ARTO LUOMA

Optimal Designs in Linear Regression Models

University of Tampere
Tampere 2000

Optimal Designs in Linear Regression Models

Acta Universitatis Tamperensis 747

ACADEMIC DISSERTATION
University of Tampere,
Department of Mathematics, Statistics and Philosophy
Finland

Distribution

University of Tampere
Sales Office
P.O. Box 617
33101 Tampere
Finland

Tel. +358 3 215 6055
Fax +358 3 215 7150
taju@uta.fi
http://granum.uta.fi

Cover design by
Juha Siro

Printed dissertation
Acta Universitatis Tamperensis 747
ISBN 951-44-4817-0
ISSN 1455-1616

Electronic dissertation
Acta Electronica Universitatis Tamperensis 33
ISBN 951-44-4818-9
ISSN 1456-954X
http://acta.uta.fi

Tampereen yliopistopaino Oy Juvenes Print
Tampere 2000

ARTO LUOMA

Optimal Designs in Linear Regression Models

■

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Economics and Administration of
the University of Tampere, for public discussion
in the auditorium A 1 of the main building
of the University of Tampere, Kalevantie 4,
Tampere, on May 25th, 2000, at 12 o'clock.

University of Tampere
Tampere 2000

Contents

OPTIMAL DESIGNS IN LINEAR REGRESSION MODELS

Arto Luoma

ABSTRACT

The thesis consists of five papers and a summary. It has two main themes in the area of experimental design. Firstly, optimal designs for estimation, prediction and inverse prediction are developed in random coefficient linear regression models. Mainly first-degree models are dealt with. The results can be applied in repeated measurements situations.

Secondly, a relatively new design criterion, the distance optimality criterion, is introduced. Its properties are studied and some results in estimation and prediction problems deduced. The criterion can be used to find optimal designs or to improve designs in polynomial regression models. It has interesting relations to traditional D- and E-optimality criteria.

Key words and phrases: repeated measurements, information matrix, stem curves, variance components, regression model, Schur-concavity, polynomial designs, A-, D- and E-optimality, distance optimality.

Preface

I had my first contact with the subject of my thesis in summer 1994 when professor Bikas Sinha from the Indian Statistical Institute visited our department. He had already cooperated with my supervisor Erkki Liski in experimental design and I had a joy to read one of their joint articles. They also welcomed me in their research group. They had developed optimal designs in growth curve models with intraclass correlation structure and a natural extension to their study were random coefficient linear regression models.

There was also another reason why we started developing optimal designs in such models. In our department we had the project ”Developing an integrated system for forest harvesting”, financed by the Ministry of Education, and we were therefore concentrating on forest applications. We had received from Kajaani-automatiikka lots of tree stem data collected by forest harvesters and somebody had observed that variation in tree stems could be nicely modeled using random coefficient regression models.

In summer 1994 the preliminary drafts of the three first papers of the thesis were written. They provided a basis on which some future work could be done. It was a credit to Sinha that he was able to formulate the problems in such a form that they could be dealt with. However, it was not easy to find complete solutions to some problems which seemed simple at the outset. There was also a third person to join our research group, Dr. Nripesh Mandal from Calcutta University, who visited our department twice.

Professor Sinha had many short visits in our department and three times he presented us the DS optimality criterion which he had discovered. He suggested that we would study the use of this criterion in the theory of experimental design. Finally, in summer 1997, during his visit to our department, some preliminary work was done and a draft of the fourth paper of the thesis was born.

It was Dr. Alexander Zaigraev from Nicholas Copernicus University of Toruń with whom Liski and I continued the study of the DS criterion after Sinha’s departure. I had obtained his address from professor Fellman after giving a presentation in one graduate seminar, and we had started scientific correspondence. Since in our department there was a project ”Statistical methods in developing an integrated control system for forest harvesting”, financed by the Academy of Finland, Zaigraev could visit us during the autumns 1997 and 1998. The results of our cooperation are found in the fifth article of the thesis.

I want to express my gratitude to professor Liski who patiently encouraged me during my postgraduate studies. I am also thankful to professor Sinha who has been an example of creative scientific work and whose enthusiasm as a teacher I have admired. Likewise I wish to thank Dr. Zaigraev with whom I could experience some mathematical adventures during our intensive cooperation and who also provided me an example of scientific work. I thank Dr. Mandal from whom I have also learned something. We also obtained inspiration from our contacts in the area of forestry. I warmly thank the referees of the thesis, professors Friedrich Pukelsheim and Kirti Shah, as well as my opponent Johan Fellman for their contribution.

I owe thanks to Robert MacGilleon, who kindly checked the language of the thesis.

I am grateful to Tampere Graduate School in Information Sciences and Engineering, since as its student I could concentrate on scientific work and postgraduate studies in full time. I also received financing from the Academy of Finland, Ministry of Education, TEKES, Tampere University Foundation and University of Tampere.

Finally, I thank the Lord who strengthened me and gave success to my work. I thank all the personnel of the department for their support as well as my parents and friends in church.

Arto Luoma Tampere, May 2000

List of original papers

The thesis comprises the following five original papers:

[1] Liski, Erkki P., Luoma, A. and Sinha, Bikas K., (1996). Optimal Designs in Random Coefficient Linear Regression Models. *Calcutta Statistical Association Bulletin*, **46**, 211-230.

[2] Liski, Erkki P., Luoma, A., Mandal, N. K. and Sinha, Bikas K. (1998). Optimal Designs for Prediction in Random Coefficient Linear Regression Models. *Journal of Combinatorics, Informatics and Systems Sciences* (J. N. Srivastava Felicitation Volume) 23(1-4), 1-16.

[3] Liski, Erkki P., Luoma, A., Mandal, N. K. and Sinha, Bikas K. (1997). Optimal Design for an Inverse Prediction Problem under Random Coefficient Regression Models. *Journal of the Indian Society of Agricultural Statistics* (Golden Jubilee Volume), **XLIX**, 277–288.

[4] Liski, E. P., Luoma, A., Mandal, N. K. and Bikas K. Sinha (1998). Pitman nearness, distance criterion and optimal regression designs. *Calcutta Statistical Association Bulletin* **48** (191-192), 179–194.

[5] Liski, E. P., Luoma, A. and Zaigraev, A. (1999). Distance optimality design criterion in linear models. *Metrika* 49(3): 193-211.

1 Introduction

The first three papers included here deal with optimum designs in random coefficient regression models (RCR-models); the other two study the distance optimality criterion.

The set of possible experimental conditions is called the *experimental domain* or *experimental region*. The thesis deals mainly with cases in which the experimental conditions may assume values continuously on a given interval. In paper [1], however, only integer values on a given interval are accepted and in paper [5] there is an example in which the experimental domain is a ball or cube in *m*-dimensional Euclidean space. *The experimental design* is defined by a finite number of values, called support points, which belong to the experimental domain, and by the proportions of observations assigned to them.

Random coefficient regression models are used in situations when repeated measurements are taken on several individuals. The design may vary from individual to individual and the overall design must include allocation of observations to each individual.

Throughout the thesis, the proportions of observations may assume values continuously on the interval [0,1]. Usually, this kind of design cannot be realized exactly with a finite number of observations and is therefore called *approximate.* The total number of measurements, denoted by N in repeated measurement situations and otherwise by n, is kept fixed, and the objective of the experimental design is to find an optimal allocation of them in the experimental domain.

Comparison of designs is based on the dispersion matrix of the unknown parameter vector. Alternatively, the starting-point may be the *information matrix*, defined as the inverse of the dispersion matrix. It is equivalent to the *Fisher information matrix* of the unknown parameter vector, provided that the best linear unbiased estimator (BLUE) is used and the vector of random terms is assumed to follow the multivariate normal distribution.

In the case of the classical linear model, $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, considerations are usually based on the *moment matrix* of the design, defined by

$$\mathbf{M} = \sum_i \xi_i \mathbf{x}_i \mathbf{x}_i',$$

where $\mathbf{x}_i$ is a regression vector and ξ_i the proportion of observations assigned to it. The information matrix is then defined using the moment matrix in such a way that the matrices are equal in the case of the full paramater system (cf. Pukelsheim, p.

63). Thus the definition is independent of n, the number of observations, and σ^2, the variance of error terms.

Comparison of information matrices (or equivalently dispersion matrices) is made using Loewner ordering. A matrix $\mathbf{A}$ is greater than or equal to a matrix $\mathbf{B}$, denoted as $\mathbf{A} \geq \mathbf{B}$, if $\mathbf{A} - \mathbf{B}$ is nonnegative definite. Since the Loewner ordering is only partial, it is not sufficient for comparison of designs. Only in rare cases can a design be found such that its information matrix is greater than or equal to the information matrix of any competing design. Hence scalar valued optimality criteria are needed. These are actually mappings from the set of information matrices to the real line, such that they preserve the Loewner ordering.

Matrix means provide a fairly large class of optimality criteria. Denote by $\lambda_1, \lambda_2, \ldots, \lambda_k$ the eigenvalues of a positive definite matrix $\mathbf{C}$. Then the matrix mean $\phi_p(\mathbf{C})$ is defined by

$$\phi_p(\mathbf{C}) = (\frac{1}{k} \sum_{i=1}^{k} \lambda_i^p)^{\frac{1}{p}}.$$

Matrix means are used as optimality criteria as $p \leq 1$. A definition which covers all nonnegative definite matrices is to be found in Pukelsheim (Section 6.7). The classical A-, E-, D- and T-optimality criteria are obtained from ϕ_p as special or limiting cases as ϕ_p is applied to information matrices. Criterion ϕ_{-1} is the A-optimality criterion. Maximizing the criterion is equivalent to minimizing the trace of the dispersion matrix. Criterion $\phi_{-\infty} = \lim_{p \to -\infty} \phi_p = \lambda_{min}$, the smallest eigenvalue of the information matrix, is called the E-optimality criterion and $\phi_0 = \lim_{p \to 0} \phi_p = (\prod_i \lambda_i)^{\frac{1}{k}}$ the D-optimality criterion. Maximizing the D criterion is equivalent to minimizing the determinant of the dispersion matrix. Criterion $\psi_1 = \frac{1}{k}\text{tr } \mathbf{C}$ is called the T or trace criterion.

2 Random Coefficient Linear Regression Models

In the first three papers of the thesis, random coefficient linear regression models are addressed. The models are applied in repeated measurement situations, where one or more measurements are taken on one or more individuals (units). The observations within an individual are correlated while two observations from different individuals are uncorrelated.

Let $\mathbf{y}_i$ be the $n_i \times 1$ response vector for the ith individual. Then the RCR model is of the form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{X}_i$ is the $n_i \times k$ model matrix and the vector $\boldsymbol{\beta}$ contains the fixed effects, $\mathbf{b}_i$ the random effects and $\boldsymbol{\varepsilon}_i$ the error terms. The dispersion matrices of $\mathbf{b}_i$ and ε_i are denoted by

$$D(\mathbf{b}_i) = \mathbf{D} \text{ and } D(\boldsymbol{\varepsilon}_i) = \sigma^2\mathbf{I}$$

for every $i = 1, 2, \ldots, n$. It is assumed that $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are uncorrelated. Consequently,

$$D(\mathbf{y}_i) = \mathbf{X}_i\mathbf{D}\mathbf{X}_i' + \sigma^2\mathbf{I} \doteq \boldsymbol{\Sigma}_i.$$

Since the responses of different individuals are uncorrelated,

$$D(\mathbf{y}) = \text{diag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_n),$$

where the vector $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_n')'$ contains the responses of all individuals. The GLS estimator (which is also BLUE) for $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} = (\sum_i \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i)^{-1} \sum_i \mathbf{X}'\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i.$$

In a case where the design for all individuals is the same, the expression for $\widehat{\boldsymbol{\beta}}$ can be simplified to the form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}, \tag{1}$$

where $\mathbf{X}$ and $\boldsymbol{\Sigma}$ stand for the common matrices $\mathbf{X}_i$ and $\boldsymbol{\Sigma}_i$, respectively, and $\bar{\mathbf{y}} = \frac{1}{n}\sum_i \mathbf{y}_i$ is the average of the individual response vectors. However, it can be shown that the GLS estimator (1) is equal to the OLS estimator

$$\widehat{\boldsymbol{\beta}}_{OLSE} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{y}},$$

since $D(\bar{\mathbf{y}}) = \frac{1}{n}\boldsymbol{\Sigma}$ has the so-called Rao simple structure with

$$\boldsymbol{\Sigma} = \mathbf{X}\mathbf{D}\mathbf{X}' + \sigma^2\mathbf{I} \tag{2}$$

(cf. Rao 1967). To prove that the GLSE (1) and OLSE are equal, let us choose a matrix $\mathbf{Z}$ such that it has full column rank and its columns span the orthogonal complement of the column space of $\mathbf{X}$. Then, using the identity

$$\mathbf{I} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}',$$

$\mathbf{\Sigma}$ can be written in the form

$$\mathbf{\Sigma} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}' + \mathbf{Z}\mathbf{\Omega}\mathbf{Z}',$$

where

$$\mathbf{\Lambda} = \mathbf{D} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}, \quad \mathbf{\Omega} = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}.$$

Substituting the expression

$$\mathbf{\Sigma}^{-1} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{\Lambda}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{\Omega}^{-1}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$$

in (1) it will be seen that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{OLSE}$ are equal.

The dispersion matrix of the OLS-estimator is

$$D(\widehat{\boldsymbol{\beta}}_{OLSE}) = \frac{1}{n}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n}\mathbf{D} + \frac{\sigma^2}{N}\mathbf{M}^{-1},$$

where $\mathbf{M} = \frac{1}{m}\mathbf{X}'\mathbf{X}$, m being the number of observations within one unit. It is clear that if n, the number of individuals measured, is fixed, the problem of minimizing tr $D(\widehat{\boldsymbol{\beta}})$ is reduced to minimizing tr $\mathbf{M}^{-1}$, and, consequently, the solution is the same as in the fixed parameter case. Thus, the dispersion matrix of the random vector has no effect on the design, as far as A-optimality is concerned. This is applied to prediction problems in papers [2] and [3].

If n is not fixed, it would appear profitable to measure as many individuals as possible to minimize $D(\widehat{\boldsymbol{\beta}})$. However, since the total number of observations N is fixed, the number of observations for each individual, $m = \frac{N}{n}$, would be minimized and usually the optimal allocation of observations within an individual would not be achieved even approximately. It turns out that the optimal design depends on the values of $\sigma^{-2}\mathbf{D}$. There is an example of this in paper [2], again in the context of predicton.

If the design is not the same for all individuals, the formulae including $\bar{\mathbf{y}}$ can no longer be used. This is the case in paper [1] and in one example in paper [2].

3 Distance Optimality Design Criterion

Sinha (1970, pp. 1-20) suggested a new kind of optimality criterion, called the distance criterion, when he studied the optimal allocation of observations in the 1-way ANOVA model. In papers [4] and [5], the criterion is studied in the context of the classical linear model $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, assuming that the error terms are normally distributed.

A design is said to be distance optimal, or DS-optimal, where D stands for 'distance' and S for 'stochastic', if it maximizes the probability

$$\Pr(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| < \varepsilon) \tag{3}$$

simultaneously for all $\varepsilon > 0$, where $\widehat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$ and $\|\cdot\|$ denotes the Euclidean distance. Such designs are found when first-degree polynomial fit models are used. In many situations, however, the distance optimality property is too strong and there exists no distance optimal design. Then the probability (3), called the DS(ε)-criterion function, can be maximized for fixed $\varepsilon > 0$. When all values $\varepsilon > 0$ are taken into account, a class of optimality criteria is obtained.

As is the case with the ϕ_p criterion, the $DS(\varepsilon)$ criterion can also be expressed using eigenvalues. The $DS(\varepsilon)$ criterion function $\psi_\varepsilon(\boldsymbol{\lambda})$ is defined by

$$\psi_\varepsilon(\boldsymbol{\lambda}) = \Pr(\frac{Z_1^2}{\lambda_1} + \frac{Z_2^2}{\lambda_2} + \ldots + \frac{Z_k^2}{\lambda_k} \leq \frac{n\varepsilon^2}{\sigma^2}), \tag{4}$$

where Z_i are independent random variables following the standard normal distribution, λ_i are the eigenvalues of the moment matrix $\mathbf{M}$ and n is the number of observations. The classical D- and E-optimality criteria are obtained from the $DS(\varepsilon)$ criterion as limiting cases. Maximizing ψ_ε as $\varepsilon \to 0$ or $\varepsilon \to \infty$ is equivalent to maximizing $\det \mathbf{M}$ or $\lambda_{min}(\mathbf{M})$, respectively. The results are proved in paper [5].

The matrix means $\phi_p(\mathbf{C})$, $p \leq \infty$ are concave functions of positive definite matrices $\mathbf{C}$ (cf. Pukelsheim, 6.13). This is a useful property, and can be applied to prove that symmetrization improves design in polynomial regression models, for example (cf. Pukelsheim, 13.1). Unfortunately, the DS criterion does not possess this property, as is shown in paper [5]. However, it does possess a weaker property, Schur concavity.

Before defining Schur concavity, the concept of vector majorization is given. Denote by $x_{[1]}, x_{[2]}, \ldots, x_{[k]}$ the elements of a vector $\mathbf{x}$ in decreasing order and by $x_{(1)}, x_{(2)}, \ldots, x_{(k)}$ the same elements in increasing order. A vector $\mathbf{a} \in \mathbf{R}^k$ is said to be majorized by $\mathbf{b}$, denote $\mathbf{a} \prec \mathbf{b}$, if

$$\begin{aligned} \sum_{i=1}^{j} a_{[i]} &\leq \sum_{i=1}^{j} b_{[i]}, \quad j = 1, 2, \ldots, k-1 \\ \sum_{i=1}^{k} a_{[i]} &= \sum_{i=1}^{k} b_{[i]}, \end{aligned}$$

or equivalently,

$$\begin{aligned} \sum_{i=1}^{j} a_{(i)} &\geq \sum_{i=1}^{j} b_{(i)}, \quad j = 1, 2, \ldots, k-1 \\ \sum_{i=1}^{k} a_{(i)} &= \sum_{i=1}^{k} b_{(i)}. \end{aligned}$$

It can be shown (Muirhead, 1903; Hardy, Littlewood, and Polya, 1934, 1952, p. 47; see also Marshall and Olkin, Lemma 2.B.1) that $\mathbf{b}$ can be obtained from $\mathbf{a}$ by making at most $k-1$ transformations of the type

$$\mathbf{a}^{new} = (a_1, \ldots, a_{i-1}, \lambda a_i + (1-\lambda) a_j, a_{i+1}, \ldots, a_{j-1}, \lambda a_j + (1-\lambda) a_i, a_{j+1}, \ldots, a_k), \tag{5}$$

where $0 \leq \lambda \leq 1$.

Majorization implies weak supermajorization. A vector $\mathbf{a}$ is said to be supermajorized weakly by $\mathbf{b}$, denote $\mathbf{a} \prec^w \mathbf{b}$, if

$$\sum_{i=1}^{j} a_{(i)} \geq \sum_{i=1}^{j} b_{(i)}, \quad j = 1, 2, \ldots, k.$$

Using induction, it can be shown that if $\mathbf{a} \prec^w \mathbf{b}$, then there exists a $\mathbf{c}$ such that $\mathbf{c} \leq \mathbf{a}$ (no element of $\mathbf{c}$ is greater than the corresponding element of $\mathbf{a}$) and $\mathbf{c} \prec \mathbf{b}$ (Marshall and Olkin, 5.A.9.a). Furthermore, it can be shown that $\mathbf{a} \prec^w \mathbf{b}$ iff

$$\sum_{i=1}^{k} f(a_{(i)}) \leq \sum_{i=1}^{k} f(b_{(i)})$$

for all continuous decreasing convex functions f (Tomić, 1949; see also Marshall and Olkin, 4.B.2).

A real-valued function f defined on a set $\mathcal{A} \in \mathbf{R}^k$ is said to be *Schur concave* if $\mathbf{x} \prec \mathbf{y}$ on $\mathcal{A}$ implies $f(\mathbf{x}) \geq f(\mathbf{y})$. By Schur concavity of the $DS(\varepsilon)$ criterion we mean that $\psi_\varepsilon(\boldsymbol{\lambda})$ is a Schur concave function of $\boldsymbol{\lambda}$. To prove the result, we assume that $\boldsymbol{\lambda} \prec \boldsymbol{\mu}$ and show that $\psi_\varepsilon(\boldsymbol{\lambda}) \geq \psi_\varepsilon(\boldsymbol{\mu})$. Since $f(x) = -\ln x$ is a continuous decreasing convex function of x, it follows from the weak supermajorization $\boldsymbol{\lambda} \prec^w \boldsymbol{\mu}$ that $\ln \boldsymbol{\lambda} \prec^w \ln \boldsymbol{\mu}$, where $\ln \boldsymbol{\lambda} = (\ln \lambda_1, \ldots, \ln \lambda_k)'$ and $\ln \boldsymbol{\mu} = (\ln \mu_1, \ldots, \ln \mu_k)'$. Since $\psi_\delta(\boldsymbol{\lambda})$ is an increasing function of the elements of $\boldsymbol{\lambda}$, it is sufficient to show that $\ln \boldsymbol{\gamma} \prec \ln \boldsymbol{\mu}$ implies $\psi_\varepsilon(\boldsymbol{\gamma}) \geq \psi_\varepsilon(\boldsymbol{\mu})$, where $\ln \boldsymbol{\gamma} = (\ln \gamma_1, \ldots, \ln \gamma_k)'$ is a vector such that $\ln \boldsymbol{\gamma} \leq \ln \boldsymbol{\lambda}$ and $\ln \boldsymbol{\gamma} \prec \ln \boldsymbol{\mu}$. The

vector $\ln \boldsymbol{\gamma}$ is obtained from $\ln \boldsymbol{\mu}$ using transformations (5), which change two elements of the vector at a time. Thus it suffices to prove the claim in the case $k = 2$. This is actually done in paper [4], Lemma 2.

Schur concavity can be used to prove that in polynomial regression models the $DS(\varepsilon)$-optimality criterion function ψ_ε increases if the design is symmetrized. There is a proof in paper [5], but to render the role of Schur concavity more transparent, an alternative proof is given.

Assume that the experimental domain is the interval [-1,1] and we have the dth degree polynomial fit model $y_x = \beta_0 + \beta_1 x + \ldots + \beta_d x^d + \epsilon_x$. The original design is described by a function $\tau(t)$ from [-1,1] to [0,1], which gives the proportions of observations at the support points of the design and is zero otherwise. The reflected design τ^R is described by the function $\tau^R(t) = \tau(-t)$ and the symmetrized design by $\bar{\tau} = \frac{1}{2}(\tau + \tau^R)$. The moments of the original design are $\mu_i = \sum_j \tau(t_j) t_j^i$, where t_j are the support points of the design. The even moments of the reflected design are equal to the corresponding moments of the original design, while the odd moments carry a reversed sign. Since $\mathbf{M}(\bar{\tau}) = \frac{1}{2}[\mathbf{M}(\tau) + \mathbf{M}(\tau^R)]$, where $\mathbf{M}(\cdot)$ denotes the moment matrix of a design, it is clear that in the symmetrized design the even moments are equal to the corresponding moments of the original design and the odd moments vanish. Since $\mathbf{M}(\tau^R)$ is obtained from $\mathbf{M}(\tau)$ by the similarity transformation

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

where $\mathbf{Q} = \mathrm{diag}(1, -1, \ldots, (-1)^d)$, it has the same eigenvalues as $\mathbf{M}(\tau)$.

The sum $\sum_{i=1}^{j} \mathbf{z}_i' \mathbf{M}(\tau) \mathbf{z}_i$, where the vectors $\mathbf{z}_i$ are orthonormal, is maximized by choosing as $\mathbf{z}_i$ the eigenvectors corresponding to j largest eigenvalues of $\mathbf{M}$. (Proof for Hermitian matrices is given in Fan, 1950; see also Marshall and Olkin, 20.A.2.) Denote by $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d+1})$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{d+1})$ the eigenvalue vectors of $\mathbf{M}(\bar{\tau})$ and $\mathbf{M}(\tau)$, respectively. We observe that

$$\sum_{i=1}^{j} \lambda_i = \sum_{i=1}^{j} \mathbf{z}_i' \mathbf{M}(\bar{\tau}) \mathbf{z}_i = \frac{1}{2} \sum_{i=1}^{j} \mathbf{z}_i' \mathbf{M}(\tau) \mathbf{z}_i + \frac{1}{2} \sum_{i=1}^{j} \mathbf{z}_i \mathbf{M}(\tau^R) \mathbf{z}_i \leq \sum_{i=1}^{j} \mu_i$$

for $j = 1, 2, \ldots, d$. Since $\mathrm{tr}\, \mathbf{M}(\tau) = \mathrm{tr}\, \mathbf{M}(\bar{\tau})$, we have $\sum_{i=1}^{d+1} \lambda_i = \sum_{i=1}^{d+1} \mu_i$. Thus $\boldsymbol{\lambda} \prec \boldsymbol{\mu}$, and the Schur concavity of the $DS(\varepsilon)$ criterion implies that the symmetrized design is better.

4 Summaries of papers

4.1 Optimal Designs in Random Coefficient Linear Regression Models

The paper deals with a repeated measurements situation, in which individuals are measured at the integer time points belonging to the interval

$$-k, -k+1, \ldots, k-1, k. \tag{6}$$

The measurements must be *consecutive*; thus the design $-2, -1, 0, 1$ is allowed for an individual while $-2, -1, 1$ is not. The first order RCR-model

$$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + \varepsilon_{ij}, \tag{7}$$

where the parameters $\beta_0,\ \beta_1$ are fixed and $b_{0i},\ b_{1i}$ random effects, is adopted for the ith individual. The random terms b_{0i}, b_{1i} and ε_{ij} are assumed to be uncorrelated for all i and j. Their variances, denoted by $\sigma_0^2 = V(b_{0i})$, $\sigma_1^2 = V(b_{1i})$ and $\sigma_\varepsilon^2 = V(\varepsilon_{ij})$, are assumed to be known.

In the paper, two design problems are solved. Optimal designs are found to 1) estimate the slope parameter β_1 and 2) predict at the time points (6). In both cases, symmetrization improves design. Considerations are therefore restricted to three types of symmetric design: 1) singleton designs, 2) single-run designs and 3) two-run designs. In singleton designs, each individual is measured only once. A proportion $p\ (\leq \frac{1}{2})$ of measurements is taken at both time points $-n$ and n, $1 \leq n \leq k$, and a proportion $1 - 2p$ at 0. In single-run designs, all individuals in the experiment are measured at the time points

$$-n, -n+1, \ldots, n-1, n.$$

In two-run designs, half of the individuals are measured at the points

$$m, m+1, \ldots, n-1, n,$$

where $-k \leq m \leq n \leq k$, and the second half at

$$-n, -n+1, \ldots, -m-1, -m.$$

In slope parameter estimation, analytic solutions for singleton and single-run cases are found. The solutions depend on k and the ratios $\gamma = \frac{\sigma_0^2}{\sigma_\varepsilon^2}$ and $\delta = \frac{\sigma_1^2}{\sigma_\varepsilon^2}$.

When different design types are compared, either singleton or two-run designs are optimal. Single-run designs are always improved by choosing the two-run design which has the same run length. For comparison of singleton and two-run designs, a method is suggested in which the two-run design is approximated using the corresponding single-run design. Tables are also given which indicate the optimal design.

In the prediction problem, the trace of the dispersion matrix is minimized for the predictor. Thus, the A-optimality criterion is used. It turns out that the symmetric singleton design with the support points $-k$ and k is always optimal. If the values of k, δ and γ satisfy the relation (3.6) (in the paper), the solution is also unique; otherwise there is an infinite number of optimal singleton designs.

The result that the singleton design is optimal is not proved in detail, since it is lengthy and demands only simple calculus. The claim in the proof stating that (3.8) is an increasing function of γ does not hold for all values of k, δ and Δ.

Finally, an example motivated by our forest applications is given to illustrate the results of the paper.

4.2 Optimal Designs for Prediction in Random Coefficient Linear Regression Models

In this paper there is once more a repeated measurements situation. The first-degree RCR model (7) is used with the same assumptions on the random terms as in paper [1]. The measurements are taken on the interval $[0, h]$ and predictions made on $(h, H]$. The problem originated from the situation where diameters of trees can be measured up to height h and should be predicted up to height H. Also the situation is tackled where the predicted region includes the experimental region.

First it is necessary to define what is meant by prediction. Denote by

$$\mathbf{x}_{H-h} = (h+1, h+2, \ldots, H)'$$

the points at which the predictions are made and by

$$\mathbf{X}_{H-h} = (\mathbf{1}\ \mathbf{x}_{H-h})$$

the corresponding model matrix. Assume that we have measured n individuals on the interval $[0, h]$. Then the object of prediction can be

(i) the response of the nth individual on $(h, H]$,
$\mathbf{y}_{H-h,n} = \mathbf{X}_{H-h}(\boldsymbol{\beta} + \mathbf{b}_n) + \boldsymbol{\varepsilon}_{H-h,n}$,

(ii) a new individual, $\mathbf{y}_{H-h,n+1} = \mathbf{X}_{H-h}(\boldsymbol{\beta} + \mathbf{b}_{n+1}) + \boldsymbol{\varepsilon}_{H-h,n+1}$,

(iii) the expectation of observations, $\mathbf{y}_{H-h} = \mathbf{X}_{H-h}\boldsymbol{\beta}$,

(iv) the average of the individuals measured, $\bar{\mathbf{y}}_{H-h} = \mathbf{X}_{H-h}(\boldsymbol{\beta} + \bar{\mathbf{b}}) + \bar{\boldsymbol{\varepsilon}}_{H-h}$ where $\bar{\mathbf{b}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{b}_i$ and $\bar{\boldsymbol{\varepsilon}}_{H-h} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\varepsilon}_{H-h,i}$.

Case (i) is not dealt with in the paper. In case (iii), it is clear that the optimal predictor (BLUP) is $\hat{\mathbf{y}}_{H-h} = \mathbf{X}_{H-h}\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is optimal for $\boldsymbol{\beta}$. As was mentioned in Section 2 of this summary, $\widehat{\boldsymbol{\beta}}$ is equal to $\widehat{\boldsymbol{\beta}}_{OLSE}$, if the design is the same for all individuals. In case (ii), the optimal predictor is the same as in (iii), since the expectation of $\mathbf{y}_{H-h,n+1}$ is $\mathbf{y}_{H-h}$ and the random terms $\mathbf{b}_{n+1}$ and $\boldsymbol{\varepsilon}_{H-h,n+1}$ are uncorrelated with the measurements. In Section 3.1 of the paper it is shown that the predictor $\mathbf{X}_{H-h}\widehat{\boldsymbol{\beta}}$ is also optimal in case (iv) if the design is the same for all individuals.

Three types of design are investigated. Firstly, it is assumed that all individuals have the same design and k, the number of measurements per individual, is fixed. The result is that all observations should be taken at the extremes 0 and h. The proportion of observations at 0 depends on h and H but not on the variance components, which could have been deduced in the way pointed out in Section 2 of this summary.

In the second type of design, all individuals have the same design, but k is no longer fixed. It emerges that the design depends on the ratios $\delta_0 = \frac{\sigma_0^2}{\sigma_\varepsilon^2}$ and $\delta_1 = \frac{\sigma_1^2}{\sigma_\varepsilon^2}$. The optimal design is rather difficult to characterize, since an infinite number of alternatives should be compared. Usually, only few measurements should be taken on one individual. Relation (3.15) is given to determine which one of the alternatives $(k_1, k_2) = (1, 1)$ and $(k_1, k_2) = (2, 1)$ is better, k_1 and k_2 being the numbers of observations taken at 0 and h, respectively.

Thirdly, an optimal singleton design is found in which the design is not the same for all units. Comparison is made with a design where two measurements are taken on all units.

The optimality criterion for the predictor is the trace of the dispersion matrix (A optimality). When studying designs of the first type, also IMSE and minimax criteria are applied. The problem of prediction at all unobservable data points is also taken up.

The paper ends with an application to tree stem data collected by a forest harvester.

4.3 Optimal Design for An Inverse Prediction Problem under Random Coefficient Regression Models

The main part of the paper deals with an inverse prediction problem under fixed first- and second-degree polynomial fit models. The experimental region is $[0, h]$ and the aim is to predict the value $x = x_0$ at which the polynomial $\eta(x)$ attains a pre-specified level η_0. In forestry applications, $\eta(x)$ might be a stem curve giving the diameter of a tree stem at height x and η_0 the minimum diameter of logs.

The value x_0 is estimated by plug-in estimators

$$\widehat{x}_0 = \frac{\eta_0 - \widehat{\beta}_0}{\widehat{\beta}_1}$$

and

$$\widehat{x}_0 = \frac{-\widehat{\beta}_1 - \sqrt{\widehat{\beta}_1^2 - 4\widehat{\beta}_2(\widehat{\beta}_0 - \eta_0)}}{2\widehat{\beta}_2}$$

in the first- and second-degree models, respectively, where $\widehat{\beta}_i$ are components of the ordinary least squares estimator.

Since the estimators are not linear in $\widehat{\beta}_i$, the δ method is used to obtain approximate variances for $\widehat{x}$. Expanding the estimator as a Taylor series about the true parameter value $\boldsymbol{\beta}$, we establish that

$$\widehat{x}_0 = x_0(\widehat{\boldsymbol{\beta}}) \approx x_0(\boldsymbol{\beta}) + \frac{\partial x_0}{\partial \boldsymbol{\beta}'}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

for small differences $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ and, consequently,

$$V(\widehat{x}_0) \approx \frac{\partial x_0}{\partial \boldsymbol{\beta}'} D(\widehat{\boldsymbol{\beta}}) \frac{\partial x_0}{\partial \boldsymbol{\beta}}.$$

After some calculations it will be seen that

$$V(\widehat{\mathbf{x}}_0) \approx \frac{1}{\beta_1^2} \mathbf{x}_0' D(\widehat{\boldsymbol{\beta}}) \mathbf{x}_0$$

or

$$V(\widehat{\mathbf{x}}_0) \approx \frac{1}{(\beta_1 + 2\beta_2 x_0)^2} \mathbf{x}_0' D(\widehat{\boldsymbol{\beta}}) \mathbf{x}_0$$

in the cases of the first- and second-degree models, respectively. Here $\mathbf{x}_0$ denotes the regression vector evaluated at x_0. It is unknown, but the expectation matrix $\mathcal{E}\mathbf{x}_0\mathbf{x}_0'$ is assumed to be known as prior information.

The design problem is reduced to minimizing the linear optimality criterion

$$L(\mathbf{M}^{-1}) = \text{tr } \mathbf{M}^{-1}\mathbf{B},$$

where $\mathbf{M}$ is the moment matrix of the design and $\mathbf{B} = \mathcal{E}\mathbf{x}_0\mathbf{x}_0'$. The equivalence theorem of the linear criterion states that a nonsingular moment matrix $\mathbf{M}_0$ is optimal iff

$$f'(\mathbf{x})\mathbf{M}_0^{-1}\mathbf{B}\mathbf{M}_0^{-1}f(\mathbf{x}) \leq \text{tr } \mathbf{M}_0^{-1}\mathbf{B}$$

for all x belonging to the experimental interval. In the case of optimality, the equality is obtained at the support points of the design. The regression vector is given by $f(\mathbf{x}) = (1\ x\ x^2)'$ in the second-degree model. The proof of the theorem is found in Fedorov(1972, Section 2.9) and Silvey (1980, Section 1.3). It also follows from the equivalence theorem for matrix means in the monograph by Pukelsheim (1993, Sections 7.20, 9.8).

The equivalence theorem is used to show that in the second-degree model, the optimal design has three support points, two of which are at the extremes of the experimental interval. As third support point the middle point of the experimental interval is chosen, although it is not exactly optimal. Regression vectors $f(\mathbf{l})$, $f(\mathbf{m})$ and $f(\mathbf{u})$ at the support points l, m and u, respectively, and vectors orthogonal to two of them are used to deduce the optimal weights for the design.

Finally, the inverse prediction problem is taken up in the context of RCR models. It is shown that the optimal design is the same as in the case of the classical linear model.

4.4 Pitman Nearness, Distance Criterion and Optimal Regression Designs

The DS Criterion was already defined in Section 3 of this summary. Pitman nearness is in a sense a predecessor of the distance criterion. It is used for comparing estimators, while the distance criterion is applied in the design of experiments.

In the paper, the first-degree regression model

$$y_x = \alpha + \beta x + \varepsilon_x$$

is assumed, where $x \in \chi = [-1, 1]$. For the error term it is assumed that $E\varepsilon_x = 0$, $V(\varepsilon_x) = \sigma^2$ and $Cov(\varepsilon_x, \varepsilon_{x'}) = 0$ for all $x \neq x'$.

Distance optimal designs are found for estimation of α and β separately and jointly. The distance criterion function for estimating the parameters jointly is the

probability

$$\Pr[(\hat{\alpha}-\alpha)^2+(\hat{\beta}-\beta)^2 \leq \epsilon^2].$$

In all cases, the design d_{0n}, which assigns equal weight to observations on the extremes -1 and 1, is found to be optimal.

Several natural variants of the distance optimality criterion are introduced for *prediction* in the interval χ:

$$\begin{aligned}
&\text{(I)} && \max_{d_n} \Pr[\max_{x\in\chi} |\hat{y}_x - y_x| < \epsilon],\\
&\text{(II)} && \max_{d_n} \min_{x\in\chi} \Pr[\,|\widehat{y}_x - y_x| < \epsilon\,],\\
&\text{(III)} && \max_{d_n} \Pr\Big[\int_{-1}^{+1} (\widehat{y}_x - y_x)^2 dx < \epsilon\Big],\\
&\text{(IV)} && \max_{d_n} E(L(d_n)),
\end{aligned}$$

where $L(d_n)$ is the length of the interval in χ under design d_n for which $|\hat{y}_x - y_x| < \epsilon$.

In cases (II) and (III) it is shown that the design d_{0n} is optimal. In case (I), it is shown that the lower limit

$$(I)' \quad \max_{d_n} \Pr[\max_{x\in\chi}\{|y-\alpha-\beta\bar{x}| + |\hat{\beta}-\beta||x-\bar{x}|\} < \epsilon]$$

is obtained by d_{0n}. Case (IV) is the most difficult and is studied using simulation.

When proving the optimality of d_{0n} in the estimation case, Okamoto's lemma (Okamoto 1960, see also Marshall and Olkin, 11.E.8.b) in the case $k = 2$ is useful. Okamoto's lemma states that

$$\Pr\Big(\frac{Z_1^2}{a_1} + \frac{Z_2^2}{a_2} + \ldots + \frac{Z_k^2}{a_k} \leq 1\Big) \leq \Pr\Big(\frac{\chi_k^2}{(\prod a_i)^{\frac{1}{k}}} \leq 1\Big), \tag{8}$$

where Z_i are independent random variables following the standard normal distribution, χ_k^2 follows the chi squared distribution with k degrees of freedom and a_i are positive constants.

When studying the prediction case (III), a generalization of Okamoto's lemma is needed in the case where $k = 2$. It states that

$$\Pr[l_1X^2 + l_2Y^2 \leq 1] \leq \Pr[m_1X^2 + m_2Y^2 \leq 1],$$

X and Y being independent standard normal variables and l_1, l_2, m_1 and m_2 positive real constants satisfying $\max(l_1, l_2) \geq \max(m_1, m_2)$ and $l_1l_2 \geq m_1m_2$.

The paper finally tackles a design prolem in which the predicted region is beyond the experimental region. As in papers [2] and [3], the experimental region is $[0, h]$ and the predicted one $(h, H]$. Using criteria (I)' and (II) with the modified experimental and predicted regions, optimal designs are found. The observations should be taken at 0 and h, but the design is no longer symmetric.

In case (III), the problem can be reduced to maximizing the probability

$$\Pr[(H-h)(Z_1\ Z_2)\mathbf{C}(Z_1\ Z_2)' \le \epsilon^2], \tag{9}$$

where $\mathbf{C}$ is a certain positive definite matrix and Z_1 and Z_2 are independent standard normal variables. The generalization of Okamoto's lemma cannot be used, since $\det \mathbf{C}$ and $\lambda_{max}(\mathbf{C})$ are not minimized by the same design. Using the results of paper [5], it is known that minimizing $\det \mathbf{C}$ corresponds to the case $\epsilon \to 0$ and minimizing $\lambda_{max}(\mathbf{C})$ to the case $\epsilon \to \infty$. Therefore, there exists no design which maximizes the expression (9) for all $\epsilon > 0$ simultaneously.

4.5 Distance Optimality Design Criterion in Linear Models

In the paper, a number of properties are established for the DS(ε) criterion. Also examples are given to elucidate its use as a design criterion.

It is noted that the DS(ε) criterion is isotonic relative to Loewner ordering. As a consequence, DS(ε)-optimal designs are found among those whose moment matrices cannot be improved relative to Loewner ordering. These designs are called *admissible*.

The Schur concavity of the DS(ε) criterion was already discussed in Section 3. Concavity for the criterion function $\psi_\varepsilon(\boldsymbol{\lambda})$ on $\mathbf{R}^k_+$ holds only in the case $k \le 2$. If $k > 2$, $\psi_\varepsilon(\boldsymbol{\lambda})$ is concave on the convex set $\mathcal{A}(\delta) = \Big\{\boldsymbol{\lambda} \in \mathbf{R}^k_+ : \lambda_i \in \Big[\frac{k-2}{\delta^2}, \infty\Big),\ 1 \le i \le k\Big\}$, where $\delta^2 = \frac{n\varepsilon^2}{\sigma^2}$. The result follows from the fact that the Hessian matrix of $\psi_\varepsilon(\boldsymbol{\lambda})$ is nonpositive definite on $\mathcal{A}(\delta)$. In the paper, the proof is limited to the case $k = 2$, since a counter-example is found which shows that $\psi_\varepsilon(\boldsymbol{\lambda})$ is not concave on $\mathbf{R}^k_+$ if $k > 2$.

An interesting relation of the DS criterion to the T and D criteria is found by considering the inequalities

$$\psi_\varepsilon(\boldsymbol{\lambda}) = \Pr\Big(\frac{Z_1^2}{\lambda_1} + \frac{Z_2^2}{\lambda_2} + \ldots + \frac{Z_k^2}{\lambda_k} \le \delta^2\Big) \le \Pr\Big(\chi_k^2 \le \delta^2\tilde{\lambda}\Big) \le \Pr\Big(\chi_k^2 \le \delta^2\bar{\lambda}\Big), \tag{10}$$

where λ_i are the eigenvalues of the moment matrix $\mathbf{M}$, $\tilde{\lambda} = (\prod \lambda_i)^{\frac{1}{k}}$ and $\bar{\lambda} = \frac{1}{k}\sum \lambda_i$. The first inequality follows from Okamoto's lemma (8) and the second from the arithmetic-geometric mean inequality. In both instances, there are equalities if and only if $\lambda_1 = \lambda_2 = \ldots = \lambda_k$, which means that $\mathbf{M}$ is a multiple of $\mathbf{I}$. Thus, if for some design tr $\mathbf{M}$ or $\det \mathbf{M}$ is maximized and $\mathbf{M}$ is a multiple of $\mathbf{I}$, then it is DS optimal.

In the case of the m-way first-degree polynomial model

$$Y_{ij} = \beta_0 + \beta_1 t_{i1} + \ldots + \beta_m t_{im} + E_{ij}$$

such designs are found. Assume that the experimental domain is an m-dimensional Euclidean ball $\mathcal{T}_{\sqrt{m}} = \{\mathbf{t} \in \mathbf{R}^m : \ ||\mathbf{t}|| \leq \sqrt{m}\}$. Denote by $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_l$ the support points of a design τ and by $p_1, p_2, \ldots, p_l$ the weights connected to them. Then the moment matrix is $\mathbf{M} = \mathbf{X}'\mathbf{D}\mathbf{X}$ where $\mathbf{D} = \text{diag}\,(p_1, p_2, \ldots, p_l)$ and $\mathbf{X}$ is the model matrix. For tr $\mathbf{M}$ we obtain

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

From the inequalities (10) it now follows that the design τ is DS optimal if $\mathbf{M} = \mathbf{I}$.

In the case of the minimal support size, $l = m + 1$, the condition

$$\mathbf{I} = \frac{1}{m+1}\mathbf{X}'\mathbf{X} = \frac{1}{m+1}\mathbf{X}\mathbf{X}'$$

implies that

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

for all $i \neq j \leq m+1$. Thus, the support points of a DS optimal design are vertices of a regular simplex.

References

[1] Fan, K. (1950). On a theorem of Weyl concerning eigenvalues of linear transformations II. *Proc. Nat. Acad. Sci. U.S.A.* **36,** 31-35. [MR **11**(1950)526;*Zbl.* **41**(1952)6]

[2] Fedorov, V. V., (1972). Theory of optimal experiments. Academic Press.

[3] Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

[4] Muirhead, R. F. (1903). Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proc. Edinburgh Math. Soc.* **21**, 144-157.

[5] Okamoto, M. (1960). An inequality for the weighted sum of χ^2-variates. *Bull. Math. Statist.* **9,** 69-70. [MR **22**(1961)10041;*Zbl.* **112**(1965)111]

[6] Pukelsheim, F. (1993). *Optimal Design of Experiments.* John Wiley & Sons, Inc., New York.

[7] Rao, C. R., (1967). Least Squares Theory using an Estimated Dispersion Matrix and Its Application to Measurement of Signals. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. I, L. M. LeCam and J. Neyman (eds), 355–372. Berkeley: University of California Press.

[8] Sinha, Bikas K. (1970). On the optimality of some designs. *Calcutta Statistical Association Bulletin*, **20**, 1–20.

[9] Silvey, S. D., (1980). Optimal design. Chapman and Hall.

[10] Tomić, M. (1949). Théorème de Gauss relatif au centre de gravité et son application (Serbian, Russian and French summaries). *Bull. Soc. Math. Phys. Serbie* **1**, 31-40. [MR **11**(1950)86; *Zbl.* **41**(1952)473]