

Distance optimality design criterion in linear models

Erkki P. Liski¹, Arto Luoma¹, Alexander Zaigraev²

¹ Department of Mathematics, Statistics and Philosophy, University of Tampere, P.O. Box 607, FIN-33101 Tampere, Finland (e-mail: epl@uta.fi; al18853@uta.fi)

² Faculty of Mathematics and Informatics, Nicholas Copernicus University, Chopin str. 12/18, PL-87-100 Torun, Poland (e-mail: alzaig@mat.uni.tosun.pl)

Received: June 1999

Abstract. Properties of the most familiar optimality criteria, for example A-, D- and E-optimality, are well known, but the distance optimality criterion has not drawn much attention to date. In this paper properties of the distance optimality criterion for the parameter vector of the classical linear model under normally distributed errors are investigated. DS-optimal designs are derived for first-order polynomial fit models. The matter of how the distance optimality criterion is related to traditional D- and E-optimality criteria is also addressed.

Key words: Schur-concavity, polynomial designs, D- and E-optimality, isotonicity, admissability

1 Introduction

There exists an extensive literature on the characterization of optimal designs under both discrete and continuous settings using the most familiar optimality criteria, for example A-, D- and E-optimality. For references see for example the monograph by Shah and Sinha (1989) and the book by Pukelsheim (1993). However, the distance optimality criterion, though briefly mentioned in the monograph by Shah and Sinha, has not drawn much attention hitherto.

Let us first consider a line fit model through the origin. Suppose we have n uncorrelated responses

$$Y_{ij} = x_i\beta + E_{ij}, \quad (1.1)$$

$i = 1, 2, \dots, l$, and $j = 1, 2, \dots, n_i$ with expectations and variances

$$E[Y_{ij}] = x_i\beta \quad \text{and} \quad V[Y_{ij}] = \sigma^2,$$

respectively. The least squares estimator (LSE) of β

$$\hat{\beta} = \sum_{i=1}^l n_i x_i \bar{Y}_i / \sum_{i=1}^l n_i x_i^2, \quad (1.2)$$

where $\bar{Y}_i = (Y_{i1} + Y_{i2} + \dots + Y_{in_i})/n_i$, has minimum variance among all linear unbiased estimators of β .

The variance $V[\hat{\beta}] = \sigma^2 / \sum_{i=1}^l n_i x_i^2$ depends on the values x_1, x_2, \dots, x_l of the regressor x and the numbers of *replications* n_1, n_2, \dots, n_l . The values x_i are chosen from a given *regression range* χ , which is usually an interval $[a, b]$. An *experimental design* $\xi^{(n)}$, or simply ξ , for sample size n is given by a finite number of regression values in χ , and nonzero integers n_1, n_2, \dots, n_l such that $\sum_{i=1}^l n_i = n$. The LSE $\hat{\beta}$ of β and its variance $V[\hat{\beta}]$ depend on ξ , that is we may write $\hat{\beta} = \hat{\beta}(\xi)$.

We search now for a design ξ^* which maximizes the probability

$$P(|\hat{\beta}(\xi) - \beta| \leq \varepsilon).$$

The idea is to minimize the *distance* between the true parameter value and its estimate in a *stochastic* sense. Hence we call this criterion *DS-optimality* criterion. A design ξ^* is *DS-optimal* for the LSE of β when

$$P(|\hat{\beta}(\xi^*) - \beta| \leq \varepsilon) \geq P(|\hat{\beta}(\xi) - \beta| \leq \varepsilon) \quad \text{for all } \varepsilon \geq 0 \quad (1.3)$$

and for any competing design ξ with a fixed sample size n .

Let χ_m^2 denote a chi-squared random variable with m degrees of freedom. Suppose that observations Y_{ij} in the model (1.1) follow a normal distribution. Then $(\hat{\beta} - \beta)^2 / V[\hat{\beta}] \sim \chi_1^2$, and hence

$$P(|\hat{\beta} - \beta| \leq \varepsilon) = P((\hat{\beta} - \beta)^2 \leq \varepsilon^2) = P\left(\chi_1^2 \leq \frac{\varepsilon^2}{\sigma^2} \sum_{i=1}^l n_i x_i^2\right). \quad (1.4)$$

Let $\chi = [a, b]$ be the regression range and let $\xi_p = \{a, b; p\}$ with $0 \leq p \leq 1$ denote a design that assigns the weights p and $1 - p$ to the regression values b and a , respectively. If $|b| > |a|$, then the unique maximum of the probability (1.4) is $P\left(\chi_1^2 \leq \frac{\varepsilon^2 n b^2}{\sigma^2}\right)$. Thus $\xi_1 = \{a, b; 1\}$ is the unique DS-optimal design.

Similarly, if $|b| < |a|$, then $\xi_0 = \{a, b; 0\}$ is the unique DS-optimal design. Finally, if $\chi = [-a, a]$, then every design $\xi_p = \{-a, a; p\}$ with any $0 \leq p \leq 1$ is DS-optimal. In fact, under the model (1.1) with normally distributed errors the statements

- (i) $P(|\hat{\beta}(\xi^*) - \beta| \leq \varepsilon) \geq P(|\hat{\beta}(\xi) - \beta| \leq \varepsilon)$ for all $\varepsilon > 0$,
- (ii) $P(|\hat{\beta}(\xi^*) - \beta| \leq \varepsilon) \geq P(|\hat{\beta}(\xi) - \beta| \leq \varepsilon)$ for some $\varepsilon > 0$,
- (iii) $V[\hat{\beta}(\xi^*)] \leq V[\hat{\beta}(\xi)]$

are equivalent (cf. Stępniański 1989).

Note that the optimality criterion (1.3) is defined via the *peakedness* of the distributions of $\hat{\beta}(\xi^*)$ and $\hat{\beta}(\xi)$. According to a definition proposed by Birnbaum (1948), a random variable Y_1 is more peaked about μ_1 than is a random variable Y_2 about μ_2 if

$$P(|Y_1 - \mu_1| \leq \varepsilon) \geq P(|Y_2 - \mu_2| \leq \varepsilon) \quad \text{for all } \varepsilon \geq 0.$$

When $\mu_1 = \mu_2 = 0$, we simply say that Y_1 is more peaked than Y_2 . This definition was generalized to the multivariate case by Sherman (1955). For k -dimensional random vectors \mathbf{Y}_1 and \mathbf{Y}_2 , \mathbf{Y}_1 is said to be more peaked than \mathbf{Y}_2 if

$$P\{\mathbf{Y}_1 \in \mathcal{A}\} \geq P\{\mathbf{Y}_2 \in \mathcal{A}\} \quad (1.5)$$

holds for all convex and symmetric (about the origin) sets $\mathcal{A} \subset \mathbf{R}^k$.

In Section 2 we characterize the DS(ε)- and DS-optimality for designs under the classical linear model. Section 3 deals mainly with majorization properties of the DS-optimality criterion. In Section 4 we briefly consider symmetric polynomial designs and derive DS-optimal designs for the mean parameter of the m -way first-order polynomial fit model. Finally, in Section 5, we discuss the behaviour of the DS(ε)-criterion as $\varepsilon \rightarrow 0$ and as $\varepsilon \rightarrow \infty$.

2 DS-Optimality in linear models

In this paper, we consider distance optimality under the classical linear model

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \quad (2.1)$$

where the $n \times 1$ response vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ follows a multivariate normal distribution, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ is the $n \times k$ model matrix, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)'$ is the $k \times 1$ parameter vector, $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ is the expectation vector of \mathbf{Y} and $D[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$ is the dispersion matrix of \mathbf{Y} , where $\sigma^2 = V[Y_i]$ and \mathbf{I}_n is the $n \times n$ identity matrix. The regression vector \mathbf{x}_i appears as the i th row of the model matrix \mathbf{X} . In the spirit of (1.3) we define now a *DS(ε)-optimality criterion*.

Definition 2.1. Let $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\beta}}(\xi_1)$ and $\hat{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{\beta}}(\xi_2)$ be the LSE's of $\boldsymbol{\beta}$ in (2.1) under the designs ξ_1 and ξ_2 , respectively, and $\|\cdot\|$ denotes the Euclidean norm in \mathbf{R}^k . If for a given $\varepsilon > 0$

$$P(\|\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}\| \leq \varepsilon) \geq P(\|\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}\| \leq \varepsilon), \quad (2.2)$$

then the design ξ_1 is at least as good as ξ_2 with respect to the DS(ε)-criterion.

A design ξ^* is said to be *DS(ε)-optimal* for the LSE of $\boldsymbol{\beta}$ in the model (2.1) if it maximizes the probability $P(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \leq \varepsilon)$. When ξ^* is DS(ε)-optimal for all $\varepsilon > 0$, we say that ξ^* is *DS-optimal*. In the particular case $k = 1$ the DS-criterion coincides with the DS(ε)-criterion for any given $\varepsilon > 0$. Sinha (1970) introduced the distance optimality criterion in a one-way ANOVA model for

optimal allocation of observations with a given total. Note that according to the usual definition of stochastic ordering for random variables (see Marshall and Olkin 1979, p. 481), $\|\hat{\beta}_1 - \beta\|$ is stochastically not greater $\|\hat{\beta}_2 - \beta\|$, if the inequality (2.2) holds for all $\varepsilon > 0$.

An experimental design $\xi^{(n)}$ specifies $l \leq n$ distinct regression vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l$ and assigns to them frequencies n_i such that $\sum_{i=1}^l n_i = n$. The regression vectors appearing in the design are called the support of $\xi^{(n)}$, that is $\text{supp } \xi^{(n)} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\}$. The matrix $\mathbf{X}'\mathbf{X} = \sum_{i=1}^l n_i \mathbf{x}_i \mathbf{x}_i'$ is the moment matrix of $\xi^{(n)}$; it is denoted by $\mathbf{M}(\xi^{(n)})$.

The dispersion matrix of the LSE of $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ and its inverse is the precision matrix of β . This matrix may be written as

$$\frac{1}{\sigma^2} \mathbf{M}(\xi^{(n)}) = \frac{n}{\sigma^2} \sum_{i=1}^l \frac{n_i}{n} \mathbf{x}_i \mathbf{x}_i' = \frac{n}{\sigma^2} \mathbf{M}(\xi^{(n)}/n),$$

where $\mathbf{M}(\xi^{(n)}/n)$ is the averaged moment matrix of $\xi^{(n)}$ (see Pukelsheim 1993, p. 25). Designs for a specified number of trials are called exact. In practice all designs are exact.

More generally, we may allow the weights (or proportions) of the regression vectors vary continuously in the interval $[0, 1]$. A design in which the distribution of trials over χ is specified by a measure ξ , regardless of n , is called continuous or approximate. The moment matrix of a continuous design ξ is defined by

$$\mathbf{M}(\xi) = \sum_{\mathbf{x}_i \in \text{supp } \xi} \xi(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i',$$

where $\xi(\mathbf{x}_i)$ denotes the weight of the regression vector \mathbf{x}_i (see Pukelsheim 1993, p. 26). The mathematical problem of finding the optimum design is simplified by considering only continuous designs, thus ignoring the constraint that the number of trials at any design point must be an integer.

Let $\mathbf{P}\mathbf{A}\mathbf{P}' = \mathbf{M}(\xi)$ be the spectral decomposition of $\mathbf{M}(\xi)$, where \mathbf{P} is an orthogonal $k \times k$ -matrix and $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ is the diagonal matrix of the eigenvalues of \mathbf{M} arranged in decreasing order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$.

Define $\mathbf{Z} = \frac{\sqrt{n}}{\sigma} \mathbf{A}^{1/2} \mathbf{P}'(\hat{\beta} - \beta)$ and note that $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I}_k)$ and $D[\hat{\beta}] = \frac{\sigma^2}{n} \mathbf{P}\mathbf{A}^{-1} \mathbf{P}'$. Since the observations \mathbf{Y} follow a normal distribution $N_n(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$, the LSE of $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ follows the k -variate normal distribution

$$\hat{\beta} \sim N_k \left[\beta, \frac{\sigma^2}{n} \mathbf{M}(\xi)^{-1} \right]. \quad (2.3)$$

As

$$P(\|\hat{\beta} - \beta\|^2 \leq \varepsilon^2) = P\left(\frac{\sigma^2}{n} \mathbf{Z}' \mathbf{A}^{-1} \mathbf{Z} \leq \varepsilon^2\right) = P\left(\sum_{i=1}^k \frac{Z_i^2}{\lambda_i} \leq \delta^2\right)$$

for all $\delta = \sqrt{n}\varepsilon/\sigma > 0$, the DS(ε)-criterion $P(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \leq \varepsilon)$ depends on \mathbf{M} only through its eigenvalues $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)'$. We define the criterion function ψ_ε , or equivalently ψ_δ , as

$$\psi_\varepsilon(\mathbf{M}) = P(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \leq \varepsilon^2) \quad \text{and} \quad \psi_\delta(\boldsymbol{\lambda}) = P\left(\sum_{i=1}^k \frac{Z_i^2}{\lambda_i} \leq \delta^2\right). \quad (2.4)$$

It is clear that $\psi_\varepsilon(\mathbf{M}) = \psi_\delta(\boldsymbol{\lambda})$ for $\delta = \sqrt{n}\varepsilon/\sigma > 0$.

As a function of δ^2 the DS(ε)-optimality criterion $\psi_\delta(\boldsymbol{\lambda})$ is the cumulative distribution function of $\sum_{i=1}^k Z_i^2/\lambda_i$ for every fixed $\boldsymbol{\lambda} \in \mathbf{R}_+^k$. A design ξ^* is DS(ε)-optimal for the LSE of $\boldsymbol{\beta}$ in (2.1), if for a given $\varepsilon > 0$, $\psi_\varepsilon[\mathbf{M}(\xi^*)] \geq \psi_\varepsilon[\mathbf{M}(\xi)]$ for all ξ . A design ξ^* is DS-optimal if it is DS(ε)-optimal for all $\varepsilon > 0$.

3 Properties of the DS-optimality criterion

The DS(ε)-optimality criterion $\psi_\varepsilon(\mathbf{M})$ is a function from the set of $k \times k$ positive definite matrices into the interval $[0, 1]$. Equally well we may consider the corresponding function from positive eigenvalues of \mathbf{M} into $[0, 1]$:

$$\psi_\delta(\boldsymbol{\lambda}) : \mathbf{R}_+^k \rightarrow [0, 1].$$

It follows directly from (2.4) that

$$\psi_\varepsilon(a\mathbf{M}) = \psi_{\sqrt{a}\varepsilon}(\mathbf{M}) \quad \text{and} \quad \psi_\delta(a\boldsymbol{\lambda}) = \psi_{\sqrt{a}\delta}(\boldsymbol{\lambda})$$

for all $a > 0$. An essential aspect of the optimality criterion ψ_ε for given $\varepsilon > 0$ is that it induces an ordering among designs and among the corresponding moment matrices of designs. We say that a design ξ_1 is at least as good as ξ_2 , relative to the criterion ψ_ε , if $\psi_\varepsilon[\mathbf{M}(\xi_1)] \geq \psi_\varepsilon[\mathbf{M}(\xi_2)]$. In this case we can also say that the corresponding moment matrix $\mathbf{M}(\xi_1)$ is at least as good as $\mathbf{M}(\xi_2)$ with respect to ψ_ε . The Loewner partial ordering $\mathbf{M}(\xi_1) \geq \mathbf{M}(\xi_2)$ in the set of $k \times k$ nonnegative definite matrices is defined by the relation

$$\mathbf{M}(\xi_1) \geq \mathbf{M}(\xi_2) \quad \Leftrightarrow \quad \mathbf{M}(\xi_1) - \mathbf{M}(\xi_2) \text{ nonnegative definite}$$

(see Pukelsheim 1993, p. 12).

3.1 Isotonicity and admissability

The function ψ_ε conforms to the Loewner ordering in the sense that it preserves the matrix ordering, i.e. ψ_ε is *isotonic* for all $\varepsilon > 0$.

Theorem 3.1. *The DS-criterion is isotonic relative to Loewner ordering, that is*

$$\mathbf{M}(\xi_1) \geq \mathbf{M}(\xi_2) > \mathbf{0} \quad \Rightarrow \quad \psi_\varepsilon[\mathbf{M}(\xi_1)] \geq \psi_\varepsilon[\mathbf{M}(\xi_2)] \quad \text{for all } \varepsilon > 0. \quad (3.1)$$

Proof. Let $\mathbf{M}_i = \mathbf{M}(\xi_i)$, $i = 1, 2$, be the moment matrices of designs ξ_1 and ξ_2 , respectively. Then for the designs ξ_1 and ξ_2 under the model (2.1) the LSE's $\hat{\beta}_i \sim N_k(\beta, \sigma^2 \mathbf{M}_i^{-1})$, $i = 1, 2$. If $\mathbf{M}_1 \geq \mathbf{M}_2 > 0$, then the antitonicity of matrix inversion yields $\mathbf{M}_2^{-1} \geq \mathbf{M}_1^{-1} > 0$ (see Pukelsheim 1993, p. 13). Hence by Anderson's theorem (Anderson 1955, cf. Tong 1990, p. 73) we have

$$\psi_\varepsilon(\mathbf{M}_1) = P(\|\hat{\beta}_1 - \beta\| \leq \varepsilon) \geq P(\|\hat{\beta}_2 - \beta\| \leq \varepsilon) = \psi_\varepsilon(\mathbf{M}_2)$$

for all $\varepsilon > 0$. □

A reasonable weakest requirement for a moment matrix \mathbf{M} is that there be no competing moment matrix \mathbf{A} which is better than \mathbf{M} in the Loewner ordering sense. We say that a moment matrix \mathbf{M} is *admissible* when every competing moment matrix \mathbf{A} with $\mathbf{A} \geq \mathbf{M}$ is actually equal to \mathbf{M} (cf. Pukelsheim 1993, Chapter 10). A design ξ is rated admissible when its moment matrix $\mathbf{M}(\xi)$ is admissible. The admissible designs form a complete class (Pukelsheim 1993, Lemma 10.3). Thus every inadmissible moment matrix may be improved. If \mathbf{M} is inadmissible, then there exists an admissible moment matrix $\mathbf{A} \neq \mathbf{M}$ such that $\mathbf{A} \geq \mathbf{M}$. Since ψ_ε is isotonic relative to Loewner ordering, DS(ε)-optimal designs as well as DS-optimal ones can be found in the set of admissible designs.

3.2 Schur-concavity of the DS-optimality criterion

The notion of majorization proves useful in a study of the function $\psi_\delta(\lambda)$. Majorization concerns the diversity of the components of a vector (cf. Marshall and Olkin 1979, p. 7). Let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)'$ and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)'$ be $k \times 1$ vectors and $\lambda_{[1]} \geq \lambda_{[2]} \geq \dots \geq \lambda_{[k]}$, $\gamma_{[1]} \geq \gamma_{[2]} \geq \dots \geq \gamma_{[k]}$ be their ordered components.

Definition 3.1. A vector λ is said to majorize γ , written $\lambda \succ \gamma$, if $\sum_{i=1}^m \lambda_{[i]} \geq \sum_{i=1}^m \gamma_{[i]}$ holds for all $m = 1, 2, \dots, k-1$ and $\sum_{i=1}^k \lambda_i = \sum_{i=1}^k \gamma_i$.

Majorization provides a partial ordering on \mathbf{R}^k . The order $\lambda \succ \gamma$ implies that the elements of λ are more diverse than the elements of γ . Then, for example,

$$\lambda \succ \bar{\lambda} = (\bar{\lambda}, \bar{\lambda}, \dots, \bar{\lambda})' \quad \text{for all } \lambda \in \mathbf{R}^k,$$

where $\bar{\lambda} = \frac{1}{k} \sum_{i=1}^k \lambda_i$. Functions which reverse the ordering of majorization are said to be Schur-concave (cf. Marshall and Olkin 1979, p. 54).

Definition 3.2. A function $f(\mathbf{x}) : \mathbf{R}^k \rightarrow \mathbf{R}$ is said to be a Schur-concave function if for all $\mathbf{x}, \mathbf{y} \in \mathbf{R}^k$ the relation $\mathbf{x} \succ \mathbf{y}$ implies $f(\mathbf{x}) \leq f(\mathbf{y})$.

Thus the value of $f(\mathbf{x})$ becomes greater when the components of \mathbf{x} become less diverse. For further details on Schur-concave functions, see Marshall and Olkin (1979).

We say that $DS(\varepsilon)$ -criterion is Schur-concave if the function $\psi_\delta(\boldsymbol{\lambda})$ defined by the equation (2.4) is a Schur-concave function of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)'$.

Theorem 3.2. *The $DS(\varepsilon)$ -criterion is Schur-concave for all $\varepsilon > 0$.*

Proof. Consider the function $\psi_\delta(\boldsymbol{\lambda})$ defined in (2.4). Since the joint density function of the components of \mathbf{Z} is Schur-concave (Tong 1990, Theorem 4.4.1), then by Proposition 7.4.2 of Tong (1990, p. 163)

$$\psi_\delta(\boldsymbol{\lambda}) = P\left(\sum_{i=1}^k \left(\frac{Z_i}{\sqrt{\lambda_i}}\right)^2 \leq \delta^2\right)$$

is a Schur-concave function of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)'$ for all $\delta > 0$. This proves the theorem. \square

Since $\boldsymbol{\lambda} \succ \bar{\boldsymbol{\lambda}} = (\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_k)'$ for all $\boldsymbol{\lambda}$, Theorem 3.2 implies the following corollary (cf. also Tong 1990, Theorem 7.4.2):

Corollary 3.1. *For the function $\psi_\delta(\boldsymbol{\lambda})$ defined by the equation (2.4) the inequality*

$$\psi_\delta(\boldsymbol{\lambda}) \leq \psi_\delta(\bar{\boldsymbol{\lambda}})$$

holds for all $\boldsymbol{\lambda} \in \mathbf{R}^k$ and all $\delta > 0$, where $\bar{\boldsymbol{\lambda}} = (\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_k)'$ and $\bar{\lambda} = \frac{1}{k} \sum_{i=1}^k \lambda_i$.

Corollary 3.2. *Let $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ denote the column vectors whose components are the eigenvalues of \mathbf{M}_1 and \mathbf{M}_2 , respectively, arranged in decreasing order. If \mathbf{M} is a moment matrix with a vector of eigenvalues $(1 - \alpha)\boldsymbol{\lambda} + \alpha\boldsymbol{\gamma}$, then*

$$\psi_\varepsilon[(1 - \alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2] \geq \psi_\varepsilon(\mathbf{M}) \quad (3.2)$$

for all $\alpha \in [0, 1]$ and all $\varepsilon > 0$.

Proof. Let $\boldsymbol{\lambda}[(1 - \alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2]$ denote the column vector of eigenvalues of $(1 - \alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2$ arranged in decreasing order. Since by Theorem G.1. (Marshall and Olkin 1979, p. 241) $(1 - \alpha)\boldsymbol{\lambda} + \alpha\boldsymbol{\gamma} \succ \boldsymbol{\lambda}[(1 - \alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2]$, the inequality (3.2) follows from Theorem 3.2. \square

Note especially that $\boldsymbol{\lambda} + \boldsymbol{\gamma} \succ \boldsymbol{\lambda}(\mathbf{M}_1 + \mathbf{M}_2)$. Therefore $\psi_\varepsilon(\mathbf{M}_1 + \mathbf{M}_2) \geq \psi_\varepsilon(\mathbf{M})$, where $\boldsymbol{\lambda} + \boldsymbol{\gamma}$ is the vector of eigenvalues of \mathbf{M} . The following theorem is, in fact, one version of the result due to Okamoto. The proof can be found in Marshall and Olkin (1979, p. 303).

Theorem 3.3. *The function $\psi_\delta(\boldsymbol{\lambda})$ is a Schur-concave function of $(\log \lambda_1, \log \lambda_2, \dots, \log \lambda_k)'$ for all $\delta > 0$.*

The following corollary is a direct consequence of Theorem 3.3.

Corollary 3.3. *The inequality*

$$\psi_\delta(\lambda) \leq \psi_\delta(\tilde{\lambda})$$

holds for all $\lambda \in \mathbf{R}_+^k$ and all $\delta > 0$, where $\tilde{\lambda} = (\tilde{\lambda}, \tilde{\lambda}, \dots, \tilde{\lambda})'$ and $\tilde{\lambda} = \prod_{i=1}^k \lambda_i^{1/k}$.

3.3 Concavity

The function ψ_δ is concave on \mathbf{R}_+^k if

$$\psi_\delta[(1 - \alpha)\lambda_1 + \alpha\lambda_2] \geq (1 - \alpha)\psi_\delta(\lambda_1) + \alpha\psi_\delta(\lambda_2)$$

for all $\alpha \in (0, 1)$ and all $\lambda_1, \lambda_2 \in \mathbf{R}_+^k$. Concavity is often regarded as a compelling property of an optimality criterion (cf. Pukelsheim 1993, p. 115). In this section we show that the DS(ε)-optimality criterion is not, in general, concave.

Theorem 3.4. *The function $\psi_\delta(\lambda)$ is concave on \mathbf{R}_+^k for every fixed $\delta > 0$ if and only if $k \leq 2$.*

Proof. Let us first consider the special case $\lambda_1 = \lambda_2 = \dots = \lambda_k$ and let c denote the joint value of λ_i 's. Then $\psi_\delta(c) = P(\chi_k^2 \leq c\delta^2)$ is a concave function of $c > 0$ if and only if for a given $\delta > 0$ the second derivative $\psi_\delta''(c) \leq 0$ for all $c > 0$ (see Marshall and Olkin 1979, 16B.3.c., p. 448). It can be shown by direct calculation that $\psi_\delta''(c) \leq 0$ if $c \geq \frac{k-2}{\delta^2}$ and $\psi_\delta''(c) > 0$ if $c < \frac{k-2}{\delta^2}$ (cf. also Lemma 5.1). Consequently, for $k \geq 3$ the function ψ_δ is not a concave function of λ everywhere on \mathbf{R}_+^k . However, ψ_δ is concave on \mathbf{R}_+ , i.e. when $k = 1$. Also we see that ψ_δ is concave on the line $\lambda_1 = \lambda_2$ in \mathbf{R}_+^2 . Next we show that ψ_δ is concave on the whole of \mathbf{R}_+^2 .

Let us now assume that $k = 2$. We show that

$$\psi_\delta(\lambda) = \frac{1}{2\pi} \int \int_{\substack{z_1^2 + z_2^2 \leq \delta^2 \\ \frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} \leq \delta^2}} e^{-(1/2)(z_1^2 + z_2^2)} dz_1 dz_2$$

is a concave function of $\lambda = (\lambda_1, \lambda_2)'$ on \mathbf{R}_+^2 . Using polar coordinates we obtain the representation

$$\begin{aligned} \psi_\delta(\lambda) &= \frac{1}{2\pi} \int_0^{2\pi} d\phi \int_0^{\delta \sqrt{g(\lambda_1, \lambda_2, \phi)}} r e^{-(1/2)r^2} dr \\ &= \frac{1}{2\pi} \int_0^{2\pi} [1 - e^{-(1/2)\delta^2 g(\lambda_1, \lambda_2, \phi)}] d\phi, \end{aligned} \quad (3.3)$$

where $g(\lambda_1, \lambda_2, \phi) = \left(\frac{\cos^2(\phi)}{\lambda_1} + \frac{\sin^2(\phi)}{\lambda_2} \right)^{-1}$. It can be shown by differentiation that the Hessian matrix $H(\lambda, \phi) = -\frac{\partial^2 g}{\partial \lambda \partial \lambda'}$ of $-g(\lambda, \phi)$ is nonnegative definite on \mathbf{R}_+^2 for all $\phi \in [0, 2\pi]$. Hence $g(\lambda, \phi)$ is a concave function of $\lambda = (\lambda_1, \lambda_2)'$ on \mathbf{R}_+^2 for all $\phi \in [0, 2\pi]$ (Marshall and Olkin 1979, p. 448).

The second order derivative of the function $f_\delta(\lambda, \phi) = 1 - e^{-(1/2)\delta^2 g(\lambda, \phi)}$ with respect to λ is

$$\frac{\partial^2 f_\delta}{\partial \lambda \partial \lambda'} = \frac{\delta^2}{2} e^{-(1/2)\delta^2 g} \left[\frac{\partial^2 g}{\partial \lambda \partial \lambda'} - \frac{\delta^2}{2} \left(\frac{\partial g}{\partial \lambda} \right) \left(\frac{\partial g}{\partial \lambda} \right)' \right]. \quad (3.4)$$

Since it is nonpositive definite, $f_\delta(\lambda, \phi)$ is a concave function of λ on \mathbf{R}_+^2 for all $\phi \in [0, 2\pi]$. Therefore, $\psi_\delta(\lambda) = \frac{1}{2\pi} \int_0^{2\pi} f_\delta(\lambda, \phi) d\phi$ is also concave on \mathbf{R}_+^2 . \square

Note that for any given $\varepsilon > 0$, $\delta^2 = n\varepsilon^2/\sigma^2 \rightarrow \infty$ as $n \rightarrow \infty$. As shown in the proof of Theorem 3.4, $\psi_\delta(c)$ is a concave function of $c = \lambda_1 = \dots = \lambda_k$ in $\left[\frac{k-2}{\delta^2}, \infty \right)$. Clearly, for any given $\varepsilon > 0$ and $k > 2$, $\left[\frac{k-2}{\delta^2}, \infty \right) \rightarrow \mathbf{R}_+$ as $n \rightarrow \infty$. It can be shown in general that for any $\varepsilon > 0$ and $k \in \mathbf{N}$, $\psi_\delta(\lambda)$ is concave on the convex set $\mathcal{A}(\delta) = \left\{ \lambda \in \mathbf{R}_+^k : \lambda_i \in \left[\frac{k-2}{\delta^2}, \infty \right), 1 \leq i \leq k \right\}$. This set approaches \mathbf{R}_+^k as $n \rightarrow \infty$.

4 DS-optimality in polynomial fit models

We posit now a polynomial fit model of degree $d \geq 1$,

$$Y_{ij} = \beta_0 + \beta_1 t_i + \dots + \beta_d t_i^d + E_{ij}, \quad (4.1)$$

where

$$E[E_{ij}] = 0 \quad \text{and} \quad V[E_{ij}] = \sigma^2$$

for $i = 1, 2, \dots, l$ and $j = 1, 2, \dots, n_i$. The responses Y_{ij} are uncorrelated and the experimental conditions t_1, t_2, \dots, t_l are assumed to lie in the interval $\mathcal{T} = [-1, 1]$, which is called the experimental domain. The corresponding regression range $\chi = \{(1, t, \dots, t^d) : t \in \mathcal{T}\}$ is a one-dimensional curve embedded in \mathbf{R}^{d+1} . Any collection $\tau = \{t_1, t_2, \dots, t_l; p_1, p_2, \dots, p_l\}$ of $l \geq 1$ distinct points $t_i \in \mathcal{T}$ and positive numbers p_i , $i = 1, 2, \dots, l$, such that $\sum_{i=1}^l p_i = 1$, induces a continuous design ξ on the regression range χ (cf. Pukelsheim 1993, p. 32). In what follows we will also call τ a design and denote by T the set of all such designs.

4.1 Symmetric polynomial designs

First we show that the function ψ_ε is invariant with respect to the reflection operation. Let $\tau \in T$ be a design for the LSE of $\beta = (\beta_0, \beta_1, \dots, \beta_d)'$ on $\mathcal{T} = [-1, 1]$ in the polynomial fit model (4.1). The reflected design τ^R is given by $\tau^R = \{-t_1, -t_2, \dots, -t_l; p_1, p_2, \dots, p_l\}$. The designs τ and τ^R have the same even moments, while the odd moments of τ^R have a reversed sign.

If $\mathbf{M}_d(\tau)$ denotes the $(d+1) \times (d+1)$ moment matrix of τ , then

$$\mathbf{M}_d(\tau^R) = \mathbf{Q}\mathbf{M}_d(\tau)\mathbf{Q},$$

where $\mathbf{Q} = \text{diag}(1, -1, 1, -1, \dots, \pm 1)$ is a diagonal matrix with diagonal elements $1, -1, 1, -1, \dots, \pm 1$. Let $\mathbf{M}_d = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$ be the spectral decomposition of \mathbf{M}_d , where the diagonal elements of the diagonal matrix $\mathbf{\Lambda}$ are the eigenvalues of \mathbf{M} and \mathbf{P} is an orthogonal matrix. It is easy to see that $\mathbf{Q}\mathbf{M}_d\mathbf{Q} = \mathbf{Q}\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\mathbf{Q}$ is the spectral decomposition of $\mathbf{Q}\mathbf{M}_d\mathbf{Q}$. Consequently, \mathbf{M}_d and $\mathbf{Q}\mathbf{M}_d\mathbf{Q}$ have the same eigenvalues. Since the value of $\psi_\varepsilon(\mathbf{M}_d)$ depends on \mathbf{M}_d only through its eigenvalues, ψ_ε is invariant under the action of \mathbf{Q} . Thus $\psi_\varepsilon(\mathbf{M}_d) = \psi_\varepsilon(\mathbf{Q}\mathbf{M}_d\mathbf{Q})$ for all moment matrices \mathbf{M}_d and for all $\varepsilon > 0$.

Theorem 4.1. *Let τ be a design and τ^R be the corresponding reflected design in a d th degree polynomial fit model with experimental domain $\mathcal{T} = [-1, 1]$. Then*

$$\psi_\varepsilon[(1 - \alpha)\mathbf{M}_d(\tau) + \alpha\mathbf{M}_d(\tau^R)] \geq \psi_\varepsilon[\mathbf{M}_d(\tau)]$$

for all $\alpha \in [0, 1]$ and all $\varepsilon > 0$.

Proof. Let λ denote the vector of eigenvalues of $\mathbf{M}_d(\tau)$. Invariance under the action of \mathbf{Q} implies that the moment matrix $\mathbf{M}_d(\tau^R)$ has the same eigenvalues as $\mathbf{M}_d(\tau)$, and the desired result follows immediately from Corollary 3.2. \square

The symmetrized design $\bar{\tau} = \frac{1}{2}(\tau + \tau^R)$ has the moment matrix

$$\mathbf{M}_d(\bar{\tau}) = \frac{1}{2}[\mathbf{M}_d(\tau) + \mathbf{Q}\mathbf{M}_d(\tau)\mathbf{Q}],$$

where all odd moments are zero. Hence the averaging operation simplifies the moment matrices by letting all odd moments vanish. By Theorem 4.1 the symmetrized design $\bar{\tau}$ is at least as good as τ with respect to the criterion ψ_ε . Symmetrization thus increases the value of ψ_ε , or at least maintains the same value.

Pukelsheim's (1993) Claim 10.7 states that τ is admissible if and only if τ has at most $d-1$ support points in the open interval $(-1, 1)$. Thus the $(d+1)$ -point designs $\tau = \{-1, t_2, \dots, t_d, 1; p_1, p_2, \dots, p_d, p_{d+1}\}$ with $t_2, t_3, \dots, t_d \in (-1, 1)$ and $\sum_{i=1}^{d+1} p_i = 1$ are admissible. However, as noted above, the corresponding symmetrized designs $\bar{\tau}$ are at least as good as τ .

4.2 First degree polynomial fit models

Let us look at an m -way first-degree polynomial fit model

$$Y_{ij} = \beta_0 + \beta_1 t_{i1} + \cdots + \beta_m t_{im} + E_{ij} \quad (4.2)$$

with m regression variables and l experimental conditions $\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{im})'$, $i = 1, 2, \dots, l$. The experiment $\tau^{(n)}$ for sample size n has n_i replications of level $\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{im})'$. We assume now that the experimental domain is an m -dimensional Euclidean ball of radius \sqrt{m} , that is $\mathcal{T}_{\sqrt{m}} = \{\mathbf{t} \in \mathbf{R}^m : \|\mathbf{t}\| \leq \sqrt{m}\}$.

If the vectors $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{m+1}$ fulfill the conditions

$$1 + \mathbf{t}'_i \mathbf{t}_i = m + 1, \quad 1 + \mathbf{t}'_i \mathbf{t}_j = 0 \quad (4.3)$$

for all $i \neq j \leq m + 1$, then the vectors span a convex body called a regular simplex. The vertices \mathbf{t}_i belong to the boundary sphere of the ball $\mathcal{T}_{\sqrt{m}}$. For $m = 2$ the support points $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ span an equilateral triangle on the sphere of $\mathcal{T}_{\sqrt{2}}$. A design which places uniform weight $1/(m + 1)$ on the vertices $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{m+1}$ of a regular simplex in \mathbf{R}^m is called a *regular simplex design* (cf. Pukelsheim 1993, p. 391). For all $m \geq 1$ it is always possible to choose vectors $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{m+1}$ such that they fulfill the conditions (4.3). Putting equal weights $\frac{1}{m + 1}$ to them yields a regular simplex design τ with $\mathbf{M}_1(\tau) = \mathbf{I}_{m+1}$. Thus a regular simplex design always exists. Note that any rotation of a regular simplex design is also a regular simplex design.

The smallest possible support size of a feasible design for the LSE of $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)'$ in an m -way first-degree model is $l = m + 1$, because $\boldsymbol{\beta}$ has $m + 1$ components.

Theorem 4.2. *Let T_{m+1} be the set of designs τ with support size $l = m + 1$ in the m -way first-degree model (4.2) on the experimental domain $\mathcal{T}_{\sqrt{m}}$. Then a design $\tau \in T_{m+1}$ is DS-optimal if and only if it is a regular simplex design.*

Proof. Let $\tau \in T_{m+1}$ be a design for the LSE of $\boldsymbol{\beta}$ with support size $l = m + 1$. Hence $\tau = \{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{m+1}; p_1, p_2, \dots, p_{m+1}\}$, where $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{m+1}$ are distinct vectors from $\mathcal{T}_{\sqrt{m}}$ with positive weights p_1, p_2, \dots, p_{m+1} such that $\sum_{i=1}^{m+1} p_i = 1$. For such a design

$$\mathbf{M}_1(\tau) = \mathbf{X}'\mathbf{D}\mathbf{X}, \quad \mathbf{D} = \text{diag}(p_1, p_2, \dots, p_{m+1})$$

and the model matrix \mathbf{X} is square. Since $|\mathbf{X}'\mathbf{D}\mathbf{X}| = |\mathbf{D}\mathbf{X}\mathbf{X}'|$, it follows from Hadamard's inequality (Horn and Johnson 1987, p. 477) that

$$|\mathbf{M}_1(\tau)| = \lambda_1 \lambda_2 \dots \lambda_{m+1} \leq \prod_{i=1}^{m+1} p_i (1 + \mathbf{t}'_i \mathbf{t}_i), \quad (4.4)$$

where $\lambda_1, \lambda_2, \dots, \lambda_{m+1}$ are the eigenvalues of $\mathbf{M}_1(\tau)$ and $p_i (1 + \mathbf{t}'_i \mathbf{t}_i)$, $i = 1, 2, \dots, m + 1$ are the diagonal elements of $\mathbf{D}\mathbf{X}\mathbf{X}'$. Equality holds if and only if the matrix $\mathbf{X}\mathbf{X}'$ is diagonal, that is if $1 + \mathbf{t}'_i \mathbf{t}_j = 0$ for all $i \neq j \leq m + 1$.

Now the arithmetic-geometric mean inequality

$$\prod_{i=1}^{m+1} p_i \leq \left(\frac{1}{m+1} \right)^{m+1}$$

and the conditions $\mathbf{t}'_i \mathbf{t}_i \leq m$; $i = 1, 2, \dots, m+1$, yield the inequality

$$\prod_{i=1}^{m+1} p_i (1 + \mathbf{t}'_i \mathbf{t}_i) \leq 1. \quad (4.5)$$

Equality holds if and only if $p_1 = p_2 = \dots = p_{m+1} = \frac{1}{m+1}$ and all \mathbf{t}_i lie on the boundary sphere of $\mathcal{T}_{\sqrt{m}}$.

By Corollary 3.3

$$\psi_\varepsilon[\mathbf{M}_1(\tau)] \leq P(\chi_{m+1}^2 \leq \delta^2 \tilde{\lambda}) \quad (4.6)$$

for all $\varepsilon > 0$, where $\delta = \sqrt{n}\varepsilon/\sigma$ and $\tilde{\lambda} = (\lambda_1 \lambda_2 \dots \lambda_{m+1})^{1/(m+1)}$. The right hand side of (4.6) is an increasing function of the bound $\delta^2 \tilde{\lambda}$. Therefore (4.4), (4.5) and (4.6) yield the inequalities

$$\psi_\varepsilon[\mathbf{M}_1(\tau)] \leq P(\chi_{m+1}^2 \leq \delta^2 \tilde{\lambda}) \leq P(\chi_{m+1}^2 \leq \delta^2) \quad (4.7)$$

for all $\varepsilon > 0$.

Let $\tau \in T_{m+1}$ be DS-optimal. It therefore maximizes $\psi_\varepsilon[\mathbf{M}_1(\tau)]$ for all $\varepsilon > 0$. If $\psi_\varepsilon[\mathbf{M}_1(\tau)]$ attains its maximum $P(\chi_{m+1}^2 \leq \delta^2)$ for all $\varepsilon > 0$, then it follows from (4.4), (4.5) and (4.7) that the support points $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{m+1}$ fulfill the conditions (4.3) and $p_1 = p_2 = \dots = p_{m+1} = \frac{1}{m+1}$, i.e. τ is a regular simplex design.

On the other hand, for a regular simplex design $\tau \in T_{m+1}$ we have $\mathbf{M}_1(\tau) = \mathbf{I}_{m+1}$. But then $\psi_\varepsilon[\mathbf{M}_1(\tau)] = P(\chi_{m+1}^2 \leq \delta^2)$ for all $\varepsilon > 0$ and hence τ is DS-optimal. This completes the proof of the theorem. \square

If $m = 1$, then we have the line fit model

$$Y_{ij} = \beta_0 + \beta_1 t_i + E_{ij} \quad (4.8)$$

with experimental domain $\mathcal{T} = [-1, 1]$. It follows from the theorem of De la Garza (1954) and Theorem 4.2 that the design $\tau_{1/2} = \{-1, 1; \frac{1}{2}\}$, which assigns weight $\frac{1}{2}$ to the points -1 and 1 , is the unique DS-optimal design for the LSE of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ in (4.8). The same result with the help of a different technique was obtained in Liski, Luoma, Mandal and Sinha (1998). An $(m+1) \times (m+1)$ matrix \mathbf{X} with entries 1 and -1 is called a *Hadamard matrix* if $\mathbf{X}'\mathbf{X} = (m+1)\mathbf{I}_{m+1}$. Thus there exists a two-level regular simplex design (and therefore a two-level DS-optimal design) with levels ± 1 if and only if there is a Hadamard matrix of order $m+1$. If $m = 2$, for example, there is no Hadamard matrix of order 3 and there exists no two-level regular simplex design with levels ± 1 in a 2-way first-order polynomial fit model.

Corollary 4.1. *Let T_l be the set of designs τ with support size $l \geq m + 1$ in the m -way first-degree model (4.2) on the experimental domain $\mathcal{T}_{\sqrt{m}}$. Then a design $\tau \in T_l$ is DS-optimal if $\mathbf{M}_1(\tau) = \mathbf{I}_{m+1}$.*

Proof. Let $\tau = \{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_l; p_1, p_2, \dots, p_l\} \in T_l$ be a design for the LSE of $\boldsymbol{\beta}$. Since $\text{tr } \mathbf{M}_1(\tau) = \text{tr } \mathbf{X}'\mathbf{D}\mathbf{X}$, we have by the properties of trace that

$$\text{tr } \mathbf{X}'\mathbf{D}\mathbf{X} = \text{tr } \mathbf{D}\mathbf{X}\mathbf{X}' = \sum_{i=1}^l p_i(1 + \mathbf{t}_i'\mathbf{t}_i) \leq m + 1.$$

The upper bound is attained if and only if $\mathbf{t}_i'\mathbf{t}_i = m$ for all $i = 1, 2, \dots, l$. Clearly, under the constraint $\text{tr } \mathbf{M}_1(\tau) = \lambda_1 + \lambda_2 + \dots + \lambda_{m+1} \leq m + 1$, $|\mathbf{M}_1(\tau)| = \lambda_1 \lambda_2 \dots \lambda_{m+1}$ attains its maximum if and only if $\lambda_1 = \lambda_2 = \dots = \lambda_{m+1} = 1$. Consequently, $|\mathbf{M}_1(\tau)|$ has its maximum for a design τ if $\mathbf{M}_1(\tau) = \mathbf{I}_{m+1}$. By the same argument as in the last paragraph of Theorem 4.2, we can conclude that τ is DS-optimal if $\mathbf{M}_1(\tau) = \mathbf{I}_{m+1}$. \square

As shown in Theorem 4.2, $\mathbf{M}_1(\tau) = \mathbf{I}_{m+1}$ holds for a regular simplex design τ . Hence a design τ satisfying the condition $\mathbf{M}_1(\tau) = \mathbf{I}_{m+1}$ exists for the minimal feasible support size $l = m + 1$. Existence of a DS-optimal design in general, for any given pair of positive integers m and $l \geq m + 1$, seems to be an unsolved problem. Nevertheless, a DS-optimal design can be found for certain values of m and $l \geq m + 1$. For example, the complete factorial design 2^m is a DS-optimal design with $l = 2^m$. It is a two-level design which assigns uniform weight $1/2^m$ to each of the $l = 2^m$ vertices of the m dimensional cube $[-1, 1]^m \subset \mathcal{T}_{\sqrt{m}}$. The moment matrix of the complete factorial design 2^m is \mathbf{I}_{m+1} , and consequently it is a DS-optimal design for the LSE of $\boldsymbol{\beta}$.

But if we have an m -way first-degree model without a constant term,

$$Y_{ij} = \beta_1 t_{i1} + \dots + \beta_m t_{im} + E_{ij}, \quad (4.9)$$

then a DS-optimal design on $\mathcal{T}_{\sqrt{m}}$ always exists. This result follows from Chow and Lii (1993), where D-optimal designs were constructed for the LSE of $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)'$ in the model (4.9). It can be shown, using Corollary 4.1, that these D-optimal designs are also DS-optimal.

5 D-, E- and DS(ε)-optimality

To open this section we consider a class of designs for mean parameters of a simple second-degree polynomial fit model. It is an example of a set of designs where no DS-optimal design exists. The main result of this section, Theorem 5.1, shows that the classical D- and E-criteria follow from the DS(ε)-criterion as ε approaches 0 and ∞ , respectively.

5.1 Quadratic regression without the first-degree term

Let us now consider the parabola fit model

$$E[Y_{ij}] = \beta_0 + \beta_1 t_i^2 \quad (5.1)$$

over the experimental domain $\mathcal{T} = [-1, 1]$. If we denote $z_i = t_i^2$, then the experimental domain of z is $\mathcal{T}_z = [0, 1]$. We may thus seek DS-optimal design for the LSE of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ over $\mathcal{T}_z = [0, 1]$.

Let τ be a design with $l \geq 2$ distinct experimental levels z_1, z_2, \dots, z_l on $\mathcal{T}_z = [0, 1]$. Then, by the theorem of De la Garza (1954), it is always possible to find a two-point design $\tau_p = \{a, b; p\}$ such that

$$\mathbf{M}_1(\tau) = \mathbf{M}_1(\tau_p),$$

where p and $q = 1 - p$ are weights at the points b and a , respectively, and $a, b \in [0, 1]$ (cf. Pukelsheim 1993, Claim 10.7). Thus the moment matrix of τ_p is

$$\mathbf{M}_1(\tau_p) = q \begin{pmatrix} 1 & a \\ a & a^2 \end{pmatrix} + p \begin{pmatrix} 1 & b \\ b & b^2 \end{pmatrix} = \begin{pmatrix} 1 & qa + pb \\ qa + pb & qa^2 + pb^2 \end{pmatrix}.$$

However, it is likewise always possible to choose a competing design $\tau_{p^*} = \{0, 1; p^*\}$ with $p^* = qa + pb$, which is at least as good as τ_p in the Loewner ordering sense. This claim can be checked straightforwardly by noting that

$$\mathbf{M}_1(\tau_{p^*}) - \mathbf{M}_1(\tau_p) = \begin{pmatrix} 0 & 0 \\ 0 & q(a - a^2) + p(b - b^2) \end{pmatrix}$$

is nonnegative definite, since $q(a - a^2) + p(b - b^2) \geq 0$ for all $a, b \in [0, 1]$. As DS-optimality criterion is isotonic relative to Loewner ordering, we have the inequality

$$\psi_\varepsilon[\mathbf{M}_1(\tau_{p^*})] \geq \psi_\varepsilon[\mathbf{M}_1(\tau_p)] \quad \text{for all } \varepsilon > 0.$$

Therefore τ_{p^*} is at least as good as τ_p with respect to the DS-optimality criterion. From this it follows that we may restrict ourselves to the class of two-point designs $\tilde{\tau}_p = \{0, 1; p\}$ where p is the proportion of replications at 1. The moment matrix is of the form

$$\mathbf{M}_1(\tilde{\tau}_p) = \begin{pmatrix} 1 & p \\ p & p \end{pmatrix}. \quad (5.2)$$

The eigenvalues of $\mathbf{M}_1(\tilde{\tau}_p)$ are

$$\lambda_{1,2} = \frac{p+1}{2} \pm \sqrt{\left(\frac{p+1}{2}\right)^2 - p(1-p)}.$$

The greatest eigenvalue $\lambda_1(p)$ is a monotonically increasing function of p , while $\lambda_2(p)$ attains its maximum at $p = \frac{2}{3}$. The product of the eigenvalues $\lambda_1(p)\lambda_2(p) = p(1-p) = |\mathbf{M}_1(\tilde{\tau}_p)|$, which has the maximum at $p = \frac{1}{2}$. Hence $\tau_{2/5} = \{0, 1; \frac{2}{5}\}$ is the E-optimal design and $\tau_{1/2} = \{0, 1; \frac{1}{2}\}$ is the D-optimal design for the LSE of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ over $\mathcal{T}_z = [0, 1]$.

Let $\mathbf{Z} = \frac{\sqrt{n}}{\sigma} \boldsymbol{\Lambda}^{1/2} \mathbf{P}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ be defined as in Section 2. Since the eigenvalues $\lambda_1(p)$ and $\lambda_2(p)$ of the matrix (5.2) are functions of the weight param-

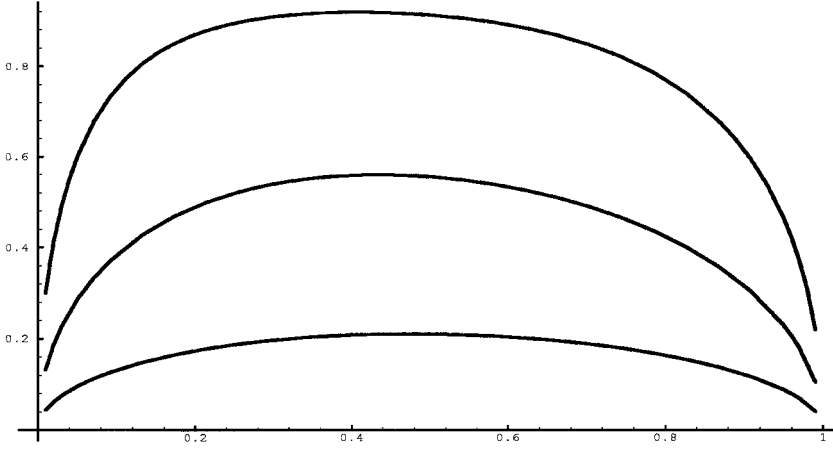


Fig. 1. The graph of $\psi_\delta(p)$ for (a) $\delta = 1$, (b) $\delta = 2$, (c) $\delta = 4$

eter p , the $\text{DS}(\varepsilon)$ -optimality criterion $\psi_\varepsilon(\mathbf{M}_1)$ depends on \mathbf{M}_1 solely through p . Hence we may write

$$\psi_\delta(\lambda) = \psi_\delta(p) = \mathbf{P}\left(\frac{Z_1^2}{\lambda_1(p)} + \frac{Z_2^2}{\lambda_2(p)} \leq \delta^2\right).$$

Note that $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, and consequently $\mathbf{Z} = (Z_1, Z_2)' \sim N_2(\mathbf{0}, \mathbf{I}_2)$. Since ψ_δ is isotonic relative to Loewner ordering and $\lambda_1(p)$ and $\lambda_2(p)$ are increasing in $[0, \frac{2}{3}]$, we may conclude that $\psi_\delta(\frac{2}{3}) \geq \psi_\delta(p)$ for all $p \in [0, \frac{2}{3}]$ and all $\delta > 0$. Consequently, if $\tilde{\tau}_p$ is DS-optimal, then $p \in [\frac{2}{3}, 1)$.

We inquire now whether there exists a DS-optimal design $\tilde{\tau}_p = \{0, 1; p\}$, $p \in [0, 1]$. If we define $V_i = Z_i/(\delta\sqrt{\lambda_i})$, then $\mathbf{V} = (V_1, V_2)'$ follows a normal distribution $N_2(\mathbf{0}, \delta^{-2}\mathbf{\Lambda}^{-1})$. The criterion ψ_δ can be written as

$$\begin{aligned} \psi_\delta(p) &= \mathbf{P}(\|\mathbf{V}\|^2 \leq 1) \\ &= \frac{\delta^2 \sqrt{\lambda_1(p)\lambda_2(p)}}{2\pi} \iint_{\|\mathbf{v}\| \leq 1} \exp\left\{-\frac{\delta^2}{2} \mathbf{v}' \mathbf{\Lambda} \mathbf{v}\right\} d\mathbf{v}, \end{aligned}$$

where $\|\mathbf{V}\|^2 = V_1^2 + V_2^2$, $\mathbf{v}' \mathbf{\Lambda} \mathbf{v} = \lambda_1(p)v_1^2 + \lambda_2(p)v_2^2$ and $d\mathbf{v} = dv_1 dv_2$. The function $\psi_\delta(p)$ is positive and continuous on $(0, 1)$ for all $\delta > 0$ and $\psi_\delta(0) = \psi_\delta(1) = 0$. Hence $\psi_\delta(p)$ has a maximum on the interval $(0, 1)$ at, say, p^* . However, the maximum point p^* depends on δ . Consequently, there exists no DS-optimal design for the LSE of $\boldsymbol{\beta}$ over $\mathcal{T}_z = [0, 1]$ in the model (5.1).

5.2 Characterization of D- and E-optimal designs using $\text{DS}(\varepsilon)$ -optimality

Let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)'$ and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)'$ be the vectors of eigenvalues of the positive definite matrices $\mathbf{M}(\xi_1)$ and $\mathbf{M}(\xi_2)$, respectively. The entries of

λ and γ are arranged in decreasing order. Since $\mathbf{M}(\xi_1)$ and $\mathbf{M}(\xi_2)$ are positive definite, $\lambda_k > 0$ and $\gamma_k > 0$. A design ξ_1 is at least as good as a design ξ_2 , with respect to the $\text{DS}(\varepsilon)$ -criterion, if

$$\psi_\varepsilon(\mathbf{M}(\xi_1)) \geq \psi_\varepsilon(\mathbf{M}(\xi_2)), \quad \text{or equivalently, } \psi_\delta(\lambda) \geq \psi_\delta(\gamma)$$

for $\delta = \sqrt{n\varepsilon}/\sigma > 0$. Denote $V_i = Z_i/(\delta\sqrt{\lambda_i})$, $i = 1, 2, \dots, k$ and $\mathbf{V} = (V_1, V_2, \dots, V_k)'$. Then \mathbf{V} follows a k -variate normal distribution $N_k(\mathbf{0}, \delta^{-2}\mathbf{\Lambda}^{-1})$. We may therefore write $\psi_\delta(\lambda)$ using (2.4) as follows:

$$\begin{aligned} \psi_\delta(\lambda) &= \mathbf{P}(\|\mathbf{V}\|^2 \leq 1) \\ &= \delta^k (2\pi)^{-k/2} |\mathbf{\Lambda}|^{1/2} \int_{\|\mathbf{v}\| \leq 1} \dots \int \exp\left(-\frac{\delta^2}{2} \mathbf{v}' \mathbf{\Lambda} \mathbf{v}\right) d\mathbf{v}, \end{aligned} \quad (5.3)$$

where $\|\mathbf{V}\|^2 = V_1^2 + V_2^2 + \dots + V_k^2$, $\mathbf{v} = (v_1, v_2, \dots, v_k)'$ and $d\mathbf{v} = dv_1 dv_2 \dots dv_k$. For any fixed λ with $\lambda_k > 0$ we may consider $\psi_\delta(\lambda)$ as a function of δ . Theorem 5.1 utilizes the following lemma.

Lemma 5.1. *Let χ_1^2 and χ_k^2 denote chi-squared random variables with 1 and k ($k > 1$) degrees of freedom, respectively. Let a and b be positive real numbers and let $\delta_0 = \delta_0(a, b, k)$ be sufficiently large. Then the following statements hold:*

- (i) *If $a \leq b$, then $\mathbf{P}(\chi_k^2 \leq a\delta) < \mathbf{P}(\chi_1^2 \leq b\delta)$ for all $\delta > 0$.*
- (ii) *If $a > b$, then $\mathbf{P}(\chi_k^2 \leq a\delta) > \mathbf{P}(\chi_1^2 \leq b\delta)$ for all $\delta > \delta_0$.*

Proof. It is clear that

$$\mathbf{P}(\chi_k^2 > c) > \mathbf{P}(\chi_1^2 > c) \quad \text{for all } c > 0,$$

since $\chi_k^2 = \sum_{i=1}^k Z_i^2$ and $\chi_1^2 = Z_1^2$. Therefore, $a \leq b$ implies $\mathbf{P}(\chi_k^2 \leq a\delta) < \mathbf{P}(\chi_1^2 \leq b\delta)$ for all $\delta > 0$. Thus we have proved the part (i).

To prove the part (ii) we consider the quotient

$$q(\delta; a, b) = \frac{H_k(\delta; a)}{H_1(\delta; b)},$$

where the functions

$$H_k(\delta; a) = \mathbf{P}(\chi_k^2 > a\delta) = c_k \int_{a\delta}^{\infty} x^{(k/2)-1} e^{-x/2} dx > 0$$

and

$$H_1(\delta; b) = \mathbf{P}(\chi_1^2 > b\delta) = c_1 \int_{b\delta}^{\infty} x^{-1/2} e^{-x/2} dx > 0$$

as functions of δ are defined on $[0, \infty)$ and $a, b, c_k = \frac{1}{\Gamma(\frac{k}{2})2^{k/2}}$ and $c_1 = \frac{1}{\sqrt{2\pi}}$ are positive constants.

We prove the lemma by showing that $q(\delta; a, b) \rightarrow 0$ as $\delta \rightarrow \infty$ for $a > b$. The limit of the quotient $q(\delta; a, b)$ is indeterminate of the form $0/0$ as $\delta \rightarrow \infty$. The functions H_k and H_1 are differentiable and their derivatives with respect to δ are

$$H'_k(\delta; a) = -c_k(a\delta)^{(k/2)-1}e^{-a\delta/2}a$$

and

$$H'_1(\delta; b) = -c_1(b\delta)^{-1/2}e^{-b\delta/2}b,$$

respectively. Therefore we test the quotient of derivatives: $H'_k(\delta; a)/H'_1(\delta; b) = c\delta^{(k-1)/2}e^{(b-a)\delta/2}$, where c is a constant not depending on δ . Then we have

$$\frac{H'_k(\delta; a)}{H'_1(\delta; b)} \rightarrow 0 \quad \text{as } \delta \rightarrow \infty \text{ if and only if } a > b.$$

Hence, by L'Hospital's rule,

$$\frac{H_k(\delta; a)}{H_1(\delta; b)} \rightarrow 0 \quad \text{as } \delta \rightarrow \infty \quad \text{for } a > b. \quad (5.4)$$

Consequently, for sufficiently large δ by (5.4)

$$H_k(\delta; a) = 1 - P(\chi_k^2 \leq a\delta) < H_1(\delta; b) = 1 - P(\chi_1^2 \leq b\delta) \quad \text{for } a > b.$$

This completes our proof. \square

We are now in a position to characterize the $DS(\varepsilon)$ -criterion, when ε approaches 0 and ∞ , respectively. These limiting cases have an interesting relationship with the traditional D- and E-optimality criteria.

Theorem 5.1. *Let $\lambda, \gamma \in \mathbf{R}^k$ denote vectors whose components are the eigenvalues of the moment matrices $\mathbf{M}(\xi_1)$ and $\mathbf{M}(\xi_2)$, respectively, arranged in decreasing order. Then the following statements hold:*

- (a) *If $\psi_\delta(\lambda) \geq \psi_\delta(\gamma)$ for all sufficiently small $\delta > 0$, then $|\mathbf{M}(\xi_1)| \geq |\mathbf{M}(\xi_2)|$; if $|\mathbf{M}(\xi_1)| > |\mathbf{M}(\xi_2)|$, then $\psi_\delta(\lambda) > \psi_\delta(\gamma)$ for all sufficiently small $\delta > 0$.*
- (b) *If $\psi_\delta(\lambda) \geq \psi_\delta(\gamma)$ for all sufficiently large δ , then $\lambda_k \geq \gamma_k$; if $\lambda_k > \gamma_k$, then $\psi_\delta(\lambda) > \psi_\delta(\gamma)$ for all sufficiently large δ .*

Proof. We first prove part (a). In view of (5.3) the inequality $\psi_\delta(\lambda) \geq \psi_\delta(\gamma)$ can be written as

$$|\Lambda|^{1/2} \int \dots \int_{\|\mathbf{v}\| \leq 1} \exp\left(-\frac{\delta^2}{2} \mathbf{v}' \Lambda \mathbf{v}\right) d\mathbf{v} \geq |\Gamma|^{1/2} \int \dots \int_{\|\mathbf{v}\| \leq 1} \exp\left(-\frac{\delta^2}{2} \mathbf{v}' \Gamma \mathbf{v}\right) d\mathbf{v}, \quad (5.5)$$

where $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ are the diagonal matrices of the eigenvalues of $\mathbf{M}(\xi_1)$ and $\mathbf{M}(\xi_2)$, respectively. Both integrals

$$\int_{\|\mathbf{v}\| \leq 1} \dots \int \exp\left(-\frac{\delta^2}{2} \mathbf{v}' \mathbf{\Lambda} \mathbf{v}\right) d\mathbf{v} \quad \text{and} \quad \int_{\|\mathbf{v}\| \leq 1} \dots \int \exp\left(-\frac{\delta^2}{2} \mathbf{v}' \mathbf{\Gamma} \mathbf{v}\right) d\mathbf{v}$$

approach $\pi^{k/2} / \Gamma\left(\frac{k}{2} + 1\right)$ as $\delta \rightarrow 0$. Consequently, if (5.5) holds for all sufficiently small $\delta > 0$, then $|\mathbf{\Lambda}|^{1/2} \geq |\mathbf{\Gamma}|^{1/2}$, or equivalently $|M(\xi_1)| = |\mathbf{\Lambda}| \geq |\mathbf{\Gamma}| = |M(\xi_2)|$. On the other hand, if $|\mathbf{\Lambda}| = |M(\xi_1)| > |M(\xi_2)| = |\mathbf{\Gamma}|$, then (5.5) with a strict inequality holds for all sufficiently small $\delta > 0$. This completes the proof of part (a).

The proof of part (b). Since $\gamma_i \geq \gamma_k > 0$ ($\gamma_i^{-1} \leq \gamma_k^{-1}$) for all $i = 1, 2, \dots, k$, Theorem 3.1 yields the inequality

$$\begin{aligned} \psi_\delta(\gamma) &= \mathbf{P}\left(\frac{Z_1^2}{\gamma_1} + \frac{Z_2^2}{\gamma_2} + \dots + \frac{Z_k^2}{\gamma_k} \leq \delta^2\right) \\ &\geq \mathbf{P}\left(\frac{Z_1^2}{\gamma_k} + \frac{Z_2^2}{\gamma_k} + \dots + \frac{Z_k^2}{\gamma_k} \leq \delta^2\right) = \mathbf{P}(\chi_k^2 \leq \gamma_k \delta^2), \end{aligned} \quad (5.6)$$

where $\mathbf{Z} \sim \mathbf{N}_k(\mathbf{0}, \mathbf{I}_k)$ and $Z_1^2 + Z_2^2 + \dots + Z_k^2 = \chi_k^2$ is a chi-squared random variable with k degrees of freedom. On the other hand, we have the inequality

$$\begin{aligned} \mathbf{P}(\chi_1^2 \leq \lambda_k \delta^2) &= \mathbf{P}\left(\frac{Z_k^2}{\lambda_k} \leq \delta^2\right) \\ &> \mathbf{P}\left(\frac{Z_1^2}{\lambda_1} + \frac{Z_2^2}{\lambda_2} + \dots + \frac{Z_k^2}{\lambda_k} \leq \delta^2\right) = \psi_\delta(\lambda), \end{aligned} \quad (5.7)$$

where χ_1^2 is a chi-squared random variable with 1 degree of freedom.

If $\psi_\delta(\lambda) \geq \psi_\delta(\gamma)$ for all sufficiently large δ , then by (5.6) and (5.7)

$$\mathbf{P}(\chi_1^2 \leq \lambda_k \delta^2) > \psi_\delta(\lambda) \geq \psi_\delta(\gamma) \geq \mathbf{P}(\chi_k^2 \leq \gamma_k \delta^2)$$

for all $\delta > \delta_0$, where δ_0 is sufficiently large. Since $\mathbf{P}(\chi_1^2 \leq \lambda_k \delta^2) > \mathbf{P}(\chi_k^2 \leq \gamma_k \delta^2)$ for all $\delta > \delta_0$, it follows from Lemma 5.1 that $\lambda_k \geq \gamma_k$.

Now let us assume that $\lambda_k > \gamma_k$. The same arguments used in derivation of the inequalities (5.6) and (5.7) also yield the inequalities

$$\psi_\delta(\lambda) \geq \mathbf{P}(\chi_k^2 \leq \lambda_k \delta^2) \quad (5.8)$$

and

$$\mathbf{P}(\chi_1^2 \leq \gamma_k \delta^2) > \psi_\delta(\gamma). \quad (5.9)$$

If $\lambda_k > \gamma_k$, then Lemma 5.1 yields the inequality

$$P(\chi_k^2 \leq \lambda_k \delta^2) > P(\chi_1^2 \leq \gamma_k \delta^2) \quad (5.10)$$

for sufficiently large δ . Hence the inequality $\psi_\delta(\lambda) > \psi_\delta(\gamma)$ follows from (5.8)–(5.10) for all sufficiently large δ . This completes the proof of part (b). \square

Note that nothing can be said about the relationship between $\psi_\delta(\lambda)$ and $\psi_\delta(\gamma)$ for all sufficiently small $\delta > 0$, if $|\mathbf{M}(\xi_1)| = |\mathbf{M}(\xi_2)|$ in part (a) of Theorem 5.1. Theorem 5.1 also shows that the $\text{DS}(\varepsilon)$ -criterion is equivalent to the D-criterion as $\varepsilon \rightarrow 0$, and the $\text{DS}(\varepsilon)$ -criterion is equivalent to the E-criterion as $\varepsilon \rightarrow \infty$. These conclusions are due to the fact that both D- and E-optimal designs are unique (cf. Hoel 1958, Pukelsheim and Studden 1993).

Acknowledgements: This paper was completed while the first author was a senior research fellow of the Academy of Finland and the second author a doctoral student at Tampere Graduate School in Information Science and Engineering. The work was started when the third author was visiting the University of Tampere in 1997 and a substantial part of the paper was prepared during his second visit in 1998. He would like to thank his colleagues for their excellent hospitality. Both visits were supported by the Academy Finland project 38113 and partly by the University of Tampere. The authors are grateful to Bikas K. Sinha who brought the distance optimality criterion to our attention.

References

- Anderson TW (1955) The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society* 6:170–176
- Birnbaum ZW (1948) On random variables with comparable peakedness. *Annals of Mathematical Statistics* 19:76–81.
- Chow KL, Lii K-S (1993) D-optimal designs with Hadamard matrix. *Journal of Mathematical Analysis and Applications* 179:76–90
- De la Garza A (1954) Spacing of information in polynomial estimation. *Annals of Mathematical Statistics* 25:123–130
- Hoel PG (1958) Efficiency problems in polynomial estimation. *Annals of Mathematical Statistics* 29:1134–1145
- Horn RA, Johnson CR (1987) *Matrix analysis*. Cambridge University Press, New York
- Liski EP, Luoma A, Mandal NK, Sinha BK (1998) Pitman nearness, distance criterion and optimal regression designs. *Calcutta Statistical Association Bulletin* 48(191–192):179–194
- Marshall AW, Olkin I (1979) *Inequalities: Theory of majorization and its applications*. Academic Press, New York
- Pukelsheim F (1993) *Optimal design of experiments*. John Wiley & Sons, Inc., New York
- Pukelsheim F, Studden WJ (1993) E-optimal designs for polynomial regression. *Annals of Statistics* 21, 1:402–415
- Shah KR, Sinha BK (1989) *Theory of optimal designs*. Springer-Verlag Lecture Notes in Statistics Series, No. 54
- Sherman S (1955) A theorem on convex sets with applications. *Annals of Mathematical Statistics* 26:763–766
- Sinha BK (1970) On the optimality of some designs. *Calcutta Statistical Association Bulletin* 20:1–20
- Stępnik C (1989) Stochastic ordering and Schur-convex functions in comparison of linear experiments. *Metrika* 36:291–298
- Tong YL (1990) *The multivariate normal distribution*. Springer-Verlag, New York