Majorization for Matrix Classes*

Geir Dahl[1,a], Alexander Guterman[2,3,4,b], Pavel Shteyner[2,3,4,c]

Published in Linear Algebra and Its Applications

Abstract

We introduce a new majorization order for classes (sets) of matrices which generalizes several existing notions of matrix majorization. Roughly, the notion says that every matrix in one class is majorized by some matrix in the other class. The motivation to study this majorization concept comes from mathematical statistics and involves the information content in experiments. This connection is briefly described. We investigate properties of this new order both of algebraic and geometric character. In particular, we establish results on so-called minimal cover classes with respect to the introduced majorization.

Key words. Matrix majorization, partial order, doubly stochastic matrix.

AMS subject classifications. 06A06, 15B51.

*The investigations of the second and the third authors were financially supported by RSF grant 16-11-10075.

[1] Department of Mathematics, University of Oslo, Norway.
[2] Lomonosov Moscow State University, Moscow, 119991, Russia.
[3] Moscow Center for Continuous Mathematical Education, Moscow, 119002, Russia.
[4] Moscow Institute of Physics and Technology, Dolgoprudny, 141701, Russia.
[a] geird@math.uio.no
[b] guterman@list.ru
[c] pashteiner@ya.ru

1 Introduction

The notion of majorization has been studied a lot, in connection with vectors (the classical notion), matrices, and even more general structures, such as measure families. The purpose of this paper is to introduce a new majorization concept for matrices, actually for classes of matrices, which extends several central notions of matrix majorization.

The motivation comes from mathematical statistics, an area called the comparison of statistical experiments (for the detailed and self-contained information on this topic we recommend the monograph [13]). It deals with the information content of experiments.

Notation: Let $M_{m,n}$ be the vector space of all real $m \times n$ matrices (where we write M_n, if $m = n$). The transpose of a matrix A is denoted by A^t. For a matrix A its j'th column is denoted by $A^{(j)}$, and the column sum vector of A is denoted by $c(A)$. The vector of a suitable length such that all its entries are 1 is denoted by e. The vector with all zero entries except 1 on the jth position is referred as the jth unit vector and is denoted by e_j. The convex hull of a set $S \subseteq \mathbb{R}^n$ is denoted by $\text{conv}(S)$. We say that a vector $x \in \mathbb{R}^n$ is *monotone* if $x_1 \geq x_2 \geq \cdots \geq x_n$. For a vector $y \in \mathbb{R}^n$ we let $y_{[j]}$ denote the jth largest number among the components of y. If $a, b \in \mathbb{R}^n$ we say that a is *majorized* by b, denoted by $a \preceq b$ (or $b \succeq a$), provided that $\sum_{j=1}^k a_{[j]} \leq \sum_{j=1}^k b_{[j]}$ for $k = 1, 2, \ldots, n$ with the equality in the case $k = n$. In [10] one may find a comprehensive study of majorization and its role in many branches of mathematics and its applications. The recent book [2] treats majorization in connection with several combinatorial classes of matrices. A matrix is called *doubly stochastic* if it is (componentwise) nonnegative and each row and column sum is 1. The set Ω_n of all $n \times n$ doubly stochastic matrices is a polytope whose extreme points are the permutation matrices (the Birkhoff - von Neumann theorem), see [2] for an in-depth survey of properties of Ω_n. A matrix is *row-stochastic* if it is componentwise nonnegative and each row sum is 1. The set of row-stochastic matrices in M_n is denoted by Ω_n^{row}. Further details on these notions can be found in [9], [12] and references therein.

We discuss some notions and facts from this theory in Section 2. We investigate properties of this new order, both of algebraic and geometric flavour. In particular, we show several results on so-called minimal cover classes.

When we discuss matrix classes (i.e., sets of matrices) we will assume that each such class is finite. Some results may be extended to infinite matrix classes.

Our paper is organized as follows: in Section 2 we present some motivation for this study. In Section 3 the majorization matrix class order is introduced and some basic properties are investigated. Section 4 is devoted to the investigations of minimal cover classes.

2 Motivation

The concepts introduced in the present paper are motivated by some basic questions in probability theory and mathematical statistics. We briefly describe this motivation.

The *theory of statistical experiments* is a mathematical treatment of fundamental concepts in mathematical statistics, and it deals with the information content of statistical experiments. The main idea is to compare two statistical experiments with respect to the notion of "risk", as explained further. The most complete treatment of this theory is the monograph by Torgersen [13]. The initial development of this area was around the 1950s, and it was originated in the works by Halmos and Bahadur [7], Blackwell [1] and LeCam [3].

Assume we want to minimize a real-valued function $L_\theta(t)$ over all $t \in T$, where T is a given (abstract) set, and θ is a given parameter in some set Θ. The function L is usually called the *loss function*. A complication is that the parameter θ is unknown. However, we can get information on θ via a statistical experiment $\mathcal{E}$ where we observe a random variable X with values in some set $\mathcal{X}$, whose distribution depends on θ. Formally, the experiment $\mathcal{E}$ is a family $(P_\theta : \theta \in \Theta)$ of probability measures on $\mathcal{X}$ (which is equipped with a σ-algebra specifying measurable sets). Thus, after the experiment is done and X is observed, say $X = x$ (the realization), we choose an element $t \in T$. Therefore, we consider $t = t(x)$ as a function of the observed value x of X, and this makes t a random variable. It makes sense to consider the (mathematical) expectation $R_\theta := \mathbb{E}L_\theta(t) = \int L_\theta(t(x))dP_\theta$ where the expectation is computed based on the probability distribution P_θ, which depends on the unknown θ. The expression R_θ, the expected loss, is called the *risk*. It depends on the unknown parameter θ and the function $x \to t$ that we used, which is called the *decision rule*. More generally, one may also consider a *randomized* decision rule (or *Markov kernel*) which, for each x, specifies a probability measure on the set T. Note that the risk function $\theta \to R_\theta$ not only depends on the decision rule, but also on the experiment $\mathcal{E}$, since this specifies the underlying probablity measures P_θ, so let us write $R_\theta^{\mathcal{E}}$ to indicate this dependence.

Now, consider two experiments $\mathcal{E}$ and $\mathcal{F}$, both with probability measures on the same parameter space Θ. We say that $\mathcal{E}$ is *more informative* than $\mathcal{F}$ if for all *finite* sets T and loss functions $L_\theta : T \to \mathbb{R}$ and for each decision rule in $\mathcal{F}$, there exists a decision rule in $\mathcal{E}$ such that

$$R_\theta^{\mathcal{E}} \leq R_\theta^{\mathcal{F}} \quad \text{for all } \theta \in \Theta.$$

This condition is a pointwise comparison of risks and, if it holds, we write $\mathcal{E} \geq \mathcal{F}$. So, we ordered the set of experiments. Informally, an experiment which is more informative, tells more about the unknown parameter than the other. Central part in the theory of comparison of experiments are theorems that characterize the property $\mathcal{E} \geq \mathcal{F}$ in several ways, some of them expressed in terms of statistical notions (such as Bayes risk). The setup above may be generalized by allowing P_θ to be a measure, or even a signed measure.

Classical majorization for vectors in $\mathbb{R}^n$ is a special case of this order relation for statistical experiments. The same is true for matrix majorization, see [5] where this concept is developed and where a brief introduction to some relevant theory of statistical experiments is presented. These special cases are obtained by letting $\mathcal{X} = \{1, 2, \ldots, n\}$, and $\Theta = \{1, 2, \ldots, m\}$. Then the (generalized) experiment $(P_\theta : \theta \in \Theta)$ corresponds to a matrix $E = [e_{ij}]$ where $e_{ij} = P_i(\{j\})$, so $\theta = i$. The case of classical vector majorization is obtained when $m = 2$ and the first row is the vector of all ones, e. If $\mathcal{E}$ is an experiment, then each P_θ is a probability measure, and E is a row-stochastic matrix. For a generalized experiment (but with $\mathcal{X}$ and Θ discrete as stated above), E is an arbitrary $m \times n$ matrix. For more on this and the majorization special case, see [5, 6].

The next example gives some intuition concerning the concepts introduced above.

Example 2.1. Assume we have two different devices for measuring some substance in the human body, where we use a simple scale of low (L), medium (M) and high (H) occurrence of the substance. Due to device quality, the output may be wrong, and the probabilities for different output based on the true underlying value can be described by a row-stochastic matrix, where both rows and columns correspond to L, M, and H. Assume the two devices are given by the two matrices E and F:

$$E = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad F = \begin{bmatrix} 0.55 & 0.39 & 0.06 \\ 0.2 & 0.6 & 0.2 \\ 0 & 0.3 & 0.7 \end{bmatrix},$$

where the (i, j)th entry provides the probability of the answer j if the true answer is i.

Consider first the device given by E. Assume the realization of a random variable in the experiment E is j. If $j = 1$ (or L), we conclude that the underlying parameter i is 1, and if $j = 3$, the only possibility is $i = 3$. If, however, the realization is $j = 2$, i can be either 1 or 2. Since $j = 2$ has probability 0.3 when $i = 1$ and 1 when $i = 2$, we would say that it is more likely that $i = 2$, although it depends on prior probabilities of $i = 1$ and $i = 2$. Thus, the experiment E gives a lot of information about the underlying parameter i, and the measuring device seems good. For the device corresponding to the matrix F, the situation is different: for each realization j there are at least two possible underlying values of i, and the probabilities do not distinguish these possibilities so well. So, device/experiment E seems better than F. In fact, one can verify that E is more informative than F, using the following method.

Let

$$X = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.6 & 0.2 \\ 0 & 0.3 & 0.7 \end{bmatrix}.$$

Note that X is also row-stochastic. Moreover, one can check that $EX = F$. This implies that E matrix majorizes F, see [5] and the next section. The so-called *randomization criterion* in the theory of statistical experiments says that (under weak assumptions that are fulfilled here) E is more informative that F if and only if there exists a row-stochastic matrix X such that $EX = F$. □

Now we extend the mentioned ordering of pairs of experiments, into comparing *sets* of experiments. Let $\mathcal{A} = \{\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_p\}$ and $\mathcal{B} = \{\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_q\}$ be two sets of experiments, all based on the same parameter space Θ. Suppose a statistician A may choose an experiment in the set $\mathcal{A}$, and another statistician B may choose an experiment in the set $\mathcal{B}$. Which one has available the most information about the unknown θ? We may introduce the following notion to model this situation: We say that $\mathcal{B}$ is *more informative* than $\mathcal{A}$ if for each $\mathcal{E}_j$ $(j \leq p)$, there exists an $i \leq q$ such that $\mathcal{F}_i$ is more informative than $\mathcal{E}_j$. In other words, whatever experiment A chooses in $\mathcal{A}$, B can choose a more informative experiment in $\mathcal{B}$, which means that B has a preferred situation for making decisions. The present paper investigates this new notion in detail.

For a comprehensive treatment of the theory of comparison of statistical experiments, including the proper framework from convexity, functional analysis and game theory, see [14]. Some shorter treatments, more directed toward majorization, may be found in [13] and [4]. In [13] the connection to the important concept of *stochastic order* is discussed.

3 The majorization matrix class order

Let $\preceq$ denote a matrix majorization order. There are several such notions, and we mention some of them next. Let $A, B \in M_{m,n}$.

- Directional majorization: $A \preceq^d B$ when $Ax \preceq Bx$ for all $x \in \mathbb{R}^n$.
- Doubly stochastic majorization: $A \preceq^{ds} B$ when there is $X \in \Omega_n$ such that $A = BX$.
- Matrix majorization: $A \preceq^m B$ when there is a row-stochastic matrix $X \in \Omega_n^{row}$ such that $A = BX$.
- Weak matrix majorization: $A \preceq^{wm} B$ when there is a row-stochastic matrix $X \in \Omega_m^{row}$ such that $A = XB$.
- Strong majorization: $A \preceq^s B$ when there is $X \in \Omega_m$ such that $A = XB$.

Classical vector majorization is the special case of strong majorization where the matrices A and B have a single column. Classical vector majorization is also a special case of matrix majorization ([5]) where the matrices A and B have two rows and the first row is the all ones vector.

Theorem 3.1. *Let $A, B \in M_{m,n}$. Then the following hold.*

(i) $A \preceq^s B$ *if and only if* $A^t \preceq^{ds} B^t$ (by the definitions),

(*ii*) $A \preceq^s \Rightarrow A \preceq^d B \Rightarrow A \preceq^{wm} B$ (See the Hardy Littlewood Pólya Theorem (Theorem 46) in [8] and Proposition 3.3 in [11] which is proved by a separation argument.)

(*iii*) $A \preceq^{ds} B \Rightarrow A \preceq^m B$ (by the definitions).

Let $A, B \in M_{m,n}$, and let $\preceq$ be either $\preceq^{ds}$, $\preceq^m$, $\preceq^{wm}$ or $\preceq^s$. Then one can check if $A \preceq B$ holds efficiently (in polynomial time) using linear optimization. In fact, for each of these orders we look for a row-stochastic or doubly stochastic matrix X that satisfies a (finite) system of linear equations, namely $AX = B$ or $XA = B$ (depending on the order $\preceq$). Thus, one needs to decide if a certain system of linear inequalities has a solution, and this can be done by linear optimization.

Let $R(A)$ denote the set of rows of a matrix A. Also define, for a matrix class $\mathcal{A} \subseteq M_{m,n}$, $R(\mathcal{A}) = \bigcup\{R(A) : A \in \mathcal{A}\}$. The set of columns of a matrix or a matrix class are denoted, similarly, by $C(A)$ and $C(\mathcal{A})$. Recall that the *convex hull* of a set X of points in a Euclidean space is the smallest convex set that contains X.

Proposition 3.2. [11, Proposition 3.3] *Let $A, B \in M_{m,n}$. Then*

$$A \preceq^{wm} B \Leftrightarrow R(A) \subseteq \text{conv}(R(B)).$$

Proposition 3.3. *The following hold:*

(*i*) *$\preceq^s$ does not imply $\preceq^{ds}$ and conversely,*

(*ii*) *$A \preceq B$, where $\preceq$ is either $\preceq^s$, $\preceq^d$, or $\preceq^{wm}$ does not imply $C(A) \subseteq$ conv$(C(B))$,*

(*iii*) *$A \preceq B$, where $\preceq$ is either $\preceq^{ds}$ or $\preceq^m$ does not imply $R(A) \subseteq$ conv$(R(B))$.*

Proof. Consider the following Example:

Example 3.4.

$$\text{Let } D = \begin{bmatrix} 3/4 & 0 & 1/4 \\ 0 & 1 & 0 \\ 1/4 & 0 & 3/4 \end{bmatrix} \in \Omega_n, \quad B = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 2 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 11/4 & -1/2 \\ 1 & 3 \\ 9/4 & 1/2 \end{bmatrix}.$$

Since $DB = A$, $A \preceq^s B$ and as a consequence $A \preceq^d B$ and $A \preceq^{wm} B$. It is easy to verify that $\begin{bmatrix} 11/4 & 1 & 9/4 \end{bmatrix}^t \notin \operatorname{conv}(C(B))$.

It follows that $R(A^t) \not\subseteq \operatorname{conv}(R(B^t)) \Rightarrow A^t \not\preceq^{wm} B^t \Rightarrow A \not\preceq^{ds} B$. Thus we have proved (ii) and a half of (i).

To conclude the proof one should do the same calculations with A^t, B^t and D^t. (In this case $A^t = B^t D^t$, so $A^t \preceq^{ds} B^t$ and $A^t \preceq^m B^t$.) □

Now, for the rest of this section, let $\preceq$ denote any of these matrix majorization orders. We define our main concept as follows.

Definition 3.5. Let $\mathcal{A}$ and $\mathcal{B}$ be two, possibly, finite classes of matrices in $M_{m,n}$. We say that $\mathcal{A}$ is *majorized* by $\mathcal{B}$ if

$$\text{for all } A \in \mathcal{A} \text{ there exists some } B_A \in \mathcal{B} \text{ such that } A \preceq B_A. \tag{1}$$

If this holds, we write $\mathcal{A} \trianglelefteq \mathcal{B}$, and we call this the *majorization matrix class order* (MMC-order). We also use the notation $\trianglelefteq^d$ when $\preceq$ is $\preceq^d$, and, similarly, $\trianglelefteq^{ds}$ corresponds to $\preceq^{ds}$ and the same for the other matrix order relations.

Remark 3.6. Here again strong majorization implies directional and directional majorization implies weak majorization, see Theorem 3.1. □

Example 3.7. (i) If $\mathcal{A} = \{A\}$ and $\mathcal{B} = \{B\}$, then $\mathcal{A} \trianglelefteq \mathcal{B}$ means $A \preceq B$. So the MMC-order generalizes the matrix majorization order above, and therefore also classical vector majorization.

(ii) Assume $\mathcal{B} = \{B\}$. Then $\mathcal{A} \trianglelefteq \mathcal{B}$ means $A \preceq B$ for all $A \in \mathcal{A}$.

(iii) Assume $\mathcal{A} = \{A\}$. Then $\mathcal{A} \trianglelefteq \mathcal{B}$ means that $A \preceq B$ for some $B \in \mathcal{B}$. □

Let us consider situation (ii) in Example 3.7 in the special case of vector majorization, so strong majorization $\preceq^s$ with $n = 1$.

Proposition 3.8. *Let $b = (b_1, b_2, \ldots, b_m)^t$ and $a^{(j)} = (a_1^{(j)}, a_2^{(j)}, \ldots, a_m^{(j)})^t$ be vectors in $\mathbb{R}^m$ $(j \le p)$. Then the following statements are equivalent:*

(i) $\{a^{(1)}, a^{(2)}, \ldots, a^{(p)}\} \trianglelefteq^s \{b\}$.

(ii) $\max_{j \le p} \sum_{i=1}^k a_{[i]}^{(j)} \le \sum_{i=1}^k b_{[i]}$ $(k < m)$ *and* $\sum_{i=1}^m a_i^{(j)} = \sum_{i=1}^m b_i$ $(j \le p)$.

(iii) *There are matrices* $X^{(j)} \in \Omega_m$ *with* $a^{(j)} = X^{(j)} b$ $(j \le p)$.

(iv) *All permutations of* $a^{(j)}$ *for* $j \leq p$ *lie in the convex hull of all permutations of* b.

Proof. This follows directly from general majorization theory, see e.g. Theorem P1 in [11]. □

Consider two matrix classes $\mathcal{U} \subseteq M_{m,n}$ and $\mathcal{V} \subseteq M_{n,k}$. Define

$$\mathcal{A} \circ \mathcal{B} = \{AB : A \in \mathcal{A},\ B \in \mathcal{B}\} \subseteq M_{m,k},$$

the set of pairwise products.

Recall that Ω_n^{row} (resp., Ω_n) denotes the set of all row-stochastic (resp., doubly-stochastic) matrices of order n. Then we observe the following for matrix classes $\mathcal{A}, \mathcal{B} \subseteq M_{m,n}$:

- Doubly stochastic majorization: $\mathcal{A} \trianglelefteq^{ds} \mathcal{B}$ if and only if $\mathcal{A} \subseteq \mathcal{B} \circ \Omega_n$.
- Matrix majorization: $\mathcal{A} \trianglelefteq^{m} \mathcal{B}$ if and only if $\mathcal{A} \subseteq \mathcal{B} \circ \Omega_n^{row}$.
- Weak matrix majorization: $\mathcal{A} \trianglelefteq^{wm} \mathcal{B}$ if and only if $\mathcal{A} \subseteq \Omega_m^{row} \circ \mathcal{B}$.
- Strong majorization: $\mathcal{A} \trianglelefteq^{s} \mathcal{B}$ if and only if $\mathcal{A} \subseteq \Omega_m \circ \mathcal{B}$.

Proposition 3.9. *Let* $\mathcal{A}$, $\mathcal{B}$ *and* $\mathcal{C}$ *be matrix classes in* $M_{m,n}$. *Then*

(i) $\mathcal{A} \trianglelefteq \mathcal{A}$.

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

(iii) *If* $\mathcal{A} \trianglelefteq \mathcal{B}$, *then*

$$\max_{A \in \mathcal{A}} \operatorname{rank} A \leq \max_{B \in \mathcal{B}} \operatorname{rank} B.$$

(iv) *If* $\mathcal{A} \trianglelefteq \mathcal{B}$, *then*

$$\max_{A \in \mathcal{A}} |\det A| \leq \max_{B \in \mathcal{B}} |\det B|.$$

Proof. (i) and (ii) follow from the definition. (iii): This follows from the fact that if matrices A and B satisfy $A \preceq B$ for any of these orders except $\preceq^d$, then, by definition, either $A = BX$ or $A = XB$. It follows that $\operatorname{rank} A \leq \operatorname{rank} B$. By the second part of Remark 3.1, directional majorization implies weak majorization, so what just stated, proves the result for $\preceq^d$ as well. The proof of (iv) follows from the fact that for all $X \in \Omega_n^{row}$, $|\det X| \leq 1$ (this follows directly by Laplace expansion on the first row). Therefore, for all $A \in \mathcal{A}$ there exists some $B \in \mathcal{B}$ such that $|\det A| \leq |\det B|$ since there exists $X \in \Omega_n^{row}$ such that either $A = BX$ or $A = XB$ (depending on the majorization type). ∎

There is an analogue of Proposition 3.2 for matrix classes.

Proposition 3.10. *Let $\mathcal{A}$, $\mathcal{B}$ be matrix classes in $M_{m,n}$. Then $\mathcal{A} \trianglelefteq \mathcal{B}$, where $\trianglelefteq$ is either $\trianglelefteq^s$, $\trianglelefteq^d$ or $\trianglelefteq^{wm}$, implies that $R(\mathcal{A}) \subseteq \operatorname{conv}(R(\mathcal{B}))$.*

Proof. For matrices A and B, if $A \preceq B$ for one of the orders $\preceq^s$ or $\preceq^d$, then $A \preceq^{wm} B$. By Proposition 3.2 each row in A is a convex combination of the rows in B. Finally, for every $x^t \in R(\mathcal{A})$ then there exists some $B \in \mathcal{B}$ such that $x^t \in \operatorname{conv}(R(B))$ and as a consequence $x^t \in \operatorname{conv}(R(\mathcal{B}))$. ∎

Let $\mathcal{A}$, $\mathcal{B}$ be matrix classes in $M_{m,n}$. Then $R(\mathcal{A}) \subseteq \operatorname{conv}(R(\mathcal{B}))$ does not imply $\mathcal{A} \preceq^{wm} \mathcal{B}$ in general as the following example shows.

Example 3.11. Let $\mathcal{A} = \{A_1\} = \{\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}\}$ and $\mathcal{B} = \{\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}\}$. Then $R(A_1) \subseteq \operatorname{conv}(R(\mathcal{B}))$, but $\begin{bmatrix} 1 & 0 \end{bmatrix} \notin \operatorname{conv}(R(B_2))$ and $\begin{bmatrix} -1 & 0 \end{bmatrix} \notin \operatorname{conv}(R(B_1))$. It follows that $A_1 \not\preceq^{wm} B_i$ and, as a result, $\mathcal{A} \not\trianglelefteq^{wm} \mathcal{B}$. ∎

Given $A \in M_{m,n}$. We denote $[A, v] \in M_{m,n+1}$ the matrix whose first (ordered) n columns are equal to those of A and its last column is the vector v. $\left[\begin{smallmatrix} A \\ v^t \end{smallmatrix}\right]$ denotes the same for rows.

For $\mathcal{A} = \{A_i : 1 \leq i \leq p\}$ the class $[\mathcal{A}, v]$ (resp. $\left[\begin{smallmatrix} \mathcal{A} \\ v^t \end{smallmatrix}\right]$) denotes the class formed by the extended matrices: $\{[A_i, v] : 1 \leq i \leq p\}$ (resp. $\{\left[\begin{smallmatrix} A_i \\ v^t \end{smallmatrix}\right] : 1 \leq i \leq p\}$). Recall that e is the vector of 1.

Proposition 3.12. *Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of matrices in $M_{m,n}$. Then,*

(i) If $\trianglelefteq$ is either $\trianglelefteq^s$, $\trianglelefteq^d$ or $\trianglelefteq^{wm}$ then $\mathcal{A} \trianglelefteq \mathcal{B}$ if and only if $[\mathcal{A}, e] \trianglelefteq [\mathcal{B}, e]$.

(ii) $\mathcal{A} \trianglelefteq^{ds} \mathcal{B}$ if and only if $\left[\begin{smallmatrix} \mathcal{A} \\ e^t \end{smallmatrix}\right] \trianglelefteq^{ds} \left[\begin{smallmatrix} \mathcal{B} \\ e^t \end{smallmatrix}\right]$.

Proof. This follows directly from the definitions. ∎

Let $A \in M_{m,n}$. For each $J \subset \{1, 2, \ldots, n\}$ we denote by $A[; J]$ the submatrix of A whose columns are indexed by the elements in J. Similarly for each $I \subset \{1, 2, \ldots, n\}$ we denote by and $A[I;]$ the submatrix of A whose rows are indexed by the elements in I.

For $\mathcal{A} = \{A_i,\ i \in \{1, 2, \dots, p\}\}$ the class $\mathcal{A}[I;]$ denotes the class formed by the corresponding submatrices $\{A_i[I;],\ i \in \{1, 2, \dots, p\}\}$ and similarly for columns.

Proposition 3.13. *Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of matrices in $M_{m,n}$, and sets I and J as above. Then the following holds.*

(i) If $\trianglelefteq$ is either $\trianglelefteq^s$, $\trianglelefteq^d$ or $\trianglelefteq^{wm}$, then $\mathcal{A} \trianglelefteq \mathcal{B} \Rightarrow \mathcal{A}[;J] \trianglelefteq \mathcal{B}[;J]$.

(ii) If $\trianglelefteq$ is either $\trianglelefteq^{ds}$ or $\trianglelefteq^m$, then $\mathcal{A} \trianglelefteq \mathcal{B} \Rightarrow \mathcal{A}[I;] \trianglelefteq \mathcal{B}[I;]$.

Proof. This proposition follows directly from the same result for matrices (see [5, Theorem 2.1(iii)] for $\preceq^{ds}$ or $\preceq^m$ and [11, Proposition 3.5] for $\preceq^s$, $\preceq^d$ or $\preceq^{wm}$) □

For a matrix $A \in M_{m,n}$ and an integer k, let $\overline{A}(k)$ denote the matrix with $\binom{m}{k}$ rows and n columns, where the rows of $\overline{A}(k)$ are all possible averages (arithmetic means) of various k-tuples of different rows of A, the arithmetic mean of k vectors $v_1, v_2, \dots, v_k$ is the vector whose ith coordinate is the arithmetical mean of ith coordinates of $v_1, v_2, \dots, v_k$.

Proposition 3.14. ([11, Corollary 3.13]) *Let $A, B \in M_{m,n}$. Then $A \preceq^d B$ if and only if*

$$\overline{A}(k) \preceq^{wm} \overline{B}(k) \quad (k \in \{1, 2, \dots, [\frac{m}{2}], m\}).$$

For $\mathcal{A} = \{A_i,\ i \in \{1, 2, \dots, p\}\}$ and an integer k, let $\overline{\mathcal{A}}(k)$ denote $\{\overline{A_i}(k),\ i \in \{1, 2, \dots, p\}\}$.

Corollary 3.15. *Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of matrices. Then*
$\mathcal{A} \trianglelefteq^d \mathcal{B} \Rightarrow \overline{\mathcal{A}}(k) \trianglelefteq^{wm} \overline{\mathcal{B}}(k)$ for $k = 1, 2, \dots, [\frac{m}{2}]$ and $k = m$.

Proof. This statement follows directly from the same result for matrices. □

The converse implication does not hold as the following example shows.

Example 3.16. Let $\mathcal{A} = \{A_1\} = \{\begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}\}$ and $\mathcal{B} = \{B_1, B_2\}$, where $B_1 = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix}$, $B_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Then it is easy to see that $A_1 \preceq^{wm} B_1$ and the column sum vectors $c(A_1) = c(B_2)$ that is $\overline{A_1}(1) \preceq^{wm} \overline{B_1}(1)$ and $\overline{A_1}(2) \preceq^{wm} \overline{B_2}(2)$. Then $\overline{\mathcal{A}}(k) \trianglelefteq^{wm} \overline{\mathcal{B}}(k)$ for $k = 1, 2$. But $c(A_1) \neq c(B_1)$ and $A_1 \npreceq^{wm} B_2$. Thus $\mathcal{A} \ntrianglelefteq^d \mathcal{B}$.

Theorem 3.17. *Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of matrices in $M_{m,n}$. If $\mathcal{A} \trianglelefteq \mathcal{B}$, where $\trianglelefteq$ is either $\trianglelefteq^s$, $\trianglelefteq^d$ or $\trianglelefteq^{wm}$, then*

$$\bigcap_{B \in \mathcal{B}} \ker B \subseteq \bigcap_{A \in \mathcal{A}} \ker A.$$

Proof. This theorem follows directly from the fact that $A \preceq^{wm} B$ implies $\ker B \subseteq \ker A$. □

4 Minimal cover classes

As explained in Section 2 a motivation for this work lies in the theory of comparison of statistical experiments ([13]), and we introduced the new ordering which compares two sets of experiments, $\mathcal{A} = \{\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_p\}$ and $\mathcal{B} = \{\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_q\}$. We say that $\mathcal{B}$ is *more informative* than $\mathcal{A}$ if for each $\mathcal{E}_j$ $(j \leq q)$, there exists an $i \leq p$ such that $\mathcal{F}_i$ is more informative than $\mathcal{E}_j$. Mathematically, this idea lies behind the MMC order of matrix classes, and it also motivates the main question in this section: how small can the set $\mathcal{B}$ be, when we require that $\mathcal{A} \trianglelefteq \mathcal{B}$. Thus, the general question is: For a given matrix class $\mathcal{A}$, determine

$$\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq \mathcal{B}\} \tag{2}$$

where we assume that $\mathcal{A} \trianglelefteq \mathcal{B}$ holds for some finite class $\mathcal{B}$. This is a kind of covering problem which is dependent on the underlying matrix majorization order $\preceq$. We say that a class $\mathcal{B}$ is a *minimal cover class* for $\mathcal{A}$ if $\mathcal{A} \trianglelefteq \mathcal{B}$ and the cardinality of $\mathcal{B}$ equals the minimum in (2).

For a matrix class $\mathcal{A}$ there may not exist a single matrix B such that $\mathcal{A} \trianglelefteq^{wm} \mathcal{B}$ where $\mathcal{B} = \{B\}$. The following example illustrates this.

Example 4.1. Let $\mathcal{A} = \{A_1, A_2\} = \{\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 2 \end{bmatrix}\}$. Then, for any $B \in M_{2,2}$, the set $\operatorname{conv}(R(B))$ is a line segment on a plane since $\operatorname{conv}(R(B))$ is a convex hull of two 2-dimensional points, and cannot contain $R(A_1) \cup R(A_2)$. Thus by Proposition 3.2 there is no any B such that $\mathcal{A} \trianglelefteq^{wm} \{B\}$. □

Remark 4.2. The construction in the example above shows that for all p there exists a matrix class $\mathcal{A} = \{A_1, A_2, \ldots, A_p\}$ such that $\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq^{wm} \mathcal{B}\} = p$. Here $|X|$ denotes the cardinality of a set or family X.

In order to show this, consider the following example.

Example 4.3. Let $\mathcal{A} = \{A_1, A_2, \dots, A_p\}$ such that $A_s = \begin{bmatrix} 0 & s \\ 1 & s \end{bmatrix}$, $s = 1, \dots, p$. Then, as in the example above, by Proposition 3.2 there is no any $\mathcal{B}$ such that $|\mathcal{B}| < p$ and $\mathcal{A} \trianglelefteq^{wm} \mathcal{B}$. □

Let $\mathcal{A}$ be a matrix class such that there exists a matrix B with $\mathcal{A} \trianglelefteq^{wm} \{B\}$. Then, in general, there may not exist a matrix C such that $\mathcal{A} \trianglelefteq^{s} \{C\}$, as the next example shows.

Example 4.4. $\mathcal{A} = \{A_1, A_2\}$ where $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $A_2 = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$. Then $\mathcal{A} \trianglelefteq^{wm} \{A_1\}$. Suppose that there exists a matrix C such that $A_1 \preceq^s C$ and $A_2 \preceq^s C$. It follows that $A_1 = \begin{bmatrix} x & 1-x \\ 1-x & x \end{bmatrix} C$ for some x with $0 \leq x \leq 1$, and $A_2 = DC$ for some doubly stochastic matrix D. Since A_1 is non-singular, $x \neq 1/2$. Then $C = \begin{bmatrix} \frac{x}{2x-1} & \frac{x-1}{2x-1} \\ \frac{x-1}{2x-1} & \frac{x}{2x-1} \end{bmatrix}$. It follows that $D = A_2 C^{-1} = \begin{bmatrix} x & 1-x \\ 1/2 & 1/2 \end{bmatrix}$. Then D is not doubly stochastic, a contradiction.
□

Classical vector majorization has a nice geometrical interpretation, given by Rado's theorem ([10]): for $x, y \in \mathbb{R}^n$, x is majorized by y if and only if x lies in the convex hull of all points obtained by permuting the components of y. Moreover, the geometry of matrix majorization was studied in [5, 6]. Our next results continue on this line of geometrical approach to majorization.

A simplex is a generalization of the notion of a triangle or tetrahedron to arbitrary dimension. Suppose the $k+1$ points $u_0, u_1, \dots, u_k \in \mathbb{R}^k$ are affinely independent, which means that the vectors $\{u_1 - u_0, u_2 - u_0, \dots, u_k - u_0\}$ are linearly independent. Then the simplex determined by these points is the set

$$\mathcal{S} = \left\{ \theta_0 u_0 + \theta_1 u_2 + \cdots + \theta_k u_k : \sum_{i=0}^{k} \theta_i = 1 \text{ and } \theta_i \geq 0 \text{ for all } i \right\}.$$

The points $u_0, u_1, \dots, u_k \in \mathbb{R}^k$ are vertices or extreme points of the simplex. It can be said also that a k-simplex is a k-dimensional polytope which is the convex hull of its $k+1$ vertices. For example, a 2-simplex is a triangle, a 3-simplex is a tetrahedron.

Remark 4.5. Let $\mathcal{A}$ be a matrix class in $M_{m,n}$ with $\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq^{wm} \mathcal{B}\} = 1$. Then each $A \in \mathcal{A}$ has all its rows in the same $(n-1)$-simplex. In fact, since there exists $B \in M_{m,n}$ such that $\mathcal{A} \trianglelefteq^{wm} \{B\}$, the points in $\mathbb{R}^n$ corresponding to the rows of matrices in $\mathcal{A}$ are contained in $\operatorname{conv}(R(B))$ and therefore in some $(n-1)$-simplex. □

Recall our standing assumption that the matrix classes considered (in majorization statements) are assumed to be finite.

Proposition 4.6. *Let $\mathcal{A}$ be a matrix class in $M_{m,n}$ where $m > n$. Then*

$$\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq^{wm} \mathcal{B}\} = 1.$$

Proof. The set $R(\mathcal{A})$, consisting of all rows of the matrices in $\mathcal{A}$, is a subset of $\mathbb{R}^n$. Choose a ball $S_r = \{x \in \mathbb{R}^n : \|x\| \leq r\}$, for some $r > 0$, such that $R(\mathcal{A}) \subseteq S_r$. Then there exists a full-dimensional simplex P containing S_r. The simplex is the convex hull of its $n+1 \leq m$ extreme points. Let B be the $m \times n$ matrix with these extreme points as rows and otherwise zero rows (there are $m-n-1 \geq 0$ of these). Then $R(\mathcal{A}) \subseteq \operatorname{conv}(R(B))$, so, by Proposition 3.2, $\mathcal{A} \trianglelefteq^{wm} \{B\}$, as desired. □

Proposition 4.7. *Let $\mathcal{A}$ be a matrix class in $M_{m,n}$. Then*

$$\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq^{d} \mathcal{B}\} \geq \Sigma_{\mathcal{A}}$$

where $\Sigma_{\mathcal{A}}$ is the number of matrices in $\mathcal{A}$ with different column sum vectors.

Proof. If $A \preceq^d B$ then the column sums of A and B coincide, so $c(A) = c(B)$. It follows that for every $A_i \in \mathcal{A}$ there exists $B_j \in \mathcal{B}$ such that $c(A_i) = c(B_j)$, and then obviously $\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq^{d} \mathcal{B}\} \geq \Sigma_{\mathcal{A}}$. □

As the following example shows the inequality in the previous proposition may be strict.

Example 4.8. Let $A_1 = \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix}$, $A_2 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$. Then column sum vectors $c(A_1) = c(A_2) = (2, 2)$, so $\Sigma_{\mathcal{A}} = 1$. However, by Proposition 3.2 there is no B such that $A_1 \preceq^{wm} B$ and $A_2 \preceq^{wm} B$, and since $\preceq^d$ is stronger than $\preceq^{wm}$, there is no B such that $A_1 \preceq^d B$ and $A_2 \preceq^d B$, as required. □

Consider again the relation $\preceq^{wm}$ and the characterization in Proposition 3.2. We assume $m \leq n$, as the case of $m > n$ was settled in Proposition 4.6.

Lemma 4.9. *Let $\mathcal{A}$ be a matrix class $\mathcal{B}$ in $M_{m,n}$. Then* $\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq^{wm} \mathcal{B}\}$ *is a minimal number of hyperplanes, covering (as above) the set of polytopes* $\operatorname{conv}(R(A))$*, where* $A \in \mathcal{A}$*.*

Proof. We may conclude, in the case of weak matrix majorization, that a minimal cover class of some matrix class $\mathcal{A} \subseteq M_{m,n}$ is a class $\mathcal{B} \subseteq M_{m,n}$ such that each polytope $\operatorname{conv}(R(A))$, where $A \in \mathcal{A}$, is contained in the polytope $\operatorname{conv}(R(B))$ for some $B \in \mathcal{B}$. Thus, we are looking for a minimal family of polytopes with this specific covering property. It follows that $\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq \mathcal{B}\}$ is a minimal number of hyperplanes, covering (as above) a set of polytopes $\operatorname{conv}(R(A))$, where $A \in \mathcal{A}$. □

Lemma 4.10. *For every matrix class $\mathcal{A}$ there exists an algorithm for finding a minimal cover class in the case of weak majorization.*

Proof. We may enumerate every partition of $\mathcal{A}$ and, for each partition, we may check that every subclass can be weakly majorized by one matrix in polynomial time. Indeed, we only have to verify that for each subclass of the partition all points corresponding to each row of each matrix in this subclass belongs to the same hyperplane. □

We remark that it is possible to reduce complexity of the algorithm in the previous proof to $2^{|\mathcal{A}|}poly(n, m, |\mathcal{A}|) + 3^{|\mathcal{A}|}poly(|\mathcal{A}|)$ using dynamic programming over subsets of $\mathcal{A}$.

Lemma 4.11. Let $\mathcal{A} \subseteq M_{m,n}$ be a matrix class such that for every $A \in \mathcal{A}$ the rows of A are vertices of an $(m-1)$-simplex. Then there is a polynomial algorithm for finding minimal cover class. □

Proof. If the rows of matrix A are vertices of an $(m-1)$-simplex, then the hyperplane corresponding to the convex hull of rows of matrix A is unique. Thus for A_1, A_2 there exists B such that $A_i \preceq^{wm} B$ if and only if the hyperplanes corresponding to A_1, A_2 coincide. □

Thus in the case of weak matrix majorization we can find, computationally, a minimal cover class $\mathcal{B}$ of any class $\mathcal{A} \subseteq M_{m,n}$.

Now, let us consider the case of directional majorization. Recall that for given $A \in M_{m,n}$, the matrix $\overline{A}(k) \in M_{\binom{m}{k},n}$ denotes the matrix, such that each its row is the average of some k different rows of A.

A *barycenter* (or a *centroid*) of a set is a center of mass of all its points. For a k-simplex with the vertices $u_0, u_1, \ldots, u_k \in \mathbb{R}^n$, its barycenter is the average point of the vertices: $\frac{1}{k+1}\sum_{i=0}^{k} u_k$.

Definition 4.12. For any matrix $X \in M_{m,n}$, let P_X be a polytope in $\mathbb{R}^n$ with m vertices corresponding to the rows of X. For the convenience we suppose that there are exactly m vertices of P_X even if some of them coincide.

Lemma 4.13. *Let $X \in M_{m,n}$. Then $\frac{1}{m}c(X)$ is the barycenter of P_X.*

Proof. By the definition of P_X, it has exactly m vertices. The rest follows from the definition of barycenter. □

Note that Lemma 4.13 may not hold for $\operatorname{conv}(R(X))$ as the following example shows:

Example 4.14. Let $X = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 3 & 0 \end{bmatrix}$. Then $\operatorname{conv}(R(X))$ is a line segment with start point at $\begin{bmatrix} 0 & 0 \end{bmatrix}$ and end point at $\begin{bmatrix} 3 & 0 \end{bmatrix}$. It follows that the barycenter of $\operatorname{conv}(R(X))$ is different from $\begin{bmatrix} 1 & 0 \end{bmatrix} = \frac{1}{3}c(X)$, and $\operatorname{conv}(R(X))$ is a polytope with only two vertices.

□

Theorem 4.15. *Let $\mathcal{A}$ be a matrix class in $M_{m,n}$. Assume that there exists a matrix $B \in M_{m,n}$ such that $\mathcal{A} \trianglelefteq^{wm} \{B\}$ and the vectors of column sums of all $A_i \in \mathcal{A}$ coincide, that is $c(A_i) = c(A_j)$ for all i, j. Then there exists $C \in M_{m,n}$ such that $\mathcal{A} \trianglelefteq^{d} \{C\}$.*

Proof. The condition $c(A_i) = c(A_j)$ for all i, j implies coincidence of the barycenters of the convex polytopes P_{A_i} corresponding to the matrices A_i. We denote this common barycenter by O. Without loss of generality we may assume that the barycenter of the polytope P_B corresponding to B is also O. Indeed, in the other case we can consider B' instead of B such that the polytope $P_{B'}$ corresponding to B' is obtained from P_B by the parallel shift moving the barycenter of P_B to O and homothety providing that $P_B \subseteq P_{B'}$. Thus the conditions of the lemma are satisfied for B' and barycenter of $P_{B'}$ is O.

It is straightforward to check that for any $X \in M_{m,n}$ the convex hull of the row space $\operatorname{conv}(R(\overline{X}(k))) \subseteq \operatorname{conv}(R(X))$ for any k with $1 \le k \le m$, since the rows of $\overline{X}(k)$ lie in the convex hull of the rows of X by the definition of the matrix $\overline{X}(k)$. Then, by Proposition 3.14, we obtain that if there is a matrix $C \in M_{m,n}$ such that $\operatorname{conv}(R(B)) \subseteq \bigcap_{k=1,\ldots,[\frac{m}{2}]} R(\overline{C}(k))$ and $c(B) = c(C)$, then $\mathcal{A} \trianglelefteq^{d} C$. Indeed, for any $A_i \in \mathcal{A}$ and $k = 1, 2, \ldots, [\frac{m}{2}]$ it holds that

$\operatorname{conv}(R(\overline{A_i}(k))) \subseteq \operatorname{conv}(R(A)) \subseteq \operatorname{conv}(R(B)) \subseteq \operatorname{conv}(R(\overline{C}(k)))$. Moreover, $c(A_i) = c(C)$, i.e., all conditions of Proposition 3.14 are satisfied, so, it is applicable and provides the required result.

It remains to show that such C always exists. We will show this constructively in several lemmas, namely Lemmas 4.16 – 4.18 below.

Note that for an arbitrary matrix D, since $c(D) \in \bigcap_{k=1,\dots,[\frac{m}{2}]} R(\overline{D}(k))$, we obtain that $\bigcap_{k=1,\dots,[\frac{m}{2}]} \overline{D}(k) \neq \emptyset$. Moreover, the barycenters of all matrices $\overline{D}(k)$, $k = 1, 2, \dots, [\frac{m}{2}]$ coincide by the definition of $\overline{D}(k)$.

Lemma 4.16. *Let $D \in M_{m,n}$. Then the subspaces $\mathbb{R}D$ and $\mathbb{R}\overline{D}(k)$, the minimal subspaces of $\mathbb{R}^n$ containing P_D and $P_{\overline{D}(k)}$, correspondingly, coincide.*

Proof. Let O be the barycenter of P_D. We denote the vertices of P_D by $D_{(i)}$, $i = 1, 2, \dots, m$ and the vertices of $P_{\overline{D}(k)}$ by $D^k_{(i)}$, $i = 1, 2, \dots, \binom{m}{k}$. By the definition of $\overline{D}(k)$ it holds that $\overrightarrow{OD^k_{(i)}} \in \mathbb{R}D$ for all i. Thus $\mathbb{R}D \supseteq \mathbb{R}\overline{D}(k)$.

Also by the definition, for each $i = 1, 2, \dots, \binom{m}{k}$ the vector

$$\overrightarrow{OD^k_{(i)}} = \frac{1}{k}\Big(\overrightarrow{OD_{(i_1)}} + \dots + \overrightarrow{OD_{(i_k)}}\Big). \tag{3}$$

Then for any fixed $j = 1, 2, \dots, m$ there are $\binom{m-1}{k-1}$ vertices $D^k_{(i)}$ such that the corresponding expression (3) contains $D_{(j)}$. We denote these vertices by ${}_jD^k_{(u)}$, $u = 1, 2, \dots, \binom{m-1}{k-1}$. Hence there are $\binom{m}{k} - \binom{m-1}{k-1} = \binom{m-1}{k}$ vertices $D^k_{(i)}$ such that the corresponding expression (3) does not contain $D_{(j)}$. We denote these vertices by ${}_{\hat{j}}D^k_{(v)}$, $v = 1, 2, \dots, \binom{m-1}{k}$. Let us sum up all vectors $\overrightarrow{O\,{}_jD^k_{(u)}}$ and express them via (3). Then we get

$$\sum_{u=1}^{\binom{m-1}{k-1}} \overrightarrow{O\,{}_jD^k_{(u)}} = s\overrightarrow{OD_{(j)}} + t(\overrightarrow{OD_{(1)}} + \dots + \overrightarrow{OD_{(j-1)}} + \overrightarrow{OD_{(j+1)}} + \dots + \overrightarrow{OD_{(m)}}) \tag{4}$$

for some s and t. Note that $s > t$, since each $\overrightarrow{O\,{}_jD^k_{(i)}}$ contains $\overrightarrow{OD_{(j)}}$. Now we sum up all vectors $\overrightarrow{O\,{}_{\hat{j}}D^k_{(v)}}$ to get

$$\sum_{v=1}^{\binom{m-1}{k-1}} \overrightarrow{O\,{}_{\hat{j}}D^k_{(v)}} = l\left(\overrightarrow{OD_{(1)}} + \dots + \overrightarrow{OD_{(j-1)}} + \overrightarrow{OD_{(j+1)}} + \dots + \overrightarrow{OD_{(m)}}\right) \tag{5}$$

for some l.

Therefore, by multiplying (4) with l and (5) with s one obtains

$$s \cdot l \cdot \overrightarrow{OD_{(j)}} = l \cdot \sum_{u=1}^{\binom{m-1}{k-1}} \overrightarrow{O_{\ j}D^k_{(u)}} - t \cdot \sum_{v=1}^{\binom{m-1}{k}} \overrightarrow{O_{\ \hat{j}}D^k_{(v)}}. \tag{6}$$

Hence, the other inclusion holds and the required equality $\mathbb{R}D = \mathbb{R}\overline{D}(k)$ is proved. ∎

Now we find s, l and t explicitly.

Lemma 4.17. *The parameters s, l and t from Lemma 4.16 can be expressed as follows*

$$s = \frac{\binom{m-1}{k-1}}{k}, \quad t = \frac{\binom{m-2}{k-2}}{k} \text{ and } l = \binom{m-2}{k-1}.$$

Proof. As it was mentioned in the beginning of the proof of Lemma 4.16 there are $\binom{m-1}{k-1}$ vectors $\overrightarrow{O_{\ j}D^k_{(u)}}$. Thus each vector $OD_{(j)}$ appears in the equality (4) for $\binom{m-1}{k-1}$ times with the scalar $1/k$. Furthermore, there are $\binom{m-2}{k-2}$ vectors $\overrightarrow{O_{\ j}D^k_{(u)}}$ with $OD_{(i)}$ and $OD_{(j)}$ in their corresponding sums for some fixed $i \neq j$ since we choose $(k-2)$ other vectors among $(m-2)$ that are left. Thus $t = \frac{\binom{m-2}{k-2}}{k}$.

Finally, there are $\binom{m-1}{k}$ vectors $\overrightarrow{O_{\ \hat{j}}D^k_{(i)}}$. As a consequence, $k\binom{m-1}{k} = l(m-1)$. Hence, $l = \binom{m-2}{k-1}$. ∎

Since the linear subspaces $\mathbb{R}D$ and $\mathbb{R}\overline{D}(k)$ coincide, as it was stated in Lemma 4.16, we can first consider $D = B$, and then elongate all sides of the polytope in order to obtain the matrix C, for which it holds that $\bigcap\limits_{k=1,\dots,[\frac{m}{2}]} R(\overline{C}(k)) \supseteq \operatorname{conv}(R(B))$.

Lemma 4.18. *Let C be a matrix in $M_{m,n}$ such that P_C is obtained from P_B by the α-homothety with the center O and $\alpha = \max\limits_{k=1,\dots,[\frac{m}{2}]}(k + \frac{(k-1)}{k}(\binom{m}{k} - 1))$. Then $\bigcap\limits_{k=1,\dots,[\frac{m}{2}]} R(\overline{C}(k)) \supseteq \operatorname{conv}(R(D))$.*

Proof. We plan to explore the equality (6). Note that $\sum\limits_i \overrightarrow{OD^k_{(i)}} = \overrightarrow{OO} = O$ since O is the barycenter. Then $-\overrightarrow{OD^k_{(j)}} = \sum\limits_{i \neq j} \overrightarrow{OD^k_{(i)}}$.

It follows that $\overrightarrow{OD_{(j)}} = \frac{1}{s}\sum \overrightarrow{O_{\,j}D^k_{(i)}} + \frac{t}{sl}\sum\limits_{q\in Q}\sum\limits_{i\neq q}\overrightarrow{OD^k_{(i)}}$ where Q is the set of all indexes q such that $\overrightarrow{OD^k_{(q)}}$ is $\overrightarrow{O_{\,\hat{j}}D^k_{(i)}}$.

Thus $\overrightarrow{OD_{(j)}}$ is a linear combination of vectors $\overrightarrow{OD^k_{(i)}}$ with positive scalars. The sum of all the scalars is equal to

$$\frac{\binom{m-1}{k-1}}{s} + \frac{t}{sl}\binom{m-1}{k}\left(\binom{m}{k} - 1\right) =: \alpha(k).$$

Finally,

$$\overrightarrow{OD_{(j)}} = \frac{\frac{1}{s}\sum \alpha(k)\overrightarrow{O_{\,j}D^k_{(i)}} + \frac{t}{sl}\sum\limits_{q\in Q}\sum\limits_{i\neq q}\alpha(k)\overrightarrow{OD^k_{(i)}}}{\alpha(k)}$$

$$\overrightarrow{OD_{(j)}} = \frac{\frac{1}{s}\sum \overrightarrow{O_{\,j}C^k_{(i)}} + \frac{t}{sl}\sum\limits_{q\in Q}\sum\limits_{i\neq q}\overrightarrow{OC^k_{(i)}}}{\alpha(k)}$$

for every j.

It follows that every $\overrightarrow{OD_{(j)}}$ is a convex combination of $\overrightarrow{OC^k_{(j)}}$.

Let $\alpha = \max\limits_{k=1,\ldots,[\frac{m}{2}]}\alpha(k)$. Then it is easy to see that, if C is such that P_C is obtained from P_B by the α-homothety with the center O, then $\overrightarrow{OD_{(j)}}$ is a convex combination of $OC^k_{(j)}$ for every $k\in\{1,\ldots,[\frac{m}{2}]\}$. Indeed we have already shown that it is a linear combination with positive scalars and by definition of α the sum of this scalars is less or equal to 1. If it is equal to 1 then it is proved, else we may add to the linear combination $\sum\limits_i OC^k_{(i)} = 0$ with the suitable scalar. Moreover,

$$\alpha(k) = \frac{\binom{m-1}{k-1}}{s} + \frac{t}{sl}\binom{m-1}{k}(\binom{m}{k} - 1) = k + \frac{\binom{m-2}{k-2}}{\binom{m-1}{k-1}\binom{m-2}{k-1}}\frac{m-1}{k}\binom{m-2}{k-1}(\binom{m}{k} - 1) =$$

$$= k + \frac{(k-1)\binom{m-2}{k-2}}{k(m-1)\binom{m-2}{k-2}}(m-1)(\binom{m}{k} - 1) = k + \frac{(k-1)}{k}(\binom{m}{k} - 1).$$

□

Finishing the proof of Theorem 4.15:

Thus $\bigcap_{k=1,\ldots,[\frac{m}{2}]} R(\overline{C}(k)) \supseteq \operatorname{conv}(R(D)) = \operatorname{conv}(R(B)) \supseteq \operatorname{conv}(R(A_i)) \supseteq \bigcup_{k=1,\ldots,[\frac{m}{2}]} R(\overline{A_i}(k))$ for any $A_i \in \mathcal{A}$. Therefore, $\mathcal{A} \trianglelefteq^d \{C\}$ as required.

□

Remark 4.19. *Let $\mathcal{A} = \{A_1, A_2\}$ such that $c(A_1) \neq c(A_2)$. Then there is no B such that $\mathcal{A} \trianglelefteq^d \{B\}$.*

Proof. Suppose that there exists some B such that $\mathcal{A} \trianglelefteq^d \{B\}$. It follows that $A_i \preceq^d B$ and by Proposition 3.14 $c(A_1) = c(B) = c(A_2)$, a contradiction. □

Remark 4.20. In order to find a matrix C as in Theorem 4.15 we do not use the class $\mathcal{A}$ itself, except the information about the barycenter. We may use just a matrix, covering $\mathcal{A}$ in terms of weak majorization. □

Theorem 4.15 allows us to find a general algorithm for finding a minimal cover class of a given $\mathcal{A} \subseteq M_{m,n}$. First of all, we split the class $\mathcal{A}$ into equivalence classes according to the column sum vectors of the matrices. Then, for each class, we find a minimal cover class in terms of weak majorization. Now we can split all of the equivalence classes for the subsets which can be covered by one matrix and with the same vectors of column sums. The number of these subsets equals to $\min\{|\mathcal{B}| : \mathcal{A} \trianglelefteq^d \mathcal{B}\}$ and an example of such a class can be found as we have shown in the proof of Theorem 4.15.

The following property allows us to find a minimal cover class in terms of strong majorization for a matrix class $\mathcal{A}$ such that for each $A \in \mathcal{A}$ its set of rows $R(A)$ is the vertex set of a simplex.

Proposition 4.21. ([11, Corollary 3.21]) *Let $A, B \in M_{m,n}$ be such that the rows of A are the vertices of a simplex. If $A \preceq^{wm} B$ and $c(A) = c(B)$, then $A \preceq^s B$.*

There are several cases when weak majorization implies strong majorization. Some of them are studied in [11, Subsection 3.2]. In addition, the following theorem allows us to present another case.

Theorem 4.22. *Let $A, B \in M_n$ where B is invertible. Then $A \preceq^d B$ if and only if $A \preceq^s B$.*

Proof. We only need to prove that $A \preceq^d B$ implies $A \preceq^s B$. So assume that $A \preceq^d B$ holds. Then $Ax \preceq Bx$ for all $x \in \mathbb{R}^n$. Since B is invertible, $y = Bx$ runs through $\mathbb{R}^n$ whenever x runs through $\mathbb{R}^n$. Therefore, $Ax = AB^{-1}y \preceq y$ for all $y \in \mathbb{R}^n$ which, by Ostrowski's theorem (see Theorem 2.A.4 in [10]), means that $AB^{-1} \in \Omega_n$. So $AB^{-1} = R$ for some $R \in \Omega_n$, and therefore $A = RB$. This means that $A \preceq^s B$, as desired. □

From the previous theorem we get the following result.

Lemma 4.23. *Let $\mathcal{A}, \mathcal{B}$ be matrix classes in M_n such that $\mathcal{A} \trianglelefteq^d \mathcal{B}$,*

$$|\mathcal{B}| = \min\{|\mathcal{C}| : \mathcal{A} \trianglelefteq^d \mathcal{C}\}$$

and each $B \in \mathcal{B}$ is invertible. Then $|\mathcal{B}| = \min\{|\mathcal{C}| : \mathcal{A} \trianglelefteq^s \mathcal{C}\}$.

Remark 4.24. It is important to note that in general for some matrix $A \in M_n$ there does not exist an invertible matrix $B \in M_n$ such that $A \preceq B$, here $\preceq$ denotes any of introduced majorizations. □

Example 4.25. Let $A = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$. Let $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $A \preceq^{wm} B$. It follows that there exists a row stochastic matrix $R = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}$ such that $\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. It is easy to verify that $a = b$ and $c = d$. Thus we have proved the required statement for weak majorization and as a consequence for strong and directional. Considering A^t instead of A allows us to complete the proof for doubly stochastic and matrix majorization. □

We now turn to the doubly stochastic majorization order $\preceq^{ds}$, and rank-constraints. Consider the doubly stochastic matrix class majorization, so $A = BX$ where $X \in \Omega_n$ and the matrices A and B belong to two matrix classes $\mathcal{A}$ and $\mathcal{B}$.

Consider $A \preceq^{ds} B$, where B is a rank one matrix, so $B = uv^T$ where $u, v \in \mathbb{R}^n$ are both nonzero vectors. This implies that A is also of rank one. Moreover

$$BX = uv^T X = uz^T$$

where $z^T = v^T X$, i.e., $z = X^T v$. Since X^T is doubly stochastic, by the Hardy-Littlewood-Pólya theorem on majorization ([10]), this means that $z \preceq v$, so

z is majorized by v. This shows that the principal ideal $M_{u,v}$ generated by $B = uv^T$ is given by

$$M_{u,v} = \{A : A \preceq^{ds} B\} = \{uz^T : z \preceq v\}.$$

This is a polyhedron, in fact a polytope, in M_n. Note that

$$M_{u,v} = \text{ conv}\{uw^T : w \text{ is a permutation of } v\}.$$

Checking membership in M_{uv}, i.e., checking if $A \preceq^{ds} B$ for given A, can be done as follows

- First, check if each column of A is parallel to u and, if so, we get (unique) numbers α_j ($j \leq n$) such that the j'th column of A is $\alpha_j u$.
- Then, check if the majorization $(\alpha_1, \alpha_2, \ldots, \alpha_n) \preceq v$ holds.

Proposition 4.26. *For given A and B, where B has rank one, the majorization $A \preceq^{ds} B$ can be checked by the algorithm above in $O(mn+n\log n)$ steps. Moreover, for classes $\mathcal{A}$ and $\mathcal{B}$ of $n \times n$ matrices, where each matrix in $\mathcal{B}$ has rank one, the algorithm tests if $\mathcal{A} \trianglelefteq \mathcal{B}$ in $O(|\mathcal{A}||\mathcal{B}|(mn + n\log n))$ steps.*

Proof. This follows from the discussion above since all computations reduce to (i) for each of m columns we need n operations to check if two vectors are parallel, and (ii) checking majorizations $z \preceq v$ for vectors of length n, and this is done by sorting the entries in each vector and use the partial sum definition of majorization, which takes $O(mn + n\log n)$ steps. □

Example 4.27. Let

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix} = uv^T$$

where $u = (1, 2, 3)$ and $v = (1, 0, 0)$. Let $z = (1/2, 1/4, 1/4)$, so $z \preceq v$, then

$$A = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1 & 1/2 & 1/2 \\ 3/2 & 3/4 & 3/4 \end{bmatrix} = uz^T$$

lies in $M_{u,v}$. □

Now, consider the same approach, but applied to the situation where B has rank k, so $B = \sum_{i=1}^{k} u_i v_i^T$ for some nonzero vectors $u_i \in \mathbb{R}^m$, $v_i \in \mathbb{R}^n$. Then

$$A = BX = \sum_{i=1}^{k} u_i v_i^T X = \sum_{i=1}^{k} u_i z_i^T$$

where $X \in \Omega_n$, $z_i = Xv_i$ that is $z_i \preceq v_i$.

Therefore we may propose the following algorithm (not sufficient yet) of checking if $A \preceq^{ds} B$. Let $u_1, u_2, \ldots, u_k$ be k linear independent columns of B. Then we can find $z_1, z_2, \ldots, z_k$ such that $A = \sum_{i=1}^{k} u_i z_i^T$ and such selection of z_i is unique because each column of A is a linear combination of linear independent vectors. It follows that for any A we can determine z_i and then we can check if $z_i \preceq v_i$ for all i. Note that this condition may not be sufficient. If in addition there exists doubly stochastic matrix X such that $z_i = Xv_i$ for every i, then $A \preceq^{ds} B$.

Therefore we only have to find a way of determining if for k pairs of vectors z_i and v_i such that $z_i \preceq v_i$ there exists an overall doubly stochastic matrix X such that $z_i = Xv_i$. Different characterizations of whenever such a matrix exists are found in [5]. Moreover, related polytopes and connection to transportation polytopes are discussed in [6].

Finally, in this paper, we define some subclasses of the set Ω_n of $n \times n$ doubly stochastic matrices.

Let $1 \leq k \leq n$. Define $\Omega_n^{(k)}$ as the set of all $n \times n$ doubly stochastic matrices whose first k rows are equal. These are matrices of the form

$$D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}$$

where $D_2 \in M_{n-k,n}$ is row stochastic with column sum vector $c := c(D_2) \leq e$ (i.e., each column sum is at most 1), and each row in D_1 is $(1/k)(e - c)^t$. If $k = n$, the matrix D_2 is void, and $c := O$ (the zero vector). Each matrix in $\Omega_n^{(k)}$ has rank at most $n - k + 1$. The extreme cases are $\Omega_n^{(1)} = \Omega_n$ and $\Omega_n^{(n)} = \{(1/n)J_n\}$, and clearly $\Omega_n^{(n)} \subseteq \Omega_n^{(n-1)} \subseteq \cdots \subseteq \Omega_n^{(1)} = \Omega_n$. For $1 \leq k < n$ define the doubly stochastic matrix $D_n^{(k)} = ((1/k)J_k) \oplus I_{n-k}$, which is the direct sum of the $k \times k$ matrix with all entries being $1/k$ and the identity matrix I_{n-k}. Also, define $D_n^{(n)} = (1/n)J_n$. The following result is easy to prove, but gives insight in the matrix majorization order based on $\preceq^{ds}$.

Proposition 4.28. *Let $\mathcal{A}$ be a matrix class in M_n, and let $1 \le k \le n$. Then*

$$\{A \in M_n : A \preceq^{ds} D_n^{(k)}\} = \Omega_n^{(k)}.$$

Thus, $\mathcal{A} \trianglelefteq^{ds} \{D_n^{(k)}\}$ if and only if $\mathcal{A} \subseteq \Omega_n^{(k)}$.

Proof. Let $X \in \Omega_n$ and partition X similar to $D_n^{(k)}$. This gives

$$D_n^{(k)} X = \begin{bmatrix} (1/k)J_k & O \\ O & I_{n-k} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} (1/k)J_k X_{11} & (1/k)J_k X_{12} \\ X_{21} & X_{22} \end{bmatrix} = K$$

The matrix K is doubly stochastic, as the product of two doubly stochastic matrices, and its first k rows are equal, namely equal to the barycenter of the first k rows of X. Therefore $K \in \Omega_n^{(k)}$. Conversely, if $D \in \Omega_n^{(k)}$, so its first k rows are equal, then

$$D_n^{(k)} D = D.$$

Thus, we have shown that $D \preceq^{ds} D_n^{(k)}$ if and only if $D \in \Omega_n^{(k)}$, and the proposition follows. ∎

This previous result means that the matrix $D_n^{(k)}$ is a maximal element w.r.t. $\preceq^{ds}$ in the set $\Omega_n^{(k)}$.

References

[1] D. Blackwell, Equivalent comparisons of experiments, Ann. Math. Statistics 24, (1953), 265–272.

[2] R.A. Brualdi, *Combinatorial Matrix Classes*, Encyclopedia of Mathematics, Cambridge University Press, 2006.

[3] L. LeCam, Sufficiency and approximate sufficiency, Ann. Math. Statistics 35, Number 4 (1964), 1419–1455.

[4] G. Dahl, Pseudo-experiments and majorization, Research report, 11 (1984), Department of Mathematics, University of Oslo.

[5] G. Dahl, Matrix majorization, *Linear Algebra Appl.*, **288** (1999), 53–73.

[6] G. Dahl, Majorization polytopes, *Linear Algebra Appl.*, **297** (1999), 157–175.

[7] P.R. Halmos, L.J. Savage, Application of the Radon–Nikodym theorem to the theory of sufficient statistics. Ann. Math. Statistics 20, (1949). 225–241.

[8] G.H. Hardy, J.E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, (1934, 1952, 1978).

[9] S.-G. Hwang and S.-S. Pyo, Matrix majorization via vector majorization. *Linear Algebra Appl.*, **332–334**, (2001), 15–21.

[10] A.W. Marshall and I. Olkin, B.C. Arnold *Inequalities: Theory of Majorization and Its Applications*, Second Edition, Springer, New York, 2011.

[11] F. D. Martínez Pería, Pedro G. Massey and Luis E. Silvestre, Weak matrix majorization. *Linear Algebra Appl.* , 403, p 343–368, (2005).

[12] S. Sherman, On a conjecture concerning doubly stochastic matrices, *Proc. Amer. Math. Soc.*, **3**, (1952), 511–513.

[13] E. Torgersen, Stochastic Orders and Comparison of Experiments, In: Stochastic Orders and Decision under Risk, IMS Lecture Notes - Monograph Series (1991)

[14] E. Torgersen, Comparison of Statistical Experiments, Encyclopedia of Mathematics and Its Applications, Vol. 36, Cambridge University Press, Cambridge, 1991.