Equivalence theorem for Schur optimality of experimental designs

Radoslav Harman
Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, 84248 Bratislava, Slovakia

Received 18 January 2006; received in revised form 2 February 2007; accepted 16 March 2007
Available online 25 May 2007

Abstract

An experimental design is said to be Schur optimal, if it is optimal with respect to the class of all Schur isotonic criteria, which includes Kiefer’s criteria of $E_k$-optimality, distance optimality criteria and many others. In the paper we formulate an easily verifiable necessary and sufficient condition for Schur optimality in the set of all approximate designs of a linear regression experiment with uncorrelated errors. We also show that several common models admit a Schur optimal design, for example the trigonometric model, the first-degree model on the Euclidean ball, and the Berman’s model.

© 2007 Elsevier B.V. All rights reserved.

MSC: 62K05

Keywords: Optimal design; Schur optimality; $E_k$-optimality; $D$-optimality; Equivalence theorem; Trigonometric regression; Berman’s model

1. Introduction

Consider a linear regression model on a compact experimental domain $X$. For any chosen design point $x \in X$, we can observe a random variable $y = f'(x) \beta + \varepsilon$, where $f: X \to \mathbb{R}^m$ is a vector of continuous and linearly independent regression functions, $\beta \in \mathbb{R}^m$ is an unknown vector of parameters, and $\varepsilon$ is a random error. For different observations, the errors are assumed to be uncorrelated, with zero mean and the same finite variance, which is assumed to be 1 without loss of generality. We will denote this model by $(f, X)$ and say that the model $(f, X)$ is $m$-dimensional.

By an approximate design we understand any probability measure $\xi$ finitely supported on $X$. For an experimenter, the value $\xi\{x\}$ determines the relative proportion of the measurements to be taken in $x$. The set of all approximate designs on $X$ will be denoted by $\Xi_X$. We say that $\xi \in \Xi_X$ is an exact design of size $n$, if it is possible to realize $\xi$ by $n$ experiments, i.e., if $\xi\{x_i\} = n_i/n$ for $x_1, \ldots, x_l \in X$ and some natural numbers $n_1, \ldots, n_l$ summing to $n$. A uniform probability on an $n$-point subset of $X$ is an example of an exact design of size $n$. For a more detailed introduction to optimal design of experiments, we refer the reader to Pázman (1986) and Pukelsheim (1993).

The information matrix of a design $\xi \in \Xi_X$ in the model $(f, X)$ is a positive semidefinite matrix defined by the formula

$$M_{f,X}(\xi) = \int_X f(x)f'(x)d\xi(x).$$

E-mail address: harman@fmph.uniba.sk

0378-3758/$-$ see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2007.05.031
If we perform \( n \) experiments according to an exact design \( \xi \) of size \( n \), and if \( \mathbf{M} = \mathbf{M}_{f,\mathbf{X}}(\xi) \) is regular, then \( n\mathbf{M} \) is equal to the inverse of the covariance matrix of the least squares estimate \( \hat{\beta} \) of the parameter \( \beta \). Moreover, if the observations are normally distributed, then \( \hat{\beta} \sim N_m(\beta, (1/n)\mathbf{M}^{-1}) \), the matrix \( \mathbf{M} \) is the normalized Fisher information matrix of \( \hat{\beta} \) and the set

\[
\mathcal{C}(\mathbf{M}) = \{ b \in \mathbb{R}^m : (b - \hat{\beta})'\mathbf{M}(b - \hat{\beta}) \leq q/n \}
\]

is the confidence ellipsoid covering \( \beta \) with probability \( P[\chi^2_m \leq q] \) (see Pázman, 1986, Section 4.2).

Let \( \mathcal{M}_+ \) be the cone of all positive semidefinite matrices of type \( m \times m \), and let \( \mathbf{M}, \mathbf{N} \in \mathcal{M}_+ \). Suppose that \( \mathbf{M} \succeq_L \mathbf{N} \), where \( \succeq_L \) is the Loewner comparison, i.e., \( \mathbf{M} - \mathbf{N} \in \mathcal{M}_+ \). Then \( \mathcal{C}(\mathbf{M}) \) is a subset of \( \mathcal{C}(\mathbf{N}) \). Hence, if the aim is to minimize the extent of the confidence ellipsoid for \( \beta \), then the design \( \xi \) with information matrix \( \mathbf{M} \) is at least as good as a design \( \xi' \) with information matrix \( \mathbf{N} \). However, except for the trivial case \( m = 1 \), there is no Loewner universally optimal design \( \xi \) satisfying \( \mathbf{M}_{f,\mathbf{X}}(\xi) \succeq_L \mathbf{M}_{f,\mathbf{X}}(\xi') \) for all \( \xi' \in \Xi_* \) (see Pukelsheim, 1993, Section 4.7).

Note that the lengths of semi-axes of the confidence ellipsoid \( \mathcal{C}(\mathbf{M}) \) are proportional to the inverse of the square roots of the eigenvalues of \( \mathbf{M} \). Thus, if the objective of the design is to minimize the extent of the confidence ellipsoid independently on its shift or an orthogonal rotation, it is reasonable to base the comparison of designs on the eigenvalues of corresponding information matrices. In this paper we focus on the Schur ordering of matrices, which is based only on the eigenvalues of the information matrix, and which does admit a universally optimal design in some nontrivial models.

For \( \mathbf{A} \in \mathcal{M}_+ \), let \( \lambda(\mathbf{A}) \) denote the vector of all eigenvalues arranged in the nondecreasing order: \( 0 \leq \lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}) \leq \cdots \leq \lambda_m(\mathbf{A}) \). On \( \mathcal{M}_+ \) we define the relation \( \succeq_S \) of Schur ordering as \( \mathbf{M} \succeq_S \mathbf{N} \) iff

\[
\sum_{i=1}^k \lambda_i(\mathbf{M}) \geq \sum_{i=1}^k \lambda_i(\mathbf{N}) \quad \text{for all } k \in \{1, \ldots, m\},
\]

cf. Hedayat (1981), Giovagnoli and Wynn (1981), Bondar (1983), and Cheng (1995). In the theory of majorization, we denote \( (1) \) by \( \lambda(\mathbf{M}) \preceq_w \lambda(\mathbf{N}) \) and say that \( \lambda(\mathbf{M}) \) is weakly supermajorized by \( \lambda(\mathbf{N}) \); see Marshall and Olkin, 1979, Chapter 1. The Schur ordering of matrices falls under the general concept of matrix majorization in the sense of the following well-known characterization (cf. Giovagnoli and Wynn, 1985; see also Pukelsheim, 1993, Chapter 14).

**Theorem 1.** Let \( \mathbf{M}, \mathbf{N} \in \mathcal{M}_+ \). Then the following statements are equivalent: (i) \( \mathbf{M} \succeq_S \mathbf{N} \); (ii) \( \mathbf{M} \succeq_L \sum_{i=1}^n \mathbf{U}_i \mathbf{N}_i \mathbf{U}_i' \) for some \( \mathbf{U}_1, \ldots, \mathbf{U}_n \in \{0, 1\} \) summing to one and some orthogonal matrices \( \mathbf{U}_1, \ldots, \mathbf{U}_n \).

An information matrix \( \mathbf{M}^* \) is said to be Schur optimal for the model \( (f, \mathbf{X}) \) in a set \( \mathcal{M} \) of information matrices, iff \( \mathbf{M}^* \in \mathcal{M} \) and \( \mathbf{M}^* \succeq_S \mathbf{N} \) for all \( \mathbf{N} \in \mathcal{M} \). By the “Schur optimal information matrix”, without specification of the set of competing information matrices, we will understand the information matrix that is Schur optimal in the set \( \mathcal{M}_{f,\mathbf{X}} = \{ \mathbf{M}_{f,\mathbf{X}}(\xi) : \xi \in \Xi_* \} \) corresponding to all approximate designs on \( \mathbf{X} \). A design \( \xi^* \in \Xi_* \) is said to be Schur optimal for the model \( (f, \mathbf{X}) \) iff \( \mathbf{M}_{f,\mathbf{X}}(\xi^*) \) is the Schur optimal information matrix for the model \( (f, \mathbf{X}) \).

An optimality criterion is a function \( \Phi : \mathcal{M}_+ \to \mathbb{R} \) measuring an aspect of the information matrix that can be captured by a single real number. An information matrix \( \mathbf{M}^* \in \mathcal{M}_{f,\mathbf{X}} \) is said to be \( \Phi \)-optimal for the model \( (f, \mathbf{X}) \) iff \( \Phi(\mathbf{M}^*) = \sup_{\mathbf{M} \in \mathcal{M}_{f,\mathbf{X}}} \Phi(\mathbf{M}) \). A design \( \xi^* \in \Xi_* \) is called \( \Phi \)-optimal for \( (f, \mathbf{X}) \) iff \( \mathbf{M}_{f,\mathbf{X}}(\xi^*) \) is a \( \Phi \)-optimal information matrix for \( (f, \mathbf{X}) \).

We will say that a criterion \( \Phi \) is Schur isotonic iff \( \Phi \) is isotonic with respect to the relation \( \succeq_S \), i.e., iff \( \mathbf{M} \succeq_S \mathbf{N} \) implies \( \Phi(\mathbf{M}) \geq \Phi(\mathbf{N}) \). Definition of the Schur ordering and Theorem 1 imply that the class of Schur isotonic criteria contains all the criteria that are Loewner isotonic and Schur concave. The most important examples of Schur isotonic criteria are the Kiefer’s criteria of \( \Phi_{p,\mathbf{M}} \)-optimality for \( -\infty < p \leq 1 \) (see, e.g., Pukelsheim, 1993, Chapter 6), in particular the criteria of \( D \)-optimality \( \Phi_0(\mathbf{M}) = (\det(\mathbf{M}))^{1/m} \), \( E \)-optimality \( \Phi_{-\infty}(\mathbf{M}) = \lambda_1(\mathbf{M}) \), \( A \)-optimality \( \Phi_{-1}(\mathbf{M}) \) and \( T \)-optimality \( \Phi_{1}(\mathbf{M}) = \text{tr}(\mathbf{M})/m \). The class of Schur isotonic criteria further includes the distance optimality criteria \( DS(\varepsilon) \) for all \( \varepsilon > 0 \) (Liski et al., 1999), concave versions of the characteristic polynomial criteria \( \Phi_{Chk} \) (Rodríguez-Díaz and López-Fidalgo, 2003), criteria \( \Psi_{p,r} \) related to principal component analysis (Dette et al., 2005) and many others.

It turns out that the assumption of Schur optimality is strong, and only special models admit a Schur optimal design (in their usual parametrization; cf. Section 3.5). On the other hand, once there exists a Schur optimal design, it inherits
favorable statistical as well as mathematical properties of all Schur isotonic criteria. For instance, $E$- and $D$-optimality of a Schur optimal design means that the design simultaneously minimizes diameter as well as volume of the confidence ellipsoid for $\beta$. Moreover, the Schur optimal information matrix is unique and regular, since these properties are satisfied by the $D$-optimal information matrix (see, e.g., Pázman, 1986, Chapter 4).

In Section 2 we formulate the central new result of this paper: a sufficient and necessary condition for Schur optimality, verification of which requires only calculation of projectors on the eigenspaces of the information matrix and maximization of explicitly given functions over the experimental domain; see Theorem 3. In this respect, the condition is similar to the equivalence theorems for real-valued differentiable criteria of optimality (cf., e.g., Pukelsheim, 1993, Chapter 7). Note that application of the theorem does not require knowledge of the theory of majorization.

In Section 3 we exhibit selected examples of models admitting a Schur optimal design, namely the trigonometric model, the model of the first-degree regression on the centered Euclidean ball, and the Berman’s model on a circle or a circular arc. By doing this, we extend known optimality results to the broad class of all Schur isotonic criteria. Finally, we show that in fact any model can be linearly reparametrized in order to admit a Schur optimal design.

2. Equivalence theorem for Schur optimality

Let $\Phi_{E_k}(M)$ denote the sum of the $k$ smallest eigenvalues of $M \in S_+^m$.

$$\Phi_{E_k}(M) = \sum_{i=1}^{k} \lambda_i(M).$$

Note that $\Phi_{E_k}$ is the criterion of $E$-optimality and $m^{-1}\Phi_{E_m}$ is the criterion of $T$-optimality. Moreover, $M \succeq S N$ if and only if $\Phi_{E_k}(M) \geq \Phi_{E_k}(N)$ for all $k = 1, \ldots, m$.

By definition, an information matrix $M^*$ (or a design $\xi^*$) is Schur optimal if and only if it is $\Phi_{E_k}$-optimal for all $k = 1, \ldots, m$. For a single $k < m$, proving $\Phi_{E_k}$-optimality of a given information matrix can be a difficult optimization problem. Nevertheless, we will show that verification whether an information matrix is $\Phi_{E_k}$-optimal simultaneously for all $k = 1, \ldots, m$, is relatively simple.

Let $\mathcal{P}_k^m$ denote the set of all $m \times m$ orthogonal projectors projecting on a $k$-dimensional linear subspace of $\mathbb{R}^m$, i.e.,

$$\mathcal{P}_k^m = \{P \in S_+^m : P^2 = P, \text{tr}(P) = k\}.$$

**Lemma 2.** Let $(f, X)$ be an $m$-dimensional model, let $\xi \in \Xi_X$, $k \in \{1, \ldots, m\}$ and let $P \in \mathcal{P}_k^m$. Then

$$\Phi_{E_k}(M_{f, X}(\xi)) \leq \max_{x \in X} \|P f(x)\|^2. \quad (2)$$

Moreover, in the case of equality in (2), the design $\xi$ is $\Phi_{E_k}$-optimal and the maximum in (2) is achieved on the support points of $\xi$.

**Proof.** Let $M = M_{f, X}(\xi)$. Any projector $P \in \mathcal{P}_k^m$ can be written in the form $P = VV'$, where $V$ is an $m \times k$ matrix with columns forming an orthonormal basis of span($P$), that is $V'V = I_k$. Using the Ky Fan’s maximum principle (see, e.g., Horn and Johnson, 1985, Corollary 4.3.18) we obtain

$$\Phi_{E_k}(M) \leq \text{tr}(VV'M) = \text{tr}(MP) = \text{tr}(MP^2) = \text{tr}(PMP) = \text{tr}\left(P \int_{x \in X} f(x) f'(x) \, d\xi(x) P'\right) = \int_{x \in X} \|P f(x)\|^2 \, d\xi(x) \leq \max_{x \in X} \|P f(x)\|^2.$$

We proved inequality (2). From the proof of (2) it is clear that the equality $\Phi_{E_k}(M) = \max_{x \in X} \|P f(x)\|^2$ implies

$$\int_{x \in X} \|P f(x)\|^2 \, d\xi(x) = \max_{x \in X} \|P f(x)\|^2.$$

We proved inequality (2). From the proof of (2) it is clear that the equality $\Phi_{E_k}(M) = \max_{x \in X} \|P f(x)\|^2$ implies

$$\int_{x \in X} \|P f(x)\|^2 \, d\xi(x) = \max_{x \in X} \|P f(x)\|^2$$

which entails the second part of the lemma. □
We remark that Lemma 2 can be used to construct bounds on the minimal efficiency of a given design with respect to the class of all Schur isotonic information functions (see Harman, 2004b for a different approach).

Define the following set associated with $\mathbf{M} \in \mathcal{M}_+^m$:

$$\mathcal{D}_M = \{k \in \{1, \ldots, m-1\} : \lambda_k(\mathbf{M}) < \lambda_{k+1}(\mathbf{M})\} \cup \{m\}.$$  

Let $\mathbf{M} = \sum_{i=1}^m \lambda_i(\mathbf{M}) u_i u_i^t$, where $u_1, \ldots, u_m$ is an orthonormal system of eigenvectors of $\mathbf{M}$. Note that for $k \in \mathcal{D}_M$ the matrix

$$P_{M,k} = \sum_{i=1}^k u_i u_i^t \in \mathcal{P}_k$$

is a uniquely defined orthogonal projector on the $k$-dimensional eigenspace of $\mathbf{M}$ corresponding to the eigenvalues $\lambda_1(\mathbf{M}), \ldots, \lambda_k(\mathbf{M})$. If $\mathbf{M}$ is a positive definite matrix, then $\mathcal{D}_M$ is the set of all indices $k$, such that $\Phi_{E_k}$ is differentiable in $\mathbf{M}$ and $P_{M,k}$ is equal to $\nabla \Phi_{E_k}(\mathbf{M})$—the gradient of $\Phi_{E_k}$ in $\mathbf{M}$ (see Harman, 2004a, Proposition 3).

**Theorem 3.** An information matrix $\mathbf{M} \in \mathcal{M}_+^m$ is Schur optimal for the model $(f, \mathcal{X})$ if and only if

$$\Phi_{E_k}(\mathbf{M}) = \max_{x \in \mathcal{X}} \|P_{M,k} f(x)\|^2 \quad \text{for all} \ k \in \mathcal{D}_M.$$  

Moreover, if $\mathbf{M}$ is the Schur optimal information matrix for $(f, \mathcal{X})$, then the maxima in (3) are achieved on the support points of any Schur optimal design.

**Proof.** Let $\mathbf{M} \in \mathcal{M}_+^m$ and let $k \in \mathcal{D}_M$. From Theorem 4 in Harman (2004a) we obtain that $\mathbf{M}$ is $\Phi_{E_k}$-optimal if and only if

$$\Phi_{E_k}(\mathbf{M}) = \max_{x \in \mathcal{X}} f'(x) \nabla \Phi_{E_k}(\mathbf{M}) f(x) = \max_{x \in \mathcal{X}} \|P_{M,k} f(x)\|^2.$$  

Therefore, we see that (i): Schur optimality implies (3) and (ii): (3) implies $\Phi_{E_k}$-optimality of $\mathbf{M}$ for all $k \in \mathcal{D}_M$. Observe that (ii) is also a direct consequence of Lemma 2.

The argument that condition (3) entails $\Phi_{E_s}$-optimality of $\mathbf{M}$ for all $s \in \{1, \ldots, m\}$, can be based on convexity of Lorenz curves (see Marshall and Olkin, 1979, Chapter 1). More precisely: for any $\mathbf{A} \in \mathcal{M}_+^m$ let $v_A : [0, m] \to \mathbb{R}$ be the piecewise linear function interpolating nodes $(k, \Phi_{E_k}(\mathbf{A}))$ for $k = 0, \ldots, m$, where $\Phi_{E_0} \equiv 0$ on $\mathcal{M}_+^m$. If $m = 1$, then $v_A$ is a linear function, and if $m > 1$, then $\lambda_k(\mathbf{A}) \leq \lambda_{k+1}(\mathbf{A})$ directly implies $v_A(k) - v_A(k-1) \leq v_A(k+1) - v_A(k)$ for all $k = 1, \ldots, m-1$. Hence, the function $v_A$ is convex on $[0, m]$.

Let $s \in \{1, \ldots, m\}$ and let $k_1, k_2 \in \{0\} \cup \mathcal{D}_M$ be such that $k_1 < s \leq k_2$ and $\lambda_{k_1+1}(\mathbf{M}) = \lambda_{k_2}(\mathbf{M})$. Choose any information matrix $\mathbf{N} \in \mathcal{M}_+^m$. Clearly, $\Phi_{E_{k_1}^{-1}}$ and $\Phi_{E_{k_2}^{-1}}$-optimality of $\mathbf{M}$ implies $\mathbf{V}_M(k_1) \succeq \mathbf{V}_N(k_1)$ as well as $\mathbf{V}_M(k_2) \succeq \mathbf{V}_N(k_2)$. But then convexity of $\mathbf{V}_N$ and linearity of $\mathbf{V}_M$ on $[k_1, k_2]$ gives $\mathbf{V}_M(s) \succeq \mathbf{V}_N(s)$, which is equivalent to $\Phi_{E_s}(\mathbf{M}) \succeq \Phi_{E_s}(\mathbf{N})$.

The second statement of the theorem follows from characterization (3) and from Lemma 2. □

Note that (3) can be also expressed using an orthonormal system $u_1, \ldots, u_m$ of eigenvectors of $\mathbf{M}$ corresponding to $\lambda_1(\mathbf{M}), \ldots, \lambda_m(\mathbf{M})$, since

$$\|P_{M,k} f(x)\|^2 = \sum_{i=1}^k (u_i^t f(x))^2.$$  

Nevertheless, knowledge of such a system of eigenvectors can be more restrictive than knowledge of the projectors on the eigenspaces, because the projectors can be trivially calculated from the eigenvectors, while the converse operation can be relatively difficult.

Theorem 3 implies that if $d$ is the number of distinct eigenvalues of $\mathbf{M} \in \mathcal{M}_+^m$, then verification of Schur optimality of $\mathbf{M}$ reduces to calculation of $d-1$ projectors on eigenspaces of $\mathbf{M}$ (the projector $P_{M,m} = \mathbf{I}_m$ is trivial) and maximization of $d$ real functions over the experimental domain. Particularly simple is the verification of Schur optimality of an orthogonal design with the information matrix proportional to $\mathbf{I}_m$, which is formulated as a stand-alone corollary of Theorem 3.
Corollary 4. Let \((f, X)\) be an \(m\)-dimensional model and let \(\lambda I_m \in \mathcal{M}_{f,X}\), where \(\lambda > 0\). Then \(\lambda I_m\) is the Schur optimal information matrix for \((f, X)\) if and only if \(\max_{x \in X} \|f(x)\|^2 = m\lambda\).

Proof. Let \(M = \lambda I_m \in \mathcal{M}_{f,X}\). Notice that \(\Sigma M = \{m\}\), \(\Phi_{E_m}(M) = \text{tr}(M) = m\lambda\) and \(P_{M_m} = I_m\). By Theorem 3, Schur optimality of \(M\) is thus equivalent to \(m\lambda = \Phi_{E_m}(M) = \max_{x \in X} \|f(x)\|^2\). □

Corollary 4 guarantees Schur optimality of many designs studied in the literature, where various techniques have been employed to prove optimality with respect to specific Schur isotonic criteria. For instance, all the following designs are Schur optimal: the \(E\)-optimal design from Example 4.2. in Dette and Studden (1993), the \(DS\)-optimal design from Theorem 4.2. in Liski et al. (1999), the \(\Psi_{p,r}\)-optimal design from the paper Dette et al. (2005), etc. In the following section several selected models admitting a Schur optimal design are described in more detail.

3. Examples of Schur optimal designs

3.1. Schur optimality in the models with constant term

Let \((f, X)\) be a model and let \(c > 0\). By \((f_c, X)\), where \(f_c = (c, f')'\), denote the model \((f, X)\) with inclusion of the constant term \(c\). Suppose that the model \((f_c, X)\) is regular in the sense that \(f_c\) is a vector of regression functions linearly independent over \(X\). If \(\xi\) is an approximate design on \(X\), then

\[
M_{f_c, X}(\xi) = \int_X \left( \begin{array}{cc} c^2 & cf'(x) \\ cf(x) & f(x)f'(x) \end{array} \right) \, d\xi(x) = \left( \begin{array}{cc} c^2 & cE_{f_c, X}(\xi) \\ cE_{f, X}(\xi) & M_{f, X}(\xi) \end{array} \right),
\]

(4)

where \(E_{f_c, X}(\xi) = \int_X f(x) \, d\xi(x)\) is the barycenter of the probability measure \(f(\xi)\). If \(E_{f, X}(\xi) = 0\), we will say that the design \(\xi\) is centered for \((f, X)\).

In general the constant term \(c\) strongly affects the eigenvalues of the information matrix in (4). Nevertheless, if \(\xi\) is centered for the model \((f, X)\), then introducing the constant term \(c\) merely adds an additional eigenvalue \(c^2\) to the spectrum of the information matrix. As we show in the following theorem, this implies that a design \(\xi\) that is Schur optimal and centered for \((f, X)\) retains its Schur optimality after inclusion of the constant term. For \(a \in \mathbb{R}\) and a \(k \times k\) matrix \(A\), we will define the \((k + 1) \times (k + 1)\) matrix

\[
D(a, A) = \left( \begin{array}{cc} a & 0_k' \\ 0_k & A \end{array} \right).
\]

Theorem 5. Let \(\xi^*\) be a centered Schur optimal design for the model \((f, X)\). Then \(\xi^*\) is a Schur optimal design and \(D(c^2, M_{f, X}(\xi^*))\) is the Schur optimal information matrix for the model \((f_c, X)\).

Proof. Choose any approximate design \(\xi\) on \(X\). We will prove the theorem by showing that, in the model \((f_c, X)\), the information matrix of \(\xi^*\) dominates the information matrix of \(\xi\) in the sense of the Schur ordering. In view of centrality of \(\xi^*\) and expression (4), we need to prove that

\[
D(c^2, M) \succeq S \left( \begin{array}{cc} c^2 & h' \\ h & N \end{array} \right),
\]

(5)

where \(M = M_{f, X}(\xi^*), N = M_{f, X}(\xi), \) and \(h = cE_{f, X}(\xi)\).

Obviously, \(M \succeq S N\), because \(M\) is Schur optimal for \((f, X)\). Thus, by Theorem 1, there exist coefficients \(\varepsilon_1, \ldots, \varepsilon_n \in [0, 1]\) summing to one and orthogonal matrices \(U_1, \ldots, U_n\), such that \(M \succeq L \sum_{i=1}^{n} \varepsilon_i U_i N U_i'\). Therefore

\[
D(c^2, M) \succeq D \left( c^2, \sum_{i=1}^{n} \varepsilon_i U_i N U_i' \right).
\]

(6)
For $i = 1, \ldots, n$ let $V_i = D(1, U_i)$ and $\beta_i = x_i/2$ and for $i = n + 1, \ldots, 2n$ let $V_i = D(1, -U_{i-n})$ and $\beta_i = x_{i-n}/2$. Note that the matrices $V_i$ are orthogonal for all $i = 1, \ldots, 2n$. It is straightforward to check that for any $h \in \mathbb{R}^m$:

$$D \left( c^2 \sum_{i=1}^n x_i U_i U_i^t \right) = 2n \sum_{i=1}^n \beta_i V_i \left( c^2 h^t \ N \right) V_i^t. \quad (7)$$

Combining (6), (7) and using Theorem 1, we obtain the required statement (5). Note that Theorem 5 can also be proved by a straightforward, but more technical application of the equivalence Theorem 3. □

We note that Theorem 5 is closely related to Lemma 1 in Schwabe (1996), stating analogous property for individual criteria of $\Phi_p$-optimality.

3.2. Trigonometric model on the full circle

Consider the trigonometric (Fourier) regression of a fixed degree $s \in \mathbb{N}$ without and with a constant term. Formally, we will analyze the models $(f, \mathfrak{F})$ and $(f_c, \mathfrak{F})$, where $\mathfrak{F} = [0, 2\pi]$, $c > 0$, and

$$f(t) = (\cos(t), \sin(t), \ldots, \cos(st), \sin(st))^t \quad \text{for} \ t \in \mathfrak{F}. \quad (8)$$

Let $\xi_n^*, n \in \mathbb{N}$, be the uniform design on $\{t_{n,1}, \ldots, t_{n,n}\}$, where

$$t_{n,j} = 2\pi(j-1)/n \quad \text{for} \ j \in \{1, \ldots, n\}. \quad (9)$$

It is a standard result in the design theory that in the model $(f_1, \mathfrak{F})$, the design $\xi_n^*$ is $\Phi_{p}$-optimal for all $-\infty < p \leq 1$ once $n \geq 2s + 1$; see, e.g., Pukelsheim (1993) Section 9.16. In Rodriguez-Díaz and López-Fidalgo (2003), optimality of $\xi_n^*$ has been proved for parameterized polynomial criteria. Finally, Proposition 6 in Harman (2004a) implies that $\xi_n^*$ is $\Phi_{Ek}$-optimal for all $k = 1, \ldots, 2s + 1$, which means that $\xi_n^*$ is Schur optimal for the model $(f_1, \mathfrak{F})$. An extension of this result to the models $(f, \mathfrak{F})$ and $(f_c, \mathfrak{F})$ can be easily proved by Corollary 4 and Theorem 5.

**Theorem 6.** Let $s \in \mathbb{N}$, $n \geq 2s + 1$. Then (i) $\frac{1}{2} I_{2s}$ is the Schur optimal information matrix and $\xi_n^*$ is a Schur optimal design for the model $(f, \mathfrak{F})$; (ii) $\text{diag}(c^2, \frac{1}{4}, \ldots, \frac{1}{4})$ is the Schur optimal information matrix and $\xi_n^*$ is a Schur optimal design for the model $(f_c, \mathfrak{F})$.

**Proof.** The information matrix of $\xi_n^*$ in $(f_1, \mathfrak{F})$ is $D(1, \frac{1}{2} I_{2s})$, which implies

$$\frac{1}{n} \sum_{j=1}^n f(t_{n,j}) f'(t_{n,j}) = \frac{1}{2} I_{2s}, \quad (10)$$

$$\sum_{j=1}^n f(t_{n,j}) = 0_{2s}. \quad (11)$$

To prove (i) for the $m = 2s$ dimensional model $(f, \mathfrak{F})$, note that equality (10) implies $M_{f, \mathfrak{F}}(\xi_n^*) = \frac{1}{2} I_m$. Schur optimality of $\frac{1}{2} I_m$ follows from Corollary 4 and the obvious equality $s = \max_{t \in \mathfrak{F}} \| f(t) \|^2$. The part (ii) is a direct consequence of (i) and Theorem 5, because (11) entails that the design $\xi_n^*$ is centered for the model $(f, \mathfrak{F})$. □

3.3. First-degree regression on the Euclidean ball

Let $\mathfrak{B}_m = \{ x \in \mathbb{R}^m : \|x\|^2 \leq m \}$ be the $m$-dimensional centered Euclidean ball with radius $\sqrt{m}$ and let $\iota : \mathbb{R}^m \to \mathbb{R}^m$ be the identity. Consider the models $(\iota, \mathfrak{B}_m)$ and $(\iota_c, \mathfrak{B}_m)$, where $\iota_c = (c, \iota')^t$ for some fixed $c > 0$, which corresponds to the first-degree regression on $\mathfrak{B}_m$ without and with a constant term. For these models, we will exhibit a large class of Schur optimal designs that are based on the functions $f : [0, 2\pi] \to \mathbb{R}^{2s}$ defined in (8) and the points $t_{n,j} \in [0, 2\pi]$ defined in (9).
Theorem 7. (i) \( I_m \) is the Schur optimal information matrix and \( \xi_{m,n}^\ast \) is a Schur optimal design for \((t, \mathcal{B}_m)\); (ii) \( D(c^2, I_m) \) is the Schur optimal information matrix and \( \xi_{m,n}^\ast \) is a Schur optimal design for \((t, \mathcal{B}_m)\).

Proof. To prove (i), notice that \( M_{1, \mathcal{B}_m}(\xi_{m,n}^\ast) = I_m \) follows from (10) and that Schur optimality of \( I_m \) for the model \((t, \mathcal{B}_m)\) is a direct consequence of Corollary 4. Obviously, equality (11) means that the designs \( \xi_{m,n}^\ast \) are centered for \((t, \mathcal{B}_m)\). Thus, (ii) follows from Theorem 5. □

The case of even \( m \). Let \( m = 2s, s \in \mathbb{N} \) and \( n \geq m + 1 \). Let \( \xi_{m,n}^\ast \) be the uniform design on the set \( \{ \sqrt{2} f(t_{n,j}) : j = 1, \ldots, n \} \).

Theorem 8. (i) \( I_m \) is the Schur optimal information matrix and \( \xi_{m,n}^\ast \) is a Schur optimal design for \((t, \mathcal{B}_m)\). If \( n \) is even, then also \( \xi_{m,n}^\ast \) is a Schur optimal design for \((t, \mathcal{B}_m)\); (ii) \( D(c^2, I_m) \) is the Schur optimal information matrix and, if \( n \) is even, then \( \xi_{m,n}^\ast \) is a Schur optimal design for \((t, \mathcal{B}_m)\).

Proof. For the part (i), notice that equalities (10) and (11) yield \( M_{1, \mathcal{B}_m}(\xi_{m,n}^\ast) = I_m \) and that Schur optimality of \( I_m \) follows from Corollary 4. To prove that, for an even \( n \), the information matrix of \( \xi_{m,n}^\ast \) is \( I_m \), we can use equalities (10) and \( \sum_{j=1}^{n} (-1)^j f(t_{n,j}) = 0 \), which is a consequence of

\[
\sum_{j=1}^{n} (-1)^j e^{i\pi.2k(j-1)/n} = \frac{e^{i\pi.2k+n} - 1}{e^{i\pi.2k/n} + 1} = 0 \quad \text{for} \ k \in \{1, 2, \ldots, n/2 - 1\}.
\]

The part (ii) follows from Theorem 5 since, if \( n \) is even, then the equalities (11) and \( \sum_{j=1}^{n} (-1)^j = 0 \) imply that \( \xi_{m,n}^\ast \) is centered for \((t, \mathcal{B}_m)\). □

We remark that \( \xi_{m,m+1}^\ast \) (if \( m \) is even) and \( \xi_{m,m+1}^\ast \) (if \( m \) is odd) are regular simplex designs, that is, they are uniform designs supported on the vertices of a regular simplex inscribed into the ball \( \mathcal{B}_m \). Note also that the parts (ii) of Theorems 7 and 8 improve the answer for the question stated in Liski et al. (1999), i.e., for which \( m \) and \( n \) there exists an exact design of size \( n \) with the information matrix \( I_{m+1} \) in the model \((t_1, \mathcal{B}_m)\). We have constructed designs with this property for all those \( m, n \) such that at least one of the numbers \( m \) and \( n \) is even.

3.4. Berman’s model on a circle or a circular arc

Let \((F, \mathcal{Q})\) denote the Berman’s bivariate linear regression model with response \((y_1, y_2)' = F'(t)\beta + (\varepsilon_1, \varepsilon_2)'\), where

\[
F'(t) = \begin{pmatrix} f_1'(t) \\ f_2'(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cos(t) & -\sin(t) \\ 0 & 1 & \sin(t) & \cos(t) \end{pmatrix}.
\]

The design point \( t \) belongs to the experimental domain \( \mathcal{Q} = [-\varpi, \varpi] \), where \( \varpi \in (0, \pi] \) is the maximal angle of the measurement, \( \beta = (\beta_1, \beta_2, \beta_3, \beta_4)' \) is the vector of parameters and the errors \( \varepsilon_1, \varepsilon_2 \) are uncorrelated with zero mean and the same variance (see Berman, 1983; cf. Wu, 1997).

For an approximate design \( \xi \) on \( \mathcal{Q} \), we define the information matrix analogously as for the univariate case:

\[
M_{F, \mathcal{Q}}(\xi) = \int_{\mathcal{Q}} F(t) F'(t) d\xi(t) = \int_{\mathcal{Q}} [f_1(t) f_1'(t) + f_2(t) f_2'(t)] d\xi(t).
\]

The set of all possible information matrices will be denoted by \( M_{F, \mathcal{Q}} \).
Define \( \psi = \min(x, \pi/2) \). Let \( \zeta^* \) be the uniform design on the set \{\psi, -\psi\}, which corresponds to an even number of alternated measurements under angles \( \psi \) and \( -\psi \). It is straightforward to verify that the information matrix of \( \zeta^* \) is

\[
M^* = M_{F,Y}(\zeta^*) = \begin{pmatrix}
1 & 0 & \cos(\psi) & 0 \\
0 & 1 & 0 & \cos(\psi) \\
\cos(\psi) & 0 & 1 & 0 \\
0 & \cos(\psi) & 0 & 1
\end{pmatrix}.
\]

We will show that \( M^* \) dominates all information matrices of \( M_{F,Y} \) in the sense of the Schur ordering, therefore \( M^* \) is optimal in the class \( M_{F,Y} \) with respect to all Schur isotonic criteria. The key idea of how to prove this claim by employing results of Section 2 is to show that \( M^* \) is Schur optimal for a univariate model \((f, \mathcal{X})\) such that \( M_{f,\mathcal{X}} \supseteq M_{F,Y} \).

Consider a modification \((f, \mathcal{X})\) of the model \((F, \mathcal{Y})\), where selection of the coordinates to be measured is a part of the experimental design, i.e., the responses are modeled by \( y = f(t, z)\beta + \epsilon \) with

\[
f(t, z) = \frac{1}{\sqrt{2}} \begin{pmatrix}
1 + z \\
1 - z \\
(1 + z) \cos t + (1 - z) \sin t \\
(1 - z) \cos t - (1 + z) \sin t
\end{pmatrix}
\]

for \((t, z)' \in \mathcal{X} = \mathcal{Y} \times \{-1, +1\}\). Note that \( z = \pm 1 \) determines the coordinate to be measured.

**Theorem 9.** (i) \( M_{f,\mathcal{X}} \supseteq M_{F,Y} \); (ii) \( M^* \) is the Schur optimal information matrix for \((f, \mathcal{X})\).

**Proof.** Let \( \zeta \) be an approximate design on \( \mathcal{Y} \). Define \( \zeta = \zeta \times \eta^* \in \Sigma_{\mathcal{X}} \), where \( \eta^* \) is the uniform probability on \{-1, +1\}. We can easily check that \( M_{f,\mathcal{X}}(\zeta) = M_{F,Y}(\zeta) \), which implies (i).

We will prove (ii). If \( \psi = \pi/2 \), then \( M^* = I_4 \). In this case, Schur optimality of \( M^* \) follows from Corollary 4 and the equality \( \| f(t, z) \|^2 = 4 \) for all \((t, z)' \in \mathcal{X} \).

To prove Schur optimality of \( M^* \) for the case \( \psi < \pi/2 \), we can use the decomposition

\[
M^* = (1 - \cos(\psi))P + (1 + \cos(\psi))(I_4 - P),
\]

where \( P \) is the \( 4 \times 4 \) projector of rank 2 having entries \( P_{ii} = \frac{1}{2} \) for all \( i \), \( P_{ij} = -\frac{1}{2} \) if \( |i - j| = 2 \), \( P_{ij} = 0 \) for all other indices \( i, j \). Hence,

\[
\lambda(M^*) = (1 - \cos \psi, 1 - \cos \psi, 1 + \cos \psi, 1 + \cos \psi)',
\]

and consequently \( \Sigma_{M^*} = \{2, 4\} \), \( P_{M^*,2} = P \), and \( P_{M^*,4} = I_4 \). But

\[
\max_{(t, z)' \in \mathcal{X}} \| P_{M^*,2} f(t, z) \|^2 = \max_{t \in [-\psi, \psi]} 2 - 2 \cos 2 t = 2 - 2 \cos \psi = \Phi_E(f)
\]

and

\[
\max_{(t, z)' \in \mathcal{X}} \| P_{M^*,4} f(t, z) \|^2 = \max_{(t, z)' \in \mathcal{X}} \| f(t, z) \|^2 = 4 = \Phi_E(f).
\]

We can close the proof using Theorem 3. \( \square \)

3.5. Orthogonalization of models based on the D-optimal information matrix

It turns out that there is no Schur optimal design for most of commonly used models, such as the polynomial regression of degrees 2 and more in the standard parametrization. Nevertheless, for any model there exists a linear reparametrization that does admit a Schur optimal design. The reparametrization of the model corresponds to orthogonalizing the D-optimal information matrix. More precisely:

**Theorem 10.** Let \( \zeta_0^* \) be a D-optimal design for the model \((f, \mathcal{X})\), let \( V \) be any orthogonal matrix of type \( m \times m \) and let \( h > 0 \). Then the design \( h\zeta_0^* \) is Schur optimal for the model \((g, \mathcal{X})\), where \( g = hVM_{f,\mathcal{X}}^{-1/2}(\zeta_0^*) f \). The information matrix of \( h\zeta_0^* \) in the model \((g, \mathcal{X})\) is \( h^2 I_m \).
Proof. Let the assumptions of Theorem 10 hold. From the definition of information matrix we directly obtain that

$M_{g,X}(x_0^\ast) = h^2 1_m$. Using the equivalence theorem for $D$-optimality of $x_0^\ast$ in $(f,X)$ we have

$$\max_{x \in X} \|g(x)\|^2 = \max_{x \in X} h^2 f'(x) M_{f,X}(x_0^\ast) f(x) = h^2 m.$$  

By Corollary 4, this is enough to guarantee Schur optimality of $x_0^\ast$ for the model $(g,X)$. □

For instance, the quadratic regression on $[-1,1]$ can be linearly transformed to the model $(g,[-1,1])$, where

$$g(x) = (\sqrt{2}, \sqrt{3}x, 3x^2 - 2)'$$

with the design uniform on $\{-1,0,1\}$ being Schur optimal.

Although the models constructed from Theorem 10 are usually artificial, the theorem underlines the crucial role of parametrization in the classical optimal design theory. One practical conclusion is that it might be difficult to justify the effort necessary for constructing optimal designs with respect to Schur isotonic criteria other than $D$-optimality. Unless, of course, we have a special reason to use a parametrization that is different from the one admitting a Schur optimal design.

Acknowledgements

I would like to thank anonymous referees for very useful comments on the first version of the paper. I am also indebted to professors A. Hedayat, B. Heiligers, A. Pázman and R. Schwabe for helping me obtain relevant literature. The research was supported by the Slovak VEGA-Grant No. 1/0264/03.

References