AD-A086 830 ILLINOIS UNIV AT CHICAGO CIRCLE DEPT OF MATHEMATICS F/G 12/2
STUDY OF OPTIMALITY CRITERIA IN DESIGN OF EXPERIMENTS.(U)
JUN 80 A HEDAYAT AFOSR-79-3050
UNCLASSIFIED AFOSR-TR-80-0514 NL

1 OF 1
AD
A086830

END
DATE
FILMED
8-80
DTIC

LEVEL II

STUDY OF OPTIMALITY
CRITERIA IN DESIGN OF EXPERIMENTS *

by

A. Hedayat
Department of Mathematics
University of Illinois, Chicago

June 3, 1980

* Invited paper presented at International Symposium on Statistics and Related Topics held during May 5-8, 1980 at Carleton University, Ottawa, Cananda.

Research supported by Grant AFOSR 79-3050[illegible].

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) AFOSR (19) TR-80-0514	2. GOVT ACCESSION NO. AD-A086 830	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) STUDY OF OPTIMALITY CRITERIA IN DESIGN OF EXPERIMENTS		5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) A./Hedayat		8. CONTRACT OR GRANT NUMBER(s) (15) AFOSR-79-3050
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Illinois Department of Mathematics Chicago, IL 60680		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) (17) 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	(11) 3	12. REPORT DATE June 1980
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 33		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

INTERNATIONAL SYMPOSIUM ON STATISTICS AND RELATED TOPICS, Carleton University, Ottawa, Canada, May 5-8, 1980

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Design, optimality criterion, A-optimality, D-optimality, L-optimality, E-optimality, Universal optimality, Φ_p-criterion and its generalization, S-optimality, (M,S)-optimality, Schur optimality.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

~~In~~ this paper ~~we~~ have rigorously studied various optimality criteria currently adopted by design specialists in choosing a best design for performing an experiment. These optimality criteria includes: G-optimality, D-optimality, L-optimality, E-optimality, S-optimality, (M,S)-optimality, Φ_p-criteria, Universal optimality, type 1 and 2 criteria and Schur optimality.

PHI sub p

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

409950

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

STUDY OF OPTIMALITY CRITERIA IN DESIGN OF EXPERIMENTS

by

A. Hedayat
Department of Mathematics
University of Illinois, Chicago

ABSTRACT

In this paper we have rigorously studied various optimality criteria currently adopted by design specialists in choosing a best design for performing an experiment. These optimality criteria includes: G-optimality, D-optimality, L-optimality, E-optimality, S-optimality, (M,S)-optimality Φ_p-criteria, Universal optimality, type 1 and 2 criteria and Schur optimality.

Research supported by Grant AFOSR 79-3050.

Study of Optimality

Criteria in Design of Experiments

by

A. Hedayat
Department of Mathematics
University of Illinois, Chicago

1. Preliminary.

We perform experiments mainly to estimate or test hypotheses about some specified unknown parameters of a given model efficiently. Different considerations lead us to different criteria for the choice of the "best" design. Although Definition 2.1 is a response function criterion, most criteria in design theory are directly related to parameter estimation. Hence the information matrices play an important role and thus by Caratheodory theorem we can limit our search to discrete designs which are supported on sets consisting of finite number of points.

To see how the optimality criteria in design theory arose, we first give an example of the very basic motivation: Let d be a design and let Y be the vector of observations obtained under d. Assume

$$E(\underset{\sim}{Y}) = X\underset{\sim}{\theta},\ \mathrm{Cov}(Y) = \sigma^2 I, \tag{1.1}$$

where $\underset{\sim}{Y}$ is an $n \times 1$ vector of observations, X is an $n \times k$ matrix with known entries specified by d, $\underset{\sim}{\theta}$ is a $k \times 1$ vector of unknown constants, and I denotes the identity matrix of order n. In many cases we are only interested in the subvector θ_1 of θ. With no loss of generality we can write $\theta' = (\theta_1' \vdots \theta_2')$, where θ_1 is a $v \times 1$ vector, $1 \leq v \leq k$. According to the partition $\theta' = (\theta_1' \vdots \theta_2')$ the Model (1.1) can be written as

$$E(Y) = (X_1 \vdots X_2) \binom{\theta_1}{\theta_2}, \ \mathrm{Cov}(Y) = \sigma^2 I. \qquad (1.1)'$$

The information matrix of θ_1 under d and the Model (1.1)′ is $X_1'X_1 - X_1'X_2(X_2'X_2)^- X_2'X_1$. We shall denote this by M_d. Note that $M_d = X'X$ when $v = k$, i.e., $\theta_1 = \theta$. Now we consider four cases:

(i) <u>To estimate each components of</u> θ:

Assume $X'X$ is nonsingular, and suppose we want to estimate each of the individual parameters. By Gauss-Markov Theorem, the best linear unbiased estimator (b.l.u.e.) $\hat{\theta}$ of $\hat{\theta}$ is given by:

$$\hat{\theta} = (X'X)^{-1}X'Y \qquad (1.2)$$

with

$$\mathrm{Cov}(\hat{\theta}) = \sigma^2(X'X)^{-1}. \qquad (1.3)$$

Let x_i be the i-th column of X and c_j be the j-th column of $X(X'X)^{-1}$, then from (1.2) and (1.3) it follows that

$$\hat{\theta}_i = c_i' Y \qquad (1.4)$$

with

$$\mathrm{Var}(\hat{\theta}_i) = \sigma^2(c_i'c_i). \tag{1.5}$$

Since $(X'X)^{-1}(X'X) = I_k$, we have $c_i'x_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta. Applying the Schwarz inequality, we obtain

$$(x_i'x_i)(c_i'c_i) \geq (x_i'c_i)^2 = 1 \tag{1.6}$$

hence

$$\mathrm{Var}(\hat{\theta}_i) \geq \sigma^2/x_i'x_i. \tag{1.7}$$

Usually, the experimenter has some amount of freedom in the choice of the k vectors x_i. If possible, we would like to select a design which estimates each of the parameters with minimum variance. Observe that the equality in (1.6) holds if and only if $c_i = cx_i$ for a constant c, which implies that $X'X$ is a diagonal matrix. Hence, theoretically speaking, the "best design" is when $X'X$ is a diagonal matrix with diagonal entries as large as possible. (e.g., if x_{ij} = 0,1, or -1; then $x_i'x_i \leq n$, the best design is the one for which $X'X = nI_k$.) But such a design does not always exist, see Hedayat and Wallis (1979). When such designs do not exist, the question arises to how a best design should be defined. A reasonable approach is to minimize the average variance of each of the estimated parameters or to minimize the generalized variance, etc.

(ii) <u>To estimate linear functions of a subvector θ_1 of θ</u>:

Suppose we want to estimate linear functions of θ_1 in the form $q_1'\underline{\theta}_1$. The b.l.u.e. of $q_1'\underline{\theta}_1$ is $q_1'\hat{\theta}_1$ with

$$\mathrm{Var}(q_1'\hat{\theta}_1) = \sigma^2 q_1' M_d^- q_1, \tag{1.8}$$

where $\hat{\theta}_1 = M_d^- Q_d$, (1.9)

and $Q_d = [X_1' - X_1'X_2(X_2'X_2)^- X_2'] Y$, (1.10)

while M_d^- is any generalized inverse of M_d.

In choosing a design for estimating $q_1'\underline{\theta}_1$ there are many criteria. One of them is based on the following inequality

$$\mu_{min} \le \frac{q_1' M_d^- q_1}{q_1' q_1} \le \mu_{max}, \tag{1.11}$$

where μ_{max} and μ_{min} are the maximum and the minimum (nonzero) eigenvalues of M_d^-, respectively. This inequality gives a bound for the variance of $q_1'\hat{\underline{\theta}}_1$:

$$\mu_{min}\, q_1'q_1\sigma^2 \le \mathrm{Var}(q_1'\hat{\underline{\theta}}_1) \le \mu_{max}\, q_1'q_1\sigma^2 \tag{1.12}$$

(iii) <u>To test hypotheses</u>:

Suppose in addition Y is multivariate normal and we want to test $\theta_1 = \theta_2 = \ldots = \theta_v = 0$ $(v \le k)$. (Assume M_d is nonsingular). Then the usual F test has a power function depending monotonically (increasing) on a parameter λ where

$$\lambda = \sigma^{-2}\, \underline{\theta}_1'\, M_d \underline{\theta}_1 \tag{1.13}$$

and thus by (1.11) and (1.13)

$$\frac{\bar{\mu}_{min}}{\sigma^2} \theta_1'\theta_1 \le \lambda \le \frac{\bar{\mu}_{max}}{\sigma^2} \theta_1'\theta_1 \qquad (1.14)$$

where $\bar{\mu}_{max}$ and $\bar{\mu}_{min}$ are the maximum and the minimum eigenvalues of M_d.

(iv) <u>To construct confidence region</u>:

Again assume Y is multivariate normal and M_d is nonsingular. A $1-\alpha$ joint confidence region for θ_1 is a solid ellipsoid:

$$(\theta_1-\hat{\theta}_1)'M_d\,(\theta_1-\hat{\theta}_1) \le \sigma^2\chi^2_\alpha(v), \text{ if } \sigma^2 \text{ is known,} \qquad (1.15)$$

where $\chi^2_\alpha(v)$ is the $1-\alpha$ percentile of the χ^2 distribution with v degrees of freedom. Or

$$(\theta_1-\hat{\theta}_1)'M_d\,(\theta_1-\hat{\theta}_1) \le vS^2F_\alpha(v,n-r), \text{ if } \sigma^2 \text{ is unknown,} \qquad (1.16)$$

where $F_\alpha(v,n-r)$ is the $1-\alpha$ percentile of the F distribution with v and $n-r$ degrees of freedom, and $S^2 = Y'[I-X(X'X)^-X'X]Y/(n-r)$ is an unbiased estimator of σ^2 (assume rank $(X'X) = r$).

We observe that:

(a) The volume (expected volume, if σ^2 is unknown) of the above ellipsoid is proportional to the square root of det M_d^{-1}:

(b) The semi-exes (expected semi-axes, if σ^2 is unknown) of the above ellipsoid is proportional to the square roots of the eigenvalues of M_d^{-1}.

In Section 2 we shall study some well-known optimality criteria. Section 3-7 will be some generalization of those in

Section 2, or some recent developments in the determination of optimal designs. Throughout this paper we write the optimality criteria as a class of convex nonincreasing functionals ϕ on the set of information matrices rather than the class of convex nondecreasing functionals ψ on the set of covariance matrices, since the former is more general than the latter. For instance, when the covariance matrix of interest is equal to M_d^- (as in (1.13)), we have $\phi(M_d) = \psi(M_d^-)$ which is convex in M_d if ψ is convex in M_d^- but not on the other hand. The strict inclusion of one class in the other is illustrated by the fact that, if $\lambda_1(M_d^-) \geq \cdots \geq \lambda_v(M_d^-)$ are the eigenvalues of M_d^-, then $\Sigma\lambda_i^{\frac{1}{2}}(M_d^-) = \Sigma\lambda_i^{-\frac{1}{2}}(M_d)$ is convex in M_d but $\Sigma\lambda_i^{\frac{1}{2}}(M_d^-)$ is not convex in M_d^-.

Notation used in the rest of this paper are listed below:

$\mathcal{B}_v$ = the class of all $v \times v$ nonnegative definite matrices.

$\mathcal{B}_{v,o}$ = the class of all $v \times v$ nonnegative definite matrices with zero row and column sums.

$\mathcal{D}$ = the class of designs under consideration.

$\mathcal{C}$ = $\{M_d,\ d \in \mathcal{D}\}$.

Also, let $\mu_{d1} \geq \mu_{d2} \geq \cdots \geq \mu_{dv}$ be the eigenvalues of M_d. Note that if $\mathcal{C} \subseteq B_{v,o}$, $\mu_{dv} = 0$, for all $d \in D$. If necessary, we let ξ denote an approximate design (a probability measure on the experimental space) and M_ξ be the associated information matrix.

To avoid messy expressions, the dimensions of matrices should be deduced from the context if they are not explicitly specified.

2. Some well-known optimality criteria.

Assume $C \subseteq \mathcal{B}_v$.

I. G-optimality.

Smith (1918) introduced a response function criterion which can be stated as follow:

Definition 2.1. A design $\xi^* \in \mathcal{A}$ is G-optimal if and only if

$$\min_{\xi \in \mathcal{A}} \max_{x \in \chi} \operatorname{var}_{\xi} \widehat{EY}_x = \max_{x \in \chi} \operatorname{Var}_{\xi^*} \widehat{EY}_x,$$

where $\widehat{EY}_x$ is the b.l.u.e. of EY_x and χ is the experimental space. Kiefer called it G-optimal (for global or minimix), since we are minimizing the maximum variance of any predicted value over the experimental space.

II. D-optimality.

Definition 3.2.2. A design $d^* \in \mathcal{A}$ is D-optimal if and only if M_{d^*} is non-singular and $\min_{d \in \mathcal{A}} \det(M_d^{-1}) = \det(M_{d^*}^{-1})$. Here, "D-" stands for determinant. The concept introduced and studied by Wald (1943) and applied by Mood (1946). This criterion has many appealing properties;

(1) under normality, if d^* is D-optimal, d^* minimizes:

(a) The volume (or expected volume, if σ^2 is unknown, and rank (M_d) is invariant under d) of the smallest invariant confidence region on $\theta_1, \theta_2, \ldots, \theta_v$ for any given confidence coefficient.

(b) The generalized variance of the estimators of parameters. (see remark below).

(2) In the class of approximate designs, D-optimality $\Leftrightarrow$ G-optimality whenever $v = k$, i.e., $\theta_1 = \theta$.

(3) The design remains D-optimal if one changes the scale of the parameters: Let $\theta_1', \theta_2', \ldots, \theta_v'$ be related to $\theta_1, \theta_2, \ldots, \theta_v$ by a non-singular linear transformation. If d^* is D-optimal for $\theta_1, \ldots, \theta_v$, then d^* is also D-optimal for $\theta_1', \ldots, \theta_v'$. The analogue for other criteria is false in even the simplest settings.

Remark: Suppose $X = (X_1, X_2, \ldots, X_n)'$ is distributed as multivariate $N(\mu, Y)$. The determinant of V is called the generalized variance of X as defined by Wilks (1932).

In the theory of linear regression, under normal assumption, $\hat{\underline{\theta}}_1 = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_v)'$ is distributed as $N(\underline{\theta}_1, M^{-1}\sigma^2)$, so the generalized variance of $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_v)$ is equal to the determinant of $M_d^{-1}\sigma^2$ which is the product of σ^{2v} and det M_d^{-1} (Assume M_d is non-singular).

III. L-optimality.

Definition 2.3. A design $d^* \in \Omega$ id linear optimal (L-optimal) if and only if $\min_{d \in \Omega} L(M_d^{-1}) = L(M_{d^*}^{-1})$ where L is a nonnegative linear functional on $\mathcal{C}$.

One of the most useful linear criteria of optimality is A-optimality defined when

$$L(M_d^{-1}) = Tr(M_d^{-1}).$$

Definition 2.4. A design $d^* \in \Omega$ is A-optimal if and only if M_{d^*} is non-singular and $\min_{d \in \Omega} Tr(M_d^{-1}) = Tr(M_{d^*}^{-1})$. "A-" stands for average. In a statistical sense, if d^* is A-optimal, it minimizes the average variances of $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_v$. This criterion was introduced and studied by Elfving (1952) and Chernoff (1953).

IV. E-optimality.

Definition 2.5. A design $d^* \in \mathcal{D}$ is E-optimal if and only if $\min_{d \in \mathcal{D}} \mu_{dv}^{-1} = \mu_{d^*v}^{-1}$. E-optimality was first considered in hypothesis testing (Wald (1943), Ehrenfield (1955)). "E-" stands for eigenvalue. It has the following properties:

(1) In hypothesis testing. Under the normality assumption, an E-optimal design maximizes the minimum power of the associated F-test of size α on the contour $\theta_1'\theta_1 = c$ for every α and c. (See (1.14)).

(2) In point estimation, An E-optimal design minimizes the maximum variance of the b.l.u.e.'s of the $q_1'\theta_1$ over all $v \times 1$ vectors q_1 with $q_1'q_1 = 1$. (See (1.12)).

(3) In interval estimation. An E-optimal design minimizes the largest semi-axis of the (hyper) ellipsoid when normality assumptions are made on the observations.

Now it seems natural to specify some optimality functional Φ on $\mathcal{C}$ and to pose the problem: Find d to minimize $\Phi(M_d)$. We call Φ an optimality criterion. The above well-known criteria are then:

$$\text{D-optimality: } \Phi_D(M_d) = \det(M_d^{-1}) = \prod_{i=1}^{v} \mu_{di}^{-1} \qquad (2.1)$$

L-optimality: $\phi_L(M_d) = L(M_d^{-1})$ (2.2)

A-optimality: $\phi_A(M_d) = Tr(M_d^{-1}) = \sum_{i=1}^{v} \mu_{di}^{-1}$ (2.3)

E-optimality: $\phi_E(M_d) = \mu_{dv}^{-1}$. (2.4)

(2.1), (2.3) and (2.4) are regarded as infinite if M_d is singular.

Note, in case $C \subseteq \mathcal{B}_{v,o}$, the definitions of D-,A-,E-optimality are similar, one can simply replace the index v in (2.1), (2.2) and (2.4) by $v-1$.

3. S-optimality and (M,S)-optimality:

Assume $C \subseteq \mathcal{B}_v$. When $Tr(M_d) = \sum_i \mu_{di} = A$ is a constant, for all $d \in \mathcal{D}$, the D-, A-, E-optimalities are attained when all the μ_{di}'s are equal (we call such a design a symmetric design). Unfortunately, symmetric designs do not always exist. Intuitively, in the absence of a symmetric design, we may want to believe that the "closest" design to the hypothetical symmetric design is a reasonable design to use. Shah (1960) proposed the Euclidean distance between the vector of eigenvalues of the designs as the measure of distance between the corresponding designs. Thus, according to Shah (1960) if there is no symmetric design in D, we should use the design d which minimizes the Euclidean distance between $(\mu_{d1},\ldots,\mu_{dv})$ and the vector of eigenvalues of the hypothetical symmetric design, $(A/v,\ldots,A/v)$, i.e.,

$$\left[\sum \mu_{di}^2 - (\sum \mu_{di})^2/v\right]^{\frac{1}{2}} \qquad (3.1)$$

Clearly, this is only a heuristic approach with no statistical justification. However, it has the merit that when $Tr(M_d)$ is a constant, the minimization of (3.1) is equivalent to that of $TrM_d^2 = \sum_{i,j} m_{dij}^2$ which is easier to handle.

Define $\phi: \mathcal{C}_v \to [0, +\infty]$ such that

$$\phi(M_d) = TrM_d^2 = \sum_i \mu_{di}^2 = \sum_{i,j} m_{dij}^2. \qquad (3.2)$$

Formally, we have:

Definition 3.1. Suppose $Tr(M_d) = A$ is a constant for all $d \in \mathcal{D}$. A design $d^* \in \mathcal{D}$ is called S-optimal if and only if d^* minimizes $\phi(M_d)$ (as in 3.2) for all $d \in \mathcal{D}$.

Motivated by Shah's criterion, Eccleston and Hedayat (1974) proposed a similar procedure in the case when TrM_d is not a constant.

Let $\mathcal{C}' \subseteq \mathcal{C}$ be such that the matrices in $\mathcal{C}'$ have maximum trace.

Definition 3.2. A design $d^* \in \mathcal{D}$ is (M,S)-optimal if and only if $M_{d^*} \in \mathcal{C}'$ and d^* minimizes $\phi(M_d)$ (as in (3.2)), for all $d \in \mathcal{D}'$, where $\mathcal{D}' = \{d \in \mathcal{D};\ M_d \in \mathcal{C}'\}$.

A geometric interpretation of (M,S)-optimality can be given as follow. Set

$$S_A = \{(\mu_{d1}, \ldots, \mu_{dv});\ \mu_{di} > 0,\ \sum_i \mu_{di} = A\},$$

and

$$S_{AB} = \{(\mu_{d1}, \ldots, \mu_{dv});\ \mu_{di} > 0,\ \sum_i \mu_{di} = A;\ \sum_i \mu_{di}^2 = B\}.$$

Then S_A is an open simplex and S_{AB} is part of a (v-2)-dimensional sphere with $(A/v, \ldots, A/v)$ as the center and the quantity

$\rho = \left[\sum_i \mu_{di}^2 - \left(\sum_i \mu_{di}\right)^2/v\right]^{\frac{1}{2}}$ as the radius, when $B \geq A^2/v$. The procedure of finding an (M,S)-optimal design is the same as to choose a simplex S_A as far away from the origin as possible, and then find a design with the vector of eigenvalues on S_A which is closest to the center of the simplex in the Euclidean sense.

In the $\mathcal{B}_{v,o}$ context, same arguments hold except replaceing v by $v-1$.

4. Φ_p-criteria.

In Keifer (1974), the following family of criteria was introduced. We shall describe it in the $\mathcal{B}_v$ context.

Let

$$\Phi_p(M_d) = \left[\frac{1}{v}\,\mathrm{Tr}(M_d^{-p})\right]^{\frac{1}{p}} = \left[\frac{1}{v}\sum_{i=1}^{v} \mu_{di}^{-p}\right]^{\frac{1}{p}}, \quad 0 < p < \infty \,. \tag{4.1}$$

Definition 4.1. A design $d^* \in \mathcal{D}$ is Φ_p-optimal if and only if d^* minimizes $\Phi_p(M_d)$, $d \in \mathcal{D}$.

When $C \subseteq \mathcal{B}_v$, we may restrict ourself to d with M_d nonsingular. The following theorem will give a connection between D-, A-, E-criterion and the Φ_p-criterion.

Theorem 4.1. (i) $\quad \Phi_1(M_d) = \frac{1}{v}\,\mathrm{Tr}(M_d^{-1}) = \frac{1}{v}\left(\sum_{i=1}^{v} \mu_{di}^{-1}\right)$

$$\text{(ii)} \quad \Phi_0(M_d) = \lim_{p\to 0} \Phi_p(M_d) = \left(\prod_{i=1}^{v} \mu_{di}^{-1}\right)^{\frac{1}{v}} \tag{4.2}$$

$$\text{(iii)} \quad \Phi_\infty(M_d) = \lim_{p\to\infty} \Phi_p(M_d) = \mu_{dv}^{-1}$$

Proof: (i) is clear

(ii) $$\Phi_p(M_d) = \left[\frac{1}{v}\sum_{i=1}^{v} \mu_{di}^{-p}\right]^{\frac{1}{p}}$$

$$\log \Phi_p(M_d) = \frac{1}{p} \log \left[\frac{1}{v}\sum_{i=1}^{v}\mu_{di}^{-p}\right] .$$

As p tends to zero, the right hand side goes to $\frac{0}{0}$, so by applying L'Hospital's rule, we obtain

$$\lim_{p\to 0} \log \Phi_p(M_d) = \lim_{p\to 0} \frac{\frac{1}{v}\left[\sum_{i=1}^{v} \mu_{di}^{-p} \log \mu_{di}^{-1}\right]}{\frac{1}{v}\left[\sum_{i=1}^{v} \mu_{di}^{-p}\right]}$$

$$= \frac{1}{v}\sum_{i=1}^{v} \log \mu_{di}^{-1}$$

$$= \frac{1}{v} \log \prod_{i=1}^{v} \mu_{di}^{-1} .$$

Hence $$\lim_{p\to 0} \Phi_p(M_d) = \left(\prod_{i=1}^{v} \mu_{di}^{-1}\right)^{\frac{1}{v}} .$$

(iii) Let $\mu'_{di} = \mu_{dv}\mu_{di}^{-1}$.

Then

$$\log \Phi_p(M_d) = \frac{1}{p} \log \left[\frac{1}{v}\sum_{i=1}^{v}(\mu'_{di}\mu_{dv}^{-1})^{p}\right]$$

$$= \frac{1}{p} \log \left[\frac{1}{v}\, \mu_{dv}^{-p} \sum_{i=1}^{v} \mu_i'^{p}\right]$$

$$= \frac{1}{p} \log \frac{1}{v} + \log \mu_{dv}^{-1} + \frac{1}{p} \log \left(\sum_{i=1}^{v} \mu'^{p}_{di}\right).$$

Since $\mu'_{di} \leq 1$, for all i,

we conclude $$0 \leq \log\left(\sum_{i=1}^{v} \mu'^{p}_{di}\right) \leq \log v .$$

Hence $$\lim_{p\to\infty} \frac{1}{p} \log\left(\sum_{i=1}^{v} \mu'^{p}_{di}\right) = 0 .$$

Therefore $$\lim_{p\to\infty} \log \Phi_p(M_d) = \log \mu_{dv}^{-1} ,$$

and consequently $$\lim_{p\to\infty} \Phi_p(M_d) = \mu_{dv}^{-1} .$$

Corollary 4.1.

(i) When $p = 1$, Φ_p-criterion is equivalent to A-optimality.

(ii) When p approaches to 0, the limiting case of Φ_p-criterion is equivalent to D-optimality.

(iii) When p approaches to ∞, the limiting case of Φ_p-criterion is equivalent to E-optimality.

Remark: The Φ_p-criterion in the $\mathcal{B}_{v,0}$ context is

$$\Phi_p(M_d) = \left[\frac{1}{v-1} \sum_{i=1}^{v-1} \mu_{di}^{-p}\right]^{\frac{1}{p}} .$$

5. Universal Optimality.

In Keifer (1975), a strong optimality criterion was considered. Here, we restrict ourself in $\mathcal{B}_{v,0}$. (Since in $\mathcal{B}_v$ context, it is easier.)

Definition 5.1. We say $d^* \in \mathcal{D}$ is a universally optimal design, if d^* minimizes $\Phi(M_d)$, $d \in \mathcal{D}$ for any $\phi : \mathcal{B}_{v,0} \longrightarrow (-\infty, +\infty]$

satisfying:

(i) ϕ is convex,

(ii) ϕ (bM) is nonincreasing in the scalar $b \geq 0$ for each $M \in \mathcal{B}_{v,0}$. (5.1)

(iii) ϕ is invariant under each permutation of rows and (the same on) columns.

Since $-Tr(M)$ satisfies (5.1), immediately we have the following theorem:

Theorem 5.1. If $d^* \in \mathcal{D}$ is universally optimal, then TrM_{d^*} is maximum.

Definition 5.2. A matrix M is called a completely symmetric (c.s.) matrix if $M = \alpha I_v + \beta J_v$ where α, β are scalars and I_v is the identity matrix, J_v consists of all 1's.

Lemma 5.1. If M_1 and M_2 are two completely symmetric matrices in $\mathcal{B}_{v,0}$, then there exists an h such that $M_2 = hM_1$.

Proof: Suppose $M_1 = \alpha_1 I_v + \beta_1 J_v$

$$M_2 = \alpha_2 I_v + \beta_2 J_v$$

$$M_i \cdot \underline{1} = 0 \Rightarrow \alpha_i + v\beta_i = 0 \quad \text{for } i = 1,2,$$

So $$M_i = -v\beta_i I_v + \beta_i J_v, \; i = 1,2.$$

Let $$h = \beta_2/\beta_1 .$$

Then $$M_2 = -v\beta_2 I_v + \beta_2 J_v = h(-v\beta_1 I_v + \beta_1 J_v) = hM_1 .$$

The following theorems are simple tools in determining such an optimal design.

Theorem 5.2. Suppose $\mathcal{C} \subseteq \mathcal{B}_{v,0}$ contains a M_{d*} for which

$$\begin{aligned} &\text{(a)} \quad M_{d*} \ \text{is c.s.} \\ &\text{(b)} \quad \mathrm{Tr}M_{d*} = \max_{d \in \mathcal{D}} \mathrm{Tr}M_d . \end{aligned} \tag{5.2}$$

Then d^* is universally optimal in $\mathcal{D}$.

Proof: From Theorem 5.1 it suffices to show that $\Phi(M_{d*})$ minimizes $\Phi(M_d)$ for all Φ satisfies (5.1), $M_d \in \mathcal{C}'$ where $\mathcal{C}' \subseteq \mathcal{C}$ consists of the matrices which have maximum trace.

For any $M_d \in \mathcal{C}'$, let τM_d be obtained from M_d by permuting rows and columns according to τ, and let $\bar{M}_d = \sum_\tau \tau M_{d/v!}$, the symmetrized version of M_d. By (5.1)(a) and (c) we have

$$\Phi(\bar{M}_d) \le \sum_\tau \frac{1}{v!} \Phi(\tau M_d) = \Phi(M_d), \tag{5.3}$$

for any Φ satisfying (5.1). Of course $\bar{M}_d$ need not be in $\mathcal{C}$, but $\bar{M}_d$ is c.s. and in $\mathcal{B}_{v,0}$. By Lemma 5.1, $\bar{M}_d$ is of the form bM_{d*} for some $b \ge 0$. Now $\mathrm{Tr}(\bar{M}_d) = \mathrm{Tr}(M_d)$. But $\mathrm{Tr}(M_d) = \mathrm{Tr}(M_{d*})$ by assumption. This implies $b = 1$ and hence $\bar{M}_d = M_{d*}$. By (5.3), $\Phi(M_{d*}) = \Phi(\bar{M}_d) \le \Phi(M_d)$ for all Φ satisfying (5.1) and $M_d \in \mathcal{C}'$. Therefore M_{d*} is universally optimal.

Theorem 5.3. Suppose an M_{d*} satisfying (5.2) exists. Let $\Phi: \mathcal{B}_{v,0} \to (-\infty, +\infty]$ be a function satisfying (5.1). If, in addtion, Φ is strickly convex (and hence also "nonincreaseing" in property (ii) is replaced by "decreasing"), then every Φ-optimal d' has $M_{d'} = M_{d*}$. (i.e., d' is also universally optimal).

Proof: Let $\bar{M}_{d'} = \sum_{\tau} M_{d'}/v!$. Since Φ is strictly convex, we have

$$\Phi(\bar{M}_{d'}) < \sum_{\tau} \frac{1}{v!} \Phi(\tau M_{d'}) = \Phi(M_{d'}). \qquad (5.4)$$

Again $\bar{M}_{d'}$ is c.s. and in $\mathcal{B}_{v,0}$, this implies that $\bar{M}_{d'} = bM_{d*}$ for some $b \geq 0$. Since M_{d*} satisfies (5.2), $Tr(M_{d*}) \geq Tr(M_{d'})$ which implies $b \leq 1$. But if $b < 1$

$$\Phi(M_{d'}) > \Phi(\bar{M}_{d'}) = \Phi(bM_{d*}) > \Phi(M_{d*}). \quad (5.5)$$

This contradicts the assumption that d' is Φ-optimal. From (5.4) and (5.5) we can conclude that

$$b = 1 \text{ and } M_{d'} = \bar{M}_{d'}$$

i.e.,

$$M_{d'} = M_{d*} .$$

And d' is indeed universally optimal. $\square$

Let Φ_1 and Φ_2 be two convex functions satisfying (5.1). Suppose d^* is Φ_1-optimal, the following theorem gives a sufficient condition for d^* to be Φ_2-optimal.

Theorem 5.4. If $\Phi_1 \leq \Phi_2$ on $\mathcal{C}$ and if $\Phi_1(M_{d*}) = \Phi_2(M_{d*})$, then d^* is Φ_2-optimal if d^* is Φ_1-optimal.

Proof: Assume d^* is Φ_1-optimal, then $\Phi_1(M_{d*}) \le \Phi_1(M_d)$ for all $d \in \mathcal{D}$.

By assumption

$$\Phi_2(M_{d*}) = \Phi_1(M_{d*}) \le \Phi_1(M_d) \le \Phi_2(M_d).$$

Hence the result.

Example 5.1. A useful family of criteria in the $\mathcal{B}_{v,0}$ context is the Φ_p-criteria, for $0 < p < \infty$, with the limiting values

$\Phi_0(M_d) = \prod_i \mu_{di}^{-\frac{1}{v-1}}$ and $\Phi_\infty(M_d) = \mu_{d(v-1)}^{-1}$. Here

$p < q \Rightarrow \Phi_p(M_d) \le \Phi_q(M_d)$ with equality if and only if all μ_{di} are equal. Hence from Theorem 5.4 if M_{d*} is c.s. and d^* is Φ_p-optimal $\Rightarrow d^*$ is Φ_q-optimal for all $q > p$.

In the absence of universal optimality, some weaker optimality results which have some useful statistical implications (for instance, include A-, E-, D-criteria and all ϕ_p-criteria, $0 < p < \infty$) has been discussed by Kiefer (1974).

Observe that (4.1) and (4.2) are equivalent to the following:

$$\text{(a)} \quad \Phi_p^*(M_d) = \sum_i \mu_{di}^{-p}, \; 0 < p < \infty;$$

$$\text{(b)} \quad \Phi_0^*(M_d) = -\sum_i \log \mu_{di} \qquad (5.6)$$

$$\text{(c)} \quad \Phi_\infty(M_d) = \mu_{dv}^{-1}.$$

Let
$$\Phi^*(M_d) = \sum_i f(\mu_{di}), \qquad (5.7)$$

where f is convex on $[0, +\infty)$. We want to find conditions under which a design d is Φ^*-optimal.

Lemma 5.2. If f is a convex function on $[0, +\infty)$, then

$$\sum_{i=1}^{v-1} f(\mu_{di}) \geq \frac{v-1}{v} \sum_{j=1}^{v-1} f\left(\frac{v}{v-1} m_{djj}\right) \qquad (5.8)$$

for any M_d in $\mathcal{B}_{v,0}$, with equality if all the μ_{di}'s are equal or M_d is c.s.

Proof: Let P be the $(v-1)\times v$ orthonormal matrix such that

$$PM_dP' = \Lambda_d = \begin{bmatrix} \mu_{d1} & & 0 \\ & \ddots & \\ 0 & & \mu_{d(v-1)} \end{bmatrix}.$$

Augment P with $(\frac{1}{\sqrt{v}}, \frac{1}{\sqrt{v}}, \ldots, \frac{1}{\sqrt{v}})$ and call the resulting matrix P^*.

$$P^*M_dP^{*\prime} = \begin{bmatrix} PM_dP' & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \Lambda_d & 0 \\ 0 & 0 \end{bmatrix}.$$

Assume $P^* = (p^*_{ij})$, and let $e_{ij} = p^{*2}_{ij}$.

Then $\sum_{j=1}^{v} e_{ij} = 1$ and $\sum_{i=1}^{v-1} e_{ij} = 1 - \frac{1}{v} = \frac{v-1}{v}$.

Also $M_d = P'\Lambda_dP \Rightarrow m_{djj} = \sum_{i=1}^{v-1} e_{ij}\,\mu_{di}$.

$$\text{Thus} \quad \frac{v-1}{v} f\left(\frac{v}{v-1} m_{djj}\right) = \frac{v-1}{v} \quad f\left(\frac{v}{v-1} \sum_{i=1}^{v-1} e_{ij}\,\mu_{di}\right)$$

$$= \frac{v-1}{v} f\left(\sum_{i=1}^{v-1} \frac{v}{v-1} e_{ij}\,\mu_{di}\right).$$

Since

$$\sum_{i=1}^{v-1} e_{ij} = \frac{v-1}{v} \Rightarrow \sum_{i=1}^{v-1} \frac{v}{v-1} e_{ij} = 1 \quad \text{and} \quad e_{ij} \geq 0.$$

The convexity of f

$$\Rightarrow \frac{v-1}{v} f\left(\sum_{i=1}^{v-1} \frac{v}{v-1} e_{ij} \mu_{di}\right) \leq \frac{v-1}{v} \sum_{i=1}^{v-1} \frac{v}{v-1} e_{ij} f(\mu_{di})$$

$$= \sum_{i=1}^{v-1} e_{ij} f(\mu_{di}).$$

Hence we have

$$\frac{v-1}{v} f\left(\frac{v}{v-1} m_{djj}\right) \leq \sum_{i=1}^{v-1} e_{ij} f(\mu_{di}). \qquad (5.9)$$

Summing on j, we obtain

$$\frac{v-1}{v} \sum_{j=1}^{v} f\left(\frac{v}{v-1} m_{djj}\right) \leq \sum_{i=1}^{v-1} f(\mu_{di}).$$

If $\mu_{d1} = \mu_{d2} = \ldots = \mu_{d(v-1)} = \mu_d$ (i.e., M_d is c.s.)

$$m_{djj} = \sum_{i=1}^{v-1} e_{ij} \mu_{di} = \mu_d\left(\frac{v-1}{v}\right).$$

Then

$$\frac{v-1}{v} \sum_{j=1}^{v} f\left(\frac{v}{v-1} m_{djj}\right) = \frac{v-1}{v} \sum_{j=1}^{v} f(\mu_d) = (v-1) f(\mu_d)$$

$$= \sum_{i=1}^{v-1} f(\mu_d) = \sum_{i=1}^{v-1} f(\mu_{di}). \qquad (5.10)$$

Theorem 5.5. If Φ^* is given by (5.7) with f convex, and if $d^* \in \mathcal{D}$ satisfies:

(i) M_d^* is c.s.

(ii) d^* minimizes $\sum_{j=1}^{v-1} f\left(\frac{v}{v-1} m_{djj}\right)$, (5.11)

then d^* is Φ^*-optimal.

Proof: Follows directly from (5.10).

Example: In the case of (5.6), we obtain,

(a) If M_d^* is c.s. and minimizes $\sum_j m_{djj}^{-p} \Rightarrow d^*$ is Φ_p^*-optimal.

(b) If M_d^* is c.s. and maximizes $\sum_j \log m_{djj} \Rightarrow d^*$ is Φ_0^*-optimal (i.e., it is D-optimal).

(c) If M_d^* is c.s. and maximizes $\min_j m_{djj} \Rightarrow d^*$ is Φ_∞-optimal (i.e., it is E-optimal).

Also, from Theorem 5.2,

(d) If M_d^* is c.s. and maximizes $\sum_j m_{djj} \Rightarrow d^*$ is Φ_p-optimal, $0 \leq p \leq \infty$ and more.

6. Type 1 and Type 2 Criteria.

Cheng (1978) refined Kiefer's criteria and defined a larger class of optimality criteria that include A-, E-, D-, all Φ_p-criteria, $0 < p < \infty$, and more.

Again, let $C \subseteq \mathcal{B}_{v,0}$. (In the $\mathcal{B}_v$ context, similar arguments hold.) Let $t_{\mathcal{D}} = \max_{d \in \mathcal{D}} \mathrm{Tr} M_d$.

Definition 6.1. A design $d^* \in \mathcal{D}$ satisfies optimality criteria of type 1 if d^* minimizes $\Phi_f(M_d) = \sum_{i=1}^{v-1} f(\mu_{di})$ where f is a real-valued function defined on $[0,t_{\mathcal{D}})$ such that

a) f is continuous, strictly convex, and strictly decreasing on $[0,t_{\mathcal{D}}]$. We include here the possibility that $f(0) = \lim_{x \to 0^+} f(x) = +\infty$. (6.1)

b) f is continuously differentiable on $(0,t_{\mathcal{D}})$, and f' is strictly concave on $(0,t_{\mathcal{D}})$, i.e., $f' < 0$, $f'' > 0$, and $f''' < 0$ on $(0,t_{\mathcal{D}})$.

Definition 6.2. A design $d^* \in \mathcal{D}$ satisfies optimality criteria of type 2, if d^* minimizes $\Phi_f(M_d) = \sum_{i=1}^{v-1} f(\mu_{di})$ where f has the same property as in Definition 6.1. Except that the strict concavity of f' is replaced by atrict convexity, i.e., $f''' > 0$ on $(0,t_{\mathcal{D}})$.

Also, a generalized optimality criterion of type i ($i = 1,2$) is defined to be the pointwise limit of a sequence of type i criteria.

From (4.2) and (5.6), the A-, D-, and Φ_p-criterion are of typ 1 and the E-criterion is a generalized criterion of type 1 (being the limit of Φ_p-criteria, as $p \to \infty$). Note that the A- and D-criteria correspond to the choices of $f(x) = x^{-1}$ and $-\log x$ respectively.

Remarks: (i) There do exist functions satisfying the requirements for a type 2 criterion. For example, let $f(x) = \epsilon x^3 - ax$ over the interval $[0,t_{\mathcal{D}}]$ of interest, when $\epsilon > 0$, $a > 0$ and ϵ compared with a, is small.

(ii) From Section 4 if there is a symmetric design which maximizes $\mathrm{Tr}M_d$ over $\mathcal{D}$, then it is optimal with respect to a very general class of criteria including both generalized type 1 and type 2 criteria. $\square$

It appears that most optimality criteria (universal optimality is an exception) which place equal emphasis on all the parameters can be formulated in terms of the eigenvalues of the information matrix. In Section 7 we shall introduce another optimality criterion of the form $\phi(\mu_{d1},\ldots,\mu_{d(v-1)})$ with ϕ Schur convex or convex symmetric.

7. Schur optimality.

The concept of Schur optimality was introduced by Magda (1979). To see how it was defined, let us recall the following:

Definition 7.1. A matrix with nonnegative entries is called doubly stochastic if the sum of the entries is 1 in every row and every column.

Definition 7.2. Let I be an interval on the real line. A function $\phi:I^n \to R$ is called Schur convex (after Schur (1923)) if

$$\phi(Sx) \leq \phi(x)$$

for all $x \in I^n$ and every doubly stochastic matrix S. A Schur convex function is not necessarily convex, e.g., $\phi(x_1,x_2) = |x_1-x_2|$. Any Schur convex funcion is symmetric, because for any permutaion matrix P we have

$$\phi(Px) \leq \phi(x) = \phi(P^{-1}Px) \leq \phi(Px).$$

Hence $\phi(Px) = \phi(x)$ as desired. We have used the fact that a permutation matrix and its inverse are examples of doubly stochastic matrices.

While symmetry is a necessary condition to have Schur convexity it is by no means sufficient. When convexity is added to symmetry we can insure Schur convexity. This is seen as follow: By Birkhoff (1946) every doubly stochastic matrix S can be written as a convex sum of permutaion matrices. Let $S = \Sigma\lambda_i Pi, (\Sigma\lambda_i = 1)$. Then

$$\phi(Sx) = \phi(\Sigma\lambda_i P_i x) \underbrace{\leq}_{\text{convexity of } \phi(\cdot)} \Sigma\lambda_i \phi(P_i x) \underbrace{=}_{\text{symmetry of } \phi(x)} \Sigma\lambda_i \phi(x)$$

$= \phi(x)$ and this proves Schur convexity.

Assume $\mathcal{C} \subseteq \mathcal{R}_{v,0}$, let $I = [0,t_{\mathcal{N}}]$ and n be the smallest integer for which $\mu_{d(n+1)} = \mu_{d(n+2)} = \ldots = \mu_{dv} = 0$ for all $d \in \mathcal{N}$.

Define $\sigma(M_d)$ to be the following vector in I^n:

$$\sigma(M_d) = \begin{pmatrix} \mu_{d1} \\ \vdots \\ \mu_{dn} \end{pmatrix} . \qquad (7.1)$$

For $d \in \mathcal{N}$ and any Schur convex function ϕ defined on I^n and nonincreasing in its arguments, set

$$\phi(M_d) = \phi(\sigma(M_d)). \qquad (7.2)$$

Schur optimality is now defined as follows:

Definition 7.5. A design $d^* \in \mathcal{D}$ is called Schur-optimal if d^* minimizes $\Phi(M_d)$, for all $d \in \mathcal{D}$, and all Schur convex functions Φ nonincreasing in their arguments.

Note that, if $\phi: I \to R$ is convex, then

$$\Phi(x) = \sum_{i=1}^{n} \phi(x_i), \; x = (x_1, x_2, \ldots, x_n) \quad (7.3)$$

is Schur convex on I^n because $\Phi(\cdot)$ is symmetric and convex. From (5.6) D-, A-, and all Φ_p-criteria defined so far on the eigenvalues of the information matrices are instances of Schur functions. As a symmetric and convex function on I_n

$$E(x_1, \ldots, x_n) = -\min_{1 \leq i \leq n} \{x_1, x_2, \ldots, x_n\}$$

is also Schur convex. This function is associated with E-optimality. Note that E-optimality is no longer a limiting case when delt with as a Schur convex function. To prove Schur optimality, we state the following very useful tool.

Theorem 7.2. (derived from Ostrowski (1952)).

Let $F(x_1, \ldots, x_n)$ be a Schur convex and nonincreasing function in its arguments on I^n. Let

$$y_1 \geq y_2 \geq \cdots \geq y_n; \; x_1 \geq x_2 \geq \cdots \geq x_n \quad (7.4)$$

satisfy the following

$$y_1 + \cdots + y_\ell \leq x_1 + \cdots + x_\ell \quad \text{for all } 1 \leq \ell \leq n. \quad (7.5)$$

Then

$$F(y_1,\ldots,y_n) \leq F(x_1,\ldots,x_n).$$

For convenience, when two vectors x and $y \in I^n$, Satisfy (7.4) and (7.5) we write $y \leq x$.

Applying the above result we can immediately conclude:

Theorem 7.3. d^* is Schur optimal if $\sigma(M_d^*) \leq \sigma(M_d)$ for all $d \in \mathfrak{D}$.

It should be pointed out that the ordered partial sums in (7.4) are examples of Schur convex functions. Further useful results can be obtained in Hardy and Littlewood (1967).

Lemma 7.2. Let $M_d \in C \subseteq B_{v,0}$ and $P_i (1 \leq i \leq n)$ be n orthogonal matrices such that $M_d^{(i)} = P_i^{-1} M_d P_i$ also satisfies $M_d^{(i)} \mathbf{1} = 0$ for all $1 \leq i \leq n$. Set $\overline{M}_d = \frac{1}{n} \sum_{i=1}^{n} M_d^{(i)}$. Then for any Schur convex function Φ nonincreasing in its arguments we have $\Phi(\overline{M}_d) \leq \Phi(M_d)$.

Proof: Since the P_i's are orthogonal, we have $\sigma(M_d^{(i)}) = \sigma(M_d)$ and hence $\Phi(M_d^{(i)}) = \Phi(M_d)$ for all $1 \leq i \leq n$. Moreover, let $\{\mu_{di}\}$ and $\{\overline{\mu}_{di}\}$ denote the eigenvalues of M_d and $\overline{M}_d$ respectively (and let them be ordered nonincreasingly.) Then it is known (see Bellman (1970)) that

$$\sum_{i=1}^{\ell} \overline{\mu}_{di} \leq \sum_{i=1}^{\ell} \mu_{di} \quad \text{for } \ell = 1,2,\ldots,v-1.$$

By Theorem 7.1 we obtain $\Phi(\overline{M}_d) \leq \Phi(M_d)$.

Remark: We call $\overline{M}_d$ (defined in Lemma 7.2), an averaged version of M_d.

Verifying the requirements of Theorem 7.3 is difficult because of the large variety of information matrices M_d. It is practically impossible to find $\sigma(M_d)$. When averaging M_d properly, however, it is easily seen that finding $\sigma(\overline{M}_d)$ is a tractable task. Hence comparing $\sigma(M_{d*})$ and $\sigma(\overline{M}_d)$ (in view of Theorem 7.4) is often time possible.

Theorem 7.4. d* is Schur optimal if $\sigma(M_{d*}) \leq \sigma(\overline{M}_d)$ for all $d \in \mathcal{D}$, where $\overline{M}_d$ is some average version of M_d.

Proof. $\Phi(M_{d*}) \leq \Phi(\overline{M}_d)$ where the first inequality holds from the assumption $\sigma(\overline{M}_{d*}) \leq \sigma(\overline{M}_d)$ and the latter from Lemma 7.2.

Closing remarks: We refer the reader to "Special Issue on Optimal Design Theory" No. 14, Vol. A7 (1978) of Communications in Statists (edited by this author) for further ideas, results and references. Currently we are preparing a book on the subject of optimal design of experiments. The book should be available for distribution within a year or so. Meanwhile, the interested reader can obtain preliminary versions of some chapters of the book.

REFERENCES/BIBLIOGRAPHY

1. Bellman, R. (1970). Introduction to Matrix Analysis. 2nd edition, McGraw Hill.

2. Birkhoff, G. (1946). Tres observations sobre el algebra lineal, Univ. nac. Tucuman, Revista Ser. A,5, 147-150.

3. Box, M.J. and Draper, N.R. (1971). Factorial designs, the $|X'X|$ criterion, and some related matters. Technometrics 13, 731-742.

4. Cheng, C.-S. (1978). A note on (M,S)-optimality. Commun. Statist. A7 (14), 1327-1338.

5. Cheng, C.-S. (1978). Optimality of certain asymmetrical experimental designs. Ann. Math. Statist, 6, 1239-1261.

6. Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Statist. 24, 586-602.

7. Cramer, H. (1974). Mathematical Methods of Statistics. 13th printing, Princeton University press.

8. Eccleston, J.A. and Hedayat, A. (1974). On the theory of connected designs: characterization and optimality. Ann. Statist. 2, 1238-1255.

9. Ehrenfeld, S. (1955). On the efficiency of experimental designs. Ann. Math. Statist. 26, 247-255.

10. Elfving, G. (1952). Optimal allocation in linear regression theory. Ann. Math. Statist. 23, 255-262.

11. Fedorov, V.V. (1972). Theory of Optimal Experiments, New York: Academic press.

12. Hardy, G.H., Littlewood, J.E. and Polya, G. (1967). Inequalities. Cambridge University press.

13. Hedayat, A. and Wallis, W.D. (1978). Hadamard matrices and their applications. Ann. Statist. 6, 1184-1238.

14. Hedayat, A. (1980). Theory of Optimal Design of Experiments. A book in preparation.

15. Kiefer, J. (1958). On the nonrandomized optimality and randomized non optimality of symmetrical designs. Ann. Math. Statist. 29, 675-699.

16. Kiefer, J. (1959). Optimum experimental designs. J.R. Statist. Soc. B21, 272-319.

17. Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. 2, 849-879.

18. Kiefer, J. (1975). Construction and optimality of generalized Youden designs. in A Survey of Statistical Design and Linear Models. (J.N. Srivastava, Ed.). Amsterdam: North Holland Publishing Co., 333-353.

19. Magda, C.G. (1979). On E-optimal block designs and Schur optimality. Ph.D. thesis. University of Illinois at Chicago Circle.

20. Mood, A.M. (1946). On Hoteling's weighing problem. Ann. Math. Statist. 17, 432-446.

21. Ostrowski, A. (1952). Sur quelques applications des fonctions convexés et concaves au sems de l. Schur, Journ. de Math. Pures et Appliq., 31, 253-292.

22. Seber, G.A.F. (1977). Linear Regression Analysis. Wiley

23. Shah, K.R. (1960). Optimality criteria for incomplete block designs. Ann. Math. Statist. 31, 791-794.

24. Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. Biometrika, 12 1-85.

25. Wald, A. (1943). On the efficient design of statistical investigations. Ann. Math. Statist. 14, 134-140.

26. Wilks, S.S. (1932). Certain generalizations in the analysis of variance. Biometrika 24, 471-494.