THE APPLICATION OF MATRIX THEORY

TO OPTIMAL DESIGN OF EXPERIMENTS

A thesis presented in partial fulfilment
of the requirements for the degree
of Doctor of Philosophy
at Massey University

Vernon John Thomas

1975

Abstract

A development of the theory of optimum experimental design is presented. The notation and proofs are in terms commonly used by statisticians, rather than in the earlier measure theory terms. The D-optimality equivalence theorem is extended to the singular case, and similar results derived for a number of other criteria. Atwood's theorem for special n-tic polynomials is extended to the case where not all parameters are of interest. Finally methods of constructing optimal designs are considered and extended to allow deletion of unsatisfactory points, and some numerical examples are included.

Acknowledgements

I would like to acknowledge, with gratitude, the advice and encouragement of Dr B.S. Weir and the forebearance and sympathy of my wife and family, during the preparation of this thesis.

Table of contents

1. INTRODUCTION

1.1 PURPOSE.

The justification for the work leading to this thesis can perhaps best be illustrated by quoting M.J. Box in his comments to two papers on the subject of design optimality, Laycock (1972) and Wynn (1972): "Tonight's two authors are to be congratulated on achieving a high standard of lucidity in what is a notoriously difficult subject mathematically. In the past, many authors working in this area have been at pains to conform to an abstruse but widely accepted style, with the result that their work has been only readily appreciated by a rather small and elite band of the already converted." Along the same lines, a quotation from Box and Draper (1971) is relevant:
"There is a tendency, on the part of many statisticians who are connected with experimental design, to omit mention of the excellent papers written by those who belong to what is popularly (and complimentarily) known as the "Kiefer School". The papers concerned are often long, intricate and involved, and most of them are fine pieces of detailed reasoning. They are, however, extremely difficult to read and the experimenter (or even the statistician!) searching for a design may quickly be deterred after reading a page or two."

As an example of the type of style referred to by Box, consider the following definition of M-optimality, quoted from a well-known paper on the subject:

> A design is said to be $M_{\alpha,c}$ optimal in Δ if, for some φ^*_{d*} in $H_{d*}(\alpha)$,
>
> $$\inf_{\Gamma_c} \beta_{\varphi^*_{d*}}(\mu,\sigma^2) = \max_{d\in\Delta} \sup_{\varphi\in H_d(\alpha)} \inf_{\Gamma_c} \beta_{\varphi}(\mu,\sigma^2)$$

Stripped of the difficult notation, the above expression states that a design is $M_{\alpha,c}$-optimal if there exists a test such that the

minimum power of the test, on the contour $\mu^T\mu/\sigma^2=c$, is a maximum over all designs and tests.

Another example is the following theorem:

> If $\theta^{(1)}$ is estimable under ξ^*, equations (2.1), (2.2), (2.17)(a), and (2.19)(a) are equivalent. Moreover, (2.1) (and thus any of the above) implies (2.17)(b), which is equivalent to (2.19)(b), while (2.1) is implied by (2.17)(c) (or, equivalently, (d)).

The equations referred to span five pages in the original paper.

The primary purpose of this thesis is to provide a general development of the subject of optimum design, in terms readily followed by the bulk of statisticians. This, in turn, has made it necessary to re-prove the basic theorems in matrix terms. As a result of this it has been possible to extend a number of these theorems.

1.2 GENERAL APPROACH.

The historical approach, following Kiefer (1959, 1961a, 1961b, 1962a, 1962b) and Kiefer and Wolfowitz (1959, 1960), has used a definition of a regression or experimental design as a probability measure with support at a finite number of points. This measure, usually denoted by ξ is used in integrations over the region of operability, thus

$$\int_R d\xi = 1$$

and, if $\underset{\sim}{x}$ represents a point in R,

$$\int_R \underset{\sim}{x}\underset{\sim}{x}^T \, d\xi = M(\xi)$$

is Fisher's information matrix for the design.

By contrast, statisticians tend to refer to a design matrix X, all of whose n rows are in R, as specifying a design.

The information matrix is now

$$\frac{1}{n} X^T X = S$$

the fact that S is a function of a particular design X, being implicit. This is the approach adopted in this thesis. It has been necessary to introduce a weighting matrix Λ, in order to take advantage of continuity arguments. Λ is a diagonal matrix, whose diagonal elements are positive (or at least non-negative) and add to one, and whose ith diagonal element represents the proportion of the sample taken at the point given by the ith row of X. Now

$$X^T \Lambda X = S$$

1.3 GENERALIZED INVERSES.

Earlier work, particularly Kiefer (1962b) and Karlin and Studden (1966), has made extensive use of game theory methods in obtaining their results. However, the same results may be obtained by the more natural use of generalized inverses. Apart from the specialization in section 6.3, the usual one-condition generalized inverse defined by $AA^-A = A$ is used. Where A is symmetric (as it usually is in the present context) A^- is, for convenience, taken to be symmetric also. This enables consideration of singular information matrices in a straightforward way.

Theorem 8.1 and its supporting lemmas, is an attempt to provide a proof in answer to the comment of Atwood (1969) that if S is singular, "there is no simple known theorem analogous to this result" (that, where interest is in s of the k parameters, D-optimality is equivalent to an analog of G-optimality.).

2. THE GENERAL LINEAR MODEL

2.1 BACKGROUND.

The general linear model has relevance in the context in which an experimenter wishes to draw conclusions about some physical process which produces some measurable result or response. This response may be single valued or multi-valued. This thesis will consider only the single valued case.

The experimenter wishes to draw conclusions about the underlying process, for any one of a number of reasons. To do so, he postulates a form of model under which the response depends on a number of independent variables, together with a random disturbance factor. We suppose that there are w independent variables, denoted by the wx1 vector $\underset{\sim}{\xi}$. The response will be denoted by y. Thus, the model is

$$y = f(\underset{\sim}{\xi}) + \epsilon \qquad (2.1)$$

where f is some scalar function, and ϵ is a random variable with zero mean.

The experimenter may go further, and suppose that y depends on a linear combination of k linearly independent functions x_i of $\underset{\sim}{\xi}$, thus $x_i = x_i(\underset{\sim}{\xi})$. With $\underset{\sim}{x}$ denoting the k x 1 vector with elements x_i, the model now becomes:

$$y = \underset{\sim}{x}^T \underset{\sim}{\beta} + \epsilon \qquad \text{with } \underset{\sim}{x} = \underset{\sim}{x}(\underset{\sim}{\xi}) \qquad (2.2)$$

where the constant but unknown elements of $\underset{\sim}{\beta}$ are referred to as the parameters of the model. This is the general linear model. Note that the model is linear in $\underset{\sim}{\beta}$ and $\underset{\sim}{x}$, not necessarily in $\underset{\sim}{\xi}$, polynomials being the commonest counter-example.

3. THE EXPERIMENTAL SITUATION

3.1 SPECIFICATION OF THE PROBLEM.

The experimenter then formulates the question to which he requires an answer. The commonest questions are:

1. Is $L\underset{\sim}{\beta} = \underset{\sim}{g}$ true, where $\underset{\sim}{g}$ is some known vector and L a known r x k matrix? (3.1)

2. What is the expected value of y for some $\underset{\sim}{\xi}$ of particular interest? (3.2)

3. What are the values of $\underset{\sim}{\beta}$, or of some transformation $R\underset{\sim}{\beta}$ of them? (3.3)

Of these, (3.1) is the hypothesis testing problem, (3.2) the prediction problem, and (3.3) the parameter estimation problem. The estimation problem may be further complicated when interest is restricted to some sub-set of the space of $\underset{\sim}{\xi}$, I, called the region of interest. I will be used indifferently, and without ambiguity, to represent the region defined on $\underset{\sim}{\xi}$ and the derived region defined on the space of $\underset{\sim}{x}$ (as well as its normal meaning of the identity matrix). This leads to

4. What is the expected value of y for all $\underset{\sim}{\xi} \in I$? (3.4)

3.2 THE EXPERIMENTAL DESIGN.

In order to satisfy his curiosity about one or more of the above questions, the experimenter designs an experiment. He is assumed to be able to select some $\underset{\sim}{\xi}$ within some space R (known as the region of operability and usually determined by physical limitations, or, in some cases by the limits of reliability of the form of the model given by (2.2)), and perform an experiment which yields a sample

value of y. The region, I, referred to above, may coincide with R, be a subset of it (the interpolation problem) or be completely disjoint (the extrapolation problem), or overlap R.

The single experiment noted above is repeated a number of times at the same or different values of $\underset{\sim}{\xi}$. We suppose that n such single experiments are performed in all. Let us suppose that the experimenter selects, at most, p distinct values of $\underset{\sim}{\xi}$, and that the ith of these values is used $\lambda_{ii}n$ times, where λ_{ii} is between 0 and 1. The complete set of n samplings will hereafter be termed the experiment.

Now assemble the p values of $\underset{\sim}{x}$, derived from the p values of $\underset{\sim}{\xi}$, as the rows of a matrix X, p x k. Also assemble the corresponding values of λ_{ii} as the diagonal elements of a p x p matrix Λ, whose non-diagonal elements are zero. The matrix X will be known as the design matrix, and Λ as the weighting matrix, the matrix pair $D = (X,\Lambda)$ being known as the design. Thus we introduce the following definition of a design:

Definition 3.2.1

A design is a matrix pair $D = (X,\Lambda)$ where $\underset{\sim}{x}_i^T$, the rows of X are in the region R, and Λ is a diagonal matrix whose diagonal elements are non-negative and add to one.

Certain theoretical requirements need the imposition of the constraint that the region R be a compact space, and this constraint will be assumed to hold. It is difficult to imagine any practical example in which it does not hold. As noted previously, $S = X^T\Lambda X$ is Fisher's information matrix.

3.3 DERIVED DESIGNS.

Given any point $\underset{\sim}{x}$, in R, a design can be constructed. Also, given any two designs, further designs may be constructed.

Lemma 3.3.1

If $\underset{\sim}{x} \in R$ then $(\underset{\sim}{x}^T, 1)$ is a design.

For the remainder of this thesis, we assume that $D = (X,\Lambda)$ is a design. Alternative designs will be indicated by superscript symbols, thus $D^* = (X^*,\Lambda^*)$.

Lemma 3.3.2

If $D = (X,\Lambda)$ is a design, then so is (X,Λ^*) where Λ^* is any weighting matrix satisfying definition 3.2.1.

Note that this may require $n \to \infty$ if some λ_{ii} is irrational. It does not, however, affect p.

Lemma 3.3.3

If $D = (X,\Lambda)$ is a design, and $\underset{\sim}{x} \in R$, then $D^* = (X^*,\Lambda^*)$ is a design, where

$$X^* = \begin{pmatrix} X \\ \underset{\sim}{x}^T \end{pmatrix} \qquad \Lambda^* = \begin{pmatrix} (1-\theta)\Lambda & 0 \\ 0 & \theta \end{pmatrix} \qquad 0 \le \theta \le 1$$

Corollary

If $D = (X,\Lambda)$ and $D^* = (X^*,\Lambda^*)$ are designs, then so is $D^+ = (X^+,\Lambda^+)$ where

$$X^+ = \begin{pmatrix} X \\ X^* \end{pmatrix} \qquad \Lambda^+ = \begin{pmatrix} (1-\theta)\Lambda & 0 \\ 0 & \theta\Lambda^* \end{pmatrix} \qquad 0 \le \theta \le 1 \quad (3.5)$$

The symbol D^+ will be used frequently for the design given by (3.5). Note that $S^+ = (1-\theta)S + \theta S^*$.

3.4 OTHER OPTIONS.

The experimenter is also able to choose the method of estimation, necessary even where the main interest is in hypothesis testing, and, where applicable, the test to be used.

While a maximum likelihood estimate, and the usual F-test, are used in the vast majority of cases, they do not exclude the possibility of alternatives being preferable in some circumstances. Any test will, of course, require some distributional assumption about the errors, ϵ, in the model.

In addition, the experimenter may be unsure of the adequacy of the model (2.2), and may wish to guard against bias arising from the possibility that some specified alternative model is the true model. For the purposes of this thesis the model (2.2) will be assumed to be correct.

3.5 DISTRIBUTION OF ERRORS.

Throughout this thesis, the errors ϵ will be assumed to be normally distributed, with zero mean.

While some work has been done on the case where errors in different response measurements are correlated, in the majority of work they are assumed to be uncorrelated, and this will be taken to be the case in this thesis.

The errors are also assumed to have a common variance, σ^2.

3.6 EXPLICIT MODEL.

Let U be an n x k matrix derived from D = (X,Λ) by repeating the ith row of X, $\lambda_{ii}n$ times. Evidently this is possible only if $\lambda_{ii}n$ is an integer for each i. If Λ has irrational elements on the diagonal, then it is not possible to perform the experiment

implied by the design $D = (X,\Lambda)$. Such a design may occur when an attempt is made to use an optimum design, in which case an approximate design will give a near optimum when n is sufficiently large. The approximate design modifies λ_{ii} in such a way that $\lambda_{ii}n$ values are integers. In developing the model it is assumed that this has been done and that U exists. Now the model, at the experimental points, is

$$\underset{\sim}{y} = U\underset{\sim}{\beta} + \underset{\sim}{\epsilon}, \qquad \underset{\sim}{\epsilon} \sim N(\underset{\sim}{0},\sigma^2 I) \tag{3.6}$$

Note that $U^TU = nX^T \Lambda X = nS$

Using maximum likelihood estimation, the normal equations are

$$nS\hat{\underset{\sim}{\beta}} = U^T\underset{\sim}{y} \tag{3.7}$$

which, since $SS^-U^T = U^T$, always has a solution (from the theory of generalized inverses). The solution is

$$\hat{\underset{\sim}{\beta}} = \frac{1}{n} S^-U^T\underset{\sim}{y} + (I-S^-S)\underset{\sim}{h}$$

where $\underset{\sim}{h}$ is arbitrary, and S^- is a one condition generalized inverse of S^- (that is, $SS^-S = S$). For convenience, S^- will be taken to be symmetric.

Suppose now that the aim of the experiment is to test the hypothesis $L\underset{\sim}{\beta} = \underset{\sim}{g}$, where L is r x k. This requires $LL^-\underset{\sim}{g} = \underset{\sim}{g}$, for consistency, and $LS^-S = L$ for $L\underset{\sim}{\beta}$ to be estimable, (see Pringle and Rayner (1971), for example).

If either of these conditions is not met, then useful results cannot be obtained.

The usual test is an F-test which tests

$$\frac{SS(H)/\text{rank } L}{SS(E)/(n-\text{rank } X)} \tag{3.9}$$

with degrees of freedom rank L and (n - rank U) = (n - rank X), where

$$SS(H) = (LS^{-}U^{T}\underset{\sim}{y} - \underset{\sim}{g})^{T}Q\,(LS^{-}U^{T}\underset{\sim}{y} - \underset{\sim}{g}) \tag{3.10}$$

where $Q = \frac{1}{n}(LS^{-}L^{T})^{-}$, and

$$SS(E) = y^{T}(I - \tfrac{1}{n}US^{-}U^{T})y \tag{3.11}$$

with non-centrality parameter

$$\frac{1}{\sigma^{2}}(L\underset{\sim}{\beta} - \underset{\sim}{g})^{T}Q^{-}QQ^{-}(L\underset{\sim}{\beta} - \underset{\sim}{g})$$

In the equivalent estimation problem, the variance of $L\hat{\underset{\sim}{\beta}}$ is given by

$$\text{var } L\hat{\underset{\sim}{\beta}} = \frac{1}{n}\sigma^{2}LS^{-}L^{T} \tag{3.12}$$

again, provided $L\underset{\sim}{\beta}$ is estimable.

3.7 DEVELOPMENT OF A CANONICAL FORM

Algebraic manipulation of expressions such as those given above can be cumbersome. To lessen this, a number of writers use a canonical form of the model. In particular Lehmann (1959) gives a general account, without existence proofs, and Kendall and Stuart (1967) Vol 2, give a development which contains basic errors. (In the equation before (24.90), the left hand side is, at most, of rank k, and the right hand side is of rank n, in general $n > k$). Being unable to find any mention of the less than full rank case, or any explicit form for the transformation, the author has developed the model given below, to fill this gap, and, in particular, to cover the singular case. The model is that of the last section, and the aim is to test the hypothesis $L\underset{\sim}{\beta} = \underset{\sim}{g}$, with $LL^{-}\underset{\sim}{g} = \underset{\sim}{g}$ and $LS^{-}S = L$ (that is, $L\underset{\sim}{\beta} = \underset{\sim}{g}$ is consistent, and $L\underset{\sim}{\beta}$ is estimable). Note that this requires rank $L \leq$ rank S = rank X.

We desire to transform $\underset{\sim}{y}$ by $\underset{\sim}{z} = C\underset{\sim}{y} + \underset{\sim}{b}$ with a number of specified characteristics. Partition $\underset{\sim}{z}$ (and C and $\underset{\sim}{b}$ correspondingly) in such a way that $\underset{\sim}{z}_1$ is (rank L) x 1, $\underset{\sim}{z}_2$ is (n - rank X) x 1, and $\underset{\sim}{z}_3$ is (rank X - rank L) x 1. The conditions we set on C and $\underset{\sim}{b}$ are, for any $\underset{\sim}{\beta}$,

$$\left.\begin{array}{ll} 1. & \text{The transformation must be reversible.} \\ 2. & \text{var } \underset{\sim}{z} = \sigma^2 I \\ 3. & E(\underset{\sim}{z}_2) = \underset{\sim}{0} \\ 4. & E(\underset{\sim}{z}_1) = \underset{\sim}{0} \text{ if and only if } L\underset{\sim}{\beta} = \underset{\sim}{g}. \end{array}\right\} \qquad (3.13)$$

Condition 1.

This implies that C must be non-singular.

Condition 2.

This implies that $CC^T = I$ and hence that $C^T = C^{-1}$, and the C_i (the partitions of C) are orthogonal and orthonormal.

Condition 3.

This imples that $C_2U\underset{\sim}{\beta} - \underset{\sim}{b}_2 = \underset{\sim}{0}$ for any $\underset{\sim}{\beta}$, and hence that $\underset{\sim}{b}_2 = \underset{\sim}{0}$ and $C_2U = 0$. Thus C_2 is any (n - rank X) x n matrix, orthonormal and in the null space of U. In particular, C_2 may be taken as any orthonormal basis for $I - \frac{1}{n} US^{-}U^T$.

Condition 4.

This requires that $C_1U\underset{\sim}{\beta} + \underset{\sim}{b}_1 = \underset{\sim}{0}$ if and only if $L\underset{\sim}{\beta} = \underset{\sim}{g}$. It is readily shown (by solving $L\underset{\sim}{\beta} = \underset{\sim}{g}$ and inserting in $C_1U\underset{\sim}{\beta} + \underset{\sim}{b}_1 = \underset{\sim}{0}$) that a necessary condition is

$$LS^{-}U^T(I - C_1^TC_1) = 0$$

and we require a full rank solution C_1, such that $C_1C_1^T = I$. Such a solution is given by

$$C_1 = PNS^{-}U^T$$

where L = MN with N of full rank and dimension rank L x k, M is of full rank and dimension r x rank L, and

P is of full rank and dimension rank L x rank L, given by

$$P^TP = n(NS^-N^T)^{-1}$$

It can be shown that $C_1U\underset{\sim}{\beta} + \underset{\sim}{b}_1 = \underset{\sim}{0}$ if and only if $L\underset{\sim}{\beta} = g$, and that $C_1C_1^T = I$, $C_1C_2^T = 0$. Also, $\underset{\sim}{b}_1 = -nPNL^-\underset{\sim}{g}$.

Reverting to condition 1, C_3 is chosen to make C of full rank. Thus any conformable matrix orthonormal and orthogonal to C_1 and C_2 will suffice. Finally, $\underset{\sim}{b}_3$ is arbitrary and may be set to zero.

3.8 SUMMARY OF CANONICAL FORM.

Summarizing the above results:

Let $$\underset{\sim}{z} = C\underset{\sim}{y} + \underset{\sim}{b} = \begin{pmatrix} \underset{\sim}{z}_1 \\ \underset{\sim}{z}_2 \\ \underset{\sim}{z}_3 \end{pmatrix} \qquad (3.20)$$

where $$\underset{\sim}{b} = \begin{pmatrix} -nPNL^-\underset{\sim}{g} \\ 0 \\ 0 \end{pmatrix}$$

$$C = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}$$

where $C_1 = PNS^-U^T$, C_2 is an orthonormal basis for $I - US^-U^T$, and C_3 is any matrix orthonormal and orthogonal to C_1 and C_2, with $L = MN$ and $P^TP = n(NS^-N^T)^{-1}$.

The problem can now be phrased: given a vector $\underset{\sim}{z}$ of independent normal variates with common variance, with $E(\underset{\sim}{z}_2) = \underset{\sim}{0}$, test the hypothesis $\underset{\sim}{z}_1 = \underset{\sim}{0}$. Under these conditions,

$$\left.\begin{aligned} SS(H) &= \underset{\sim}{z}_1^T\underset{\sim}{z}_1 \\ SS(E) &= \underset{\sim}{z}_2^T\underset{\sim}{z}_2 \end{aligned}\right\} \qquad (3.21)$$

with non-centrality parameter $\underset{\sim}{\mu}^T\underset{\sim}{\mu}/\sigma^2$ where $\underset{\sim}{\mu} = E(\underset{\sim}{z}_1)$.

3.9 SIMPLIFIED MODEL.

It is difficult, with the above canonical form, to see the immediate application of results. For this reason an alternative and intermediate form will be used in this thesis. With the notation of sections 3.6 and 3.7, let $L = MN_1$ and augment N_1 to a non-singular square matrix N.

Now let

$$\underset{\sim}{\theta} = N\underset{\sim}{\beta} - \begin{pmatrix} N_1L^-\underset{\sim}{g} \\ \underset{\sim}{0} \end{pmatrix} \qquad (3.22)$$

whence

$$\underset{\sim}{\beta} = N^{-1}\underset{\sim}{\theta} + N^{-1}\begin{pmatrix} N_1L^-\underset{\sim}{g} \\ \underset{\sim}{0} \end{pmatrix} \qquad (3.23)$$

and we use the model

$$y - UN^{-1}\begin{pmatrix} N_1L^-\underset{\sim}{g} \\ \underset{\sim}{0} \end{pmatrix} = UN^{-1}\underset{\sim}{\theta} + \underset{\sim}{\epsilon} \qquad (3.24)$$

If $L\underset{\sim}{\beta} = \underset{\sim}{g}$, then $MN\underset{\sim}{\beta} - MNL^-\underset{\sim}{g} = \underset{\sim}{0}$, whence premultiplication by $(M^TM)^{-1}M^T$ shows that $\underset{\sim}{\theta}_1 = \underset{\sim}{0}$. Conversely, if $\underset{\sim}{\theta}_1 = \underset{\sim}{0}$, $N_1\underset{\sim}{\beta} = N_1L^-\underset{\sim}{g}$ and premultiplication by M gives $L\underset{\sim}{\beta} = \underset{\sim}{g}$.
Thus $L\underset{\sim}{\beta} = \underset{\sim}{0}$ is equivalent to $\underset{\sim}{\theta}_1 = \underset{\sim}{0}$.

We may thus, without loss of generality, take the hypothesis to be $\underset{\sim}{\beta}_1 = \underset{\sim}{0}$, where $\underset{\sim}{\beta}_1$ is s x 1, with s = rank L. The above formulae give the means to proceed to the full model, where necessary. The form of the model is unchanged from that given in section 3.6, and the formulae of that section apply, with

$$N = L = (I \; 0) \qquad \underset{\sim}{\beta} = \underset{\sim}{0} \tag{3.25}$$

We will partition U (and X in the general model) and S to conform to the partitioning of $\underset{\sim}{\beta}$. Pringle and Rayner (1971) give a generalized inverse of a partitioned matrix. Applying this to S gives

$$S^- = \begin{pmatrix} Q^- & -Q^- S_{12} S_{22}^- \\ -S_{22}^- S_{21} Q^- & S_{22}^- + S_{22}^- S_{21} Q^- S_{12} S_{22}^- \end{pmatrix} \tag{3.26}$$

where $Q = S_{11} - S_{12} S_{22}^- S_{21}$ and has the same meaning in the present context as Q in section 3.6.

Hence $$\hat{\underset{\sim}{\beta}}_1 = \frac{1}{n}(Q^- U_1^T - Q^- S_{12} S_{22}^- U_2^T)\underset{\sim}{y}$$

Now $$E(\hat{\underset{\sim}{\beta}}_1) = \frac{1}{n}(Q^- U^T - Q^- S_{12} S_{22}^- U_2^T)(U_1 \underset{\sim}{\beta}_1 + U_2 \underset{\sim}{\beta}_2)$$

$$= \underset{\sim}{\beta}_1$$

Combining (3.25) and (3.26) gives $LS^- L^T = Q^-$, which, if $\underset{\sim}{\beta}_1$ is to be estimable, must be non-singular, hence, in this case $Q^- = Q^{-1}$, and (3.10) becomes

$$\left.\begin{aligned} SS(H) &= \hat{\underset{\sim}{\beta}}_1^T Q \, \hat{\underset{\sim}{\beta}}_1 \\ SS(E) &= \underset{\sim}{y}^T (I - \tfrac{1}{n} U S^- U^T) \underset{\sim}{y} \end{aligned}\right\} \tag{3.27}$$

with non-centrality parameter $\underset{\sim}{\beta}_1^T Q^{-1} \underset{\sim}{\beta}_1 / \sigma^2$, and

$$\text{var } \hat{\underset{\sim}{\beta}}_1 = \frac{1}{n}\sigma^2 Q^{-1} \qquad (3.28)$$

This is the form of the model which will be used in the succeeding sections.

On occasion we may refer to the case in which $L\underset{\sim}{\beta}_1$ is not estimable. In such a case the formulae in (3.27) are meaningless, but

$$\text{var } \hat{\underset{\sim}{\beta}}_1 = \frac{1}{n}\sigma^2 Q^- Q Q^-$$

where Q^- is, for convenience, taken to be symmetric.

3.10 SIZE OF THE EXPERIMENT.

As a practical consideration, the experimenter does not want to have the number of distinct points in his experiment to be indefinitely large. It will be seen later that the practical design criteria depend only on S, rather than explicitly on Λ or X. Thus it is sufficient to consider the set of possible S matrices. The following theorem enables a limit to be placed on the number of rows in X.

Caratheodory's Theorem (from Fedorov (1972))

Each point $\underset{\sim}{y}^*$ in the convex hull V* of any subset V of m-dimensional space can be represented in the form

$$\underset{\sim}{y}^* = \sum_{i=1}^{m+1} \alpha_i \underset{\sim}{v}_i$$

where $\alpha_i \geq 0$, $\sum_{i=1}^{m+1} \alpha_i = 1$, $\underset{\sim}{v}_i \in V$.

Also, if $\underset{\sim}{y}^*$ is a boundary point of the set V*, then $\alpha_{m+1} = 0$.

The matrix S has $k(k+1)/2$ distinct elements and may thus be considered a vector in $m=k(k+1)/2$ dimensional space. Consider now the set of S matrices arising from experiments concentrated at a single point, that is, the set of S matrices such that $S = \underset{\sim}{x}\underset{\sim}{x}^T$. Let this be the subset V in Caratheodory's theorem. Now, any general S matrix has the form

$$S = \sum_{i=1}^{p} \lambda_{i\,i} \underset{\sim}{x}_i \underset{\sim}{x}_i^T$$

where each $\underset{\sim}{x}_i \underset{\sim}{x}_i^T$ is represented by a vector in V. Thus S is within the convex hull of V, and thus, by Caratheodory's theorem, can be represented by

$$S = \sum_{i=1}^{\frac{1}{2}k(k+1)+1} \alpha_i \underset{\sim}{x}_i \underset{\sim}{x}_i^T$$

Thus, any S can be derived from at most $k(k+1)/2 + 1$ distinct points, and hence $p \leq k(k+1)/2 + 1$.

4. OPTIMALITY AND TESTS OF HYPOTHESIS

4.1 GENERAL CONSIDERATIONS.

Unfortunately, optimality is not a uniquely defined quality. In the area of hypothesis testing - in general, the experimenter requires his test, assumed to be of size α ($0 < \alpha < 1$), to be as powerful as possible. However, it can be shown that, for the general linear model, a universally most powerful (UMP) test does not exist, unless r=1 where r is the number of rows of L. This has the effect of reducing L to a vector, or in the simplified model, of reducing the test to a test of a single parameter. This result was established by Kolodzieczyk (1935). Where the hypothesis is simple, that is, where r=1, the ordinary F-test is UMP.

Kiefer (1958) has summarized most of the following results.

4.2 UNBIASED TESTS.

Certain tests may have optimal properties. A test whose power function is not less than α, the size of the test, for any value of the parameters is said to be unbiased. Note that, since the test is of size α, at the null hypothesis its power is not greater than α. Hence, for an unbiased test, it must be exactly α at the null hypothesis. It should be noted that the term "bias" in this context is completely distinct from the meaning in the estimation context. A biased test is one which is more likely to reject the null hypothesis when it is true, than when it is false, for some specific alternative hypothesis. The importance of this criterion is that it is a natural seeming requirement, and by restricting attention to unbiased tests, in cases where no UMP test exists, a test which is UMP among unbiased tests (UMPU) may sometimes be found. It can, unfortunately, be shown that no UMPU test exists for the general linear hypothesis.

4.3 LOCALLY UNBIASED TESTS.

The situation is improved if the requirement of the last section is relaxed somewhat. An alternative is that the power of the test at the null hypothesis be equal to α, and that the power be non-decreasing in the neighbourhood of the null hypothesis. Thus, if π is the power of the test, a function of $\underset{\sim}{\beta}_1$ and σ^2 in the simplified model, a test is locally unbiased if

1. $\pi = \alpha$ at H_0 (4.1)

2. $\dfrac{\partial \pi}{\partial \underset{\sim}{\beta}_1} = \underset{\sim}{0}$ at H_0 (4.2)

3. The matrix $\dfrac{\partial^2 \pi}{\partial \underset{\sim}{\beta}_1 \partial \underset{\sim}{\beta}_1^T}$ is non-negative definite at H_0 (4.3)

The determinant of the matrix in condition 3 is the Gaussian curvature of the power function. If this is positive definite, the test is said to be strictly locally unbiased.

4.4 SIMILAR TESTS.

A similar test is one for which $\pi = \alpha$ at H_0. Since all unbiased tests are similar tests, consideration of similar tests may sometimes lead to UMPU tests.

4.5 TYPE D AND E TESTS.

Isaacson (1951) introduced the notion of type D tests, extending the ideas of Neyman and Pearson (1936 and 1938). A type D test is one which is strictly locally unbiased, and for which the Gaussian curvature at the null hypothesis is a maximum. An important characteristic of type D tests is that they are invariant under transformations of the parameter space which are locally one-to-one and which are twice differentiable with continuous partial derivatives. Isaacson extended his idea of type D tests to include type E tests which have the same requirements, but

which include nuisance parameters. E type tests are therefore appropriate to the simplified model, D type to the model where L is of rank k (the number of parameters in the model).

4.6 INVARIANCE.

Suppose the problem is put in the canonical form of section 3.6. The form of the problem suggests that we might reasonably require certain properties from the test used. For example we would not like the test to give a different result if we reorganized the various matrices in such a way that the same values of $\underset{\sim}{z}_1$ occurred, but in a different order. Thus we require the test to be invariant over all the orderings of the elements of $\underset{\sim}{z}_1$. A similar remark applies to the elements of $\underset{\sim}{z}_2$ and $\underset{\sim}{z}_3$.

Considering the method of derivation of the canonical form, the matrix C could be premultiplied by a block diagonal orthogonal matrix without violating any of the requirements for C. Thus it is reasonable to expect the test to be invariant under orthogonal transformations of $\underset{\sim}{z}_1$ and $\underset{\sim}{z}_2$. In addition, $\underset{\sim}{b}_3$ was arbitrary, and the test should therefore be invariant under the addition of a constant vector to $\underset{\sim}{z}_3$. Finally, multiplication of $\underset{\sim}{z}$ by a constant affects only σ^2, and we expect the test to reflect this.

Thus we might reasonably expect that the test be invariant under the above group of transformations. It must be emphasized that invariance is not a necessary criterion. Kiefer (1958) has given examples in which tests which are not invariant in the above sense have greater power, in some circumstances, than the corresponding invariant test. In each case, however, some subset of $\underset{\sim}{\beta}_1$ is selected for special treatment.

4.7 THE F-TEST.

The ordinary F-test is derived as the likelihood ratio test for the general linear hypothesis. It has the characteristic that its

power function is a function of $\lambda = \underset{\sim}{\beta}_1^T Q \underset{\sim}{\beta}_1/\sigma^2$ alone, given by

$$\pi = \int_{F'_\alpha(\nu_1,\nu_2,0)}^{\infty} dG[F'(\nu_1,\nu_2,\lambda)] \tag{4.4}$$

where ν_1, and ν_2 are the degrees of freedom, F'_α is the $100(1-\alpha)$ percentage point of the distribution, and $dG(F')$ is the density function of the non-central F distribution, given by

$$dG(F') = e^{-\frac{1}{2}\lambda} \sum_{i=0}^{\infty} \frac{\frac{(\frac{1}{2}\lambda)^i}{i!} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}\nu_1+i}}{B(\frac{1}{2}\nu_1+i, \frac{1}{2}\nu_2)} \frac{(F')^{\frac{1}{2}\nu_1+i-1}}{\left(1+\frac{\nu_1}{\nu_2}F'\right)^{\frac{1}{2}(\nu_1+\nu_2)+i}} dF' \tag{4.5}$$

B being the appropriate beta function. Evidently

$$\pi = \sum_{i=0} a_i\lambda^i = \alpha + \sum_{i=1} a_i\lambda^i \tag{4.6}$$

by evaluating the constant term.

Condition (4.1), $\pi = \alpha$ at the null hypothesis ($\lambda = 0$, in this case) is thus satisfied. Also

$$\frac{\partial\pi}{\partial\underset{\sim}{\beta}_1} = \frac{\partial\pi}{\partial\lambda}\frac{\partial\lambda}{\partial\underset{\sim}{\beta}_1} = \frac{2}{\sigma^2}\sum_{i=1}^{\infty} i a_i\lambda^{i-1} Q\underset{\sim}{\beta}_1 \text{ since } \frac{\partial\lambda}{\partial\underset{\sim}{\beta}_1} = \frac{2}{\sigma^2}Q\underset{\sim}{\beta}_1 \tag{4.7}$$

and, since $\underset{\sim}{\beta}_1 = \underset{\sim}{0}$ at the null hypothesis, so is $\frac{\partial\pi}{\partial\underset{\sim}{\beta}_1}$.

$$\frac{\partial^2\pi}{\partial\underset{\sim}{\beta}_1\partial\underset{\sim}{\beta}_1^T} = \frac{2}{\sigma^4}\sum_{i=2}^{\infty} i(i-1)a_i\lambda^{i-2}Q\underset{\sim}{\beta}_1\underset{\sim}{\beta}_1^T Q + \frac{2}{\sigma^2}\sum_{i=1} i a_i\lambda^{i-1}Q \tag{4.8}$$

and at the null hypothesis this becomes

$$\frac{2a_1}{\sigma^2}Q \tag{4.9}$$

Now π is a strictly increasing function of $\lambda \geq 0$, and hence $a_1 > 0$. Thus the matrix (4.9) is non-negative definite, since Q is. Also, if $\underset{\sim}{\beta}_1$ is estimable the matrix is positive definite. Thus the F-test is locally unbiased or strictly locally unbiased, according to whether $\underset{\sim}{\beta}_1$ is not estimable or estimable.

In addition, note that the F-test statistic is invariant under the transformations given in section 4.6. Lehmann (1959) has shown that it is UMP among tests which are invariant in this sense.

4.8 WALD'S THEOREM.

The following result is due to Wald (1942), with simpler proofs given by Wolfowitz (1949) and Lehmann (1950):

> The F-test maximizes the integral of the power function on the surface $\lambda = c > 0$, among similar tests of size α.

4.9 HSU'S THEOREM.

Hsu (1941) showed that the F-test is UMP among tests of size α whose power function is a function of λ only. As noted by Kiefer (1958) this may also be derived from Wald's Theorem.

4.10 F-TEST OF TYPE D OR E.

Kiefer (1958) has shown, using Wald's Theorem, that the F-test is of D or E type, depending on whether all the parameters are being tested.

4.11 USE OF THE F-TEST.

The comments of the previous sections give no incontrovertible justification for the use of the F-test. Of the qualities

noted, invariance seems to be the most compelling. In addition there are general reasons for using the F-test, for example it is well understood, easily calculated, and well tabulated. All this should not blind one to the fact that, in a given situation, it is not necessarily completely optimal.

5. DESIGN OPTIMALITY CRITERIA FOR HYPOTHESIS TESTING

5.1 M-OPTIMALITY.

This is the most general type of optimality, defined by Kiefer (1958). A design is said to be M-optimal if there exists a test whose minimum power on the contour $\underset{\sim}{\beta}_1^T W \underset{\sim}{\beta}_1/\sigma^2 = c$, with $c > 0$, is at least as great as the minimum power on the same contour for any design and any test and for any value of c. Unfortunately, M-optimal designs are very difficult to characterize, and the criterion has little practical application. W defines the contour on which the power function is to be examined.

5.2 L-OPTIMALITY.

This criterion, also defined by Kiefer (1958), is essentially a local (that is, near to the null hypothesis $\underset{\sim}{\beta}_1 = 0$) version of M-optimality. Letting the greatest minimum power on the $\underset{\sim}{\beta}_1^T W \underset{\sim}{\beta}_1 = c$, $c > 0$, contour, referred to above, be π_{max}, a design is said to be L-optimal if there exists a test of size α whose minimum power on the same contour, denoted by π is such that

$$\lim_{c \to 0} \frac{\pi - \alpha}{\pi_{max} - \alpha} = 1$$

If this holds for all α $(0 < \alpha < 1)$ then the design is L-optimal.

This criterion is slightly more tractable than M-optimality.

5.3 D-OPTIMALITY.

This, in the hypothesis testing context, is a restricted version of L-optimality, in which the tests are restricted to type D or E. In this case, a D-optimal design is one which maximizes the Gaussian curvature of the power of the test. By (4.9) this is the design which maximizes det Q. Wald (1943) and Ehrenfeld (1955)

considered this type of optimality. When a subset of the $\underset{\sim}{\beta}$ is considered, as in the simplified model, this criterion is sometimes (for example Whittle (1973)) called D_s-optimality.

5.4 E-OPTIMALITY.

Wald (1943) and Ehrenfeld (1955) also considered E-optimality (named, apparently, after Ehrenfeld, and having no connection with E type tests). Suppose attention is restricted to F-tests only, then an M-optimal design, within this limitation, is termed an E-optimal design.

Since the test is now fixed, and the power of the test is a monotone increasing function of λ, M-optimality now requires that the design be such that the minimum value of λ on the contour $\underset{\sim}{\beta}_1^T W \underset{\sim}{\beta}_1/\sigma^2 = c$ be a minimum for all c, simultaneously. A Lagrange multiplier θ is used to minimize

$$\underset{\sim}{\beta}_1^T Q \underset{\sim}{\beta}_1/\sigma^2 - \theta(\underset{\sim}{\beta}_1^T W \underset{\sim}{\beta}_1/\sigma^2 - c)$$

At this minimum $Q\underset{\sim}{\beta}_1 = \theta W \underset{\sim}{\beta}_1$ whence $\theta = \underset{\sim}{\beta}_1^T Q \underset{\sim}{\beta}_1/c\sigma^2$ is, taking W=I, proportional to the maximum eigenvalue of Q, and an E-optimal design is one which minimizes this greatest eigenvalue.

5.5 COMPARISON OF CRITERIA.

Kiefer (1958) pointed out that E-optimality makes an ad hoc assumption that the F-test is to be used. He then quoted a rather artificial example (intended purely to make this point) for which E-optimality was achieved, and then showed that the design was neither M-optimal nor L-optimal, and finally concluded that L- or D-optimality, though local, seem preferable to E-optimality.

6. DESIGN OPTIMALITY CRITERIA FOR ESTIMATION

6.1 GENERAL CONSIDERATIONS.

Reverting to the discussion on the experimenter's aims, given in section 3.1 and in succeeding sections, problems (3.2), (3.3) and (3.4) have not yet been covered. In each of these we attempt to estimate some value or vector of values. In such a case it is reasonable to require that an optimal design be a design that minimizes the variance of the estimate, or the generalized variance where interest lies in more than one estimate, and we will first consider such criteria. The problems now become

for (3.3): minimize: $$\text{var}(R\hat{\underset{\sim}{\beta}}) = \frac{1}{n}\sigma^2 RS^{-}R^{T} \qquad (6.1)$$

for (3.2): minimize: $$\text{var}(\frac{1}{n}\underset{\sim}{x}^{T}S^{-}U^{T}y) = \sigma^2 \underset{\sim}{x}^{T}S^{-}SS^{-}\underset{\sim}{x} \qquad (6.2)$$

where $\underset{\sim}{x}$ is a known vector.

For (3.4) we may minimize the average variance or the maximum variance, thus

$$\text{minimize:} \int_{\underset{\sim}{x}\in I} \underset{\sim}{x}^{T}S^{-}SS^{-}\underset{\sim}{x}\, d\underset{\sim}{x} \qquad (6.3)$$

(known as Q-optimality) or, alternatively,

$$\text{minimize:} \max_{\underset{\sim}{x}\in I} \underset{\sim}{x}^{T}S^{-}SS^{-}\underset{\sim}{x} \qquad (6.4)$$

Evidently (6.1) is not meaningful in itself, since one cannot, in general, minimize a matrix. Thus, normally, a function of the matrix is used, the generalized variance being the most common, thus

$$\text{minimize:} \quad \det RS^{-}R^{T} \qquad (6.5)$$

This is D-optimality (in the simplified model of section 3.9, the matrix becomes Q^{-1}).

Alternatively, one might seek to minimize the average variance of the parameters, or the maximum variance, giving

$$\text{minimize:} \quad \frac{1}{r} \sum_{i=1}^{r} \left[\frac{1}{n} \sigma^2 \, RS^{-}R^{T}\right]_{ii}$$

$$\text{minimize:} \quad \max_{i=1,r} \left[\frac{1}{n} \sigma^2 \, RS^{-}R^{T}\right]_{ii}$$

The first of these is equivalent to

$$\text{minimize:} \quad \text{trace } RS^{-}R^{T} = \text{trace } S^{-}R^{T}R. \tag{6.6}$$

and second to

$$\text{minimize:} \quad \max_{i=1,r} \left[RS^{-}R^{T}\right]_{ii} \tag{6.7}$$

6.2 SIMPLIFIED MODEL.

Criteria (6.5), (6.6) and (6.7) may be put in terms of the simplified model, as follows

for (6.5),	maximize:	$\det Q$	(6.8)
for (6.6),	minimize:	$\text{trace } Q^{-1}$	(6.9)
for (6.7),	minimize:	$\max_{i=1,r} \left[Q^{-1}\right]_{ii}$	(6.10)

As noted above, (6.8) is D-optimality, discussed in section 5.3. The other two have been discussed by various writers. Criterion (6.9), called A-optimality, has been considered by Elfving (1952), Stone (1959), and Kiefer (1959), and (6.10) by Murty (1971). In addition, Kiefer (1959) mentions a criterion (R-optimality) related to A-optimality,

$$\text{maximize:} \quad \text{trace } Q \tag{6.11}$$

Note that, although "minimize: $\det Q^{-1}$" and (6.8) are equivalent, (6.9) and (6.11) are not.

Stone (1959) considered an optimality criterion based on Shannon's information measure. This criterion was first mentioned by Lindley (1956). Stone took the information about $\underset{\sim}{\beta}$ after the experiment less the information before the experiment as his measure. In order to develop his results, he had to assume that $\underset{\sim}{\beta}$ was a random variable with a multivariate normal distribution, with non-singular covariance matrix. Since this is not the case in the present model, his results will not be considered further.

Another optimality criterion, c-optimality, discussed by Wynn (1972) seeks to minimize the variance of $\underset{\sim}{c}^T\underset{\sim}{\beta}$, where $\underset{\sim}{c}$ is known. This may be considered as a special case of most of the above, and, in particular of D-optimality, with s=1.

6.3 RESPONSE ESTIMATION.

Criteria (6.2), (6.3) and (6.4) cannot be couched in terms of the simplified model, referring as they do to the estimate of y at the point $\underset{\sim}{x}$, which, in turn, uses all of $\underset{\sim}{\beta}$.

All these criteria involve the expression $\underset{\sim}{x}^T S^- S S^- \underset{\sim}{x}$. Now the generalized inverse used for S^- is any solution of $SS^-S = S$. If P is such that (and such a P always exists)

$$PSP^T = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

then it can be shown (Pringle and Rayner (1971), for example) that S^- must be of the form

$$S^- = P^T \begin{pmatrix} I & U \\ V & W \end{pmatrix} P$$

where U, V, and W are completely arbitrary.

We have already imposed the restriction that the generalized inverse be symmetric. This requires $V = U^T$ and $W = W^T$. If, in addition, we require $W = U^TU$, then $S^-SS^- = S^-$. Making this requirement simplifies (6.2), (6.3) and (6.4) accordingly. No generality is lost by this process.

Note also that $\underset{\sim}{x}^TS^-\underset{\sim}{x} = \text{trace } S^-\underset{\sim}{x}\underset{\sim}{x}^T$. Thus (6.2) and (6.3) (as well as A-optimality ((6.6) or (6.9)) and c-optimality) are specializations of

$$\text{minimize trace } S^-B \tag{6.12}$$

where B is a known square matrix, often of rank 1. This area will be considered in Chapter 9, where (6.12) is termed B-optimality.

By contrast, (6.4) is something of a special case, known as G-optimality. It is not directly defined (in terms of variance of estimates) for the simplified model, but an analogous expression which reduces to G-optimality when s=k is given by

$$\text{minimize} \quad \max_{\underset{\sim}{x}\in R} (\underset{\sim}{x}^TS^-\underset{\sim}{x} - \underset{\sim}{x}_2^TS_{22}^-\underset{\sim}{x}_2) \tag{6.13}$$

The expression in parentheses will be denoted by $d(\underset{\sim}{x},D)$, D being the design which gives rise to S. It will be shown in a later section that G-optimality, defined in this manner, is equivalent to D-optimality, that is, a design which is G-optimal is also D-optimal, and vice versa. Note, for future reference, that, by expanding S^- as a partitioned matrix,

$$d(\underset{\sim}{x},D) = \underset{\sim}{x}^TS^-\underset{\sim}{x} - \underset{\sim}{x}_2^TS_{22}^-\underset{\sim}{x}_2$$

$$= (\underset{\sim}{x}_1 - S_{12}S_{22}^-\underset{\sim}{x}_2)^TQ^{-1}(\underset{\sim}{x}_1 - S_{12}S_{22}^-\underset{\sim}{x}_2) \tag{6.14}$$

7. GENERAL OPTIMALITY

7.1 φ-OPTIMALITY.

Many of the criteria mentioned in the last chapter are special cases of what Whittle (1973) has called φ-optimality.

Suppose φ is a scalar function of the design $D = (X, \Lambda)$. Assume that φ is concave. This implies that

$$\varphi^+ \geq (1 - \theta)\varphi + \theta\varphi^* \qquad 0 \leq \theta \leq 1 \qquad (7.1)$$

where φ and φ* are defined on the designs D and D*, and φ^+ is the function of $D^+ = (X^+, \Lambda^+)$ with

$$X^+ = \begin{pmatrix} X \\ X^* \end{pmatrix} \qquad \Lambda^+ = \begin{pmatrix} (1-\theta)\Lambda & 0 \\ 0 & \theta\Lambda^* \end{pmatrix}$$

A sufficient condition for concavity is that

$$\left.\frac{\partial^2 \varphi^+}{\partial \theta^2}\right|_{\theta=0+} \leq 0 \text{ for all } D, D^* \qquad (7.2)$$

the derivative being assumed to exist.

Define

$$\Phi = \left.\frac{\partial \varphi}{\partial \theta}\right|_{\theta=0+} \qquad (7.3)$$

Note that $\Phi = 0$ when $D = D^*$.

The maximal rate of ascent of φ from D towards D^*, with increasing θ, is given by

$$\rho = \sup_{D^*} \Phi \tag{7.4}$$

Whittle's theorem is now

> If φ is a concave scalar function of the design D, then a φ-optimal design can be characterized by any of the following three conditions:
>
> D maximizes φ (7.5)
>
> D minimizes ρ (7.6)
>
> $\rho = 0$ (7.7)

Proof:

> The continuity of the elements of $\underset{\sim}{x}$ and the compactness of R ensure that a design that maximizes φ does exist. At such a point $\Phi \leq 0$ for all D^*, hence $\rho \leq 0$. However, $\Phi = 0$ when $D = D^*$, hence $\rho = 0$. Thus (7.5) implies (7.7). Furthermore $\rho \geq \Phi$ for any D^*, and $\Phi = 0$ for $D = D^*$, hence $\rho \geq 0$. But the lower bound is reached for at least one design (that which maximizes φ) and hence (7.6) and (7.7) are equivalent. Finally, in $\rho \geq 0$, equality holds only at a maximum of φ, and hence (7.7) implies (7.5).

7.2 APPLICATIONS.

As noted above, many of the criteria considered are special cases of Whittle's theorem. The application to particular criteria will be considered in succeeding chapters.

8. APPLICATION OF φ-OPTIMALITY TO D-OPTIMALITY.

8.1 PRELIMINARY RESULTS

First note that det Q is a maximum if and only if log det Q is a maximum. Let φ = log det Q, where Q is as mentioned in criterion (6.8). Throughout this section, φ will be so defined. In addition we assume det Q > 0 for some design, for if this is not the case, any design is D-optimal.

A number of preliminary lemmas are required.

Lemma 8.1.1

If A_i is contained in the row space of B_i and B_i is non-negative definite, where A_i and B_i are r x r, for all i, then

$$W = \Sigma\ A_i B_i^{-} A_i^T - (\Sigma\ A_i)(\Sigma\ B_i)^{-}(\Sigma\ A_i^T) \qquad (8.1)$$

is non-negative definite. Summations are taken over the n values of i.

Proof

B_i can be expressed as $P_i^T P_i$ and A_i as $K_i P_i$.
Let

$$D = \begin{pmatrix} P_1(P_1^TP_1)^{-}P_1^T & & 0 \\ & \ddots & \\ 0 & & P_n(P_n^TP_n)^{-}P_n^T \end{pmatrix}$$

$$P = \begin{pmatrix} P_1 \\ \vdots \\ P_n \end{pmatrix} \qquad K = (K_1 \ \ldots \ K_n) \qquad E = P(P^TP)^{-}P^T$$

Now $W = K(D - E)K^T$

Also $DP = P$, $P^TD = P^T$, $D^2 = D$, $E^2 = E$ and hence $DE = E = ED$, and thus $D - E$ is idempotent, or $W = VV^T$ where $V = K(D - E)$ which establishes the result.

Corollary 1 $W = 0$ if and only if $A_i = (\Sigma A_i)(\Sigma B_i)^- B_i$

Proof $W = 0$ implies that $KD = KE$ or $A_i B_i^- P_i^T = (\Sigma A_i)(\Sigma B_i)^- P_i^T$. The proof in one direction follows by post-multiplication by P_i, noting that $A_i B_i^- B_i = A_i$. The proof of the converse is obtained by postmultiplying the equation on the right of the corollary by $B_i^- A_i$ and summing over i.

Corollary 2 Kiefer's lemma (Kiefer (1959), lemma 3.2) Kiefer's result is a special case, with B_i non-singular, since in such a case A_i is necessarily contained in the row space of B_i. Kiefer also included an arbitrary weighting function which can be absorbed in the matrices.

Lemma 8.1.2

$$\varphi^+ = \log \det Q^+ \text{ is concave in } \theta$$

Proof

1. $\det Q^+ \geq \det[(1 - \theta)Q + \theta Q^*]$

Note that

$$Q^+ - (1 - \theta)Q - \theta Q^* = (1 - \theta)S_{12}S_{22}^- S_{21} + \theta S_{12}^* S_{22}^{*-} S_{21}^*$$

$$- [(1 - \theta)S_{12} + \theta S_{12}^*][(1 - \theta)S_{22} + \theta S_{22}^*]^-[(1-\theta)S_{21}+\theta S_{21}^*]$$

which is non-negative definite by lemma 8.1.1. Hence the result follows, since $\det(A - B) \geq 0$ implies $\det A \geq \det B$.

2. $\det[(1 - \theta)Q + \theta Q^*] \geq (1 - \theta)\det Q + \theta \det Q^*$

This relationship holds for $\theta = 1$. Now take $0 \leq \theta < 1$. Let $W = (1 - \theta)Q + \theta Q^*$. Observe that $\det W > 0$ for $0 \leq \theta < 1$.

$$\frac{\partial \log \det W}{\partial \theta} = \frac{1}{\det W} \sum_{i,j} \frac{\partial \det W}{\partial w_{ij}} \frac{\partial w_{ij}}{\partial \theta}$$

$$= \frac{1}{\det W} \sum_{i,j} W^{ij}(q^*_{ij} - q_{ij})$$

$$= \sum_{i,j} w^{ij}(q^*_{ij} - q_{ij})$$

where W^{ij} is the cofactor of w_{ij} and w^{ij} the corresponding element of W^{-1}.

$$\frac{\partial^2 \log \det W}{\partial \theta} = \sum_{i,j,u,v} \frac{\partial w^{ij}}{\partial w_{uv}} \frac{\partial w_{uv}}{\partial \theta} (q^*_{ij} - q_{ij})$$

$$= -\sum_{i,j,u,v} w^{iu} w^{jv}(q^*_{ij} - q_{ij})(q^*_{uv} - q_{uv})$$

$$= -\operatorname{trace}[W^{-1}(Q^* - Q)]^2 \leq 0$$

hence log det W and thus det W are concave.

3. Finally φ^+ is concave if and only if $\det Q^+$ is concave.

Lemma 8.1.3

$$\frac{\partial (S^+)^-}{\partial \theta} = -(S^+)^-(S^* - S)(S^+)^-$$

for $0 < \theta < 1$ or for $0 \leq \theta < 1$ if S^* is contained in the row space of S.

Proof

It is readily verified that, with $0 < \theta < 1$, both S and S* are contained in the rwo space of S^+. (The proof involves putting S^+ in the form V^TV and showing that $SV^-V = S$, and hence that $S(S^+)^-(S^+) = S$, with a similar result for S*). Thus

$$S^+(S^+)^-(S^* - S) = S^* - S = (S^* - S)(S^+)^-S^+ \qquad (8.2)$$

Now consider the identity $S^+(S^+)^-S^+ = S^+$.
Differentiating

$$\frac{\partial S^+}{\partial \theta} (S^+)^-S^+ + S^+ \frac{\partial (S^+)^-}{\partial \theta} S^+ + S^+(S^+)^- \frac{\partial S^+}{\partial \theta} = \frac{\partial S^+}{\partial \theta}$$

or

$$(S^* - S)(S^+)^-S^+ + S^+ \frac{\partial (S^+)^-}{\partial \vartheta} S^+ + S^+(S^+)^-(S^* - S) = S^* - S$$

from which

$$S^+ \frac{\partial (S^+)^-}{\partial \theta} S^+ = -(S^* - S)$$

From (8.2) and generalized inverse theory, this has a solution for the derivative. Without loss of generality the solution may be taken to be the expression in the statement of the lemma.

If S* is contained in the row space of S, (8.2), and hence the remainder of the proof, holds for $\theta = 0$.

Lemma 8.1.4

$$\frac{\partial \log \det Q^+}{\partial \theta} = \operatorname{trace}[(S^+)^-(S^*-S)] - \operatorname{trace}[(S_{22}^+)^-(S_{22}^*-S_{22})] \qquad (8.3)$$

Proof

Let $u = \frac{\partial \log \det Q^+}{\partial\theta} = \text{trace } (Q^+)^{-1} \frac{\partial Q^+}{\partial\theta}$

from the proof of lemma 8.1.2, where the derivative was evaluated for W.

Expanding

$$u = \text{trace}[(Q^+)^{-1} \frac{\partial}{\partial\theta} (S^+_{11} - S^+_{12}(S^+_{22})^- S^+_{21})]$$

$$= \text{trace}[(Q^+)^{-1}[S^*_{11}-S_{11}-(S^*_{12}-S_{12})(S^+_{22})^- S^+_{21}$$

$$+S^+_{12}(S^+_{22})^-(S^*_{22}-S_{22})(S^+_{22})^- S^+_{21} - S^+_{12}(S^+_{22})^-(S^*_{21}-S_{21})]] \qquad (8.4)$$

on differentiating and using lemma 8.1.3 for $\frac{\partial(S_{22})^-}{\partial\theta}$.

Note that the result holds at $\theta=0$ by continuity arguments, see Appendix III.

Now, from the theory of generalized inverses,

$$(S^+)^- = \begin{pmatrix} (Q^+)^{-1} & -(Q^+)^{-1}S^+_{12}(S^+_{22})^- \\ -(S^+_{22})^- S^+_{21}(Q^+)^{-1} & (S^+_{22})^- + (S^+_{22})^- S^+_{21}(Q^+)^{-1}S^+_{12}(S^+_{22})^- \end{pmatrix}$$

On using this to expand the expression on the right of (8.3), noting that the second term cancels, the right hand side of (8.3) is equal to

$$\text{trace}[(Q^+)^{-1}[S^*_{11}-S_{11}-S^+_{12}(S^+_{22})^-(S^*_{21}-S_{21})]]$$

$$+ \text{trace}[(S^+_{22})^- S^+_{21}(Q)^{-1}[-(S^*_{12}-S_{12})+S^+_{12}(S^+_{22})^-(S^*_{22}-S_{22})]]$$

$$= u$$

using the fact that matrices commute under the trace operator to place $(Q^{+})^{-1}$ at the beginning of the second term, giving this expression identical to (8.4). This establishes the result.

Lemma 8.1.5

$$\Phi = \frac{\partial \log \det Q}{\partial \theta}\bigg|_{\theta=0+} = \text{trace } S^{-}S^{*} - \text{trace } S_{22}^{-}S_{22}^{*} - \text{rank } Q$$

Proof

From lemma 8.1.4

$$\Phi = \text{trace } S^{-}S^{*} - \text{trace } S_{22}^{-}S_{22}^{*} - (\text{trace } S^{-}S - \text{trace } S_{22}^{-}S_{22})$$

The last term is equal to rank S - rank S_{22}, and this is equal to rank Q by inspection of the identity

$$\begin{pmatrix} I & -S_{12}S_{22}^{-} \\ 0 & I \end{pmatrix} S \begin{pmatrix} I & 0 \\ -S_{22}^{-}S_{21} & I \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & S_{22} \end{pmatrix}$$

Lemma 8.1.6

If D^{*} is any design, then it may be replaced by a design matrix consisting of a single row, such that Φ will be at least as great.

Proof

Noting that $S^{*} = \Sigma_i \lambda_{ii}^{*} x_i^{*} x_i^{*T}$ where the x_i^{*} are the rows of X^{*},

and similarly, $S_{22}^{*} = \Sigma_i \lambda_{ii}^{*} x_{i2}^{*} x_{i2}^{*T}$,

$$\Phi = \Sigma_i \lambda^*_{ii}(\underset{\sim}{x}^{*T}_i S^- \underset{\sim}{x}^*_i - \underset{\sim}{x}^{*T}_{i2} S^-_{22} \underset{\sim}{x}^*_{i2}) - \text{rank } Q \qquad (8.5)$$

If we select that value of i which maximizes the expression in brackets, and replace X* by that $\underset{\sim}{x}^{*T}_1$ alone, the resulting Φ is at least as great as in (8.5).

Lemma 8.1.7

$$\rho = \max_{\underset{\sim}{x}} (\underset{\sim}{x}^T S^- \underset{\sim}{x} - \underset{\sim}{x}^T_2 S^-_{22} \underset{\sim}{x}_2) - \text{rank } Q$$

Proof

This is an immediate consequence of lemma 8.1.6.

Lemma 8.1.8

If $\max_{\underset{\sim}{x}} (\underset{\sim}{x}^T F \underset{\sim}{x} - \underset{\sim}{x}^T_2 H \underset{\sim}{x}_2) = \text{trace } (FS - HS_{22})$

where F and H are any conformable matrices, then

$$\underset{\sim}{x}^T_i F \underset{\sim}{x}_i - \underset{\sim}{x}^T_{i2} H \underset{\sim}{x}_{i2} = \max_{\underset{\sim}{x}} (\underset{\sim}{x}^T F \underset{\sim}{x} - \underset{\sim}{x}^T_2 H \underset{\sim}{x}_2)$$

for i = 1,...,p

Proof

$$\text{trace } (FS - HS_{22}) = \sum_i \lambda_{ii} (\underset{\sim}{x}^T_i F \underset{\sim}{x}_i - \underset{\sim}{x}^T_{i2} H \underset{\sim}{x}_{i2})$$

and if this is equal to the maximum over $\underset{\sim}{x}$, then each of the $\underset{\sim}{x}^T_i F \underset{\sim}{x}_i - \underset{\sim}{x}^T_{i2} H \underset{\sim}{x}_{i2}$ must be equal to the maximum.

Note that if ρ from lemma 8.1.7 is zero, then the above result holds with $F = S^-$ and $H = S_{22}^-$, since rank Q = trace S^-S - trace $S_{22}^-S_{22}$.

8.2 EQUIVALENCE THEOREM

Theorem 8.1

Where a design for which Q is non-singular exists, the following criteria are equivalent

1. D maximizes det Q (8.6)
2. D minimizes $\max_{\underset{\sim}{x}}(\underset{\sim}{x}^T S^- \underset{\sim}{x} - \underset{\sim}{x}_2^T S_{22}^- \underset{\sim}{x}_2)$ (8.7)
3. $\max_{\underset{\sim}{x}}(\underset{\sim}{x}^T S^- \underset{\sim}{x} - \underset{\sim}{x}_2^T S_{22}^- \underset{\sim}{x}_2) = \text{rank } Q = s$ (8.8)
$= \underset{\sim}{x}_i^T S^- \underset{\sim}{x}_i - \underset{\sim}{x}_{i2}^T S_{22}^- \underset{\sim}{x}_{i2}$ for i=1,...,p

Proof

This result follows from Whittle's theorem and from the preceding lemmas.

This theorem was first proved, for the case in which s=k, with S non-singular, by Kiefer and Wolfowitz (1960), and extended to the case $s < k$, but with S still non-singular, by Kiefer (1962b), using a difficult games theory argument. The most appealing and simplest proof for that case appears to be that given by Federov (1972). As far as I am aware, the above result is the first time the theorem has been extended, in a reasonably simple form, to the case in which S and S_{22} may be singular. The form of d(x,D) given in (6.14), (8.7) and (8.8) is new in that the generalized inverses extend $d(\underset{\sim}{x},D)$ to the singular case.

9. APPLICATION OF φ-OPTIMALITY TO OTHER CRITERIA

9.1 GENERAL COMMENT

As mentioned in section 6.3, many optimality criteria are special cases of

minimize trace $S^{-}B$

and Whittle's theorem may be applied to this. The theorem can also be applied to R-optimality (maximize trace S).

By contrast, E-optimality (minimize the maximum eigenvalue of S) and the criterion that minimizes the largest diagonal element of S^{-} (or of $R^{T}S^{-}R$, see equation (6.7)), are intractable. In particular, for any reasonable choice of optimizing function, φ,

$$\frac{\partial^2 \varphi^+}{\partial \theta^2}$$

may not exist at all points since the choice of eigenvalue or diagonal element may change with θ.

9.2 R-OPTIMALITY

Let φ = trace S, and φ^+ = trace S^+. Now

$$\frac{\partial \varphi^+}{\partial \theta} = \sum_{u,v} \frac{\partial \text{trace} S^+}{\partial s^+_{uv}} \frac{\partial s^+_{uv}}{\partial \theta}$$

$$= \sum_{u,v} \delta_{uv} (s^*_{uv} - s_{uv})$$

$$= \sum_{u} (s^*_{uu} - s_{uu})$$

$$= \text{trace}(S^* - S)$$

where δ_{uv} = 1 if u=v, and otherwise zero.

Note that this derivative is independent of θ, and hence that the second derivative, with respect to θ, is zero. φ is thus concave.

Also, for the same reason, Φ from Whittle's theorem is given by

$$\Phi = \text{trace}(S^* - S)$$

Now, since $S^* = \sum_i \lambda^*_{ii} \underset{\sim}{x}^*_i \underset{\sim}{x}^{*T}_i$, there is some value of i such that

$$\text{trace } S^* \leq \text{trace } \underset{\sim}{x}^*_i \underset{\sim}{x}^{*T}_i = \underset{\sim}{x}^{*T}_i \underset{\sim}{x}^*_i$$

and thus, in turn, any X^* may be replaced by a single vector $\underset{\sim}{x}$, such that Φ is at least as large. Hence

$$\rho = \max_{S^*} \Phi = \max_{\underset{\sim}{x}} \underset{\sim}{x}^T \underset{\sim}{x} - \text{trace } S$$

Now suppose that $\rho=0$. In this case

$$\max_{\underset{\sim}{x}} \underset{\sim}{x}^T \underset{\sim}{x} = \text{trace } S$$

$$= \underset{\sim}{x}^T_i \underset{\sim}{x}_i \quad \text{for } i = 1,\ldots,\rho$$

from lemma 8.1.8.

This now implies that Λ is arbitrary, and the selection of the points is the only important factor. If, however, the $\underset{\sim}{x}$ for which $\underset{\sim}{x}^T \underset{\sim}{x}$ is a maximum is unique then S has rank one, with rather unfortunate results for estimation purposes.

However, for completness, Whittle's theorem gives:

The following criteria are equivalent, and each implies R-optimality:

1. D maximizes trace S
2. D minimizes $\max_{\underset{\sim}{x}} \underset{\sim}{x}^T \underset{\sim}{x}$ - trace S
3. $\max_{\underset{\sim}{x}} \underset{\sim}{x}^T \underset{\sim}{x} = \text{trace } S = \underset{\sim}{x}^T_i \underset{\sim}{x}_i$ for all $i = 1,\ldots,p$.

Note that these imply that an R-optimum design is not invariant of location changes.

In any case, the remarks above regarding the rank of S suggest strongly that R-optimality is an unsatisfactory criterion.

9.3 B-OPTIMALITY

At this point, we define the term B-optimality, in which the aim is to minimize the quantity

$$\text{trace } S^{-}B$$

where B is a known, non-negative definite matrix. As noted in section 6.3, A-optimality, c-optimality, Q-optimality and what might be called point optimality (in which it is desired to minimize $\underset{\sim}{x}^{T}S^{-}\underset{\sim}{x}$ for a given $\underset{\sim}{x}$) are special cases of this criterion.

Let $\varphi = -\text{trace } S^{-}B$, and we seek to maximize (algebraically) φ. Also, to improve legibility, let

$$G = (S^{+})^{-}$$

for this section. Now $\varphi^{+} = -\text{trace } GB$.

Note that, if W is dependent on θ, and B is not

$$\frac{\partial \text{ trace } WB}{\partial\theta} = \sum_{uv} \frac{\partial \text{ trace } WB}{\partial w_{uv}} \quad \frac{\partial w_{uv}}{\partial\theta}$$

$$= \sum_{ijuv} \frac{\partial w_{ij} b_{ji}}{\partial w_{uv}} \quad \frac{\partial w_{uv}}{\partial\theta}$$

$$= \sum_{ijuv} \frac{\partial w_{ij}}{\partial w_{uv}} \quad \frac{\partial w_{uv}}{\partial\theta} b_{ij}$$

$$= \sum_{ij} \frac{\partial w_{ij}}{\partial\theta} b_{ij}$$

$$= \text{trace } \frac{\partial W}{\partial\theta} B$$

Hence

$$\frac{\partial \varphi^+}{\partial \theta} = - \text{trace } \frac{\partial G}{\partial \theta} B$$

$$= \text{trace } G(S^* - S)GB \qquad (9.1)$$

from lemma 8.1.3, for $0 < \theta < 1$, and at $\theta=0$ by continuity.

$$\frac{\partial^2 \varphi^+}{\partial \theta^2} = \text{trace} \left[\frac{\partial G}{\partial \theta} (S^* - S)GB + G(S^* - S)\frac{\partial G}{\partial \theta} B \right]$$

$$= -2 \text{ trace } G(S^* - S)G(S^* - S)GB$$

≤ 0 provided B is non-negative definite, and hence φ is convex, as required by the theorem. Setting $\theta=0$ in (9.1) gives

$$\Phi = \text{trace } S^-(S^* - S)S^-B$$

$$= \text{trace } S^-S^*S^-B - \text{trace } S^-SS^-B$$

$$= \text{trace } S^-(\sum_i \lambda^*_{ii} \underset{\sim}{x}^*_i \underset{\sim}{x}^{*T}_i)S^-B - \text{trace } S^-SS^-B$$

and, as in previous sections, we may select the value of i which maximizes

$$\underset{\sim}{x}^{*T}_i S^- B S^- \underset{\sim}{x}^*_i$$

and replace X* by this $\underset{\sim}{x}_j$ alone, giving

$$\rho = \max_{\underset{\sim}{x}} \underset{\sim}{x}^T S^- B S^- \underset{\sim}{x} - \text{trace } S^-SS^-B$$

Now, in a manner similar to that noted in section 9.2, if $\rho=0$, from lemma 8.1.8,

$$\underset{\sim}{x}^T_i S^- B S^- \underset{\sim}{x}_i = \max_{\underset{\sim}{x}} \underset{\sim}{x}^T S^- B S^- \underset{\sim}{x} \quad \text{for } i = 1,\ldots,p$$

Whittle's theorem now gives:

The following criteria are equivalent, and each implies B-optimality:

1. D minimizes trace $S^{-}B$
2. D minimizes $\max_{\underset{\sim}{x}} \underset{\sim}{x}^{T}S^{-}BS^{-}\underset{\sim}{x}$ - trace $S^{-}SS^{-}B$
3. $\max_{\underset{\sim}{x}} \underset{\sim}{x}^{T}S^{-}BS^{-}\underset{\sim}{x}$ = trace $S^{-}SS^{-}B$
   = $\underset{\sim}{x}_{1}^{T}S^{-}BS^{-}\underset{\sim}{x}_{1}$ for i = 1,...,p

Note that, for a suitable selection of S^{-}, trace $S^{-}SS^{-}B$ = trace $S^{-}B$. This also holds if B is contained in the row space of S. As a practical matter, the various quadratic forms are not unique if $\underset{\sim}{x}$ and B are not contained in the row space of S. Thus it is reasonable to make the requirement that $\underset{\sim}{x}$ and B be contained in the row space of S.

10. D-OPTIMA IN SPECIAL CASES

10.1 SPECIAL N-TIC POLYNOMIAL

A form of the model, termed a special n-tic polynomial, was defined by Scheffé (1958). In this case, the elements of $\underset{\sim}{x}$ are (with the possible inclusion of a constant) products of the elements of $\underset{\sim}{\xi}$, with the restriction that no element of $\underset{\sim}{\xi}$ occurs at higher than the first degree, and the further requirement that all such products occur. Thus, if $\underset{\sim}{\xi}^T = (\xi_1 \xi_2 \xi_3)$,

$$\underset{\sim}{x}^T = (1 \ \xi_1 \ \xi_2 \ \xi_3 \ \xi_1\xi_2 \ \xi_1\xi_3 \ \xi_2\xi_3 \ \xi_1\xi_2\xi_3)$$

the initial 1 being optional.

10.2 ATWOOD'S THEOREM

Atwood (1969) proved that, for a special n-tic polynomial, and s=k, a D-optimal design can have no experimental points in the interior of any line segment in the space of $\underset{\sim}{\xi}$, on which all the variables but one are held constant.

The following theorem strengthens Atwood's result.

Theorem 10.2.1

In the case where $s \leq k$, and S is not necessarily of full rank, a D-optimal design can include no points in the interior of any line segment in the space of $\underset{\sim}{\xi}$, on which all the variables but one are constant, unless, on the whole of that line segment

$$\underset{\sim}{x}_1 - S_{12}S_{22}^{-}\underset{\sim}{x}_2$$

is independent of the element of $\underset{\sim}{\xi}$ that is not held constant.

Proof

From (6.14), $d(\underset{\sim}{x},D) = (\underset{\sim}{x}_1 - S_{12}S_{22}^{-}\underset{\sim}{x}_2)^T Q^{-1}(\underset{\sim}{x}_1 - S_{12}S_{22}^{-}\underset{\sim}{x}_2)$

Since Q is at least non-negative definite, $d(\underset{\sim}{x},D)$ is never negative. However, from the special n-tic form of $\underset{\sim}{x}$, $d(\underset{\sim}{x},D)$ is at most quadratic in each ξ_i, and holding all but one of these ξ_i constant makes $d(\underset{\sim}{x},D)$ a non-negative quadratic and hence convex. If $\underset{\sim}{x}_1 - S_{12}S_{22}^{-}\underset{\sim}{x}_2$ is not independent of the ξ_i which varies, $d(\underset{\sim}{x},D)$ is a strictly convex quadratic, and cannot have a maximum in the interior of any segment of the line.

Corollary If s=k, $\underset{\sim}{x}_1 - S_{12}S_{22}^{-}\underset{\sim}{x}_2$ becomes $\underset{\sim}{x}$ which cannot be independent of any ξ_i, and hence the exception condition does not hold. (This is Atwood's result).

The exception condition has caused considerable difficulty. The author has been unable to find an example in which the condition is met, and hence to show that such exceptional designs do exist. It also appears to be particularly difficult to characterize such designs.

11. CONSTRUCTION OF D-OPTIMAL DESIGNS

11.1 PREVIOUS WORK

Wynn (1970, 1972) has provided methods in the full rank case for obtaining D-optimum designs by adding to some starting design the point $\underset{\sim}{x}$ which maximizes

$$\underset{\sim}{x}^T S^{-1} \underset{\sim}{x} - \underset{\sim}{x}_2^T S_{22}^{-1} \underset{\sim}{x}_2$$

with a new Λ given by

$$\begin{pmatrix} \frac{p}{p+1}\Lambda & 0 \\ 0 & \frac{1}{p+1} \end{pmatrix}$$

Whittle (1973) improved the method of augmenting Λ by suggesting

$$\begin{pmatrix} (1-\alpha)\Lambda & 0 \\ 0 & \alpha \end{pmatrix}$$

where α is chosen to maximize the optimization criterion.

11.2 RECALCULATION OF Λ

The methods noted above while suitable for sequential design, suffer from the disadvantage of not deleting unsatisfactory points when the design is constructed in advance. The following sections give a refinement whose derivation began with an attempt to select, given X, that Λ which maximizes det Q.

11.3 EXPLICIT RESULTS

Lemma 11.3.1

If $S^* = \underset{\sim}{x}\underset{\sim}{x}^T$ and $\theta < 1$ then

$$\frac{\partial \log\det Q^+}{\partial \theta} = \frac{d(\underset{\sim}{x},D^+)-s}{1-\theta}$$

Proof First note that, since $S^+ = (1-\theta)S + \theta S^*$

$$S = \frac{S^+ - \theta S^*}{1-\theta} \qquad S^* - S = \frac{S^* - S^+}{1-\theta}$$

Hence, in the present case

$$\text{trace}(S^+)^-(S^* - S) = \frac{1}{1-\theta}\,\text{trace}(S^+)^-(\underset{\sim}{x}\underset{\sim}{x}^T - S^+)$$

$$= \frac{1}{1-\theta}\,(\underset{\sim}{x}^T(S^+)^-\underset{\sim}{x} - \text{rank } S^+)$$

with a similar result involving S_{22}^+.

From lemma 8.1.4

$$\frac{\partial \log\det Q^+}{\partial \theta} = \text{trace}(S^+)^-(S^* - S) - \text{trace}(S_{22}^+)^-(S_{22}^* - S_{22})$$

and the result follows on noting that

$$s = \text{rank } S^+ - \text{rank } S_{22}^+$$

Theorem 11.3.1

A design $D = (X,\Lambda)$ may be improved by the addition of a given point $\underset{\sim}{x}$ if and only if

$$d(\underset{\sim}{x},D) > s$$

Proof This follows from lemma 11.3.1 setting $\theta=0$.

Corollary The design may be improved by increasing λ_{ii} if and only if $d(\underset{\sim}{x}_i,D) > s$.

Theorem 11.3.2

A design $D = (X,\Lambda)$ may be improved by reducing the weighting of some point $\underset{\sim}{x}_i$ and rescaling the remainder, if and only if

$$d(\underset{\sim}{x}_i, D) < s$$

Proof This follows from lemma 11.3.1, with $\theta = \lambda_{ii}$, on writing S for S^+, with the ith diagonal element of Λ set to zero.

We now assume that $\underset{\sim}{x}$ is, in fact, one of the rows of X, and wish to adjust λ_{ii} by

$$(1-\theta)\lambda_{ii} + \theta$$

and the other diagonal elements of Λ by $(1-\theta)\lambda_{jj}$ with $j \neq i$. Note that this implies that $\underset{\sim}{x}$ is contained in the row space of S. Now θ is to be selected in such a way that

$$d(\underset{\sim}{x}, D^+) = s$$

From Pringle and Rayner (1971), theorem 2.18,

$$(S^+)^- = \frac{1}{1-\theta}\left[S^- - \frac{\theta}{1-\theta+\theta w} S^- \underset{\sim}{x}\underset{\sim}{x}^T S^-\right]$$

where $w = \underset{\sim}{x}^T S^- \underset{\sim}{x}$. Now

$$w^+ = \underset{\sim}{x}^T (S^+)^- \underset{\sim}{x} = \frac{1}{1-\theta}\left[w - \frac{\theta w^2}{1-\theta+\theta w}\right]$$

$$+ \frac{w}{1-\theta+\theta w}$$

with a similar expression (with $w_2 = \underset{\sim}{x}_2^T S_{22}^- \underset{\sim}{x}_2$) for w_2^+.

Thus, optimizing in terms of $\underset{\sim}{x}$ requires a solution of the equation

$$\frac{w}{1-\theta+\theta w} - \frac{w_2}{1-\theta+\theta w_2} = s \qquad (11.2)$$

in terms of θ. Thus, once $\underset{\sim}{x}$ has been chosen, θ may be calculated as the solution of a quadratic.

Viewed as a function of θ, det Q^+ has two turning points, and discontinuities at

$$\theta = 1, \quad \frac{1}{1-w_2}, \quad \frac{1}{1-w}$$

in descending order. Hence there will be one solution between 1 and $\frac{1}{1-w_2}$ of which the sign will depend on the sign of $w-w_2-s$, which in turn indicates whether the weighting of a point should be increased or decreased. If a point has its weighting decreased below zero, the weighting should be set to zero, and the point deleted from the design.

11.4 IMPROVEMENT TO DET Q

If the intention is to improve det Q by replacing Q with Q^+, assuming for the moment that S is of full rank, then

$$\det Q^+ = \frac{\det S^+}{\det S_{22}^+}$$

$$= \frac{\det[(1-\theta)S + \theta \underset{\sim}{x}\underset{\sim}{x}^T]}{\det[(1-\theta)S_{22} + \theta \underset{\sim}{x}_2 \underset{\sim}{x}_2^T]}$$

$$= \frac{\det[(1-\theta)S](1 + \frac{\theta}{1-\theta} w)}{\det[(1-\theta)S_{22}](1 + \frac{\theta}{1-\theta} w_2)}$$

$$= (1-\theta)^s \frac{\det S}{\det S_{22}} \frac{1-\theta+\theta w}{1-\theta+\theta w_2}$$

$$= (1-\theta)^s \det Q \frac{1-\theta+\theta w}{1+\theta+\theta w_2}$$

and

$$\frac{\det Q^+}{\det Q} = (1-\theta)^s \frac{1-\theta+\theta w}{1-\theta+\theta w_2} \qquad (11.1)$$

The same result is obtained if the restriction of a full rank S matrix is relaxed. The proof involves a rather lengthy expansion of Q^+, but is otherwise straightforward.

If $\underset{\sim}{x}$ is already a point, $\underset{\sim}{x}_i$, in the design, then θ may be negative, if θ is sufficiently negative that λ_{ii} becomes negative, the point is simply deleted from the design, and the other λ values rescaled.

In this case

$$S^+ = \frac{1}{1-\lambda_{ii}} (S - \lambda_{ii} \underset{\sim}{x}_i \underset{\sim}{x}_i^T)$$

and by a process similar to that above,

$$\frac{\det Q^+}{\det Q} = (1-\lambda_{ii})^{-s} \frac{1-\lambda_{ii} w_i}{1-\lambda_{ii} w_{i2}} \qquad (11.3)$$

In theory, (11.1) and (11.3) give the means to select the best available point for alteration. In practice, it may be simpler to use some arbitrary, but reasonable criterion for selection, since the decision making process derived from the above equations may involve more calculation than the adjustment process. In the examples quoted later, that $\underset{\sim}{x}_i$ for which the absolute value of

$$d(\underset{\sim}{x}_i, D) - s$$

was greatest was chosen, with good results. As noted below, the method of selecting the maximum $d(x_i, D)$ leads to very slow convergence.

11.5 ALL PARAMETERS OF INTEREST

If all parameters are of interest, that is, if s=k, the above expressions simplify considerably.

In particular (11.2) becomes

$$\frac{w}{1-\theta+\theta w} = k$$

or $$\theta = \frac{w-k}{(w-1)k}$$

and

$$\frac{\det Q^+}{\det Q} = \left(\frac{k-1}{w-1}\right)^k \left(\frac{w}{k}\right)^{k+1} \qquad (11.4)$$

Note that, if $\underset{\sim}{x} = \underset{\sim}{x}_i$, and

$$k - w > \lambda_{ii} w(k-1)$$

then an attempt will be made to reduce λ_{ii} below zero. In this situation

$$\frac{\det Q^+}{\det Q} = \frac{1-\lambda_{ii} w_i}{(1-\lambda_{ii})^k}$$

Note that (11.4) is an increasing function of w, hence any w_i between the largest w_i and k, need not be considered as a candidate.

12. CONSTRUCTION OF B-OPTIMAL DESIGNS

12.1 IMPROVEMENT OF Λ

By direct analogy to the previous chapter, a design may be improved in terms of B-optimality, as follows.

From (9.1), using $G = (S^+)^-$

$$\frac{\partial \text{ trace } GB}{\partial \theta} = \text{trace } G(\underset{\sim}{x}\underset{\sim}{x}^T - S)GB$$

and this is to be equated to zero and solved for θ. Once again it is assumed that $\underset{\sim}{x}$ is linearly dependent on S. This gives

$$G = \frac{1}{1-\theta}\left(S^- - \frac{\theta}{1-\theta+\theta w} S^- \underset{\sim}{x}\underset{\sim}{x}^T S^-\right)$$

where $w = \underset{\sim}{x}^T S^- \underset{\sim}{x}$

Straightforward multiplication, noting that $SS^-\underset{\sim}{x} = \underset{\sim}{x}$, gives

$$G(\underset{\sim}{x}\underset{\sim}{x}^T - S)G = \frac{1}{(1-\theta)^2} S^- \left[\frac{1-\theta^2+\theta^2 w}{(1-\theta+\theta w)^2} \underset{\sim}{x}\underset{\sim}{x}^T - S\right] S^-$$

and on postmultiplying by B, taking traces and equating to zero

$$\text{trace } S^-SS^-B = \frac{1-\theta^2+\theta^2 w}{(1-\theta+\theta w)^2} \underset{\sim}{x}^T S^- BS^- \underset{\sim}{x}$$

Note that the left hand side is equal to trace S^-B for S^- suitably chosen, or for B linearly dependent on S.

13. APPLICATIONS

13.1 BACKGROUND

A computer program was written to test convergence of the D-optimality criterion under the conditions of the last section. Since the computer (a PDP11/45) was not equipped with double precision arithmetic, numerical instability tended to occur. For this reason, full rank models with all parameters of interest, and with small values of k were chosen.

In each case four points were chosen at random from the region of operability, and the iteration scheme used to improve Λ. When this had converged, a new point was chosen at random, and if, for this point, $\underset{\sim}{x}^T S^{-1} \underset{\sim}{x} > k$, it was included in the design, and the iteration continued, otherwise new random points were selected until the condition was met. This was repeated until 16 extra points had been selected.

In addition, as described in section 13.4, arising from a local problem, an aisle quadratic was analysed, thus providing a practical example of the application of the iterative scheme.

Finally, although it was not used in the examples noted above, a generalized inverse routine, as described in Appendix I, was developed. This was tested and found quite satisfactory.

13.2 LINEAR MODEL

In this test example,

$$\underset{\sim}{x}^T = (1 \ \xi_1 \ \xi_2) \text{ with } 0 \leq \xi_1, \xi_2 \leq 1$$

The optimum is given by weightings of .25 each at the points $(\xi_1,\xi_2) = (1,1),(1,-1),(-1,1),(-1,-1)$, for which design det S = 1. The design was considered to have converged when $\Sigma(\underset{\sim}{x}^T S^{-1} \underset{\sim}{x} - 3)^2/p \leq .25E\text{-}8$.

Ten runs were carried out, and the final values of det S, together with the number of iterations required are given. By contrast, the strategy of selecting max $\underset{\sim}{x}^T S^{-1} \underset{\sim}{x}$ rather than max abs$(\underset{\sim}{x}^T S^{-1} \underset{\sim}{x} - 3)$ required, for run 1, 333 iterations to converge for the first four points, and had not converged for the fifth 2000 iterations later.

Figure 1 gives the order in which points were eliminated in run 1, illustrating the manner in which points were selected closer and closer to the periphery.

Lastly, the final design is given in table 2, for run 1 (the best), run 8 (the worst) and run 9 (the quickest convergence).

A straightforward calculation reveals that if, in the optimal design, the point $(\xi_1,\xi_2) = (1,1)$ is replaced by (θ,θ), the derivative of det S with respect to θ, at $\theta = 1$ is 1, and hence, as the point moves inwards from the corner, det S drops at the same rate. If all four points have 1 replaced by θ in the same way, then det S = θ^4, which falls off rapidly with decreasing θ. This explains what might be thought to be low values of det S.

TABLE 1

Results for linear model.

Run	Final det S	Iterations required for first 4 points	Iterations required for 20 points
1	.8439	25	1338
2	.6711	2(11)	620
3	.5666	2(17)	861
4	.6435	2(5)	1163
5	.7610	13	759
6	.7216	2(39)	599
7	.8385	2(12)	971
8	.4495	26	706
9	.7141	2(11)	362
10	.7402	12	664

Note: In the number of iterations required column, the bracketed figures give the iterations required for the fifth point, where one of the first four was unsuitable and immediately eliminated.

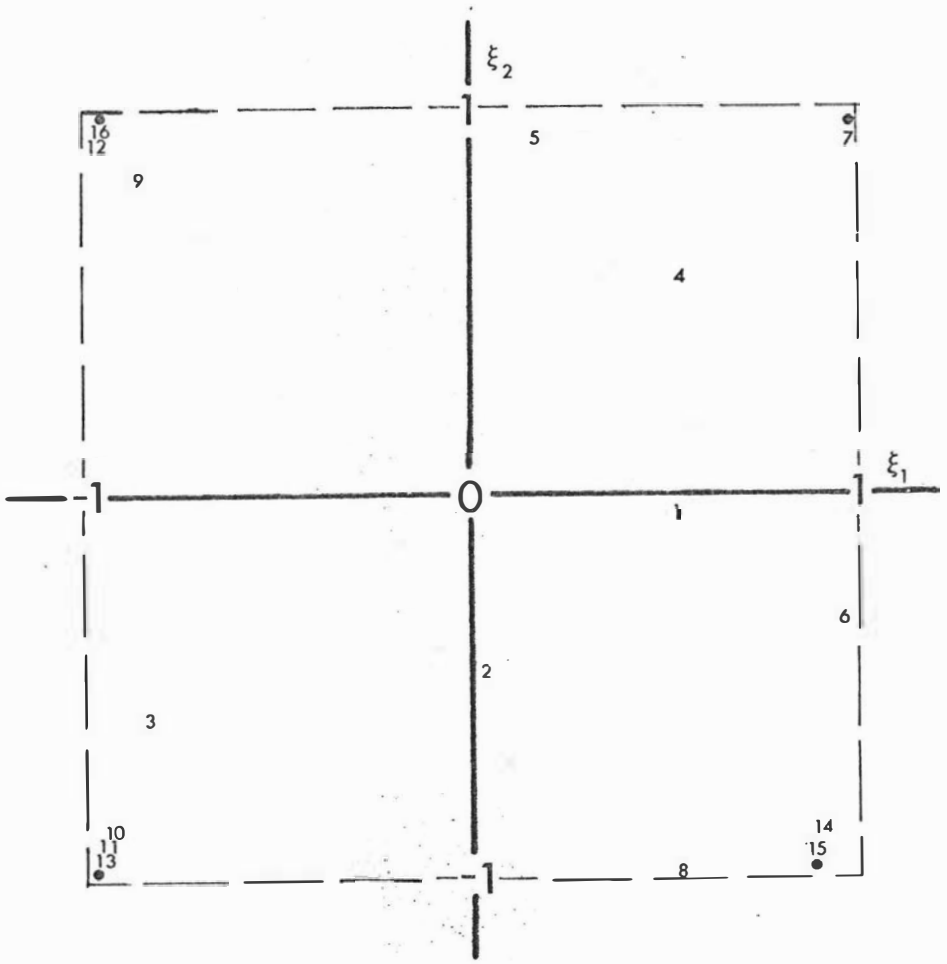

Fig 1. Order of elimination of points in linear model, run 1. Dots are points left in the final design.

TABLE 2

Final points and lambda values for three runs in the linear model.

Run	λ	ξ_1	ξ_2
1	.2563	.998	.932
(best	.2358	.873	-.945
det S = .8439)	.2634	-.992	.961
	.2460	-.975	-.986
8	.2695	.973	.856
(worst	.0810	.021	.999
det S = .4495)	.3066	.875	-.939
	.2441	-.997	-.054
	.0988	-.753	-.823
9	.2538	.853	.997
(quickest	.2677	.859	-.986
convergence,	.2294	-.863	.849
det S = .7141)	.2491	-.971	-.974

13.3 POLYNOMIAL MODEL

In this example, the model was

$$\underset{\sim}{x}^T = (1 \; \xi_1 \; \xi_1^2 \; \xi_1^3)$$

The theoretical optimum has four points, with $\xi_1 = -1, -1/\sqrt{5}, 1/\sqrt{5}, 1$, and each $\lambda = .25$, with det S = 16/3125 = 5.12×10^{-3}.

Two refinements were added to the procedure for this model. If p=k, then X is square, hence det S = $(\det X)^2 \pi \lambda_{ii}$, which is maximized in terms of Λ by $\lambda_{ii} = 1/p$ for i = 1,...,p. Thus whenever the number of points in the design dropped to k, Λ was immediately set to this value and a new point selected. The other refinement arose from the observation that most iterations arose from the difficulty in separating, and rejecting one of, two very close points. However, such points do not affect det S greatly, and another point may be added without loss of efficiency. Hence, if convergence had not been reached 100 iterations after the last new point, a further point was added. The unsatisfactory point of the close pair was then usually rejected in the next series of iterations.

It is interesting to consider the behaviour of det S under departures from the optimum design. Let the design be $\xi_1 = -\alpha, -\beta, \beta, \alpha$, with equal weightings.
Now

$$\det S = \frac{1}{16} \alpha^2\beta^2(\alpha^2-\beta^2)^4$$

Differentiating with respect to α gives

$$\frac{1}{8} \alpha\beta^2(\alpha^2-\beta^2)^3(5\alpha^2-\beta^2)$$

With $\alpha=1$, $\beta=1/\sqrt{5}$, this gives 192/3125 or 12 det S, for the derivative, indicating great sensitivity at the extremes. On the other hand, with $\alpha = 1/\sqrt{5}$, $\beta = 1$, the value is zero, substantiating

the observed insensitivity to the exact intermediate values. In fact, with $\alpha = 1$, $\beta = .5$, det S is .9656 of the optimum.

In view of the above, the value of det S achieved depended almost entirely on how close the randomly selected end points came to to plus or minus one. However, table 3 illustrates two representive runs. The second was much the worst seen, for most, det S ranged from 4.1×10^{-3} to 4.5×10^{-3}.

13.4 AISLE QUADRATIC

The method described in sections 13.2 and 13.3 was intended more as an illustrative device than as a serious approach to design optimization. The situation described in this section arose from a practical problem with a constraint imposed by the physical environment. The aim of the experiment was to estimate quadratic growth contours in a glasshouse. Thus with two physical dimensions, the model becomes

$$\underset{\sim}{x}^T = (1 \; \xi_1 \; \xi_2 \; \xi_1^2 \; \xi_2^2 \; \xi_1\xi_2)$$

with $-1 \leq \xi_1, \xi_2 \leq 1$. However, the normal layout of a glasshouse imposed another restriction - the aisle down the centre required abs $\xi_2 \geq \alpha$ where α is a constant depending on the particular glasshouse.

In approaching this problem, the twelve points illustrated in figure 2 were selected as the starting points for a design, and the iterative procedure used to refine Λ. In the particular problem considered, $\alpha = 1/3$.

The final design thus generated is given in table 4. Note the assymmetry.

Using the symmetry of the situation, 3 similar optimum designs may be produced. Averaging these 4 optimum designs leads to a symmetric optimum, given in the last column of table 4.

TABLE 3

Final points and λ values for two runs in the polynomial model.

det S	λ	ξ_1
4.95×10^{-3}	.2494	.9946
(best)	.2502	.4435
	.2494	-.4384
	.0008	-.4564
	.2501	-.9999
2.16×10^{-3}	.25	.9279
(worst)	.25	.4117
	.25	-.4381
	.25	-.9335

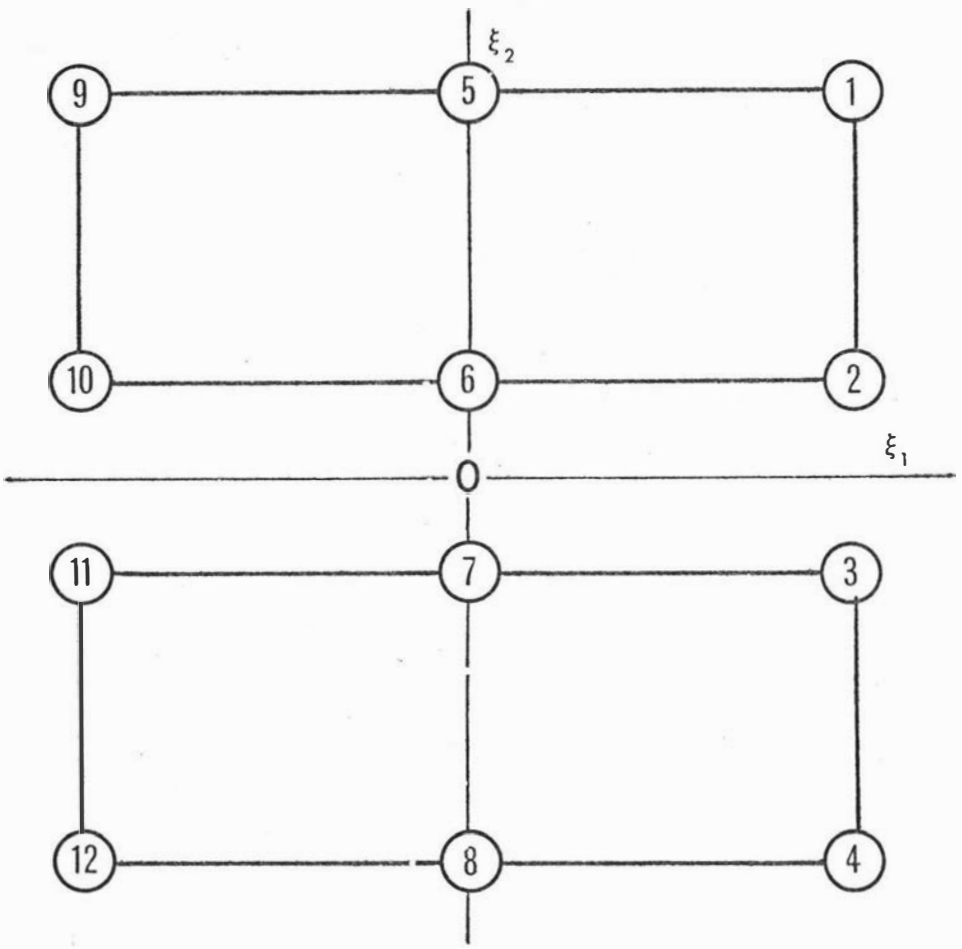

Fig 2. Layout of selected points for aisle quadratic. The heavy rectangles represent the region of operability.

TABLE 4

Optimum aisle quadratic design with points from figure 2 as starting values.

Point	ξ_1	ξ_2	λ from computer run	Symmetric λ
1	1	1	.1463	.1420
2	1	.3333	.0314	.0442
3	1	-.3333	.0570	.0442
4	1	-1	.1377	.1420
5	0	1	.0699	.0785
6	0	.3333	.0748	.0492
7	0	-.3333	.0235	.0492
8	0	-1	.0871	.0785
9	-1	1	.1463	.1420
10	-1	.3333	.0314	.0442
11	-1	-.3333	.0570	.0442
12	-1	-1	.1377	.1420

Regardless of α, for such a symmetric design, S has the form

$$S = \begin{pmatrix} s_{11} & 0 & 0 & s_{14} & s_{15} & 0 \\ 0 & s_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & s_{33} & 0 & 0 & 0 \\ s_{14} & 0 & 0 & s_{44} & s_{45} & 0 \\ s_{15} & 0 & 0 & s_{45} & s_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{66} \end{pmatrix} \qquad (13.1)$$

and hence S^{-1} has the same form. Using superscripts to denote the elements of S^{-1}, and noting that $\underset{\sim}{x}_i^T S^{-1} \underset{\sim}{x}_i = 6$ for the points in the design, we have

$$\begin{aligned} s^{11} &= 6 + \alpha^2 s^{55} \\ s^{22} + 2s^{14} &= -s^{44} \\ s^{33} + 2s^{15} &= -(1+\alpha^2)s^{55} \\ s^{66} + 2s^{45} &= 0 \end{aligned}$$

and using these relationships

$$\underset{\sim}{x}^T S^{-1} \underset{\sim}{x} - 6 = s^{44}\xi_1^2(\xi_1^2 - 1) + s^{55}(\xi_2^2 - 1)(\xi_2^2 - \alpha^2) \qquad (13.2)$$

It is now easy to see that any point (ξ_1, ξ_2) in the allowable area, other than those in the design, must reduce the value of (13.2) below zero. Hence the points considered produce a true optimum.

The equations to be solved to give Λ for a given α proved quite intractable. However, a polynomial approximation for the optimum value of s_{33} was obtained, and from this the other values may be derived as indicated in Appendix II.

If interest is concentrated on the constant and linear terms only, a slightly different problem arises.

In this case, and using the same 12 points, again, by considerations of symmetry, S has the form given in (13.1). In this case, we are interested in

$$d(\underset{\sim}{x},D) = \underset{\sim}{x}^T \left[S^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & S_{22}^{-1} \end{pmatrix} \right] \underset{\sim}{x}$$

in which W, the matrix in square brackets, has the same form as S, except that the bottom right element is zero.

Again, calculating $d(\underset{\sim}{x}_i,D) = 3$ for the points in the design leads to

$$w_{11} = 3 + \alpha^2 w_{55}$$

$$w_{22} + 2w_{14} = -w_{44}$$

$$w_{33} + 2w_{15} = -(1+\alpha^2)w_{55}$$

$$w_{45} = 0$$

and

$$d(\underset{\sim}{x},D) - 3 = w_{44}\,\xi_1^2(\xi_1^2-1) + w_{55}(\xi_2^2-1)(\xi_2^2-\alpha^2)$$

with the same conclusions as previously.

14. SUMMARY AND CONCLUSIONS

As noted in the introduction the main aim of this thesis has been to rephrase the existing optimum design theory in terms more familiar to a statistician. For this reason Chapter 2 gives the background to the matrix formulation of the general linear model.

Chapter 3 then discusses the various questions the experimenter might pose, then shows how, if the interest is in testing a linear hypothesis, the general hypothesis is equivalent, under a canonical transformation, to a simpler hypothesis. Because of the difficulty in interpreting this canonical form, an intermediate form, termed the simplified model has been developed. For completeness, the chapter concludes with a proof (from Federov) that the number of distinct points required is bounded.

In Chapters 4 and 5, a general development of hypothesis testing optimality theory, based largely on Kiefer (1958), has been presented, and, in particular, the development of D-optimality in this context. Chapter 6 considers various optimality criteria applied to parameter estimation.

Chapter 7 merely reiterates Whittle's theorem with slight modifications to matrix notation. Chapter 8 presents a proof for the equivalence theorem for D-optimality in its most general form, and similarly Chapter 9 derives equivalence theorems for R-optimality, and a general class of criteria, termed in this thesis, B-optimality.

Chapter 10 extends one of the theorems of Atwood (1969), indicating how a matrix notation is adaptable to the type of special case he considered.

Chapters 11 and 12 give methods of design refinement which can be used in particular cases again emphasizing the notation chosen.

Chapter 13 gives some examples of these methods in practice.

In all the above topics, Chapter 8 presents the most difficult mathematics. It should be noted, however, that this arises from the generality of the results. The proof simplifies accordingly for more restricted cases. Chapters 8 to 12 are those which illustrate the utility of the notation normally used in linear model theory when applied to the subject of design optimization.

15. REFERENCES

ATWOOD, C.L. (1969).
"Optimal and efficient designs of experiments."
Ann. Math. Statist., 40, 1570 - 1603

BOX, M.J. & DRAPER, N.R. (1971).
"Factorial designs, the $|X^TX|$ criterion and some related matters."
Technometrics, 13, 731 - 742

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

ELFVING, G. (1952).
"Optimum allocation in regression theory."
Ann. Math. Statist., 23, 255 - 262

FEDEROV, V.V. (1972).
Theory of Optimal Experiments
Academic Press, N.Y. & London

HSU, P.L. (1941).
"Analysis of variance from the power function standpoint."
Biometrika 32, 62 - 69

ISAACSON, S. (1951).
"On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters."
Ann. Math. Statist., 22, 217 - 234

KARLIN, S. & STUDDEN, W.J. (1966).
"Optimum experimental designs."
Ann. Math. Statist., 37, 783 - 815

KENDALL, M.G. & STUART, A. (1967).
The Advanced Theory of Statistics.
3 Vols., Griffin, London.

KIEFER, J. (1958).
"On the nonrandomized optimality and randomized nonoptimality of symmetric designs."
Ann. Math. Statist., 29, 675 - 699

KIEFER, J. (1959).
"Optimum experimental designs."
J.R. Statist. Soc. B, 21, 272 - 319

KIEFER, J. (1961a).
"Optimum designs in regression problems, II."
Ann. Math. Statist., 32, 298 - 335

KIEFER, J. (1961b).
"Optimum experimental designs V, with applications to systematic and rotatable designs."
Proc. Fourth Berkely Symp. Math. Statist. Prob., 1, 381 - 405

KIEFER, J. (1962a).
"An extremum result."
Canad. J. Math., 14, 597 - 601

KIEFER, J. (1962b).
"Two more criteria equivalent to D-optimality of designs."
Ann. Math. Statist., 33, 792 - 796

KIEFER, J. & WOLFOWITZ, J. (1959).
"Optimum designs in regression problems."
Ann. Math. Statist., 30, 271 - 294

KIEFER, J. & WOLFOWITZ, J. (1960).
"The equivalence of two extremum problems."
Canad. J. Math., 12, 363 - 366

KOŁODZIECZYK, S. (1935).
"On an important class of statistical hypotheses."
Biometrika 27, 161 - 190

LAYCOCK, P.J. (1972).
"Convex loss applied to design in regression problems."
J.R. Statist. Soc. B, 34, 148 - 186

LEHMANN, E.L. (1950).
"Some principles of the theory of testing hypotheses."
Ann. Math. Statist., 21, 1 - 26

LEHMANN, E.L. (1959).
Testing Statistical Hypotheses.
Wiley, New York

LINDLEY, D.V. (1956).
"On a measure of the information provided by an experiment."
Ann. Math. Statist., 27, 986 - 1005

MURTY, V.N. (1971).
"Minimax designs."
J. Am. Statist. Assoc., 66, 319 - 320

NEYMAN, J. & PEARSON, E.S. (1936).
"Contributions to the theory of testing statistical hypotheses, Part I."
Statist. Res. Memoirs, 1, 1 - 37

NEYMAN, J. & PEARSON, E.S. (1938).
"Contributions to the theory of testing statistical hypotheses, Parts II and III."
Statist. Res. Memoirs, 2, 25 - 57

PRINGLE, R.M. & RAYNER, A.A. (1971).
Generalized Inverse Matrices with Applications to Statistics.
Griffin, London

RAO, C.R. & MITRA, S.J. (1971).
Generalized Inverse of Matrices and its Applications
Wiley, New York

SCHEFFÉ, H. (1958).
"Experiments with mixtures."
J.R. Statist. Soc., B, 20, 344 - 366

STONE, M. (1959).
"Application of a measure of information to the design and comparison of regression experiments."
Ann. Math. Statist., 30, 55 - 70.

WALD, A. (1942).
"On the power function of the analysis of variance test."
Ann. Math. Statist., 13, 434 - 439

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

WHITTLE, P. (1973).
"Some general points in the theory of optimal experimental design."
J.R. Statist. Soc. B, 35, 123 - 130

WOLFOWITZ, J. (1949).
"The power of the classical tests associated with the normal distribution."
Ann. Math. Statist., 20, 540 - 551

WYNN, H.P. (1970).
"The sequential generation of D-optimum experimental designs."
Ann. Math. Statist., 41, 1655 - 1664

WYNN, H.P. (1972).
"Results in the theory and construction of D-optimum experimental designs."
J.R. Statist. Soc. B, 34, 133 - 147, 170 - 186

APPENDIX I

GENERALIZED INVERSE CALCULATION

In calculating generalized inverses in the application of the results of section 11, a modified Cholesky algorithm was used.

Theorem

If X is partitioned as $(X_1\ X_2)$ with X_1 of full rank and $X_2 = X_1K$, then

$$S^- = \begin{pmatrix} S_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

with partitioning conformable to that of X, is a generalized inverse of S with the additional properties that S^- is symmetric and $S^-SS^- = S^-$.

Proof

$$S = \begin{pmatrix} S_{11} & S_{11}K \\ K^TS_{11} & K^TS_{11}K \end{pmatrix}$$

Evaluation of SS^-S and S^-SS^- then establishes the result.

Note that the partition above is conditioned only by the rank of X, and does not (necessarily) correspond to the partition used elsewhere in this thesis.

Evidently the columns of X may be permuted, and the result holds with similar permutations on the rows and columns of S and S^-.

The practical result of the above theorem is that, in the present situation, a generalized inverse of S may be obtained by the

ordinary Cholesky decomposition, with the additional rule that, whenever a zero diagonal element is encountered, the whole of the corresponding row and column are set to zero, and thereafter treated as any other row.

As a practical consideration, if using a computer, a decision has to be made as to whether a particular diagonal element is zero or not. Thus a lower threshold needs to be set for this value.

APPENDIX II

D-OPTIMUM AISLE QUADRATIC

Note that the first coefficient in the expansion of s_{23} is exact, to 7 places, derived from an analytic solution for $\alpha=0$. The other coefficients provide a polynomial approximation (the best obtainable on the available computer)

$$s_{33} = s_{15} = .7434853 + .0015168\alpha + .077491\alpha^2 + .11364\alpha^3 - .17851\alpha^4 + .19538\alpha^5 + .048651\alpha^6$$

$$s_{22} = s_{44} = s_{14} = \frac{s_{33}^2 - \alpha^2}{(1+\alpha^2)s_{33} - 2\alpha^2}$$

$$s_{55} = (1+\alpha^2)s_{33} - \alpha^2$$

$$s_{66} = s_{45} = \frac{1}{8s_{33}} \left\{ 13s_{33}^2 - 5(1+\alpha^2)s_{33} - 3\alpha^2 + 8\alpha^2 s_{22} \right\}$$

$$\lambda_1 = \lambda_4 = \lambda_9 = \lambda_{12} = s_{22}/4$$

$$\lambda_2 = \lambda_3 = \lambda_{10} = \lambda_{11} = (s_{66} - s_{22})/4\alpha^2$$

$$\lambda_5 = \lambda_8 = \frac{1}{2}\left(\frac{s_{33} - \alpha^2}{1 - \alpha^2} - s_{22} \right)$$

$$\lambda_6 = \lambda_7 = \frac{1}{2}\left(\frac{1-s_{33}}{1-\alpha^2} - \frac{s_{66}-s_{22}}{\alpha^2} \right)$$

With the numbering given in figure 2.

Given s_{33}, all but the first of the expressions are exact.

APPENDIX III

CONTINUITY THEOREM

The following theorem is somewhat outside the general theme of this thesis, and is included merely for completeness.

Theorem If $\underset{\sim}{a}$ is contained in the column space of A then

$$[(1-\theta)\underset{\sim}{a} + \theta\underset{\sim}{b}]^T[(1-\theta)A + \theta B]^-[(1-\theta)\underset{\sim}{a} + \theta\underset{\sim}{b}]$$

is continuous in the interval $0 \leq \theta < 1$ where A and B are non-negative definite.

Proof

There exists a non-singular transformation U, such that

$$UAU^T = D \text{ and } UBU^T = E$$

where D and E are diagonal, and the diagonal elements of D are zero or one. (See, for example, Rao and Mitra (1971), theorem 6.2.3).

Thus $U[(1-\theta)A + \theta B]U^T = (1-\theta)D + \theta E$ and from the properties of generalized inverses (Pringle and Rayner (1971), theorem 1.3),

$$[(1-\theta)A + \theta B]^- = U^T[(1-\theta)D + \theta E]^-U$$

$$A^- = U^TD^-U$$

noting that $DD^- = D$.

Now the diagonal elements of $[(1-\theta)D + \theta E]^-$ are of the form (with d_{ii}, e_{ii} non-negative)

$$\frac{1}{(1-\theta)d_{ii}+\theta e_{ii}} \quad \text{or } 0 \quad (\text{if } d_{ii}=e_{ii}=0)$$

and hence $\theta^r[(1-\theta)D + \theta E]^-$, for $r \geq 1$, is continuous in the range $0 \leq \theta < 1$. Thus the terms in $\theta \underset{\sim}{b}$ are continuous.

It is sufficient, therefore to consider the term (neglecting the $(1-\theta)^2$ factor)

$$\underset{\sim}{a}^T[(1-\theta)A + \theta B]^- \underset{\sim}{a}$$

Note that, since $\underset{\sim}{a}$ is contained in the column space of A,

$$\underset{\sim}{a} = AA^-\underset{\sim}{a} = U^{-1}DU^{-T}U^TD^-U\underset{\sim}{a}$$

$$= U^{-1}DU\underset{\sim}{a}$$

Hence

$$\underset{\sim}{a}^T[(1-\theta)A + \theta B]^-\underset{\sim}{a} = \underset{\sim}{a}^TU^T[(1-\theta)D + \theta E]^-U(U^{-1}DU\underset{\sim}{a})$$

$$= \underset{\sim}{a}^TU^T[(1-\theta)D + \theta E]^-DU\underset{\sim}{a}$$

Now the typical element of $[(1-\theta)D + \theta E]^-D$ is

$$\frac{d_{ii}}{(1-\theta)d_{ii}+\theta e_{ii}}$$

which is zero if d_{ii} is zero and otherwise

$$\frac{1}{1-\theta + \theta e_{ii}}$$

with the denominator positive in the range $0 \leq \theta < 1$, and thus is continuous.

This establishes the result.

Evidently the theorem extends to the case $\theta=1$ if $\underset{\sim}{b}$ is linearly dependent on B.

The extension of

$$\frac{\partial \log \det Q}{\partial \theta}$$

to the case $\theta=0$ in lemma 8.1.4, as well as similar results elsewhere, depends on this theorem.