Journal of Statistical Planning and Inference 17 (1987) 159–171
North-Holland

GROUP INVARIANT ORDERINGS AND EXPERIMENTAL DESIGNS

Alessandra GIOVAGNOLI

Università degli Studi – Perugia, 06100 Perugia, Italy

Friedrich PUKELSHEIM

Universität Augsburg, D-8900 Augsburg, Federal Republic of Germany

Henry P. WYNN

The City University, London, England

Received 9 September 1985; revised manuscript received 1 December 1986
Recommended by Ching-Shui Cheng

Abstract: Recent work by Giovagnoli and Wynn and by Eaton develops the theory of G-majorization with application to matrix orderings. Using this theory much of the work begun by Kiefer on 'universally' optimal designs of experiments can be better understood. The technique is to combine a group ordering (G-majorization) with another invariant ordering, such as the Loewner ordering, to define upper weak G-majorization on the information matrices of the experiments. Using an idea from previous work of Giovagnoli and Wynn combined with work by Pukelsheim and Styan on the matrix concavity of information matrices a general theory of weak G-majorization for linear models is developed which includes orderings for subsets of estimable functions.

AMS Subject Classification: 62K05, 62K10, 06F99.

Key words and phrases: Majorization; Orthogonal groups; Partial orderings; Optimal experimental design.

1. Introduction

Majorization is a fairly recent innovation to the theory of experimental design made roughly simultaneously by a number of authors, cf. Giovagnoli and Wynn (1980, 1981, 1985a), Hedayat (1981), and Bondar (1983). Much of the motivation of all this work was to develop and understand the seminal work of Kiefer (1975) on the 'universal' optimality of experimental design such as balanced incomplete block designs and Latin squares.

A full description of G-majorization, i.e. group majorization, with special emphasis on matrix majorization was given in Giovagnoli and Wynn (1985b) which we refer to as GW (1985b). Indeed the present paper is the promised second half of that work. The central idea of that paper is that a type of G-majorization for matrices combined with another ordering such as the Loewner orderings gives a very natural ordering on the information matrices of experiments. G-majorization is also discussed by Eaton (1984).

The idea of ordering experiments in some way is old and goes back at least to Blackwell (1951) and Le Cam (1964). Recent work by Torgersen (1985) discusses applications to linear models, see also Hansen and Torgersen (1974). Papers by Pukelsheim (1980, 1983, 1987) develop a duality theory for the Loewner ordering and for certain kinds of order preserving functionals, cf. our Section 3.

2. G-majorization and GW-majorization

We shall recall here essential results of G- and GW-majorization following GW (1985b) and Eaton (1984). Classical majorization is a particular case of what follows by taking G to be the group of permutation matrices, see Marshall and Olkin (1979).

Let G be a closed subgroup of the orthogonal group Orth(k) of $k\times k$ matrices Q for which $Q^{\mathrm{T}}Q$ is equal to the identity matrix. For $x, y\in\mathbf{R}^k$ we define

$$y <G\, x \quad\Leftrightarrow\quad y\in \operatorname{conv} Gx,$$

where conv denotes the convex hull and $Gx=\{Qx \mid Q\in G\}$ is the orbit of x under G. The relation $<G$ is called G-majorization and is a preordering in $\mathbf{R}^k$.

The ordering G will be combined with another (partial or pre-)ordering denoted by $<\cdot$ which is assumed to satisfy a single *compatibility condition*, namely

$$y<\cdot\, x \quad\Rightarrow\quad \sum_{i\in I}\alpha_i Q_i y <\cdot \sum_{i\in I}\alpha_i Q_i x,$$

for all finite sets of $\alpha_i\geq 0$ satisfying $\sum\alpha_i=1$, and of $Q_i\in G$. We thus define *lower weak G-majorization* $<_{\mathrm{GW}}$ by

$$y <_{\mathrm{GW}} x \quad\Leftrightarrow\quad y<\cdot\, v \text{ and } v <G\, x, \text{ for some } v\in\mathbf{R}^k.$$

Similarly *upper weak G-majorization* $<^{\mathrm{GW}}$ is defined by

$$y <^{\mathrm{GW}} x \quad\Leftrightarrow\quad v<\cdot\, y \text{ and } v <G\, x, \text{ for some } v\in\mathbf{R}^k.$$

An element x is minimal with respect to $<G$ if $y<G\,x$ implies $y\in Gx$. The minimal elements of $<G$ constitute the *centre* of the ordering. They actually coincide with the G-invariant elements of $\mathbf{R}^k$. For each $x\in\mathbf{R}^k$ there is a unique G-invariant $\bar{x}$ for which $\bar{x}<G\,x$, and more strongly: if $y<G\,x$ then $\bar{x}<G\,y$. The element $\bar{x}$ is simply the average of x with respect to the Haar probability measure on G.

Theorem 1 of GW (1985b) establishes the equivalence of (A) $y<G\,x$ and (B) $f(y)\leq f(x)$, for all G-invariant convex functions f. The major unresolved problem is to give (under suitable conditions) a statement similar to (B) equivalent to $<_{\mathrm{GW}}$ or $<^{\mathrm{GW}}$. Certainly if $y<_{\mathrm{GW}}x$ ($y<^{\mathrm{GW}}x$) then $f(y)\leq f(x)$ for all G-invariant, convex,

$<\cdot$-increasing (decreasing) functions f, but the converse statement remains unproved in general. It does hold when the majorization is the usual one and $<\cdot$ is the componentwise ordering (of the ordered entries), as well as in the following case:

Denote by $[x_0, x]$ the straight line between x_0 and x, i.e. $[x_0, x] = \mathrm{conv}\{x_0, x\}$. Fix some G-invariant vector x_0. Then the *ray ordering*

$$y \leq_{\mathrm{R}} x \quad \Leftrightarrow \quad y \in [x_0, x]$$

is G-compatible. In this ordering two vectors x and y are comparable if and only if they lie on a common ray emanating from x_0. An important special case is the contraction ordering

$$y \leq_{\mathrm{K}} x \quad \Leftrightarrow \quad y = \varrho x, \text{ for some } \varrho \in [0, 1]$$

obtained from $x_0 = 0$.

Theorem 1. *For $<\cdot = \leq_{\mathrm{R}}$ the following three statements are equivalent:*

(A) $y <_{\mathrm{GW}} x$ ($y <^{\mathrm{GW}} x$);

(B) $f(y) \leq f(x)$ *for all G-invariant, quasi-convex, $<\cdot$-increasing (decreasing) functions* f;

(C) $f(y) \leq f(x)$ *for all G-invariant, convex functions f which have a minimum in* x_0.

Proof. Theorem 3 of GW (1985b) shows that (A) implies (B). Next (B) implies (C) since the functions in (C) are a fortiori quasi-convex, and $<\cdot$-increasing: when $y <\cdot x$, i.e. $y = \varrho x_0 + (1-\varrho)x$, then $f(y) \leq \max\{f(x_0), f(x)\} = f(x)$.

Now let $\delta^*(z \mid C) = \sup_{x \in C} z'x$ be the support function of a convex set C. Assume property (C), fix a vector z, and define the function $f(x) = \delta^*(z \mid \mathrm{conv}\{x_0\} \cup Gx)$. By Lemma 3 of GW (1985b), f is G-invariant and convex, and f has a minimum in x_0. Applying property (C) to such functions f obtained from arbitrary vectors z we get $\delta^*(z \mid \mathrm{conv}\{x_0\} \cup Gy) \leq \delta^*(z \mid \mathrm{conv}\{x_0\} \cup Gx)$ for all z, i.e. $\mathrm{conv}\{x_0\} \cup Gy \subseteq \mathrm{conv}\{x_0\} \cup Gx$. In particular $y \in \mathrm{conv}\{x_0\} \cup Gx$, and this is the same as $y <_{\mathrm{GW}} z$. □

Theorem 1 actually remains true for the more general ordering

$$y \leq x \quad \Leftrightarrow \quad y \in \mathrm{conv}(S_0 \cup \{x\})$$

with some G-invariant set S_0. The proof is similar to the one given above and therefore omitted.

Next we show that a related result holds in a restricted version for $\bar{y}$ rather than y when $<\cdot$ is a *G-invariant cone ordering*, i.e. $y <\cdot x \Leftrightarrow x - y \in K$, where K is a closed convex cone in $\mathbf{R}^k$ which is G-invariant. Again we identify a large class (B) of order-preserving functions, of which a smaller subclass (C) is sufficient to characterize the ordering (A).

Theorem 2. *Assume $<\cdot$ is a G-invariant cone ordering. The following three state-*

ments are equivalent for every vector x and every centred vector $\bar{y}$:

(A) $\bar{y} <_{\mathrm{GW}} x$ ($\bar{y} <^{\mathrm{GW}} x$);

(B) $f(\bar{y}) \leq f(x)$ *for all G-invariant, quasi-convex,* $<\cdot$ *-increasing (decreasing) functions* f;

(C) $f(\bar{y}) \leq f(x)$ *for all G-invariant, linear,* $<\cdot$ *-increasing (decreasing) functions* f.

Proof. Assume (A). Then (B) follows from the properties of monotonicity, quasi-convexity, and invariance:

$$f(\bar{y}) \leq f(v) = f\left(\sum \alpha_i Q_i x\right) \leq \max f(Q_i x) = f(x).$$

Evidently (B) implies (C). It remains to show that (C) implies (A), and here the assumption that $<\cdot$ is a G-invariant cone ordering becomes essential.

Assume (C). Along with the order cone $K = \{x \in \mathbf{R}^k \mid x \cdot\!> 0\}$ we shall consider the dual cone $K^{\mathrm{dual}} = \{z \in \mathbf{R}^k \mid x^{\mathrm{T}} z \geq 0, \text{ for all } x \in K\}$. Let $\bar{y}$ be the centre of conv Gy, i.e. $\bar{y}$ is the average of the orbit Gy under the Haar probability measure. Being an average, the centring operation $y \to \bar{y}$ is a linear mapping. Now fix a vector z in the dual cone K^{dual}, and define the function $f(x) = \bar{x}^{\mathrm{T}} z$. This choice of f evidently is linear and G-invariant. It is also $<\cdot$ -increasing as we shall show next. Indeed, $y <\cdot\, x$ means $x - y \in K$. As the cone K is assumed to be invariant under the action of the group G, and as G contains linear transformations only, we obtain $Qx - Qy \in K$, for all $Q \in G$. Then also $\bar{x} - \bar{y} \in K$ since K is closed. Finally we obtain $f(x) - f(y) = (\bar{x} - \bar{y})^{\mathrm{T}} z \geq 0$.

Applying property (C) to all such functions f which are obtained from arbitrary vectors z in the dual cone K^{dual} we get $f(\bar{y}) = \bar{y}^{\mathrm{T}} z \leq f(x) = \bar{x}^{\mathrm{T}} z$, i.e. $(\bar{x} - \bar{y})^{\mathrm{T}} z \geq 0$, for all $z \in K^{\mathrm{dual}}$. In other words, $\bar{x} - \bar{y}$ lies in the second dual $(K^{\mathrm{dual}})^{\mathrm{dual}}$. But since dualisation is an idempotent operation for closed convex cones we have $(K^{\mathrm{dual}})^{\mathrm{dual}} = K$. Thus $\bar{y} <\cdot\, \bar{x}$. Since $\bar{x}$ lies in conv Gx we finally can conclude that $\bar{y} <_{\mathrm{GW}} x$. □

An important idea for this paper is that of *induced* GW orderings. Let ψ be a function defined from $\mathbf{R}^k$ to $\mathbf{R}^s$ which is concave with respect to an ordering $<\cdot$ defined on $\mathbf{R}^s$, namely:

$$\sum \alpha_i \psi(x_i) <\cdot\ \psi\left(\sum \alpha_i x_i\right).$$

Under suitable conditions weak majorization on $\mathbf{R}^k$ then *induces* weak majorization on $\mathbf{R}^s$. The proof of the following theorem parallels that of Theorem 4 of GW (1985b) and is therefore omitted. Note that we only require $G \subseteq \mathrm{GL}(k)$ rather than $G \subseteq \mathrm{Orth}(k)$.

Theorem 3. *Let a group* $G \subseteq \mathrm{GL}(k)$ *and a G-compatible ordering* $<\cdot$ *give* $<^{\mathrm{GW}}$ *on* $\mathbf{R}^k$. *Suppose* $^{-} : G \to \mathrm{GL}(s)$ *is a group homomorphism, and set* $\bar{G} = \{\bar{Q} \mid Q \in G\}$. *Let the induced group* $\bar{G} \subseteq \mathrm{GL}(s)$ *and a* $\bar{G}$*-compatible ordering which for simplicity we indicate again by* $<\cdot$ *give* $<^{\mathrm{GW}}$ *on* $\mathbf{R}^s$. *Then for all functions* $\psi : \mathbf{R}^k \to \mathbf{R}^s$ *which are*

G–$\bar{G}$-equivariant (i.e. $\psi(Qx)=\bar{Q}\psi(x)$) we have:

(i) *If ψ is convex (concave), then*

$$y \prec G\, x \quad\Rightarrow\quad \psi(y) \prec_{\mathrm{GW}} \psi(x) \ \ (\psi(y) \prec^{\mathrm{GW}} \psi(x)).$$

(ii) *If ψ is convex and increasing (decreasing), then*

$$y \prec_{\mathrm{GW}} x \ \ (y \prec^{\mathrm{GW}} x) \quad\Rightarrow\quad \psi(y) \prec_{\mathrm{GW}} \psi(x).$$

(iii) *If ψ is concave and increasing (decreasing), then*

$$y \prec^{\mathrm{GW}} x \ \ (y \prec_{\mathrm{GW}} x) \quad\Rightarrow\quad \psi(y) \prec^{\mathrm{GW}} \psi(x). \qquad \square$$

All the above results can be applied to the following version of matrix majorization. Let $G \subseteq \mathrm{Orth}(k)$ act on $k\times k$ real matrices by congruence, i.e. $Q\in G$ carries $A\in\mathbf{R}^{k\times k}$ into QAQ^{T}. For matrices $A,B\in\mathbf{R}^{k\times k}$ thus G-majorization is defined by

$$A \prec G\, B \quad\Leftrightarrow\quad A\in \mathrm{conv}\{QBQ^{\mathrm{T}} \mid Q\in G\},$$

cf. Sections 3 and 4 of GW (1985b). The $<\cdot$ ordering can be any G-compatible ordering. When we restrict A and B to the symmetric matrices, the most common ones are

- the Loewner ordering $\le_{\mathrm{L}}$: $A\le_{\mathrm{L}} B$ if $B-A$ is nonnegative definite, and
- the 'contradiction ordering' $\le_{\mathrm{K}}$: $A=\varrho B$, for some $\varrho\in[0,1]$.

When A and B are symmetric and nonnegative definite then

$$A\le_{\mathrm{K}} B \quad\Rightarrow\quad A\le_{\mathrm{L}} B.$$

This shows that the contraction ordering is coarser than the Loewner ordering in that the former orders fewer pairs A,B than the latter.

Example. For $A=aa^{\mathrm{T}}$ and $B=bb^{\mathrm{T}}$ having rank 1 and for any $G\subseteq\mathrm{Orth}(k)$ and $<\cdot\; = \;\le_{\mathrm{L}}$:

(i) $A \prec G\, B \;\Rightarrow\; B\in GA$.

(ii) $A \prec^{\mathrm{GW}} B$ ($A \prec_{\mathrm{GW}} B$) $\;\Leftrightarrow\;$ $a=\varrho b$ for some $|\varrho|\ge 1$ ($|\varrho|\le 1$).

Of special importance for this paper is the matrix version of Theorem 3 which we state in the GL(k) case. Let $G\subseteq \mathrm{GL}(k)$, and let Sym($k$) denote the set of symmetric $k\times k$ matrices. Suppose $C:\mathrm{Sym}(k)\to\mathrm{Sym}(s)$ is a matrix function such that a congruence transformation of an argument induces a congruence transformation of the associated image; more precisely let $^{-}: G\to \mathrm{GL}(s)$ be a homomorphism inducing the subgroup $\bar{G}\subseteq\mathrm{GL}(s)$ such that C is G–$\bar{G}$-equivariant, i.e. $C(QAQ^{\mathrm{T}})=\bar{Q}C(A)\bar{Q}^{\mathrm{T}}$.

Theorem 4. *With respect to the Loewner ordering $<\cdot\; = \;\le_{\mathrm{L}}$ we have:*

(i) *If C is matrix-concave then*

$$A \prec G\, B \quad\Rightarrow\quad C(A) \prec^{\mathrm{GW}} C(B).$$

(ii) *If C is matrix-concave and Loewner-increasing then*

$$A <^{GW} B \quad \Rightarrow \quad C(A) <^{GW} C(B).$$

In both cases the second orderings are with respect to $\bar{G}$ *and* $\leq_L$ *on* Sym(s). □

If a matrix M is G-invariant and $G' \subseteq G$ is a subgroup then M is a fortiori G'-invariant. If M^* is $<\cdot$-minimal and $\rightarrow$ is an ordering coarser than $<\cdot$ (as for instance when $\rightarrow = \leq_K$ and $<\cdot = \leq_L$) then M^* is also $\rightarrow$-minimal.

The material in Section 4 of GW (1985b) provides a fair illustration of these interrelations. Moreover that paper emphasizes that there are quite a few ways of obtaining the famous Proposition 1 of Kiefer (1975) as a special case of the general considerations as set out above. The next section prepares the stage for a transition to the proper experimental design problem.

3. Information functionals

Although the theme of this paper is orderings, discussion is needed of the implications for individual functionals. We specialize now to matrices $\mathcal{M} \subseteq \mathrm{NND}(k)$ which we shall call 'information matrices'. In the context of design of experiments, or in more general problems, the emphasis is on minimizing or maximizing functionals $\Phi(M)$ over $M \in \mathcal{M} \subseteq \mathrm{NND}(k)$.

In some cases, as we shall see later, $<^{GW}$, for some G and $<\cdot$, is the natural ordering to consider for such matrices in that $M_1 <^{GW} M_2$ means that M_1 is better (has 'more' information and/or is more 'balanced') than M_2.

We shall consider only functionals which are order preserving with respect to $<^{GW}$, i.e. $M_1 <^{GW} M_2 \Rightarrow \Phi(M_1) \leq \Phi(M_2)$. An important subclass of such functions as we have seen are those which are (1) G-invariant, (2) convex, and (3) $<\cdot$-decreasing. We shall usually take $G \subseteq \mathrm{Orth}(k)$ and $<\cdot = \leq_L$, or $<\cdot = \leq_K$. As explained above it seems very difficult to characterize $<^{GW}$ in this way, namely to establish that if $\Phi(M_1) \leq \Phi(M_2)$, for all Φ satisfying (1), (2), and (3), then $M_1 <^{GW} M_2$. The matrix version of Theorem 2 is the strongest result we have in general. For the orthogonal group and $\leq_L$ we *can* make the reverse statement, the characterization following from Lemma 7 of GW (1985b), see also Karlin and Rinott (1981). In this case $M_1 <^{GW} M_2$ means $\lambda(M_1) <^{W} \lambda(M_2)$ where $\lambda(\cdot)$ is the vector of ordered eigenvalues and $<^{W}$ is ordinary upper weak majorization as in Marshall and Olkin (1979).

A word of explanation concerning the functionals Φ is due. It may seem more natural to consider G-invariant concave increasing Φ so that higher Φ is associated with more information. For example

$$\Phi_p(M) = (\mathrm{tr}\, M^p)^{1/p}, \quad p \leq 1,\ p \neq 0,$$

$$\Phi_0(M) = (\det M)^{1/k}, \qquad \Phi_{-\infty}(M) = \lambda_{\min}(M)$$

are such functionals. But by taking $(\Phi_p(M))^{-1}$ or $-\log \Phi_p(M)$ we have the property we need, and dealing with convex decreasing functions Φ is merely a notational accident.

A more important question is why use M, rather than consider M^{-1} (or M^+, the Moore–Penrose generalized inverse) and convex *increasing* functions of M^{-1} or M^+? We shall advance two reasons for this. Firstly, the rationale for *not* relying on M^{-1} is that (under suitable conditions)

$$M_1 \prec^{\mathrm{GW}} M_2 \quad \Rightarrow \quad M_1^+ \prec_{\mathrm{GW}} M_2^+,$$

see Lemma 9 of GW (1985b). Hence the $\prec^{\mathrm{GW}}$ ordering on the M matrices is coarser than the $\prec_{\mathrm{GW}}$ on the M^+ matrices. This justifies the use of the phrase 'weak universal optimality' in Kiefer and Wynn (1981) for an M^* such that $(M^*)^+$ is minimal with respect to the $\prec_{\mathrm{GW}}$-ordering on the set of M^+-matrices for $M \in \mathcal{M}$, which is a consequence of, but does not imply minimality in the $\prec^{\mathrm{GW}}$-ordering on $\mathcal{M}$ itself. The second reason is built on the passage from the grand $k \times k$ information matrices M to reduced $s \times s$ information matrices $f(M) = (K^{\mathrm{T}} M^- K)^{-1}$ which becomes essential in the experimental design setting, as opposed to the similar passage from the grand $k \times k$ dispersion matrices M^{-1} to the reduced $s \times s$ dispersion matrices $g(M) = K^{\mathrm{T}} M^- K$. Pukelsheim and Styan (1983) have given a derivation of the concavity of the information transformation f, and the convexity of the dispersion transformation g, for a set of M matrices which is sufficiently large for the experimental design problem. That derivation visibly demonstrates that f plays a more primary role than g, again favouring information matrices over dispersion matrices. The experimental design point of view is to be explained in greater detail in the next section.

4. G-majorization for the linear model

Let us take a linear regression model of the standard form

$$Y(x) = x^{\mathrm{T}} \beta + \varepsilon, \quad \beta \in \mathbf{R}^k.$$

Let $\mathscr{X} \subseteq \mathbf{R}^k$ be the design region and ξ the design. Then under the usual second order assumptions,

$$M(\xi) = \int x x^{\mathrm{T}} \, \mathrm{d}\xi$$

is the information matrix of ξ. Given a subgroup of $\mathrm{GL}(k)$ a first central assumption is *invariance of* $\mathscr{X}$, i.e.

$$Q(\mathscr{X}) = \mathscr{X}, \quad \text{for all } Q \in G.$$

Then we can think of a new measure, ξ_Q, obtained from ξ by acting with Q first. Thus G induces an action by congruence on the set of information matrices, namely

$$M(\xi_Q)=\int xx^{\mathrm{T}}\,\mathrm{d}\xi_Q=\int_{\mathscr{X}} Qx(Qx)^{\mathrm{T}}\,\mathrm{d}\xi,$$

$$M(\xi)\to QM(\xi)Q^{\mathrm{T}},\quad \text{for all } Q\in G.$$

The same result can also be obtained by looking at G as transformations on the parameter space $\mathbf{R}^k$. Then linearity of $M(\xi)$ in the design measure establishes the following.

Lemma. *G-majorization of the design measure induces* (*matrix*) *G-majorization for the information matrix.* □

If we order elements of $\mathscr{M}$, the set of information matrices, according to G-majorization or GW-majorization, an improvement in the design (in the sense of making it smaller with respect to G-majorization) will lead to an improved information matrix. For instance if $G=\mathrm{Perm}(k)$ is the group of $k\times k$ permutation matrices, a symmetric design will give an information matrix which is G-minimal. In most cases we shall be interested in some s-dimensional parameter system $K^{\mathrm{T}}\beta$ where K is some $k\times s$ matrix. The components of $K^{\mathrm{T}}\beta$ are estimable under a design ξ if and only if the range (column space) of $M(\xi)$ contains the range of K. We define the reduced information matrix for $K^{\mathrm{T}}\beta$ through

$$C(M)=\begin{cases}(K^{\mathrm{T}}M^{-}K)^{+} & \text{if range}(M)\supseteq\text{range}(K),\\ 0 & \text{otherwise.}\end{cases}$$

This matrix function is (a) nonnegative definite, (b) positively homogeneous, and (c) concave. Hence it can be shown to be matrix-increasing, i.e.

$$M_1\leq_{\mathrm{L}} M_2 \quad\Rightarrow\quad C(M_1)\leq_{\mathrm{L}} C(M_2);$$

see also Theorem 3 of Pukelsheim and Styan (1983).

Given a group G as above a second central assumption is *range invariance of* K, i.e.

$$Q(\text{range } K)\subseteq \text{range } K,\quad \text{for all } Q\in G.$$

This is the same kind of condition as in invariant hypothesis testing.

Range invariance of K implies that the map $Q\to\bar{Q}=K^{+}QK$ is a group homomorphism. For upon recalling that the inclusion $Q(\text{range } K)\subseteq\text{range } K$ is equivalent to the identity $KK^{+}QK=QK$, we evidently have $\overline{PQ}=K^{+}PQK=K^{+}PKK^{+}QK=\bar{P}\cdot\bar{Q}$. Hence the group G induces a group $\bar{G}$ of linear transformation on $\mathbf{R}^s$ such that for all $Q\in G$ we have that $QK=K\bar{Q}$ with $\bar{Q}=K^{+}QK\in\bar{G}$.

It also implies an action by congruence of $\bar{G}$ on the reduced information matrices, namely $C(QMQ^{\mathrm{T}})=\bar{Q}C(M)\bar{Q}^{\mathrm{T}}$, for all $Q\in G$. For upon realizing that $Q^{-1\mathrm{T}}M^{-}Q^{-1}$ is a generalized inverse of QMQ^{T} and utilizing the homomorphic properties of overbarring we obtain the chain of equalities

$$C(QMQ^{\mathrm{T}}) = (K^{\mathrm{T}}[QMQ^{\mathrm{T}}]^{-}K)^{+} = (K^{\mathrm{T}}Q^{-1\mathrm{T}}M^{-}Q^{-1}K)^{+} = (\overline{Q^{-1\mathrm{T}}K^{\mathrm{T}}M^{-}KQ^{-1}})^{+}$$
$$= (\overline{Q^{-1}})^{-1}(K^{\mathrm{T}}M^{-}K)^{+}(\overline{Q^{-1\mathrm{T}}})^{-1} = \bar{Q}C(M)\bar{Q}^{\mathrm{T}}.$$

In summary, range invariance of K turns $^{-}$ into a homomorphism from G to $\bar{G}$, and C is G–$\bar{G}$-equivariant.

By Theorem 4 in Section 2 the above mentioned assumptions of invariance of $\mathscr{X}$ and range invariance of K imply that

$$M_1 <^{\mathrm{GW}} M_2 \quad \Rightarrow \quad C(M_1) <^{\mathrm{GW}} C(M_2), \quad \text{for } M_1, M_2 \in \mathscr{M}.$$

Thus the result of the lemma given above extends to reduced information matrices: improving the design via G-majorization leads to an improvement for *all* the reduced information matrices.

We now turn to the optimality of a design ξ. Optimality for an estimable set of parameters is insured if the reduced information matrix $C(M(\xi))$ is minimal with respect to upper GW-majorization, for a fixed group $\bar{G}$ induced by G.

If the class of possible designs is such that the set of reduced information matrices contains the centre with respect to $<G$ with group $\bar{G}$, it is enough to look among those designs which have invariant information matrices, and compare the reduced matrices with respect to the other ordering, typically the Loewner ordering $\leq_{\mathrm{L}}$. The results in Proposition 2 and 3 of GW (1985b) are obtained in this way.

If the group $\bar{G}$ is too small, however, the procedure may not end in a minimal matrix. Consider for instance the case of quadratic regression

$$Y = \beta_0 + x\beta_1 + x^2\beta_2 + \varepsilon, \quad x \in [-1, 1].$$

The group G contains only two elements, identity and reflection about the origin. Comparison of the information matrices of symmetric designs $\bar{\xi}$ for polynomial regression, that is matrices of the type

$$\begin{pmatrix} 1 & 0 & \int x^2\,\mathrm{d}\bar{\xi} \\ 0 & \int x^2\,\mathrm{d}\bar{\xi} & 0 \\ \int x^2\,\mathrm{d}\bar{\xi} & 0 & \int x^4\,\mathrm{d}\bar{\xi} \end{pmatrix}$$

has then to be carried out by the Loewner ordering. The solution in this case is that the set of best ('admissible') designs which cannot be improved under $<^{\mathrm{GW}}$ consists of all designs which put masses $\frac{1}{2}\alpha$ at $x = \pm 1$ and $1-\alpha$ at $x=0$, for some $0 \leq \alpha \leq 1$. Thus while in this example the improvement procedure does not directly lead to an optimal design, it does simplify the problem drastically in that it ends in a one-dimensional class of designs with parameter α, which then may be analysed by a direct approach.

We conclude this section with some implications of the $<^{\mathrm{GW}}$-ordering for matrices in $\mathscr{M}$. Given $M \in \mathscr{M}$ we can think of $V = M^{+}$ as a general 'dispersion' matrix. More precisely, for an estimable set of parameters $K^{\mathrm{T}}\beta$ based on an experiment with n observations we have

$$\sigma^{-2} n \operatorname{Var}(K^{\mathrm{T}}\hat{\beta}) = K^{\mathrm{T}} M^{+} K = C(M)^{+},$$

where σ^2 is the error variance, and $\hat{\beta}$ is the least-squares estimator of β. Then range invariance of K entails the following.

Let $M_1, M_2 \in \mathcal{M}$ be different information matrices of $K^{\mathrm{T}}\beta$ under designs ξ_1, ξ_2, respectively. If $K^{\mathrm{T}}\beta$ is estimable under both designs then

$$M_1 <^{\mathrm{GW}} M_2 \quad \Rightarrow \quad \operatorname{Var}_1(K^{\mathrm{T}}\hat{\beta}) <_{\mathrm{GW}} \operatorname{Var}_2(K^{\mathrm{T}}\hat{\beta}). \tag{$*$}$$

Here Var_i indicates the dispersion matrix under the design ξ_i $(i = 1, 2)$, and the $<_{\mathrm{GW}}$-ordering is taken with respect to the group $\bar{G}$ induced by G. Statement $(*)$ follows by piecing together the implication

$$M_1 <^{\mathrm{GW}} M_2 \quad \Rightarrow \quad C(M_1) <^{\mathrm{GW}} \mathrm{C}(\mathrm{M}_2)$$

from Theorem 4, and the implication

$$C(M_1) <^{\mathrm{GW}} C(M_2) \quad \Rightarrow \quad C(M_1)^{+} <_{\mathrm{GW}} C(M_2)^{+}$$

from Lemma 9 in GW (1985b).

A further consequence is obtained by applying Lemmas 9 and 8 of GW (1985b) which yield

$$\begin{aligned} M_1 <^{\mathrm{GW}} M_2 \quad &\Rightarrow \quad \sup_{Q \in G} \operatorname{tr}(K^{\mathrm{T}} Q M_1^{+} Q^{\mathrm{T}} K) \le \sup_{Q \in G} \operatorname{tr}(K^{\mathrm{T}} Q M_2^{+} Q^{\mathrm{T}} K) \\ &\Rightarrow \quad \sup_{Q \in G} \operatorname{tr} \operatorname{Var}_1(K^{\mathrm{T}}(Q\hat{\beta})) \le \sup_{Q \in G} \operatorname{tr} \operatorname{Var}_2(K^{\mathrm{T}}(Q\hat{\beta})). \end{aligned}$$

This is particularly meaningful when K is just a vector, so that we are estimating contrasts of G-transformations of the parameters and comparing the maximum variances.

5. Improving block designs

In this section we apply the material on the general regression model in Section 4 to treatment block designs. Following Giovagnoli and Wynn (1981) we consider a weight matrix

$$N = \{w_{ij}\} \quad (i = 1, \ldots, v; j = 1, \ldots, b),$$

which gives the proporion w_{ij} of observations on treatment i in block j. In line with the 'continuous' theory of the last section we assume that $\min w_{ij} = 0$ and $\sum \sum w_{ij} = 1$. This can be interpreted as a measure on the set of possible locations (i, j). Under the usual additive model $y_{ijk} = \alpha_i + \beta_j + \varepsilon_{ijk}$ the full information matrix for estimation of all α_i and β_j can be written as

$$M(N) = \begin{bmatrix} \Delta_r & N \\ N^{\mathrm{T}} & \Delta_s \end{bmatrix},$$

where the treatment 'replication' vector r consists of the row sums of N, and the block 'size' vector s consists of the column sums, while Δ_x denotes the diagonal matrix with x on the diagonal. Take a 'large' group G_0 acting on the locations (i, j), called the relabeling group. This $G_0 = \text{Perm}(v) \times \text{Perm}(b)$ where $\text{Perm}(v)$ and $\text{Perm}(b)$ are the permutation groups on the treatments (i) and blocks (j), respectively. If a pair (R, S) with $R \in \text{Perm}(v)$ and $S \in \text{Perm}(b)$ is applied to the design represented by N, we obtain

$$N_Q = RNS^{\mathrm{T}}, \qquad M(N_Q) = QM(N)Q^{\mathrm{T}},$$

where Q is block-diagonal with top left block R and bottom right block S. Define G, then, as

$$G = \left\{ \begin{bmatrix} R & 0 \\ 0 & S \end{bmatrix} \middle| R \in \text{Perm}(v), S \in \text{Perm}(b) \right\}.$$

Now using the linearity of $M(N)$ in the measure N we have more generally that

$$N' = \sum_{i \in I} \alpha_i R_i N S_i^{\mathrm{T}} \quad \Rightarrow \quad M(N') <_G M(N).$$

I.e., G-majorization of the N matrices implies G-majorization of the M matrices with respect to the same group G which, however, acts through RNS^{T} on the N matrices and through QMQ^{T} on the M matrices.

We are now in a position to *induce* results for subsets of contrasts. Most obviously we can consider treatment contrasts of the form $\alpha_i - \bar{\alpha}$ by taking

$$K = \begin{bmatrix} I_{v \times v} - v^{-1} J_{v \times v} \\ 0_{b \times v} \end{bmatrix}.$$

It can be checked that K satisfies the range invariance condition of Section 4. The matrix $C(N) = (K^{\mathrm{T}} M(N)^{-} K)^{+}$ is the normalized version of the usual C-matrix, the information matrix for the estimation of treatment contrasts, and can be written explicitly as

$$C(N) = \Delta_r - N \Delta_s^{-1} N^{\mathrm{T}}$$

(assuming $s_j > 0$ for $j = 1, \ldots, b$). Moreover, $C(RNS^{\mathrm{T}}) = RC(N)R^{\mathrm{T}}$ so that $\bar{G}$ is just the permutation group $\text{Perm}(v)$ itself. Theorem 4 just leads to the following.

Theorem 5. *If* $N' = \sum_{i \in I} \alpha_i R_i N S_i^{\mathrm{T}}$ *then* $C(N') <^{\mathrm{GW}} C(N)$. □

An implication of this result is that $\lambda(C(N')) <^{\mathrm{w}} \lambda(C(N))$. By taking $R_i = I_{v \times v}$ and $S_i = I_{b \times b}$ in separate stages we obtain essentially Theorem 7 of Giovagnoli and Wynn (1981).

One can also consider subgroups of $\text{Perm}(v)$. One such subgroup is that which leaves the first treatment $(i = 1)$ fixed. This would arise when the treatment is a control. A suitable K matrix for comparison of treatments $(v > 1)$ with the control would be

$$\tilde{K}=\begin{bmatrix} -1_{v-1} & \vdots & I_{v-1} \\ & 0_{b\times v} & \end{bmatrix}.$$

The information matrix is

$$\tilde{C}(N)=(\tilde{K}^{\mathrm{T}}M(N)^{-}\tilde{K})^{-1}=\Delta_{\tilde{r}}-\tilde{N}\Delta_{s}^{-1}\tilde{N}^{\mathrm{T}},$$

where $\tilde{N}$ is obtained from N by omitting the first row, while $\tilde{r}$ is obtained from r by omitting the first element. Applying Theorem 5 again we see that leaving the control fixed (except for permutation between blocks) improvement is made by 'improvement' in the test treatments ($i>1$) alone. This is essentially the idea used in Giovagnoli and Wynn (1985a). The extension to multi-way layouts is straightforward. The range invariance condition for K must be checked carefully in any problem, for example if sets of contrasts from several factors are included.

Acknowledgements

Early versions of this work were given at the NSF Workshop on Efficient Data Collection in July 1984 at the University of California at Los Angeles. Cooperation on this work was also supported by a grant of the British Council, and by a grant of the Italian Ministry of Education. Thanks are due to the referee for valuable suggestions.

References

Blackwell, D. (1951). Comparison of experiments. In: J. Neyman, Ed., *Proc. Second Berkeley Symp. Math. Statist. Probab.* Univ. California Press, Berkeley, CA, 93–102.

Bondar, J.V. (1983). Universal optimality of experimental designs: definitions and a criterion. *Canad. J. Statist.* 11, 325–331.

Eaton, M.L. (1984). On group induced orderings, monotone functions and convolution theorems. In: Y.L. Tong, Ed., *Inequalities in Statistics and Probability*. IMS Lecture Notes Series, Vol. 5, 13–25.

Giovagnoli, A. and H.P. Wynn (1980). A majorization theorem for the C-matrices of binary designs. *J. Statist. Plann. Inference* 4, 145–154.

Giovagnoli, A. and H.P. Wynn (1981). Optimum continuous block designs. *Proc. Roy. Soc. London Ser. A* 377, 405–416.

Giovagnoli, A. and H.P. Wynn (1985a). Schur-optimal continuous block designs for treatments with a control. In: L.M. Le Cam and R.A. Olshen, Eds., *Proceedings of the Berkeley Conference in Honor of Jerzy Newman and Jack Kiefer, Vol. 1*. Wadsworth, Belmont, CA, 418–433.

Giovagnoli, A. and H.P. Wynn (1985b). G-majorization with applications to matrix orderings. *Linear Algebra Appl.* 67, 111–135.

Hansen, O.H. and E.N. Torgersen (1974). Comparison of linear normal experiments. *Ann. Statist.* 2, 367–373.

Hedayat, A. (1981). Study of optimality criteria in design of experiments. In: M. Csörgő, D.A. Dawson, J.N.K. Rao and A.K.Md.E. Saleh, Eds., *Statistics and Related Topics* (*Ottawa, Ont.* 1980). North-Holland, Amsterdam, 39–56.

Karlin, S. and Y. Rinott (1981). Entropy inequalities for classes of probability distributions II. The multivariate case. *Adv. in Appl. Probab.* 13, 325–351.

Kiefer, J.C. (1975). Constructions and optimality of generalized Youden designs. In: J.N. Srivastava, Ed., *A Survey of Statistical Design and Linear Models*. North-Holland, Amsterdam, 333–353.

Le Cam, L.M. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* 35, 1419–1455.

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

Pukelsheim, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* 4, 339–364.

Pukelsheim, F. (1983). On information functions and their polars. *J. Optim. Theory Appl.* 41, 533–546.

Pukelsheim, F. (1987). Information increasing orderings in experimental design theory. *Internat. Statist. Rev.* 55 (2).

Pukelsheim, F. and G.P.H. Styan (1983). Convexity and monotonicity properties of dispersion matrices of estimators in linear models. *Scand. J. Statist.* 10, 145–149.

Torgersen, E.N. (1985). Majorization and approximate majorization for families of measures. In: T. Calinski and W. Klonecki, Eds., *Linear Statistical Inference, Proceedings, International Conference, Poznán (Poland) 1984*. Lecture Notes in Statistics 35. Springer, Berlin–Heidelberg–New York.