ON SCHUR OPTIMALITY

by

Gregory M. Constantine
Department of Mathematics
Indiana University
Bloomington, Indiana 47405

March, 1980

Research supported by the Air Force Office of Scientific Research under the grant AFOSR 76-3050.
Schur-optimality is defined (in the general setting of a linear model) as a generalization of the well-known D-, A-, and E-optimality criteria. Techniques to establish Schur-optimality are outlined, based chiefly on a process of averaging information matrices and on vector majorization. A design with a completely symmetric information matrix of maximal trace is shown to be Schur-optimal. A design with an information matrix of maximal trace and exactly two distinct nonzero eigenvalues is proved Schur-better than a large class of designs. One description of a subcollection of designs over which...
Schur-optimality holds is given only in terms of the diagonal elements of the information matrices. Consequences of this are then examined in the setting on block designs.
ON SCHUR OPTIMALITY

by

Gregory M. Constantine
Department of Mathematics
Indiana University
Bloomington, Indiana 47405
(In amintirea lui Grigore)

SUMMARY

Schur-optimality is defined (in the general setting of
a linear model) as a generalization of the well-known D-,
A- and E-optimality criteria. Techniques to establish
Schur-optimality are outlined, based chiefly on a process
of averaging information matrices and on vector majoriza-
tion. A design with a completely symmetric information
matrix of maximal trace is shown to be Schur-optimal.
A design with an information matrix of maximal trace and
exactly two distinct nonzero eigenvalues is proved Schur-
better than a large class of designs. One description
of a subcollection of designs over which Schur-optimality
holds is given only in terms of the diagonal elements of
the information matrices. Consequences of this are then
examined in the setting of block designs.

Keywords: SCHUR-OPTIMALITY; SCHUR-CONVEXITY; D-,A-,E-OPTI-
MALITY CRITERIA; LINEAR MODEL; INCOMPLETE BLOCK DESIGNS;
MAJORIZATION; EIGENVALUES

AMS 1970 subject classifications: Primary 62K05; Secondary 62K10
1. INTRODUCTION

Oftentimes, in experiments, interest arises in estimating parameters with equal (or near equal) precision. A design balanced as much as possible for the parameters of interest is intuitively felt to be the right choice since no parameter ought to be left to disadvantage. But precision (formulated in terms of the covariance matrix of the estimates) turns out not to be related to the concept of balance neither in a direct way nor in an obvious one. While balance relates directly to the entries of the information matrix, precision is closely connected to its spectrum. In the present paper a connection between balance and precision is made with the help of the concept of Schur-optimality. Certain convex operations on the entries of the information matrix (as trends for balance) turn out to be compatible with the Schur-convex functions defined on the spectrum of the information matrix, as measures of precision. By compatibility we roughly mean that the more balanced a design tends to be, the closer to a minimum the values of the Schur-convex functions get. Our goal is to eventually find a design for which these Schur-convex functions reach their minima. As we define it, the concept of Schur-optimality is very strong and a design can only seldom be proved Schur-optimal over the collection of all possible designs. It is however quite convenient to show that whenever a design with
many desirable symmetries exist, it is better (in a very strong sense) than a large number of less symmetric designs. This is the main scope of the paper.

A design $d$ describes the way a certain statistical experiment is to be conducted. It generally specifies where the observations are to be taken and in what proportion at each location. The design is, within limits, subject to choice by the experimenter.

Let $\Omega$ be the collection of all possible designs. For $d \in \Omega$ we assume the following expected linear response:

$$E(Y) = X_d \theta$$

where $Y$ is the $m \times 1$ vector of uncorrelated observations with common variance $\sigma^2$, $X_d$ is the design matrix of dimension $n \times p$ and $\theta$ is a $p \times 1$ vector of unknown parameters. Usually statistical interest arises in estimating linear functions of $\theta_1$, a $v \times 1$ subvector of $\theta$. Then the reduced normal equations for $\theta_1$ can be written as

$$C_d \theta_1 = Q_d' Y$$

with $Q_d$ a $m \times v$ matrix and $C_d$ a $v \times v$ nonnegative definite matrix, called the information matrix of the design $d$ (for $\theta_1$). Unless otherwise specified, all the information matrices $C_d$ in the sequel will be with reference to the subvector $\theta_1$.

In the block design setting, for example, we are to compare $v$ varieties (labeled 1, 2, ..., $v$) via $b$ blocks
of size \( k \times v \). A design \( d \) in this case is a \( k \times b \) array with varieties as entries and blocks as columns. The collection of all designs is denoted by \( \Omega_{v,b,k} \). The usual additive model, under which these designs are considered, specifies the expectation on variety \( i \) in block \( j \) as \( \alpha_i + \beta_j \) where \( \alpha_i \) is the (unknown) effect of variety \( i \) and \( \beta_j \) is the (unknown) effect of the \( j^{th} \) block. Let 

\[
\theta = (\alpha_1, \ldots, \alpha_v, \beta_1, \ldots, \beta_b)',
\]

be the vector of unknown parameters. We are interested only in the subvector \( \alpha = (\alpha_1, \ldots, \alpha_v)' \) of variety effects. The information matrix for \( \alpha \), when the design \( d \) is used for estimation, is

\[
C_d = \text{diag}(r_{d1}, \ldots, r_{dv}) - \frac{1}{K} N_d N_d'
\]

where \( r_{di} \) is the number of replications of variety \( i \) in \( d \) and \( N=(n_{dij}) \), with \( n_{dij} \) signifying the number of times variety \( i \) occurs in block \( j \). The above information matrix is nonnegative definite with row sums zero, for all \( d \in \Omega_{v,b,k} \). The row sums being zero reflects the fact that only linear contrasts (i.e. functions \( t'\alpha \) with \( t'1 = 0 \), where \( 1 \approx (1, \ldots, 1)' \)) are estimable under any design \( d \). This is very often the case in discrete settings. The collection of information matrices \( C_d \), with \( d \in \Omega_{v,b,k} \), has therefore a common kernel generated by \( 1 \approx \).

Denote by \( u(C_d) \) the nondecreasingly ordered vector of eigenvalues of \( C_d \) outside an eventual common kernel (which
is of no help in distinguishing among designs. Let \( u(C_d) \) be a vector of length \( n(\leq v) \). In most relevant instances \( n \) is either \( v \) or \( v-1 \). Designs \( d \) for which \( u(C_d) \) has some zero entries are ruled out as bad designs, generally because they are disconnected. We shall therefore focus our attention to designs \( d \) for which \( u(C_d) \) has all its entries positive. In summary, to a design \( d \in \Omega \) we associate an information matrix \( C_d \) and the nondecreasingly ordered vector \( u(C_d) \) of its eigenvalues associated with the eigenvectors of \( C_d \) outside a kernel common to all \( C_d \) with \( d \in \Omega \), i.e.,

\[
\begin{align*}
\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\ qua
teria are all genuine instances of Schur-convex functions which are nonincreasing in their arguments. We then define the concept of a Schur-optimal design and, in section 3, outline general methods of establishing Schur-optimality. The techniques which we mention rely chiefly on convexity (averaging information matrices) and on vector majorization. In the last section of the paper we illustrate the material presented in section 3 by showing that a design with a completely symmetric matrix of maximal trace is Schur-optimal. This result very much resembles Proposition 1 of Kiefer (1975) on Universal optimality. There is no direct relationship between Schur-optimality and Universal optimality but the concepts are discussed comparatively in section 4. We then show that a design with an information matrix of maximal trace and exactly two distinct nonzero eigenvalues is Schur-optimal over classes of designs which satisfy a verifiable condition on the entries of their information matrices. For a subclass of designs a sufficient condition is conveniently formulated only in terms of the diagonal entries of the information matrices. In the block design setting corollaries are derived for binary designs with exactly two nonzero eigenvalues. Examples of such designs are the Partially Balanced Incomplete Block designs with two associate classes, as well as extended and abridged both Balanced Incomplete Block designs and Group Divisible designs.
2. DEFINITIONS

Let us recall the following:

A square matrix is called doubly stochastic if it has nonnegative entries with row and column sums equal to 1.

Let $I$ be an interval on the real line $\mathbb{R}$. A function $\dagger$ defined on $I^n$ with real values is called Schur-convex if

$$\dagger(Sx) \leq \dagger(x)$$

for all $x \in I^n$ and all $S$ doubly stochastic.

A real function $\dagger$ defined on $I^n$ is said to be nonincreasing in its arguments if it is a nonincreasing function when restricted to each of its arguments.

A function $F$ defined on $I^n$ is called symmetric if $F(Px) = F(x)$ for all $x \in I^n$ and all permutation matrices $P$.

A function $\phi$ defined on a convex set $A$ in $\mathbb{R}^n$ with real values is called convex if

$$\phi(\alpha x + (1-\alpha)y) \leq \alpha \phi(x) + (1-\alpha)\phi(y)$$

for all $x, y \in A$ and all $0 \leq \alpha \leq 1$.

By observing that a permutation matrix (and its inverse) is doubly stochastic one can see that a Schur-convex function is always symmetric. The converse is of course false, but the well-known result of Birkhoff and Von Neumann which states that the collection of doubly stochastic matrices is the convex span of permutation matrices provides us with many examples of Schur-convex functions. By this result it immediately follows that any symmetric and convex function is Schur-convex. Lastly, a Schur-convex function need not be convex, e.g.,

$$\dagger(x_1, x_2) = |x_1 - x_2|^{1/2}.$$
Let $b$ be a constant such that $u_{d_{i}} < b$ for all $1 \leq i \leq n$ and all $d \in \Omega$, e.g., $b$ can be the maximal trace of $C_{d}$ with $d \in \Omega$. The constant $b$ is always finite, either because $\Omega$ is finite or because of arguments involving compactness. Let $I = [0, b]$. Then $u(C_{d}) \in I^{n}$ for all $d \in \Omega$. So we can define $\hat{\Psi}(u(C_{d}))$ for $\hat{\Psi}$ Schur-convex.

For convenience we set by definition $\hat{\Psi}(C_{d}) = \hat{\Psi}(u(C_{d}))$ for $\hat{\Psi}$ Schur-convex and nonincreasing in its arguments. We are now ready to define Schur-optimality.

**Definition 2.1.** A design $d^{*}$ is said to be Schur-better than another design $d$ (notation $d^{*} \succ d$) if

$$\hat{\Psi}(C_{d^{*}}) \leq \hat{\Psi}(C_{d})$$

for all Schur-convex functions $\hat{\Psi}$ nonincreasing in their arguments.

**Definition 2.2.** A design $d^{*}$ in $\Omega$ is called Schur-optimal over $\Omega$ if $\hat{\Psi}(C_{d^{*}}) \leq \hat{\Psi}(C_{d})$

for all Schur-convex functions $\hat{\Psi}$ nonincreasing in their arguments and all designs $d$ in $\Omega$.

Letting $\hat{\Psi}(C_{d}) = -\log \sum_{i=1}^{n} u_{d_{i}}$, we obtain the well-known functions associated with $D$-, $A$-, and $E$-optimality criteria, respectively. The above functions are Schur-convex because they all are symmetric and convex.
\[ \Phi(c_d) = \sum_{i=1}^{n} f(u_{di}) \] with \( f \) convex nonincreasing and the \( \phi_1 \) criteria defined by Kiefer (1974) are also Schur-convex for the same reason. It is clear that all these functions are nonincreasing in their arguments. Note that the E-criterion is not a limiting case when formulated in terms of Schur-convexity.

Schur-optimality is a very strong criterion, as it implies all of the above. As the above examples indicate, it probably quite satisfactory to look at just symmetric and convex functions of eigenvalues, rather than Schur-convex. But the techniques that we shall outline readily apply to Schur-convex functions and this motivates the extension.

3. ON AVERAGING AND MAJORIZATION

The principal tool that we shall employ when searching for Schur-optimal designs is contained in Theorem 3.2. This theorem relies, in turn, on a fundamental result on majorization due to Hardy, Littlewood and Polya (1934), which was later extended by Ostrowski (1952).

Whenever we write \( x \preceq y \) for two vectors \( x = (x_1, x_2, \ldots, x_n) \) and \( y = (y_1, y_2, \ldots, y_n) \) in \( \mathbb{R}^n \) we assume that
\[ x_1 \leq x_2 \leq \cdots \leq x_n, \quad y_1 \geq y_2 \geq \cdots \geq y_n \]
and that \( \sum_{i=1}^{m} y_i \leq \sum_{i=1}^{m} x_i \) for all \( 1 \leq m \leq n \).

If \( x \preceq y \) we say that \( y \) majorizes \( x \).
The useful concept of majorization has been considered in the context of design optimality by Cheng (1979) to block designs with 4 varieties and an arbitrary number of blocks.

The following result can be easily derived from Ostrowski (1952):

**Theorem 3.1.** Let \( x, y \in \mathbb{R}^n \). If \( x \preceq y \) then \( \psi(x) \leq \psi(y) \) for all Schur-convex functions \( \psi \) nonincreasing in their arguments.

Denote by \( \sigma(A) \) the nondecreasingly ordered vector of eigenvalues of the matrix \( A \). The result we state next is essentially due to Ky Fan (1951) although he has not formulated it in this form.

**Lemma 3.1.** Let \( A_i (1 \leq i \leq m) \) be nonnegative definite matrices. Then
\[
\sigma( \sum a_i A_i ) \preceq \sum a_i \sigma(A_i)
\]
where \( 0 \leq a_i \leq 1 \) and \( \sum a_i = 1 \).

Before we proceed, we need to introduce some notation. For a \( v \times v \) matrix \( A \) and a permutation \( \sigma \) on the symbols \( 1, 2, \ldots, v \) we denote by \( A^\sigma \) the matrix obtained from \( A \) after performing the row and (same) column permutations as indicated by \( \sigma \). That is \( A^\sigma = P A P' \), where \( P \) is the \( v \times v \) matrix representation of \( \sigma \). Since \( A \) and \( A^\sigma \) are
similar matrices, their ordered spectra \( \sigma(A) \) and \( \sigma(A^\sigma) \) are identical for any permutation \( \sigma \). We say that \( A^\sigma \) is a conjugate of \( A \).

The following is an immediate but useful consequence (see also Magda (1979), Lemma 3.2.1).

**Proposition 3.1.** Any convex combination \( \sum a_i A^\sigma_i \) of conjugates of a nonnegative definite matrix \( A \) satisfies

\[
\sigma\left( \sum a_i A^\sigma_i \right) \leq \sigma(A)
\]

Now we can state and prove

**Theorem 3.2.** A design \( d^* \in \Omega \) is Schur-better than \( d \) (\( d^* \gg d \)) if

\[
\sigma(C_{d^*}) \leq \sigma\left( \sum a_i C_{d_i}^{\sigma_i} \right)
\]

for some convex combination of conjugates of \( C_d \).

**Proof.** Firstly, observe that by Theorem 3.1 \( \mu(C_{d^*}) \leq \mu(C_d) \) implies \( d^* \gg d \). Since the last \( n \) components of \( \mu(C_d) \) and \( \sigma(C_d) \) are the same and the first \( v-n \) components of \( \sigma(C_d) \) are zero (for all \( d \in \Omega \) \( \mu(C_{d^*}) \leq \mu(C_d) \) is equivalent with \( \sigma(C_{d^*}) \leq \sigma(C_d) \). So \( \sigma(C_{d^*}) \leq \sigma(C_d) \) implies \( d^* \gg d \).

Using the assumption and then Proposition 3.1 we have

\[
\sigma(C_{d^*}) \leq \sigma\left( \sum a_i C_{d_i}^{\sigma_i} \right) \leq \sum a_i \sigma(C_{d_i}^{\sigma_i}) = \sigma(C_d).
\]

The last equality is true because \( C_{d_i}^{\sigma_1} \) and \( C_d \) are conjugates and hence have the same spectrum. This concludes the proof.
Theorem 3.2 is helpful in the following way. Suppose that a design \( d^* \in \Omega \) with a lot of symmetries is believed optimal in some broad sense for the intuitive reasons mentioned in the introduction. It would be very satisfactory to show that for a large class of designs \( d \in \Omega \), \( d^* \) is Schur-better than \( d \) (i.e. \( d^* \triangleright d \)). Because of its balance \( d^* \) has an information matrix for which \( \sigma(C_{d^*}) \) can be computed. But \( \sigma(C_d) \) for an arbitrary \( d \) is impossible to calculate. It is very often possible, however, to compute

\[
\sigma(\frac{1}{m} \sum_{i=1}^{m} C_{d_i}^q)\]

for a convex combination of certain conjugates of \( C_d \). The entries in the convex combination tend to even out and this generally facilitates the computation of the spectrum. The spectrum of the convex combination is a helpful intermediary between \( \sigma(C_{d^*}) \) and \( \sigma(C_d) \) and the content of Theorem 3.2 becomes of assistance. We shall illustrate this in the next section.

Of particular importance is the average \( \overline{C_d} = \frac{1}{m} \sum_{i=1}^{m} C_{d_i}^q \)

which we call an averaged version of \( C_d \). We will be making extensive use of averaged versions in the next section and find it convenient to rely on the following:

**Theorem 3.3.** A design \( d^* \in \Omega \) is Schur-better than \( d \) if
\[ \sigma(C_{d*}) \leq \sigma(\overline{C}_d) \] for some averaged version \( \overline{C}_d \) of \( C_d \).

Clearly \( d^* \) is Schur-optimal over \( \Omega \) if it is Schur-better than all the designs \( d \) in \( \Omega \).

4. RESULTS ON SCHUR-OPTIMALITY

Let \( d^* \in \Omega \) be a design for which \( u(C_{d*}) \) has all its entries equal to \( u_{d*} \). Assume also that the trace of \( C_d \) is at most equal to the trace of \( C_{d*} \), for all \( d \in \Omega \).

We claim that \( u(C_{d*}) \not\preceq u(C_d) \). Let \( u(C_d) = (u_{d1}, \ldots, u_{dn})' \).

Then if \( \sum_{i=1}^{n} u_{d1} > k u_{d*} \) for some \( k \leq n \) we have \( u_{d1} \geq u_{d*} \) for all \( i \geq k + 1 \) and hence also \( \sum_{i=k+1}^{n} u_{d1} \geq (n-k) u_{d*} \).

This implies that \( \text{tr} C_d = \sum_{i=1}^{k} u_{d1} + \sum_{i=k+1}^{n} u_{d1} > k u_{d*} + (n-k) u_{d*} = nu_{d*} = \text{tr} C_{d*} \), a contradiction. We therefore have \( u(C_{d*}) \not\preceq u(C_d) \).

Theorem 3.1 gives now

**Theorem 4.1.** If there exists a design \( d^* \) in \( \Omega \) such that \( u(C_{d*}) \) has all its entries equal and trace \( C_{d*} \geq \text{trace} C_d \), for all \( d \in \Omega \), then \( d^* \) is Schur-optimal over \( \Omega \).

There are two particularly useful consequences. Before we state them, let us call a matrix completely symmetric if all its diagonal entries are equal and all its off-diagonals are also equal. The following two propositions readily
satisfy the assumptions of Theorem 4.1.

**Proposition 4.1.** Let $\mathcal{D}$ consist of designs $d$ for which $C_d$ has zero row sums. Then a design $d^* \in \mathcal{D}$ with a completely symmetric matrix of maximal trace is Schur-optimal over $\mathcal{D}$.

**Proposition 4.2.** If $d^* \in \mathcal{D}$ is a design whose information matrix is a multiple of the identity matrix and has maximal trace, then $d^*$ is Schur-optimal over $\mathcal{D}$.

The above propositions are reformulations of Proposition 1 and Proposition 1' of Kiefer (1975) in terms of Schur-optimality. Kiefer phrased the aforementioned results in terms of Universal-optimality, a concept that proves to be very valuable especially when dealing with optimality in regular settings which permit symmetric designs. The essential difference between Universal-optimality and Schur-optimality lies in the relaxation of monotonicity in a scalar to that in each individual component of $\mu(C_d)$. This permits immediate connections to the results on hermitian matrices by Ky Fan and results on majorization by Hardy, Littlewood and Polya. Schur-optimality is applicable in less regular settings especially when showing that a design with desirable symmetries is Schur-better than large classes of less symmetric designs. We illustrate this next.
Throughout the remainder of the paper, $\Omega$ is assumed to consist of designs $d$ for which the information matrix $C_d$ has row sums zero. Moreover, $u(C_d)$ will be a vector with $v-1$ components, as is the case in most discrete settings.

For a $m\times m$ matrix $C = (c_{ij})$, let us denote by $\Delta(C)$ the quantity $(m-1) \sum_{i=1}^{m} c_{ii} - \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij}$. We can now state and prove

**Theorem 4.2.** If $\Omega$ contains a design $d^*$ such that $u(C_{d^*})$ has exactly two distinct components, $u_{d^*1} < u_{d^*2}$, with $u_{d^*1}$ of multiplicity $r$ and $u_{d^*2}$ of multiplicity $s$ ($r+s = v-1$), then $d^*$ is Schur-optimal over the collection of designs $d$ in $\Omega$ which satisfy $\text{Trace } C_d \leq \text{Trace } C_{d^*}$ and either (a) or (b) below:

(a) $\Delta(M_d) \leq r(r+1) u_{d^*1}$ for some $(r+1) \times (r+1)$ principal minor $M_d$ of $C_d$.

(b) $\Delta(M_d) \geq s(s+1) u_{d^*2}$ for some $(s+1) \times (s+1)$ principal minor $M_d$ of $C_d$.

**Proof:** Let $d \in \Omega$ satisfy (a). Write the information matrix $C_d$ in such a way that $M_d$ is in the upper left hand corner. Average $C_d$ over the first $r+1$ rows and columns and then over the remaining $v-r-1$ rows and columns. Let
\( \overline{c}_d \) be the average version of \( c_d \) so obtained. Explicitly

\[
\overline{c}_d = \frac{1}{(r+1)! (v-r-1)!} \sum c_d^i
\]

where \( \sigma_i \) is a product of a permutation on the first \( r+1 \) rows and columns of \( c_1 \) with a permutation on the last \( v-r-1 \) rows and columns. The shape of \( \overline{c}_d \) is

\[
\overline{c}_d = \begin{bmatrix}
(a_d + \alpha_d)I - \alpha_dJ & -\beta_dJ \\
-\beta_dJ & (b_d + \nu_d)I - \nu_dJ
\end{bmatrix}
\]

where the upper left hand corner is \((r+1) \times (r+1)\) and \( I \) and \( J \) are respectively the identity matrix and the matrix with all its entries 1. By adding the \( v \times v \) matrix \( \beta_dJ \) to \( \overline{c}_d \), the eigenvalues of \( \overline{c}_d \) are found to be 0, \( a_d + \alpha_d \), \( \nu_d \) and \( b_d + \nu_d \) of multiplicities 1, \( r \), 1 and \( v-r-2 \), respectively. By Proposition 3.1 all these eigenvalues are nonnegative. Next we show that \( u(c_{d*}) \leq u(\overline{c}_d) \). To achieve this let us denote by \( u(\overline{c}_d) \) the vector whose first \( r \) entries are equal to the average of the first \( r \) entries of \( u(\overline{c}_d) \), and whose last \( s \) entries are equal to the average of the last \( s \) entries of \( u(\overline{c}_d) \). It is easy to see that \( u(\overline{c}_d) \leq u(\overline{c}_d) \). Since \( \lambda(M_d) = r(r+1)(a_d + \alpha_d) \) and since \( d \)
satisfies (a) we have $a_d + c_d \leq \mu_d^*$. This implies that each of the first $r$ entries of $\mu(C_d)$ is at most $\mu_d^*$. Since $\text{Trace } C_d \leq \text{Trace } C_d^*$, it now follows without much difficulty that $\mu(C_d^*) \leq \mu(C_d)$. This shows that $\mu(C_d^*) \leq \mu(C_d)$ and hence also that $\sigma(C_d^*) \leq \sigma(C_d)$. By Theorem 3.3 we now conclude that $d^*$ is Schur-better than $d$. When $d \in \Omega$ satisfies (b) a similar averaging yields $d^* \succ d$. This concludes the proof of the theorem.

One important remark is in order here. Theorem 4.2 does not impose any condition on having $\mu_d^*$ and $\mu_d^*$ close together. It is easy to see, however, that when they are close to each other more designs satisfy (a) or (b) in the theorem and hence $d^*$ is Schur-optimal over a larger collection of designs. A convenient way to ensure the closeness of $\mu_d^*$ and $\mu_d^*$ is to demand that $\text{Trace}(C_d^2) \leq \text{Trace}(C_d^2)$ for all $d \in \Omega$. Whenever such a design exists, Theorem 4.2 ensures its Schur-optimality over large subfamilies of designs in $\Omega$. This is a helpful fact, as it eases the search for the $D-, A-$ and $E$-optimal designs in $\Omega$.

Another important observation relates to the selection of a principal minor $M_d$ for which (b) holds. Since $C_d$ is nonnegative definite the diagonal elements of $C_d$ are relatively large. Moreover, the diagonal elements of $M_d$ carry a lot of weight (each one is multiplied by $s$) in $\Delta(M_d)$. 
The off-diagonals of $M_d$ are therefore of much less importance when it comes to maximizing $\Delta(M_d)$ over $M_d$. Whence it makes sense to first choose the principal minor $M_d$ of maximal trace and check if condition (b) in the above theorem is satisfied for this particular principal minor.

There is one special case of Theorem 4.2 that deserves mention. It is the setting in which all the designs in $\Omega$ have information matrices with nonpositive off-diagonals. Such is the case in the setting of block designs or two way elimination, for example.

Denote by $c_{dij}$ the entries of the information matrix $C_d$ in such a way that $c_{dll} \leq c_{d22} \leq \ldots \leq c_{dv,v}$.

**Theorem 4.3.** Let $\Omega$ consist of designs whose information matrices have zero row sums and nonpositive off-diagonals. If $\Omega$ contains a design $d^*$ which satisfies the assumptions of Theorem 4.2 then $d^*$ is Schur-optimal over the class of designs $d$ in $\Omega$ which satisfy either

$$\sum_{i=1}^{r+1} c_{dii} \leq r \mu_{d^*1} \quad \text{or} \quad \sum_{i=v-s}^{v} c_{dii} \geq (s+1) \mu_{d^*2}$$

**Proof:** Since $c_{dij} \leq 0$, for $i \neq j$, and the row sums of $C_d$ are zero, we have

$$\Delta(M_d) = \sum_{i=1}^{r+1} c_{dii} + \sum_{i=1}^{r+1} c_{dii} \leq \sum_{i=1}^{r+1} c_{dii} = (r+1) \mu_{d^*1} \leq r(r+1) \mu_{d^*1}$$

where
Md is the \((r+1)\times(r+1)\) principal minor of \(C_d\) whose diagonal entries are \(c_{di}, i = 1,2,\ldots,r+1\). We thus satisfy condition (a) in Theorem 4.2, and hence \(d^* \succeq d\). By letting \(M_d\) be the \((s+1)\times(s+1)\) principal minor in the lower right hand corner of \(C_d\) we have (by simply using the fact that \(c_{dij} \leq 0\) for \(i \neq j\)) \(\Delta(M_d) \geq \sum_{i=v-s}^{v} c_{di} \geq s(s+1)\mu_{d*2}\). We are now done by (b) of Theorem 4.2. This ends the proof.

Theorem 4.2 and 4.3 have a number of consequences when considered under specific linear models. We shall examine next a corollary in the block design setting.

Order the replication numbers in a design \(d \in \Omega_{v,b,k}\) such that \(r_{d1} \leq r_{d2} \leq \cdots \leq r_{dv}\). By observing that the \(i\)th diagonal entry of the information matrix of a binary design design \(d\) is \(\frac{k-1}{k}r_{di}\), one has the following reformulation of Theorem 4.2 in terms of the replication numbers \(r_{di}\).

**Corollary 4.1.** If \(d^*\) is a binary design in \(\Omega_{v,b,k}\) whose information matrix has exactly two distinct nonzero eigenvalues, \(\mu_{d*1} < \mu_{d*2}\) the former being of multiplicity \(r\) and the latter of multiplicity \(s\) (\(r+s = v-1\)), then \(d^*\) is Schur-better than all the binary designs \(d\) which satisfy either

\[
\sum_{i=1}^{r+1} r_{di} \leq \frac{kr_{di}}{k-1} \mu_{d*1} \quad \text{or} \quad \sum_{i=v-s}^{v} r_{di} \geq \frac{k(s+1)}{k-1} \mu_{d*2}.
\]
Designs \( d^* \) which satisfy the assumptions of Theorem 4.1 exist in many settings. Among block designs we mention the Partially Balanced Incomplete Block designs with two associate classes, extended and abridged Balanced Incomplete Block designs with any number of disjoint binary blocks and Group Divisible designs with \( \lambda_2 = \lambda_1 + 1 \) adjoined or abridged by disjoint binary blocks compatible with the partition of the groups (see Constantine (1980)).

When a pair of varieties occurs in either \( \lambda \) or \( \lambda + 1 \) blocks (which ensures the closeness of \( \mu_{d^*1} \) and \( \mu_{d^*2} \)), Theorem 4.2 asserts that all these designs are Schur-optimal over large classes of designs. We now examine one such instance more closely.

Let \( d^* \in \Omega_{v,b,k} \) be a Group Divisible design with \( \lambda_2 = \lambda_1 + 1 \) (where \( \lambda_1 \) is the number of blocks containing two varieties that are in the same group) and \( m \) groups of size \( n \).

Examples show that \( d^* \) is not Schur-optimal (but very likely both D- and A-optimal) over all designs. For \( m=2 \) a very strong optimality statement (including D- and A-) was proved by Cheng (1978); the E-optimality has been obtained by Takeuchi (1961) for general \( m \).

For a design \( d \in \Omega_{v,b,k} \) let \( N_d N_d' = (\lambda_{dij}) \). Let furthermore \( (\mathcal{M}_i)_{i=1, \ldots, m} \) be an arbitrary partition of
the $v$ varieties in $m$ groups of size $n$ each and set

$$r_d = \frac{1}{v} \sum_{i=1}^{v} (kr_{di} - \lambda_{dii})$$

and

$$\lambda_d = \frac{1}{v(n-1)} \sum_{t=1}^{m} \sum_{i,j \in M_t} \lambda_{dij}.$$  

It is not very hard to show that the matrix $kC_d$ having $m$ diagonal blocks of size $n$, all equal to $(r_d + \lambda_d)I - \lambda_dJ$ and with all the entries outside these blocks equal, is an averaged version of $kC_d$. One can show (see Magda(1979)) that $r_d + \lambda_d$ is an eigenvalue of $kC_d$ and has multiplicity $v-m$. Since $d^*$ is binary, we have $r_d \leq (k-1)r$ where $r$ is the replication of any variety in $d^*$. Now, if $\lambda_d \leq \lambda_1$ we have $r_d + \lambda_d \leq (k-1)r + \lambda_1$. But $(k-1)r + \lambda_1$ is the smallest eigenvalue of $kC_{d^*}$ and it has multiplicity $v-m$. With the assumption that $\lambda_d \leq \lambda_1$ and using the fact that $kC_{d^*}$ has maximal trace, it can be now readily verified that $\sigma(C_{d^*}) \leq \sigma(C_d)$. By Theorem 3.3 we therefore have:

**Theorem 4.3.** A Group Divisible design $d^* \in \Omega_{v,b,k}$ with $m$ groups of size $n$ and $\lambda_2 = \lambda_1 + 1$ is Schur-optimal over the class of designs $d$ which satisfy $\lambda_d \leq \lambda_1$ for $\lambda_d$ associated with some partition of the varieties.
REFERENCES


