Proc. Second International Tampere Conference in Statistics
(Tampere, Finland, 1-4 June 1987)
Tarmo Pukkila and Simo Puntanen, *Editors*

pp. 261 - 274

INVITED PAPER

Majorization orderings for linear regression designs

Friedrich PUKELSHEIM
University of Augsburg, F.R. Germany

Key words and phrases: Group majorization, invariant design problems, information increasing orderings, universal optimality, simultaneous optimality.

ABSTRACT

Classical vector majorization captures the idea of whether the entries of a vector are more nearly equal than those of another one. Much of experimental design theory revolves around the same idea, and is called "balance" here. In the present paper we outline the use of majorization techniques for a general concept of balancedness: Matrix majorization replaces vector majorization, and linear transformation groups which leave the design problem invariant take the place of the permutation group for vector majorization.

1. INTRODUCTION

Majorization has emerged as a powerful tool to describe the notion of *balancedness* in experimental design theory. Its possible usefulness was already alluded to by Kiefer (1974, p. 862). The papers of Giovagnoli & Wynn (1981) and Bondar (1983) study majorization properties of the vector of eigenvalues associated with the information matrices of the designs. Further work of Giovagnoli &

Wynn (1985a,b) suggests a transition from vector majorization to matrix majorization, as elaborated in Giovagnoli, Pukelsheim & Wynn (1987). Applications of this theory to the design problem has been surveyed by Pukelsheim (1987a,b). In the present paper we sketch the essential steps of this recent development.

In Section 2 we first identify the various levels of the design problem:

- designs ξ, i.e. discrete probability distributions on an experimental domain $\mathcal{X} \subset \mathbb{R}^k$,
- moment matrices $M(\xi)$, i.e. $k \times k$ nonnegative definite matrices depending on a design ξ,
- information matrices $C(M)$, i.e. $s \times s$ nonnegative definite matrices representing the information for the s-dimensional parameter system of interest, as a function of moment matrices M, and
- objective functions $\phi(C)$, i.e. real functions of the information matrices C with information-like properties.

Invariance of the design problem under a group $\mathcal{Q}$ of linear transformations has its impact on each of these levels, as discussed in Section 3. We wish to stress that the groups act on moment matrices and on information matrices by congruence, not by similarity.

In Section 4 we turn to the desired orderings for experimental designs. Majorization relative to the group $\mathcal{Q}$ leads to the notion of when one moment matrix is *more balanced* than another one. However, for design applications this has to be built up into a two-stage preordering also involving the Löwner ordering of nonnegative definite matrices. The resulting *information increasing ordering* $\gg$ generalizes Kiefer's (1975) notion of *universal optimality*, and is intimately related to *simultaneous optimality* relative to all objective functions which are invariant.

2. MAXIMIZING INFORMATION

Suppose the *experimental conditions* are given through some k-dimensional vector from a compact *experimental domain* $\mathcal{X} \subset \mathbb{R}^k$. Under experimental conditions $x \in \mathcal{X}$ we may draw a single real-valued observation

$$Y(x) = x'\theta + \sigma e,$$

where we assume that the error e has unit variance, and that the errors are uncorrelated between observations under different experimental conditions as well as between repeated observations under the same experimental conditions. A *design* ξ then is taken to be a discrete probability distribution on the experimental domain $\mathcal{X}$, determining allocation and proportion of the experimental conditions. However, the problem is not really one of dealing with probability distributions ξ themselves, but to study the behaviour of certain matrices associated with designs ξ.

2.1 Moment matrices

The mean vector parameter θ is of dimension k. Accordingly we associate with a design ξ its $k \times k$ *moment matrix*

$$M(\xi) = \int_{\mathcal{X}} xx' d\xi = \sum_{i=1}^{l} \xi(x_i) x_i x_i'.$$

The set of all moment matrices forms a convex compact subset of nonnegative definite matrices, due to the assumed compactness of the experimental domain $\mathcal{X}$. Formally we simply assume to start with a convex compact *feasible set* $\mathcal{M}$ of $k \times k$ nonnegative definite matrices. The primary choice for $\mathcal{M}$ certainly is the set of *all* moment matrices, but other choices are of interest. For instance, $\mathcal{M}$ may be a set of moment matrices obtained from designs with certain restrictions on their marginals, or with restrictions on the support, here $\mathcal{M}$ would be a proper subset of all moment matrices. Bayesian design problems call for a choice of $\mathcal{M}$ to be the set of all moment matrices shifted by a prior information matrix R, see Chaloner (1984, p. 286). Through the introduction of the set $\mathcal{M}$ we are in a position to allow for such cases and others. Optimality will thus always be meant relative to the set $\mathcal{M}$.

2.2 Information matrices

We shall assume that the parameter system of interest $K'\theta$ is given through the $k \times s$ matrix K of rank s. For a nonsingular moment matrix M the $s \times s$ *information matrix for* $K'\theta$ is defined to be

$$C(M) = (K'M^{-1}K)^{-1}.$$

In our earlier work (Pukelsheim 1980, p. 341) we have chosen to define the information matrix for $K'\theta$ to be 0 when $K'\theta$ is not identifiable. The resulting discontinuity is unnecessarily strong, as has transpired in recent work of Gaffke (1985b, p. 73) and Müller-Funk, Pukelsheim & Witting (1985, p. 23). Moreover, the old definition fails to measure identifiability of subsystems of $K'\theta$. A refined definition for singular moment matrices M is

$$C(M) = \lim_{\epsilon \downarrow 0} (K'(M + \epsilon I_k)^{-1}K)^{-1} = \min_L LML',$$

where the minimum is taken over all left inverses L of K, relative to the Löwner matrix ordering. Now identifiability holds if and only if the matrix $C(M)$ in the refined definition is nonsingular, and in this case $C(M)$ admits the simpler representation $(K'M^{-}K)^{-1}$ where M^{-} is an arbitrary generalized inverse of M. Thus identifiability is entirely encoded in the rank behaviour of the information matrix, under the refined definition.

2.3 Objective functions

Over the years a considerable amount of work has gone into exploring the frontiers of the class of functionals which may rightly serve as optimality criteria for the experimental design problem. The classical criteria of D-, A-, E-optimality have been embedded into the continuous class of p-means (Kiefer 1974, p. 865). Pukelsheim (1980) admits information functionals which are defined to be concave, positive, and homogeneous; Gaffke (1985a, p. 385; 1985b, p. 69) presents a subgradient theorem covering functionals which are concave and isotonic (under the Löwner ordering).

Such a bewildering variety of optimality criteria does not please every human mind. Yet it serves its purpose. For instance let us discuss uniform optimality of some moment matrix M, i.e. in the Löwner ordering – the 'usual' ordering between symmetric matrices – we have $M \geq A$ for all $A \in \mathcal{M}$. Then optimality is inherited by the huge class of isotonic (i.e. increasing) criterion functions, since evidently $\phi(M) \geq \phi(A)$ whenever ϕ is isotonic. Conversely, if M is ϕ-optimal for every function ϕ in the relatively small class $\phi(C) = z'Cz$, with $z \in \mathbb{R}^s$, then M is uniformly optimal (Pukelsheim 1980, p. 344).

More generally let $\gg$ be a partial ordering for information matrices. Then it is useful to know which class of objective functions is order preserving, and the bigger the class the better. On the other hand it may be helpful to identify a subclass of functionals as small as possible so that simultaneous optimality over the subclass implies optimality under the partial ordering $\gg$. Hence we study wide classes of criteria in order to move away from any single particular criterion in the direction of statistically more reasonable partial orderings $\gg$. Section 4 will illuminate this situation further, but first we must briefly digress into when a design problem is invariant.

3. INVARIANT DESIGN PROBLEMS

The groups which determine our majorization relations originate from the invariance properties of the design problem. Assume that a subgroup $\mathcal{Q}$ of the general linear group $GL(k)$ acts linearly on the experimental conditions x, i.e.

$$x \to Qx, \text{ with } Q \in \mathcal{Q} \subset GL(k).$$

For this to make sense we require that the experimental domain is invariant, i.e. $Q(\mathcal{X}) \subset \mathcal{X}$ for all $Q \in \mathcal{Q}$.

3.1 Moment congruence

The linear group action on the experimental conditions induces a *congruence action* on moment matrices:

$$M(\xi) = \int_{\mathcal{X}} xx'd\xi \longrightarrow \int_{\mathcal{X}} Qxx'Q'd\xi = QM(\xi)Q'$$

For this to make sense we require that the feasible set $\mathcal{M}$ is invariant, i.e. $Q\mathcal{M}Q' \subset \mathcal{M}$ for all $Q \in \mathcal{Q}$. This invariance property is clearly satisfied for instance when $\mathcal{M}$ is the set of all moment matrices because then $QM(\xi)Q' = M(\eta)$, say, with η being the distribution of Qx under ξ.

3.2 Information congruence

The final invariance property focuses on the parameter system of interest $K'\theta$, stipulating

$$Q(\text{range}K) = \text{range}K \qquad \text{for all } Q \in \mathcal{Q}.$$

This simply means that the linear hypothesis $K'\theta = 0$ is invariant (or that the group $\mathcal{Q}$ acts on the parameter space in such a way that the set of parameters of interest remains invariant as does the set of nuisance parameters).

Range invariance of K entails that for every $Q \in \mathcal{Q}$ the $s \times s$ matrix $\check{Q} = K^+QK$ is nonsingular, where K^+ is the Moore-Penrose inverse of K. Furthermore we have

$$QK = K\check{Q}, \quad \text{and} \quad C(QMQ') = \check{Q}C(M)\check{Q}'.$$

Since the set $\check{\mathcal{Q}} = \{\check{Q} \in GL(s) \mid Q \in \mathcal{Q}\}$ forms a subgroup of $GL(s)$ the latter property means that the transition C from moment matrices to information matrices is equivariant under the groups $\mathcal{Q}$ and $\check{\mathcal{Q}}$.

3.3 Invariant objective functions

Once on the level of information matrices the reduced group $\check{\mathcal{Q}}$ has been determined it is clear that for an optimality criterion ϕ to be invariant we require

$$\phi(\check{Q}C\check{Q}') = \phi(C) \quad \text{for all } \check{Q} \in \check{\mathcal{Q}}.$$

As the group $\check{\mathcal{Q}}$ becomes larger the class of invariant criteria will evidently become smaller. For instance for the trivial group $\check{\mathcal{Q}} = \{I_s\}$ all criteria are invariant, for the orthogonal group $\check{\mathcal{Q}} = \mathrm{Orth}(s)$ an invariant criterion $\phi(C)$ must be a function of the ordered eigenvalues of C, and for the group of unimodular linear transformations $\check{\mathcal{Q}} = \mathrm{Unim}(s)$ the only invariant information functional is $(\det C)^{1/s}$.

It ought to be acknowledged that invariance is **not** automatically built into any given design problem. The preceding exposition has dwelt on its mathematical prerequisites. Whether it is meaningful from a statistical point of view must be decided on the ground of the practical poblem in question.

4. INFORMATION INCREASING ORDERINGS

Group majorization underlies the idea that given a point A any other point M in the convex hull of the orbit of A is an average and as such is more balanced. For instance the group $\mathcal{Q}$ acts on the designs themselves through $\xi \to \xi \circ Q^{-1}$. Hence a design ξ is *more balanced* than η when

$$\xi = \sum \alpha_i \, \eta \circ Q_i^{-1},$$

for a finite number of transformations $Q_i \in \mathcal{Q}$, where $\min \alpha_i \geq 0$ and $\sum \alpha_i = 1$. For a finite group $\mathcal{Q} = \{Q_1, \cdots, Q_n\}$ of order n we may choose $\alpha_i = 1/n$ and average over *all* transformations Q_i to obtain a design ξ which is invariant. For an infinite compact group $\mathcal{Q}$ a similar averaging procedure is possible with respect to Haar probability measure, but the resulting invariant measure may no longer be discrete. For instance an equidistant design on the circle averaged over all rotations yields Lebesgue measure on the circle.

Fortunately we rarely work on the level of designs ξ. Rather we let a design ξ inherit its performance properties from the moment matrix $M(\xi)$, and the information matrix $C(M(\xi))$. As an example recall that a design ξ is uniformly optimal for $K'\theta$ when in the Löwner ordering $C(M(\xi)) \geq C(M(\eta))$ for all competing designs η. Thus orderings of moment matrices and of information matrices are of greater interest.

4.1 Balancedness among moment matrices

We shall call a moment matrix B *more balanced than* another moment matrix A when B lies in the convex hull of the orbit of A, i.e.

$$B = \sum \alpha_i Q_i A Q_i'$$

for a finite number of transformation $Q_i \in \mathcal{Q}$, where $\min \alpha_i \geq 0$ and $\sum \alpha_i = 1$. This is the usual concept of group majorization where B is considered "smaller" than A. In the design context B carries more information and therefore we reverse the majorization notation and consider B to be "larger" than A, from the information point of view.

When the group $\mathcal{Q}$ is compact then each moment matrix A has in the convex hull of its orbit a unique invariant and hence most balanced matrix $\overline{A}$, obtained from averaging with respect to Haar probability measure dQ according to

$$\overline{A} = \int_{\mathcal{Q}} QAQ' \, dQ.$$

Due to compactness and convexity the feasible set $\mathcal{M}$ must contain any such matrix $\overline{A}$. Hence the invariant matrix $\overline{A}$ may be obtained as the moment matrix of a design ξ, without necessitating the invariance of ξ! For instance in trigonometric regression (Pukelsheim 1980, p. 360) every uniform distribution on an arbitrary set of equidistant support points leads to the same moment matrix as Lebesgue measure on the circle which is the unique rotation invariant distribution.

An ordering which is always available for comparing information is the Löwner ordering among moment matrices. It is natural to try and combine these two concepts.

The combination appropriate for the design problem produces the *information increasing ordering*, as follows. A moment matrix M is called *at least as informative* as another moment matrix A, denoted by $M \gg A$, when M is larger in the Löwner ordering than some matrix B which is more balanced than A, i.e.

$$M \geq B \in \text{convex hull of the orbit of } A \text{ under } \mathcal{Q}, \qquad \text{for some } B.$$

The information increasing ordering is *transitive*, in that $M \gg A$ and $A \gg F$ imply $M \gg F$. When the group $\mathcal{Q}$ is compact the ordering is also *antisymmetric 'modulo $\mathcal{Q}$'*, i.e. $M \gg A$ and $A \gg M$ entail $M = QAQ'$ for some $Q \in \mathcal{Q}$ (rather than $M = A$).

Other combinations of the two orderings are feasible and are discussed by Giovagnoli, Pukelsheim & Wynn (1987). That the present combination is appropriate for the design problem becomes apparent as we continue the discussion on the information matrix level.

4.2 Balancedness among information matrices

An information matrix C is called *at least as informative* as another information matrix D, denoted by $C \gg D$, when C is larger in the Löwner ordering than some matrix E which is more balanced than D, i.e.

$$C \geq E \in \text{ convex hull of the orbit of } D \text{ under } \check{\mathcal{Q}}, \qquad \text{for some } E.$$

Notice that we do not insist that the intermediate matrix E lies in $C(\mathcal{M})$: On the level of moment matrices the intermediate matrix B automatically lies in the feasible set $\mathcal{M}$, but due to the lack of convexity of $C(\mathcal{M})$ we have no such knowledge about E.

Consistency of the information increasing ordering from the level of moment matrices to the level of information matrices is shown by the following.

Theorem 1 *(Giovagnoli, Pukelsheim & Wynn 1987, Thm. 4). If M is at least as informative as A then $C(M)$ is at least as informative as $C(A)$.*

Proof. If $M \geq B = \sum \alpha_i Q_i A Q_i'$ then monotonicity, concavity, and equivariance of C yield

$$\begin{aligned} C(M) &\geq C(B) = C(\textstyle\sum \alpha_i Q_i A Q_i') \\ &\geq \textstyle\sum \alpha_i C(Q_i A Q_i') = \sum \alpha_i \check{Q}_i C(A) \check{Q}_i' = E, \text{ say}. \end{aligned}$$

Evidently E lies in the convex hull of the orbit of $C(A)$. $\square$

An analysis of the proof shows that when B is more balanced than A it does not generally follow that $C(B)$ is more balanced than $C(A)$, while it does follow that $C(B)$ is more informative than $C(A)$. A similar remark pertains to the "opposite" two-stage ordering: When M is more balanced than some matrix B which is larger in the Löwner ordering than A then $C(M)$ is more informative than $C(A)$ rather than inheriting the "opposite" ordering property.

Using a notion first introduced by Kiefer (1975) we now define an information matrix C to be *universally optimal* when C is at least as informative as D for all competing information matrices D. When the group $\check{\mathcal{Q}}$ is compact then a given information matrix C may be averaged with respect to Haar probability measure $d\check{Q}$ to obtain

$$\overline{C} = \int_{\check{\mathcal{Q}}} \check{Q} C \check{Q}' \, d\check{Q}.$$

However, this matrix $\overline{C}$ need not be a feasible information matrix as the set $C(\mathcal{M})$ may fail to be convex. On the other hand we have the following result.

Theorem 2. *Suppose the group $\check{\mathcal{Q}}$ is compact and the information matrix C is invariant. Then C is universally optimal if and only if C is larger in the Löwner ordering than the invariant matrices $\overline{D}$ obtained from the competing information matrices D.*

Proof. If C is invariant and universally optimal then $C = \overline{C} \geq \overline{E} = \overline{D}$, for every competing information matrix D. Conversely, if $C \geq \overline{D}$ then obviously $C \gg D$, since $\overline{D}$ lies in the convex hull of the orbit of D. $\square$

As an example let $\check{\mathcal{Q}}$ be the permutation group. Then an invariant information matrix is completely symmetric, i.e. it has the form $\alpha\overline{J}+\beta(I-\overline{J})$ where $\overline{J}$ is the $s\times s$ matrix with all entries equal to $1/s$. Comparison in the Löwner ordering thus reduces the task to comparing the two eigenvalues α and β.

4.3 Invariant objective functions

If an information matrix C is at least as informative as D, then $\phi(C) \geq \phi(D)$, for every criterion function ϕ which is isotonic, concave, and invariant. This simply follows from a repetition of the steps used to establish Theorem 1. In many cases the subclass of linear criteria is sufficient to establish universal optimality:

Theorem 3 *(Giovagnoli, Pukelsheim & Wynn 1987, Thm. 2). Suppose the group $\check{\mathcal{Q}}$ is compact and the information matrix C is invariant. Then C is universally optimal if and only if C is ϕ-optimal simultaneously for all criteria ϕ which are linear, isotonic, and invariant.*

Proof. The functions $\phi(D) = z'\overline{D}z$ are linear, isotonic, and invariant. Hence $C \geq \overline{D}$, for all competing information matrices D, and we may invoke Theorem 2. □

Whether universal optimality under a noncompact group $\check{\mathcal{Q}}$ is equivalent to simultaneous optimality relative to an appropriate subclass of criteria remains an open question.

Acknowledgement. Sincere thanks go to James V. Bondar and the referees for their helpful remarks on the original draft of the paper.

REFERENCES

Bondar, J.V. (1983). Universal optimality of experimental designs: definitions and a criterion. *The Canadian Journal of Statistics*, 11, 325-331.

Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *The Annals of Statistics*, 12, 283-300.

Gaffke, N. (1985a). Directional derivatives of optimality criteria at singular matrices in convex design theory. *Statistics*, 16, 373-388.

Gaffke, N. (1985b). Singular information matrices, directional derivatives, and subgradients in optimal design theory. In: T. Calinski & W. Klonecki (Eds.) *Linear Statistical Inference, Proceedings, Poznan 1984*, pp. 61-77. Lecture Notes in Statistics Vol. 35. Springer, Berlin Heidelberg New York Tokyo.

Giovagnoli, A. & Wynn, H.P. (1981). Optimum continuous block designs. *Proceedings of the Royal Society London Series A*, 377, 405-416.

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

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

Giovagnoli, A., Pukelsheim, F. & Wynn, H.P. (1987). Group invariant orderings and experimental designs. *Journal of Statistical Planning and Inference*, forthcoming.

Kiefer, J.C. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, 2, 849-879.

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

Müller-Funk, U., Pukelsheim, F. & Witting, H. (1985). On the duality between locally optimal tests and optimal experimental designs. *Linear Algebra and Its Applications*, 67, 19-34.

Pukelsheim, F. (1980). On linear regression designs which maximize information. *Journal of Statistical Planning and Inference*, 4, 339-364.

Pukelsheim, F. (1987a). Information increasing orderings in experimental design theory. *International Statistical Review*, forthcoming.

Pukelsheim, F. (1987b). Ordering experimental designs. In: *Proceedings of the First World Congress of the Bernoulli Society, Tashkent, September 1986*, forthcoming. VNU Science Press, Utrecht.

Received 12 March 1987
Revised 1 July 1987

Institut für Mathematik
Universität Augsburg
Memminger Strasse 6
D-8900 Augsburg
F.R. Germany