UNIVERSITY OF OSLO
Department of Informatics

Matrix majorization

Geir Dahl

Report 263,
ISBN 82-7368-195-5

June 1998

Matrix majorization

Geir Dahl *

June 1998

Abstract

We study the concept *matrix majorization*: for two real matrices $\mathbf{A}$ and $\mathbf{B}$ having m rows we say that $\mathbf{A}$ majorizes $\mathbf{B}$ if there is a row-stochastic matrix $\mathbf{X}$ with $\mathbf{AX} = \mathbf{B}$. A special case is classical notion of vector majorization. Several properties and characterizations of matrix majorization are given. Moreover, interpretations of the concept in mathematical statistics are discussed and some combinatorial questions are studied.
Keywords: Majorization, row-stochastic matrices, convexity.

1 Introduction

Vector majorization is a much studied concept in linear algebra and its applications. If $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ one says that $\mathbf{a}$ *majorizes* $\mathbf{b}$, denoted by $\mathbf{a} \succ \mathbf{b}$, provided that $\sum_{j=1}^{k} a_{[j]} \geq \sum_{j=1}^{k} b_{[j]}$ for $k = 1, \ldots, n-1$ and $\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j$. (Here $a_{[j]}$ denotes the jth largest number among the components of $\mathbf{a}$). Several generalizations of this concept have also been introduced, and one such direction is to define majorization for matrices. In fact, in the basic book on majorization of Marshall and Olkin [8], one finds a chapter devoted to multivariate majorization where different orderings are presented. One such possibility is to say that $\mathbf{A}$ majorizes $\mathbf{B}$ if there is a doubly stochastic matrix $\mathbf{X}$ satisfying $\mathbf{AX} = \mathbf{B}$; we then write $\mathbf{A} \succ_d \mathbf{B}$. This is motivated by the theorem of Hardy-Littlewood and Pólya saying that, for row vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, $\mathbf{a} \succ \mathbf{b}$ if and only if $\mathbf{aX} = \mathbf{b}$ for some doubly stochastic matrix $\mathbf{X}$. This ordering was studied in detail in [1], see also [7]. Further references on multivariate majorization are found in the survey paper [2] and in [3].

Majorization in a very general setting was also studied by Torgersen in connection with his development of the theory of comparison of statistical experiments, see [9], [10]. This theory is motivated by the question: when does one statistical experiment provide more information about the underlying unknown parameter than another experiment does. If one restricts the attention to certain simple experiments, a connection to the classical notion of vector majorization is

*University of Oslo, Department of Informatics, P.O.Box 1080, Blindern, 0316 Oslo, Norway (Email:geird@ifi.uio.no)

obtained (see also [6] for this). Related work on comparison of measure families and multivariate majorization was done in [7]. However, these generalizations have been oriented towards statistical studies and do not seem to be well known in the linear algebra literature. It is a purpose of this paper to introduce and study this generalization of (vector) majorization as applied to the interesting case of matrices with m rows. At this level of generality one may develop old and new results using some tools from convexity and polyhedral theory.

Some of our notation is explained next. $\mathbb{R}^{m,n}$ is the vector space of real $m \times n$ matrices. Let $\mathbf{A} \in \mathbb{R}^{m,n}$. Then the jth column vector of $\mathbf{A}$ is denoted by $\mathbf{a}^j$ and the ith row vector is denoted by $\mathbf{a}_i$. Moreover, $\text{cone}(\mathbf{A})$ denotes the polyhedral cone generated by the column vectors of $\mathbf{A}$, this is the set of all nonnegative linear combinations of $\mathbf{a}^1, \ldots, \mathbf{a}^n$. We let $\text{span}(\mathbf{A})$ denote the subspace spanned by the columns of $\mathbf{A}$. Moreover, $\mathbf{I}_n$ denotes the identity matrix of order n. A column vector (suitably dimensioned) of all ones will be denoted by $\mathbf{e}$ and $\mathbf{0}$ denotes a matrix or vector with all components being zero. K_n is the standard simplex in $\mathbb{R}^n$, i.e., $K_n = \{\mathbf{x} \in \mathbb{R}^n_+ : \sum_{j=1}^n x_j = 1\}$. We let $\mathcal{M}_{m,n}$ denote the set of all row-stochastic $m \times n$ matrices (so each of the m row vectors lie in K_n). If $m = n$ we write $\mathcal{M}_n$ instead of $\mathcal{M}_{n,n}$. The set of all doubly stochastic matrices of order n is denoted by Ω_n. Both $\mathcal{M}_n$ and Ω_n are polytopes in the vector space of appropriately dimensioned matrices. If $S \subseteq \mathbb{R}^n$ the convex hull (resp. conical hull) of S is denoted by $\text{conv}(S)$ (resp. $\text{cone}(S)$).

2 Matrix majorization

Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices with m rows, say $\mathbf{A} \in \mathbb{R}^{m,n}$ and $\mathbf{B} \in \mathbb{R}^{m,p}$. We say that $\mathbf{A}$ *majorizes* $\mathbf{B}$, and write $\mathbf{A} \succ \mathbf{B}$ (or $\mathbf{B} \prec \mathbf{A}$), provided that there exists a matrix $\mathbf{X} \in \mathcal{M}_{n,p}$ such that

$$\mathbf{AX} = \mathbf{B}. \tag{1}$$

Note that the number of columns in the two matrices $\mathbf{A}$ and $\mathbf{B}$ may be different. This is in contrast to the ordering $\succ_d$ mentioned in the introduction (saying that $\mathbf{A} \succ_d \mathbf{B}$ if there is a *doubly* stochastic matrix $\mathbf{X}$ satisfying $\mathbf{AX} = \mathbf{B}$). Since $\mathbf{Xe} = \mathbf{e}$ whenever $\mathbf{X} \in \mathcal{M}_{n,p}$, we see that if $\mathbf{A} \succ \mathbf{B}$ then $\mathbf{Be} = \mathbf{AXe} = \mathbf{Ae}$, so the ith rowsum in $\mathbf{A}$ and $\mathbf{B}$ coincide for $i = 1, \ldots, m$. Corresponding to $\mathbf{A} \in \mathbb{R}^{m,n}$ and $\mathbf{B} \in \mathbb{R}^{m,p}$ we define the *majorization polytope*

$$\mathcal{M}_{n,p}(\mathbf{A} \succ \mathbf{B}) = \{\mathbf{X} \in \mathcal{M}_{n,p} : \mathbf{AX} = \mathbf{B}\}. \tag{2}$$

This set is nonempty iff $\mathbf{A} \succ \mathbf{B}$, and in that case, it is a bounded polyhedron, i.e., a polytope in the vector space $\mathbb{R}^{n,p}$.

An interesting special case of matrix majorization arises when $n = p$ and, say, the first row in both $\mathbf{A}$ and $\mathbf{B}$ is the row vector $\mathbf{e}^T \in \mathbb{R}^n$ of all ones. So assume that

$$\mathbf{A} = \begin{bmatrix} \mathbf{e}^T \\ \mathbf{A}^* \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{e}^T \\ \mathbf{B}^* \end{bmatrix}.$$

Note that a matrix $\mathbf{X}$ in $\mathcal{M}_n$ is doubly stochastic, i.e., it lies in Ω_n, if and only if $\mathbf{e}^T\mathbf{X} = \mathbf{e}^T$. Thus, we obtain that $\mathbf{A} \succ \mathbf{B}$ if and only if there is a doubly stochastic matrix $\mathbf{X}$ with $\mathbf{A}^*\mathbf{X} = \mathbf{B}^*$, i.e., $\mathbf{A}^* \succ_d \mathbf{B}^*$. Thus matrix majorization ($\succ$) generalizes the ordering $\succ_d$ discussed above. If we here further specialize to the case where $\mathbf{A}^*$ consists of a single row vector $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{B}^*$ the row vector $\mathbf{b} \in \mathbb{R}^n$, the matrix majorization corresponds to $\mathbf{a} \succ \mathbf{b}$, i.e., that $\mathbf{a}$ majorizes $\mathbf{b}$. Thus, the classical concept of vector majorization is a (very) special case of our notion of matrix majorization. For vector majorization, say $\mathbf{a} \succ \mathbf{b}$, the majorization polytope $\Omega(\mathbf{a} \succ \mathbf{b})$ consisting of all $\mathbf{S} \in \Omega_n$ satisfying $\mathbf{x} = \mathbf{S}\mathbf{y}$ was studied in [4]. Several combinatorial properties of this polytope were established. A related study for majorization polytopes in connection with the ordering $\succ_d$ is found in [5].

If $\mathbf{A} \in \mathbb{R}^{m,n}$ and $I \subseteq \{1, \ldots, m\}$ we let $\mathbf{A}[I]$ denote the submatrix of $\mathbf{A}$ induced by the rows indexed by the elements in I. Some basic properties of matrix majorization are collected in the following proposition. The first two properties mean that $\succ$ is a *preorder* on the set of matrices with m rows; it is a reflexive, transitive relation, but it is not antisymmetric (and therefore not a partial order).

Theorem 2.1 *For each $\mathbf{A} \in \mathbb{R}^{m,n}, \mathbf{B} \in \mathbb{R}^{m,p}, \mathbf{C} \in \mathbb{R}^{m,q}$ the following statements hold.*

(i) $\mathbf{A} \succ \mathbf{A}$.

(ii) *If $\mathbf{A} \succ \mathbf{B}$ and $\mathbf{B} \succ \mathbf{C}$, then $\mathbf{A} \succ \mathbf{C}$.*

(iii) *If $\mathbf{A} \succ \mathbf{B}$, then $\mathbf{A}[I] \succ \mathbf{B}[I]$ for each $I \subseteq \{1, \ldots, m\}$.*

(iv) *If $\mathbf{A} \succ \mathbf{B}$ and $\mathbf{H} \in \mathbb{R}^{m,m}$, then $\mathbf{HA} \succ \mathbf{HB}$.*

(v) *If $\mathbf{A} \succ \mathbf{B}$ and $\mathbf{P} \in \mathbb{R}^{n,n}$ and $\mathbf{Q} \in \mathbb{R}^{p,p}$ are two permutation matrices, then $\mathbf{AP} \succ \mathbf{BQ}$.*

(vi) *If $\mathbf{A} \succ \mathbf{B}$, then* $\text{cone}(\mathbf{A}) \supseteq \text{cone}(\mathbf{B})$ *and* $\text{rank}(\mathbf{A}) \geq \text{rank}(\mathbf{B})$.

(vii) *Assume that $m = n$ and that $\mathbf{A}$ is nonsingular. Then $\mathbf{A} \succ \mathbf{B}$ if and only if* $\text{span}(\mathbf{A}) \supseteq \text{span}(\mathbf{B})$, *and in that case the majorization polytope is given by $\mathcal{M}_{n,p}(\mathbf{A}, \mathbf{B}) = \{\mathbf{A}^{-1}\mathbf{B}\}$.*

(viii) *For each $k \geq 1$ the set $\{\mathbf{Z} \in \mathbb{R}^{m,k} : \mathbf{Z} \prec \mathbf{A}\}$ is a polyhedron, in particular, it is convex.*

(ix) *Let $\mathbf{A}'$ be a matrix obtained by augmenting $\mathbf{A}$ with some columns of all zeros. Then $\mathbf{A} \succ \mathbf{A}' \succ \mathbf{A}$.*

Proof. All these results are obtained using elementary arguments, and therefore omitted. A fact that is used is that the set of row-stochastic matrices is closed under matrix products. □

Note that there are examples where $\mathbf{A} \not\succ \mathbf{B}$, but $\mathbf{PA} \succ \mathbf{B}$ for a suitable permutation matrix $\mathbf{P}$. Thus, reordering of rows may influence majorization.

Remark: Due to property (ix) of Theorem 2.1 one may assume that $\mathbf{A}$ and $\mathbf{B}$ have the same number of columns (by augmenting the smaller matrix with columns of all zeros). However, we prefer to allow a different number of columns in our subsequent discussions.

How can we interpret the property that $\mathbf{A}$ majorizes $\mathbf{B}$? Whenever $\mathbf{A}$ is nonsingular, this simply means that the column space of $\mathbf{A}$ contains that of $\mathbf{B}$ (see (vii) in Proposition 2.1). More generally, $\mathbf{A} \succ \mathbf{B}$ reflects that "the columns of $\mathbf{A}$ are more spread out than the columns of $\mathbf{B}$". Thus, if $\mathbf{A} \succ \mathbf{B}$ we know by property (vi) of Proposition 2.1 that $\text{cone}(\mathbf{A}) \supseteq \text{cone}(\mathbf{B})$. Conversely, if $\text{cone}(\mathbf{A}) \supseteq \text{cone}(\mathbf{B})$ there are nonnegative numbers $x_{i,j}$ for $i \le n$ and $j \le p$ such that $\mathbf{b}_j = \sum_{i=1}^n x_{i,j}\mathbf{a}_i$ for each $j \le p$, i.e., $\mathbf{AX} = \mathbf{B}$ where $\mathbf{X} := [x_{i,j}]$. Here the each number $x_{i,j}$ may be considered as a weight associated with the vector $\mathbf{a}_i$, one for each column $\mathbf{b}_j$ of $\mathbf{B}$. Thus, $\mathbf{A} \succ \mathbf{B}$ occurs precisely when the sum of all these weights associated with $\mathbf{a}_i$ add up to 1, for each $i \le n$. A main goal in this paper is to clarify the notion of matrix majorization, and in section 3 we give a number of equivalent conditions for $\mathbf{A}$ to majorize $\mathbf{B}$.

A *disjoint-row-support matrix* is a matrix such that the supports of its rows are pairwise disjoint. (The support of a vector in $\mathbb{R}^n$ is the set of indices of nonzero components).

Proposition 2.2 *Let $\mathbf{A} \in \mathcal{M}_{m,n}, \mathbf{D} \in \mathcal{M}_{m,p}, \mathbf{E} \in \mathcal{M}_{m,q}$ be row-stochastic matrices such that (i) $\mathbf{D}$ is a disjoint-row-support matrix and (ii) all the row vectors of $\mathbf{E}$ are equal. Then we have*

$$\mathbf{D} \succ \mathbf{A} \succ \mathbf{E}.$$

Proof. For $i \le m$ let $S_i = \{j \le p : d_{i,j} > 0\}$ be the support of the ith row in $\mathbf{D}$, and let $S_0 = \{1, \ldots, p\} \setminus \cup_{i \le m} S_i$. Thus $S_0, \ldots, S_m$ is a partition of $\{1, \ldots, p\}$. Define the matrix $\mathbf{X} = [x_{i,j}] \in \mathbb{R}^{p,n}$ as follows. Let $i \le p$. If $i \in S_k$ for some $k \ge 1$ let $x_{i,j} = a_{k,j}$ for every j. If $i \in S_0$ let $x_{i,j} = 1$ if $j = 1$ and $x_{i,j} = 0$ if $j > 1$. $\mathbf{X}$ is nonnegative as $\mathbf{A}$ is. Clearly $\sum_j x_{i,j} = 1$ when $i \in S_0$ and when $i \notin S_0$ there is a unique $k \ge 1$ with $i \in S_k$ and then $\sum_j x_{i,j} = \sum_j a_{k,j} = 1$. Thus, $\mathbf{X}$ is a row-stochastic matrix. Furthermore, for each $k \le m$ and $j \le n$, $(\mathbf{DX})_{k,j} = \sum_{i \in S_k} d_{k,i}x_{i,j} = \sum_{i \in S_k} d_{k,i}a_{k,j} = a_{k,j}\sum_{i \in S_k} d_{k,i} = a_{k,j}$, so we conclude that $\mathbf{DX} = \mathbf{A}$ and $\mathbf{D} \succ \mathbf{A}$. Let each row of $\mathbf{E}$ be the vector $(w_1, \ldots, w_q) \in K_q$, so $\mathbf{E} = [w_1\mathbf{e}, \ldots, w_q\mathbf{e}]$. Let $\mathbf{X}$ be the row-stochastic $n \times q$ matrix where each row equals $(w_1, \ldots, w_q)$. Then $\mathbf{AX} = \mathbf{A}[w_1\mathbf{e}, \ldots, w_q\mathbf{e}] = [w_1\mathbf{Ae}, \ldots, w_q\mathbf{Ae}] = [w_1\mathbf{e}, \ldots, w_q\mathbf{e}] = \mathbf{E}$, so $\mathbf{A} \succ \mathbf{E}$. □

Thus there are maximal and minimal elements with respect to $\succ$ but these elements are not unique.

Even in the case when $m = 1$ there are some interesting aspects of the majorization $\mathbf{A} \succ \mathbf{B}$. Let $\mathbf{a} \in K_n$ and $\mathbf{b} \in K_p$, so each vector is nonnegative and the sum of its components is one. Define $\mathbf{A} = [\mathbf{a}^T]$ and $\mathbf{B} = [\mathbf{b}^T]$. Then $\mathbf{A} \succ \mathbf{B}$; for if we define $\mathbf{X} = \mathbf{eb}^T$ we see that $\mathbf{X} \ge \mathbf{0}$, $\mathbf{Xe} = \mathbf{eb}^T\mathbf{e} = \mathbf{e}$ and $\mathbf{a}^T\mathbf{X} = \mathbf{a}^T\mathbf{eb}^T = \mathbf{b}^T$. Thus, each row-stochastic matrix with only one row matrix majorizes every other such matrix. The interesting part, however, concerns the majorization polytope for a given majorization. To explain this, assume initially that $a_j > 0$ for each $j \leq n$. Let $\mathbf{D}(\mathbf{a})$ denote the diagonal matrix with $d_{j,j} = a_j$ for $j \leq n$. For $\mathbf{X} \in \mathcal{M}_{n,p}$ we have that $\mathbf{a}^T\mathbf{X} = \mathbf{e}^T\mathbf{D}(\mathbf{a})\mathbf{X} = \mathbf{e}^T\mathbf{Y}$ where we define $\mathbf{Y} = \mathbf{D}(\mathbf{a})\mathbf{X} \in \mathbb{R}^{n,p}$. Then we also calculate $\mathbf{Ye} = \mathbf{D}(\mathbf{a})\mathbf{Xe} = \mathbf{D}(\mathbf{a})\mathbf{e} = \mathbf{a}$. It follows that

$$\mathcal{M}_{n,p}(\mathbf{A} \succ \mathbf{B}) = \{\mathbf{D}(\mathbf{a})^{-1}\mathbf{Y} : \mathbf{Y} \in \mathcal{T}_{n,p}(\mathbf{a}, \mathbf{b})\}$$

where

$$\mathcal{T}_{n,p}(\mathbf{a}, \mathbf{b}) = \{\mathbf{Y} \in \mathbb{R}^{n,p}_{+} : \mathbf{e}^T\mathbf{Y} = \mathbf{b}^T, \mathbf{Ye} = \mathbf{a}\}.$$

The polytope $\mathcal{T}_{n,p}(\mathbf{a}, \mathbf{b})$ is the well-known *transportation polytope* which is widely studied, see e.g. [11]. The 1-scheleton of this polytope is known, and its vertices correspond to spanning trees in the complete bipartite graph $K_{n,n}$. Note that $\mathcal{T}_{n,p}(\mathbf{a}, \mathbf{b})$ is nonempty as $\sum_j a_j = \sum_j b_j$. It follows that, in the case of $m = 1$, the majorization polytope is isomorphic (affinely equivalent) to the transportation polytope; and the correspondence is simply appropriate scaling of the rows of the matrix. A similar correspondence exists in the general case where one allows components of $\mathbf{a}$ to be zero. In that case the rows in $\mathbf{X} \in \mathcal{M}_{n,p}(\mathbf{A} \succ \mathbf{B})$ corresponding to the zero components are arbitrary vectors in K_p while the $k \times p$ submatrix consisting of the remaining rows is related to "transportation matrices" of size $k \times p$.

For any matrices $\mathbf{A} \in \mathbb{R}^{m,n}$ and $\mathbf{B} \in \mathbb{R}^{m,p}$ we have

$$\mathcal{M}_{n,p}(\mathbf{A} \succ \mathbf{B}) = \cap_{i=1}^{m} \mathcal{M}_{n,p}(\mathbf{a}_i \succ \mathbf{b}_i)$$

so, in general, the majorization polytope $\mathcal{M}_{n,p}(\mathbf{A} \succ \mathbf{B})$ is the intersection of m polytopes each being a "scaling" of a transportation polytope.

3 Characterizations

The purpose of this section is to give a number of equivalent conditions for the majorization $\mathbf{A} \succ \mathbf{B}$. We equip the real vector space $\mathbb{R}^{m,k}$ with the inner product $\langle \mathbf{V}, \mathbf{W} \rangle = \sum_{i,j} v_{i,j} w_{i,j} = \mathrm{Tr}(\mathbf{V}^T\mathbf{W})$ for $m \times k$ matrices $\mathbf{V} = [v_{i,j}]$ and $\mathbf{W} = [w_{i,j}]$. The Euclidean inner product between two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ is also denoted by $\langle \mathbf{v}, \mathbf{w} \rangle$ (and equal to $\sum_{i=1}^{n} v_i w_i$). We define the function $\diamond : \mathbb{R}^{m,k} \times \mathbb{R}^{m,k} \to \mathbb{R}^m$ such that for $\mathbf{V}, \mathbf{W} \in \mathbb{R}^{m,k}$ the ith component of the vector $\mathbf{V} \diamond \mathbf{W}$ is the Euclidean inner product of the ith row of $\mathbf{V}$ and the ith row of $\mathbf{W}$. This function is linear in each of the two arguments. Observe that $\langle \mathbf{V}, \mathbf{W} \rangle = \mathbf{e}^T(\mathbf{V} \diamond \mathbf{W})$. For a matrix $\mathbf{A} \in \mathbb{R}^{m,n}$ and a positive integer k we define the set

$$\mathcal{M}(\mathbf{A}; k) = \{\mathbf{AM} : \mathbf{M} \in \mathcal{M}_{n,k}\}$$

which we call a *Markotope* associated with $\mathbf{A}$ (a matrix in $\mathcal{M}_{n,k}$ is sometimes called a Markov matrix). Thus, by definition of matrix majorization, we have that $\mathcal{M}(\mathbf{A}; k) = \{\mathbf{B} \in \mathbb{R}^{m,k} : \mathbf{B} \prec \mathbf{A}\}$. Markotopes are of interest in getting alternative descriptions of matrix majorization. A *zonotope* is a vector sum of line segments (each being the convex hull of two points). Zonotopes are special polytopes with several interesting properties.

Proposition 3.1 *$\mathcal{M}(\mathbf{A};k)$ is a polytope in $\mathbb{R}^{m,k}$. Each vertex of $\mathcal{M}(\mathbf{A};k)$ may be written*

$$[\ \textstyle\sum_{j\in J_1}\mathbf{a}^j,\ldots,\sum_{j\in J_k}\mathbf{a}^j\ ]\,\mathbf{P} \tag{3}$$

where $\mathbf{P}\in\mathbb{R}^{k,k}$ is a permutation matrix and $J_1,\ldots,J_k$ is a partition of $\{1,\ldots,n\}$ (some of the sets may be empty in which case the vector sum should be understood as the zero vector). Moreover, $\mathcal{M}(\mathbf{A};2)$ is a zonotope in $\mathbb{R}^{m,2}$:

$$\mathcal{M}(\mathbf{A};2)=[\ \mathbf{0},\textstyle\sum_{j=1}^n\mathbf{a}^j\ ]+\displaystyle\sum_{j=1}^n \mathrm{conv}(\{\mathbf{0},[\mathbf{a}^j,-\mathbf{a}^j]\}).$$

Proof. The vertices of the polytope $\mathcal{M}_{n,k}$ are the (0,1)-matrices of dimension $n\times k$ with exactly one nonzero (a one) in each row; let $\mathbf{X}^1,\ldots,\mathbf{X}^N$ denote these vertices. Then $\mathcal{M}(\mathbf{A};k)$ consists of matrices of the form $\mathbf{A}\sum_{j=1}^N\lambda^j\mathbf{X}^j=\sum_{j=1}^N\lambda^j\mathbf{A}\mathbf{X}^j$ for $\lambda^1,\ldots,\lambda^N\geq 0$ and $\sum_j\lambda^j=1$. This gives the description of the vertices of $\mathcal{M}(\mathbf{A};k)$. When $k=2$ each matrix in $\mathcal{M}_{n,2}$ has the form $[\mathbf{x},\mathbf{e}-\mathbf{x}]$ where $\mathbf{x}\in[0,1]^n$ and a calculation completes the proof. □

From this proposition we obtain a result which resembles a result of Rado (see [8]) saying that, for $\mathbf{a}\in\mathbb{R}^n$, the set $\{\mathbf{b}\in\mathbb{R}^n:\mathbf{b}\prec\mathbf{a}\}$ is the convex hull of the vectors obtained by permuting the components of $\mathbf{a}$.

Corollary 3.2 *The Markotope $\mathcal{M}(\mathbf{A};k)=\{\mathbf{B}\in\mathbb{R}^{m,k}:\mathbf{B}\prec\mathbf{A}\}$ is the convex hull of matrices obtained from $\mathbf{A}$ as specified in (3).*

Support functions of compact convex sets play an important role in the study of matrix majorization. For a nonempty compact subset C of the vector space $\mathbb{R}^{m,k}$ its *support function* $\psi_C:\mathbb{R}^{m,k}\to\mathbb{R}$ (with respect to the inner product $\langle\cdot,\cdot\rangle$) is given by

$$\psi_C(\mathbf{A})=\max\{\langle\mathbf{A},\mathbf{C}\rangle:\mathbf{C}\in C\}\quad\text{for }\mathbf{A}\in\mathbb{R}^{m,k}.$$

(The support function is also defined for arbitrary sets, but we only consider compact sets here). It is a useful fact that for two compact convex sets C and D in $\mathbb{R}^{m,k}$ we have that $C\subseteq D$ if and only if $\psi_C\leq\psi_D$. Each support function ψ defined on $\mathbb{R}^m$ (or $\mathbb{R}^{m,n}$) is *sublinear*, i.e., it is positively homogeneous ($\psi(\lambda\mathbf{x})=\lambda\psi(\mathbf{x})$ for each $\mathbf{x},\mathbf{y}\in\mathbb{R}^m$ and real number $\lambda\geq 0$) and convex (or equivalently: positively homogeneous and subadditive, i.e., $\psi(\mathbf{x}+\mathbf{y})\leq\psi(\mathbf{x})+\psi(\mathbf{y})$). If $l_1,\ldots,l_k$ are linear functionals on $\mathbb{R}^m$ the function ψ defined as the pointwise maximum of $l_1,\ldots,l_k$, i.e., $\psi(\mathbf{x})=\max_{t\leq k}l_t(\mathbf{x})$, is sublinear. In fact ψ is then the support function of a polytope P with at most k vertices (if $l_t(\mathbf{x})=\langle\mathbf{v}_t,\mathbf{x}\rangle$ then P may be taken as the convex hull of the points $\mathbf{v}_1,\ldots,\mathbf{v}_k$). We let $\Psi_k(\mathbb{R}^m)$ denote the set of sublinear functionals that may be written as a maximum of at most k linear functionals.

The following theorem summarizes some equivalent conditions for matrix majorization. It follows from a general result of [9] (see also [10]), but we give a new direct proof using techniques from convexity. In section 5 we give statistical interpretations of this result (in the case when $\mathbf{A}$ and $\mathbf{B}$ are row-stochastic).

Below an inequality $\leq$ between vectors means that the inequality holds for each component.

Theorem 3.3 *Let* $\mathbf{A} = [a_{i,j}] \in \mathbb{R}^{m,n}$ *and* $\mathbf{B} = [b_{i,j}] \in \mathbb{R}^{m,p}$. *Then the following five statements are equivalent.*

(i) $\mathbf{A} \succ \mathbf{B}$.

(ii) $\mathcal{M}(\mathbf{A};k) \supseteq \mathcal{M}(\mathbf{B};k)$ *for each positive integer* k.

(iii) *For each* $k \geq 1$ *and* $\mathbf{L} \in \mathbb{R}^{m,k}$ *we have that*

$$\min\{\langle \mathbf{A}', \mathbf{L}\rangle : \mathbf{A}' \in \mathcal{M}(\mathbf{A};k)\} \leq \min\{\langle \mathbf{B}', \mathbf{L}\rangle : \mathbf{B}' \in \mathcal{M}(\mathbf{B};k)\}.$$

(iv) *For each* $k \geq 1$, $\mathbf{L} \in \mathbb{R}^{m,k}$ *and* $\mathbf{N} \in \mathcal{M}_{p,k}$ *there is an* $\mathbf{M} \in \mathcal{M}_{n,k}$ *such that*

$$(\mathbf{AM}) \diamond \mathbf{L} \leq (\mathbf{BN}) \diamond \mathbf{L}.$$

(v) *For each* $k \geq 1$ *and* $\psi \in \Psi_k(\mathbb{R}^m)$ *we have that*

$$\sum_{j=1}^{n} \psi(\mathbf{a}^j) \geq \sum_{j=1}^{p} \psi(\mathbf{b}^j).$$

Proof. Let $k \geq 1$ and assume that $\mathbf{A} \succ \mathbf{B}$, so there is a matrix $\mathbf{X} \in \mathcal{M}_{n,p}$ with $\mathbf{AX} = \mathbf{B}$. For each $\mathbf{N} \in \mathcal{M}_{p,k}$ we get $\mathbf{BN} = \mathbf{AXN}$. Here $\mathbf{XN} \in \mathcal{M}_{n,k}$, so we conclude that (i) implies (ii). Conversely, if (ii) holds we choose $k = p$ and because $\mathbf{I}_p \in \mathcal{M}_p$ there exists an $\mathbf{X} \in \mathcal{M}_{n,p}$ such that $\mathbf{AX} = \mathbf{BI}_p = \mathbf{B}$, i.e., $\mathbf{A} \succ \mathbf{B}$. Thus, (i) and (ii) are equivalent.

By multiplying the inequality in (iii) by -1 we see that (iii) is equivalent to the fact that $\psi_{\mathcal{M}(\mathbf{A};k)} \geq \psi_{\mathcal{M}(\mathbf{B};k)}$. Thus, (ii) and (iii) are equivalent due to the property of the support function mentioned above.

It is easy to see that (i) implies (iv). For assume that (i) holds and let $\mathbf{X} \in \mathcal{M}_{n,p}$ be such that $\mathbf{AX} = \mathbf{B}$. Let $k \geq 1$, $\mathbf{L} \in \mathbb{R}^{m,k}$ and $\mathbf{N} \in \mathcal{M}_{p,k}$. Define $\mathbf{M} = \mathbf{XN}$. Then $\mathbf{M} \in \mathcal{M}_{n,k}$ and $(\mathbf{BN}) \diamond \mathbf{L} = (\mathbf{AXN}) \diamond \mathbf{L} = (\mathbf{AM}) \diamond \mathbf{L}$ so (iv) holds (even with equality).

The fact that (iv) implies (iii) is due to the following observation. If $\mathbf{B}' \in \mathcal{M}(\mathbf{B};k)$, then $\mathbf{B}' = \mathbf{BN}$ for some $\mathbf{N} \in \mathcal{M}_{p,k}$ and therefore $\langle \mathbf{B}', \mathbf{L}\rangle = \langle \mathbf{BN}, \mathbf{L}\rangle = \mathbf{e}^T((\mathbf{BN}) \diamond \mathbf{L})$.

Thus, we now have that conditions (i), (ii), (iii) and (iv) are equivalent. To complete the proof we show that (v) and (iii) are equivalent. We calculate

$$\begin{aligned} &\min\{\langle \mathbf{A}', -\mathbf{L}\rangle : \mathbf{A}' \in \mathcal{M}(\mathbf{A};k)\} = -\max\{\langle \mathbf{A}', \mathbf{L}\rangle : \mathbf{A}' \in \mathcal{M}(\mathbf{A};k)\} = \\ -\; &\max\{\langle \mathbf{AX}, \mathbf{L}\rangle : \mathbf{X} \in \mathcal{M}_{n,k}\} = -\max\{\textstyle\sum_j \sum_t x_{j,t} \sum_i l_{i,t} a_{i,j} : \mathbf{X} \in \mathcal{M}_{n,k}\} = \\ &\textstyle\sum_j \max_{t \leq k} \sum_i l_{i,t} a_{i,j} = \sum_j \psi(\mathbf{a}^j) \end{aligned}$$

where $\psi(\mathbf{x})$ is defined as the (pointwise) maximum of the k linear functionals on $\mathbb{R}^m$ where the tth functional is given by $\mathbf{x} \to \sum_i l_{i,t} x_i$. This proves the equivalence of (iii) and (v) because varying $\mathbf{L}$ (or $-\mathbf{L}$) corresponds to all functionals in $\Psi_k(\mathbb{R}^m)$. □

There are some interesting consequences of the theorem that involve convex functions. Let $\Psi(\mathbb{R}^m)$ denote the set of all positively homogeneous convex functions $\psi : \mathbb{R}^m \to \mathbb{R}$ (i.e., the sublinear functions).

Corollary 3.4 *Let* $\mathbf{A} \in \mathbb{R}^{m,n}$ *and* $\mathbf{B} \in \mathbb{R}^{m,p}$. *Then* $\mathbf{A} \succ \mathbf{B}$ *if and only if*

$$\sum_{j=1}^{n} \psi(\mathbf{a}^j) \geq \sum_{j=1}^{p} \psi(\mathbf{b}^j) \tag{4}$$

holds for every positively homogeneous convex function $\psi : \mathbb{R}^m \to \mathbb{R}$.

Proof. From the fact that $\Psi_k(\mathbb{R}^m) \subset \Psi(\mathbb{R}^m)$ it follows that $\mathbf{A} \succ \mathbf{B}$ whenever (4) holds for all $\psi \in \Psi(\mathbb{R}^m)$ (see Theorem 3.3 (v)). To see the converse, we recall from convex analysis that $\Psi(\mathbb{R}^m)$ coincides with the support functions of compact convex sets. Thus, if $\psi \in \Psi(\mathbb{R}^m)$ then $\psi = \psi_C$ for a compact convex set $C \subset \mathbb{R}^m$. Then there is a sequence of polytopes $\{P^{(k)}\}_{k=1}^{\infty}$ such that $\rho(P^{(k)}, C) \to 0$ as $k \to \infty$. Here $\rho(D, C) = \max\{\sup_{\mathbf{d}\in D} \inf_{\mathbf{c}\in C} \|\mathbf{c} - \mathbf{d}\|, \sup_{\mathbf{c}\in C} \inf_{\mathbf{d}\in D} \|\mathbf{c} - \mathbf{d}\|\}$ is the Hausdorff distance between sets C and D. Assume now that $\mathbf{A} \succ \mathbf{B}$ so, by Theorem 3.3

$$\sum_{j=1}^{n} \psi_{P^{(k)}}(\mathbf{a}^j) \geq \sum_{j=1}^{p} \psi_{P^{(k)}}(\mathbf{b}^j)$$

holds for $k = 1, 2, \ldots$. By letting $k \to \infty$ we obtain the desired inequality (4). We here used the fact that $D \to \psi_D$ is continuous (using Hausdorff distance for sets and supremum norm for functions). ☐

We may also obtain a simple proof the following interesting result of [7].

Corollary 3.5 *([7]) Let* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m,n}$. *Then there is a doubly stochastic matrix* $\mathbf{X}$ *of order* n *satisfying* $\mathbf{AX} = \mathbf{B}$ *if and only if*

$$\sum_{j=1}^{n} \phi(\mathbf{a}^j) \geq \sum_{j=1}^{n} \phi(\mathbf{b}^j)$$

holds for every (continuous) convex function $\phi : \mathbb{R}^m \to \mathbb{R}$.

Proof. As noted in the introduction there is a doubly stochastic $n \times n$-matrix $\mathbf{X}$ satisfying $\mathbf{AX} = \mathbf{B}$ if and only if $\begin{bmatrix} \mathbf{e}^T \\ \mathbf{A} \end{bmatrix} \succ \begin{bmatrix} \mathbf{e}^T \\ \mathbf{B} \end{bmatrix}$. Due to Corollary 3.4 this is further equivalent to that $\sum_{j=1}^{n} \psi(1, \mathbf{a}^j) \geq \sum_{j=1}^{n} \psi(1, \mathbf{b}^j)$ for all $\psi \in \Psi(\mathbb{R}^{m+1})$. But the class of functions $\mathbf{x} \to \psi(1, \mathbf{x})$ where $\psi \in \Psi(\mathbb{R}^{m+1})$ coincides with the set of convex functions on $\mathbb{R}^m$ (and they are all continuous) and we are done. ☐

Comparing the two corollaries we see that the effect of the stronger requirement that $\mathbf{X}$ is *doubly* stochastic in $\mathbf{AX} = \mathbf{B}$ is that the inequality (4) also holds for convex functions that are not positively homogeneous.

These results may be seen as a generalization of the well-known result of Schur (see [8]): for vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ we have that $\mathbf{a} \succ \mathbf{b}$ if and only if $\sum_{j=1}^{n} \phi(a_j) \geq \sum_{j=1}^{n} \phi(b_j)$ for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$. Assume next that $\psi : \mathbb{R}^m \to \mathbb{R}$ is sublinear and extend the definition of ψ to take matrix arguments by letting $\psi(\mathbf{A}) = \sum_{j=1}^{n} \psi(\mathbf{a}^j)$ (and $\psi(\mathbf{B}) = \sum_{j=1}^{p} \psi(\mathbf{b}^j)$). From Corollary 3.4 we have that

$$\mathbf{A} \succ \mathbf{B} \;\; \Rightarrow \;\; \psi(\mathbf{A}) \geq \psi(\mathbf{B})$$

i.e., that the functional ψ is isotone with respect to the preorder $\succ$. We remark that the implication just stated is easy to prove directly. In fact, if $\mathbf{A} \succ \mathbf{B}$ and ψ is sublinear we obtain $\sum_j \psi(\mathbf{b}^j) = \sum_j \psi(\sum_i \mathbf{a}^i x_{i,j}) \leq \sum_j \sum_i \psi(\mathbf{a}^i x_{i,j}) = \sum_j \sum_i x_{i,j} \psi(\mathbf{a}^i) = \sum_i \psi(\mathbf{a}^i) \sum_j x_{i,j} = \sum_i \psi(\mathbf{a}^i)$. By choosing different functions ψ one obtains different matrix inequalities based on a matrix majorization. As an illustration, consider a matrix $\mathbf{A} \in \mathcal{M}_{m,n}$. The following inequality holds for any vector norm $\|\cdot\|$:

$$\|\mathbf{e}\| \leq \sum_{j=1}^{n} \|\mathbf{a}^j\| \leq \sum_{j=1}^{m} \|\mathbf{e}_j\|$$

where $\mathbf{e}_j$ is the jth unit vector in $\mathbb{R}^m$ and $\mathbf{e}$ is the all ones vector in $\mathbb{R}^m$. In particular, with the Euclidean norm $\|\cdot\|_2$ the inequality becomes $\sqrt{m} \leq \sum_{j=1}^{n} \|\mathbf{a}^j\|_2 \leq m$. The inequality is obtained from the choice $\psi(\mathbf{x}) = \|\mathbf{x}\|$ (which is sublinear) and the majorizations $\mathbf{I}_m \succ \mathbf{A} \succ \mathbf{e}$ due to Proposition 2.2. The bounds are best possible.

For two given vectors $\mathbf{a}$ and $\mathbf{b}$ one easily checks whether $\mathbf{a} \succ \mathbf{b}$ according to its definition in terms of the partial sums of the form $\sum_{j=1}^{k} a_{[j]}$. For matrix majorization things are more complicated, but the problem of deciding if $\mathbf{A} \succ \mathbf{B}$ may be solved as a linear programming problem. Let $\mathbf{A} \in \mathbb{R}^{m,n}$ and $\mathbf{B} \in \mathbb{R}^{m,p}$ be given. Then the matrix equation $\mathbf{AX} = \mathbf{B}$ may be written as

$$\begin{bmatrix} \mathbf{A} & & \\ & \ddots & \\ & & \mathbf{A} \\ \mathbf{I}_n & \ldots & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{x}^1 \\ \vdots \\ \mathbf{x}^p \end{bmatrix} = \begin{bmatrix} \mathbf{b}^1 \\ \vdots \\ \mathbf{b}^p \\ \mathbf{e} \end{bmatrix} \tag{5}$$

i.e., $\mathbf{A}\mathbf{x}^j = \mathbf{b}^j$ for $j \leq p$ and $\mathbf{x}^1 + \ldots + \mathbf{x}^p = \mathbf{e}$. With suitable definitions of the matrix $\bar{\mathbf{A}} \in \mathbb{R}^{(mp+n),np}$ and the vectors $\bar{\mathbf{x}} \in \mathbb{R}^{np}$, $\bar{\mathbf{b}} \in \mathbb{R}^{mp+n}$, we may write (5) as $\bar{\mathbf{A}}\bar{\mathbf{x}} = \bar{\mathbf{b}}$. We may now apply Farkas' lemma to this new system and thereby obtain another characterization of matrix majorization. For a $n \times n$ matrix $\mathbf{Z} = [z_{i,j}]$ define $\rho(\mathbf{Z}) = \sum_{j=1}^{n} \max_i z_{i,j}$. Note that $\rho(\mathbf{Z}) = \sum_{j=1}^{n} \psi(\mathbf{z}^j)$ where ψ is the sublinear functional on $\mathbb{R}^n$ given by $\psi(\mathbf{z}) = \max_{i \leq n} z_i$.

Theorem 3.6 *Let $\mathbf{A} \in \mathbb{R}^{m,n}$ and $\mathbf{B} \in \mathbb{R}^{m,p}$. Then $\mathbf{A} \succ \mathbf{B}$ if and only if*

$$\rho(\mathbf{Y}^T\mathbf{A}) \geq \langle \mathbf{B}, \mathbf{Y} \rangle \quad \textit{for all } \mathbf{Y} \in \mathbb{R}^{m,p}.$$

Proof. With the notation introduced above, $\mathbf{A} \succ \mathbf{B}$ holds if and only if $\bar{\mathbf{A}}\bar{\mathbf{x}} = \bar{\mathbf{b}}$ has a nonnegative solution $\bar{\mathbf{x}}$. By Farkas' lemma this is true if and only if for all $\mathbf{y}^i \in \mathbb{R}^m$, $i \leq p$ and $\mathbf{z} \in \mathbb{R}^n$ with $(\mathbf{y}^i)^T\mathbf{A} + \mathbf{z}^T \geq \mathbf{0}$ for $i \leq p$ it holds that $\sum_{i=1}^{p}(\mathbf{y}^i)^T\mathbf{b}^i + \mathbf{z}^T\mathbf{e} \geq \mathbf{0}$. It is easy to see that the last statement is true iff $\sum_{i=1}^{p}(\mathbf{y}^i)^T\mathbf{b}^i + \mathbf{z}^T\mathbf{e} \geq \mathbf{0}$ holds for $z_j = \max_{i\leq p}\langle -\mathbf{y}^i, \mathbf{a}^j\rangle$. We may here replace $\mathbf{y}^i$ by $-\mathbf{y}^i$ and thereby obtain the equivalent condition $\rho(\mathbf{Y}^T\mathbf{A}) = \sum_{j=1}^{n}\max_{i\leq p}\langle \mathbf{y}^i, \mathbf{a}^j\rangle \geq \sum_{i=1}^{p}(\mathbf{y}^i)^T\mathbf{b}^i = \langle \mathbf{B}, \mathbf{Y}\rangle$ for all $m \times p$ matrices $\mathbf{Y}$ (with columns $\mathbf{y}^1, \ldots, \mathbf{y}^p$). □

Another way of expressing essentially the same conditions as in Theorem 3.6 is to consider the polyhedral cone

$$C = \{(\mathbf{y}^1, \ldots, \mathbf{y}^p, \mathbf{z}) : (\mathbf{y}^i)^T\mathbf{A} + \mathbf{z}^T \geq \mathbf{0} \text{ for } i \leq p\}.$$

According to general polyhedral theory, C is also a finitely generated cone, i.e., it is the set of nonnegative linear combinations of a finite subset C_0 of C. Here C_0 contains generators of C, i.e., a (direction) vector for each extreme ray of the cone C. It follows that $\mathbf{A} \succ \mathbf{B}$ if and only if $\sum_{i=1}^{n}(\mathbf{y}^i)^T\mathbf{b}^i + \mathbf{z}^T\mathbf{e} \geq \mathbf{0}$ holds for all $(\mathbf{y}^1, \ldots, \mathbf{y}^n, \mathbf{z}) \in C_0$. In the case of a general matrix $\mathbf{A}$ it is difficult to find explicitly the generators of C, but for subclasses of matrices this may be possible. For given matrices $\mathbf{A}$ and $\mathbf{B}$ we can decide (computationally) if $\mathbf{A} \succ \mathbf{B}$ by checking the consistency of the linear system $\bar{\mathbf{A}}\bar{\mathbf{x}} = \bar{\mathbf{b}}$, $\bar{\mathbf{x}} \geq \mathbf{0}$. This may be done in standard ways by any linear programming algorithm.

4 The case with two rows

In this section we study matrix majorization in the special case where the matrices are nonnegative and have two rows. Then $\mathbf{A} \succ \mathbf{B}$ simplifies considerably and a nice geometric description in the plane may be given. We also get descriptions that generalize the partial sum description of vector majorization: $\mathbf{a} \succ \mathbf{b}$ means that $\sum_{j=1}^{k} a_{[j]} \geq \sum_{j=1}^{k} b_{[j]}$ for $k = 1, \ldots, n$, and $\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j$. Throughout this section we consider matrices

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & \ldots & a_{1,n} \\ a_{2,1} & \ldots & a_{2,n} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{1,1} & \ldots & b_{1,p} \\ b_{2,1} & \ldots & b_{2,p} \end{bmatrix} \tag{6}$$

that are nonnegative (all elements are nonnegative). The main reduction is given in the following result.

Theorem 4.1 *Let* $\mathbf{A} \in \mathbb{R}^{2,n}$ *and* $\mathbf{B} \in \mathbb{R}^{2,p}$ *be nonnegative matrices with* $\mathbf{Ae} = \mathbf{Be}$. *Then* $\mathbf{A} \succ \mathbf{B}$ *if and only if*

$$\sum_{j=1}^{n} \psi(\mathbf{a}^j) \geq \sum_{j=1}^{p} \psi(\mathbf{b}^j) \tag{7}$$

holds for all $\psi \in \Psi_2(\mathbb{R}^2)$.

Proof. The necessity of the condition is due to Theorem 3.3 (property (i) implies property (v)). Conversely, assume that (7) holds for all $\psi \in \Psi_2$ (we omit writing $\Psi_\cdot(\mathbb{R}^2)$ in this proof). If $\psi \in \Psi_1$, i.e., it is a linear functional, say $\psi(\mathbf{y}) = \mathbf{c}^T\mathbf{y}$, we get $\sum_j \psi(\mathbf{a}^j) = \mathbf{c}^T \sum_j \mathbf{a}^j = \mathbf{c}^T \sum_j \mathbf{b}^j = \sum_j \psi(\mathbf{b}^j)$ so (7) holds (with equality). Let $\psi \in \Psi_k$ for some $k \geq 3$, so ψ may be written as $\psi(\mathbf{y}) = \max_{t \leq k}\langle \mathbf{v}^t, \mathbf{y}\rangle$ ($\mathbf{y} \in \mathbb{R}^2$) for suitable points $\mathbf{v}^1, \ldots, \mathbf{v}^k \in \mathbb{R}^2$. Equivalently, $\psi = \psi_P$ where ψ_P is the support function of the polytope $P = \text{conv}(\{\mathbf{v}^1, \ldots, \mathbf{v}^k\})$. We shall simplify ψ by constructing a "simple" set Z such that $\psi(\mathbf{y}) = \psi_Z(\mathbf{y})$ for all nonnegative vectors $\mathbf{y}$. To do so, consider the unbounded (and pointed) polyhedron $P' = P - \mathbb{R}^2$ consisting of those vectors $\mathbf{x}$ satisfying $\mathbf{x} \leq \mathbf{y}$ for some $\mathbf{y} \in P$. Then the vertex set of P' is some subset of $\{\mathbf{v}^1, \ldots, \mathbf{v}^k\}$, say for notational simplicity, $\{\mathbf{v}^1, \ldots, \mathbf{v}^s\}$. It is clear that $\psi(\mathbf{y}) = \psi_{P'}(\mathbf{y})$ for all $\mathbf{y} \in \mathbb{R}^2$ with $\mathbf{y} \geq \mathbf{0}$. Consider $I^1 = \{\mathbf{v}^1\}$ and the line segments $I^t = \text{conv}(\{\mathbf{0}, \mathbf{v}^t - \mathbf{v}^{t-1}\})$ for $t = 2, \ldots, s$ and define the vector sum $Z = I^1 + I^2 + \ldots + I^s$ (which is a zonotope). It is straightforward to verify that $\psi(\mathbf{y}) = \psi_{P'}(\mathbf{y}) = \psi_Z(\mathbf{y})$ for all $\mathbf{y} \geq \mathbf{0}$. Moreover, the support function is additive with respect to sets, so $\psi_Z = \sum_{t=1}^s \psi_{I^t}$. Here we note that $\psi_{I^1} \in \Psi_1$ and $\psi_{I^t} \in \Psi_2$ for $t = 2, \ldots, s$ (in fact, $\psi_{I^t}(\mathbf{y}) = \max\{\mathbf{0}, \mathbf{v}^t - \mathbf{v}^{t-1}\}$). Thus, by assumption, the inequality (7) holds for each ψ_{I^t} and since each $\mathbf{a}^j$ and $\mathbf{b}^j$ is nonnegative we obtain $\sum_j \psi(\mathbf{a}^j) = \sum_j \psi_Z(\mathbf{a}^j) = \sum_j \sum_t \psi_{I^t}(\mathbf{a}^j) \geq \sum_j \sum_t \psi_{I^t}(\mathbf{b}^j) = \sum_j \psi_Z(\mathbf{b}^j) = \sum_j \psi(\mathbf{b}^j)$. This proves that (7) holds for all $\psi \in \Psi_k$, $k \geq 1$, so $\mathbf{A} \succ \mathbf{B}$ (due to Theorem 3.3) and the proof is complete. □

We see from the proof that a sufficient condition for $\mathbf{A} \succ \mathbf{B}$ is that (7) holds for all sublinear functionals of the form $\max\{\mathbf{0}, \mathbf{z}\}$ where $\mathbf{z} \in \mathbb{R}^2$.

Let again $\mathbf{A} \in \mathbb{R}^{2,n}$ and $\mathbf{B} \in \mathbb{R}^{2,p}$ be nonnegative matrices with $\sum_j \mathbf{a}^j = \sum_j \mathbf{b}^j$. In order to present a geometric characterization of matrix majorization we introduce the set

$$Z_{\mathbf{A}} = \sum_{j=1}^{n} \text{conv}(\{\mathbf{0}, \mathbf{a}^j\}) \tag{8}$$

associated with $\mathbf{A}$. The set $Z_{\mathbf{B}}$ is defined similarly. These sets are zonotopes (vector sum of line segments). Moreover, $Z_{\mathbf{A}} \subset \mathbb{R}^2$ and it is symmetric around the point $(1/2)\sum_{j=1}^n \mathbf{a}^j$. In Fig. 1 we see two zonotopes $Z_{\mathbf{A}}$ and $Z_{\mathbf{B}}$ (shaded) where $n = p = 3$. As $\mathbf{Ae} = \mathbf{Be}$, the zonotopes $Z_{\mathbf{A}}$ and $Z_{\mathbf{B}}$ have the same point of symmetry. Let $\alpha_i := \sum_{j=1}^n a_{i,j}$ for $i = 1, 2$. The "upper boundary" of $Z_{\mathbf{A}}$ may be seen as the graph of a function $\beta_{\mathbf{A}} : [0, \alpha_1] \to \mathbb{R}$ given by

$$\begin{aligned}&\beta_{\mathbf{A}}(\alpha) = \max\{y : (\alpha, y) \in Z_{\mathbf{A}}\} = \\ &\max\{\textstyle\sum_{j=1}^n a_{2,j}\delta_j : \sum_{j=1}^n a_{1,j}\delta_j \leq \alpha, \ \ 0 \leq \delta_j \leq 1 \ \text{ for } j = 1, \ldots, n\}\end{aligned}$$

for $0 \leq \alpha \leq \alpha_1$. The function $\beta_{\mathbf{A}}$ is piecewise linear, concave, nondecreasing and continuous. Moreover, $\beta_{\mathbf{A}}(0) = \sum_{j:a_{1,j}=0} a_{2,j} \geq 0$ and $\beta_{\mathbf{A}}(\alpha_1) = \alpha_2$.

Corollary 4.2 *The following conditions are equivalent for nonnegative matrices $\mathbf{A} \in \mathbb{R}^{2,n}$ and $\mathbf{B} \in \mathbb{R}^{2,p}$ with $\mathbf{Ae} = \mathbf{Be}$:*

(i) $\mathbf{A} \succ \mathbf{B}$.
(ii) $\mathcal{M}(\mathbf{A}; 2) \supseteq \mathcal{M}(\mathbf{B}; 2)$.
(iii) $Z_{\mathbf{A}} \supseteq Z_{\mathbf{B}}$.
(iv) $\beta_{\mathbf{A}} \geq \beta_{\mathbf{B}}$.

Proof. From the proof of Theorem 3.3 we see that conditions (ii) and (v) of Theorem 3.3 are equivalent *for every fixed* $k \geq 1$. Combining this with Theorem 4.1 the equivalence of (i) and (ii) above follows.

Let $\mathbf{c} = \sum_j \mathbf{a}^j = \sum_j \mathbf{b}^j$. Then the Markotope $\mathcal{M}(\mathbf{A}; 2)$ consists of the $n \times 2$-dimensional matrices of the form $\left[\ \sum_j \delta_j \mathbf{a}^j, \mathbf{c} - \sum_j \delta_j \mathbf{a}^j\ \right]$ where $0 \leq \delta_j \leq 1$ for $j = 1, \ldots, n$. But such vectors $\sum_j \delta_j \mathbf{a}^j$ are precisely the elements in the zonotope $Z_{\mathbf{A}}$ so the equivalence of (ii) and (iii) is clear. Moreover, (iv) is just a rewriting of (iii) due to the symmetry of the zonotopes. □

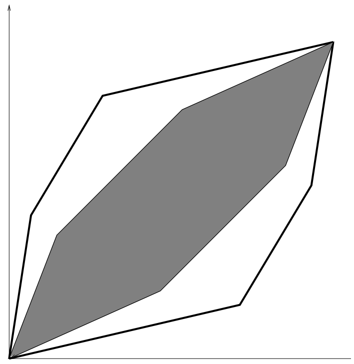

Figure 1: Zonotopes $Z_{\mathbf{A}}$ and $Z_{\mathbf{B}}$ (shaded).

In the example shown in Fig. 1 we have that $Z_{\mathbf{A}} \supseteq Z_{\mathbf{B}}$, so $\mathbf{A} \succ \mathbf{B}$.

Remark: More generally, for matrices $\mathbf{A}$ and $\mathbf{B}$ with m rows, we may also define $Z_{\mathbf{A}}$ and $Z_{\mathbf{B}}$ as in (8). Again, we have that $\mathbf{A} \succ \mathbf{B}$ implies $Z_{\mathbf{A}} \supseteq Z_{\mathbf{B}}$. But the converse implication is false when $m \geq 3$.

Our final goal in this section is to derive a partial sum characterization of matrix majorization (again for two rows). To simplify the presentation we will assume that $a_{1,j}, b_{1,j} > 0$ for all j. We may also assume, possibly after a column permutation applied to $\mathbf{A}$, that $a_{2,1}/a_{1,1} \geq a_{2,2}/a_{1,2} \geq \ldots \geq a_{2,n}/a_{1,n}$, i.e., the slope of the line segment $\text{conv}(\{\mathbf{0}, \mathbf{a}^j\})$ is nonincreasing in j. Then the vertices of the zonotope $Z_{\mathbf{A}}$ are of the form $\sum_{j=1}^k \mathbf{a}^j$ and (by symmetry) $\mathbf{c} - \sum_{j=1}^k \mathbf{a}^j$ for $j = 1, \ldots, n$. (Some of these may not be vertices; this happens when some slopes are equal). Moreover, the function $\beta_{\mathbf{A}}$ satisfies

$$\beta_{\mathbf{A}}(\sum_{j=1}^k a_{1,j}) = \sum_{j=1}^k a_{2,j} \quad \text{for } k = 0, 1, \ldots, n$$

and it is linear on each interval $[\sum_{j=1}^k a_{1,j}, \sum_{j=1}^{k+1} a_{1,j}]$. Similar remarks hold for $\mathbf{B}$ and $\beta_{\mathbf{B}}$. This gives the following partial sum characterization of matrix majorization.

Corollary 4.3 *Let* $\mathbf{A} \in \mathbb{R}^{2,n}$ *and* $\mathbf{B} \in \mathbb{R}^{2,p}$ *be nonnegative matrices with* $\mathbf{Ae} = \mathbf{Be}$ *and* $a_{1,j}, b_{1,j} > 0$ *for all* $j \leq n$. *Then* $\mathbf{A} \succ \mathbf{B}$ *if and only if*

$$\beta_{\mathbf{A}}(\sum_{j=1}^{k} b_{1,j}) \geq \sum_{j=1}^{k} b_{2,j} \quad \text{for } k = 1, \ldots, p-1. \tag{9}$$

A similar result holds without the assumption that elements in the first rows are strictly positive. Consider now the special case where $n = p$ and $a_{1,j} = b_{1,j} = 1$ for all $j \leq n$, so we consider the matrices (dropping some unnecessary indices)

$$\mathbf{A} = \begin{bmatrix} 1 & \ldots & 1 \\ a_1 & \ldots & a_n \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & \ldots & 1 \\ b_1 & \ldots & b_n \end{bmatrix}.$$

As noted in section 2, $\mathbf{A} \succ \mathbf{B}$ is then equivalent to the vector majorization $(a_1, \ldots, a_n) \succ (b_1, \ldots, b_n)$ as row-stochastic matrices with $\mathbf{e}^T\mathbf{X} = \mathbf{e}^T$ are doubly stochastic. We now note that (9) reduces to the well-known partial sum description of vector majorization: $\sum_{j=1}^{k} a_j \geq \sum_{j=1}^{k} b_j$ for $k = 1, \ldots, n-1$. Recall that $\sum_j a_j = \sum_j b_j$ and that the ordering of slopes now correspond to $a_1 \geq \ldots \geq a_n$ and $b_1 \geq \ldots \geq b_n$. Thus we see that matrix majorization in the case of two rows provides a natural generalization of vector majorization.

Finally, we mention that one may construct a lattice from the set of nonnegative matrices $\mathbf{A}$ with two rows and with $\sum_j \mathbf{a}^j$ equal to a fixed vector. The lattice operations $\vee$ and $\wedge$ correspond to operations on the associated zonotopes: $Z_{\mathbf{A}\vee\mathbf{B}} = \text{conv}(Z_{\mathbf{A}} \cup Z_{\mathbf{B}})$ and $Z_{\mathbf{A}\wedge\mathbf{B}} = Z_{\mathbf{A}} \cap Z_{\mathbf{B}}$ (here $\mathbf{A} \vee \mathbf{B}$ resp. $\mathbf{A} \wedge \mathbf{B}$ is "the" smallest upper resp. lower bound of $\mathbf{A}$ and $\mathbf{B}$).

5 Matrix majorization in statistics

Matrix majorization is of interest in mathematical statistics. In fact, there is a theory of comparison of statistical experiments which provides a generalization of the present framework. We briefly explain some of these ideas as applied to the case of discrete experiments. The more general situation involves families of (probability) measures on a general measurable space where Markov kernels play the role of row-stochastic matrices. For a comprehensive treatment of these topics we refer to [10].

Let $\mathbf{A} \in \mathcal{M}_{m,n}$, i.e., it is nonnegative with all rowsums being 1. Then $\mathbf{A}$ represents a *statistical experiment* (or statistical model) where rows correspond to the possible values of an unknown parameter θ (the state of "nature") and the columns correspond to the possible values of a random variable Z that may be observed. More precisely, we interpret $a_{i,j}$ as the probability that $Z = j$ given that the state θ is i. Thus, for each possible state $i \leq m$ we have a probability distribution on the sample space $\{1, \ldots, n\}$ given by the ith row of $\mathbf{A}$. A *statistical decision problem* is to make a decision based on an observation in the experiment so as to minimize (in some sense) a certain loss function. More precisely, assume that the decision is to choose an element in a set $T = \{1, \ldots, k\}$

(called a decision space) and that the "loss" of choosing $t \in T$ when (the unknown state is) $\theta = i$ is given by the real number $L_i(t)$. For instance, if the problem is to estimate θ, so $T = \{1, \ldots, m\}$ (the decision space and the parameter space coincide) we might be interested in the loss function $|t - \theta|$. The statistical decision problem is to choose a best possible decision rule, i.e., a function $\delta : \{1, \ldots, n\} \to K_k$. Here $\delta_j \in K_k$ is the density of a discrete probability measure on $T = \{1, \ldots, k\}$ and it specifies the probability of making the different decisions when $Z = j$ is the outcome of the experiment. Thus the matrix Δ with jth row being the vector δ_j is row-stochastic. In particular, when the matrix Δ is integral, the decision rule is non-randomized and defines a map from the sample space $\{1, \ldots, n\}$ to T. The *risk* $R_A(i, \delta)$ when $\theta = i$ of a decision rule δ (w.r.t. to the loss function L) is defined as the expected loss

$$R_A(i, \delta) = \sum_{j=1}^{n} a_{i,j} \sum_{t=1}^{k} L_i(t) \delta_{j,t}. \tag{10}$$

Note that the risk depends on the experiment $\mathbf{A}$. Risk may be expressed in terms of matrix products as we have $R_A(i, \delta) = \sum_{t=1}^{k} L_i(t) \sum_{j=1}^{n} a_{i,j} \delta_{j,t} = \sum_{t=1}^{k} L_i(t) (\mathbf{A}\Delta)_{i,t} = \langle \mathbf{L}_i, (\mathbf{A}\Delta)_i \rangle$. Here we use $\mathbf{L}_i := (L_i(1), \ldots, L_i(k))$ and $(\mathbf{A}\Delta)_i$ denotes the ith row of the matrix $\mathbf{A}\Delta$. The row-stochastic matrix $\mathbf{A}\Delta$ corresponds to the *performance function* which for each $i \leq m$ specifies the probability of making decision t when the decision rule δ is used.

In the theory of comparison of statistical experiments one studies the notion of one experiment being "more informative" than another. The idea is to have a concept that reflects that an experiment $\mathbf{A}$ gives more information about the unknown parameter θ than another experiment $\mathbf{B}$ does and therefore $\mathbf{A}$ should be preferred for decision-making purposes. In fact, several natural such concepts may be introduced, and a main result in the theory is that these notions are equivalent. This is the content of Theorem 3.3 when it is given a statistical interpretation as discussed next. Let $\mathbf{A}$ and $\mathbf{B}$ be two experiment matrices with the same parameter space (but possibly different sample spaces), say $\mathbf{A} \in \mathcal{M}_{m,n}$ and $\mathbf{B} \in \mathcal{M}_{m,p}$.

Condition (iv) of the theorem says that for every finite decision space $T = \{1, \ldots, k\}$ and every loss function L defined on $\{1, \ldots, m\} \times T$ and for every decision rule δ (confer $\mathbf{N}$) in the experiment $\mathbf{B}$, there is a decision rule μ (confer $\mathbf{M}$) in $\mathbf{A}$ such that $R_A(i, \mu) \leq R_B(i, \delta)$ for all $i \leq m$. Thus, every decision problem may be solved in $\mathbf{A}$ with a risk function which is *uniformly* as good as the risk in $\mathbf{B}$. Condition (iv) involves a pointwise comparison of risk, and, if it holds, one says that the experiment $\mathbf{A}$ is *more informative* than the experiment $\mathbf{B}$.

In order to explain condition (ii) of Theorem 3.3 we note that the Markotope $\mathcal{M}(\mathbf{A}; k) = \{\mathbf{AM} : \mathbf{M} \in \mathcal{M}_{n,k}\}$ may be seen as the set of matrices corresponding to performance functions in the experiment $\mathbf{A}$ when the decision space has k elements. Thus, condition (iv) says that every performance function in $\mathbf{B}$ may also be obtained in $\mathbf{A}$.

The number $\min\{\langle \mathbf{A}', \mathbf{L}\rangle : \mathbf{A}' \in \mathcal{M}(\mathbf{A}; k)\}$ (when divided by m) may be seen as the average risk in $\mathbf{A}$ when L is the loss function. This is also called the minimum Bayes risk w.r.t. the uniform prior distribution on the parameter space. We see that condition (iii) of Theorem 3.3 says that the minimum Bayes risk in $\mathbf{A}$ is no greater than the minimum Bayes risk in $\mathbf{B}$ and this holds for all loss functions (on a finite decision space).

The condition (v) of the theorem says that the maximum Bayes utility (negative loss) in $\mathbf{A}$ is no smaller than the maximum Bayes utility in $\mathbf{B}$. This is a rewriting of the minimum Bayes risk criterion (iii), as explained in the proof of Theorem 3.3.

Finally, we have condition (i) of Theorem 3.3 which is our definition of matrix majorization. It says that there is a row-stochastic matrix $\mathbf{X}$ such that $\mathbf{AX} = \mathbf{B}$, i.e., the experiment $\mathbf{B}$ may be constructed from the experiment $\mathbf{A}$ by a chance mechanism $\mathbf{X}$ which is independent of the unknown parameter θ. Thus, if Z_A and Z_B are random variables associated with the experiments $\mathbf{A}$ and $\mathbf{B}$, respectively, then we may construct another random variable Z from Z_A having the same distribution as Z_B (the distribution of Z for $\theta = i$ is the ith row of $\mathbf{AX}$). Condition (i) is called the randomization criterion for comparison of experiments.

We remark that a proof of the mentioned generalization of Theorem 3.3 for (general) statistical experiments (see [10]) relies on a minmax theorem of two-person, zero-sum games and a weak compactness result in in functional analysis.

The simplifications discussed in section 4 may also be given a statistical interpretation. Experiments corresponding to row-stochastic matrices with two rows are called *dichotomies*. Decision problems where the decision space T has two elements are called *testing problems*. Thus, Theorem 4.1 says that comparison of dichotomies reduces to comparison for testing problems. Moreover, the zonotope $Z_{\mathbf{A}}$ is the set of power functions (the probability of rejecting a null hypothesis, viewed as a function of the parameter θ). The function $\beta_{\mathbf{A}}$ is sometimes called the *Neyman-Pearson function*, and its function value $\beta_{\mathbf{A}}(\alpha)$ is the power of the best test for testing $\theta = 1$ against $\theta = 2$ among all tests with prescribed level α of significance. Thus, Corollary 4.2 says that $\mathbf{A} \succ \mathbf{B}$ if and only if testing problems may be solved at least as good when information is based on $\mathbf{A}$ as when it is based on $\mathbf{B}$.

6 Combinatorial majorization

In this final section we study some combinatorial aspects of matrix majorization. We focus on majorization between certain classes of $(0,1)$-matrices and want to interpret majorization combinatorially.

Let $\mathbf{A} \in \mathcal{M}_{m,n}$ be an *integral* matrix, i.e., it is a $(0,1)$-matrix with a unique 1 in each row, and let $\mathcal{M}^1_{m,n}$ denote the class of such matrices. Then $\mathbf{A}$ may be represented by the following partition $\mathcal{P}(\mathbf{A}) = \{P_1, \ldots, P_n\}$ of the row index set $\{1, \ldots, m\}$: $P_j = \{i \le m : a_{i,j} = 1\}$ for $j \le n$. Some of these sets may be empty, but they are pairwise disjoint. For two partitions $\mathcal{P} = \{P_1, \dots, P_n\}$ and $\mathcal{Q} = \{Q_1, \dots, Q_p\}$ of the set $\{1, \dots, m\}$ we write $\mathcal{P} \leq \mathcal{Q}$ if each set P_i is a subset of some set Q_j, i.e., $\mathcal{P}$ is *finer* than $\mathcal{Q}$.

Proposition 6.1 *Let $\mathbf{A} \in \mathcal{M}^1_{m,n}$ and $\mathbf{B} \in \mathcal{M}^1_{m,p}$. Then $\mathbf{A} \succ \mathbf{B}$ if and only if $\mathcal{P}(\mathbf{A}) \leq \mathcal{P}(\mathbf{B})$. Moreover, if the ith row of $\mathbf{A}$ ($\mathbf{B}$) is $\mathbf{e}_{r_i}$ ($\mathbf{e}_{s_i}$) for $i \leq m$ and $\mathbf{A} \succ \mathbf{B}$, then the majorization polytope $\mathcal{M}_{n,p}(\mathbf{A} \succ \mathbf{B})$ consists of all those row-stochastic $n \times p$ matrices whose r_ith row equals $\mathbf{e}_{s_i}$ for $i \leq m$ while the remaining rows are arbitrary. In particular, $\dim(\mathcal{M}_{n,p}(\mathbf{A} \succ \mathbf{B})) = (p-1)(n - |\{r_i : i \leq m\}|)$.*

Proof. When the ith row of $\mathbf{A}$ ($\mathbf{B}$) is $\mathbf{e}_{r_i}$ ($\mathbf{e}_{s_i}$) for $i \leq m$ we see that $\mathbf{AX} = \mathbf{B}$ translates into $\mathbf{x}_{r_i} = \mathbf{e}_{s_i}$ for $i \leq m$. Thus, $\mathbf{A} \succ \mathbf{B}$ holds if and only if $r_i = r_j$ implies $s_i = s_j$, but this condition means that $\mathcal{P}(\mathbf{A}) \leq \mathcal{P}(\mathbf{B})$. The form of the majorization polytope now follows. □

Let $n = p$. From Proposition 6.1 it follows that $\mathcal{M}^1_{m,n}$ equipped with the majorization ordering $\succ$ is essentially a partially ordered set, in fact a lattice. More precisely, say that $\mathbf{A}$ and $\mathbf{B}$ are *equivalent* if $\mathbf{A} \succ \mathbf{B}$ and $\mathbf{B} \succ \mathbf{A}$ and consider the set Γ consisting of one representative for each equivalence class. Then $(\Gamma, \succ)$ is a lattice. Moreover the ordering is precisely the opposite of the "standard" partial order of partitions on $\{1, \dots, m\}$ (i.e., $\mathcal{P} \leq \mathcal{Q}$).

We turn to $(0,1)$-matrices of dimension $m \times n$ with exactly two ones in each row. (The subsequent discussion also holds for matrices with different number of columns; confer the remark after Theorem 2.1). Each such matrix $\mathbf{A}$ may be viewed as the incidence matrix of an "ordered graph" with m edges and n nodes. More precisely, view the column index set $\{1, \dots, n\}$ as a set of nodes. Then $\mathbf{A}$ corresponds to an ordered family $(f^A_1, \dots, f^A_m)$ of edges where each edge f^A_k is a 2-set; a subset of two nodes. Similarly, if the $(0,1)$-matrix $\mathbf{B} \in \mathbb{R}^{m,n}$ has exactly two ones in each row, we have the associated edge family $(f^B_1, \dots, f^B_m)$. For each $k \leq m$ we define the *square* $S_k = \{(i,j) : i \in f^A_k,\ j \in f^B_k\}$. Note that S_k induces a 2×2 submatrix of a $n \times n$ matrix, namely the submatrix with row index in the 2-set f^A_k and column index in the 2-set f^B_k. We also define the graph $H_{A,B} = (V, E)$ with node set $V = \{(i,j) : i,j \leq n\}$ (a node for each entry in a matrix of order n) and edges obtained by "drawing" the squares in the matrix. More precisely, the nodes (i,j) and (i',j') are adjacent if and only if either (i) $i = i'$ and $f^B_k = [j,j']$ for some $k \leq m$, or (ii) $j = j'$ and $f^A_k = [i,i']$ for some $k \leq m$. An example is shown in Fig. 2. We say that the node (i,j) is in row i and in column j. A node $(i,j) \in V$ is called a *crossnode* if it is the common endnode of two nonparallel edges in E both having the other endnode in row i. The graph $H_{A,B}$ splits into a number of connected components. A component is called *nonvertical* if it contains some crossnode, otherwise it is called *vertical*. For a vertical component all its nodes lie in two columns. We say that two edges are disjoint if they do not have a common endnode.

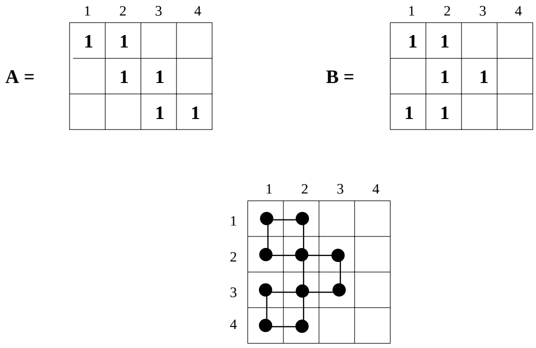

Figure 2: Matrices $\mathbf{A}$ and $\mathbf{B}$ and the graph $H_{A,B}$.

Theorem 6.2 *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m,n}$ *be* $(0,1)$*-matrices with exactly two ones in each row. Then* $\mathbf{A} \succ \mathbf{B}$ *if and only if the following conditions hold:*

(i) *If the edges* f_k^B *and* f_l^B *are disjoint (where* $k, l \le m$*), then* f_k^A *and* f_l^A *are disjoint.*

(ii) *Each nonvertical, connected component of* $H_{A,B}$ *is bipartite and all its crossnodes belong to the same color class.*

Proof. We first rewrite the matrix equation $\mathbf{AX} = \mathbf{B}$. Let, for $k \le m$, $f_k^A = [i_{k,1}^A, i_{k,2}^A]$ and $f_k^B = [i_{k,1}^B, i_{k,2}^B]$ (so the two ones in row k of $\mathbf{A}$ are in columns $i_{k,1}^A$ and $i_{k,2}^A$). The matrix equation $\mathbf{AX} = \mathbf{B}$ is equivalent to the property that, for each $k \le m$, the sum of rows $i_{k,1}^A$ and $i_{k,2}^A$ of $\mathbf{X}$ is equal to the vector $\mathbf{e}_{i_{k,1}^B} + \mathbf{e}_{i_{k,2}^B}$. This means that for each $k \le m$ the 2×2 submatrix of $\mathbf{X}$ corresponding to the square S_k (i.e. the submatrix induced by the rows $i_{k,1}^A$ and $i_{k,2}^A$ and the columns $i_{k,1}^B$ and $i_{k,2}^B$) has the form

$$\begin{bmatrix} \alpha_k & 1-\alpha_k \\ 1-\alpha_k & \alpha_k \end{bmatrix}$$

i.e., it is doubly stochastic, and, moreover, all other elements in rows $i_{k,1}^A$ and $i_{k,2}^A$ of $\mathbf{X}$ are zero.

Assume now that $\mathbf{A} \succ \mathbf{B}$. From the preceeding discussion it is clear that property (i) of the theorem holds (otherwise some rowsum in $\mathbf{X}$ would be two or more). In order to prove property (ii), consider a nonvertical, connected component V^* of $H_{A,B}$. Observe that for every crossnode (i,j) and every solution $\mathbf{X} \in \mathcal{M}_n$ of the system $\mathbf{AX} = \mathbf{B}$ we have that $x_{i,j} = 1$ (as all elements of $\mathbf{X}$ in row i that are outside each square intersecting row i must be zero). The connectivity of the component V^* combined with the fact that all 2×2 submatrices associated with the squares are doubly stochastic imply that every such submatrix of $\mathbf{X}$ is $(0,1)$ with the two ones in opposite diagonal positions.

But then we have a coloring of V^* (with colors 0 and 1) such that no edge joins two nodes of the same color and where all the crossnodes have the same color (namely 1). This proves that property (ii) holds.

To prove the converse, we assume that properties (i) and (ii) both hold. We shall find an $\mathbf{X} \in \mathcal{M}_n$ with $\mathbf{AX} = \mathbf{B}$. Observe that each row i contains nodes from at most one nontrivial component of $H_{A,B}$; this follows from property (i). Thus the row index set $\{1, \ldots, n\}$ may be partitioned into three sets: the rows I_v intersecting a vertical component, the rows I_n intersecting a nonvertical component, and the remaining rows I_r. We define a matrix $\mathbf{X}$ as follows. Consider a nonvertical component of $H_{A,B}$. By property the component is bipartite and its node set has a partition (U, W) where all the crossnodes (of the component) lie in U. We then define $x_{i,j} = 1$ for all nodes $(i, j) \in U$ and $x_{i,j} = 0$ for all $(i, j) \in W$. Also let $x_{i,j} = 1/2$ for each node (i, j) in a vertical component, and for each $i \in I_r$ define $x_{i,1} = 1$. All other entries in $\mathbf{X}$ are zero.

We prove that all rowsums in $\mathbf{X}$ are 1. Due to the observation above rowsums are clearly 1 for rows in $I_v \cup I_r$ so consider an $i \in I_n$. There is a unique nontrivial component of $H_{A,B}$ with nodes in row i, and it is nonvertical. If the component has only two nodes in row i, then these nodes lie in different color classes (as they are adjacent) and exactly one of the corresponding entries in $\mathbf{X}$ is 1. Otherwise, the component has at least three nodes and (from property (i)) the edges of E in row constitute a star, i.e., a set of edges all having, say, node (i, j) as a common endnode (being the center of the star). But then (i, j) is a crossnode and therefore $x_{i,j} = 1$ while each incident node (i, t) belongs to the other color class so $x_{i,t} = 0$. Thus, in both cases the rowsum is 1 and we conclude that $\mathbf{X}$ is row-stochastic.

Finally, we must verify that $\mathbf{AX} = \mathbf{B}$. Consider a square S_k (where $k \leq m$) and let $\mathbf{X}'$ denote the corresponding 2×2-submatrix of $\mathbf{X}$. The subgraph of $H_{A,B}$ induced by the four nodes in S_k is a cycle with four edges. In particular all these nodes are in the same component. If this component is vertical, all entries in the submatrix $\mathbf{X}'$ are $1/2$. If, on the other hand, the component is nonvertical then $\mathbf{X}'$ is a 2×2 permutation matrix. In any case $\mathbf{X}'$ is doubly stochastic and according to the initial discussion above the equation $\mathbf{AX} = \mathbf{B}$ holds. Thus $\mathbf{A} \succ \mathbf{B}$. □

Note that property (i) of the theorem is equivalent to $\mathbf{A}[I] \succ \mathbf{B}[I]$ for each $I \subseteq \{1, \ldots, m\}$ with $|I| = 2$. Another way of stating property (ii) of the theorem is to say that every nonvertical, connected component of $H_{A,B}$ contains no odd cycles and every path between two crossnodes has even length. In the example shown in Fig. 2 the graph $H_{A,B}$ is bipartite, but the crossnodes $(2, 2)$ and $(3, 2)$ are adjacent so condition (ii) of Theorem 6.2 is violated.

We may also in this situation find a complete description of the majorization polytope. Assume that $\mathbf{A} \succ \mathbf{B}$ where both $\mathbf{A}$ and $\mathbf{B}$ are $(0, 1)$-matrices with two ones in each row. Let $\mathbf{X} \in \mathcal{M}_{n,n}(\mathbf{A} \succ \mathbf{B})$. As explained in the proof above the rows of $\mathbf{X}$ may be partitioned into three sets: the rows I_v intersecting a vertical component, the rows I_n intersecting a nonvertical component, and the remaining rows I_r. Now, all the rows in I_n are completely determined as in the proof: for each such component of $H_{A,B}$ the entries of $\mathbf{X}$ are 1 for the nodes in the bipartition class of the crossnodes, and all other entries in these rows are zero. Next, each row in I_r is an arbitrary nonnegative vector with sum 1. Finally, consider the rows in I_v and it suffices to consider one such vertical component. One may check that if this component is bipartite, say with color classes U and W, then $\mathbf{X}$ satisfies $x_{i,j} = \alpha$ for all $(i,j) \in U$ and $x_{i,j} = 1 - \alpha$ for all $(i,j) \in W$ where $0 \leq \alpha \leq 1$. On the other hand, if the component is not bipartite (it has some odd cycle of the form $(i_1, j), (i_2, j), \ldots, (i_t, j), (i_1, j)$ with t odd) then $x_{i,j} = 1/2$ for all nodes (i,j) in that component. From this we also see that the vertices of $\mathcal{M}_{n,n}(\mathbf{A} \succ \mathbf{B})$ are obtained by choosing α to be either 0 or 1 for each bipartite, vertical component. In particular, this means that all the vertices of $\mathcal{M}_{n,n}(\mathbf{A} \succ \mathbf{B})$ are $(0, 1/2, 1)$-integral (entries being 0, 1/2 or 1). Moreover, the majorization polytope is integral (has only integral vertices) if and only if each vertical component of $H_{A,B}$ is bipartite. The dimension of $\mathcal{M}_{n,n}(\mathbf{A} \succ \mathbf{B})$ equals $|I_r|(n-1) + b$ where b denotes the number of bipartite, vertical components of $H_{A,B}$. Finally, $\mathcal{M}_{n,n}(\mathbf{A} \succ \mathbf{B})$ has $2^b n^{|I_r|}$ vertices.

Based on Theorem 6.1 and Theorem 6.2 one may construct simple combinatorial algorithms that determine whether $\mathbf{A} \succ \mathbf{B}$ holds in the two situations ($(0,1)$-matrices with 1 nonzero in each row, or $(0,1)$-matrices with 2 nonzeros in each row).

A natural question is to interpret majorization between $(0,1)$-matrices of dimension $m \times n$ with exactly two ones in each *column* (i.e., node-edge incidence matrices of graphs). In this case, however, the answer is not very interesting: $\mathbf{A} \succ \mathbf{B}$ if and only if $\mathbf{A}$ and $\mathbf{B}$ are equal up to a permutation of the columns.

More generally, let $\mathbf{A}$ and $\mathbf{B}$ be $(0,1)$-matrices of dimension $m \times n$. These matrices may be seen as incidence matrices of set configurations. For $j \leq n$ define the set $A_j = \{i \leq m : a_{i,j} = 1\}$ so the columns of $\mathbf{A}$ are the incidence vectors of $A_1, \ldots, A_n$ as subsets of $I = \{1, \ldots, m\}$. Similarly, the sets $B_1, \ldots, B_n$ correspond to the columns of $\mathbf{B}$. It would be interesting to have a nice characterization of $\mathbf{A} \succ \mathbf{B}$ in terms of relations between the sets A_j and B_j for $j = 1, \ldots, n$. This, however, seems difficult, but some simple observations can be made. Assume that $\mathbf{A} \succ \mathbf{B}$. Then (see section 2) the ith rowsum in $\mathbf{A}$ and the one in $\mathbf{B}$ coincide, which means that each element $i \in I$ lies in the same number of A_j-sets as the number of B_j-sets. Moreover, cone($\mathbf{A}$) $\supseteq$ cone($\mathbf{B}$) which translates into the condition that each set B_j is the union of some of the A_j-sets.

References

[1] P.M. Alberti and A. Uhlmann. *Stochasticity and partial order*, volume 9 of *Mathematics and Its Applications*. D. Reidel Publishing Comp., Berlin, 1982.

[2] T. Ando. Majorization and inequalities in matrix theory. *Linear algebra and its Applications*, 199:17–67, 1994.

[3] J.V. Bondar. Comments and complements to: *Inequalities: Theory of Majorization and Its Applications*, by Albert W. Marshall and Ingram Olkin. *Linear algebra and its Applications*, 199:115–130, 1994.

[4] R. A. Brualdi. The doubly stochastic matrices of a vector majorization. *Linear algebra and its Applications*, 61:141–154, 1984.

[5] Gi-Sang Cheon and You-Ho Lee. The doubly stochastic matrices of a multivariate majorization. *J. Korean Math. Soc.*, 32(4):857–867, 1995.

[6] G. Dahl. Pseudo experiments and majorization. Master's thesis, University of Oslo, Department of Mathematics, 1984.

[7] S. Karlin and Y. Rinott. Comparison of measures, multivariate majorization, and applications to statistics. In S. Karlin et al., editor, *Studies in econometrics, time series, and multivariate statistics. (In commemoration of T.W. Anderson's 65th birthday)*, pages 465–489. Academic Press, 1983.

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

[9] E. Torgersen. Majorization and approximate majorization for families of measures, applications to local comparison of experiments and the theory of majorization of vectors in $\mathbb{R}^n$ (Schur convexity). In Calinski et al., editor, *Linear statistical inference, Proc. Int. Conf., Poznan/Pol. 1984*, volume 35 of *Lecture Notes in Statistics*, pages 259–310. Springer, 1985.

[10] E. Torgersen. *Comparison of Statistical Experiments*, volume 36 of *Encyclopedia of Mathematics and its applications.* Cambridge University Press, Cambridge, 1992.

[11] V.A. Yemelichev, M.M. Kovalev, and M.K. Kravtsov. *Polytopes, graphs and optimization.* Cambridge University Press, Cambridge, 1984.