# Information Increasing Orderings in Experimental Design Theory

**Friedrich Pukelsheim**

*Institut für Mathematik, Universität Augsburg, D-8900 Augsburg, Federal Republic of Germany*

**Summary**

**A survey is given on recent results to identify order relations for experimental designs which appropriately describe when one design is more informative than another one. The technique is to augment the usual Loewner ordering of information matrices through group majorization where the group is such that it reflects the symmetries inherent in the underlying problem. Information increasing orderings appear to be a helpful tool to systematically improve on a given design, they may also be used to motivate special criteria such as the classical determinant criterion, and they sometimes aid in identifying optimal designs or at least complete classes.**

*Key words:* Continuous and discrete design theory; *D*-, *A*-, *E*-optimality; General equivalence theory; Generalized means; Group majorization; Information functionals and their polars; Loewner ordering; Moment and information matrices; Simple block designs; Treatment and block relabelling group; Universal and simultaneous optimality.

## 1 Introduction

Experimental designs aim at providing ways and means for efficient data collection. To this end it is essential that we are able to decide whether one design is better than another one. Orthogonality or a high degree of symmetry are some features which have an immediate appeal; another possibility is to specify an optimality criterion and to compare two designs through the values which they achieve under this criterion.

Given a particular design for an experimental study the question arises whether we can do better, i.e. whether we can improve upon the given design in order to get closer to orthogonality, to obtain more symmetry, or to improve upon the value of the optimality criterion. A rule of thumb is that a design with more symmetry performs better. Information increasing orderings provide a means to make this idea more precise.

There also transpires some kind of reconciliation between the more aesthetic features of a design, such as orthogonality or symmetry, and the more formal approach through optimality criteria. Kiefer's (1975) notion of 'universal optimality' illustrates this point: complete symmetry of information matrices, i.e. equal on-diagonal elements and equal off-diagonal elements, appears side by side with optimality under a wide class of criteria. Universal optimality and simultaneous optimality, as introduced below, will throw more light on this point.

Information increasing orderings will be introduced in § 3. They are determined by a group under which the design problem remains invariant. In order to see how the group acts on the model parameterization and on the parameters of interest it is helpful to distinguish the various stages of the experimental design problem; this is done in § 2. Section 4 uses information increasing orderings to obtain far-reaching optimality properties of block designs, without taking any recourse to the general equivalence theory. The situation is not quite so favourable for quadratic regression over the experimental region $[-1, +1]$; here information increasing orderings lead to a complete class only. This example, and a brief review of the appropriate general equivalence theory, is presented in §§ 5 and 6. Section 7 concludes the paper with a survey of the relevant literature.

## 2 Continuous designs in the classical linear model

In the discrete design theory, a design $\xi_n$ for $n$ observations determines in an experimental domain $\mathscr{X}$ a finite number of points $x_i$ $(i = 1, \ldots, l)$ and assigns to these points weights $\xi_n(x_i)$ of the form $n_i/n$ which sum to 1. Then $\xi_n(x_i) = n_i/n$ directs the experimenter to make $n_i$ observations under experimental condition $x_i$, in a sample of size $n$.

In the continuous design theory, a design $\xi$ on $\mathscr{X}$ is taken to be a probability distribution with finite support. While its support points $x_i$ $(i = 1, \ldots, l)$, say, still determine a finite number of experimental conditions for experimentation, the weights $\xi(x_i)$ need not be rational and may attain any value between 0 and 1, specifying the proportion of observations under condition $x_i$. In general, then, a continuous design $\xi$ only provides an approximation to a discrete design $\xi_n$ which is realizable.

In order to be able to decide which of two given designs is better we must detail the underlying statistical model. As usual we shall assume a *classical linear model*

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

where $Y(x)$ is the observed yield, or response, under experimental condition $x$, and linearly decomposes into a fixed effects term and a random error term. The fixed effects term $f(x)'\beta$ depends on the $\mathbb{R}^k$-valued *regression function* $f$ which determines the effect of the experimental condition $x$ on the expected yield, while $\beta$ is a real $k$-dimensional vector of unknown parameters. A prime denotes transposition. The error term $e$ is random with zero mean and unit variance, scaled with an unknown factor $\sigma > 0$.

### 2.1 *Moment matrices*

In a classical linear model a natural measure for the performance of an experimental design $\xi$ is its *moment matrix*

$$M(\xi) = \int f(x)f(x)'\, d\xi = \sum_{i=1}^{l} \xi(x_i)f(x_i)f(x_i)'.$$

Fisher information of an exact design $\xi_n$ for the full parameter vector $\beta$ is equal to $(n/\sigma^2)M(\xi_n)$ provided $M(\xi_n)$ is nonsingular, where $n$ is the sample size and $\sigma^2$ is the model variance. The $k \times k$ matrix $M(\xi)$ is nonnegative-definite, and depends linearly on $\xi$.

The set of moment matrices obtained from *all* designs is well known to be convex, due to the passage from discrete to continuous designs, and compact. However, in many cases we are not interested in all designs, but may wish to prescribe the marginals or delimit the support. Thus let $\mathscr{M}$ be the set of moment matrices of those designs which in a given situation are taken to compete for optimality; we assume that the set $\mathscr{M}$ is convex and compact.

### 2.2 *Information matrices*

Often we are interested, not in the full parameter vector $\beta$, but in an $s$-dimensional subsystem $K'\beta$ where $K$ is a given $k \times s$ matrix of rank $s$. For instance, the full parameter vector $\beta$ may decompose into $s$ components of interest and $k-s$ nuisance parameters. The *information matrix* for $K'\beta$

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

is best explained using covariance matrices. The simple least-squares estimator for $\beta$ has covariance matrix $(\sigma^2/n)M^{-1}$, that is Fisher information and the covariance matrix are inverse to each other. Similarly, since the simple least-squares estimator for $K'\beta$ has covariance matrix $(\sigma^2/n)K'M^{-1}K$, the information matrix for $K'\beta$ is $(n/\sigma^2)J(M)$.

As is well known an optimal design $\xi$ for an $s$-dimensional parameter system may have a moment matrix $M(\xi)$ which is singular. It is easy to see why this may happen. At times we may increase the information on the parameters of interest at the cost of decreasing information on the nuisance parameters. In some instances this is pushed to a point where the nuisance parameters are no longer identifiable (estimable, testable) in which case $M(\xi)$ becomes singular. Of course, the parameters of interest always must stay identifiable.

More formally, a parameter set $K'\beta$ is identifiable under a design $\xi$ if and only if the moment matrix $M(\xi)$ is such that its range (column space) contains the range of $K$. Denoting by

$$\mathscr{A}(K) = \{A \in \text{NND}(k) \mid \text{range}\, A \supset \text{range}\, K\}$$

the set of all nonnegative-definite $k \times k$ matrices $A$ whose range contains the range of $K$, we extend the definition of $J$ according to

$$J(A) = \begin{cases} (K'A^-K)^{-1} & \text{for } A \in \mathscr{A}(K), \\ 0 & \text{for } A \notin \mathscr{A}(K). \end{cases}$$

Thus identifiability leads to a reduced $s \times s$ positive-definite matrix $(K'A^-K)^{-1}$, while nonidentifiability leads to 0. We shall assume that $\mathscr{A}(K)$ meets $\mathscr{M}$, that is the feasible set $\mathscr{M}$ contains at least one moment matrix $M$ under which $K'\beta$ is identifiable.

### 2.3 *Information functionals*

The most popular optimality criteria are $D$-, $A$- and $E$-optimality, given by

$$\begin{array}{ll} \det C & \text{(Determinant optimality)}, \\ \text{trace}\, C^{-1} & \text{(Average-variance optimality)}, \\ \lambda_{\min}(C) & \text{(Eigenvalue optimality)}. \end{array}$$

They correspond to the three particular cases $p = 0, -1, -\infty$ of the one-parameter family

$$\begin{array}{ll} \phi_0(C) = (\det C)^{1/s} & p = 0, \\ \phi_p(C) = (\text{trace}\, C^p/s)^{1/p} & 0 \neq p \leqslant 1, \\ \phi_{-\infty}(C) = \lambda_{\min}(C) & p = -\infty. \end{array}$$

This family constitutes the generalized means of order $p \in [-\infty, +1]$ of the eigenvalues of information matrices. When we inquire into which properties make these means appropiate for measuring information, we come up with (a) nonnegativity, (b) positive homogeneity, and (c) concavity. We shall call a real function $\phi$ on the set NND($s$) of nonnegative-definite matrices an *information functional* provided it satisfies properties (a), (b) and (c). The class of all information functionals will be denoted by $\Phi$. It may seem worthwhile to pause and comment on these properties in somewhat greater detail.

To be precise, property (a) requires an information functional $\phi$ to be nonnegative on the set $\mathrm{NND}(s)$ of nonnegative-definite matrices, and to be positive on its subset $\mathrm{PD}(s)$ of positive-definite matrices. This is in line with the fact that information is bounded from below by null, and that a positive-definite information matrix indicates positive information.

Property (b) is essential in view of the proportionality factor $n/\sigma^2$, since then

$$\phi((n/\sigma^2)J) = (n/\sigma^2)\phi(J).$$

Hence the factor $n/\sigma^2$, being common to all designs under question, does not aid in comparing any two given designs. Positive homogeneity thus passes on to information functionals the appealing feature of Fisher information of being additive on independent replications, and inversely proportional to the model variance. However, this also means that our considerations account neither for sample size $n$ nor for the model variance $\sigma^2$; all they do is to distribute the proportions of observations over the feasible experimental sites in $\mathscr{X}$.

The third property (c), concavity, reflects the natural requirement that information cannot be increased through interpolation. Lack of concavity is fatal if a functional is to serve as a measure of information.

In view of positive homogeneity, property (c) is the same as superadditivity, that is $\phi(C+D) \geqslant \phi(C) + \phi(D)$. If $C \geqslant D$ then superadditivity and nonnegativity give

$$\phi(C) = \phi(C - D + D) \geqslant \phi(C - D) + \phi(D) \geqslant \phi(D).$$

Therefore every information functional $\phi$ is increasing in the Loewner ordering.

All in all the defining properties (a), (b), (c) form a minimum set of requirements which optimality criteria for experimental designs ought to satisfy. On the other hand they are so weak that they result in abundancy of information functionals. The question is whether such a bewildering variety is desirable.

I think the situation is best compared with loss functions. Although squared error loss is the one used most commonly, other loss functions do exist, and awareness of other loss functions helps distinguish squared error loss. Some procedures may even be optimal with respect to a wide class of loss functions, delimited for instance by convexity or boundedness. Knowning such properties is reassuring, even though it may not become visible when the task is to solve a practical problem.

Quite similarly the discussion of general information functionals provides proper evidence of the consequences which the choice of a particular criterion entails. Two points are worth mentioning. First we may be able to identify designs which perform well under a wide class of criteria. For instance an information matrix is maximal in the Loewner ordering if and only if it is optimal with respect to *all* information functionals. This is further elaborated in the following section on invariance where we shall distinguish between universal optimality and simultaneous optimality.

Secondly we do obtain further insight that the traditional criterion of determinant optimality rightly plays such a distinguished role, quite similarly to squared error loss. The determinant criterion is distinguished through the general theory in that it is the only one which is self-polar (see § 5 for details), and that it is the last criterion to be left over when the class of information functionals is narrowed down through invariance as set out in § 3.

## 3 Information increasing orderings

Designs which show more symmetry tend to be more informative. The mathematical expression of symmetry is invariance under a suitable group of transformations. The essence of the argument is as follows. Suppose $\check{G}$ is a group of transformations $\check{g}$ acting on the $s \times s$ information matrices $C$. Assume that $\phi$ is an information functional which is $\check{G}$-invariant, that is $\phi(\check{g}C) = \phi(C)$. Then symmetrization increases information since

$$\phi(\sum \alpha_i \check{g}_i C) \geqslant \sum \alpha_i \phi(\check{g}_i C) = \phi(C)$$

whenever $\min \alpha_i \geqslant 0$ and $\sum \alpha_i = 1$. The inequality follows from concavity, and equality from invariance.

Unfortunately details of this are somewhat more laborious. First we start with a group $G$ acting on the experimental domain $\mathscr{X}$, and then we deal with the induced group $\tilde{G}$ which acts on the model parameterization, and $\check{G}$ which acts on the parameters of interest. Invariance considerations for estimation and testing problems require the same detour.

### 3.1 *Invariance*

The starting point is a group $G$ which acts on the experimental domain $\mathscr{X}$. Since the experimental conditions $x$ enter into the fixed effects through the $k$-dimensional regression function $f$, we assume that there is a group $\tilde{G}$ on $\mathbb{R}^k$ such that the actions translate according to

$$f(gx) = \tilde{g}f(x),$$

i.e. the regression function $f$ is $G$–$\tilde{G}$-equivariant. We are dealing with linear models, and hence our essential assumption is that the group $\tilde{G}$ is a group of linear transformations, i.e. a subgroup of the general linear group $\mathrm{GL}(k)$. It is convenient to denote the members of $\tilde{G}$ by $Q$ rather than by $\tilde{g}$ whence the action on $\mathbb{R}^k$ is $y \rightarrow Qy$ as usual. For moment matrices $M$ this induces the congruence transformation

$$M(\xi) \rightarrow \int Qf(x)(Qf(x))' \, d\xi(x) = QM(\xi)Q'.$$

Next we make sure that the quantities which enter into the optimal design problem remain invariant under the group $\tilde{G}$. Firstly we require that the set $\mathscr{M}$ of competing moment matrices is invariant, i.e.

$$Q\mathscr{M}Q' = \mathscr{M} \quad \text{for all } Q \in \tilde{G}.$$

When the parameter system of interest is $K'\beta$ we secondly demand that the range (column space) of $K$ is invariant, i.e.

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

This invariance property is well known from hypothesis testing. For since expected yield is $f(x)'\beta$, the action $f(x) \rightarrow Qf(x)$ on the regressors induces the action $\beta \rightarrow Q'\beta$ on the parameters. Thus a linear hypothesis $K'\beta = 0$ is invariant if and only if the null spaces of $K'$ and of $K'Q'$ are equal, that is $K$ and $QK$ have the same range.

Since we measure information through information matrices $J(M)$ we need to evaluate terms like $J(QMQ')$. Now range invariance of $K$ guarantees that for every $Q \in \tilde{G}$ there exists some $s \times s$ matrix $\check{Q}$ such that $QK = K\check{Q}$, namely $\check{Q} = (K'K)^{-1}K'QK$. The set

$$\check{G} = \{\check{Q} \in \mathrm{GL}(s) \mid Q \in \tilde{G}\}$$

forms a subgroup of $\mathrm{GL}(s)$. It is not hard to show that

$$J(QMQ') = \check{Q}J(M)\check{Q}' \quad \text{for all } Q \in \tilde{G}.$$

In other words: the congruence action of the group $\tilde{G}$ on the set of moment matrices induces a congruence action of the group $\check{G}$ on the set of information matrices, and the mapping $J$ is $\tilde{G}$–$\check{G}$-equivariant.

As a final invariance property we shall use information functionals $\phi$ which are invariant under $\check{G}$, that is

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

We now resume our original theme of studying the increase in information due to symmetrization.

### *3.2 Information preorderings*

Suppose that the moment matrix $M$ lies in the convex hull of the orbit $\tilde{G}A = \{QAQ' \mid Q \in \tilde{G}\}$ of another moment matrix $A$, that is

$$M = \sum \alpha_i Q_i A Q_i',$$

with $\min \alpha_i \geq 0$ and $\sum \alpha_i = 1$. We shall then say that $M$ is *more centered under $\tilde{G}$ than $A$*. Being reflexive and transitive this relation is a preordering, known as *group majorization*. By concavity of $J$ we get

$$J(M) = J(\sum \alpha_i Q_i A Q_i') \geq \sum \alpha_i \check{Q}_i J(A) \check{Q}_i'.$$

Now consider an information functional $\phi$ which is $\check{G}$-invariant. By monotonicity, concavity, and invariance of $\phi$ we finally obtain

$$\phi(J(M)) \geq \phi(\sum \alpha_i \check{Q}_i J(A) \check{Q}_i') \geq \sum \alpha_i \phi(\check{Q}_i J(A) \check{Q}_i') = \phi(J(A)).$$

This shows that group majorization increases $\phi$-information, for every $\check{G}$-invariant information functional $\phi$.

The strongest statistically meaningful information ordering is the Loewner ordering among moment matrices. Group majorization and Loewner ordering are complementary in that they never apply simultaneously, at least when all transformations $Q \in \tilde{G}$ are orthogonal: if $M = \sum \alpha_i Q_i A Q_i'$ and either $M \geq A$ or $A \geq M$ then, having identical traces, $M$ and $A$ must coincide. This suggests an amalgamation into a two-stage *information preordering*, denoted by $\gg$, as follows.

Given two moment matrices $M$ and $A$ we define $M$ to be *at least as informative under $\tilde{G}$ as $A$* if $M$ is larger in the Loewner ordering than some matrix $B$ which is more centered under $\tilde{G}$ than $A$. Formally:

$$M \gg A \Leftrightarrow M \geq B \in \operatorname{conv} \tilde{G}A \quad \text{for some } B \in \text{NND}(k).$$

The set of information matrices is equipped with the corresponding preordering $\gg$ relative to the induced group $\check{G}$.

If among moment matrices $M$ is at least as informative as $A$ then we have shown above that among information matrices $J(M)$ is at least as informative as $J(A)$, and that $\check{G}$-invariant information functionals preserve this order.

Heritability from one stage down to the next fails to hold for the relation of being more centered: if $M$ is more centered under $\tilde{G}$ than $A$ then it follows that $J(M)$ is, not more centered, but more informative under $\check{G}$ than $J(A)$. This indicates that the information preordering is more natural for the underlying problem than mere group majorization, even though it is slightly more involved.

### *3.3 Universal optimality versus simultaneous optimality*

Our next goal is to clarify the relation between maximizing the information preordering, and optimality simultaneously for all invariant information functionals. Kiefer's (1975) result is the first in this direction. From that paper we borrow the notion of *universal optimality,* but confine it here to mean maximization in the information preordering.

Thus a moment matrix $M \in \mathcal{M}$ is called universally optimal for $K'\beta$ in $\mathcal{M}$ if $J(M)$ is invariant and most informative under $\check{G}$, that is if

$$\check{Q}J(M)\check{Q}' = J(M) \quad \text{for all } \check{Q} \in \check{G},$$
$$J(M) \gg J(A) \quad \text{for all } A \in \mathcal{M}.$$

The situation becomes particularly transparent in the case when $\check{G}$ is a compact subgroup of the orthogonal group Orth$(s)$. Then every information matrix $C$ has in the convex hull of its orbit a unique invariant and hence most centered matrix $C$. Indeed, $\bar{C}$ may be obtained through the linear operation of centering with respect to Haar probability measure $d\check{Q}$,

$$\bar{C} = \int_{\check{G}} \check{Q}C\check{Q}'\, d\check{Q}.$$

For universal optimality it is then sufficient to study the restriction of the Loewner preordering to those information matrices which are invariant, and this often greatly facilitates the problem. In this case universal optimality coincides with *simultaneous optimality* with respect to all $\check{G}$-invariant information functionals; i.e. a moment matrix $M \in \mathcal{M}$ is universally optimal for $K'\beta$ in $\mathcal{M}$ if and only if $M$ has $\mathcal{M}$-maximal $\phi$-information for $K'\beta$, for every information functional $\phi$ which is $\check{G}$-invariant.

The concept of universal optimality appears to be more restrictive than the concept of simultaneous optimality, in that the latter more easily extends to groups which are neither compact nor amenable. An example of a nonamenable group is the group Unim$(s) = \{\check{Q} \in \mathrm{GL}(s) \mid \det \check{Q} = \pm 1\}$ of transformations which are unimodular, i.e. volume preserving.

Here is a list of some known cases. As the group $\check{G}$ grows, the class of invariant information functionals shrinks. The ultimate survivor is the determinant criterion.

(i) When $\check{G} = \{I_s\}$ is trivial, then all information functionals are invariant, and the information preordering coincides with the Loewner preordering.

(ii) When $\check{G} = $ Perm$(s)$ is the group of $s \times s$ permutation matrices, we are lead to Kiefer's original notion of universal optimality; no characterizations of the class of invariant information functionals nor of the information preordering is available.

(iii) When $\check{G} = $ Orth$(s)$ is the orthogonal group, then an information functional is invariant if and only if it is a function of the ordered eigenvalues, and the information preordering is upper weak majorization of the ordered eigenvalues.

(iv) When $\check{G} = $ Unim$(s)$ is the group of unimodular transformations, then the unique, up to positive proportionality, invariant information functional is the determinant criterion $\phi_0$.

(v) Little can be hoped for beyond this point; no information functional is invariant under the full general linear group $\mathrm{GL}(s)$.

In summary we see that the centering operation yields more informative designs. In the classical block design setting it in fact leads to optimal designs. Section 4 will illustrate this approach.

## 4 Optimality of simple block designs

Our model is the fixed effects two-way classification

$$Y_{ijk} = \alpha_i + \gamma_j + \sigma e_{ijk},$$

with treatment effects $\alpha = (\alpha_1, \ldots, \alpha_v)' \in \mathbb{R}^v$, and block effects $\gamma = (\gamma_1, \ldots, \gamma_b)' \in \mathbb{R}^b$. Thus the full parameter vector for the mean is

$$\beta = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} \in \mathbb{R}^{v+b}.$$

The regression function $f$ takes the values $f(i, j)$ which is a vector of dimension $v + b$ consisting of zeroes except for the $i$th of the first $v$ entries and the $j$th of the last $b$ entries both of which are unity. Designs $\xi$ are identified with $v \times b$ *weight matrices* $W$ with nonnegative entries summing to 1, an entry $w_{ij}$ giving the proportion of observations to be allocated with the $i$th treatment in the $j$th block. In other words $W = N/n$ is the continuous version of the incidence matrix $N$ for $n$ observations. The treatment replication vector $r = W1_b$ and the block-size vector $s = W'1_v$ will be called *treatment marginals* and *block marginals*, respectively.

### 4.1 Moment matrices

Writing $\Delta_r$ for the diagonal matrix with the vector $r$ down the diagonal, the moment matrix of $W$ turns out to be

$$M(W) = \begin{bmatrix} \Delta_r & W \\ W' & \Delta_s \end{bmatrix}.$$

We shall consider the sets of moment matrices $\mathcal{M}$, $\mathcal{M}(r, .)$, $\mathcal{M}(., s)$ obtained from all designs, all designs with treatment marginals $r$, and all designs with block marginals $s$, respectively.

Relabelling treatments or blocks should not matter. The *treatment relabelling group* is

$$\left\{ \begin{bmatrix} R & 0 \\ 0 & I_b \end{bmatrix} \,\middle|\, R \in \text{Perm}\,(v) \right\},$$

where Perm $(v)$ is the group of $v \times v$ permutation matrices. The group action is easily seen to be

$$QM(W)Q' = \begin{bmatrix} R & 0 \\ 0 & I_b \end{bmatrix} \begin{bmatrix} \Delta_r & W \\ W' & \Delta_s \end{bmatrix} \begin{bmatrix} R' & 0 \\ 0 & I_b \end{bmatrix} = \begin{bmatrix} \Delta_{Rr} & RW \\ (RW)' & \Delta_s \end{bmatrix} = M(RW).$$

Hence the set $\mathcal{M}$ is invariant. The action on the set of all weight matrices is multivariate majorization from the left.

Also of interest is the *block relabelling group*

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

Since $QM(W)Q' = M(WS')$ again $\mathcal{M}$ is invariant, and the action on the weight matrices is multivariate majorization from the right.

### 4.2 *Treatment contrasts*

First we concentrate on the symmetrized treatment contrasts $(\alpha_1 - \bar{\alpha}_., \ldots, \alpha_v - \bar{\alpha}_.)'$, that is $K'\beta$ with

$$K = \begin{bmatrix} K_v \\ 0 \end{bmatrix}, \quad K_v = I_v - \bar{J}_v,$$

where $\bar{J}_v$ is the $v \times v$ matrix with all entries equal to $1/v$.

For the treatment relabelling group we find $QK = KR$. Hence the induced group $\check{G}$ is the permutation group Perm $(v)$, and simultaneous optimality covers those information functionals which are permutationally invariant.

A given weight matrix $W$ may be improved upon through its centered version $\bar{J}_v W = \bar{1}_v s'$. Designs of the form $\bar{1}_v s'$ may be called *equi-replicated product designs* since the uniform treatment marginals $\bar{1}_v = (1/v)1_v$ indicate equal replication for each treatment, and since the joint distribution $\bar{1}_v s'$ on all treatment block combinations is the product of the treatment marginals and the block marginals.

The information matrix for the symmetrized treatment contrasts is often called the *C-matrix*, and equals

$$C(W) = J(M(W)) = \Delta_r - W\Delta_s^- W'.$$

In the special case of equi-replicated product designs $\bar{1}_v s'$ we obtain $C(\bar{1}_v s') = (1/v)K_v$, independently of the block marginals $s$. Altogether we have proved the following result.

RESULT 1. *The equi-replicated product designs have $\mathcal{M}$-maximal $\phi$-information for the symmetrized treatment contrasts, for all those information functionals $\phi$ which are permutationally invariant.*

For the block relabelling group we find that $QK = K$ whence the induced group $\check{G} = \{I_s\}$ is trivial. Here simultaneous optimality covers all information functionals, and hence coincides with optimality in the Loewner ordering. Given a weight matrix $W$ the centering operation $W\bar{J}_b = r\bar{1}_b'$ yields an improvement in the Loewner ordering of $C$-matrices, but stays in the class $\mathcal{M}(r, .)$ corresponding to designs with given treatment marginals $r$.

For the product designs $W = rs'$ the $C$-matrix is found to be $C(rs') = \Delta_r - rr'$, independently of $s$. As for identifiability we must demand that all components of the treatment marginals $r$ are positive. Hence we have shown Result 2.

RESULT 2. *Suppose $r \in \mathbb{R}^v$ is a vector of positive treatment marginals. Then the product designs with treatment marginals $r$ have $\mathcal{M}(r, .)$-maximal $\phi$-information for the symmetrized treatment contrasts, for all information functionals $\phi$.*

### 4.3 *A maximal parameter system*

Finally we discuss a parameter system of maximal dimension $v + b - 1$, namely $K'\beta$ with

$$K = \begin{bmatrix} K_v & \bar{J}_v \\ 0 & I_b \end{bmatrix}.$$

When $\tilde{G}$ is the treatment relabelling group we easily see that $QK = KQ$, whence the induced group $\check{G} = \tilde{G}$ does not differ from the original one. Again, a passage from $W$ to the centered version $\bar{J}_v W = \bar{1}_v s'$ brings improvement, and identifiability necessitates positive block marginals $s$. The information matrix is found to be

$$J(M(\bar{1}_v s')) = \begin{bmatrix} K_v/v & 0 \\ 0 & \Delta_s \end{bmatrix},$$

and hence improvement is restricted to the class $\mathcal{M}(\,.\,, s)$ of designs with given block marginals $s$. This yields Result 3.

RESULT 3. *Suppose $s \in \mathbb{R}^b$ is a vector of positive block marginals. Then the equi-replicated product design $\bar{1}_v s'$ has $\mathcal{M}(\,.\,, s)$-maximal $\phi$-information for the maximal parameter system given above, for all those information functionals $\phi$ which are invariant under the treatment relabelling group.*

A similar statement holds for the block relabelling group and is omitted. If as a direct sum we take the treatment block relabelling group

$$\left\{ \begin{bmatrix} R & 0 \\ 0 & S \end{bmatrix} \;\middle|\; R \in \mathrm{Perm}\,(v),\, S \in \mathrm{Perm}\,(b) \right\},$$

we obtain the final result.

RESULT 4. *The uniform design $\bar{1}_v \bar{1}_b'$ has $\mathcal{M}$-maximal $\phi$-information for the maximal parameter system given above, for all those information functionals $\phi$ which are invariant under the treatment block relabelling group.*

Results 1 and 2 and, restricted to the means $\phi_p$, also Results 3 and 4 may be derived from the general equivalence theory as well. The present approach would seem to involve less technicalities, besides being more powerful conceptually. However, it remains an open question whether it also extends to improve upon or establish optimality of incomplete block designs, i.e. designs with a restricted support set.

Comparing Results 1 and 2, or 3 and 4 it becomes obvious that a shrinking class of competing designs comes with a growing class of optimality criteria for simultaneous optimality. Anyway, when the experimental domain $\mathscr{X}$ fails to be finite one can no longer hope that centering is powerful enough to lead to optimality. The quadratic regression model will serve as an example.

## 5 Optimality of quadratic regression designs

Here we consider the regression function $f(x) = (1, x, x^2)'$ on the symmetric experimental domain $\mathscr{X} = [-1, +1]$, with underlying linear model

$$Y(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \sigma e.$$

All designs $\xi$ are taken to be feasible, i.e. the class $\mathcal{M}$ of moment matrices is as large as possible.

### *5.1 Symmetric three-point designs*

By symmetry we feel that an optimal design will place mass $1-\alpha$, say, at 0 and divide the remaining $\alpha$ mass equally on $\pm 1$. Hence define the *symmetric three-point design* $\xi_\alpha$ through

$$\xi_\alpha(-1) = \xi_\alpha(+1) = \alpha/2, \quad \xi_\alpha(0) = 1-\alpha, \quad \alpha \in [0, 1].$$

We shall now justify our feeling and explore its domain of validity.

### *5.2 Sign-change groups*

Let $G$ be the transformation group on $\mathscr{X} = [-1, +1]$ consisting of the identity and the sign change $gx = -x$. Let

$$\tilde{G} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

be the two-element group of linear transformations of $\mathbb{R}^3$ consisting of the identity, and of a sign change in the second component. Evidently the regression function $f$ is $G$–$\tilde{G}$-equivariant, and the class $\mathcal{M}$ of all moment matrices is $\tilde{G}$-invariant.

All subsets of components of the parameter vector $\beta$ are found to have a coefficient matrix $K$ which is range invariant. For instance consider the constant-linear case

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = K'\beta, \quad \text{with } K = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

The range of $K$ consists of all 3-dimensional vectors whose third component is 0. Range invariance means that a sign change of the second component should not change the range, as is indeed the case. Furthermore the induced group $\check{G}$ turns out to be

$$\check{G} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

Since

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} a & -b \\ -b & c \end{pmatrix},$$

the $\check{G}$-invariant information functionals $\phi$ are such that they remain the same under an off-diagonal sign change. A similar discussion pertains to the other parameter subsets and is omitted.

### *5.3 Design improvement*

A first improvement is made by passing from a given design $\xi(x)$ to the centered version $\bar{\xi}(x) = (\xi(x) + \xi(-x))/2$, with moment matrix

$$M(\bar{\xi}) = \begin{pmatrix} 1 & 0 & \alpha \\ 0 & \alpha & 0 \\ \alpha & 0 & \int x^4 \, d\xi \end{pmatrix}, \quad \alpha = \int x^2 \, d\xi.$$

Since $\int x^4 \, d\xi \leqslant \alpha$, any such matrix can be improved upon in the Loewner ordering by

$$M_\alpha = \begin{pmatrix} 1 & 0 & \alpha \\ 0 & \alpha & 0 \\ \alpha & 0 & \alpha \end{pmatrix}.$$

As it happens $M_\alpha$ is the moment matrix of the symmetric three-point design $\xi_\alpha$. Thus the intuitively appealing restriction to symmetric three-point designs is quite legitimate.

*For every design $\xi$ there exists a symmetric three-point design $\xi_\alpha$ which is at least as informative as $\xi$.*

When we are interested in orthogonally invariant criteria only, we may narrow down the class of designs even further. It is not hard to show that the three eigenvalues of the moment matrices $M_\alpha$ are increasing for $\alpha \in (0, \frac{2}{5})$, whence an optimal weight $\alpha$ must satisfy $\alpha \geqslant \frac{2}{5}$. The case $\alpha = 1$ is not feasible since none of the parameter subsets is identifiable under $\xi_1$. Of the remaining class of symmetric three-point designs $\{\xi_\alpha \mid \frac{2}{5} \leqslant \alpha < 1\}$ one can show that indeed each member appears as an optimal design, for some parameter subset and for some mean $\phi_p$.

In summary, quadratic regression provides an example for which invariance considerations lead to a considerable simplification of the problem, namely to the one-parameter subclass of all symmetric three-point designs. This reduces the problem to one of real analysis, namely to maximize the real concave function

$$g(\alpha) = \phi(J(M_\alpha)), \quad \alpha \in [0, 1].$$

Cubic regression already behaves somewhat differently, and one has to take recourse to the general equivalence theory. We append an outline of the theory in such generality as is needed here.

## 6 General equivalence theory

An optimality criterion other than $D$-optimality but of similar statistical import is the globally oriented $G$-optimality which calls for minimization of $\max_{x \in \mathscr{X}} f(x)' M(\xi)^{-1} f(x)$. Kiefer & Wolfowitz (1960) proved the Equivalence Theorem of the continuous theory, which says that the two criteria of $D$-optimality for the full parameter $\beta$, and of $G$-optimality lead to the same class of optimal designs. The result was preceded by the special case of polynomial regression where the $G$-optimal designs of Guest (1958) and the $D$-optimal designs of Hoel (1958) were observed to coincide. This came as a surprise to the people working in the field, or as Kiefer put it: 'In fact the startling coincidence is that these two people have the same first two initials (P.G.) and you can compute the odds of that!!'.

It is not longer the case that the general equivalence theory shows the equivalence of two criteria each of which being statistically appealing and important in its own right. Rather, the theory seeks to exhibit necessary and sufficient conditions for optimality which are easy to verify. We now outline some of the general results, for designs $\xi$ which maximize information for a parameter system $K'\beta$ when information is measured by an information functional $\phi$. Since for a parameter system $K'\beta$ the information matrix is $J(M)$, we must maximize the composite function $\phi(J(M(\xi)))$ when $\xi$ varies over a subset of designs $\Xi$ feasible for the problem under question. As in the block design setting, $\Xi$ may be the set of all designs, of all designs with given marginals, of all designs with prescribed support, etc.

### *6.1 The General Equivalence Theorem*

A first step consists in singling out the matrix part of the problem. To this end let $\mathcal{M}$ be the set of feasible moment matrices $M(\Xi)$. Assume that $\mathcal{M}$ is convex and compact, and that it intersects $\mathscr{A}(K)$. For a given information functional $\phi$ the matrix problem then reads:

$$\text{Maximize } \phi(J(M)) \quad \text{subject to } M \in \mathcal{M}.$$

An optimal solution to this problem is said to have *$\mathcal{M}$-maximal $\phi$-information for $K'\beta$.*

In order to characterize optimality we need the polar functional given by

$$\phi^0(D) = \inf_{C \in PD(s)} \operatorname{trace} CD/\phi(C).$$

Very little can be said about its statistical meaning except that it, too, is an information functional. More can be said in special cases. The mean $\phi_p$ has *polar function* proportional to $\phi_q$ where $p$ and $q$ are conjugate over $[-\infty, +1]$; that is $p+q=pq$. The unique self-polar member is the mean $\phi_0$ thus providing another distinctive view of determinant optimality. Optimality may now be characterized as follows.

General Equivalence Theorem. *Let $M \in \mathcal{M}$ be a moment matrix under which $K'\beta$ is identifiable, that is $M \in \mathcal{A}(K)$. Abbreviate the information matrix $J(M) = (K'M^-K)^{-1}$ by $C$. Then $M$ has $\mathcal{M}$-maximal $\phi$-information for $K'\beta$ if and only if*

(i) *there exists some matrix $D \in$ NND$(s)$ solving*

$$\phi(C)\phi^0(D) = \operatorname{trace} CD = 1,$$

*and*

(ii) *there exists some $k \times k$ matrix $G$ solving $MGM = M$ such that, upon setting $B = G'KCDCK'G$, we have*

$$\operatorname{trace} AB \leqslant 1 \quad \textit{for all } A \in \mathcal{M}.$$

The characterization of optimal moment matrices is thus split into two parts, according to the fact that the objective function is a composition of the functions $\phi$ and $J$. The first part is in terms of the $s \times s$ matrices $C$ and $D$, and involves the polar information functional $\phi^0$. In many cases condition (i) determines a unique and explicit solution $D$ and hence simplifies drastically. For instance, for the generalized means $\phi_p$ condition (i) has the unique solution $D = C^{p-1}/\operatorname{trace} C^p$.

Thus all the emphasis is on the second part of the theorem concerning the $k \times k$ matrices $G$, $B$ and $A$. The key feature is that the competing moment matrices $A$ enter condition (ii) *linearly*, while inversions and other more involved computations are required of the optimality candidate $M$ and the associated matrix $B$ only.

### *6.2 Corollaries*

Of the many consequences of the General Equivalence Theorem we mention but three. The first corollary characterizes uniform optimality of $M$, that is in the Loewner ordering of information matrices we have $J(M) \geqslant J(A)$, for all $A \in \mathcal{M}$. Again the competing moment matrices $A$ enter the condition linearly.

Corollary on uniform optimality. *Let $M \in \mathcal{M}$ be a moment matrix with maximal rank. Then $M$ is uniformly optimal for $K'\beta$ in $\mathcal{M}$ if and only if $K'M^-AM^-K \leqslant K'M^-K$, for all $A \in \mathcal{M}$.*

As pointed out before, uniform optimality is the same as simultaneous optimality with respect to all information functionals. Hence the corollary may be used to prove Result 2 of § 4.2, but the argument given there seems to be simpler.

The second corollary gives a sufficient condition for the existence of an optimal information matrix, based on the polar information functional $\phi^0$.

Corollary on existence. *Suppose the information functional $\phi$ is such that its polar functional $\phi^0$ is strictly isotonic. Then there exists a moment matrix $M \in \mathcal{M}$ which has $\mathcal{M}$-maximal $\phi$-information for $K'\beta$.*

As already mentioned the mean $\phi_p$ has polar functional proportional to $\phi_q$ where $p+q=pq$. Hence when $p<1$ then $\phi_p$ has a strictly isotonic polar functional, and there exists a moment matrix with $\mathcal{M}$-maximal $\phi_p$-information for $K'\beta$.

The polar function of $\phi_1$ is $\phi_{-\infty}$ and fails to be strictly isotonic. The quadratic regression model provides a simple example where no moment information matrix exists. For the full parameter vector $\beta$ we must maximize $\phi_1(M_\alpha)=\frac{1}{3}\operatorname{trace} M_\alpha=\frac{2}{3}\alpha$. The maximal value would be $\alpha=1$ except that $\beta$ is no longer identifiable under $M_1$. Therefore no $\phi_1$-optimal design for $\beta$ exists.

The third corollary shows that given a particular optimal moment matrix all other optimal moment matrices may be obtained as solutions of an inhomogeneous linear matrix equation.

COROLLARY ON MULTIPLICITY. *Suppose the information functional $\phi$ is strictly concave. Let $M \in \mathcal{M}$ be a moment matrix which has $\mathcal{M}$-maximal $\phi$-information for $K'\beta$, and let $G$ be some $k \times k$ matrix as stipulated in the General Equivalence Theorem. Then any other moment matrix $A \in \mathcal{M}$ also has $\mathcal{M}$-maximal $\phi$-information for $K'\beta$ if and only if $AG'K=K$.*

This corollary was actually used to establish the multiplicities stated in Results 1–4 of § 4, but again the straightforward argument given there seems to be more appealing. For the quadratic regression model uniqueness of the symmetric three-point designs $\xi_\alpha$ may be proved using the corollary, if one so desires.

### 6.3 *Conclusion*

In conclusion it may be appropriate to enumerate the various degrees of freedom obtained.

(a) We must decide on the underlying statistical model, i.e. prescribe the regression function $f$.
(b) We may choose the parameter system of interest, i.e. fix some $k \times s$ matrix $K$ of rank $s$.
(c) We need an optimality criterion, i.e. pick some information functional $\phi$.
(d) We must delimit the class of competing designs, i.e. decide on a convex and compact set $\mathcal{M}$ of moment matrices which are feasible.

Given a practical problem these points will certainly vary in their importance, but all of them are supported by the general theory.

Similar reservations apply to invariance considerations. *If* a problem shows symmetries and *if* the transformations of the induced group $\tilde{G}$ are linear *then* the information preordering will be a helpful tool. But, of course, a problem need not show any symmetries, or it may fail to lead to linear transformations.

Whether invariance applies to a particular problem is a matter of practical consideration as well. In the quadratic regression model the design variable $x$ may be an indicator of location, varying from the 'left end' $x=-1$ through the 'midpoint' $x=0$ towards the 'right end' $x=+1$; invariance under a sign change will then be a reasonable requirement. In another practical application $x$ may be time. Then a sign change would mean exchanging 'yesterday' and 'tomorrow' which is absurd.

## 7 Notes and remarks

The preceding sections leave out a number of important aspects: combinatorial designs, computational algorithms, sequential experimentation, Bayesian designs, practical considerations. The reviews of Atkinson (1982) and Steinberg & Hunter (1984) provide information on these points.

*Section* 2. The set-up is standard, see Kiefer (1959), Silvey (1980). The continuous theory, as the name indicates, emphasizes the theoretical part of the problem; for more practical considerations, see Box (1982), Hahn (1984). The import and role of reduced information is apparent from Kiefer (1974a), see also § 16.E.7 of Marshall & Olkin (1979). The extension to singular information matrices which retain identifiability is given by Pukelsheim & Styan (1983); see also Gaffke & Krafft (1979). The means of order $p$ are introduced on page 865 of Kiefer (1974a). A short proof of the concavity properties of these means is given by Gaffke & Krafft (1979). Information functionals were introduced by Pukelsheim (1980). The class of all information functionals is closed under forming averages and finite pointwise minima; its richness is visualized through convex sets by Pukelsheim (1983b). Relations of classical criteria with testing and estimation problems are discussed in the recent work of Fedorov & Khabarov (1986).

*Section* 3. Invariance permeates the subject from its very beginning (Kiefer, 1959). Our presentation has been greatly influenced by Giovagnoli & Wynn (1985b), Giovagnoli, Pukelsheim & Wynn (1986), and by personal discussions with H.P. Wynn. The arrangement of starting with a group $G$ and inducing the groups $\bar{G}$ and $\check{G}$ is novel, as is the discussion of universal versus simultaneous optimality. The relevance of the Loewner ordering for the present problem follows from Theorem 3.1 of Kiefer (1959); it also fits into the wider context of comparison of experiments as shown by Hansen & Torgersen (1974). The notion of universal optimality was coined by Kiefer (1975).

An example of a criterion which is permutationally invariant, but fails to be orthogonally invariant is $\phi(C) = \min C_{ii}$, as already listed as (d) by Kiefer (1960, p. 383). An instant where authors on purpose choose a criterion which fails to be invariant is given by Conlisk & Watts (1979).

That majorization relative to the orthogonal group leads to upper weak majorization of the ordered eigenvalues is proved by Karlin & Rinott (1981). Bondar & Milnes (1981, p. 115) mention that the special linear group is not amenable, i.e. that it does not have an invariant mean. The relevance of this group for $D$-optimality is clear from Kiefer (1959, § 2.E). Other considerations concerning orderings are put forward by Gaffke (1981).

*Section* 4. A discussion of simple block designs in the continuous theory was initiated by Kurotschka (1971); see also Krafft (1978). The investigations were continued independently by Giovagnoli & Wynn (1981, 1985a), and by Pukelsheim (1983a). The latter paper contains Results 1 and 2, while Results 3 and 4 are new. An instant of a trade-off between information on the parameters of interest versus information on the nuisance parameters is given by Pukelsheim & Titterington (1986).

*Section* 5. The example of quadratic regression has been discussed at various places in the literature (Galil & Kiefer, 1977; Studden, 1980; Pukelsheim, 1980; Preitschopf & Pukelsheim, 1986).

*Section* 6. The optimal design problem is here treated as a problem of maximizing information rather than minimizing risk, a similar point of view is taken by Silvey (1980). A fairly detailed presentation of the state of the art is given by Pazman (1986). The information point of view leads to a consistency of exposition which the present author failed to reach otherwise. The approach of minimizing a convex functional is employed in the recent report of Bandemer, Näther & Pilz (1986) who give a detailed review of Bayesian experimental design and experimental design in case of correlated observations.

On page 113 of Farrell, Kiefer & Walbran (1967) $G$-optimality is read as global optimality. The original Equivalence Theorem dates back to Kiefer & Wolfowitz (1960). The quotation on the initials 'P.G.' is from Kiefer (1974b). The General Equivalence Theorem, as presented here, is taken from Pukelsheim (1980); various proofs based on convex analysis are given by Pukelsheim & Titterington (1983). The corollary on existence using polar functionals is adapted from Müller-Funk, Pukelsheim & Witting (1985). The corollaries on uniform optimality and multiplicity are presented by Pukelsheim (1980). That paper also discusses cardinality, location, and weights of the support of an optimal design $\xi$. Special criteria such as the means $\phi_p$, or the linear functionals $\phi_L(C) = \text{trace}\, CL$ are also treated as are simultaneous optimality with respect to all $p$-means, admissibility, and bounds for the optimal information.

## References


Atkinson, A.C. (1982). Developments in the design of experiments. *Int. Statist. Rev.* **50,** 161–177.

Bandemer, H., Näther, W. & Pilz, J. (1986). Once more: Optimal experimental design for regression models. *Statistics.* To appear.

Bondar, J.V. & Milnes, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahr. verw. Geb.* **57,** 103–128.

Box, G.E.P. (1982). Choice of response surface design and alphabetic optimality. *Utilitas Math.* **21B,** 11–55.

Conlisk, J. & Watts, H. (1979). A model for optimizing experimental designs for estimating response surfaces. *J. Econometrics* **11,** 27–42.

Farrell, R.H., Kiefer, J.C. & Walbran, A. (1967). Optimum multivariate designs. *Proc. 5th Berkeley Symp.* **1,** 113–138.

Fedorov, V. & Khabarov, V. (1986). Duality of optimal designs for model discriminating and parameter estimation problems. *Biometrika* **73,** 183–190.

Gaffke, N. (1981). Some classes of optimality criteria and optimal designs for complete two-way layouts. *Ann. Statist.* **9,** 893–898.

Gaffke, N. & Krafft, O. (1979). Matrix inequalities in the Loewner ordering. In *Modern Applied Mathematics: Optimization and Operations Research,* Ed. B. Korte, pp. 595–622. Amsterdam: North-Holland.

Galil, Z. & Kiefer, J.C. (1977). Comparison of rotatable designs for regression on balls, I (quadratic). *J. Statist. Plan. Infer.* **1,** 27–40.

Giovagnoli, A. & Wynn, H.P. (1981). Optimum continuous block designs. *Proc. R. Soc. Lond.* A **377,** 405–416.

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

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

Giovagnoli, A., Pukelsheim, F. & Wynn, H.P. (1986). Group invariant orderings and experimental designs. *J. Statist. Plan. Infer.* To appear.

Guest, P.G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist.* **29,** 294–299.

Hahn, G.J. (1984). Experimental design in the complex world. *Technometrics* **26,** 19–31.

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

Hoel, P.G. (1958). Efficiency problems in polynomial estimation. *Ann. Math. Statist.* **29,** 1134–1145.

Karlin, S. & Rinott, Y. (1981). Entropy inequality for classes of probability distributions II. The multivariate case. *Adv. Appl. Prob.* **13,** 325–351.

Kiefer, J.C. (1959). Optimum experimental designs. *J.R. Statist. Soc.* B **21,** 272–319.

Kiefer, J.C. (1960). Optimum experimental designs V, with applications to systematic and rotabable designs. *Proc. 4th Berkeley Symp.* **1,** 381–405.

Kiefer, J.C. (1974a). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2,** 849–879.

Kiefer, J.C. (1974b). Lectures on Design Theory. Mimeograph Series No. 397, Department of Statistics, Purdue University.

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

Kiefer, J.C. & Wolfowitz, J. (1960). The equivalence of two extremum problems. *Can. J. Math.* **12,** 363–366.

Krafft, O. (1978). *Lineare statistische Modelle und optimale Versuchspläne.* Göttingen: Vandenhoeck & Ruprecht.

Kurotschka, V. (1971). Optimale Versuchspläne bei zweifach klassifizierten Beobachtungsmodellen. *Metrika* **17,** 215–232.

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

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

Pazman, A. (1986). *Foundations of Optimum Experimental Design*. Dordrecht: Reidel.

Preitschopf, F. & Pukelsheim, F. (1986). Optimal designs for quadratic regression. *J. Statist. Plan. Infer.* To appear.

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

Pukelsheim, F. (1983a). On optimality properties of simple block designs in the approximate design theory. *J. Statist. Plan. Infer.* **8**, 193–208.

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

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

Pukelsheim, F. & Titterington, D.M. (1983). General differential and Lagrangian theory for optimal experimental designs. *Ann. Statist.* **11**, 1060–1068.

Pukelsheim, F. & Titterington, D.M. (1986). Improving multi-way block designs at the cost of nuisance parameters. *Statist. Prob. Letters* **4**, 261–264.

Silvey, S.D. (1980). *Optimal Design*. London: Chapman and Hall.

Steinberg, D.M. & Hunter, W.G. (1984). Experimental design: review and comment. *Technometrics* **26**, 71–97.

Studden, W.J. (1980). $D_s$-optimal designs for polynomial regression using continued fractions. *Ann. Statist.* **8**, 1132–1141.

## Résumé

Cet article présente un résumé des résultats récents des relations d'ordre dans la théorie de planification d'expériences. Ces relations décrivent quand un plan d'expérience est plus informatif qu'un autre, et elles se composent d'une combinaison de l'ordre de Loewner pour les matrices informationelles, et d'une $G$-majorization où le groupe $G$ reflète les symmétries du problème. Ces ordres qui augmentent l'information peuvent être utilisés pour améliorer systématiquement un plan donné, pour distinguer certains critériums d'optimalité comme le critérium de déterminant, et pour trouver des plans optimaux ou au moins des classes complètes.