(\lambda_k),$$

we see that a batch of $\xi_{t+1}^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ cells are destined for the ℓ^{th} output port during time slot $[t, t+1)$.

During any time slot at most one cell can be transmitted, or equivalently, served. Let $Q_t^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ denote the number of cells present at the beginning of time slot $[t, t+1)$ in the ℓ^{th} output buffer, $\ell = 1,\ldots,L$. If we assume the system to be initially empty at time $t = 0$, then the queue size process evolves according to the recursion

$$\begin{aligned} Q_0^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) &= 0; \\ Q_{t+1}^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) &= \left[Q_t^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) - 1\right]^+ + \xi_{t+1}^\ell(\boldsymbol{\lambda}, \boldsymbol{R}), \quad t = 0,1,\ldots. \end{aligned} \tag{3.1}$$

For each $\ell = 1, \ldots, L$ and $n = 1, 2, \ldots$, let $D_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ denote the delay of the n^{th} cell to arrive at the ℓ^{th} output port, i.e., $D_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ represents the time that elapses between the arrival of the n^{th} cell at the ℓ^{th} output port and the end of its transmission. At each of the output queues, we assume that batches are processed in order of arrival, i.e., all cells in the m^{th} batch are served before the cells in the $(m+1)^{rst}$ batch, $m = 1, 2, \ldots$, but the order of service within a given batch is random. As a result, the delay process of the n^{th} cell can be decomposed into two successive stages, and we can write

$$D_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) = W_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) + B_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$$

where the rv $W_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ counts the number of slots required for transmitting all the cells in the batches which have arrived before that containing the n^{th} cell, and the rv $B_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ denotes the number of slots that the n^{th} cell needs to wait before it is served, once the batch to which it belongs starts being served.

The recursions (3.1) are very similar to the Lindley recursion for single server queues, and by arguments similar to those used in that context, we can show the following facts [4, 5, 6]: We define the offered load to the ℓ^{th} output buffer by

$$\rho_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \equiv \sum_{k=1}^{K} \lambda_k r_{k\ell}, \quad \ell = 1, \ldots, L. \tag{3.2}$$

Whenever the conditions $\rho_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) < 1$, $\ell = 1, \ldots, L$, are satisfied simultaneously, there exists an $\mathbb{N}^L$–valued rv $\boldsymbol{Q}(\boldsymbol{\lambda}, \boldsymbol{R}) \equiv (Q^1(\boldsymbol{\lambda}, \boldsymbol{R}), \ldots, Q^L(\boldsymbol{\lambda}, \boldsymbol{R}))$ such that $\boldsymbol{Q}_t(\boldsymbol{\lambda}, \boldsymbol{R}) \equiv (Q_t^1(\boldsymbol{\lambda}, \boldsymbol{R}), \ldots, Q_t^L(\boldsymbol{\lambda}, \boldsymbol{R})) \Longrightarrow_t \boldsymbol{Q}(\boldsymbol{\lambda}, \boldsymbol{R})$ (with $\Longrightarrow_t$ denoting weak convergence [1]). In such circumstances, the system is termed *stable* and $\boldsymbol{Q}(\boldsymbol{\lambda}, \boldsymbol{R})$ is called the steady–state queue size vector or the queue size in statistical equilibrium. If for some $\ell = 1, \ldots, L$, we only have $\rho_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) < 1$, then the one–dimensional convergence $Q_t^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \Longrightarrow_t Q^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ still takes place and the ℓ^{th} output queue is then said to be stable. In that case, we also have $D_n^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \Longrightarrow_n D^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ for some rv $D^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ given by

$$D^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) =_{st} Q^\ell(\boldsymbol{\lambda}, \boldsymbol{R}) + B^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$$

where $B^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ is the forward recurrence time associated with $\xi_1^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ and $Q^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ and $B^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ are independent rvs.

4 Comparison Results and One–Dimensional Bounds

We now present several stochastic comparison results that describe how changes in arrival rates and routing probabilities affect the various performance measures; these results were obtained in the companion paper [5, 6]. To simplify the presentation, for each rate vector $\boldsymbol{\lambda}$ and routing matrix $\boldsymbol{R}$, we write

$$\gamma_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \equiv (\lambda_1 r_{1\ell}, \ldots, \lambda_K r_{K\ell}), \quad \ell = 1, \ldots, L.$$

We focus here only on results concerning performance measures that are associated with a single output destination. Results concerning the delay measures associated with input ports are also available in [4, 5] but are omitted in for the sake of brevity. Throughout the notation $\leq_{icx}$ is used to denote the convex increasing ordering on the collection of distributions [8].

Theorem 4.1 *Assume that for some $\ell = 1, \ldots, L$, the comparison*

$$\gamma_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \prec^w \gamma_\ell(\boldsymbol{\lambda}', \boldsymbol{R}') \tag{4.3}$$

holds. Then we have $Q_t^\ell(\boldsymbol{\lambda}', \boldsymbol{R}') \leq_{icx} Q_t^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ for all $t = 0, 1, \ldots$. If in addition $\rho_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) < 1$, then in statistical equilibrium we have $Q^\ell(\boldsymbol{\lambda}', \boldsymbol{R}') \leq_{icx} Q^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$, $B^\ell(\boldsymbol{\lambda}', \boldsymbol{R}') \leq_{st} B^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$ and $D^\ell(\boldsymbol{\lambda}', \boldsymbol{R}') \leq_{icx} D^\ell(\boldsymbol{\lambda}, \boldsymbol{R})$.

Under (4.3), the stability condition $\rho_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) < 1$ implies $\rho_\ell(\boldsymbol{\lambda}', \boldsymbol{R}') < 1$, so that the ℓ^{th} output queue is stable in both systems and the comparisons have a well–defined meaning. Furthermore, in the comparison of Theorem 4.1, if the total load (3.2) to the ℓ^{th} output queue is constrained to some given value, then condition (4.3) is equivalent to

$$\gamma_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \prec \gamma_\ell(\boldsymbol{\lambda}', \boldsymbol{R}'). \tag{4.4}$$

Theorem 4.1 thus suggests a way to obtain lower and upper bounds on the queue size metrics (among other things) by seeking the "extremizers" in the conditions (4.4) under certain load constraints. As the reader may have already realized, this leads to the generic problems presented in Section 1.

For the remainder of the discussion, we fix some $\ell = 1, \ldots, L$ and consider two situations which are both associated with the ℓ^{th} output queue.

Problem A – For a given arrival vector $\boldsymbol{\lambda}$, we seek the routing matrix $\boldsymbol{R}$ which minimizes (resp. maximizes) the performance measures at the ℓ^{th} output queue subject to the total load (3.2) to the ℓ^{th} output queue being constrained to some given value, say ρ_ℓ. In view of Theorem 4.1 (and remarks following it) it suffices to identify routing matrices $\boldsymbol{R}^-$ and $\boldsymbol{R}^+$ such that

$$\gamma_\ell(\boldsymbol{\lambda}, \boldsymbol{R}^-) \prec \gamma_\ell(\boldsymbol{\lambda}, \boldsymbol{R}) \prec \gamma_\ell(\boldsymbol{\lambda}, \boldsymbol{R}^+) \tag{4.5}$$

amongst the routing matrices $\boldsymbol{R}$ which satisfy the load equation

$$\sum_{k=1}^{K} \lambda_k r_{k\ell} = \rho_\ell. \tag{4.6}$$

Since we are concerned only with the ℓ^{th} output queue, we need only specify the ℓ^{th} column of the routing matrices involved, and the problem thus reduces to finding vectors $\boldsymbol{c}^-$ and $\boldsymbol{c}^+$ in the set $\mathcal{A}(\boldsymbol{\lambda}; \rho_\ell)$ such that

$$\gamma(\boldsymbol{\lambda}, \boldsymbol{c}^-) \prec \gamma(\boldsymbol{\lambda}, \boldsymbol{c}) \prec \gamma(\boldsymbol{\lambda}, \boldsymbol{c}^+), \quad \boldsymbol{c} \in \mathcal{A}(\boldsymbol{\lambda}; \rho_\ell).$$

With this notation, $\boldsymbol{c}^-$, $\boldsymbol{c}$ and $\boldsymbol{c}^+$ represent the ℓ^{th} column of the routing matrices $\boldsymbol{R}^-$, $\boldsymbol{R}$ and $\boldsymbol{R}^+$, respectively, appearing in (4.5). By invoking the results of Section 2 we can now easily characterize $\boldsymbol{c}^-$ and $\boldsymbol{c}^+$, and we do so under the assumptions $0 < \lambda_1 \leq \ldots \leq \lambda_K$ and $0 < \rho_\ell < \sum_{k=1}^{K} \lambda_k$.

By specializing Theorem 2.1, we find

$$\boldsymbol{c}^+ \equiv (\underbrace{0, \ldots, 0}_{m-2}, a^+, \underbrace{1, \ldots, 1}_{K-m+1})$$

with m and a^+ given by

$$m \equiv \min\{i = 2, \ldots, K : \sum_{k=i}^{K} \lambda_k \leq \rho_\ell\} \quad \text{and} \quad a^+ \equiv \frac{\rho_\ell - \sum_{k=m}^{K} \lambda_k}{\lambda_{m-1}}.$$

On the other hand, Theorem 2.2 immediately yields the following: If $\rho_\ell \leq K\lambda_1$, then $\boldsymbol{c}^- = (\frac{\rho_\ell}{K\lambda_1}, \ldots, \frac{\rho_\ell}{K\lambda_K})$, whereas if $K\lambda_1 < \rho_\ell$, then there exists t ($t = 1, \ldots, K-1$) such that

$$\lambda_t < a^- \equiv \frac{\rho_\ell - \sum_{k=1}^{t} \lambda_k}{K - t} \leq \lambda_{t+1} \tag{4.7}$$

and

$$\boldsymbol{c}^- = (\underbrace{1, \ldots, 1}_{t}, \frac{a^-}{\lambda_{t+1}}, \ldots, \frac{a^-}{\lambda_K}). \tag{4.8}$$

In sum, for a given arrival vector $\boldsymbol{\lambda}$, any routing matrix whose ℓ^{th} column is given by $\boldsymbol{c}^+$ (resp. $\boldsymbol{c}^-$) will minimize (resp. maximize) the performance measures at the ℓ^{th} output queue subject to the load constraint (4.6). Here minimization and maximization are understood in the sense of the stochastic ordering $\leq_{icx}$.

Of particular interest is the situation where the input ports are *equiloaded*, i.e., $\boldsymbol{\lambda} \equiv \frac{\lambda}{K}(1,\ldots,1)$ for some $\lambda > 0$. The feasibility constraint now reads $\rho_\ell < \lambda$, and additional simplifications occur: We have

$$m = \lceil K(1 - \frac{\rho_\ell}{\lambda})\rceil + 1 \quad \text{and} \quad a^+ = \frac{K\rho_\ell - (K-m+1)\lambda}{\lambda}$$

while, as would be expected, we find that (4.7)–(4.8) specialize to $\boldsymbol{c}^- = \frac{\rho_\ell}{\lambda}(1,\ldots,1)$.

Problem B – For a given routing matrix $\boldsymbol{R}$, we now seek the arrival vector $\boldsymbol{\lambda}$ which minimizes (resp. maximizes) the performance measures at the ℓ^{th} output queue subject to the load constraint (4.6) at the ℓ^{th} output queue. Again, with $\boldsymbol{c}_\ell$ denoting the ℓ^{th} column of the routing matrix $\boldsymbol{R}$, we need only identify arrival vectors $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$ such that $\gamma_\ell(\boldsymbol{\lambda}^-,\boldsymbol{R}) \prec \gamma_\ell(\boldsymbol{\lambda},\boldsymbol{R}) \prec \gamma_\ell(\boldsymbol{\lambda}^+,\boldsymbol{R})$ for all arrival vectors $\boldsymbol{\lambda}$ satisfying (4.6), and the problem thus reduces to finding vectors $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$ in the set $\mathcal{A}(\boldsymbol{c}_\ell;\rho_\ell)$ such that

$$\gamma(\boldsymbol{\lambda}^-,\boldsymbol{c}_\ell) \prec \gamma(\boldsymbol{\lambda},\boldsymbol{c}_\ell) \prec \gamma(\boldsymbol{\lambda}^+,\boldsymbol{c}_\ell), \quad \boldsymbol{\lambda} \in \mathcal{A}(\boldsymbol{c}_\ell;\rho_\ell). \tag{4.9}$$

In order to characterize $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$, we again invoke the results of Section 2 under the assumptions $0 < r_{1\ell} \leq \ldots \leq r_{K\ell}$ and $0 < \rho_\ell < \sum_{k=1}^K r_{k\ell}$. This time, we have

$$\boldsymbol{\lambda}^+ = (\underbrace{0,\ldots,0}_{m-2}, b^+, \underbrace{1,\ldots,1}_{K-m+1})$$

with m and b^+ given by

$$m \equiv \min\{i = 2,\ldots,K : \sum_{k=i}^K r_{k\ell} \leq \rho_\ell\} \quad \text{and} \quad b^+ \equiv \frac{\rho_\ell - \sum_{k=m}^K r_{k\ell}}{r_{m-1,\ell}}.$$

Theorem 2.2 implies the following: If $\rho_\ell \leq K r_{1\ell}$, then $\boldsymbol{\lambda}^-$ is given by

$$\boldsymbol{\lambda}^- = (\frac{\rho_\ell}{K r_{1\ell}},\ldots,\frac{\rho_\ell}{K r_{K\ell}}).$$

Finally, if $K r_{1\ell} < \rho_\ell$, then there exists $t = 1,\ldots,K-1$ such that

$$r_{t\ell} < b^- \equiv \frac{\rho_\ell - \sum_{k=1}^t r_{k\ell}}{K-t} \leq r_{t+1,\ell}$$

and

$$\boldsymbol{\lambda}^- = (\underbrace{1,\ldots,1}_{t}, \frac{b^-}{r_{t+1,\ell}},\ldots,\frac{b^-}{r_{K\ell}}).$$

Therefore, for a given routing matrix $\boldsymbol{R}$, in the sense of the stochastic ordering $\leq_{icx}$, the arrival vector $\boldsymbol{\lambda}^+$ (resp. $\boldsymbol{\lambda}^-$) minimizes (resp. maximizes) the performance measures at the ℓ^{th} output queue subject to the load constraint (4.6) at that queue. These results are only one–dimensional, and in general are not independent of ℓ, so that the arrival vectors $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$ obtained earlier do *not* simultaneously satisfy (4.9) under (4.6) for *all* $\ell = 1, \ldots, L$.

We now consider the often–studied situation where the addressing scheme is *input independent* in the sense that $\boldsymbol{R}$ has all its row identical with $\boldsymbol{r}_k = \boldsymbol{r}$, $k = 1, \ldots, K$, for some vector $\boldsymbol{r} = (r_1, \ldots, r_L)$ in $\mathcal{S}_L$. In that case the constraint (4.6) takes the form $\sum_{k=1}^K \lambda_k = \frac{\varrho_\ell}{r_\ell} \equiv \lambda$, with $0 < \lambda < K$. Under this constraint, the vectors $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$ given by $\boldsymbol{\lambda}^- = \frac{\lambda}{K}(1, \ldots, 1)$ and

$$\boldsymbol{\lambda}^+ = (\underbrace{0, \ldots, 0}_{m-2}, \lambda - (K - m + 1), \underbrace{1, \ldots, 1}_{K-m+1})$$

where $m \equiv \lceil K - \lambda \rceil + 1$ do satisfy the inequalities (4.5) simultaneously for *all* $\ell = 1, \ldots, L$.

References

[1] P. Billingsley, *Convergence of Probability Measures*, J. Wiley & Sons, New York (NY), 1968.

[2] C.–S. Chang, X.L. Chao and M. Pinedo, "A note on queues with Bernoulli routing," Proceedings of the 29th IEEE Conference on Decision and Control, Honolulu (HI), December 1990, pp. 897–902.

[3] C.–S. Chang, "A new ordering for stochastic majorization: Theory and applications," *Advances in Applied Probability* **24** (1992), pp. 604–634.

[4] Y.–B. Kim, *On the Performance of Crossbar Switches: Stochastic Comparison and Large Size Asymptotics*, Electrical Engineering Department, University of Maryland, College Park (MD). Expected May 1996.

[5] Y.–B. Kim and A.M. Makowski, "Stochastic comparison results for non–blocking switches with output queueing," *Queueing Systems – Theory & Applications*, submitted 1996.

[6] Y.–B. Kim and A.M. Makowski, "Load balancing for non–blocking switches with output queueing," in preparation.

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

[8] S. Ross, *Stochastic Processes*, J. Wiley & Sons, New York (NY), 1984.