

Studies in experimental design and data analysis with special
reference to computational problems.

by

B. A. Worthington.

A Thesis submitted to the University of Kent at Canterbury for
the degree of Doctor of Philosophy.

November 1975

Index

	Page
Chapter 1. Introduction	1.
Chapter 2. The Analysis of a General Block Design	5
2.1 Introduction	5.
2.2 Notation and Model	9.
2.3 Prerequisite Theory on Singular Designs	10.
2.4 Generalized Inverses	12.
2.5 Calculation of the Eigensystem of a Design and the Generalized Inverse A^-	15.
2.6 The Inter-block Analysis	23.
2.7 Partitioning a Sum of Squares according to Contrasts of Interest	26.
2.8 Derivation of Basic Contrasts in Certain Special Cases	33.
2.9 A Program for the Analysis of a General Block Design	34.
2.10 Discussion	
Chapter 3. The Dual of a Block Design	
3.1 Introduction	37.
3.2 The Eigensystem of the Dual Design	37.
3.3 The Analysis of the Dual Design	42.
3.4 Relationships Between the Mean Efficiencies of a Design and Those of its Dual	48.
3.5 Discussion	52.
Chapter 4. The Construction of Optimal Block Designs when Contrasts of Interest have been given	
4.1 Introduction	53.
4.2 A Criterion based on D-optimality	54.
4.3 Iterative Method of Constructing Generalized D-optimal Designs	56.
4.4 Excursions	59.

4.5	Fixed Block Size	64.
4.6	Over-specification in the Model for Block Designs	65.
4.7	A Computer Program	66.
4.8	A Criterion Based on A-optimality	67.
4.9	Construction of A-optimal Designs when Contrasts and Weights have been given	67.
4.10	Excursions for A-optimal Designs	70.
4.11	Inherent Singularity in the Model	71.
4.12	A Criterion based on the Simple or Arithmetic Mean of the Efficiency Factors	74.
4.13	Construction of Simple Mean Optimal Designs with Contrasts	76.
4.14	Discussion	80.
Chapter 5.	Summary	82.
	References	86.
	Acknowledgements	90.

Abstract

An experimenter, designing or analysing an experiment, frequently finds that not all can be done according to the text-book. Considerations outside his control may have to be taken into account that affect the design; he may not be interested in all his treatments equally but perhaps in some unusual contrasts between them. Again, unforeseen circumstances and happenings can afterwards force him to discard some of his results, so upsetting the balance or orthogonality of the design.

The aim of this thesis is to help an experimenter in such a situation. The first part is concerned with analysing a block experiment that is in general unbalanced and non-orthogonal. Two different methods, one iterative, one non-iterative, are derived for obtaining the analysis, each with its own advantages. The non-iterative method basically is derived from the actual design and produces matrices, which can then operate on any suitable data supplied. The iterative method, however, found in appendix A, is applied directly to the data from the start, to produce the treatment effects directly. Although the iterative method is easier to apply and can also be used with a wider class of design than can the non-iterative method, the inter-block analysis and the analysis of the dual become easier using the non-iterative method.

Certain contrasts are related to the design in special ways, and, if known, make the analysis of the design easier. The implications are discussed in chapter 2, which is also concerned with finding the contribution to the sum of squares for these and other, more general, contrasts of interest.

The dual of a design is defined as that design formed from the original design when treatments and blocks classifications are interchanged, i.e. treatments become blocks and vice-versa in the new design. It is useful for studying block differences eliminating those

due to treatments, which may, for example, be required if the blocking system arose from the possible residual effects of treatments from some previous experiment on the same material. The second part of the thesis, in chapter 3, is concerned with the analysis of the dual. It is shown that there is no need to start again from the beginning when analysing the dual if the original design had already been analysed, because the analysis can provide information about that of the dual. The method is especially easy when the non-iterative method of analysing block designs, discussed in chapter 2, has been used for the original design.

The experimenter will often be more interested in some contrasts between treatments than in others and a design can be selected to give more precise information about these contrasts. The construction of such designs is discussed in the third part of the thesis. Various measures can be used to judge which design is best as regards contrasts of interest. Algorithms for finding the optimal design according to these measures are derived and discussed in chapter 4.

Listings and flowcharts of a program to carry out the non-iterative analysis of chapter 2 and of a program to construct optimal block designs appear in appendices B and C.

Chapter 1 Introduction

Three problems are considered in this thesis:

- 1) The analysis of general block designs,
 - 2) The analysis of the dual of a design
- and 3) The construction of optimal block designs.

The first problem was thought to be worthy of consideration because current methods all lack certain requirements that a general method of analysis should have. There are three main methods in use:

a) The method of Tocher (1952)

This method is to calculate directly an inverse of the information matrix of the treatments. Although the information matrix is always singular; if the design is connected, the only completely confounded contrast being that of the mean, then a simple addition to the matrix will cause it to become non-singular and an inverse can be found. This inverse, termed Ω , is then a generalized inverse (Rao and Mitra, 1971) of the original information matrix and so yields correct results.

The disadvantage of this method arises when some disconnectedness is present, it then being no longer obvious as to what to add to the information matrix so as to cause it to become non-singular - the addition depending upon the contrasts confounded. Also, for large experiments the size of the matrix to be inverted, equal to the number of treatments, can cause excessive calculation.

b) The method of Kuiper (1952)

This is an iterative method, calculating the treatment vector by a sequence of vectors tending to the zero vector, the sum of the sequence giving the correct estimate of the treatment vector. This method will also work when some confounding is present in the design. If a high degree of partial confounding exists however, the procedure takes a long time to converge. The dispersion matrix for treatment effects can be calculated if the design is connected by the extension due to Calinski

(1971); else if some knowledge of the eigensystem of the design is known, the extension due to Pearce, Calinski and Marshall (1974) can be used for disconnected designs.

c) The method of Wilkinson (1970)

This method, which can be applied to any design, uses a series of 'sweeps' across the data; each sweep, corresponding to either blocks or treatments, calculates simple means for the factor, divides these by an appropriate efficiency factor and subtracts them from the current input vector. The next sweep is then applied to the resulting vector and so on until the vector of residuals appears. The efficiency factors, if not known, can be found, but not their multiplicities, by a suitable dummy analysis. However, in complex experiments this gives rise to calculating the roots of a polynomial of degree equal to the order of balance of the design. A disadvantage of Wilkinson's method lies in the difficulties involved when calculating variance information (James and Wilkinson, 1971).

It was therefore decided to find a method which completely analysed all designs fairly efficiently and would take full advantage of any inherent simplicity that may exist in the design. The method should also be well suited to programming onto a computer.

Two solutions were found: one detailed in chapter 2, the other in appendix A. The first solution, a non-iterative method, is more useful in situations where a number of experiments, each having the same design, are carried out; whereas the other, an iterative method, is more easily suited to hand computation if necessary and has the advantage of being applicable to designs other than block experiments, row and column designs for example. The first method has a further advantage if the dual is to be used, since it provides the analysis for the dual at the same time as that for the original design.

The second problem, the analysis of the dual, can arise if treatments from a previous experiment are used as a blocking system for the current experiment. The experimenter may then wish to dualise

his design and study the effects of the previous treatments eliminating those of the current set (see, for example, Pearce, 1970). This can easily result in an unbalanced design even if the original design was balanced and the experimenter may then be faced with having to analyse a possibly difficult dual design in addition to the original design. It was decided to find whether the analysis for the original design could be used for the dual design also.

The theory behind the dual of a design is derived in chapter 3 and it is shown that the analysis for the dual can indeed be found almost directly from the method of analysis presented in chapter 2. Also, if the experimenter has chosen a design that is good with respect to some measure of efficiency, he may wish to know how good the dual design is with respect to that measure also. Relations between measures for the design and its dual, for various common measures, are therefore derived in this chapter.

The third problem arises when an experimenter, wishing to carry out some experiments, finds, due to considerations of space or available plots or whatever, that no standard design exists that he can use. He may attempt to invent a suitable design himself, but would find it difficult to know whether he had found the best design possible. Procedures do exist for helping the experimenter choose the 'optimal' design (e.g. Mitchell, 1974), where 'optimal' depends upon the criterion of optimality chosen, but they tend not to take into account the fact that the experimenter will often be interested in certain contrasts between treatments more than, or even to the exclusion of, other contrasts. It was decided then, that an algorithm be found which provided the experimenter with the design that 'best' met his needs as regards his interests concerning contrasts between the treatments; 'best', of course, depending on the criterion of optimality chosen.

Algorithms have therefore been found for each of three possible

criteria and the relative advantages of each given. These are detailed in chapter 4.

The main aim of the thesis is to help those experimenters finding themselves with a design to construct and/or analyse which does not fit into the realm of the standard designs.

Chapter 2 The Analysis of a General Block Design

2.1 Introduction

If observations \underline{y} are known to have expected values $\underline{X}\underline{\theta}$, where \underline{X} is known and $\underline{\theta}$ are unknown parameters, and it is also known that the error involved with each observation has the same variance σ^2 and is uncorrelated with the other observations, then the best linear unbiased estimate of $\underline{\theta}$, $\hat{\underline{\theta}}$ say, is obtained by the principle of least squares (Gauss, 1821; collected works 1873) as

$$\hat{\underline{\theta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}, \quad \text{and } D(\hat{\underline{\theta}}) = (\underline{X}'\underline{X})^{-1}\sigma^2$$

if \underline{X} is a matrix of full rank.

If the observations are not uncorrelated, but have a dispersion matrix of $\underline{V}\sigma^2$, where \underline{V} is known, then the estimate of $\hat{\underline{\theta}}$ becomes

$$\hat{\underline{\theta}} = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{y}, \quad \text{and } D(\hat{\underline{\theta}}) = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\sigma^2 \text{ (Aitken, 1934).}$$

Suppose now that \underline{X} is not of full rank but that $\underline{V} = \underline{I}$. Then an unbiased estimate $\hat{\underline{\theta}}$ of $\underline{\theta}$ cannot now be found. Plackett (1960) imposes a set of linearly independent conditions on $\underline{\theta}$, i.e. $\underline{B}\underline{\theta} = \underline{0}$ where \underline{B} is such that $\text{rank}(\underline{X}';\underline{B}')$ equals the number of unknown parameters. Then an unbiased estimate $\hat{\underline{\theta}}$ of $\underline{\theta}$, subject to $\underline{B}\hat{\underline{\theta}} = \underline{0}$, is given by

$$\hat{\underline{\theta}} = (\underline{X}'\underline{X} + \underline{B}'\underline{B})^{-1}\underline{X}'\underline{y},$$

with

$$D(\hat{\underline{\theta}}) = (\underline{X}'\underline{X} + \underline{B}'\underline{B})^{-1}\underline{X}'\underline{X}(\underline{X}'\underline{X} + \underline{B}'\underline{B})^{-1}\sigma^2.$$

The restrictions actually form the set of non-estimable functions of $\underline{\theta}$.

An alternative method (Graybill, 1961 p292) is to augment $\underline{X}'\underline{X}$, i.e. use

$$\begin{bmatrix} \underline{X}'\underline{X} & \underline{B}' \\ \underline{B} & \underline{0} \end{bmatrix} = \underline{A}^{-1} \text{ say,}$$

and invert this. Writing the inverted matrix \underline{A} as

$$\begin{bmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{bmatrix},$$

then \underline{A}_{11} is used for the solution, viz.

$$\hat{\underline{\theta}} = \underline{A}_{11}^{-1} \underline{X}' \underline{y}, \quad \text{and} \quad D(\hat{\underline{\theta}}) = \underline{A}_{11}^{-1} \sigma^2.$$

Both these methods are using generalized inverses (Rao and Mitra, 1971) to give a consistent solution to the normal equations subject to the restrictions $\underline{B}'\hat{\underline{\theta}} = \underline{0}$. The generalized inverses are not equal (John, 1964) and in fact an infinite set of generalized inverses exists giving a consistent solution. The generalized inverse may be considered as the dispersion matrix of $\hat{\underline{\theta}}$ if estimable contrasts only are required (John, 1964).

Now suppose that estimates of some of the parameters only are required, the remainder being unwanted or 'nuisance' variables. Without loss of generality let the parameters be partitioned as

$$\underline{\theta} = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix}$$

where $\underline{\theta}_1$ is the vector of the required parameters and $\underline{\theta}_2$ are the nuisance variables. It will be seen that the partitions $(\underline{X}'\underline{X})^{11}$ and $(\underline{X}'\underline{X})^{12}$ from

$$(\underline{X}'\underline{X})^{-1} = \begin{bmatrix} (\underline{X}'\underline{X})^{11} & (\underline{X}'\underline{X})^{12} \\ (\underline{X}'\underline{X})^{21} & (\underline{X}'\underline{X})^{22} \end{bmatrix}$$

are all that is necessary, i.e. the whole inversion is not required. Schur's identity can therefore be used and this is essentially what Tocher (1952) does, first adding $\underline{B}'\underline{B}$ to $\underline{X}'\underline{X}$ because of the inherent singularity of the simple block experiment model. Tocher assumes that this is the only cause of the singularity, none arising from any confounding or disconnection in the design, and so gives a complete algebraic solution for all such designs. When other causes of singularity are present however, the restrictions \underline{B} to be placed upon $\underline{\theta}$ are in general unknown and depend on precisely how the singularity of the design arises.

Rather than calculate the required inverse directly, iterative methods may be used. The method presented by Kuiper (1952) essentially

does this, actually calculating $\hat{\beta}_1$ as a sequence of vectors which add up to the required answer. The Kuiper procedure converges even when the design is confounded, since it effectively calculates the generalized inverse $(\underline{X}'\underline{X})^{-1}$ automatically using the constraints supplied by the eigenvectors of the original matrix which correspond to the zero eigenvalues (Worthington, 1975). Since the speed of convergence is dependent upon the value of the least non-zero eigenvalue (Worthington, 1975), if this is small, the procedure can be slow. The treatment sum of squares is easily obtained.

Variance information can be obtained when no confounding exists by Calinski's extension (1971) to Kuiper's method. This involves analysing suitable dummy variates so as to produce a row of the dispersion matrix as the estimated parameters.

If some knowledge of the eigenvectors corresponding to the zero efficiency factors (Jones, 1959) of a design is available then Pearce, Calinski and Marshall (1975) have extended Kuiper's method further to produce the dispersion matrix for singular designs also.

A procedure which will always converge in a fixed number of steps is that given by Wilkinson (1970). This uses 'sweeps' across the data, each sweep, corresponding to either blocks or treatments, calculates a set of means for the factor, divides them by an appropriate efficiency factor and subtracts them from the current input vector. The next sweep is then applied to the resulting vector and so on until the vector of residuals appears. These efficiency factors are not in general known beforehand and must be calculated using a dummy analysis of pseudo-variates. This in general however, requires the solution of a polynomial equation of degree equal to the order of balance of the experiment to be found. As in Kuiper's method, this procedure will work for a confounded design.

A disadvantage of Wilkinson's method is the difficulty associated with calculating variance information (James and Wilkinson, 1971).

Finally, all present methods require a complete re-analysis for computing the inter-block analysis of a design.

No method exists which

- a) can be applied to any design, whether confounded or unbalanced or whatever,
- b) gives the dispersion matrix for estimable contrasts
- and c) computes the inter-block analysis without re-starting the procedure.

An attempt has therefore been made to provide a reasonably efficient method which complies with all the above conditions. With this method, it is necessary to use the efficiency factors (Jones, 1959) of the design; therefore, due to the absence of a simple method of finding them, a standard routine will have to be used. A matrix inversion is also necessary, but the size of the matrix, equal to the order of balance, is in general small.

The method, given the efficiency factors of a design, produces enough further information about the eigensystem to allow an analysis to be made. An alternative method, derived from Pease (1965), is compared.

Now, certain contrasts between treatments may be of interest to the experimenter; for example, the difference between the control treatment and the other treatments, or, whether a quadratic or higher order response exists between different levels or amounts of some treatment.

Providing these contrasts are independent of each other (in effect, that the covariance between them is zero), the experimenter will then require his treatment sum of squares to be partitioned, with each contrast having a sum of squares associated with it, so as to

be able to apply an F-test to each. If these contrasts satisfy certain conditions then they may lead to a simplification of the analysis.

2.2 Notation and Model

The model used is

$$\underline{y} = \alpha \underline{1} + \underline{D}'\underline{\beta} + \underline{\Delta}'\underline{\gamma} + \underline{\eta}$$

where \underline{y} is a vector of yields of length N , N being the number of plots; α is a general parameter; \underline{D}' is an $(N \times b)$ design matrix for blocks, b being the number of blocks; $\underline{\beta}$ is the vector of block parameters of length b ; $\underline{\Delta}'$ is an $(N \times v)$ design matrix for treatments, v being the number of treatments; $\underline{\gamma}$ is the vector of treatment parameters of length v ; and $\underline{\eta}$ is a vector of residuals of length N . The dispersion matrix of the residuals is assumed to be $\underline{I}\sigma^2$.

The incidence matrix, \underline{n} , of treatments to blocks is then equal to $\underline{\Delta D}'$; the vector of block sizes, \underline{k} , equals $\underline{D}\underline{1}$ and the vector of treatment replications, \underline{r} , equals $\underline{\Delta}\underline{1}$.

The matrix $\underline{a}^{b\delta}$ denotes the vector \underline{a} , with elements raised to the power b , written in diagonal form; its inverse is $\underline{a}^{-b\delta}$. Then

$$\underline{\Delta\Delta}' = \underline{r}^{\delta} \quad \text{and} \quad \underline{DD}' = \underline{k}^{\delta}.$$

Also $\underline{\Delta}'\underline{1} = \underline{D}'\underline{1} = \underline{1}$.

Minimising $\underline{\eta}'\underline{\eta}$ leads to the Normal equations for treatment effects, being

$$\underline{A}\hat{\underline{\gamma}} = \underline{Q},$$

where $\underline{A} = \underline{r}^{\delta} - \underline{nk}^{-\delta}\underline{n}'$ (2.2.1)

and $\underline{Q} = \underline{T} - \underline{nk}^{-\delta}\underline{B}$,

$\underline{T} = \underline{\Delta y}$, being the vector of treatment totals and $\underline{B} = \underline{Dy}$, the vector of block totals. The matrix \underline{A} is the same as the matrix \underline{C} used by Chakrabarti (1962).

The normal equations are singular, since $\underline{A}\underline{1} = \underline{0}$. Tocher (1952)

therefore assumed the constraints $\underline{r}'\hat{\underline{y}} = 0$, in which case the Normal equations can be written

$$(\underline{A} + \underline{r}\underline{r}'/N)\hat{\underline{y}} = \underline{Q} = \underline{\Omega}^{-1}\hat{\underline{y}} \quad \text{say.}$$

For connected designs $\underline{\Omega}^{-1}$ is not singular and a solution is immediately available as $\hat{\underline{y}} = \underline{\Omega}\underline{Q}$.

However, in general, more than one constraint is needed; these constraints cannot be arbitrary but depend upon the matrix, \underline{A} , and the problem is that of finding these constraints.

2.3 Prerequisite Theory on Singular Designs

It is necessary to prove a preliminary result involving the eigensystem of a design.

Definition

The eigensystem of a design is the system of eigenvectors and eigenvalues of $\underline{A}\underline{r}^{-\delta}$.

Theorem 2.3

The matrix $\underline{A}\underline{r}^{-\delta}$ always has a complete set of linearly independent eigenvectors.

Proof

Let $\underline{P} = \underline{r}^{-\frac{1}{2}\delta} (\underline{A}\underline{r}^{-\delta}) \underline{r}^{\frac{1}{2}\delta}$; then \underline{P} is symmetric and has, therefore, a complete set of linearly independent eigenvectors.

Let \underline{X} be the matrix having the eigenvectors of \underline{P} as its columns.

Then $\underline{P}\underline{X} = \underline{X}\underline{\lambda}$ where $\underline{\lambda}$ is the vector of eigenvalues of \underline{P} .

Since the eigenvectors are linearly independent, \underline{X} is non-singular; therefore

$$\underline{X}^{-1}\underline{P}\underline{X} = \underline{\lambda} \quad ,$$

i.e.

$$\underline{X}^{-1} \underline{r}^{-\frac{1}{2}\delta} (\underline{A}\underline{r}^{-\delta}) \underline{r}^{\frac{1}{2}\delta} \underline{X} = \underline{\lambda} \quad ,$$

or

$$(\underline{r}^{\frac{1}{2}\delta} \underline{X})^{-1} \underline{A}\underline{r}^{-\delta} (\underline{r}^{\frac{1}{2}\delta} \underline{X}) = \underline{\lambda} \quad ,$$

which implies that the columns of $\underline{r}^{\frac{1}{2}\delta} \underline{X}$ are the eigenvectors of $\underline{A}\underline{r}^{-\delta}$.

But, $(\underline{r}^{-\frac{1}{2}} \underline{X})$ is non-singular, therefore the eigenvectors are linearly independent. This completes the proof.

If the matrix \underline{C} has the eigenvectors of $\underline{Ar}^{-\delta}$ as its columns, and is scaled such that

$$\underline{C}' \underline{r}^{-\delta} \underline{C} = \underline{I}, \quad (2.3.1)$$

then, since

$$(\underline{Ar}^{-\delta}) \underline{C} = \underline{C} \underline{\epsilon}$$

(where $\underline{\epsilon}$ is the vector of eigenvalues), $\underline{Ar}^{-\delta}$ can be written as $\underline{C} \underline{\epsilon} \underline{C}^{-1}$, or, using (2.3.1), $\underline{C} \underline{\epsilon} \underline{C}' \underline{r}^{-\delta}$.

Therefore

$$\underline{A} = \underline{C} \underline{\epsilon} \underline{C}' = \sum_i \underline{c}_i \underline{c}_i' \epsilon_i, \quad (2.3.2)$$

if $\underline{C} = \{\underline{c}_i\}$, and $\underline{\epsilon} = \{\epsilon_i\}$.

Note that, from (2.2.1),

$$\underline{Ar}^{-\delta} = (\underline{I} - \underline{M}'),$$

where $\underline{M} = \underline{r}^{-\delta} \underline{nk}^{-\delta} \underline{n}'$

and is the matrix used by Jones (1959). Now, the eigenvalues of \underline{M} , or equivalently, \underline{M}' , are equal to one minus the efficiency factors of the design (Jones, 1959). But the eigenvalues of \underline{M}' are equal to one minus those of $\underline{Ar}^{-\delta}$, therefore, the ϵ_i 's correspond to the efficiency factors of the design.

It will be seen that one vector, \underline{r}/\sqrt{N} , \underline{c}_1 say, is always an eigenvector of $\underline{Ar}^{-\delta}$, with efficiency factor (eigenvalue) zero.

A contrast between the treatment parameters is defined as $\underline{c}' \underline{y}$, where $\underline{c}' \underline{1} = \underline{0}$. Since all eigenvectors must satisfy (2.3.1), and $\underline{c}_1 = \underline{r}/\sqrt{N}$, then this gives the following restriction on the other eigenvectors: that $\underline{c}_i' \underline{r}^{-\delta} \underline{c}_i = \underline{0}$ for $i \neq 1$ or, substituting for \underline{c}_1 , $\underline{1}' \underline{c}_i = \underline{0}$ for $i \neq 1$. The eigenvectors then, can be considered as contrasts. Due to their special nature as eigenvectors of $\underline{Ar}^{-\delta}$, they will be termed 'basic contrasts'. These are then identical with one form of the basic

contrasts used by Pearce, Calinski and Marshall (1974). The eigenvalues, then, correspond to the degree of confounding of the basic contrasts. Note that $\underline{c}_1 = \underline{r}/\sqrt{N}$, which corresponds to the mean, is always confounded, i.e. ϵ_1 always equals zero. It will be seen that when no other confounding exists, that is when no efficiencies other than ϵ_1 equal zero, then the addition of $\underline{c}_1 \underline{c}'_1 = \underline{r} \underline{r}'/N$ to expression (2.3.2) will cause $(\underline{A} + \underline{r} \underline{r}'/N)$ to be non-singular, because of the linear independence of the eigenvectors. The matrix $(\underline{A} + \underline{r} \underline{r}'/N)$ is, of course, the matrix $\underline{\Omega}^{-1}$ used by Tocher (1952) for analysing unconfounded designs.

2.4 Generalized Inverses

A generalized inverse \underline{G} of a singular matrix \underline{A} is a matrix having similar properties to that of a normal inverse of a non-singular matrix. The relation $\underline{AG} = \underline{I}$ cannot exist for singular \underline{A} , but a unique matrix can be found satisfying the relations:

$$\underline{AGA} = \underline{A} \quad (2.4.1)$$

$$\underline{GAG} = \underline{G} \quad (2.4.2)$$

$$(\underline{GA})' = \underline{GA} \quad (2.4.3)$$

$$(\underline{AG})' = \underline{AG} . \quad (2.4.4)$$

The original concept of a generalized inverse is due to Moore (1920), although he defined it in terms of projection operators over a complex field. Penrose (1955), unaware of Moore's work also developed a generalized inverse, defined as satisfying the conditions (2.4.1) - (2.4.4) above. These two definitions are equivalent if a suitable choice is made of the projection operators mentioned above when applied to vectors (Rao and Mitra, 1971). The generalized inverse defined as above is generally known as the Moore - Penrose inverse. For common applications, not all conditions need be satisfied, as will be seen later.

Suppose now the model

$$E(\underline{y}) = \underline{X}\underline{\theta}$$

is used, with errors ϵ having expected value zero and dispersion matrix $\underline{I}\sigma^2$. The matrix \underline{X} is assumed in general to have rank less than the number of parameters. Goldman and Zelen (1964) show that the minimum variance unbiased linear estimate of an estimable linear function of $\underline{\theta}$ can be found using

$$\hat{\underline{\theta}} = (\underline{X}^-)\underline{y},$$

where \underline{X}^- is the weak generalized inverse of \underline{X} , satisfying only the conditions (2.4.1) - (2.4.3).

An easier way of finding $\hat{\underline{\theta}}$ is to derive the (over specified) Normal equations

$$(\underline{X}'\underline{X})\hat{\underline{\theta}} = \underline{X}'\underline{y},$$

and use a generalized inverse of $(\underline{X}'\underline{X})$ to solve them. Rao and Mitra (1971) show that only condition (2.4.1) need be satisfied by the generalized inverse so as to produce a consistent solution

$$\hat{\underline{\theta}} = \{(\underline{X}'\underline{X})^- \underline{X}'\} \underline{y}.$$

In fact, $(\underline{X}'\underline{X})^- \underline{X}'$ is one choice for \underline{X}^- above (Rao and Mitra, 1971 p140).

Returning to the model for the block design then, with Normal equations

$$\underline{A}\hat{\underline{y}} = \underline{Q},$$

it is required to find a generalized inverse of \underline{A} satisfying only (2.4.1). In the unconfounded case, Tocher's $\underline{\Omega}$ (1952) is one possible solution.

Now, since \underline{A} can be written in the form (2.3.2), then, using (2.3.1) enables a generalized inverse of \underline{A} , satisfying (2.4.1), to be written

$$\underline{A}^- = \underline{r} \underline{C} \underline{\epsilon}_*^{-\delta} \underline{C}' \underline{r}^{-\delta},$$

where

$$\{\underline{\epsilon}_*\}_i = \begin{cases} \epsilon_i, & \text{if } \epsilon_i \neq 0 \\ r_i, & \text{if } \epsilon_i = 0 \end{cases},$$

the η_i 's being any non-zero numbers.

A consistent solution for the Normal equations is then

$$\hat{\underline{y}} = \underline{A}^{-1} \underline{Q} = \underline{r}^{-\delta} \underline{C} \underline{\epsilon}_*^{-\delta} \underline{C}' \underline{r}^{-\delta} \underline{Q} = \underline{r}^{-\delta} \sum_i \left(\frac{c_i c_i'}{\epsilon_{*i}} \right) \underline{r}^{-\delta} \underline{Q}, \quad (2.4.5)$$

and the best linear unbiased estimate of an estimable function $\underline{t}' \underline{Y}$ is $\underline{t}' \hat{\underline{y}}$. The dispersion matrix, for use in calculating variance information of estimable contrasts can be considered as $\underline{A}^{-1} \sigma^2$.

Not only are all estimable functions of $\hat{\underline{y}}$ unique (e.g. Rao and Mitra, 1971 p140), but the following theorem also applies:

Theorem 2.4

The vector $\hat{\underline{y}}$ is unique, that is, it is independent of the choice of the η_i used in the definition of $\underline{\epsilon}_*$.

Proof

The theorem will be proved by showing that $\underline{c}_i' \underline{r}^{-\delta} \underline{Q}$ is zero if \underline{c}_i corresponds to a zero eigenvalue, since then the η_i will have no effect upon the value of $\hat{\underline{y}}$.

Now,

$$\begin{aligned} \underline{Q} &= \underline{T} - \underline{nk}^{-\delta} \underline{B} = (\underline{A} - \underline{nk}^{-\delta} \underline{D}) \underline{Y} \\ &= \underline{A} (\underline{I} - \underline{D}' \underline{k}^{-\delta} \underline{D}) \underline{Y} \\ &= \underline{A} \underline{\phi} \underline{Y} \quad \text{say.} \end{aligned}$$

Note that $\underline{\phi}$ is idempotent.

Consider the expression $(\underline{c}_i' \underline{r}^{-\delta} \underline{A} \underline{\phi}) (\underline{c}_i' \underline{r}^{-\delta} \underline{A} \underline{\phi})'$, which can be written as

$$\underline{c}_i' \underline{r}^{-\delta} (\underline{A} \underline{\phi} \underline{A}') \underline{r}^{-\delta} \underline{c}_i. \quad (2.4.6)$$

But $(\underline{A} \underline{\phi} \underline{A}')$ equals $\underline{r}^{\delta} - \underline{nk}^{-\delta} \underline{n}' = \underline{A} = \sum_i (\underline{c}_i \underline{c}_i' \underline{\epsilon}_i)$.

Therefore (2.4.6) can be written as

$$\underline{c}_i' \underline{r}^{-\delta} \sum_j (\underline{c}_j \underline{c}_j' \underline{\epsilon}_j) \underline{r}^{-\delta} \underline{c}_i.$$

Using (2.3.1), this reduces to $\underline{\epsilon}_i$. Now, if \underline{c}_i corresponded to a zero eigenvalue, this implies that the sum of the elements of $(\underline{c}_i' \underline{r}^{-\delta} \underline{A} \underline{\phi})$ squared is zero. Since the elements are real, this shows that $(\underline{c}_i' \underline{r}^{-\delta} \underline{A} \underline{\phi} \underline{y})$ or

$\underline{c}_i' \underline{r}^{-\delta} \underline{Q}$ is equal to the zero vector whenever ϵ_i is zero. This concludes the theorem.

The theorem could also have been proved by showing that the constraints applied to the parameters by using \underline{A}^- are independent of the τ_i , and that these constraints, \underline{H} say, where $\underline{H}\hat{\underline{y}} = \underline{0}$, are such that $\text{rank}(\underline{A}:\underline{H}')$ equals the number of parameters, v . The solution to the Normal equations, subject to $\underline{H}\hat{\underline{y}} = \underline{0}$, is then unique (see, for example, Plackett, 1960 p42).

Corollary

The constraints imposed upon $\hat{\underline{y}}$, due to the using of \underline{A}^- , are $\underline{c}_i' \hat{\underline{y}} = \underline{0}$ for all \underline{c}_i such that $\epsilon_i = 0$.

Proof

From (2.4.5),

$$\underline{c}_i' \hat{\underline{y}} = \frac{\underline{c}_i' \underline{r}^{-\delta} \underline{Q}}{\epsilon_{*i}} = \frac{\underline{c}_i' \underline{r}^{-\delta} \Delta \underline{\phi} \underline{y}}{\epsilon_{*i}},$$

but, by the theorem, $\underline{c}_i' \underline{r}^{-\delta} \Delta \underline{\phi} = \underline{0}$ if $\epsilon_i = 0$, therefore $\underline{c}_i' \hat{\underline{y}} = 0$ for all values of \underline{y} .

These constraints are equivalent to those imposed by Pearce, Calinski and Marshall (1974), who made the point that they correspond to linear combinations of the treatment parameters that cannot be estimated from the intra-block analysis, that is, they correspond to the confounded contrasts.

2.5 Calculation of the Eigensystem of a Design and the Generalized

Inverse \underline{A}^-

As in section 4,

$$\underline{A} = \sum_i (\underline{c}_i \underline{c}_i' \epsilon_i)$$

and
$$\underline{A}^- = \underline{r}^{-\delta} \sum_i \left(\frac{\underline{c}_i \underline{c}_i'}{\epsilon_{*i}} \right) \underline{r}^{\delta}$$

where

$$\epsilon_{*i} = \begin{cases} \epsilon_i & , \text{ if } \epsilon_i \neq 0 \\ \eta_i \neq 0 & , \text{ if } \epsilon_i = 0 \end{cases} .$$

Since, from the previous section, the choice of the η_i 's is immaterial, they will be chosen so as to minimise the computation involved.

It will be seen that letting the $\eta_i \rightarrow \infty$ is equivalent to ignoring the contribution from those eigenvectors having zero eigenvalues when forming \underline{A}^- . In future, this choice of \underline{A}^- will always be used; it can be written as

$$\underline{A}^- = \underline{r}^{-\delta} \sum_i \left(\frac{\underline{c}_i \underline{c}'_i}{\epsilon_i} \right) \underline{r}^{\delta} , \quad (2.5.0)$$

where the summation covers only those values of i for which $\epsilon_i \neq 0$. This is a different procedure from that used by Pearce, Calinski and Marshall (1975), who effectively defined the η_i 's as 1.

Expanding $\underline{A} \underline{r}^{-\delta}$ in terms of its principal idempotents enables \underline{A} to be written as

$$\underline{A} = \sum_{i=1}^h (\underline{L}_i \epsilon_i) \underline{r}^{\delta} , \quad (2.5.1)$$

where

$$\underline{L}_i = \sum_j (\underline{c}_j \underline{c}'_j) \underline{r}^{-\delta} \quad (i = 1, \dots, h)$$

are the h principal idempotents of $\underline{A} \underline{r}^{-\delta}$. The j summation covers those eigenvectors \underline{c}_j having eigenvalues equal to ϵ_i . Note that the \underline{L}_i are mutually orthogonal, idempotent and are such that

$$\sum_i \underline{L}_i = \underline{I} .$$

The matrix \underline{A}^- can then be written as

$$\underline{A}^- = \underline{r}^{-\delta} \sum_{i=1}^h (\underline{L}_i / \epsilon_i) ,$$

where the summation does not include the principal idempotent corresponding to the zero eigenvalues.

There is, then, no need to calculate each separate \underline{c}_i (which are not unique anyway if there are multiple eigenvalues), just the

principal idempotents (which are unique).

The number of principal idempotents, h , equals the order of balance of the design when a simple treatment model is used (Nelder, 1965).

To find the principal idempotents, it is necessary to find the eigenvalues. Many standard routines exist for doing this, but only one has been specifically developed for experimental designs, that of Wilkinson (1970). This, however, requires a full analysis of dummy data, along with computing the solutions to a polynomial equation of degree equal to h , the order of balance. No other method could be derived and so a standard routine must be used to find the eigenvalues. If Householder's method (1964) is used to reduce the matrix to tridiagonal form, followed by the bisection method of finding the eigenvalues, the total number of multiplications is of the order $(n^3 + 2n^2t)$ (Wilkinson, 1965, p306), where t is the number of steps used in the bisection sequence and n is the order of the matrix.

It is probably easier to find the eigenvalues of $\underline{I} - \underline{Ar}^{-\delta}$ than of $\underline{Ar}^{-\delta}$, since most standard routines find the eigenvalues in decreasing order. Since the eigenvalues of $\underline{Ar}^{-\delta}$, corresponding to efficiencies, are less than or equal to one, those of $\underline{I} - \underline{Ar}^{-\delta}$ are greater than or equal to zero. The routine can then be stopped as soon as a zero eigenvalue is found because any remaining eigenvalues will also be zero. Since eigenvalues of $\underline{Ar}^{-\delta}$ are equal to one minus those of $\underline{I} - \underline{Ar}^{-\delta}$, the zero eigenvalues of $\underline{I} - \underline{Ar}^{-\delta}$ therefore correspond to contrasts estimated with full efficiency. The matrix $\underline{I} - \underline{Ar}^{-\delta}$ is the transpose of the matrix \underline{M} used by Jones (1959) and therefore will in future be written \underline{M}' .

The eigenvalue routine should actually be applied to the symmetric matrix

$$\underline{H} = \underline{r}^{-\frac{1}{2}\delta} \underline{M}' \underline{r}^{\frac{1}{2}\delta},$$

since most standard routines are written for symmetric matrices.

However, since \underline{H} is similar to \underline{M}' , it will have the same eigenvalues (see, for example, Graybill, 1969, p46). The number of distinct eigenvalues, excluding those corresponding to zero efficiency, equals the order of balance of the design, under a simple treatment model.

It is now necessary to find the \underline{L}_i matrices. Two methods will be given.

Method 1

From (2.5.1), \underline{M}' can be written as

$$\sum_{i=1}^h (\mu_i \underline{L}_i), \text{ where } \mu_i = 1 - \epsilon_i. \quad (2.5.2)$$

Since the \underline{L}_i 's are idempotent and mutually orthogonal, $(\underline{M}')^n$ can be written as

$$(\underline{M}')^n = \sum_{i=1}^h (\mu_i^n \underline{L}_i).$$

A set of simultaneous equations in \underline{L}_i can now be set down:

$$\begin{aligned} \underline{I} &= \sum_i \underline{L}_i \\ \underline{M}' &= \sum_i (\mu_i \underline{L}_i) \\ (\underline{M}')^2 &= \sum_i (\mu_i^2 \underline{L}_i) \\ &\vdots \\ (\underline{M}')^{h-1} &= \sum_i (\mu_i^{h-1} \underline{L}_i). \end{aligned} \quad (2.5.3)$$

These equations could possibly be simplified if some of the eigenvectors (basic contrasts) were given. This situation can arise since the experimenter can fairly often guess one or more basic contrasts. It is quite easy to check whether a given contrast is basic and also to find its efficiency by simply pre-multiplying by $\underline{Ar}^{-\delta}$ or, equivalently, by \underline{M}' . If some basic contrasts are found, then their contribution to the set of equations (2.5.3) can be removed. This is analogous to Nelder (1965) testing whether parts of his treatment model are possible spectral matrices of the Normal equations; if all spectral matrices are found then he calls the design balanced under the (complete)

treatment model and the analysis follows on directly. The contributions from any basic contrasts given can be removed in the following way:

Suppose that contrasts c_1, \dots, c_p are known (c_1 is taken to correspond to the mean and will always equal \underline{r}/\sqrt{N} , with $\mu_1 = 1$). Then equations (2.5.3) are re-written as

$$\begin{aligned} \underline{I} - \sum_{j=1}^p (c_j c_j') \underline{r}^{-\delta} &= \sum_{i=1}^h \underline{L}_i^* \\ \underline{M}' - \sum_{j=1}^p (\mu_j c_j c_j') \underline{r}^{-\delta} &= \sum_{i=1}^h (\mu_i \underline{L}_i^*) \\ &\vdots \\ (\underline{M}')^{h-1} - \sum_{j=1}^p (\mu_j^{(h-1)} c_j c_j') \underline{r}^{-\delta} &= \sum_{i=1}^h (\mu_i^{(h-1)} \underline{L}_i^*) \end{aligned} \quad (2.5.4)$$

where

$$\underline{L}_i^* = \underline{L}_i - \sum_k (c_k c_k') \underline{r}^{-\delta},$$

the k summation covering those of the basic contrasts having efficiency equal to ε_i .

The point is that if all the basic contrasts having efficiency ε_i are given, then that \underline{L}_i has been completely determined (i.e. $\underline{L}_i^* = \underline{0}$) and the number of simultaneous matrix equations can be reduced by one for each such \underline{L}_i . This can easily be determined since the multiplicity of each ε_i is known from the application of the standard eigenvalue routine. If the number of basic contrasts having efficiency equal to ε_i equals the multiplicity of ε_i , then that \underline{L}_i has been completely determined.

It should be noted that the adjustments, in (2.5.4), need only be applied to the first equation, since the j^{th} such equation can be generated by multiplying the $(j-1)^{\text{th}}$ by \underline{M}' (because of the orthogonality of the eigenvectors).

Now, if

$$\underline{A}_k = (\underline{M}')^k - \sum_{j=1}^p (\mu_j^k c_j c_j') \underline{r}^{-\delta} \quad (k = 0, \dots, h-1),$$

where $(\underline{M}')^0 = \underline{I}$, then (2.5.4) can be written as

$$\begin{pmatrix} \underline{A}_0 \\ \underline{A}_1 \\ \vdots \\ \underline{A}_{h-1} \end{pmatrix} = \begin{pmatrix} \underline{I} & \underline{I} & \dots & \underline{I} \\ \underline{\mu}_1 & \underline{\mu}_2 & \dots & \underline{\mu}_h \\ \vdots & \vdots & & \vdots \\ \underline{\mu}_1^{h-1} & \underline{\mu}_2^{h-1} & \dots & \underline{\mu}_h^{h-1} \end{pmatrix} \begin{pmatrix} \underline{L}_1^* \\ \underline{L}_2^* \\ \vdots \\ \underline{L}_h^* \end{pmatrix} \quad (2.5.5)$$

where

$$\underline{\mu}_i^j = \mu_i^j \underline{I}.$$

If

$$\underline{\mu} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \mu_1 & \mu_2 & \dots & \mu_h \\ \vdots & \vdots & & \vdots \\ \mu_1^{h-1} & \mu_2^{h-1} & \dots & \mu_h^{h-1} \end{pmatrix},$$

then the matrix of coefficients in (2.5.5) can be written as

$$\underline{\mu}_E = \underline{\mu} \otimes \underline{I},$$

where \otimes signifies direct product. The inverse of $\underline{\mu}_E$ can then be written

$$(\underline{\mu}_E)^{-1} = \underline{\mu}^{-1} \otimes \underline{I},$$

and the solution to (2.5.5) becomes

$$\begin{pmatrix} \underline{L}_1^* \\ \underline{L}_2^* \\ \vdots \\ \underline{L}_h^* \end{pmatrix} = (\underline{\mu}^{-1} \otimes \underline{I}) \begin{pmatrix} \underline{A}_0 \\ \underline{A}_1 \\ \vdots \\ \underline{A}_{h-1} \end{pmatrix},$$

or

$$\underline{L}_i^* = (\underline{\mu}^{-1})_{ij} \cdot \underline{A}_{j-1}.$$

The size of the actual matrix to be inverted is therefore equal to the number of unknown distinct eigenvalues.

The matrix $\underline{\mu}$ is in fact Van der Monde's matrix and has

a determinant equal to

$$\pm \prod_{i < j} (\mu_i - \mu_j) .$$

Since the μ_i 's are distinct, it can be seen that the matrix $\underline{\mu}$ is never singular.

The formula for \underline{A}^{-1} is then

$$\underline{A}^{-1} = \underline{r}^{-1} \sum_{j=1}^p \left(\frac{\underline{c}_j \underline{c}_j'}{\epsilon_j} \right) + \underline{r}^{-1} \sum_i \left(\frac{\underline{L}_i^*}{\epsilon_i} \right),$$

and the analysis follows directly.

Method 2

An alternative method of determining the \underline{L}_i 's is to use the formula

$$\underline{L}_i = \prod_{j \neq i} \frac{(\underline{M}' - \mu_j \underline{I})}{(\mu_i - \mu_j)} ,$$

(see, for example, Pease, 1965, p271).

The formula is derived in the following way:

From (2.5.3),

$$\sum_i \underline{L}_i = \underline{I} \tag{2.5.6}$$

and

$$\sum_i (\mu_i \underline{L}_i) = \underline{M}' . \tag{2.5.7}$$

Multiplying (2.5.6) by μ_j gives

$$\mu_j \sum_i \underline{L}_i = \mu_j \underline{I} . \tag{2.5.8}$$

Then (2.5.7) minus (2.5.8) gives

$$\sum_{i \neq j} [(\mu_i - \mu_j) \underline{L}_i] = \underline{M}' - \mu_j \underline{I} , \tag{2.5.9}$$

that is, the matrix \underline{L}_j has been removed in expression (2.5.9).

Producing equations analogous to (2.5.9) for all $j \neq k$, where \underline{L}_k is the matrix required, and multiplying them all together gives

$$\prod_{j \neq k} [(\mu_k - \mu_j) \underline{L}_k] = \prod_{j \neq k} (\underline{M}' - \mu_j \underline{I}) ,$$

because \underline{L}_k is then the only such matrix common to all the equations and the \underline{L}_i are mutually orthogonal and idempotent. The required formula follows.

This method may also be modified to include any known basic contrasts by changing formulae (2.5.6) and (2.5.7) to

$$\sum_i \underline{L}_i^* = \underline{I} - \sum_{i=1}^p (\underline{c}_i \underline{c}_i') \underline{r}^{-\delta} = \underline{A}_0$$

and

$$\sum_i (\mu_i \underline{L}_i^*) = \underline{M}' - \sum_{i=1}^p (\mu_i \underline{c}_i \underline{c}_i') \underline{r}^{-\delta} = \underline{A}_1$$

Proceeding as before, but noting that the analogous equation to (2.5.9) for removing \underline{L}_j^* is not required if all basic contrasts having efficiency ε_j have been given since \underline{L}_j^* is then $\underline{0}$, gives the formula

$$\underline{L}_i^* = \prod_{j \neq i} \frac{(\underline{A}_1 - \mu_j \underline{A}_0)}{(\mu_i - \mu_j)}$$

The second (Pease) method is probably to be preferred, if hand computation is to be used, if there are few \underline{L}_i^* matrices to compute. In general however, a computer will be used and it will be necessary to examine the number of operations involved in each method.

The first method requires $(h-2)$ matrix multiplications at t^3 operations each to form the \underline{A}_i matrices. The inversion will require $o(h^3)$ operations and there will be a final ht^2 operations involved in forming $(\underline{\mu}^{-1} \otimes \underline{I}) \underline{A}$, giving a total of the order $(h-2)t^3 + h^3 + ht^2$.

The second method requires $(h-2)$ matrix multiplications at t^3 operations each for each \underline{L}_i matrix, the final matrix being divided by $\prod_{i \neq j} (\mu_i - \mu_j)$, that is, a further $(h-2)t + t^2$ operations. The total, therefore, is $h \left\{ (h-2)t^3 + (h-2)t + t^2 \right\}$ operations.

It will be seen that there are always fewer operations in the first method than in the second, since h is always less than or equal to t .

2.6 The Inter-block Analysis

The inter-block analysis is used if the arrangement into blocks in a simple block experiment has been to some extent ineffective. Additional information about the treatments can then be obtained from the differences between blocks if certain assumptions are made as long as sufficient degrees of freedom exist (Fisher and Yates (1963) advise ten degrees of freedom or more for balanced incomplete block designs).

Yates (1939) showed that 'interblock estimators' could be obtained from the block totals and that they were uncorrelated with the intra-block estimators. They could therefore be combined with weights inversely proportional to the respective error variances, this then giving the best linear unbiased estimator (Yates, (1940); Tocher, 1952).

The assumptions to be made are that the block effects are random, with zero mean and variance $\sigma_{\beta_i}^2$ say, for the i^{th} block. The model, now called the 'mixed model', can then be written (with a possible reordering of plots) as

$$y = \underline{A}'\underline{\delta} + \eta$$

with

$$D(\eta) = \sigma^2 \begin{bmatrix} \underline{C}_1 & & & & \\ & \underline{C}_2 & & \underline{0} & \\ & & \ddots & & \\ & \underline{0} & & \ddots & \\ & & & & \underline{C}_b \end{bmatrix} \tag{2.6.1}$$

and $E(\eta) = \underline{0}$,

where $\underline{C}_i = \underline{I} + \frac{\sigma_{\beta_i}^2}{\sigma^2} \underline{1}\underline{1}'$, is a $(k_i \times k_i)$ matrix.

A complete analysis could proceed from here, using the model (2.6.1) to get a single estimate of $\hat{\underline{\delta}}$ (see, for example, Patterson and Thompson, 1971, for the case $\sigma_{\beta_i}^2 = \sigma_{\beta}^2$). However, difficulties arise because of the need to use estimates of the block variance σ_{β}^2 and the error variance σ^2 .

The analysis is simpler if separate estimates of $\hat{\underline{\delta}}$ are made, one

inter- and one intra-block, in the fashion of Yates (1939). For the estimation of the inter-block effects, the model is now based upon blocks rather than plots, i.e. the model is

$$\underline{B} = \underline{n}'\underline{y} + \underline{q} \quad ,$$

where the q_i 's are uncorrelated, with the variance of q_i equal to

$$k_i (1 + k_i \frac{\sigma_{\beta_i}^2}{\sigma^2}) \sigma^2.$$

Now, if the block size is constant, and $\sigma_{\beta_i}^2$ is constant ($= \sigma_{\beta}^2$, say), then $D(q)$ will equal $k \sigma^2 \underline{I}$, where $\sigma^2 = \sigma^2 + k \sigma_{\beta}^2$. If block size varies, but $\frac{\sigma_{\beta_i}^2}{k_i}$ is constant, then no simple analysis is possible. However, if $\sigma_{\beta_i}^2$ were to be inversely proportional to the block size, i.e. $\sigma_{\beta_i}^2 = \sigma_{\beta}^2 / k_i$, then $D(q)$ would equal $k \sigma^2 \underline{I}$, where $\sigma^2 = \sigma^2 + \sigma_{\beta}^2$.

This last proposition is reasonable in as far as it implies that large blocks tend to vary less than small ones. For example, an experimenter can be reasonably sure of the amount of yield from a large block, but realises that unforeseen circumstances can vary the yield from a small block quite considerably.

It should be remembered though, that no physical randomisation is possible with unequal block sizes. The model may still be acceptable however, if, for instance, blocks were randomly selected from larger blocks.

Assuming then, that the dispersion matrix of the errors is $k \sigma_0^2 \underline{I}$, where σ_0^2 depends upon the assumptions made, the analysis can proceed.

A close analogy exists between the inter-block analysis and split plot designs, since the main plot analysis in a split plot design can be made using a similar model with the dispersion matrix $k \{ (1 + (k-1)\rho) \sigma^2 \} \underline{I}$, where ρ is the correlation between experimental errors in the same block.

Using least squares theory, the Normal equations for a model

$$\underline{B} = \underline{n}'\underline{y}_I + \underline{q} \quad \text{with} \quad D(q) = k \sigma_0^2 \underline{I}, \quad (2.6.2)$$

are
$$(\underline{nk}^{-\delta} \underline{n}') \hat{\underline{Y}}_I = \underline{nk}^{-\delta} \underline{B} . \quad (2.6.3)$$

The parameter $\hat{\underline{Y}}_I$ has been given a subscript I to differentiate between the inter and intra-block estimates of the treatment parameter.

Now,

$$\underline{nk}^{-\delta} \underline{n}' = \underline{r}^{\delta} - \underline{A} = \underline{A}_0 \text{ say,} \quad (2.6.4)$$

where \underline{A} is the matrix used in the intra-block analysis and can be written

$$\underline{A} = \sum_i (\underline{c}_i \underline{c}_i' \varepsilon_i) = \underline{C} \underline{\varepsilon} \underline{C}', \quad (\text{from 2.3.2})$$

where $\underline{C} = \{\underline{c}_i\}$ is the matrix having the eigenvectors of $\underline{A} \underline{r}^{-\delta}$ as its columns with eigenvalues ε_i . The eigenvectors are scaled such that

$$\underline{C}' \underline{r}^{-\delta} \underline{C} = \underline{I} . \quad (\text{from 2.3.1})$$

Since \underline{C} is non-singular, premultiplying (2.3.1) by \underline{C} and postmultiplying by \underline{C}^{-1} leads to

$$\underline{C} \underline{C}' = \underline{r}^{\delta} = \sum_i (\underline{c}_i \underline{c}_i') . \quad (2.6.5)$$

Using (2.6.5) and (2.3.2) enables equation (2.6.4) to be written as

$$\begin{aligned} \underline{A}_0 &= \sum_i (\underline{c}_i \underline{c}_i') - \sum_i (\underline{c}_i \underline{c}_i' \varepsilon_i), \\ &= \sum_i [(1 - \varepsilon_i) \underline{c}_i \underline{c}_i'], \\ &= \sum_i (\mu_i \underline{c}_i \underline{c}_i') = \underline{C} \underline{\mu} \underline{C}' , \end{aligned} \quad (2.6.6)$$

where $\mu_i = 1 - \varepsilon_i$.

The Normal equations, (2.6.3), can now be written

$$(\underline{C} \underline{\mu} \underline{C}') \hat{\underline{Y}}_I = \underline{nk}^{-\delta} \underline{B} . \quad (2.6.7)$$

Notice that the basic contrasts, \underline{c}_i , again play an important role, but their efficiencies' are now μ_i instead of ε_i , corresponding to the fact that the inter-block stratum is being used.

It will be seen that the same \underline{L}_i matrices that were used in the intra-block analysis, that is $\underline{L}_i = \sum_j (\underline{c}_j \underline{c}_j') \underline{r}^{-\delta}$, the summation being over

those eigenvectors having eigenvalue ε_i , can be used in the inter-block analysis. This is because a generalized inverse of (2.6.6), needed to give a consistent solution to the Normal equations (2.6.7), can be written as

$$\underline{A}_0^- = \underline{r}^{-\delta} \sum_i \frac{(\underline{c}_i \underline{c}_i')}{\mu_i} \underline{r}^{-\delta} = \underline{r}^{-\delta} \sum_j \frac{(\underline{L}_j)}{\mu_j}, \quad (2.6.8)$$

where the i summation covers only those basic contrasts having non-zero inter-block efficiencies μ_i , and the j summation is over all \underline{L}_j except the matrix corresponding to zero efficiencies. As in the intra-block case, this choice of generalized inverse is the one involving the least computation. It is in general not the same as the $\underline{\Omega}_0$ of Pearce, Calinski and Marshall (1974) who effectively defined $1/\mu_i$ as 1 when $\mu_i = 0$.

Because the \underline{L}_i matrices will already have been found from the previous intra-block analysis, then \underline{A}_0^- can be calculated and hence the complete inter-block analysis can be carried out without difficulty. The total sum of squares from the model (2.6.2) is $\underline{B}' \underline{k}^{-\delta} \underline{B}$, corresponding to the block sum of squares plus that for the mean in the original analysis of variance table, and it is this which will be partitioned to give the inter-block treatment sum of squares with a corresponding residual sum of squares and that for the mean.

The weights required for combining the two estimates of the treatment parameters can now be estimated from the two residual mean squares; one estimating σ_e^2 , the other, σ^2 . If there are not enough degrees of freedom for the error mean squares, then the method of Nelder (1968) for estimating weights may be used.

2.7 Partitioning a Sum of Squares According to Contrasts of Interest

It frequently happens that an experimenter will wish to know the contribution to the treatment sum of squares given by some component of the treatment parameters. For example, he may be interested in

whether or not there is a significant difference between a control treatment and some other treatment, or whether, in a factorial experiment, the average response can be considered linear or whether quadratic or even higher orders need be taken into account. For standard designs this is reasonably simple, but for complicated designs it can be difficult. One way, for calculating the sum of squares for the difference between two treatments, would be to analyse the design and compute the treatment sum of squares, and then to merge the two treatments together, analyse this new design and compute the new treatment sum of squares. The difference between the two is then the sum of squares for the difference between the two treatments.

A formula for calculating the sum of squares corresponding to some general hypothesis is given by Pringle and Rayner (1971) and extended by John and Smith (1974). An alternative derivation of this formula, using an extended result of Rao (1966) will now be given.

The sum of squares corresponding to some linear hypothesis or system of linear constraints on the parameters can be defined as the difference between the residual sum of squares with the constraints applied and the residual sum of squares without them. An equivalent definition is that used by Plackett (1960, p54), who defined the sum of squares corresponding to a linear hypothesis as the difference between the relative minimum of

$$W = (\underline{y} - \underline{X}\underline{\theta})'(\underline{y} - \underline{X}\underline{\theta}) \quad (\text{using the model } \underline{y} = \underline{X}\underline{\theta} + \underline{\eta})$$

when the hypothesis is true and the absolute minimum of W .

Expanding W as

$$\{(\underline{y} - \underline{X}\hat{\underline{\theta}}) + \underline{X}(\hat{\underline{\theta}} - \underline{\theta})\}' \{(\underline{y} - \underline{X}\hat{\underline{\theta}}) + \underline{X}(\hat{\underline{\theta}} - \underline{\theta})\},$$

where $\hat{\underline{\theta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$, the least squares estimate of $\underline{\theta}$ (without the constraints), gives

$$W = (\underline{y} - \underline{X}\hat{\underline{\theta}})'(\underline{y} - \underline{X}\hat{\underline{\theta}}) + (\hat{\underline{\theta}} - \underline{\theta})'\underline{X}'\underline{X}(\hat{\underline{\theta}} - \underline{\theta}) + 2(\hat{\underline{\theta}} - \underline{\theta})'\underline{X}'(\underline{y} - \underline{X}\hat{\underline{\theta}}).$$

But,

$$\begin{aligned} \underline{X}'(\underline{y} - \underline{X}\hat{\underline{\theta}}) &= \underline{X}'(\underline{y} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}) \\ &= (\underline{X}' - (\underline{X}'\underline{X})(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{y} . \end{aligned}$$

Now, consider

$$\begin{aligned} &\left\{ \underline{X}(\underline{X}'\underline{X})^{-1}(\underline{X}'\underline{X}) - \underline{X} \right\}' \left\{ \underline{X}(\underline{X}'\underline{X})^{-1}(\underline{X}'\underline{X}) - \underline{X} \right\} \\ &= \left\{ (\underline{X}'\underline{X})^{-1}(\underline{X}'\underline{X}) - \underline{I} \right\}' \underline{X}' \left\{ \underline{X}(\underline{X}'\underline{X})^{-1}(\underline{X}'\underline{X}) - \underline{X} \right\} \\ &= \left\{ (\underline{X}'\underline{X})^{-1}(\underline{X}'\underline{X}) - \underline{I} \right\}' \left\{ (\underline{X}'\underline{X})(\underline{X}'\underline{X})^{-1}(\underline{X}'\underline{X}) - \underline{X}'\underline{X} \right\} , \end{aligned}$$

and the second factor vanishes by definition (2.4.1) of a generalized inverse, which implies that $\underline{X}' - (\underline{X}'\underline{X})(\underline{X}'\underline{X})^{-1}\underline{X}' = \underline{0}$.

The expansion of W can now be written as

$$W = (\underline{y} - \underline{X}\hat{\underline{\theta}})'(\underline{y} - \underline{X}\hat{\underline{\theta}}) + (\hat{\underline{\theta}} - \underline{\theta})'\underline{X}'\underline{X}(\hat{\underline{\theta}} - \underline{\theta}) .$$

Because the least squares estimate of $\underline{\theta}$ when the hypothesis is true is obtained by minimising W under those constraints, then, since the first expression in the expansion of W above is constant, it will be seen that the desired estimate can also be obtained by minimising

$$(\hat{\underline{\theta}} - \underline{\theta})'\underline{X}'\underline{X}(\hat{\underline{\theta}} - \underline{\theta}) \tag{2.7.1}$$

subject to the linear constraints upon $\hat{\underline{\theta}}$.

The actual minimum obtained is then equal to the sum of squares corresponding to the hypothesis, h say, since the absolute minimum of W is clearly

$$(\underline{y} - \underline{X}\hat{\underline{\theta}})'(\underline{y} - \underline{X}\hat{\underline{\theta}}) ,$$

the first term in the expansion of W above, as (2.7.1) is a positive definite quadratic form.

Rao (1966) gives the infimum of a quadratic form $\underline{X}'\underline{\Lambda}\underline{X}$ under the constraint $\underline{B}'\underline{X} = \underline{U}$, where $\underline{\Lambda}$ is a positive definite matrix, as $\underline{U}'\underline{S}^{-1}\underline{U}$, where $\underline{S} = \underline{B}'\underline{\Lambda}^{-1}\underline{B}$. This can be extended to a singular positive semi-definite $\underline{\Lambda}$, if certain restrictions are placed upon the choice of \underline{B} , in the following way:

Let $\underline{\Lambda}^{-}$ be the generalized inverse of $\underline{\Lambda}$ and \underline{S}^{-} the generalized

inverse of $\underline{B}'\underline{\Lambda}^{-}\underline{B}$, satisfying only condition (2.4.1).

It is assumed that $\underline{B}'\underline{X} = \underline{U}$ is consistent, then if \underline{B} is such that $\underline{B}' = \underline{C}\underline{\Lambda}$, i.e. estimable, $\text{rank}(\underline{B}'\underline{\Lambda}^{-}\underline{B}) = \text{rank}(\underline{B})$, and so the equation $\underline{B}'\underline{\Lambda}^{-}\underline{B}\underline{\lambda} = \underline{U}$ is consistent and the following hold, where $\underline{X}^* = \underline{\Lambda}^{-}\underline{B}\underline{S}^{-}\underline{U}$,

$$\underline{B}'\underline{X}^* = \underline{B}'\underline{\Lambda}^{-}\underline{B}\underline{S}^{-}\underline{U} = \underline{S}\underline{S}^{-}\underline{S}\underline{\lambda} = \underline{S}\underline{\lambda} = \underline{U} ,$$

and

$$\underline{X}^*\underline{\Lambda}\underline{X}^* = (\underline{U}'\underline{S}^{-}'\underline{B}'\underline{\Lambda}^{-}')\underline{\Lambda}(\underline{\Lambda}^{-}\underline{B}\underline{S}^{-}\underline{U}) ,$$

which, using the facts that $\underline{B}' = \underline{C}\underline{\Lambda}$, $\underline{S}\underline{S}^{-}\underline{U} = \underline{U}$ and $\underline{A}\underline{A}^{-}'\underline{A} = \underline{A}$ for a symmetric generalized inverse \underline{A} , reduces to

$$\underline{X}^*\underline{\Lambda}\underline{X}^* = \underline{U}'\underline{S}^{-}\underline{U} .$$

It is now necessary to show that $\underline{X}'\underline{A}\underline{X} \geq \underline{U}'\underline{S}^{-}\underline{U}$ for any other \underline{X} satisfying the condition $\underline{B}'\underline{X} = \underline{U}$.

Consider the identity

$$(\underline{Y}_1 - \underline{Y}_2)'(\underline{Y}_1 - \underline{Y}_2) = \underline{Y}_1'\underline{Y}_1 - \underline{Y}_1'\underline{Y}_2 - \underline{Y}_2'\underline{Y}_1 + \underline{Y}_2'\underline{Y}_2 .$$

The L.H.S. is clearly positive and so

$$\underline{Y}_1'\underline{Y}_1 \geq \underline{Y}_1'\underline{Y}_2 + \underline{Y}_2'\underline{Y}_1 - \underline{Y}_2'\underline{Y}_2 .$$

Now, let $\underline{Y}_1 = \underline{\Lambda}^{\frac{1}{2}}\underline{X}$, $\underline{Y}_2 = \underline{\Lambda}^{\frac{1}{2}}\underline{X}^*$, then

$$\underline{Y}_1'\underline{Y}_2 = \underline{X}'\underline{B}\underline{S}^{-}\underline{U} = \underline{U}'\underline{S}^{-}\underline{U} = \underline{Y}_2'\underline{Y}_2 .$$

Therefore,

$$\underline{X}'\underline{X} \geq \underline{U}'\underline{S}^{-}\underline{U} + \underline{U}'\underline{S}^{-}\underline{U} - \underline{U}'\underline{S}^{-}\underline{U} = \underline{U}'\underline{S}^{-}\underline{U} .$$

The minimum of $\underline{X}'\underline{X}$, where $\underline{B}'\underline{X} = \underline{U}$, is then $\underline{U}'\underline{S}^{-}\underline{U}$ if \underline{B} is such that $\underline{B}'\underline{\Lambda}^{-}\underline{B} = \underline{U}$ is consistent. This minimum is unique under the choices of $\underline{\Lambda}^{-}$ and \underline{S}^{-} .

Applying this result to (2.7.1), with the hypothesis h as $\underline{L}'\underline{\theta} = \underline{z}$ ($\underline{U} = \underline{L}'\underline{\hat{\theta}} - \underline{z}$), the hypothesis being consistent, gives the sum of squares as

$$(\underline{L}'\underline{\hat{\theta}} - \underline{z})'(\underline{L}'(\underline{X}'\underline{X})^{-}\underline{L})^{-}(\underline{L}'\underline{\hat{\theta}} - \underline{z}) ,$$

which is the same as John and Smith's (1974) result.

Now, let $\underline{\epsilon}$ be partitioned as

$$\underline{\epsilon} = \begin{bmatrix} \underline{\gamma} \\ \dots \\ \underline{\beta} \end{bmatrix},$$

where $\underline{\gamma}$ is the parameter vector for treatments and $\underline{\beta}$ are nuisance parameters. The hypothesis to be tested is $\underline{C}'\underline{\gamma} = \underline{0}$. This gives $\underline{B} = (\underline{C}' : \underline{0})$ and $\underline{U} = \underline{C}'\hat{\underline{\gamma}}$. If $\underline{X}'\underline{X}$ is partitioned conformably, then $(\underline{X}'\underline{X})^{-1}$ can be written as

$$(\underline{X}'\underline{X})^{-1} = \begin{bmatrix} \underline{A}^{-1} & \underline{Y} \\ \underline{Y}' & \underline{Z} \end{bmatrix},$$

where \underline{A}^{-1} is the matrix (2.5.0) used in the intra-block analysis; that is, the dispersion matrix for the treatment parameters (e.g. Rhode, 1968, p 245).

The minimum of (2.7.1) can then be written as

$$(\underline{C}'\hat{\underline{\gamma}})'(\underline{C}'\underline{A}^{-1}\underline{C})^{-1}(\underline{C}'\hat{\underline{\gamma}}).$$

The conditions placed upon \underline{B} were that $\underline{B}'\underline{A}^{-1}\underline{B}\underline{\lambda} = \underline{U}$ must be consistent, i.e. \underline{C} must be such that $\underline{C}'\underline{A}^{-1}\underline{C}\underline{\lambda} = \underline{C}'\hat{\underline{\gamma}}$ is consistent. This is satisfied if $\underline{C}'\underline{A}^{-1}\underline{C}$ is of full rank (though not a necessary condition). A particular case occurs if \underline{C} is simply a vector, whence $\underline{C}'\underline{A}^{-1}\underline{C}$ is a scalar, being non-zero if $\text{var}(\underline{C}'\hat{\underline{\gamma}})$ is non-zero, i.e. if $\underline{C}'\hat{\underline{\gamma}}$ is estimable. If $\underline{C} = [c]$, then the sum of squares for the above case is

$$\frac{(\underline{c}'\hat{\underline{\gamma}})^2}{\underline{c}'\underline{A}^{-1}\underline{c}}.$$

Suppose now that \underline{C} is such that $\underline{C}'\underline{A}^{-1}\underline{C} = \underline{I}_{(t \times t)}$, where $\text{rank}(\underline{A}^{-1}) = t$.

Now \underline{C} can be written as $\underline{S}\underline{D}$, where \underline{S} is the matrix of basic contrasts, for some \underline{D} . If \underline{S} is partitioned as $\begin{bmatrix} \underline{S}_n \\ \underline{S}_z \end{bmatrix}$, where \underline{S}_z holds the basic contrasts with zero efficiencies, then \underline{D} can be conformably partitioned as $\begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix}$, where \underline{Y} is a square matrix. Using the expansion (2.5.0) for \underline{A}^{-1} and the relation $\underline{S}'\underline{r}^{-1}\underline{S} = \underline{I}$ between basic contrasts implies that

$$\underline{C}'\underline{A}^{-1}\underline{C} = \underline{Y}'\underline{r}^{-1}\underline{Y} = \underline{I},$$

where $\underline{\epsilon}^{-\delta}$ consists of the non-zero efficiencies only. The matrix \underline{Y} is therefore of full rank and so $\underline{Y}\underline{Y}' = \underline{\epsilon}^{\delta}$.

Then $\underline{C}\underline{C}'$ can be expanded as

$$\begin{aligned} \underline{S}\underline{D}\underline{D}'\underline{S}' &= \underline{S} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} \begin{bmatrix} \underline{Y}' & \underline{Z}' \end{bmatrix} \underline{S}' = \begin{bmatrix} \underline{S}_n & \underline{S}_z \end{bmatrix} \begin{bmatrix} \underline{\epsilon}^{\delta} & \underline{E} \\ \underline{E}' & \underline{F} \end{bmatrix} \begin{bmatrix} \underline{S}'_n \\ \underline{S}'_z \end{bmatrix} \\ &= \underline{S}_n \underline{\epsilon}^{\delta} \underline{S}'_n + \underline{S}_n \underline{E} \underline{S}'_z + \underline{S}_z \underline{E}' \underline{S}'_n + \underline{S}_z \underline{F} \underline{S}'_z \\ &= \underline{A} + \underline{S}_n \underline{E} \underline{S}'_z + \underline{S}_z \underline{E}' \underline{S}'_n + \underline{S}_z \underline{F} \underline{S}'_z . \end{aligned}$$

Since $\hat{\underline{Y}} = \underline{A}^{-1}\underline{Q}$, $\hat{\underline{Y}}'\underline{C}\underline{C}'\hat{\underline{Y}}$ can be written as $\underline{Q}'\underline{A}^{-1}\underline{C}\underline{C}'\underline{A}^{-1}\underline{Q}$, and because $\underline{A}^{-1}\underline{S}_z = \underline{0}$, $\hat{\underline{Y}}'\underline{C}\underline{C}'\hat{\underline{Y}}$ reduces to $\hat{\underline{Y}}'\underline{A}\hat{\underline{Y}}$, the normal treatment sum of squares.

The sums of squares, therefore, corresponding to the $\underline{c}_i \hat{\underline{Y}}$ together form a complete partition of the sum of squares. This implies that if contrasts are chosen such that $\underline{c}_i' \underline{A}^{-1} \underline{c}_j$ is zero if $i \neq j$, and is non-zero for $i = j$, then these contrasts form a complete partition of the treatment sum of squares, the only condition being that the covariance between them is zero and the variances are non-zero. Note that this condition is sufficient but not necessary.

For the partition to be useful, it is necessary that

$$(\underline{C}'\underline{Y})'(\underline{C}'\underline{A}^{-1}\underline{C})^{-1}(\underline{C}'\underline{Y})$$

has a chi-square distribution under the assumption that

$$\underline{C}'\underline{Y} \sim N(\underline{0}, (\underline{C}'\underline{A}^{-1}\underline{C})) ,$$

for some \underline{C} where $\underline{C}'\underline{A}^{-1}\underline{C}$ is perhaps singular.

Rao (1966) gives a sufficient condition as

$$\underline{\Lambda}(\underline{A}\underline{\Lambda}\underline{A} - \underline{A})\underline{\Lambda} = \underline{0} ,$$

where the quadratic form is $\underline{Y}'\underline{A}\underline{Y}$ and $D(\underline{Y}) = \underline{\Lambda}$. Inserting $(\underline{C}'\underline{A}^{-1}\underline{C})$ as $\underline{\Lambda}$ and $(\underline{C}\underline{A}^{-1}\underline{C})^{-1}$ as \underline{A} shows that the condition does indeed hold. The degrees of freedom corresponding to the quadratic form is $\text{tr}(\underline{A}\underline{\Lambda})$, again from Rao (1966). In this case, this gives $\text{tr}((\underline{C}'\underline{A}^{-1}\underline{C})^{-1}(\underline{C}'\underline{A}^{-1}\underline{C}))$. If $\underline{C}'\underline{A}^{-1}\underline{C}$ is non-singular, the number of degrees of freedom is equal to the number of columns of \underline{C} , i.e. the number of separate contrasts. If there is a remainder part of the sum of squares, then this is also distributed as Chi-square since there must exist some contrasts to augment those

given to produce a matrix \underline{C} such that $\underline{C}'\underline{A}^{-}\underline{C}$ is of rank equal to that of \underline{A}^{-} (since such a matrix always exists). The augmented part of \underline{C} would then form a quadratic form with a chi-square distribution.

A similar argument can also be applied to the inter-block treatment sum of squares. Here the model is

$$E(\underline{B}) = \underline{n}'\underline{y}_1,$$

$$D(\underline{B}) = \underline{k}\sigma_0^2.$$

Because the dispersion matrix is not $\underline{I}\sigma_0^2$, the analogous equation to (2.7.1) for minimising is

$$(\hat{\underline{y}}_1 - \underline{y}_1)' \underline{nk}^{-\delta} \underline{n}' (\hat{\underline{y}}_1 - \underline{y}_1). \quad (2.7.2)$$

The minimum of this, under the constraints $\underline{C}'\hat{\underline{y}}_1 = \underline{0}$ is

$$(\underline{C}'\hat{\underline{y}}_1)' (\underline{C}'\underline{A}_0^{-}\underline{C})^{-} (\underline{C}'\hat{\underline{y}}_1),$$

where $\underline{A}_0 = \underline{nk}^{-\delta} \underline{n}'$, if \underline{C} is such that $\underline{C}'\underline{A}_0^{-}\underline{C}\underline{\lambda} = \underline{C}'\hat{\underline{y}}_1$ is consistent.

This can be satisfied if $\underline{C}'\underline{A}_0^{-}\underline{C}$ is of full rank (not a necessary condition). The particular case then of \underline{C} being a vector and $\underline{c}'\underline{A}_0^{-}\underline{c}$ being a non-zero scalar gives the sum of squares according to the hypothesis $\underline{c}'\underline{y}_1 = 0$ as

$$\frac{(\underline{c}'\hat{\underline{y}}_1)^2}{\underline{c}'\underline{A}_0^{-}\underline{c}}.$$

In an exactly analogous proof to that used in the intra-block stratum, it follows that the general sum of squares

$$(\underline{C}'\hat{\underline{y}}_1)' (\underline{C}'\underline{A}_0^{-}\underline{C})^{-} (\underline{C}'\hat{\underline{y}}_1)$$

has a chi-square distribution with the number of degrees of freedom equal to $\text{tr}[(\underline{C}'\underline{A}_0^{-}\underline{C})^{-}(\underline{C}'\underline{A}_0^{-}\underline{C})]$, or in the case of $(\underline{C}'\underline{A}_0^{-}\underline{C})$ being of full rank, the number of columns of \underline{C} .

If the contrast used corresponds to a basic contrast, then $\underline{c}'\underline{Dc}_j$ always equals zero, where \underline{D} is either \underline{A}^{-} or \underline{A}_0^{-} depending on the stratum involved. Also, if the efficiency of a contrast in the intra-block

stratum is not zero, then $\underline{c}'\underline{A}^{-1}\underline{c}$ equals $1/\epsilon$, and if non-zero in the inter-block stratum equals $1/\mu$. The sum of squares corresponding to \underline{c} is then $(\underline{c}'\hat{\underline{y}})^2/\epsilon$ or $(\underline{c}'\hat{\underline{y}}_1)^2/\mu$ depending on the stratum.

2.8 Derivation of Basic Contrasts in Certain Special Cases

From the previous sections it has been shown that if some basic contrasts are known then a simplification of the analysis may result. Now it may happen that in a design certain treatments are what is known as proportionate (Pearce, 1971). Two treatments are termed proportionate if

$$n_{1j} = \theta n_{2j} ,$$

for all j ; the treatments being 1 and 2, and n_{ij} an element of the incidence matrix. Pearce shows that differences between the two proportionate treatments are estimated with full efficiency. In fact, such a contrast is basic, and if several treatments are proportionate, i.e.

$$n_{1j} = \theta n_{2j} = \phi n_{3j} = \gamma n_{4j} \quad \text{etc.},$$

then an orthogonal set of basic contrasts between them could be written as

$$\begin{bmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}, \begin{bmatrix} 1 \\ \theta \\ -(\theta+1) \\ 0 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}, \begin{bmatrix} 1 \\ \theta \\ \phi \\ -(\theta+\phi+1) \\ 0 \\ \vdots \\ \vdots \end{bmatrix}, \quad \text{etc..}$$

A further property treatments may have is defined by Pearce as equivalence. Two or more treatments are said to be equivalent if interchanging them has no effect on the analysis of the design. Pearce shows that contrasts between such treatments are estimated with efficiency $(\alpha - \beta)/r$, where α is the diagonal element and β the off-diagonal element of Tocher's $\underline{\Omega}^{-1}$ (1952), or, equivalently, \underline{A} , corresponding to the equivalent treatments, and r is the replication of one of the treatments. Such contrasts also turn out to be basic.

2.9 A program for the analysis of a general block design

A program, GEN, for the analysis of a general block design has been written in ALGOL 60 to be run on the ICL 4130. A listing of the program, together with some example runs are given in appendix B.

The program requires only the incidence matrix of the design to be analysed and the data from the experiment. Contrasts may optionally be given to the program. If some contrasts are basic, the program will attempt to simplify the analysis using these. The treatment sum of squares in both strata are partitioned according to the basic contrasts. The non-basic contrasts have their corresponding sum of squares given separately and a matrix giving the variances and correlations of the contrasts enables the experimenter to partition his treatment sum of squares further according to the non-basic contrasts if they are independent of the basic contrasts (that is, $c_i'Ac_j = 0$ $i \neq j$). If some are not independent, the sum of squares can still be of use since the experimenter knows the correlation between them and the other contrasts.

The program will always give the matrix M' , the order of balance of the experiment, the efficiency factors, the grand mean, treatment means and block means, the treatment effects from the inter- and intra-block analyses, the block effects if the mixed model is not used, the analysis of variance table and the dispersion matrix for the estimable contrasts of the treatments, i.e. A^- .

2.10 Discussion

The original problem was to find a general method for the complete analysis of any block design, whether confounded or unbalanced. The method given may seem to be rather a difficult one in that the eigenvalues of the design must be obtained and a matrix inversion calculated.

However, for designs that are not connected, information about the eigensystem of the design seems essential if non-iterative methods are to be used. Wilkinson's method (1970) can yield the distinct efficiency

factors, but not only requires a complete analysis of the design with dummy variates substituted for the data, but then requires the roots of a polynomial of degree equal to the balance of the design. The finding of these roots, which are the efficiencies of the design, can present difficulties if some degree of imbalance exists since iterative methods of solution would be needed for orders of balance greater than four. It is true that the method presented here requires the inversion of a matrix of order equal to the balance, but that can proceed directly since the matrix is known not to be singular.

Comparing this direct method with the iterative method presented by Kuiper (1952), it may be seen that the work involved in the direct method is proportional to the degree of imbalance, whereas in the iterative method it may be shown (Worthington, 1975) that the work involved is proportional to the highest amount of incomplete confounding that exists, that is, the lowest non-zero efficiency factor governs the speed of convergence - the lower the value, the slower the convergence.

However, the direct method lends itself much more readily to the analysing of many experiments, all of which have the same design, since most of the work is done in calculating the L_1 matrices, these depending on the design only and therefore need only be calculated the once. A further advantage is that once the L_1 matrices have been found, the inter-block analysis and (as will be shown in chapter 3) that of the dual immediately follow on.

The use of treatment contrasts when analysing a design is useful since not only do they permit a possible simplification of the analysis if some are found to be basic, but also several, not necessarily related, questions may be asked of the experiment, each having its own estimate of reliability due to the fact that an F-test can be made for each contrast.

An improvement to the method could be made by finding a better method of determining the efficiency factors. It seems wasteful that when so much is known of the structure of the matrix \underline{M} , recourse has to be made to complicated methods.

3.1 Introduction

The dual of a block design is defined as that design generated from the original when the treatments are used for the blocking system and the blocks are taken as treatments. The incidence matrix is therefore the transpose of the original.

The idea of dual designs appears to date back to Bose and Nair (1939) who used them as a means of generating new designs. However, dual designs sometimes have a practical importance as can be seen in the case when the blocks of a design actually correspond to the residual effects of former treatments. It may now be important to study the block differences eliminating treatments and Pearce (1968, 1970) makes the point that this is achieved by considering the dual of the design. It will be shown that the dispersion matrices for the 'treatment' effects in the inter and intra block analyses for the dual are directly related to a combination of those of the original design; a separate calculation of the analysis of variance is therefore not required.

Relationships between the efficiency factors (Jones, 1959) and basic contrasts of one design and those of the other are given, as are relationships between the various measures of mean efficiency of one design and those of the other.

3.2 The eigensystem of the dual design

The following theorem relates the eigensystem of the dual with that of the original design.

Theorem 3.2

If an eigenvalue of a design is not equal to unity then it is also an eigenvalue of the dualised design and vice versa. Further, there is a direct relation between the eigenvectors corresponding to that eigenvalue in the two designs; namely, if \underline{c} is an eigenvector of the original design with non-unit eigenvalue, ϵ , then

$$\underline{c}_* = \underline{n}' \underline{r}^{-\delta} \underline{c} / \sqrt{(1-\epsilon)}$$

is the corresponding eigenvector for the dual.

Proof

The eigensystem of a design was defined in chapter 2 as the system of eigenvectors and eigenvalues of $\underline{A}r^{-\delta}$. Then

$$(\underline{A}r^{-\delta})\underline{C} = \underline{C}\underline{\epsilon}^{\delta} \quad (3.2.1)$$

where \underline{C} is the matrix having eigenvectors \underline{c}_i as its columns, with eigenvalues ϵ_i held in $\underline{\epsilon}^{\delta}$.

The matrix $\underline{A}r^{-\delta}$ can be written as

$$\underline{A}r^{-\delta} = \underline{I} - \underline{n} \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \quad (3.2.2)$$

from the definition of \underline{A} , (2.2.1).

The analogue of \underline{A} in the dual design, \underline{A}_* , is defined in the manner of \underline{A} , except that \underline{n} is replaced by \underline{n}' , \underline{r} by \underline{k} and \underline{k} by \underline{r} , giving

$$\underline{A}_* = \underline{k}^{\delta} - \underline{n}' \underline{r}^{-\delta} \underline{n}$$

and
$$\underline{A}_* \underline{k}^{-\delta} = \underline{I} - \underline{n}' \underline{r}^{-\delta} \underline{n} \underline{k}^{-\delta} \quad (3.2.3)$$

The eigensystem of the dual is then the system of eigenvectors and eigenvalues of $\underline{A}_* \underline{k}^{-\delta}$.

Now, pre-multiplying (3.2.2) by $\underline{n}' \underline{r}^{-\delta}$ and post-multiplying (3.2.3) by the same, yields the result

$$\underline{n}' \underline{r}^{-\delta} \underline{A}r^{-\delta} = \underline{A}_* \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \quad (3.2.4)$$

But, from (3.2.1)

$$\underline{n}' \underline{r}^{-\delta} \underline{A}r^{-\delta} \underline{C} = \underline{n}' \underline{r}^{-\delta} \underline{C} \underline{\epsilon}^{\delta} \quad (3.2.5)$$

and using (3.2.4), (3.2.5) can be written

$$\underline{A}_* \underline{k}^{-\delta} (\underline{n}' \underline{r}^{-\delta} \underline{C}) = (\underline{n}' \underline{r}^{-\delta} \underline{C}) \underline{\epsilon}^{\delta}.$$

Non-zero columns of the $(b_{\times v})$ matrix $(\underline{n}' \underline{r}^{-\delta} \underline{C})$, therefore, are eigenvectors of $\underline{A}_* \underline{k}^{-\delta}$, with eigenvalues given by $\underline{\epsilon}^{\delta}$, thus, by definition, they form part of the eigensystem of the dual.

However, these non-zero columns could be linearly dependent, and to prove they are not, it is sufficient to show that they satisfy $\underline{B}'\underline{k}^{-\delta}\underline{B}$ being a diagonal matrix, where \underline{B} is a matrix of eigenvectors of the dual.

Consider then,

$$(\underline{n}'\underline{r}^{-\delta}\underline{c})'\underline{k}^{-\delta}(\underline{n}'\underline{r}^{-\delta}\underline{c}) \quad (3.2.6)$$

which equals

$$\underline{c}'\underline{r}^{-\delta}(\underline{nk}^{-\delta}\underline{n}')\underline{r}^{-\delta}\underline{c} . \quad (3.2.7)$$

But, from (3.2.2)

$$(\underline{nk}^{-\delta}\underline{n}') = \underline{r}^{\delta} - \underline{A} .$$

Therefore (3.2.7) can be written as

$$\underline{c}'\underline{r}^{-\delta}(\underline{r}^{\delta} - \underline{A})\underline{r}^{-\delta}\underline{c}$$

or $\underline{c}'\underline{r}^{-\delta}\underline{c} - \underline{c}'\underline{r}^{-\delta}(\underline{Ar}^{-\delta})\underline{c}$,

which simplifies to

$$\underline{I} - \underline{\varepsilon}^{\delta} = \underline{\mu}^{\delta} \quad (3.2.8)$$

on using (3.2.1) and the fact that

$$\underline{c}'\underline{r}^{-\delta}\underline{c} = \underline{I} . \quad (2.3.1)$$

Thus non-zero columns of $(\underline{n}'\underline{r}^{-\delta}\underline{c})$ are indeed independent.

It is now required to show that a column of $(\underline{n}'\underline{r}^{-\delta}\underline{c})$ is non-zero if and only if the corresponding eigenvalue is non-unity; or, equivalently, that a column is zero if and only if the corresponding eigenvalue is equal to unity.

Assume that $\underline{n}'\underline{r}^{-\delta}\underline{c}_i$ is zero, then $(\underline{n}'\underline{r}^{-\delta}\underline{c}_i)'\underline{k}^{-\delta}(\underline{n}'\underline{r}^{-\delta}\underline{c}_i)$ must be zero and hence, from (3.2.6) and (3.2.8), ε_i must be equal to unity.

Assume now that ε_i equals unity, then from (3.2.8) and (3.2.6) $(\underline{n}'\underline{r}^{-\delta}\underline{c}_i)'\underline{k}^{-\delta}(\underline{n}'\underline{r}^{-\delta}\underline{c}_i)$ must be zero. But $\underline{k}^{-\delta}$ is a positive definite matrix, therefore $(\underline{n}'\underline{r}^{-\delta}\underline{c}_i)$ must be a zero vector.

Therefore, since a non-unit eigenvalue of the original design

corresponds to a non-null vector, and therefore to an eigenvector of the dual design, all non-unit eigenvalues of the original design are eigenvalues of the dualised design. Also, since the argument can equally well be applied starting with the dual design and forming the dual of that (that is, the original design), any non-unit eigenvalue of the dual is an eigenvalue of the original design.

Finally, since eigenvectors of the dual are scaled to satisfy

$$\underline{b}'\underline{k}^{-\delta}\underline{b} = 1,$$

then, from (3.2.8), non-null columns of $(\underline{n}'\underline{r}^{-\delta}\underline{c})$ should be scaled to $\underline{n}'\underline{r}^{-\delta}\underline{c}_i / \sqrt{(1 - \varepsilon_i)^2}$. This concludes the theorem.

Some important consequences can be drawn from the above theorem:

- 1) If there are more treatments than blocks in a design, then at least $(v - b)$ of the efficiencies will be unity, with the analogous result that if there are more blocks than treatments, then at least $(b - v)$ of the efficiencies of the dual will be unity.

Proof

The proof can be made a little easier by considering the matrices

$$\underline{M} = \underline{I} - \underline{A}\underline{r}^{-\delta}$$

and
$$\underline{M}_* = \underline{I} - \underline{A}_*\underline{k}^{-\delta}.$$

The eigenvalues of these matrices are then equal to one minus the corresponding eigenvalues of $\underline{A}\underline{r}^{-\delta}$ or $\underline{A}_*\underline{k}^{-\delta}$. The eigenvectors are unchanged. Now

$$\text{tr}(\underline{M}) = \text{tr}(\underline{M}_*) \tag{3.2.9}$$

using the definitions of \underline{A} and \underline{A}_* and the fact that $\text{tr}(\underline{AB}) = \text{tr}(\underline{BA})$.

Since the trace of a matrix is the sum of its eigenvalues, (3.2.9)

gives the result

$$\sum_{i=1}^v \mu_i = \sum_{i=1}^b \mu_{*i} \tag{3.2.10}$$

where $\mu_i = 1 - \varepsilon_i$

and $\mu_{*i} = 1 - \varepsilon_{*i}$,

the values ε_{*i} being the eigenvalues of $\underline{A}_* \underline{k}^{-\delta}$.

However, theorem 3.2 states that all non-unit eigenvalues of one design are eigenvalues of the other. Applying this result to the μ_i , gives all non-zero eigenvalues of M are non-zero eigenvalues of M_* . To make (3.2.10) hold therefore, requires that if $v > b$, then at least $(v - b)$ of the μ_i 's are zero, that is, ε_i 's are unity, and if $b > v$, then at least $(b - v)$ of the μ_i 's are zero, that is, ε_i 's are unity. Note, however, that some other efficiencies may also be unity.

2) The number of efficiencies equal to unity in a design equals $v - \text{rank}(\underline{n}_i)$.

Proof

The total number of non-unity efficiencies equals the number of independent columns of $(\underline{n}' \underline{r}^{-\delta} \underline{C})$, which equals $\text{rank}(\underline{n}' \underline{r}^{-\delta} \underline{C})$. Since $\underline{r}^{-\delta} \underline{C}$ is of full rank,

$$\text{rank}(\underline{n}' \underline{r}^{-\delta} \underline{C}) = \text{rank}(\underline{n}') = \text{rank}(\underline{n}) .$$

The number of efficiencies equal to unity, therefore equals the total number of efficiencies, v , minus the number of non-unity efficiencies, $\text{rank}(\underline{n})$.

3) The order of balance of a design (in the Nelder (1965) sense for a simple treatment model) is less than or equal to $\text{rank}(\underline{n})$.

Proof

The order of balance of a design, with a simple treatment model, is defined as the number of distinct efficiency factors, excluding that corresponding to the mean (which equals zero). But the total number of non-unity ε_i 's equals $\text{rank}(\underline{n})$, from 2) above, or $\text{rank}(\underline{n}) - 1$ when the efficiency factor for the mean, ε_1 , say, is excluded. The total number of distinct ε_i 's then, without ε_1 , must be less than or equal to $\text{rank}(\underline{n}) - 1$ plus one for the unit efficiencies, that is, the total is less than or equal to $\text{rank}(\underline{n})$.

4) If the number of treatments, v , equals the number of blocks, b , then the order of balance (in the sense of part 3)) of the original design will be equal to that of the dual. If $v > b$, then the order of balance of the dual will be equal to, or one less than, that of the original design, and if $b > v$, the order of balance of the dual will be equal to, or one more than, that of the original design.

Proof

This follows from the fact, proved in the theorem, that non-unit efficiencies are identical in both designs. The remaining $|b - v|$ unit efficiencies of the design having the most 'treatments' will be either distinct or not from those of the other design. The result follows.

3.3 The analysis of the dual design.

The direct analysis of any design revolves around obtaining the generalized inverse \underline{A}^- , used to obtain consistent solutions to the normal equations, and to provide a dispersion matrix for the calculation of the variances, etc., of estimable functions of the treatment parameters.

In the case of the dual, the matrix required is \underline{A}_*^{-2} , where \underline{A}_* is the analogue of \underline{A} defined for the original design. To find the generalized inverse, it is necessary to find the eigensystem of the dual. But all the eigenvectors corresponding to non-unit eigenvalues can be found from those of the original design. If \underline{B} holds the eigenvectors of the dual as its columns, then partition \underline{B} as $\left(\begin{array}{c} \underline{B}_{-n} \\ \underline{B}_u \end{array} \right)$, where, for some ordering, \underline{B}_{-n} holds those eigenvectors with non-unit efficiencies, and \underline{B}_u those with unit efficiencies. The matrix \underline{B}_{-n} is easily found since it is the non-zero columns of $\underline{n}'\underline{r}^{-\delta} \underline{C}_\mu^{-\delta/2}$, when $1/\mu_i$ is defined as zero if μ_i equals zero, since those μ_i correspond to zero columns of $\underline{n}'\underline{r}^{-\delta} \underline{C}$. It will be shown later that it is not necessary to know \underline{B}_u itself, just $\underline{B}_{-n} \underline{B}_{-n}'$.

To find $\underline{B}_u \underline{B}_u'$, it is necessary to use the scaling requirement applied to the eigenvectors, that is

$$\underline{B}' \underline{k}^{-\delta} \underline{B} = \underline{I},$$

or, equivalently since \underline{B} is non-singular

$$\underline{B} \underline{B}' = \underline{k}^{\delta}.$$

Partitioning \underline{B} , and substituting for \underline{B}_n gives

$$\left[\begin{array}{c} \underline{n}' \underline{r}^{-\delta} \underline{C}_{\mu}^{-\delta/2} \vdots \\ \vdots \\ \underline{B}_u \end{array} \right] \left[\begin{array}{c} \underline{\mu}^{-\delta/2} \underline{C}' \underline{r}^{-\delta} \underline{n} \\ \vdots \\ \underline{B}_u' \end{array} \right] = \underline{k}^{\delta}$$

which is allowable since extending \underline{B} by the zero columns of $\underline{n}' \underline{r}^{-\delta} \underline{C}^{-\delta/2}$ makes no difference to $\underline{B} \underline{B}'$.

Expanding the above equation gives

$$\underline{B}_u \underline{B}_u' = \underline{k}^{\delta} - \underline{n}' \underline{r}^{-\delta} \underline{C}_{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} \underline{n}. \quad (3.3.1)$$

But

$$\underline{r}^{-\delta} \underline{C}_{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} = \underline{r} \sum_i \left[\frac{c_i c_i'}{\mu_i} \right] \underline{r}^{-\delta}$$

Where the summation does not cover those values of μ_i which are zero, since $1/\mu_i$ was defined as zero if μ_i were zero. Therefore, by (2.6.8),

$$\underline{r}^{-\delta} \underline{C}_{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} = \underline{A}_0^{-}, \quad (3.3.2)$$

the matrix \underline{A}_0^{-} being the dispersion matrix for the inter-block analysis of the original design.

Substituting for \underline{A}_0^{-} in (3.3.1) gives

$$\underline{B}_u \underline{B}_u' = \underline{k}^{\delta} - \underline{n}' \underline{A}_0^{-} \underline{n}.$$

Now, again by analogy with the original design,

$$\underline{A}_* = \underline{B} \underline{\varepsilon}_* \underline{B}',$$

where ε_{*i} are the eigenvalues of the dual, a generalized inverse of which will be chosen as

$$\underline{A}_*^{-} = \underline{k}^{-\delta} \underline{B} \underline{\varepsilon}_*^{-\delta} \underline{B}' \underline{k}^{-\delta},$$

where $1/\varepsilon_i$ is taken to be zero if ε_i is zero.

Partitioning \underline{B} as $\begin{bmatrix} \underline{B}_n & \vdots & \underline{B}_u \end{bmatrix}$ again, and $\underline{\varepsilon}_*^{-\delta}$ as $\begin{bmatrix} \underline{\varepsilon}_{*n}^{-\delta} \\ \vdots \\ \underline{I} \end{bmatrix}$,

gives
$$\underline{A}_*^{-} = \underline{k}^{-\delta} \underline{B}_n \underline{\varepsilon}_{*n}^{-\delta} \underline{B}_n' \underline{k}^{-\delta} + \underline{k}^{-\delta} \underline{B}_u \underline{B}_u' \underline{k}^{-\delta} .$$

It will now be seen that \underline{B}_u by itself is indeed not required, just $\underline{B}_u \underline{B}_u'$. Substituting directly $\underline{n}' \underline{r}^{-\delta} \underline{C}_\mu^{-\delta/2}$ for \underline{B}_n cannot be done since in general it would lead to inconsistencies in dimensions involving $\underline{\varepsilon}_{*n}^{-\delta}$. However, augmenting \underline{B}_n by zero columns corresponding to the zero columns in $\underline{n}' \underline{r}^{-\delta} \underline{C}_\mu^{-\delta/2}$, and correspondingly adding extra unit diagonal elements in $\underline{\varepsilon}_{*n}^{-\delta}$ enables not only the substitution for \underline{B}_n to be made but also enables the augmented $\underline{\varepsilon}_{*n}^{-\delta}$ to be replaced by $\underline{\varepsilon}^{-\delta}$, since by the theorem, non-unit ε_i 's have equal counterparts in the dual, and unit ε_i 's produce corresponding zero columns in $\underline{n}' \underline{r}^{-\delta} \underline{C}_\mu^{-\delta/2}$.

The matrix \underline{A}_*^{-} can now be written as

$$\begin{aligned} \underline{A}_*^{-} &= \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \underline{C}_\mu^{-\delta/2} \underline{\varepsilon}^{-\delta} \underline{\mu}^{-\delta/2} \underline{C}' \underline{r}^{-\delta} \underline{n} \underline{k}^{-\delta} + \underline{k}^{-\delta} - \underline{k}^{-\delta} \underline{n}' \underline{A}_0^{-} \underline{n} \underline{k}^{-\delta} \\ &= \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \underline{C}_\mu^{-\delta} (\underline{C}' \underline{r}^{-\delta} \underline{C}) \underline{\varepsilon}^{-\delta} \underline{C}' \underline{r}^{-\delta} \underline{n} \underline{k}^{-\delta} + \underline{k}^{-\delta} - \underline{k}^{-\delta} \underline{n}' \underline{A}_0^{-} \underline{n} \underline{k}^{-\delta} \end{aligned} \tag{3.3.3}$$

using $\underline{C}' \underline{r}^{-\delta} \underline{C} = \underline{I}$.

Now
$$\underline{r}^{-\delta} \underline{C} \underline{\varepsilon}^{-\delta} \underline{C}' \underline{r}^{-\delta} = r^{-\delta} \sum_i \left(\frac{C_i C_i'}{\varepsilon_i} \right) \underline{r}^{-\delta}$$

where the summation does not cover values of i for which ε_i equals zero.

Therefore, from (2.5.0),

$$\underline{r}^{-\delta} \underline{C} \underline{\varepsilon}^{-\delta} \underline{C}' \underline{r}^{-\delta} = \underline{A}^{-} . \tag{3.3.4}$$

Substituting this and (3.3.2) into (3.3.3) gives

$$\begin{aligned} \underline{A}_*^{-} &= \underline{k}^{-\delta} \underline{n}' \underline{A}_0^{-} \underline{r}^{\delta} \underline{A}^{-} \underline{n} \underline{k}^{-\delta} + \underline{k}^{-\delta} - \underline{k}^{-\delta} \underline{n}' \underline{A}_0^{-} \underline{n} \underline{k}^{-\delta} \\ &= \underline{k}^{-\delta} + \underline{k}^{-\delta} \underline{n}' \underline{A}_0^{-} (\underline{r}^{\delta} \underline{A}^{-} - \underline{I}) \underline{n} \underline{k}^{-\delta} \end{aligned} \tag{3.3.5}$$

If a design is connected, and therefore has no zero-efficiencies apart from that corresponding to the mean, ε_1 , then Tocher's $\underline{\Omega}^{-1}$ matrix (1952) instead of \underline{A} , may be used, which effectively changes the efficiency ε_1 to unity. The matrix $\underline{\Omega}^{-1}$ can then be inverted directly, giving $\underline{\Omega}$,

which can then be used for analysing the experiment in the same way as was \underline{A}^- . Now, if the original design was connected, then so will be the dual since in the definition of connectedness, blocks and treatments are symmetrical. The analogue of $\underline{\Omega}$ then, $\underline{\Omega}_*$, say, may be used for the analysis of the dual if $\underline{\Omega}$ was used in the original design. In this case it appears that

$$\underline{\Omega}_* = \underline{k}^{-\delta} + \underline{k}^{-\delta} \underline{n}' \underline{A}_0^- (\underline{r}^\delta \underline{\Omega} - \underline{I}) \underline{n} \underline{k}^{-\delta}.$$

This result then enables the analysis of the dual to proceed directly from that of the inter-block analysis of the original design.

If the inter-block analysis of the dual is required, then the analogous matrix to \underline{A}_0^- is needed, \underline{A}_{*0}^- say. The expression for \underline{A}_{*0}^- , in terms of the eigensystem is

$$\underline{A}_{*0}^- = \underline{B} \underline{\mu}_*^{-\delta} \underline{B}' , \quad \text{where } \mu_{*i} = 1 - \epsilon_{*i}$$

by analogy with that for the original design.

A generalized inverse will be chosen as

$$\underline{A}_{*0}^- = \underline{k}^{-\delta} \underline{B} \underline{\mu}_*^{-\delta} \underline{B}' \underline{k}^{-\delta} \tag{3.3.6}$$

where $1/\mu_{*i}$ is taken to be zero if μ_{*i} is zero.

Partitioning \underline{B} as $\left[\begin{array}{c} \underline{B}_n \\ \underline{B}_u \end{array} \right]$ again, and correspondingly $\underline{\mu}_*^{-\delta}$ as $\left[\begin{array}{c|c} \mu_{*n}^{-\delta} & \underline{0} \\ \hline \underline{0} & \underline{0} \end{array} \right]$, since unit ϵ_{*i} 's correspond to zero μ_{*i} 's, enables

(3.3.6) to be written as

$$\underline{A}_{*0}^- = \underline{k}^{-\delta} \underline{B} \begin{array}{c} \mu_{*n}^{-\delta} \\ \underline{0} \end{array} \underline{B}' \underline{k}^{-\delta} . \tag{3.3.7}$$

As before, substituting directly $\underline{n}' \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta/2}$ for \underline{B}_n cannot be done because of the possible inconsistencies with the dimensions involving $\mu_{*n}^{-\delta}$. It is required then, to augment \underline{B}_n by zero columns corresponding to the zero columns in $\underline{n}' \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta/2}$, and correspondingly to add additional zero diagonal elements to $\mu_{*n}^{-\delta}$. This then causes the augmented $\mu_{*n}^{-\delta}$ to be equal to $\underline{\mu}^{-\delta}$, since zero columns of $\underline{n}' \underline{r}^{-\delta} \underline{C}$

corresponded to zero μ_i 's in the original design, and $1/\mu_i = 0$ if $\mu_i = 0$.

Equation (3.3.7) can now be written as

$$\underline{A}_{*0}^{-} = \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta/2} \underline{\mu}^{-\delta/2} \underline{C}' \underline{r}^{-\delta} \underline{n} \underline{k}^{-\delta}$$

which equals $\underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta} (\underline{C}' \underline{r}^{-\delta} \underline{C}) \underline{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} \underline{n} \underline{k}^{-\delta}$ (3.3.8)

because of (2.3.1).

Substituting (3.3.2) into (3.3.8) gives the required formula for \underline{A}_{*0}^{-} as

$$\underline{A}_{*0}^{-} = \underline{k}^{-\delta} \underline{n}' \underline{A}_0^{-} \underline{r} \underline{A}_0^{-} \underline{n} \underline{k}^{-\delta}. \quad (3.3.9)$$

If the method of analysis given by chapter 2 had been carried out for the original design, that is, the \underline{L}_i matrices had been found, then equations (3.3.5) and (3.3.9) can both be simplified as follows:

For the Intra-block \underline{A}_*^{-} , consider the expansion of $\underline{A}_0^{-} (\underline{r} \underline{A}^{-} - \underline{I})$.

Now

$$\underline{A}_0^{-} = \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} \quad \text{from (3.3.2)}$$

and $\underline{A}^{-} = \underline{r}^{-\delta} \underline{C} \underline{\epsilon}^{-\delta} \underline{C}' \underline{r}^{-\delta}$ from (3.3.4)

where in both cases the reciprocal of zero is taken as zero.

The expression $\underline{A}_0^{-} \underline{r} \underline{A}^{-} - \underline{A}_0^{-}$ is then expanded as

$$\underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} \underline{C} \underline{\epsilon}^{-\delta} \underline{C}' \underline{r}^{-\delta} - \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta}$$

which, since $\underline{C}' \underline{r}^{-\delta} \underline{C} = \underline{I}$, equals

$$\begin{aligned} & \underline{r}^{-\delta} \underline{C} (\underline{\mu}^{-\delta} \underline{\epsilon}^{-\delta} - \underline{\mu}^{-\delta}) \underline{C}' \underline{r}^{-\delta} \\ & = \underline{r}^{-\delta} \sum_i \left[c_i c_i' \left(\frac{1}{\mu_i \epsilon_i} - \frac{1}{\mu_i} \right) \right] \underline{r}^{-\delta} \end{aligned} \quad (3.3.10)$$

where reciprocals of zero are taken as zero.

Expansion of the term in brackets gives

$$\frac{1}{\mu_i (1 - \mu_i)} - \frac{1}{\mu_i}$$

which equals

$$\frac{1}{1 - \mu_i} \text{ or } \frac{1}{\epsilon_i}$$

for the case when μ_i and ε_i are both non-zero.

If $\mu_i = 0$ both terms are zero while if $\varepsilon_i = 0$ the first term is ignored while the second is not.

Equation (3.3.10) therefore may be written as

$$\underline{A}_0^- \underline{r}^\delta \underline{A}^- - \underline{A}_0^- = \underline{r}^{-\delta} \sum_i \left[\frac{c_i c_i'}{\varepsilon_i} \right] \underline{r}^{-\delta} - \underline{r}^{-\delta} \sum_j (c_j c_j') \underline{r}^{-\delta} \quad (3.3.11)$$

where the i summation is over those eigenvectors for which ε_i is not zero or unity, while the j summation is over those eigenvectors having ε_i zero.

Now, the matrices \underline{L}_i are defined as

$$\underline{L}_i = \sum_k c_k c_k' \underline{r}^{-\delta}$$

the summation being over those contrasts having efficiency ε_i .

Substituting the \underline{L}_i matrices in (3.3.11) gives

$$\underline{A}_0^- \underline{r}^\delta \underline{A}^- - \underline{A}_0^- = \underline{r}^{-\delta} \sum_i \frac{\underline{L}_i}{\varepsilon_i} - \underline{r}^{-\delta} \underline{L}_1$$

where \underline{L}_1 corresponds to zero efficiencies, and the i summation is as above, in (3.3.11).

Substituting into equation (3.3.5) gives

$$\underline{A}_*^- = \underline{k}^{-\delta} + \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \left[\sum_i \frac{\underline{L}_i}{\varepsilon_i} - \underline{L}_1 \right] \underline{nk}^{-\delta}$$

or
$$\underline{A}_*^- = \underline{k}^{-\delta} + \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \sum_i \frac{\underline{L}_i \underline{nk}}{\varepsilon_i}$$

with ε_i' 's such that $1/\varepsilon_i'$ is zero if $\varepsilon_i = 1$, -1 if $\varepsilon_i = 0$, and normal otherwise.

For the interblock \underline{A}_{*0}^- , consider the expansion of $\underline{A}_0^- \underline{r}^\delta \underline{A}_0^-$.

Substituting for \underline{A}_0^- (from (3.3.2)) gives

$$\begin{aligned} \underline{A}_0^- \underline{r}^\delta \underline{A}_0^- &= \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-\delta} \underline{C}' \underline{r}^{-\delta} \text{ with } 1/\mu_i = 0 \text{ if } \mu_i = 0. \\ &= \underline{r}^{-\delta} \underline{C} \underline{\mu}^{-2\delta} \underline{C}' \underline{r}^{-\delta} \end{aligned}$$

since $\underline{C}' \underline{r}^{-\delta} \underline{C} = \underline{I}$, and so

$$\underline{A}_0^{-1} \underline{r}^\delta \underline{A}_0^{-1} = \underline{r}^{-\delta} \sum_i \frac{c_i c_i'}{\mu_i^2} \underline{r}^{-\delta}$$

the summation being over those values of i for which μ_i is non-zero.

Substituting the \underline{L}_i matrices gives

$$\underline{A}_0^{-1} \underline{r}^\delta \underline{A}_0^{-1} = \underline{r}^{-\delta} \sum_i \frac{\underline{L}_i}{\mu_i^2}$$

and therefore

$$\underline{A}_{*0}^{-1} = \underline{k}^{-\delta} \underline{n}' \underline{r}^{-\delta} \left(\sum_i \frac{\underline{L}_i}{\mu_i^2} \right) \underline{n} \underline{k}^{-\delta}$$

where the summation does not cover zero μ_i 's.

Since both \underline{A}_*^{-1} and \underline{A}_{*0}^{-1} are both now known, the intra and inter block analyses can proceed without difficulty.

3.4 Relationships between the mean efficiencies of a design and those of its dual.

The efficiency factors (Jones, 1959) of a design give an indication of how good a design is compared to an orthogonal design with the same treatment replications. If there are many distinct efficiency factors the problem arises of trying to find some sort of 'mean' efficiency with which to judge the design, and also, so as to be able to compare the design with another, different one. When the dual is required it is useful to know the mean efficiency of the dual, given that of the original design, and to know whether or not it has a higher or lower 'mean' efficiency.

Pearce (1968, 1970) defined the mean efficiency, ε , for adjusted treatment means as

$$\varepsilon = \frac{v}{\text{tr}(\underline{r}^\delta \underline{\Omega})} \tag{3.4.1}$$

where v is the number of treatments, and $\underline{\Omega}$ is the matrix used by Tocher (1952) for connected designs. This is the harmonic mean of the

values Φ_i/θ_i , where

$\Phi_i\sigma^2$ is the variance of the i th mean, that is, Ω_{ii} and

$\theta_i\sigma^2$ is the variance of the i th mean if an orthogonal design had been possible, that is $1/r_i$.

Now $\underline{\Omega}^{-1}$ is equal to $\left(\underline{A} + \frac{rr'}{N}\right)$ and so it follows that $\underline{\Omega}^{-1}\underline{r}^{-\delta}$ has the same eigensystem as $\underline{A}\underline{r}^{-\delta}$ except for the eigenvalue ϵ_1 corresponding to the eigenvector \underline{r}/\sqrt{N} . This has the value 1 for $\underline{\Omega}^{-1}\underline{r}^{-\delta}$ as opposed to 0 for $\underline{A}\underline{r}^{-\delta}$. For a connected design then, $\underline{\Omega}^{-1}\underline{r}^{-\delta}$ has no zero eigenvalues and the matrix $\underline{r}\underline{\Omega}$, therefore has eigenvalues $1/\epsilon_i$, where $\epsilon_1 = 1$.

Equation (3.4.1), then, can be written

$$\mathcal{E} = \frac{v}{\sum \frac{1}{\epsilon_i}} \quad (3.4.2)$$

that is, the harmonic of the efficiency factors, where ϵ_1 has been taken as unity.

If the dual is considered, then by analogy with the previous argument, the mean efficiency for adjusted 'treatment' means of the dual, ϵ_* say, is

$$\mathcal{E}_* = \frac{b}{\sum \frac{1}{\epsilon_{*i}}} \quad (3.4.3)$$

where ϵ_{*i} are the efficiency factors of the dual, again with ϵ_* taken as unity.

However, theorem 3.2 and its first corollary proved that non-unit efficiencies of one design had equal counterparts in the other, and any excess efficiency factors, from the larger design, were unity.

The following results therefore follow:

$$\prod_{i=1}^v \epsilon_i = \prod_{i=1}^b \epsilon_{*i} \quad (3.4.4)$$

$$\sum_{i=1}^v \epsilon_i + (b-v) = \sum_{i=1}^b \epsilon_{*i} \quad (3.4.5)$$

and
$$\sum_{i=1}^v \frac{1}{\xi_i} + (b - v) = \sum_{i=1}^b \frac{1}{\xi_{*i}} \quad (3.4.6)$$

Substituting (3.4.6) into (3.4.3) gives

$$\xi_* = \frac{b}{\sum_{i=1}^v \frac{1}{\xi_i} + (b - v)}$$

which, using (3.4.2) can be written as

$$\xi_* = \frac{b}{\frac{v}{\xi} + b - v}$$

that is,
$$\xi_* = \frac{b}{v + (b - v)\xi}$$

in agreement with Pearce's (1970) result. In fact, Pearce essentially derived result (3.4.6), but in a different form. The E used in his

proof is $\sum_{i=1}^v \frac{1}{\xi_i} - v$, or equivalently, $\sum_{i=1}^b \frac{1}{\xi_{*i}} - b$.

However, results (3.4.4) and (3.4.5) enable the relationship between the mean efficiencies to be calculated under different definitions of the term 'mean efficiency'.

Suppose mean efficiency is now defined as the arithmetic mean of the efficiency factors, that is

$$\xi = \frac{\sum \xi_i}{v} \quad (3.4.7)$$

This is meaningful since

$$\sum \xi_i = \text{tr}(\underline{Q}^{-1} \underline{r}^{-S}),$$

that is, it is now not the ratios of the diagonal elements of the dispersion matrix of the design, with those of the dispersion matrix of the orthogonal design that are being used, as before, but the ratios of the diagonal elements of the information matrices of the design and the orthogonal design.

The corresponding formula for the dual is

$$\xi_* = \frac{\sum \xi_{*i}}{b} \quad (3.4.8)$$

Using (3.4.5),

$$\varepsilon_* = \frac{\sum \varepsilon_i + (b - v)}{b}$$

and substituting (3.4.7) gives

$$\varepsilon_* = \frac{v\varepsilon + b - v}{b}$$

A possible third definition of mean efficiency is the geometric mean of the efficiency factors, that is

$$\varepsilon = (\prod \varepsilon_i)^{\frac{1}{v}} \quad (3.4.9)$$

This measure, too is meaningful since, because the determinant of a matrix is the product of its eigenvalues,

$$\begin{aligned} \varepsilon &= \left| \Omega^{-1} r^{-\delta} \right|^{\frac{1}{v}} \\ &= \left(\frac{|\Omega^{-1}|}{|r^{\delta}|} \right)^{\frac{1}{v}} \end{aligned}$$

that is, it involves the ratio of the determinant of the information matrix of the design to that of the orthogonal design. It will be seen that optimality procedures which seek to maximise the determinant of the information matrix are maximising the mean efficiency as defined in (3.4.9).

The corresponding formula for the dual is

$$\varepsilon_* = \left[\prod \varepsilon_{*i} \right]^{\frac{1}{b}}, \quad (3.4.10)$$

and substituting (3.4.4) into (3.4.10) gives

$$\varepsilon_* = \left[\prod \varepsilon_i \right]^{\frac{1}{b}}$$

which reduces, using (3.4.9), to

$$\varepsilon_* = (\varepsilon)^{\frac{v}{b}}.$$

Note that, in all three definitions of mean efficiency given, if $b > v$, then $\varepsilon_* > \varepsilon$ (since ε must be less than or equal to one). Therefore, as Pearce (1970) pointed out for the harmonic mean, the design with lower actual replication for the group, blocks or treatments, will have the higher mean efficiency.

3.5 Discussion

The theorem giving the relationship between the efficiency factors of a design and those of its dual can be seen to have an important consequence, the analysis of a dual design can now follow on directly from that of the original design with the minimum of computation. The theorem also proves useful for finding relations between measures of efficiency for adjusted treatment means from each design. However, Pearce (1968, 1970) also defined a measure for the mean efficiency with respect to differences of means. This is then defined as the harmonic mean of the (new) values Φ_{ij}/Θ_{ij} where

$$\begin{aligned}\Theta_{ij} &= \text{variance of difference between treatments } i \text{ and } j \\ &= \Omega_{ii} + \Omega_{jj} - 2\Omega_{ij}\end{aligned}$$

and Φ_{ij} = variance of difference between treatments i and j for an orthogonal design

$$= \frac{1}{r_i} + \frac{1}{r_j}$$

It can then be shown that in the equi-replicate case, this definition is equivalent to the harmonic mean of the efficiency factors excluding the efficiency corresponding to the mean (ξ_1 in section 3.4), which is a reasonable definition. Pearce's relations between the mean efficiency for a design and that for its dual follow readily from the application of (3.4.6).

However, in the non-equi-replicate case, the definition does not reduce to being equal to the harmonic mean of the efficiencies excluding ξ_1 as above. Consequently it has not been possible to prove or disprove Pearce's (incorrectly proven) result concerning the relation between the mean efficiency with respect to differences among treatments and the mean efficiency with respect to adjusted treatment means. It does not seem clear at the moment whether or not Pearce's definition is in some way 'unnatural' or whether its relationship to efficiencies has not been understood.

Chapter 4 The construction of optimal block designs when contrasts of interest have been given.

4.1 Introduction

When designing an experiment a statistician has available a number of criteria for choosing between possible designs (for example, that of mean efficiency, mentioned in the previous chapter). Also available, but which, because of computational difficulties (to be mentioned later), tend not to be used, are methods for constructing designs that are optimal with respect to some criterion.

Apart from the computational difficulties, many such procedures are ill-suited to the design of block experiments for two reasons:

a) Much of the work in the field has been concerned with 'continuous' designs, that is, where experimental conditions can take any number of different values, as opposed to discrete designs, where only certain possible values can be taken (for example, replication of a treatment must be integral). Most of the procedures therefore decide what proportions of the total number of observations should be allowed at each possible experimental point. When the total number of observations is specified this in general leads to non-integral replication at each point. If the total number is not specified, it may require that the total number be large so as to either obtain optimal integral replications or to have integral replication that closely approximate the optimal ones.

b) In the block design case, the procedures attach equal importance to all possible contrasts between treatments and, as Pearce (1974) pointed out, they will then give similar status to all treatments and differences of treatments. That is, the 'optimal design' they therefore evaluate is either a randomised block design, a totally balanced design, or one in which treatments occur either p or $p + 1$ times per block, (see Box 1968, Wynn 1972). In practice not only may the experimenter not be interested at all in certain contrasts, but could well ascribe different levels

of importance to those contrasts that he is interested in. What is required then, is to use an optimality criterion which will take into account the contrasts of interest to the experimenter and the weighting attached to each. It is necessary to find a method which will construct the corresponding optimal discrete design for a specified number of observations.

As was mentioned before, the computational difficulties of such procedures are sometimes prohibitive, since a search through the space of the independent variables is usually entailed to find the point which maximises some function (see, for example, Federov, 1972, p 164), that point then being added to the design, replacing the present 'worst' point. Box and Draper (1971) used a 'direct search' routine for the construction of D-optimal designs, but reported satisfactory results for small designs only (less than 30 co-ordinate values).

Because of this, the approach presented by Dykstra (1971) will be used. Dykstra advocates using a given set of candidates for the sequential construction of a D-optimal design, choosing that point from the candidates which results in the largest possible increase in the appropriate function. Mitchell (1974) lists the following advantages of using candidate points:

- a) The search procedure can be programmed very easily,
- b) It is easy to exclude points that are not experimentally feasible or desirable,
- c) Variables can be either qualitative or quantitative
- and d) The number of different levels for each factor can be kept low.

Particularly because of points (a) and (c), it was decided to use the candidate method when constructing designs optimal for criteria other than D-optimality. *as well*

Three possible criteria will be examined; two based on the well known D-optimality and A-optimality criteria and one put forward by

Pearce (1974). In each case an iterative procedure will be derived with which optimal designs may be constructed.

4.2 A Criterion Based upon D-optimality

D-optimality is a very common criterion for use in constructing optimal designs, and one of the most popular for comparing designs, having its origin in a paper by Smith (1918).

Suppose the model used is

$$y = X\beta + \varepsilon ,$$

where β is the vector of parameters and ε independently and identically distributed random variables with zero mean and variance σ^2 .

The D-optimal design is then defined as that design having the minimum $|(X'X)^{-1}|$ or, equivalently, the maximum $|X'X|$. Box and Lucas (1959) gave a sampling theory justification of the criterion, and also showed that the D-optimal design has the smallest confidence region for the estimates of the parameters. Keifer (1961) showed that a D-optimal design is also a minimax design, that is, it minimises the maximum variance of a predicted value. The above justifications, however, only apply in the continuous case.

A disadvantage, mentioned by Box and Draper (1971), is that D-optimality is a 'variance criterion', and effectively assumes that the chosen model is the true underlying one. However, because of the work of Stigler (1970), they deduce that D-optimality is not unrealistic if the design is restricted to a region of interest.

Now assume that contrasts of interest have been given and are held columnwise in the matrix V , then the generalized D-optimality criterion will be, for a specific number of observations, to choose X so as to minimise $|V'DV|$ where $D = (X'X)^{-1}$ (Federov, 1972, p53). The matrix $V'DV$ is the variance - covariance matrix of the contrasts between the parameters rather than the parameters themselves.

An important disadvantage of this criterion, compared to those

presented later, is the fact that if weights are attached to the contrasts, then this has no effect on the choice of the design. This can be seen by supposing weights \underline{w} are given, that is \underline{Vw}^{δ} is to be used instead of \underline{V} , and it is therefore required to minimise $|\underline{w}^{\delta} \underline{V}' \underline{DVw}^{\delta}|$. But, since \underline{w}^{δ} is non-singular,

$$|\underline{w}^{\delta} \underline{V}' \underline{DVw}^{\delta}| = |\underline{w}^{2\delta}| |\underline{V}' \underline{DV}| = K |\underline{V}' \underline{DV}|$$

where K is a constant over all possible choices of design, therefore it is still required to minimise $|\underline{V}' \underline{DV}|$.

It should be mentioned that in certain cases it may be simpler to use a restricted model rather than use contrasts applied to the full one, for example, if only first and second order interactions in a factorial experiment were required.

4.3 Iterative method of constructing generalized D-optimal designs

The basic procedure is essentially that advocated by Dykstra (1971), and extended by Mitchell (1974), that is

- a) Scan the candidates for the point which, when added to the design, results in the maximum possible decrease in the determinant, then
- b) Remove from the design, that point which effects the minimum possible increase.

(The above steps could equally well be reversed.)

It will readily be seen that this procedure must result in a decrease of at least zero, since the 'worst' that can happen must be to subtract the point (or an equivalent point) that had just been added. These steps are then repeated until no further improvement can be made.

A formula for deriving the appropriate point to add or subtract from the design can be found as follows:

It will be seen that rows of the matrix \underline{X} correspond to the experimental points of the design, the actual form of the row being determined by the model used. Adding a point in the design therefore corresponds to

adding an extra row, \underline{x}' , say, and the new $(\underline{X}'\underline{X})$ can be found from the old $(\underline{X}'\underline{X})$ by simply adding \underline{xx}' , that is,

$$(\underline{X}'\underline{X})_{\text{new}} = (\underline{X}'\underline{X}) + \underline{xx}'.$$

By simple matrix algebra, the inverse of $(\underline{X}'\underline{X})_{\text{new}}$ can be found as

$$\underline{D}_{\text{new}} = (\underline{I} + \underline{Dxx}')^{-1} \underline{D} \quad , \quad (4.3.1)$$

where

$$\underline{D}_{\text{new}} = \{(\underline{X}'\underline{X})_{\text{new}}\}^{-1}$$

and

$$\underline{D} = (\underline{X}'\underline{X})^{-1} \quad .$$

Using the identity

$$(\underline{I} + \underline{AB})^{-1} = \underline{I} - \underline{A}(\underline{I} + \underline{BA})^{-1}\underline{B}$$

in equation (4.3.1), with \underline{A} as \underline{Dx} and \underline{B} as \underline{x} gives

$$\underline{D}_{\text{new}} = (\underline{I} - \underline{Dx}(\underline{I} + \underline{x}'\underline{Dx})^{-1}\underline{x}')\underline{D} \quad .$$

But $\underline{x}'\underline{Dx}$ is a scalar, therefore, writing $v(\underline{x})$ as $\underline{x}'\underline{Dx}$ as in Mitchell (1973), gives

$$\underline{D}_{\text{new}} = \underline{D} - \frac{\underline{Dxx}'\underline{D}}{1 + v(\underline{x})} \quad . \quad (4.3.2)$$

Since $\hat{\underline{y}} = \underline{X}'\hat{\underline{\beta}}$ and $D(\hat{\underline{\beta}}) = \underline{D}\hat{\sigma}^2$, then

$$\text{var}(y_i) = \text{var}(x_i'\hat{\underline{\beta}}) = x_i'\underline{Dx}_i\hat{\sigma}^2 \quad ,$$

and so $v(\underline{x})$ is the variance of the observation \hat{y} corresponding to the point giving rise to the row \underline{x}' .

Post and premultiplying (4.3.2) by \underline{V} and \underline{V}' respectively gives

$$\underline{V}'\underline{D}_{\text{new}}\underline{V} = \underline{V}'\underline{DV} - \frac{\underline{V}'\underline{Dxx}'\underline{DV}}{1 + v(\underline{x})} \quad . \quad (4.3.3)$$

Substituting the identity

$$|\underline{A} - \underline{CB}| = |\underline{I} - \underline{BA}^{-1}\underline{C}| |\underline{A}|$$

in equation (4.3.3) with \underline{A} as $\underline{V}'\underline{DV}$, \underline{C} as $\underline{V}'\underline{Dx}/\sqrt{(1 + v(\underline{x}))}$, and \underline{B} as \underline{C}'

gives

$$\begin{aligned} |\underline{V}'\underline{D}_{\text{new}}\underline{V}| &= \left(1 - \frac{\underline{x}'\underline{D}\underline{V}(\underline{V}'\underline{D}\underline{V})^{-1}\underline{V}'\underline{D}\underline{x}}{1 + v(\underline{x})}\right) |\underline{V}'\underline{D}\underline{V}| \\ &= \left(1 - \frac{\underline{q}'\underline{R}\underline{q}_x}{1 + v(\underline{x})}\right) |\underline{V}'\underline{D}\underline{V}| \end{aligned} \quad (4.3.4)$$

where $\underline{R} = (\underline{V}'\underline{D}\underline{V})^{-1}$

and $\underline{q}_x = \underline{V}'\underline{D}\underline{x}$.

If \underline{V} is replaced by \underline{I} , then (4.3.4) reduces to the same formula as that used by Dykstra (1971) for sequential designs.

A similar formula can be found for the case of removing or subtracting a point. For this case

$$(\underline{X}'\underline{X})_{\text{new}} = (\underline{X}'\underline{X}) - \underline{x}\underline{x}'$$

Proceeding in a similar manner as before,

$$\underline{D}_{\text{new}} = \frac{\underline{D}\underline{x}\underline{x}'\underline{D}}{1 - v(\underline{x})} \quad (4.3.5)$$

$$\text{and } |\underline{V}'\underline{D}_{\text{new}}\underline{V}| = \left(1 + \frac{\underline{q}'\underline{R}\underline{q}_x}{1 - v(\underline{x})}\right) |\underline{V}'\underline{D}\underline{V}| \quad (4.3.6)$$

with \underline{R} and \underline{q}_x as before.

Since the object is to reduce $|\underline{V}'\underline{D}_{\text{new}}\underline{V}|$ by as much as possible when adding a point and to increase by as little as possible when subtracting a point, in both cases the point chosen must be that which minimises the expression

$$1 - \frac{s(\underline{q}'\underline{R}\underline{q}_x)}{1 + sv(\underline{x})}$$

where

$$s = \begin{cases} +1 & \text{when adding a point} \\ -1 & \text{when subtracting a point.} \end{cases}$$

Having chosen a point, a reasonably simple way of updating \underline{R} can be found by noting the identity obtained from equation (4.3.2), that is

$$(\underline{A} + \underline{z}\underline{z}')^{-1} = \underline{A}^{-1} - \frac{\underline{A}^{-1}\underline{z}\underline{z}'\underline{A}^{-1}}{1 + \underline{z}'\underline{A}^{-1}\underline{z}} \quad (4.3.7)$$

Since the new \underline{R} after adding a point can be found as the inverse of the L.H.S. of equation (4.3.3), substituting \underline{R}^{-1} for \underline{A} , and \underline{z} as $\underline{q}_x / \sqrt{(1 + v(\underline{x}))}$ for \underline{z} , gives

$$\underline{R}_{\text{new}} = \frac{\underline{R} + \frac{\underline{R}\underline{q}_x \underline{q}_x' \underline{R}}{1 + mv(\underline{x})}}$$

for adding a point where m is the multiplier used for updating the determinant $|\underline{V}'\underline{DV}|$, that is

$$m = 1 - \frac{\underline{q}_x' \underline{R} \underline{q}_x}{1 + v(\underline{x})}$$

The new \underline{R} can also be found after subtracting a point by taking \underline{A} as \underline{R}^{-1} and \underline{z} as $\underline{q}_x / \sqrt{(1 - v(\underline{x}))}$ in equation (4.3.7), giving

$$\underline{R}_{\text{new}} = \underline{R} - \frac{\underline{R}\underline{q}_x \underline{q}_x' \underline{R}}{1 - mv(\underline{x})}$$

for subtracting a point where m is again the multiplier used for updating $|\underline{V}'\underline{DV}|$, that is

$$m = 1 + \frac{\underline{q}_x' \underline{R} \underline{q}_x}{1 - v(\underline{x})}$$

The matrix \underline{D} is easily updated using formula (4.3.2) or (4.3.5) according to whether adding or subtracting.

It should be noticed that \underline{R} always exists if \underline{D} does since if \underline{D} is positive definite and \underline{V} is of full rank (that is the chosen contrasts are not linearly dependent), then $\underline{V}'\underline{DV}$ is positive definite.

4.4 Excursions

Since in discrete designs, the probability of a procedure converging to a locally optimum design instead of the global one is found to be quite high, Mitchell (1974) advocates the use of 'excursions' to improve the chances of not being trapped at a local optimum. The restriction that after adding a point to the design, the procedure must subtract a point is relaxed and an 'excursion' is permitted where designs

of various sizes are constructed, each time adding or subtracting one point, eventually returning to the original size, n say, The rule is this: if it is known that returning to n points would produce no improvement, then another point is added, if not, a point is subtracted. The determinants of all those designs, of size greater than n , constructed on an excursion which led to no improvement in the required n -point design are remembered and it is these, the failure designs, which the procedure checks against when deciding to add or subtract a point. An exactly analogous procedure exists when starting off by subtracting rather than adding a point; in this case the excursion is always over designs with less than n points.

When using contrasts, the determinants checked will be $|\underline{V}'\underline{DV}|$ rather than $|\underline{X}'\underline{X}|$. Each time an excursion leads to an improvement in the determinant, the set of failure designs is cleared and the procedure starts anew from the new 'best' design.

Mitchell advocates setting limits at plus or minus six points away from n for the excursions, else the excursion wastes too much time and rarely results in an improvement. Limits less than these would not give enough freedom for the excursions.

A disadvantage of the excursions is that singularity can result if too few plots are being used (if n is less than the number of parameters plus six). To stop this, Mitchell considers $(\underline{X}'\underline{X} + \alpha\underline{X}'_0\underline{X}_0/N_0)$ instead of $\underline{X}'\underline{X}$, i.e. $\underline{M}(\alpha)$ instead of $\underline{M}(0)$ say, where α is a small positive number, \underline{X}_0 the design matrix with all of the specified candidates present, and N_0 the total number of candidates. He then minimises $|\underline{M}(\alpha)|$ instead of $|\underline{X}'\underline{X}|$. When $\underline{X}'\underline{X}$ is singular, $\underline{M}(\alpha)^{-1}$ is an approximation to a generalized inverse of $\underline{X}'\underline{X}$ as defined by Chernoff (1953).

There is obviously an error involved in doing this and Mitchell shows that the proportional error in considering $\underline{M}(\alpha)$ instead of $\underline{X}'\underline{X}$ is approximately \bar{v} , where \bar{v} is the average variance of \hat{y}_i taken over

all candidates. The quantity \bar{v} is usually less than one, and indeed, Federov (1972; p.71) shows that in the continuous case, D-optimality minimises the maximum variance of $\hat{y}_i, v(\underline{x}_i)$.

When using contrasts a similar procedure can be followed, minimising $|\underline{V}'\underline{D}(\alpha)\underline{V}|$ instead of $|\underline{V}'\underline{D}\underline{V}|$, where

$$\underline{D}(\alpha) = (\underline{M}(\alpha))^{-1}$$

The matrix $\underline{D}(0)$ is then \underline{D} when $(\underline{X}'\underline{X})$ is not singular.

As before, an error is involved and it can be approximated the following way:

Expand $\log |\underline{V}'\underline{D}(\alpha)\underline{V}|$ about $\alpha = 0$ in a Taylor series, that is

$$\log |\underline{V}'\underline{D}(\alpha)\underline{V}| \sim \log |\underline{V}'\underline{D}\underline{V}| + \alpha \left(\frac{\partial \log |\underline{V}'\underline{D}(\alpha)\underline{V}|}{\partial \alpha} \right) \Big|_{\alpha=0} \quad (4.4.1)$$

$$\text{Now } \frac{\partial (\log |\underline{Y}|)}{\partial \alpha} = \text{tr} \left(\underline{Y}^{-1} \frac{\partial \underline{Y}'}{\partial \alpha} \right), \quad (\text{Graybill (1969), p.266}) \quad (4.4.2)$$

therefore

$$\frac{\partial \log |\underline{V}'\underline{D}(\alpha)\underline{V}|}{\partial \alpha} = \text{tr} \left[(\underline{V}'\underline{D}(\alpha)\underline{V})^{-1} \frac{\partial (\underline{V}'\underline{D}(\alpha)\underline{V})}{\partial \alpha} \right]. \quad (4.4.3)$$

Also,

$$\frac{\partial (\underline{V}'\underline{D}(\alpha)\underline{V})}{\partial \alpha} = \underline{V}' \frac{\partial \underline{D}(\alpha)\underline{V}}{\partial \alpha},$$

and, from Graybill (1969), p.267,

$$\frac{\partial \underline{X}^{-1}}{\partial \alpha} = -\underline{X}^{-1} \frac{\partial \underline{X}}{\partial \alpha} \underline{X}^{-1}, \quad (4.4.4)$$

therefore,

$$\frac{\partial \underline{D}(\alpha)}{\partial \alpha} = \frac{\partial \left(\underline{X}'\underline{X} + \frac{\alpha \underline{X}'\underline{X}}{N_0} \right)^{-1}}{\partial \alpha} = - \left(\underline{X}'\underline{X} + \frac{\alpha \underline{X}'\underline{X}}{N_0} \right)^{-1} \frac{\underline{X}'\underline{X}}{N_0} \left(\underline{X}'\underline{X} + \frac{\alpha \underline{X}'\underline{X}}{N_0} \right)^{-1}.$$

Expression (4.4.3) can now be written as

$$\frac{\partial \log |\underline{V}'\underline{D}(\alpha)\underline{V}|}{\partial \alpha} = -\text{tr} \left[(\underline{V}'\underline{D}(\alpha)\underline{V})^{-1} \underline{V}'\underline{D}(\alpha) \frac{\underline{X}'\underline{X}}{N_0} \underline{D}(\alpha)\underline{V} \right],$$

which at $\alpha = 0$ becomes

$$- \frac{1}{N_0} \text{tr} \left[\underline{X}_0 \underline{D}\underline{V} (\underline{V}'\underline{D}\underline{V})^{-1} \underline{V}'\underline{D}\underline{X}'_0 \right]$$

since $\underline{D}(0) = \underline{D}$,

and finally

$$-\frac{1}{N_0} \sum \frac{q'Rq}{X'X},$$

where the summation covers all the candidates.

Now, for small α ,

$$\frac{\underline{D}(\alpha) - \underline{D}(0)}{\underline{D}(0)} \approx \log \underline{D}(\alpha) - \log \underline{D}(0),$$

which, using equation (4.4.1) means that the proportional error involved in using $\underline{D}(\alpha)$ rather than \underline{D} is approximately

$$-\alpha \cdot \text{average} \left(\frac{q'Rq}{X'X} \right).$$

For many experiments, the average $\frac{q'Rq}{X'X}$ will be less than one, therefore taking α as 0.005 would give an approximate error of less than 0.5%.

Continuous D-optimal designs with given contrasts actually minimise the maximum value of $\frac{q'Rq}{X'X}$, a result which will be proved by generalizing to include contrasts Federov's (1972, p71) result about normalised D-optimal designs minimising the maximum $v(\underline{x})$. A normalised design is simply a design in which the $\underline{X}'\underline{X}$ matrix is altered to equal $\frac{1}{N} \underline{X}'\underline{X}$. In general a normalised design will be designated as $\underline{\varepsilon}$, and the matrices corresponding to that design will be written $\underline{D}(\underline{\varepsilon})$ etc.

Now, let the optimum design be $\underline{\varepsilon}^*$ and some arbitrary design be $\underline{\varepsilon}$. Then Federov (1972, p66) shows that any design $\tilde{\underline{\varepsilon}}$ can be considered as corresponding to the linear combination of the designs $\underline{\varepsilon}^*$ and $\underline{\varepsilon}$,

$$\tilde{\underline{\varepsilon}} = (1 - \alpha)\underline{\varepsilon}^* + \alpha \underline{\varepsilon} \quad 0 < \alpha < 1.$$

Now,

$$\frac{\partial}{\partial \alpha} \log |\underline{V}'\underline{D}(\tilde{\underline{\varepsilon}})\underline{V}| = \text{tr} \left\{ (\underline{V}'\underline{D}(\tilde{\underline{\varepsilon}})\underline{V})^{-1} \underline{V}' \frac{\partial}{\partial \alpha} \underline{D}(\tilde{\underline{\varepsilon}}) \cdot \underline{V} \right\},$$

and

$$\frac{\partial \underline{D}(\tilde{\underline{\varepsilon}})}{\partial \alpha} = -\underline{D}(\tilde{\underline{\varepsilon}}) \frac{\partial \underline{M}(\tilde{\underline{\varepsilon}})\underline{D}(\tilde{\underline{\varepsilon}})}{\partial \alpha}, \quad \text{where } \underline{M}(\tilde{\underline{\varepsilon}}) = \underline{D}^{-1}(\tilde{\underline{\varepsilon}}) = \underline{X}'(\tilde{\underline{\varepsilon}})\underline{X}(\tilde{\underline{\varepsilon}}),$$

using equations (4.4.2) and (4.4.4).

Now, if $\tilde{\underline{\varepsilon}} = (1 - \alpha)\underline{\varepsilon}^* + \alpha \underline{\varepsilon}$, then

$$\underline{M}(\tilde{\underline{\varepsilon}}) = (1 - \alpha)\underline{M}(\underline{\varepsilon}^*) + \alpha \underline{M}(\underline{\varepsilon}), \quad \text{from the definition of } \underline{M}.$$

Therefore,

$$\frac{\partial \underline{M}(\tilde{\epsilon})}{\partial \alpha} = \underline{M}(\epsilon) - \underline{M}(\epsilon^*)$$

and

$$\begin{aligned} \frac{\partial \log |\underline{V}'\underline{D}(\tilde{\epsilon})\underline{V}|}{\partial \alpha} \Big|_{\alpha=0} &= -\text{tr} \left\{ \underline{R}(\tilde{\epsilon})\underline{V}'\underline{D}(\tilde{\epsilon}) \left[\underline{M}(\tilde{\epsilon}) - \underline{M}(\epsilon^*) \right] \underline{D}(\tilde{\epsilon})\underline{V} \right\} \Big|_{\alpha=0} \\ &= -\text{tr} \left\{ \underline{R}(\epsilon^*)\underline{V}'\underline{D}(\epsilon^*) \left[\underline{M}(\epsilon) - \underline{M}(\epsilon^*) \right] \underline{D}(\epsilon^*)\underline{V} \right\} \end{aligned}$$

which, using $\underline{M}(\epsilon^*)\underline{D}(\epsilon^*) = \underline{I}$, and $\underline{R}(\epsilon^*)\underline{V}\underline{D}(\epsilon^*)\underline{V} = \underline{I}$,

$$= -\text{tr} \left\{ \underline{R}(\epsilon^*)\underline{V}'\underline{D}(\epsilon^*)\underline{M}(\epsilon)\underline{D}(\epsilon^*)\underline{V} \right\} + p, \quad (4.4.5)$$

where p is the number of contrasts.

However, since ϵ^* is optimum, that is, it has the smallest $|\underline{V}'\underline{D}\underline{V}|$, the value of the derivative must be greater than or equal to zero.

Now assume, without loss of generality, that the arbitrary design consists only of one point, that is,

$$\underline{M}(\tilde{\epsilon}) = \underline{x}\underline{x}',$$

and that this point is one of those contained in ϵ^* , then, from (4.4.5)

$$\begin{aligned} \text{tr} \left\{ \underline{R}(\epsilon^*)\underline{V}'\underline{D}(\epsilon^*)\underline{x}\underline{x}'\underline{D}(\epsilon^*)\underline{V} \right\} - p &= \text{tr} \left\{ \underline{x}'\underline{D}(\epsilon^*)\underline{V}\underline{R}(\epsilon^*)\underline{V}'\underline{D}(\epsilon^*)\underline{x} \right\} - p \\ &= \underline{q}'\underline{R}(\epsilon^*)\underline{q}_x - p \leq 0. \end{aligned} \quad (4.4.6)$$

Now consider $\sum_i \frac{\underline{q}'\underline{R}\underline{q}_x}{N}$, where the summation is over all N points of some general normalised design, that is, consider

$$\begin{aligned} \sum_i \frac{\underline{x}_i'\underline{D}\underline{V}\underline{R}\underline{V}'\underline{D}\underline{x}_i}{N} &= \text{tr}(\underline{D}\underline{V}\underline{R}\underline{V}'\underline{D} \sum_i \underline{x}_i\underline{x}_i'/N), \\ &= \text{tr}(\underline{D}\underline{V}\underline{R}\underline{V}'\underline{D}\underline{M}), \end{aligned}$$

which, since $\underline{D}\underline{M} = \underline{I}$,

$$= \text{tr}(\underline{V}'\underline{D}\underline{V}\underline{R}) = p.$$

That is, the average $\frac{\underline{q}'\underline{R}\underline{q}_x}{N}$ equals p for any normalised design.

Therefore the maximum $\frac{\underline{q}'\underline{R}\underline{q}_x}{N}$ must be greater than or equal to p . In particular,

$$\max(\underline{q}'_X R(\mathcal{E}^*) \underline{q}_X) \geq p \quad (4.4.7)$$

Comparison of (4.4.6) and (4.4.7) shows that the optimum design \mathcal{E}^* minimises the maximum $(\underline{q}'_X R(\mathcal{E}^*) \underline{q}_X)$.

For unnormalised designs, $\underline{q}'_X R \underline{q}_X$ will be $1/N$ times that of the normalised design, that is, the average $\underline{q}'_X R \underline{q}_X$, for the continuous design, would be p/N .

4.5 Fixed block size

A useful restriction to place on procedures for constructing optimal block designs is to enforce block sizes to equal some fixed values given by the experimenter. This might be needed, for instance, if the experimenter is working with pots on a shelf, or plants with shoots as blocks and leaves as plots.

This facility can easily be incorporated into a procedure such as that outlined in sections 3 and 4 in the following manner: All blocks have a condition number attached to them; this number is zero initially when the original design is set up with correct block sizes. Whenever the block size is changed by adding or subtracting a point, then the corresponding condition number is altered, being always equal to the difference between the present and correct block sizes. If the condition number is negative, therefore, a block is undersized, and if positive, oversized.

Now suppose that the procedure is in the middle of an excursion with a design having more design points than the original. Then if it is decided to add a further point, this proceeds exactly as before, and the condition number of the effected block is increased by one. If it is decided to subtract a point, then the search over the candidates should be conducted over only those candidates which correspond to blocks having a positive condition number.

If the procedure is in the middle of an excursion with a design

having fewer design points than the original, then points are subtracted exactly as in the original procedure, and the affected block condition number is decreased by one; but when adding points, the search is conducted over only those candidates corresponding to blocks which have a negative condition number.

This new procedure must still give a positive or zero decrease in $|\underline{V}'\underline{D}\underline{V}|$ since the worst that can happen is for the procedure to subtract the same points from the design that were added (or vice versa).

4.6 Over-specification in the Model for Block Designs

In the ordinary block experiment the model

$$\underline{y} = \underline{1}\alpha + \underline{D}'\underline{\beta} + \underline{A}'\underline{\gamma} + \underline{q} \quad (\text{from chapter 2})$$

is used. The matrix \underline{X} for this model is

$$\underline{X} = \left[\underline{1} \mid \underline{D}' \mid \underline{A}' \right],$$

which is not of full rank because

$$\underline{D}'\underline{1} = \underline{A}'\underline{1} = \underline{1}.$$

The matrix $\underline{X}'\underline{X}$ is therefore always singular.

Mitchell (1973) re-defines the model as

$$\underline{y} = \underline{1}\alpha^* + \underline{D}^*\underline{\beta}^* + \underline{A}^*\underline{\gamma}^* + \underline{q},$$

where

$$\alpha^* = \alpha + \beta_b + \gamma_v$$

$$\beta_i^* = \beta_i - \beta_b \quad (i = 1, \dots, b-1)$$

$$\gamma_i^* = \gamma_i - \gamma_v \quad (i = 1, \dots, v-1)$$

$$\underline{D}^* = \underline{D} \text{ with the } b^{\text{th}} \text{ row missing}$$

$$\underline{A}^* = \underline{A} \text{ with the } v^{\text{th}} \text{ row missing}$$

b = number of blocks

v = number of treatments.

Because D-optimality is independent of any non-degenerate linear

transformation (Federov, 1972, p80), any other re-definition makes no difference to the choice of the optimal design. Any contrasts to be applied to the treatments will have the v^{th} element missing. This makes no difference to the value of the contrast since, if \underline{c} is a contrast with elements c_i , then

$$\sum_{i=1}^{v-1} c_i \gamma_i^* = \sum_{i=1}^{v-1} c_i (\gamma_i - \gamma_v) = \left(\sum_{i=1}^v c_i \gamma_i - c_v \gamma_v \right) - \gamma_v \sum_{i=1}^{v-1} c_i .$$

But $\sum_{i=1}^v c_i = 0$ by definition of a contrast, therefore $\sum_{i=1}^{v-1} c_i = -c_v$, and

$$\sum_{i=1}^{v-1} c_i \gamma_i^* = \sum_{i=1}^v c_i \gamma_i - c_v \gamma_v + c_v \gamma_v = \sum_{i=1}^v c_i \gamma_i .$$

4.7 A Computer Program

A FORTRAN IV program, VDVMIN, based upon Mitchell's DETMAX (1974), has been written to produce designs minimising $|\underline{V}'\underline{DV}|$. A listing and example run are given in appendix C. Essentially able to be applied to any model, it is at the moment in the form for optimising block designs, the only alterations needed being the replacement of a subroutine defining the model and the removal of the restriction to fixed block sizes. Because of the possibility of finding a locally optimum design only, the program is applied several times starting with a randomly chosen design each time; also, there is the option of starting with a design chosen by the user.

When augmenting a design, the original points will need to be retained, and the problem is to find the new, larger, design without altering the points from the old one. This is achieved by reading into the program any 'protected' points separately from the other candidates and only choosing points to add or subtract from the unprotected candidates.

4.8 A criterion based on A-optimality

An alternative, and possibly a more natural criterion to use when designing experiments is that of A-optimality, or minimising $\text{tr}\{\underline{D}\}$, that is minimising the mean dispersion of the estimates of the parameters. Chernoff (1953) gives a statistical justification of the use of the A-optimality criterion for the general case where the information matrix depends on the values of the unknown parameters. Federov (1972) states that $(\text{tr}\{\underline{D}\}/N)^{\frac{1}{2}}$, where $\underline{D} = (X'X)^{-1}$, $N =$ number of points, equals half the length of the diagonal of a cuboid circumscribed around the ellipse of dispersion.

Although A-optimality is a very easy criterion to apply between two designs to decide which is the better, procedures for constructing A-optimal designs are more complicated than those for constructing D-optimal designs. A further disadvantage is that the A-optimal design is not invariant under non-degenerate linear transformations of the parameters.

However a relative computational advantage does not occur when contrasts of interest amongst the parameters are given and require to be estimated in some optimal manner. Further, A-optimality has the overriding advantage over D-optimality in that weighting factors can be attached to the contrasts when constructing the optimal design.

The criterion used for judging optimality when contrasts and weights are given will be that a design is optimal for a given number of points, N , if it has the minimum $\text{tr}\{\underline{w}^6 \underline{V}' \underline{D} \underline{V} \underline{w}^6\}$ over all possible designs, where \underline{V} is the matrix having the given contrasts of interest, held columnwise, and \underline{w}^6 is the diagonal matrix holding the corresponding (arbitrary) weights.

The program VDVMIN (see § 4.7), can also produce A-optimal designs.

4.9 Construction of A-optimal designs when contrasts and weights have been given

Federov (1972, p169) gives an exchange algorithm for constructing A-optimal discrete designs which could be generalized to include

contrasts and weights. However, it requires a search routine to find the maximum of a function and would therefore entail much computation, the actual amount depending upon the size of the design and the number of unknown parameters. The procedure could be simplified somewhat by the use of candidates as advocated by Dykstra (1971) for the construction of D-optimal designs; also, rather than exchange two points directly, it is simpler, and leads readily to the use of 'excursions' as suggested by Mitchell (1974), to alter the design one point at a time. That is, in the simplest case, to add on that point which results in the maximum possible decrease in $\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta})$ and then to subtract the point giving the minimum possible increase in $\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta})$, where \underline{D} now corresponds to the design of $N+1$ points containing the point added previously. This must result in a decrease of at least zero in $\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta})$. The criterion for choosing the point to be added can be found in the following way.

Let $(\underline{X}'\underline{X})_{\text{new}}$ be the information matrix corresponding to the design which has just had a point added, that is,

$$(\underline{X}'\underline{X})_{\text{new}} = (\underline{X}'\underline{X}) + \underline{x}\underline{x}' ,$$

where $(\underline{X}'\underline{X})$ is the old information matrix and \underline{x}' is the row of the design matrix corresponding to the point just added. Then $\underline{D}_{\text{new}}$, the inverse of $(\underline{X}'\underline{X})_{\text{new}}$ can then be written as

$$\underline{D}_{\text{new}} = \underline{D} - \frac{\underline{D}\underline{x}\underline{x}'\underline{D}}{1 + v(\underline{x})} \quad (4.9.1)$$

where \underline{D} is the inverse of $(\underline{X}'\underline{X})$ and $v(\underline{x})$ equals $\underline{x}'\underline{D}\underline{x}$, the variance of the estimated response at that point.

Therefore,

$$\underline{w}^{\delta} \underline{V}' \underline{D}_{\text{new}} \underline{V} \underline{w}^{\delta} = \underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta} - \frac{\underline{w}^{\delta} \underline{V}' \underline{D} \underline{x} \underline{x}' \underline{D} \underline{V} \underline{w}^{\delta}}{1 + v(\underline{x})} .$$

Taking the trace,

$$\begin{aligned} \text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D}_{\text{new}} \underline{V} \underline{w}^{\delta}) &= \text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta}) - \frac{\underline{x}' \underline{D} \underline{V} \underline{w}^{\delta} \underline{V}' \underline{D} \underline{x}}{1 + v(\underline{x})} \\ &= \text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta}) - \frac{\underline{q}'_{\underline{x}} \underline{w}^{\delta} \underline{q}_{\underline{x}}}{1 + v(\underline{x})}, \end{aligned}$$

where $\underline{q}_{\underline{x}} = \underline{V}' \underline{D} \underline{x}$.

The point to be chosen, therefore, is the candidate that maximises

$$\frac{\underline{q}'_{\underline{x}} \underline{w}^{\delta} \underline{q}_{\underline{x}}}{1 + v(\underline{x})}.$$

For subtracting a point,

$$(\underline{X}' \underline{X})_{\text{new}} = (\underline{X}' \underline{X}) - \underline{x} \underline{x}',$$

where \underline{x}' is now a row of \underline{X} .

An analogous procedure to that presented above yields

$$\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta}) = \text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta}) + \frac{\underline{q}'_{\underline{x}} \underline{w}^{\delta} \underline{q}_{\underline{x}}}{1 - v(\underline{x})},$$

therefore the point to be subtracted would be the one having the minimum

$$\frac{\underline{q}'_{\underline{x}} \underline{w}^{\delta} \underline{q}_{\underline{x}}}{1 - v(\underline{x})}.$$

Having changed the design, the matrix \underline{D} can be updated using (4.9.1) when adding a point or the equation

$$\underline{D}_{\text{new}} = \underline{D} - \frac{\underline{D} \underline{x} \underline{x}' \underline{D}}{1 - v(\underline{x})} \quad (4.9.2)$$

if the point was subtracted.

Since the sequence formed by the successive values of $\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta})$ for the N-point design is monotonically decreasing, and since an absolute lower bound exists equal to the trace of the actual A-optimal design, the sequence must converge. However, there is no guarantee that it will converge to the true, or global minimum and not to some local minimum. It is necessary, therefore, to repeat the procedure

several times with different starting designs each time.

4.10 Excursions for A-optimal Designs

The use of excursions (Mitchell, 1974) follows on naturally from the basic procedure in an analogous way as that for D-optimality with contrasts. When checking for a failure design, that is, one which was constructed during an excursion which led to no improvement in the design, it is the traces that are compared instead of the determinants (since comparing the whole design exactly would be time and space consuming).

As in D-optimality, singularity of the design can result if too many points are removed from a design during an excursion. As before then, instead of using $(\underline{X}'\underline{X})$ to form \underline{D} and hence the trace, the matrix $(\underline{X}'\underline{X} + \frac{\underline{X}'_0\underline{X}_0}{N_0})$ will be used, where \underline{X}_0 is the design matrix consisting of all N_0 candidates and α is a small positive number. If the inverse is termed $\underline{D}(\alpha)$, then the error involved in using $\text{tr}\{\underline{w}'\underline{V}'\underline{D}(\alpha)\underline{V}\underline{w}\}$ instead of $\text{tr}\{\underline{w}'\underline{V}'\underline{D}\underline{V}\underline{w}\}$ for nonsingular \underline{D} can be approximated in the following way:

Writing \underline{A} for $\underline{V}\underline{w}'\underline{V}'$, then

$$\text{tr}\{\underline{w}'\underline{V}'\underline{D}(\alpha)\underline{V}\underline{w}\} = \text{tr}\{\underline{A}\underline{D}(\alpha)\}.$$

Expanding $\text{tr}\{\underline{A}\underline{D}(\alpha)\}$ in a Taylor series about $\text{tr}\{\underline{A}\underline{D}\}$,

$$\text{tr}\{\underline{A}\underline{D}(\alpha)\} \approx \text{tr}\{\underline{A}\underline{D}\} + \alpha \left. \frac{\partial \text{tr}\{\underline{A}\underline{D}(\alpha)\}}{\partial \alpha} \right|_{\alpha=0}$$

since $\underline{D}(0) = \underline{D}$.

Because matrix multiplication and trace are linear functionals,

$$\frac{\partial \text{tr}\{\underline{A}\underline{D}(\alpha)\}}{\partial \alpha} = \text{tr}\left\{\underline{A} \frac{\partial \underline{D}(\alpha)}{\partial \alpha}\right\}.$$

But,

$$\begin{aligned} \frac{\partial \underline{D}(\alpha)}{\partial \alpha} &= -\underline{D}(\alpha) \frac{\partial}{\partial \alpha} \left[\underline{X}'\underline{X} + \frac{\underline{X}'_0\underline{X}_0}{N_0} \right] \underline{D}(\alpha), \quad (\text{Graybill, 1969, p267}) \\ &= -\underline{D}(\alpha) \frac{\partial}{\partial \alpha} \frac{\underline{X}'_0\underline{X}_0}{N_0} \underline{D}(\alpha). \end{aligned}$$

Therefore,

$$\begin{aligned}
 \alpha \frac{\partial \text{tr} \left\{ \underline{AD}(\alpha) \right\}}{\partial \alpha} \Big|_{\alpha=0} &= -\alpha \text{tr} \left[\underline{AD}(\alpha) \frac{\underline{X}'\underline{X}_0 \underline{D}(\alpha)}{N_0} \right] \Big|_{\alpha=0} \\
 &= -\alpha \text{tr} \left[\underline{Vw}^{2\delta} \underline{V}' \underline{D}(\alpha) \frac{\underline{X}'\underline{X}_0 \underline{D}(\alpha)}{N_0} \right] \Big|_{\alpha=0} \\
 &= -\frac{\alpha}{N_0} \text{tr} \left\{ \underline{X}_0 \underline{D}(\alpha) \underline{Vw}^{2\delta} \underline{V}' \underline{D}(\alpha) \underline{X}'_0 \right\} \Big|_{\alpha=0} \\
 &= -\frac{\alpha}{N_0} \sum q_x' w^{2\delta} q_x \\
 &= -\alpha \cdot \text{average} \left(q_x' w^{2\delta} q_x \right) ,
 \end{aligned}$$

where the summation and average are over all the candidates.

If the average were to cover only those points included in a design with matrix \underline{D} , then

$$\begin{aligned}
 \text{average} \left(q_x' w^{2\delta} q_x \right) &= \frac{1}{N} \text{tr} \left(\underline{XDVw}^{2\delta} \underline{V}' \underline{DX}' \right) = \frac{1}{N} \text{tr} \left(\underline{Vw}^{2\delta} \underline{V}' \underline{DX}' \underline{XD} \right) \\
 &= \frac{1}{N} \text{tr} \left(\underline{Vw}^{2\delta} \underline{V}' \underline{D} \right) = \frac{1}{N} \text{tr} \left(\underline{AD} \right) .
 \end{aligned}$$

If the candidates are reasonable, then the average over all the candidates should be roughly equal to the average over those points in the design only, that is, $\text{tr}(\underline{AD})/N$.

The proportional error involved in using $\underline{D}(\alpha)$ therefore is

$$\begin{aligned}
 \frac{\text{tr} \left\{ \underline{AD}(\alpha) \right\} - \text{tr}(\underline{AD})}{\text{tr}(\underline{AD})} &\approx -\alpha \frac{\text{average} \left(q_x' w^{2\delta} q_x \right) \text{ over all candidates}}{N \cdot \text{average} \left(q_x' w^{2\delta} q_x \right) \text{ over the design points}} \\
 &\approx -\frac{\alpha}{N} .
 \end{aligned}$$

For α equal to 0.005, this represents an error of approximately $\frac{0.5\%}{N}$.

4.11 Inherent Singularity in the Model

The above method only holds if the model is normally non-singular. In the case of a block design with the usual model this is not true and this is where the big drawback of A-optimality lies. It is not in

general true, as it was for D-optimality, that the A-optimal design is invariant under a non-degenerate linear transformation. Therefore, the procedure used in the case of singularity for D-optimality, that of re-defining the model so as to obtain a non-singular one, cannot be used, since the optimal design obtained will depend on the transformation used. Federov (1972, p118) suggests the use of generalized inverses, having the property that eigenvalues of the generalized inverse are reciprocals of those of the original matrix, except for the zero eigenvalues which will correspond to zero eigenvalues in the generalized inverse.

Chernoff (1953) suggests the use of

$$\lim_{\lambda \rightarrow 0^+} (\underline{X} + \lambda \underline{Y})^{-1}, \text{ where } \underline{Y} \text{ is a symmetric matrix,}$$

such that $(\underline{X} + \lambda \underline{Y})$ is positive definite, as a pseudo-inverse of \underline{X} , when \underline{X} is singular.

The disadvantage of generalized inverses is that no simple formula analogous to (4.9.1) or (4.9.2) exists for updating the generalized inverse after each addition or subtraction of a point. To calculate the generalized inverse afresh each time would, of course, involve too much computation.

The case of block experiments, the usual model for which is always singular, can be rescued somewhat if it is considered that the following model is in some sense a 'natural' reduction of the original model:

$$\underline{y} = \underline{D}'\underline{\beta} + \underline{A}'\underline{x}_0 + \underline{q} \quad (4.11.1)$$

where

$$\underline{x}_0 = \underline{x} + \underline{1}\alpha,$$

that is, the treatment parameters now have the general mean embedded in them. This of course makes no difference to treatment contrasts. If this model (which is still singular), is further reduced by altering, or reparameterising, the block parameters in some way, then, for the

resulting model, an optimal design is independent of any non-degenerate transformation applied to the block parameters only. This can be seen by considering the model, which has been reduced in the general case to

$$\underline{y} = (\underline{D}'\underline{R})\underline{\beta}_0 + \underline{\Delta}'\underline{\delta}_1 + \underline{\eta} \quad (4.11.2)$$

where $\underline{\beta}_0 = \underline{T}\underline{\beta}$

and $\underline{\delta}_1 = \underline{S}\underline{\beta} + \underline{\alpha}_0$

are now the unknown parameters to be estimated and \underline{R} , \underline{S} and \underline{T} are transformation matrices satisfying

$$\underline{D}'\underline{R}\underline{T} + \underline{\Delta}'\underline{S} = \underline{D}'$$

and such that $[\underline{D}'\underline{R} : \underline{\Delta}']$ is of full rank.

If a non-degenerate linear transformation \underline{Y} is now applied to the block parameters, then the transformation \underline{L} to be applied to the parameters as a whole must be of the form

$$\underline{L} = \begin{bmatrix} \underline{Y} & \underline{O} \\ \underline{Z} & \underline{I} \end{bmatrix}, \quad (4.11.3)$$

if the parameters are partitioned thus: $\begin{bmatrix} \underline{\beta}_0 \\ \underline{\delta}_1 \end{bmatrix}$.

Now, as the contrasts of interest, \underline{V} , are concerned only with the treatment parameters, \underline{V} is of the form

$$\underline{V} = \begin{bmatrix} \underline{O} \\ \underline{C} \end{bmatrix},$$

where \underline{C} hold the treatment contrasts.

Now consider $\text{tr}(\underline{w}^\delta \underline{V}' \underline{D}_* \underline{V} \underline{w}^\delta)$, where \underline{D}_* is the dispersion matrix for the new model, that is, $\underline{D}_* = \underline{L}' \underline{D} \underline{L}$, where \underline{D} is the dispersion matrix for the model used before the transformation \underline{L} was applied. Then

$$\begin{aligned} \text{tr}(\underline{w}^\delta \underline{V}' \underline{D}_* \underline{V} \underline{w}^\delta) &= \text{tr}(\underline{w}^\delta \underline{V}' \underline{L}' \underline{D} \underline{L} \underline{V} \underline{w}^\delta) \\ &= \text{tr}(\underline{L} \underline{V} \underline{w}^{2\delta} \underline{V}' \underline{L}' \underline{D}) \end{aligned}$$

$$\begin{aligned}
\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta}) &= \text{tr} \left\{ \begin{array}{c} -74- \\ \left[\begin{array}{c|c} \underline{Y} & \underline{O} \\ \hline \underline{Z} & \underline{I} \end{array} \right] \left[\begin{array}{c|c} \underline{O} & \underline{O} \\ \hline \underline{I} & \underline{C} \end{array} \right] \underline{w}^{2\delta} (\underline{O} \mid \underline{C}') \left[\begin{array}{c|c} \underline{Y}' & \underline{Z}' \\ \hline \underline{O} & \underline{I} \end{array} \right] \underline{D} \end{array} \right\}, \\
&= \text{tr} \left\{ \begin{array}{c} \left[\begin{array}{c|c} \underline{Y} & \underline{O} \\ \hline \underline{Z} & \underline{I} \end{array} \right] \left[\begin{array}{c|c} \underline{O} & \underline{O} \\ \hline \underline{O} & \underline{O} \end{array} \right] \left[\begin{array}{c|c} \underline{O} & \\ \hline \underline{C}_w^{2\delta} & \underline{C}' \end{array} \right] \left[\begin{array}{c|c} \underline{Y}' & \underline{Z}' \\ \hline \underline{O} & \underline{I} \end{array} \right] \underline{D} \end{array} \right\}, \\
&= \text{tr} \left\{ \begin{array}{c} \left[\begin{array}{c} \underline{O} \\ \hline \underline{O} \end{array} \right] \left[\begin{array}{c|c} \underline{O} & \\ \hline \underline{C}_w^{2\delta} & \underline{C}' \end{array} \right] \underline{D} \end{array} \right\}, \\
&= \text{tr} \left\{ \begin{array}{c} \left[\begin{array}{c} \underline{O} \\ \hline \underline{C} \end{array} \right] \underline{w}^{2\delta} (\underline{O} \mid \underline{C}') \underline{D} \end{array} \right\}, \\
&= \text{tr}(\underline{V} \underline{w}^{2\delta} \underline{V}' \underline{D}) = \text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta}).
\end{aligned}$$

Therefore, $\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta})$ is independent of the transformation \underline{L} when of the form (4.11.3).

An easy reduction of the original block design model to the model (4.11.2) is brought about by simply removing the rows in the design matrix \underline{X} corresponding to the mean and some arbitrary block j . The treatment parameters are then

$$\underline{y}_1 = \underline{y} + 1\alpha + 1\beta_j,$$

and the $(b-1)$ block parameters are

$$\beta_{oi} = \beta_i - \beta_j \quad (i = 1, \dots, j-1, j+1, \dots, b).$$

It should be stressed that A-optimality in this model is independent only of transformations which are not applied to the treatment parameters. Under the transformations used by Mitchell (1973), therefore, which affected both the treatment and block parameters, the A-optimal design would be dependent upon the treatment transformation used.

4.12 A Criterion Based on the Simple or Arithmetic Mean of the Efficiency Factors

A third criterion for differentiating between designs is the simple or arithmetic mean of the efficiency factors (Jones, 1959); that is, a design is said to be optimal in this sense if it has the

maximum mean of its efficiencies over all possible designs. This has some statistical justification in the fact that it is equivalent to minimising Pearson's χ^2 statistic for goodness of fit, when applied to the design matrix for treatments to blocks, \underline{n} , with the expected value chosen as the value from an orthogonal experiment, that is, $\underline{n} = \underline{rk}'/N$.

This can be proved as follows:

The statistic χ^2 can be written as

$$\chi^2 = \sum_{i=1}^v \sum_{j=1}^b \frac{(n_{ij} - r_{i \cdot k_j}/N)^2}{r_{i \cdot k_j}/N}$$

where $\underline{n} = [n_{ij}]$ etc. Expanding,

$$\begin{aligned} \chi^2 &= \sum_i \sum_j (Nn_{ij}^2 - 2n_{ij}r_{i \cdot k_j} + r_{i \cdot k_j}^2/N) / r_{i \cdot k_j} \\ &= \sum_i \sum_j Nn_{ij}^2/r_{i \cdot k_j} - 2\sum_i \sum_j n_{ij} + \sum_i \sum_j r_{i \cdot k_j}/N \\ &= N\sum_i \sum_j n_{ij}^2/r_{i \cdot k_j} - 2N + N \\ &= N(\sum_i \sum_j n_{ij}^2/r_{i \cdot k_j} - 1) \\ &= N(\text{tr}[\underline{r}^{-\delta} \underline{n} \underline{n}' - \underline{\delta} \underline{n}']) - 1 \\ &= N(\text{tr}[\underline{M}] - 1) \\ &= N(\sum \mu_i - 1) \end{aligned}$$

where \underline{M} is the matrix used by Jones (1959) with eigenvalues μ_i .

The quantities $(1 - \mu_i)$ are the efficiency factors of the design.

Minimising the χ^2 statistic, therefore, is equivalent to minimising

$\sum \mu_i$, or equivalently, maximising $\sum \varepsilon_i$, where $\varepsilon_i = 1 - \mu_i$, are the

efficiency factors. The use of the simple mean is also proposed by

Pearce (1974), who makes the point that if an efficiency is zero,

corresponding to some contrasts between the treatments being confounded,

then the overall mean, $\sum \varepsilon_i$, does not also become zero.

When contrasts and weights are given, it is not obvious how the

criterion should be amended. However, consideration of D and A-optimality will help here. In D-optimality, the criterion used was $|\underline{V}'\underline{D}\underline{V}|$; if $\underline{V}' = (\underline{0} \mid \underline{V}'_t)$, where \underline{V}_t applies to treatments only, then the criterion can be written as $|\underline{V}'_t \underline{A}^- \underline{V}_t|$ where \underline{A}^- is the dispersion matrix of the treatments as in chapter 2. From (2.5.0), $\underline{A}^- = \underline{r}^{-\delta} \underline{C} \underline{\epsilon}^{-\delta} \underline{C}' \underline{r}^{-\delta}$, where reciprocals of zero are taken as zero. The matrix \underline{C} holds the eigenvectors of the design, the basic contrasts, as its columns. Since \underline{C} is of full rank, \underline{V}_t can be written as $\underline{C}\underline{\theta}$. The criterion for D-optimality can therefore be written as

$$\begin{aligned} |\underline{V}'\underline{D}\underline{V}| &= |\underline{\theta}' \underline{C}' \underline{r}^{-\delta} \underline{C} \underline{\epsilon}^{-\delta} \underline{C}' \underline{r}^{-\delta} \underline{C} \underline{\theta}| \\ &= |\underline{\theta}' \underline{\epsilon}^{-\delta} \underline{\theta}| \quad \text{since } \underline{C}' \underline{r}^{-\delta} \underline{C} = \underline{I} \text{ from (2.3.1)}. \end{aligned}$$

If contrasts were not used the criterion would have been $|\underline{\epsilon}^{-\delta}|$.

Similarly, for A-optimality, the criterion can be written as

$$\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{D} \underline{V} \underline{w}^{\delta}) = \text{tr}(\underline{w}^{\delta} \underline{\theta}' \underline{\epsilon}^{-\delta} \underline{\theta} \underline{w}^{\delta}),$$

or, for no contrasts, $\text{tr}(\underline{\epsilon}^{-\delta})$, with reciprocals of zero taken as zero, as before.

It will be seen then, that the criterion used for the simple mean optimality, $\sum \epsilon_i$, or $\text{tr}(\underline{\epsilon}^{\delta})$, should be altered to $\text{tr}(\underline{w}^{\delta} \underline{\theta}' \underline{\epsilon}^{\delta} \underline{\theta} \underline{w}^{\delta})$ when contrasts are used. This can be written as

$$\text{tr}(\underline{w}^{\delta} \underline{\theta}' \underline{C}' \underline{r}^{-\delta} \underline{C} \underline{\epsilon}^{\delta} \underline{C}' \underline{r}^{-\delta} \underline{C} \underline{\theta} \underline{w}^{\delta}) \quad \text{since } \underline{C}' \underline{r}^{-\delta} \underline{C} = \underline{I}. \quad (4.12.1)$$

But $\underline{A} = \underline{C} \underline{\epsilon}^{\delta} \underline{C}'$, from (2.3.2), and using $\underline{V}_t = \underline{C} \underline{\theta}$, enables (4.12.1) to be written as

$$\text{tr}(\underline{w}^{\delta} \underline{V}'_t \underline{r}^{-\delta} \underline{A} \underline{r}^{-\delta} \underline{V}_t \underline{w}^{\delta}). \quad (4.12.2)$$

If the contrasts \underline{V}_t are such that $\underline{V}'_t \underline{r}^{-\delta} \underline{V}_t = \underline{I}$, then (4.12.2) becomes

$$\text{tr}\{\underline{w}^{\delta} (\underline{V}'_t \underline{A} \underline{V}_t)^{-1} \underline{w}^{\delta}\},$$

that is, the 'information matrix' for the contrasts is being used.

4.13 Construction of Simple Mean Optimal Designs with Contrasts

The matrix \underline{A} equals $\underline{r}^{\delta} - \underline{nk}^{\delta} \underline{n}'$ (from 2.2.1), therefore, for fixed

\underline{r} , maximising (4.12.2) is equivalent to minimising

$$\text{tr}(\underline{w}^{\delta} \underline{V}' \underline{r}^{-\delta} \underline{nk}^{-\delta} \underline{n}' \underline{r}^{-\delta} \underline{Vw}^{\delta}) ,$$

(dropping the t suffix on the \underline{V}_t matrix), or, since $\text{tr}(\underline{AB}) = \text{tr}(\underline{BA})$,

to minimising

$$\text{tr}(\underline{r}^{-\delta} \underline{Vw}^{2\delta} \underline{V}' \underline{r}^{-\delta} \underline{nk}^{-\delta} \underline{n}') . \quad (4.13.1)$$

This is then equal to the criterion used by Pearce (1974). If the incidence matrix \underline{n} is partitioned thus:

$$\underline{n} = (\underline{n}_1 \vdots \underline{n}_2 \vdots \underline{n}_3 \vdots \dots \vdots \underline{n}_b) ,$$

the partitions corresponding to blocks, of size k_i , then (4.13.1)

can be written as

$$\sum_{i=1}^b \text{tr}(\underline{S} \underline{n}_i \underline{n}_i' / k_i) ,$$

where $\underline{S} = \underline{r}^{-\delta} \underline{Vw}^{2\delta} \underline{V}' \underline{r}^{-\delta}$. Note that \underline{S} is independent of the design once \underline{r} has been fixed.

To find the optimum value of \underline{r} , Pearce (1974) suggests minimising the weighted sum of the variance of the scaled contrasts, assuming an orthogonal design were possible. The scaling is such that $\underline{c}_i' \underline{c}_i$ is the same for all contrasts. The vector \underline{r} is chosen therefore to minimise

$$\sum_i \left\{ \frac{w_i \underline{c}_i' \underline{r}^{-\delta} \underline{c}_i}{\underline{c}_i' \underline{c}_i} \right\} \text{ subject to } \sum r_i = N. \quad (4.13.2)$$

Pearce does not give the algebraic solution to this problem, but it can be obtained as follows:

Using the method of Lagrangian multipliers, the value of \underline{r} which gives the minimum of (4.13.2) is obtained as the solution to

$$\frac{\partial}{\partial r_j} \left\{ \sum_i \frac{(w_i \underline{c}_i' \underline{r}^{-\delta} \underline{c}_i)}{\underline{c}_i' \underline{c}_i} + \lambda (\sum r_j - N) \right\} = 0 \quad \text{for each } j.$$

$$\text{i.e.} \quad \frac{\partial}{\partial r_j} \left\{ \sum_i \sum_j \frac{c_{ij}^2 w_i}{a_i r_j} + \lambda (\sum r_j - N) \right\} = 0$$

where $a_i = \underline{c}_i' \underline{c}_i$ and $[c_{ij}] = \underline{c}_i$

Therefore,

$$-\sum_i \frac{c_{ij}^2 w_i}{a_i r_j^2} + \lambda = 0,$$

or

$$r_j = \sqrt{\sum_i \frac{c_{ij}^2 w_i}{a_i \lambda}}.$$

That is, the r_j must be chosen proportional to $\sqrt{\sum_i c_{ij}^2 w_i / a_i}$.

Having chosen \underline{r} , it is now required to find the values of the vectors \underline{n}_i that minimise

$$\sum_i \text{tr}(\underline{S} \underline{n}_i \underline{n}_i' / k_i).$$

Now, suppose that \underline{n}_i is altered to $(\underline{n}_i + \underline{x})$, and \underline{n}_j to $(\underline{n}_j - \underline{x})$, for some i and j , where \underline{x} is a vector of zeroes except for two elements, these being +1 and -1. Let these non-zero elements be in the k^{th} and l^{th} positions respectively. A restriction on the choice of these positions is that the new vectors \underline{n}_i and \underline{n}_j must have zero or positive elements only. Altering \underline{n}_i and \underline{n}_j in this way leaves the vectors \underline{r} and \underline{k} unchanged. The vector \underline{k} is in general fixed and known beforehand. The change brought about to $\sum_i \text{tr}(\underline{S} \underline{n}_i \underline{n}_i' / k_i)$ is

$$\begin{aligned} & \text{tr} \left\{ \frac{\underline{S}(\underline{n}_i + \underline{x})(\underline{n}_i + \underline{x})'}{k_i} + \frac{\underline{S}(\underline{n}_j - \underline{x})(\underline{n}_j - \underline{x})'}{k_j} - \frac{\underline{S} \underline{n}_i \underline{n}_i'}{k_i} - \frac{\underline{S} \underline{n}_j \underline{n}_j'}{k_j} \right\} \\ = & \text{tr} \left\{ \underline{S} \left[\frac{\underline{n}_i \underline{n}_i'}{k_i} + \frac{\underline{n}_i \underline{x}'}{k_i} + \frac{\underline{x} \underline{n}_i'}{k_i} + \frac{\underline{x} \underline{x}'}{k_i} + \frac{\underline{n}_j \underline{n}_j'}{k_j} - \frac{\underline{n}_j \underline{x}'}{k_j} - \frac{\underline{x} \underline{n}_j'}{k_j} + \frac{\underline{x} \underline{x}'}{k_j} - \frac{\underline{n}_i \underline{n}_i'}{k_i} - \frac{\underline{n}_j \underline{n}_j'}{k_j} \right] \right\} \end{aligned}$$

Using the identity

$$\text{tr}(\underline{S} \underline{p} \underline{q}') = \text{tr}(\underline{q}' \underline{S} \underline{p}) = \text{tr}(\underline{p}' \underline{S} \underline{q}) = \text{tr}(\underline{S} \underline{q} \underline{p}'),$$

since \underline{S} is symmetric, gives the difference as

$$\begin{aligned} & \text{tr} \left\{ \frac{\underline{n}_i' \underline{S} \underline{x}}{k_i} + \frac{\underline{n}_i' \underline{S} \underline{x}}{k_i} + \frac{\underline{x}' \underline{S} \underline{x}}{k_i} - \frac{\underline{n}_j' \underline{S} \underline{x}}{k_j} - \frac{\underline{n}_j' \underline{S} \underline{x}}{k_j} + \frac{\underline{x}' \underline{S} \underline{x}}{k_j} \right\} \\ = & 2 \left\{ \frac{\underline{n}_i}{k_i} - \frac{\underline{n}_j}{k_j} \right\}' \underline{S} \underline{x} + \left(\frac{1}{k_i} + \frac{1}{k_j} \right) \underline{x}' \underline{S} \underline{x}. \end{aligned} \tag{4.13.3}$$

This is required to be negative and as large as possible, that is, to result in the largest possible decrease in (4.13.1). Note that since \underline{S} is non-negative definite, the second term in (4.13.3), is always greater than or equal to zero.

Now, since an orthogonal experiment has efficiencies equal to one (except for that corresponding to the mean, which equals zero), and therefore results in the maximum value of (4.13.2), or the minimum of (4.13.1), the lowest value of $\text{tr}(\underline{S}n_i n_i' / k_i)$ occurs when $\underline{n} = \underline{r}k' / N$. The actual minimum depends on whether the contrasts for the mean, \underline{r} , has been included, but, if so the weighting attached is usually zero, and so the absolute minimum of (4.13.1) can be taken as zero. In general, of course, this will be unobtainable.

To choose the blocks to be altered therefore, find the two that have the largest $\text{tr}(\underline{S}n_i n_i' / k_i)$. This is not an error free method in that it can sometimes happen that no improvement can be made using these two blocks, but could have been made using others. However it has the advantages of simplicity. Suppose the blocks chosen were i and j .

Now, (4.13.3) can be written as

$$\underline{b}'\underline{x} + \underline{x}'\underline{C}\underline{x}$$

where
$$\underline{b}' = 2 \begin{pmatrix} \frac{n_i}{k_i} & - \frac{n_j}{k_j} \end{pmatrix}' \underline{S}$$

and
$$\underline{C} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \end{pmatrix} \underline{S}$$

which expands, using the definition of \underline{x} , to

$$\begin{aligned} & b_k - b_l + c_{kk} - 2c_{kl} \\ & = (c_{kk} + b_k) + (c_{ll} - b_l) - 2c_{kl} \end{aligned} \tag{4.13.4}$$

Since \underline{C} is non-negative definite, then

$$c_{kk} + c_{ll} - 2c_{kl} \geq 0$$

and
$$c_{kk} + c_{ll} + 2c_{kl} \geq 0$$

which together imply that

$$2c_{kk} + 2c_{ll} \geq c_{kk} + c_{ll} - 2c_{kl} .$$

Equation (4.13.4) is then less than or equal to

$$(2c_{kk} + b_k) + (2c_{ll} - b_l) . \quad (4.13.5)$$

Therefore, as a 'coarse' minimising procedure, choose treatment k so as to minimise $(2c_{kk} + b_k)$ and treatment l so as to minimise $(2c_{ll} - b_l)$ subject to \underline{n}_i and \underline{n}_j being positive or zero at each element. This procedure requires only $o(v)$ steps. After each 'adjustment' calculate the new $\text{tr}(\underline{S}_{i, \underline{n}_i} / k_i)$ for each block and so choose the new i and j .

When no more progress can be made, equation (4.13.4) can be minimised instead of (4.13.5). This, however, called the 'fine' minimisation procedure, requires $o(v(v-1))$ steps in the search routine because of the presence of the element c_{kl} .

As usual in discrete optimality procedures, the method cannot be guaranteed to produce a global or absolute minimum. The 'full' minimisation procedure would involve finding the minimum of (4.13.3) over all pairs of blocks as well and so would involve $o(b(b-1)v(v-1))$ steps.

4.14 Discussion

Current literature on the subject of optimality (Wynn, 1970; Dykstra, 1971; Box and Draper, 1971; Mitchell, 1974) indicates a preference for the determinant criterion for choosing between designs and for constructing them. Although certain advantages are claimed for D-optimality, not least the fact that linear transformations leave the choice of the D-optimal design unchanged; when contrasts of interest to the experimenter are introduced, the fact that weighting cannot be applied using D-optimality is a very large disadvantage. Also, the claim that constructing A-optimal designs leads to more computation (Federov, 1972, p138) than for D-optimal designs is seen not to hold

when contrasts are specified, and although some trouble may arise due to the non-uniqueness under linear transformations of the optimum design, the advantage of weighting would seem to be overriding.

The advantages of the third criterion, i.e. that based on χ^2 , are its simplicity of use and its application to designs having a zero efficiency factor, that is, unconnected or fully confounded designs. Its disadvantages lie in the fact that when the full minimisation procedure is not used it is too often trapped on a very local minimum, and if the full minimisation is used, the computation becomes excessive. Possibly, also, not enough freedom is given to the procedure by not allowing treatment replications to vary during the construction. Perhaps a similar approach to that used for D and A-optimality may prove useful, where points are added to the design for the maximum improvement and subtracted for the minimum deterioration in the criterion. This proves difficult, however, since the matrix \underline{S} would then not be constant and much of the simplicity would be lost.

A disadvantage shared by all procedures dealing with discrete designs is the high probability of being trapped at a local minimum (or maximum). It would seem that the only solution is to use various, reasonably 'seperated' in some sense, starting designs in the hope that one, at least, will give the true minimum.

In this chapter the main results of the thesis are summarised and the advantages or disadvantages of the various methods used are given.

The method of analysing block experiments given in chapter 2 relies on the use of the eigenvectors of the matrix $\underline{Ar}^{-\delta}$. Accordingly, it is proved that a complete set of eigenvectors does indeed exist for this matrix. These eigenvectors and their corresponding eigenvalues, once known, enable the generalized inverse of \underline{A} to be found and thus the analysis of the design follows. If the eigenvalues are known, from the use of some standard routine, then expanding powers of a matrix, derived from $\underline{Ar}^{-\delta}$, in terms of its principal idempotents, \underline{L}_i , leads to a set of simultaneous matrix equations in the \underline{L}_i . These equations can often be simplified if the experimenter knows or can guess which contrasts between treatments are basic, that is, eigenvectors of the matrix $\underline{Ar}^{-\delta}$. The number of unknowns in the equations, the \underline{L}_i matrices, equals the number of distinct eigenvalues, that is, the order of balance of the design in the Nelder (1965) sense. It appears then that it is sufficient to invert a matrix of order equal to the order of balance only, to solve the equations. An alternative method of obtaining the \underline{L}_i matrices, due to Pease (1965) is compared but is shown to involve more computation.

It is shown that the inter-block analysis can also be obtained from the \underline{L}_i matrices calculated for the intra-block analysis. Once the intra-block analysis has been calculated therefore, no problem arises for the inter-block analysis.

A formula for testing some general hypotheses concerning the treatments is also given in chapter 2. The formula, which is the same as that given by John and Smith (1974), who had extended a formula derived by Pringle and Rayner (1971), was found using a different method.

A program, GEN, written in ALGOL 60, has been written to perform the above analyses and a listing together with an example run appears in appendix B. A flowchart of the program is also provided.

An alternative method of analysing a design, based on the work of Kuiper (1952) and utilising results by Nelder (1965) was derived and appears in Worthington (1975). A copy is given in appendix A. The method, which can be applied to any design with orthogonal block structure, iteratively produces the treatment effects for each stratum. It is proved that the procedure will always converge to the correct result. The basic procedure can easily be extended to produce the variance-covariance matrix for each stratum. A disadvantage of the procedure is that no information is produced about degrees of freedom. The method will also take a long time to converge if a low non-zero efficiency factor (Jones, 1959) exists.

In chapter 3, theory concerning the dual of a block design is given. The main result is that non-unit efficiencies for one design agree in value and multiplicity with those of its dual. Further, a specific relationship exists between basic contrasts of one design and those of the other. Consequences of this result include the fact that a certain minimum number of efficiencies must be equal to unity in the design with the greater number of 'treatments'; and that the analysis for the dual can be obtained from that for the original design due to the relationships between the basic contrasts of the two designs being known. If the analysis of chapter 2 had been used for analysing the original design, and therefore the L_1 matrices found, then it is shown that the analysis for the dual also depends on these matrices and the analysis follows on very simply. Finally, relationships between the mean efficiency of one design and that of the other are given for various measures of mean efficiency.

Chapter 4 is concerned with the construction of optimal designs. Although procedures already exist for doing this (for example, Mitchell,

1974), very few use the fact that some contrasts between treatments will be of more interest to the experimenter than others.

On this basis, an algorithm for constructing D-optimal designs when contrasts of interest are given was developed. The method requires inverses and determinant to be calculated only once, at the beginning, after which the required determinant and matrices are updated when required by using simple formulae. The method uses 'excursions' as advocated by Mitchell (1974); because this can result in singularity the information matrix is altered slightly to prevent this. A formula for the resulting error is obtained and the value of this error shown to be small. A program, VDVMIN, written in FORTRAN IV, has been written and appears in appendix C together with an example run and flowchart. The main disadvantage of D-optimality is that weighting the contrasts makes no difference to the criterion the procedure attempts to minimise. All those contrasts that are given therefore, have always in effect equal weighting factors ascribed to them.

This disadvantage is not shared by the second criterion considered, that of A-optimality. When contrasts, held columnwise in V , with corresponding weights w , are given, then the A-optimal design is said to be the design minimising $\text{tr}(w^{\delta} V' D V w^{\delta})$, D being the covariance matrix of the design (a D-optimal design would minimise $|V' D V|$). An algorithm has also been derived for this criterion. As in D-optimality, once the initial D matrix has been calculated, it is updated using a simple formula each time a point of the design is changed. When excursions are used, because singularity may result, the information matrix is altered in the same way as for D-optimality. A formula for the error involved in doing this is calculated and the value shown to be small.

Because A-optimality is not invariant under linear transformations of the parameters, the case of the usual block design model requires special consideration. In general, depending on the reduction used to make it non-singular, different A-optimal designs will result. This

is overcome by assuming that contrasts of interest will be applied to treatments only. It is then shown that if the model is reduced by altering the non-treatment parameters only, then A-optimality under this model is invariant under any linear transformation applied to the non-treatment parameters. That is, A-optimality is invariant of the particular reduction used as long as the reduction applies to the non-treatment parameters only. The program VDVMIN, mentioned above, is also suitable for constructing A-optimal designs, and an example run is given in appendix C.

The third criterion considered is that of simple mean optimality; in its simplest form this involves maximising the arithmetic or simple mean of the efficiency factors. This is shown to be equivalent to minimising a suitably defined χ^2 statistic for the design. The criterion can be generalized to include the case of contrasts of interest and associated weights being given. An algorithm is given for constructing the optimal design and works in three modes: full, fine and coarse, the full mode involving the most calculation and coarse, the least. Coarse mode is thus applied first to some starting design and when the algorithm stops at a possibly local maximum, fine mode is applied, under which, the extra freedom of movement should enable the procedure to escape a local maximum. The full mode in general requires too much calculation and it is probably better to run the algorithm several times with different starting designs and use fine and coarse modes only. All three algorithms can become trapped at a local maximum or minimum but because of the discrete nature of the problem (replications of treatments must be integral, etc.) this seems to be an unavoidable drawback of the method.

REFERENCES

- AITKEN, A.C. (1934). On least squares and linear combinations of observations. Proc. Roy. Soc. Edinburgh 55, 42-48.
- BOSE, R.C. & NAIR, R.R. (1939). Partially balanced incomplete block designs. Sankhya 4, 337-73.
- BOX, G.E.P. & LUCAS, H.L. (1959). Design of experiments in non-linear situations. Biometrika 46, 77-90.
- BOX, M.J. (1968). The occurrence of replications in optimal designs of experiments to estimate parameters in non-linear models. J. R. Statist. Soc. B 30, 290-302.
- BOX, M.J. & DRAPER, N.R. (1971). Factorial designs, the $|X'X|$ criterion and some related matters. Technometrics 13, 731-42.
- CALINSKI, T. (1971). On some desirable patterns in block designs. Biometrics 27, 275-92.
- CHAKRABARTI, M.C. (1963). On the C-matrix in design of experiments. J. Indian Statist. Ass. 1, 8-23.
- CHERNOFF, H. (1953). Locally optimum designs for estimating parameters. Ann. Math. Stat. 24, 586-602.
- DYKSTRA, O. (Jr) (1971). The augmentation of experimental data to maximise $|X'X|$. Technometrics 13, 682-8.
- FEDEROV, V.V. (1972). Theory of Optimal Experiments. New York and London: Academic Press.
- FISHER, R.A. & YATES, F. (1963). Statistical Tables for Biological, Agricultural and Medical Research. Edinburgh: Oliver and Boyd.
- GAUSS, C.F. (1873). Theoria combinationis observationum erroribus minimis obnoxiae. Werke 4, 3-93.
- GOLDMAN, A.J. & ZELEN, M. (1964). Weak generalized inverses and minimum variance linear unbiased estimation. J. Res. Nat. Bur. Stand. B (Maths and Math. Phys.) 68B, 151-72.
- GRAYBILL, F.A. (1961). An Introduction to Linear Statistical Models. Vol. 1. New York and London: McGraw - Hill.
- GRAYBILL, F.A. (1969). Introduction to Matrices with Applications in Statistics. California: Wadsworth Publishing Co..

- HOUSEHOLDER, A.S. (1964). The Theory of Matrices in Numerical Analysis. New York: Blaisdell.
- JAMES, A.T. & WILKINSON, G.N. (1971). Factorization of the residual operator and canonical decomposition of non-orthogonal factors in the analysis of variance. *Biometrika* 58, 279-95.
- JOHN, J.A. & SMITH, T.M.F. (1974). Sum of squares in non-full rank general linear hypotheses. *J. R. Statist. Soc. B* 36, 107-8.
- JOHN, P.W.M. (1964). Pseudo-inverses in the analysis of variance. *Annals Math. Stat.* 35, 895-6.
- JONES, R.M. (1959). On a property of incomplete blocks. *J.R. Statist. Soc. B* 21, 172-9.
- KIEFER, J. (1961). Optimum experimental designs V, with applications to systematic and rotatable designs. *Proc. 4th Berkeley Symp. Math. Stat. and Prob. Vol 1*, 381-405.
- KUIPER, N.H. (1952). Variante-analyse. *Statistica* 6, 151-94.
- MITCHELL, T.J. (1973). Computer construction of small 'D-optimal' incomplete block designs. *Contributed papers, 39th session of the Int. Stat. Inst.*, 732-8.
- MITCHELL, T.J. (1974). An algorithm for the construction of 'D-optimal' experimental designs. *Technometrics* 16, 203-10.
- MOORE, E.H. (1920). On the reciprocal of the general algebraic matrix (abstract). *Bull. Amer. Math. Soc.* 26, 394-5.
- NELDER, J.A. (1965). The analysis of randomised experiments with ~~out~~ orthogonal block structure (parts I and II). *Proc. Roy. Soc. Lond. A* 283, 147-78.
- NELDER, J.A. (1968). The combination of information in generally balanced designs. *J. R. Statist. Soc. B* 30, 303-11.
- PATTERSON, H.D. & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-54.
- PEARCE, S.C. (1968). The mean efficiency of equi-replicate designs. *Biometrika* 55, 251-3.
- PEARCE, S.C. (1970). The efficiency of block designs in general. *Biometrika* 57, 339-46.
- PEARCE, S.C. (1971). Precision in block experiments. *Biometrika* 58 161-7.

- PEARCE, S.C. (1974). Paper presented to the 8th International Biometric conference, Constanta, Romania, 1974.
- PEARCE, S.C., CALINSKI, T. & MARSHALL, T.F. de C. (1974). The basic contrasts of an experimental design with special reference to the analysis of data. *Biometrika* 61, 449-60.
- PEASE, M.C. (1965). *Methods of Matrix Algebra*. New York and London: Academic Press.
- PENROSE, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.* 51, 406-13.
- PLACKETT, R.L. (1960). *Principles of Regression Analysis*. Oxford: Clarendon Press.
- PRINGLE, R.M. & RAYNER, A.A. (1971). *Generalized Inverse Matrices with Applications to Statistics*. London: Griffin.
- RAO, C.R. (1966). Generalized inverse for matrices and its applications to mathematical statistics. *Research papers in statistics. Festschrift for J. Neyman*. Wiley & Sons. 263-79 (Ed. F. N. David)
- RAO, C.R. & MITRA, S.K. (1971). *Generalized Inverse of Matrices and its applications*. New York and London: Wiley and Sons.
- RHODE, C.A. (1968). Special applications to the theory of generalized matrix inversion to statistics. *Proc. Symp. Thy. and Applic. of Gen. Inv. Matrices*. 239-66.
- SMITH, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* 12, 1-85.
- STIGLER, S.M. (1970). Optimal experimental designs for polynomial regression. *J. Amer. Stat. Assoc.* 66, 311-8.
- TOCHER, K.D. (1952). The design and analysis of experiments. *J. R. Stat. Soc. B* 14, 45-91.
- WILKINSON, G.N. (1970). A general recursive procedure for analysis of variance. *Biometrika* 57, 19-46.

- WILKINSON, J.H. (1965). The Algebraic Eigenvalue Problem.
Oxford: Clarendon Press.
- WORTHINGTON, B.A. (1975). General iterative method for analysis of
variance when block structure is orthogonal.
Biometrika 62, 113-20.
- WYNN, H.P. (1970). The sequential generation of D-optimum experimental
designs. Ann. Math. Stat. 41, 1655-64.
- WYNN, H.P. (1972). Results in the theory and construction of
D-optimum experimental designs. J. R. Statist. Soc. B 34,
133-47.
- YATES, F. (1939). The recovery of inter-block information in variety
trials arranged in three dimensional lattices.
Ann. Eugen. 2, 136-56.
- YATES, F. (1940). The recovery of inter-block information in
balanced incomplete block designs. Ann Eugen. 10, 317-25.

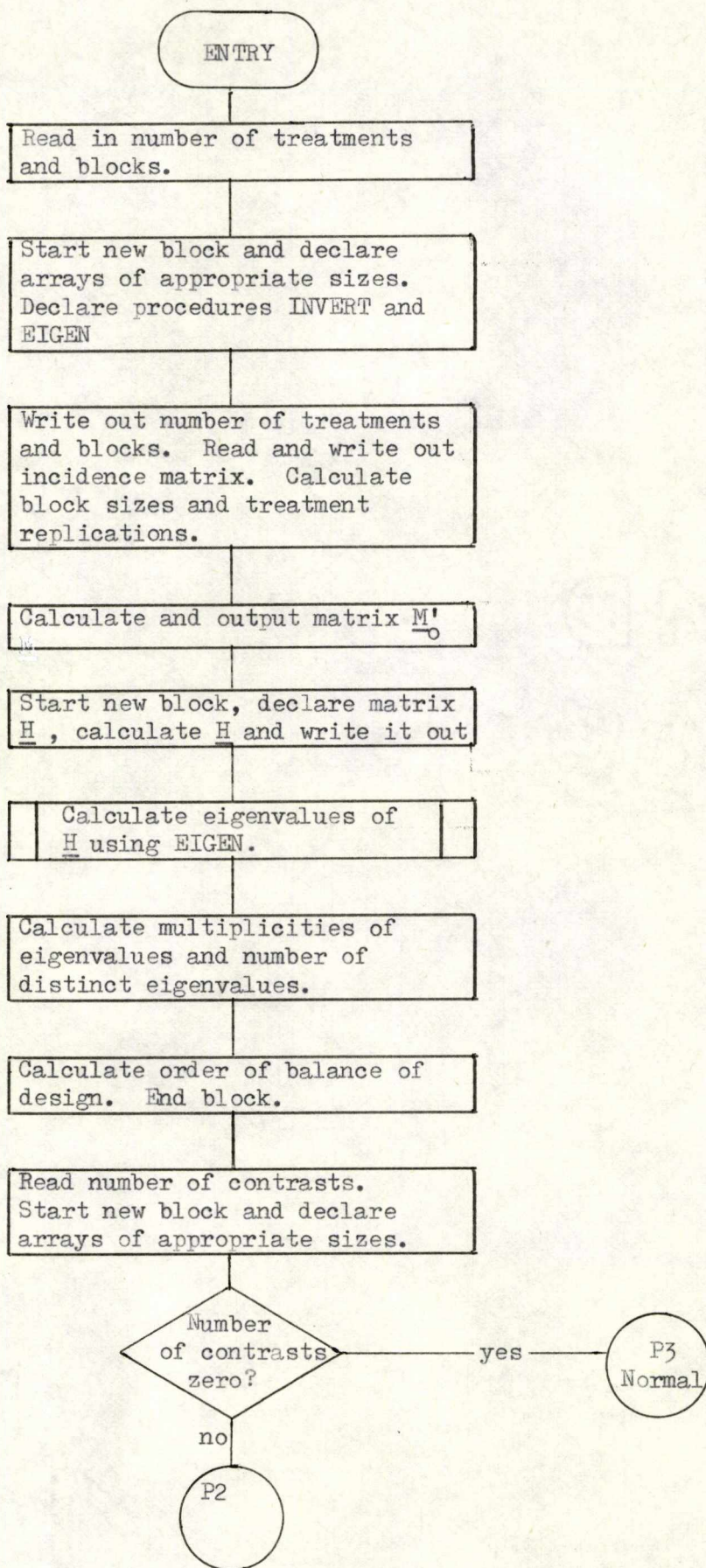
Acknowledgements.

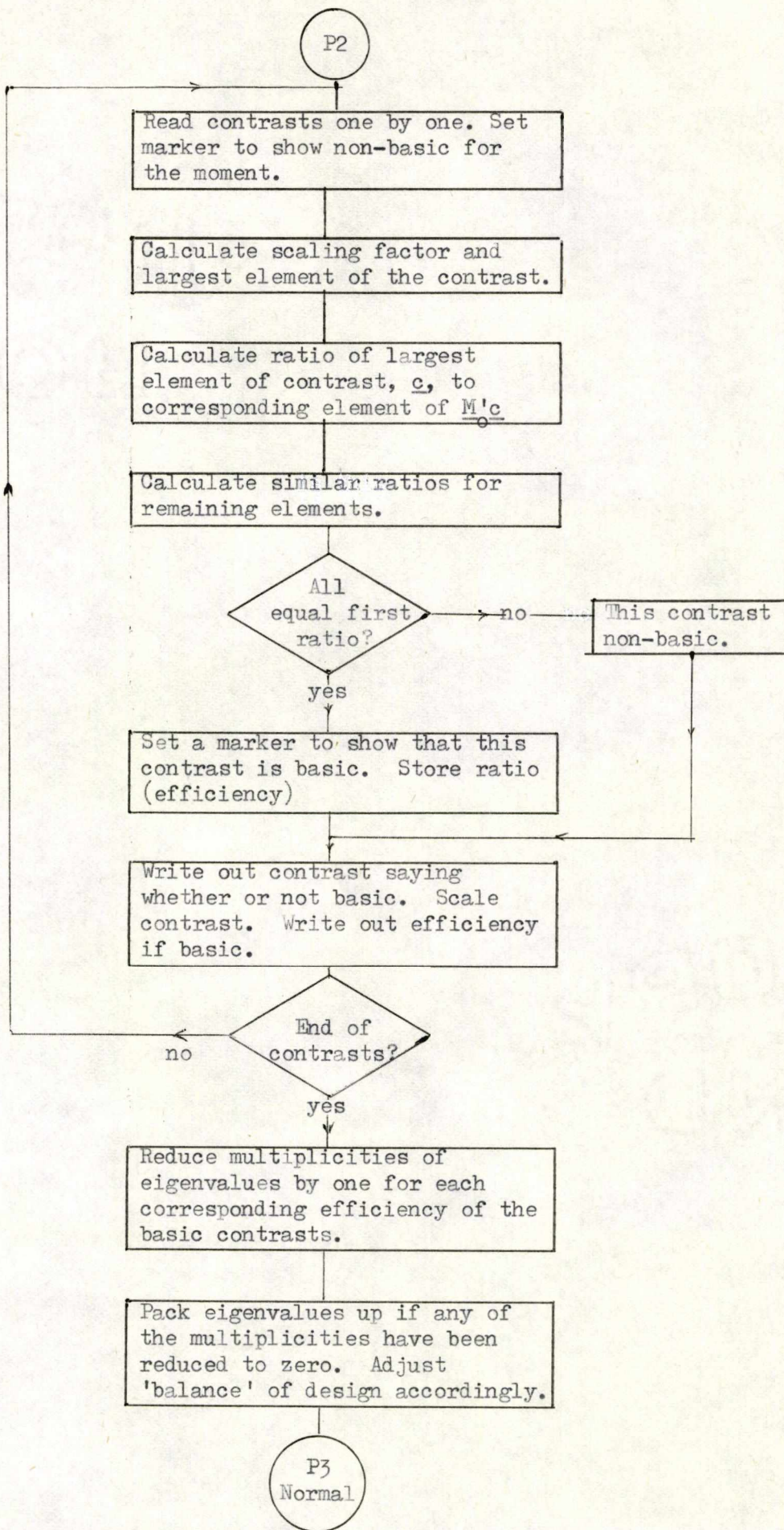
I thank my supervisor, Professor S. C. Pearce, for his valuable encouragement and advice, without which this thesis would never have appeared. I am also grateful to the Science Research Council under whose financial support the work was undertaken.

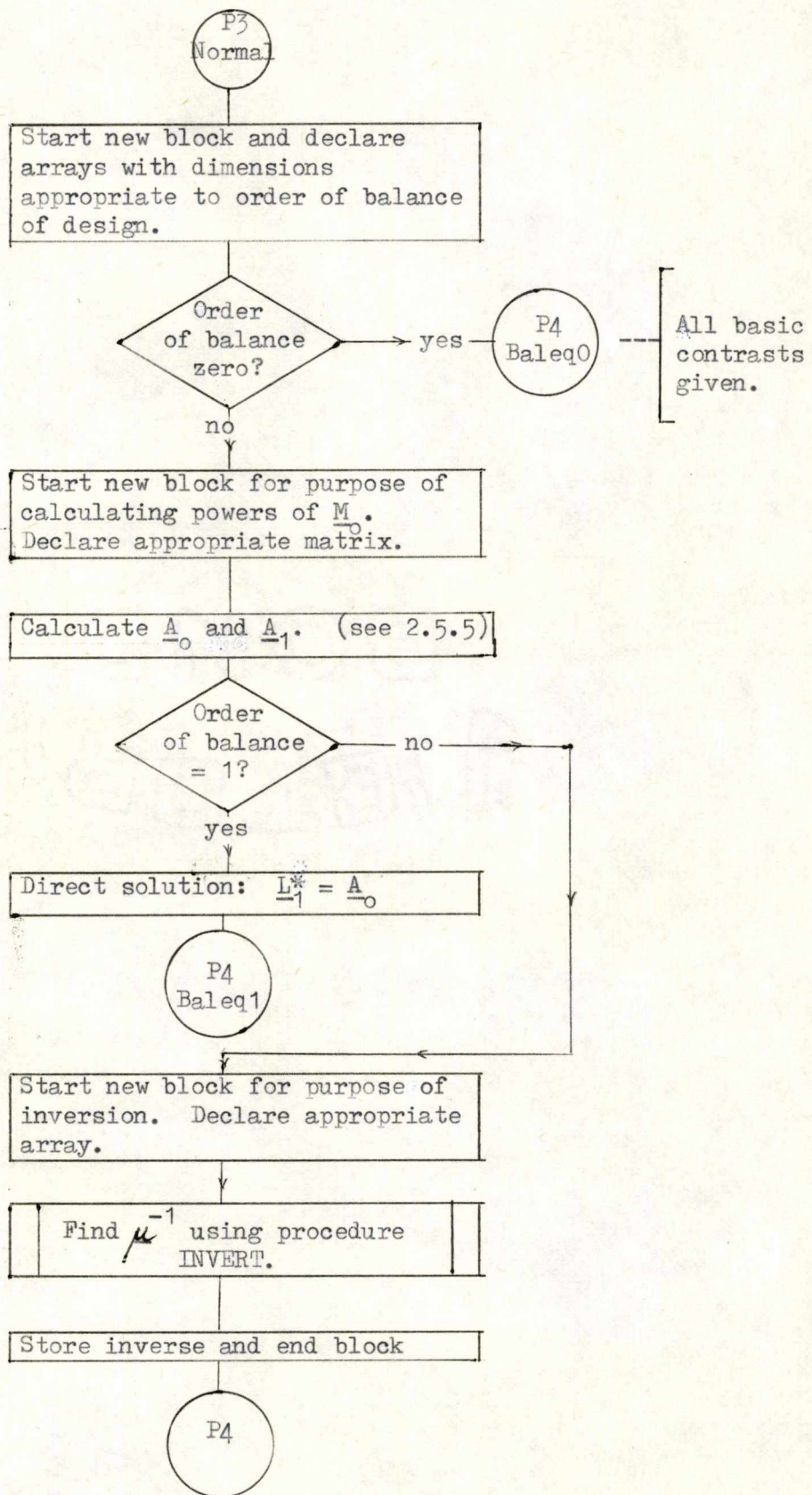
In particular, I am deeply grateful to my wife, who persuaded me to keep going and finish it the many times I felt like giving up.

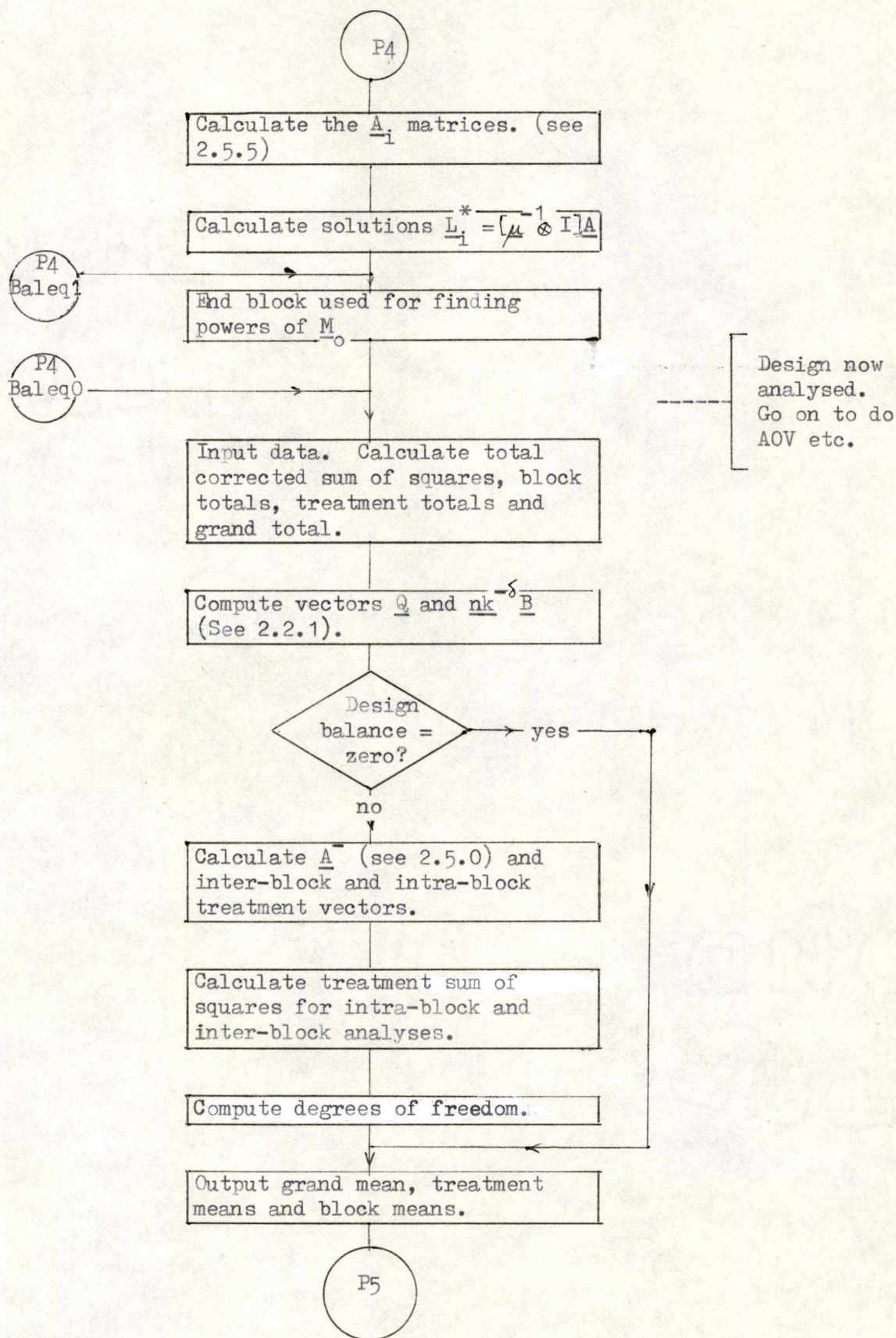


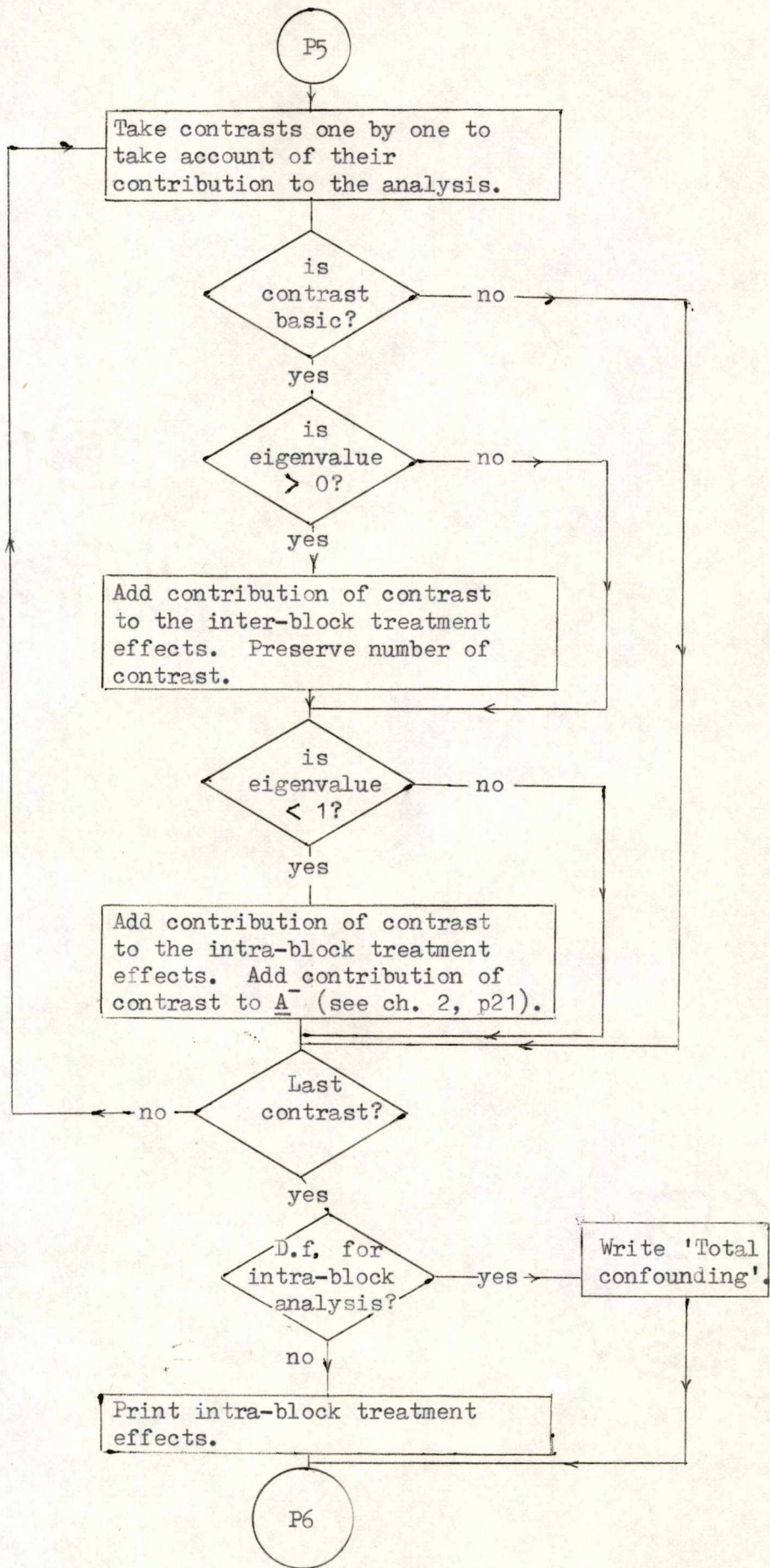
Appendix B Flowchart for program GEN

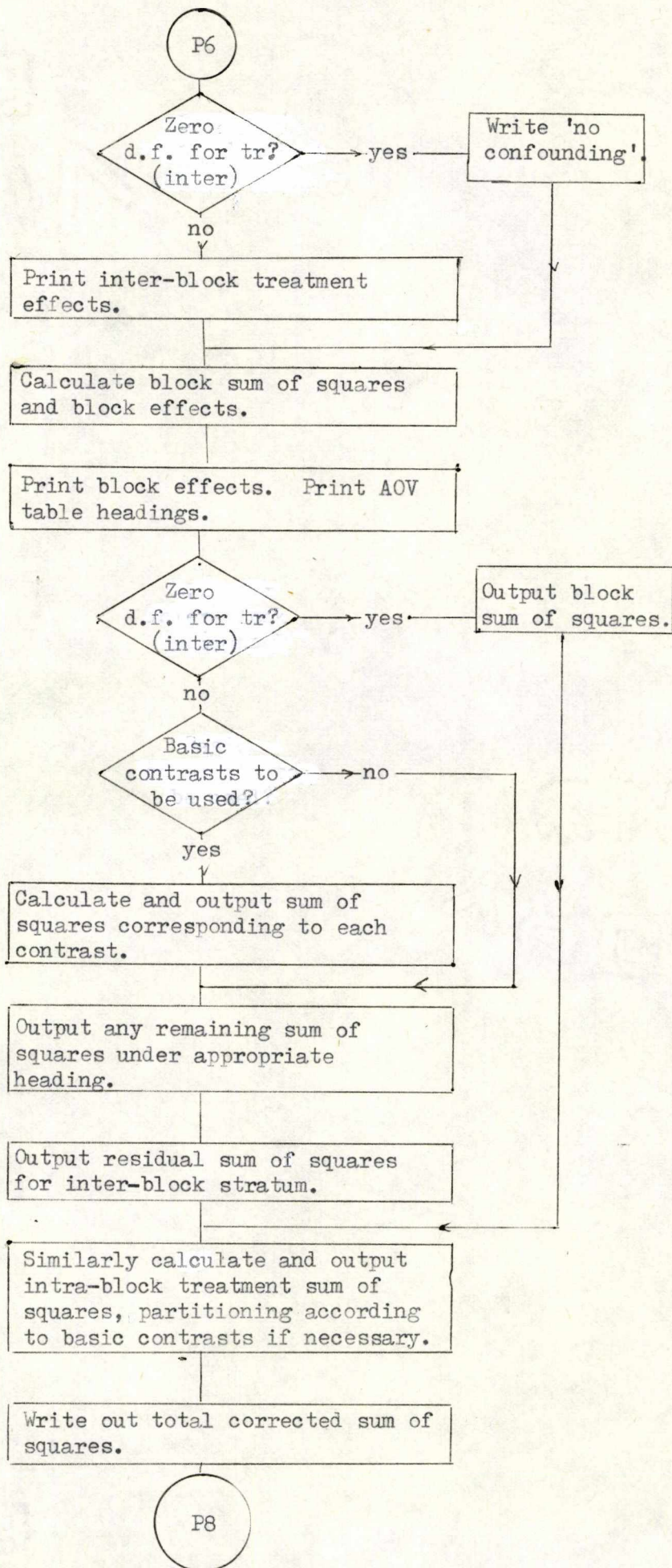


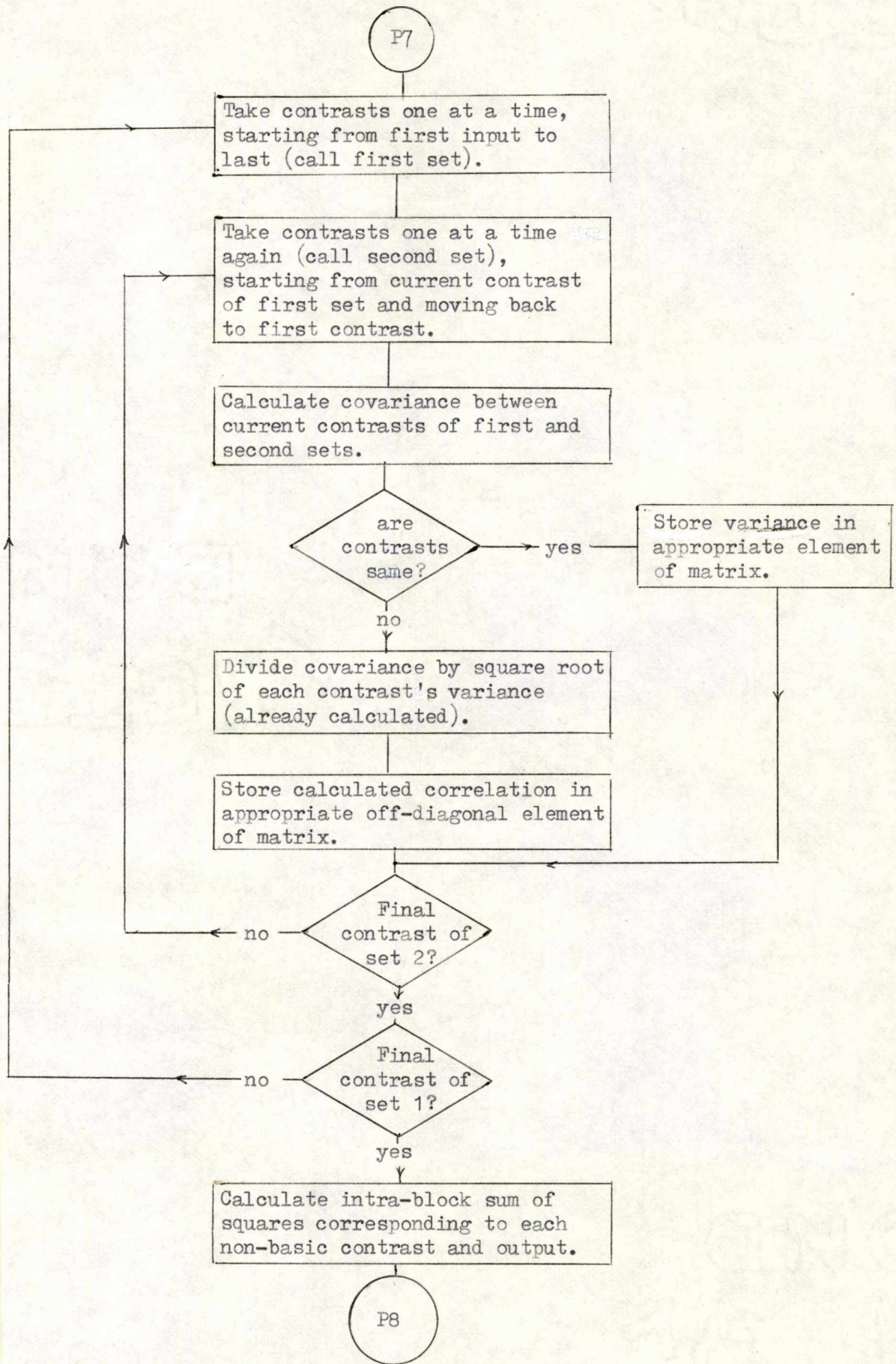


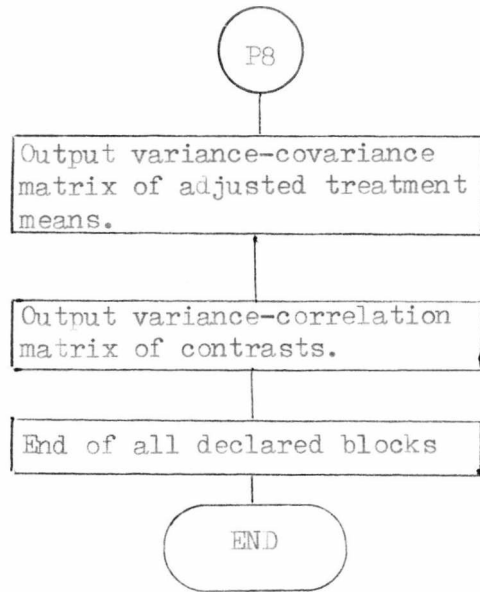






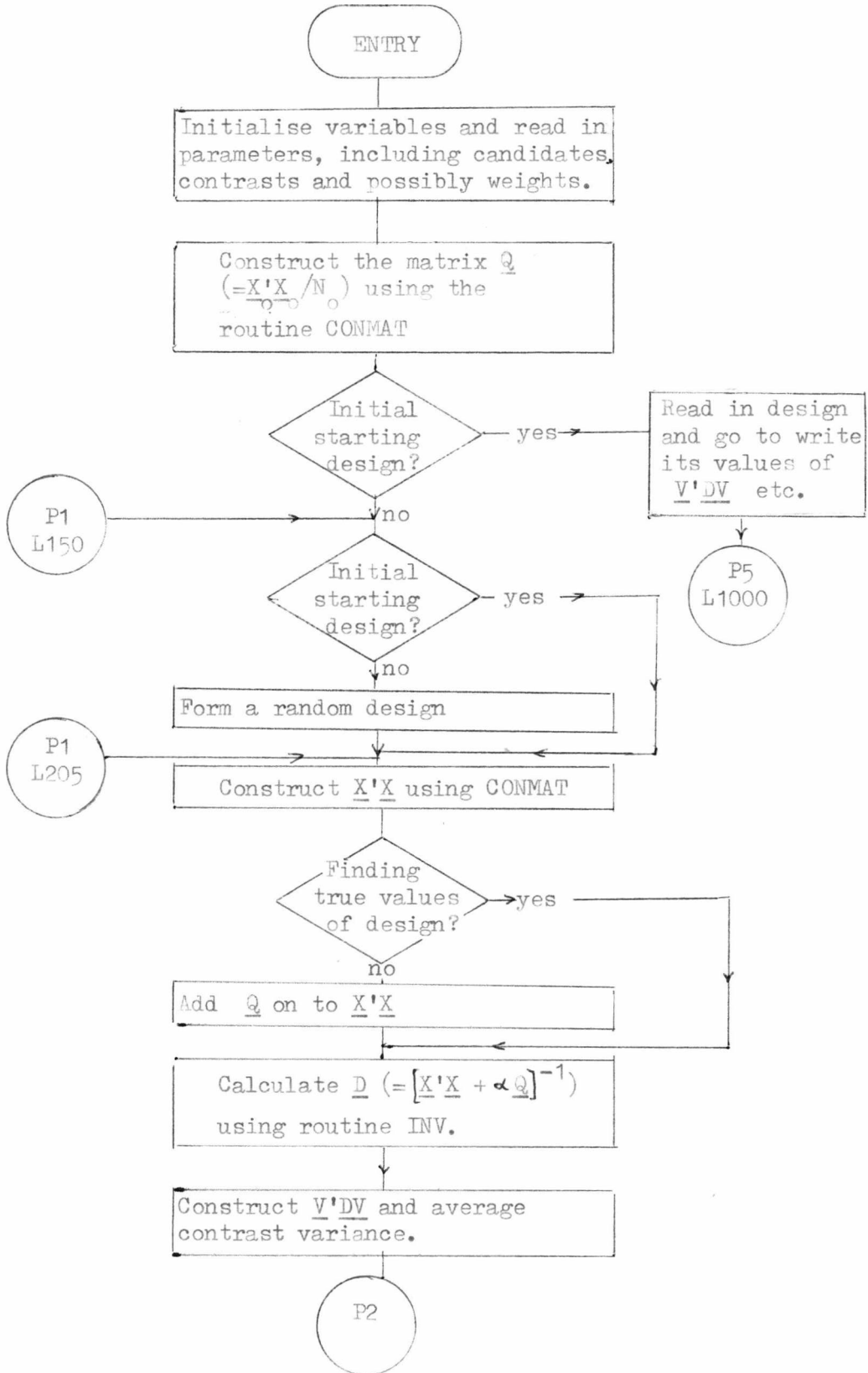


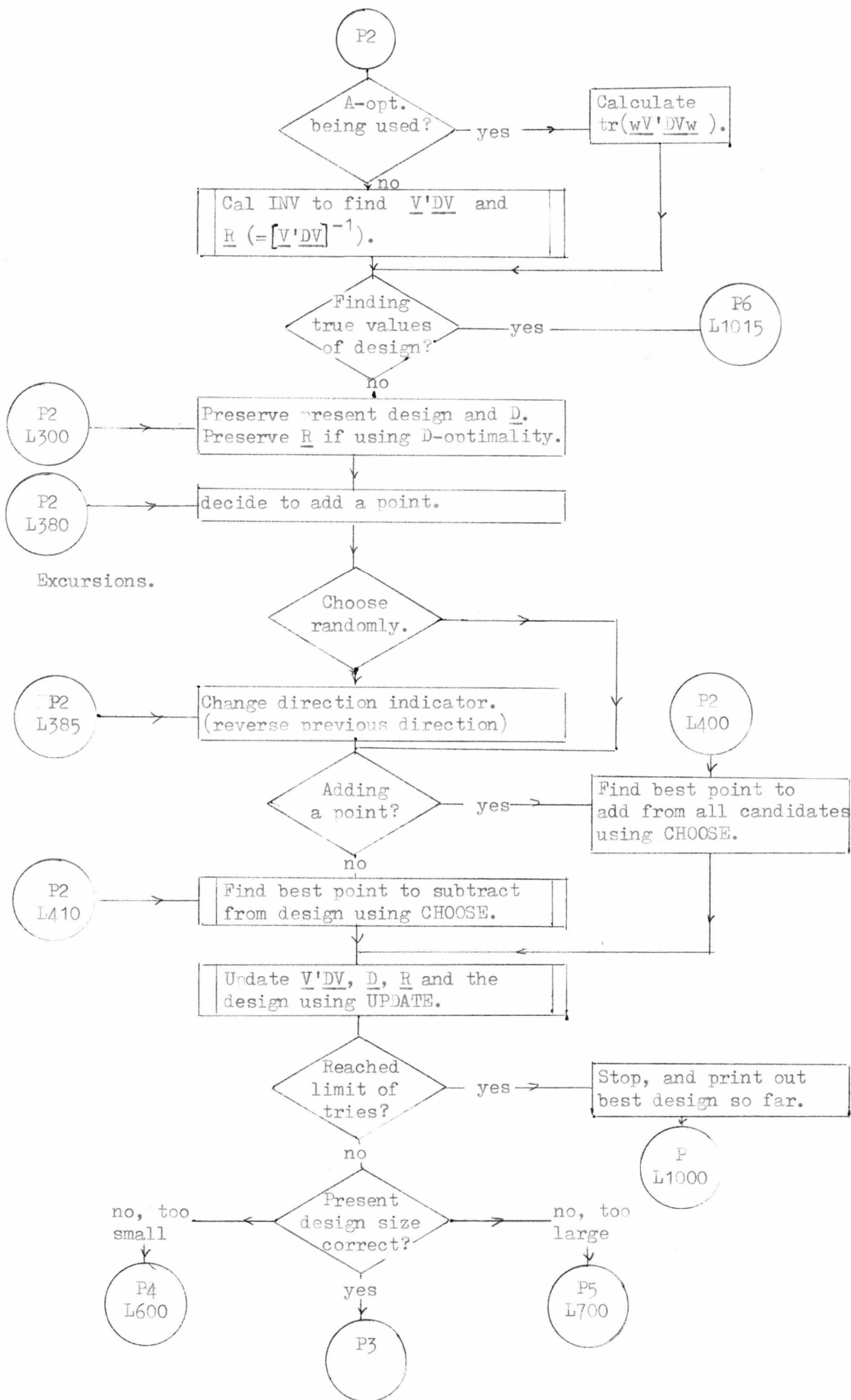




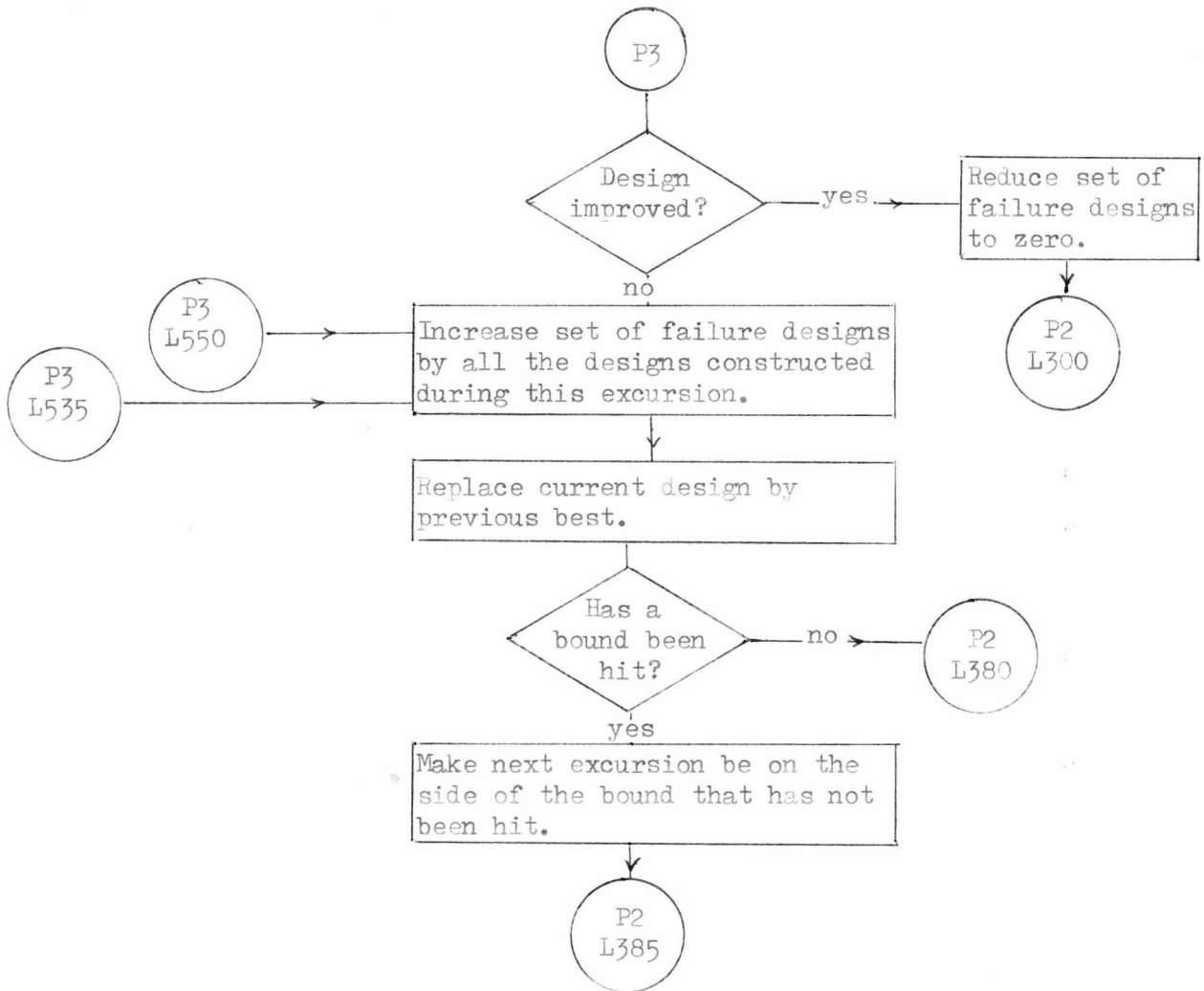
Appendix C Flowchart for program VDVMIN.

Initialisation.

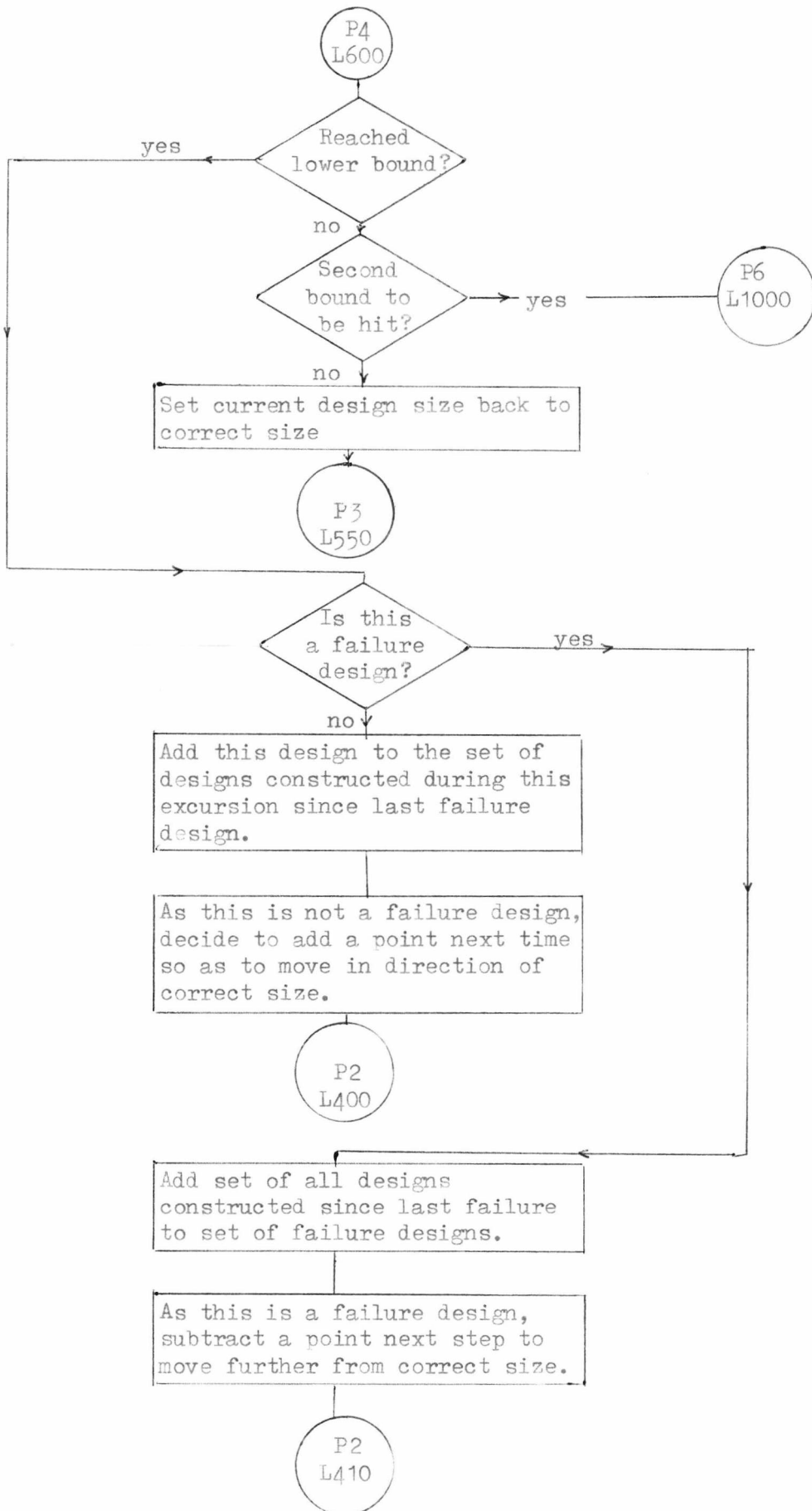




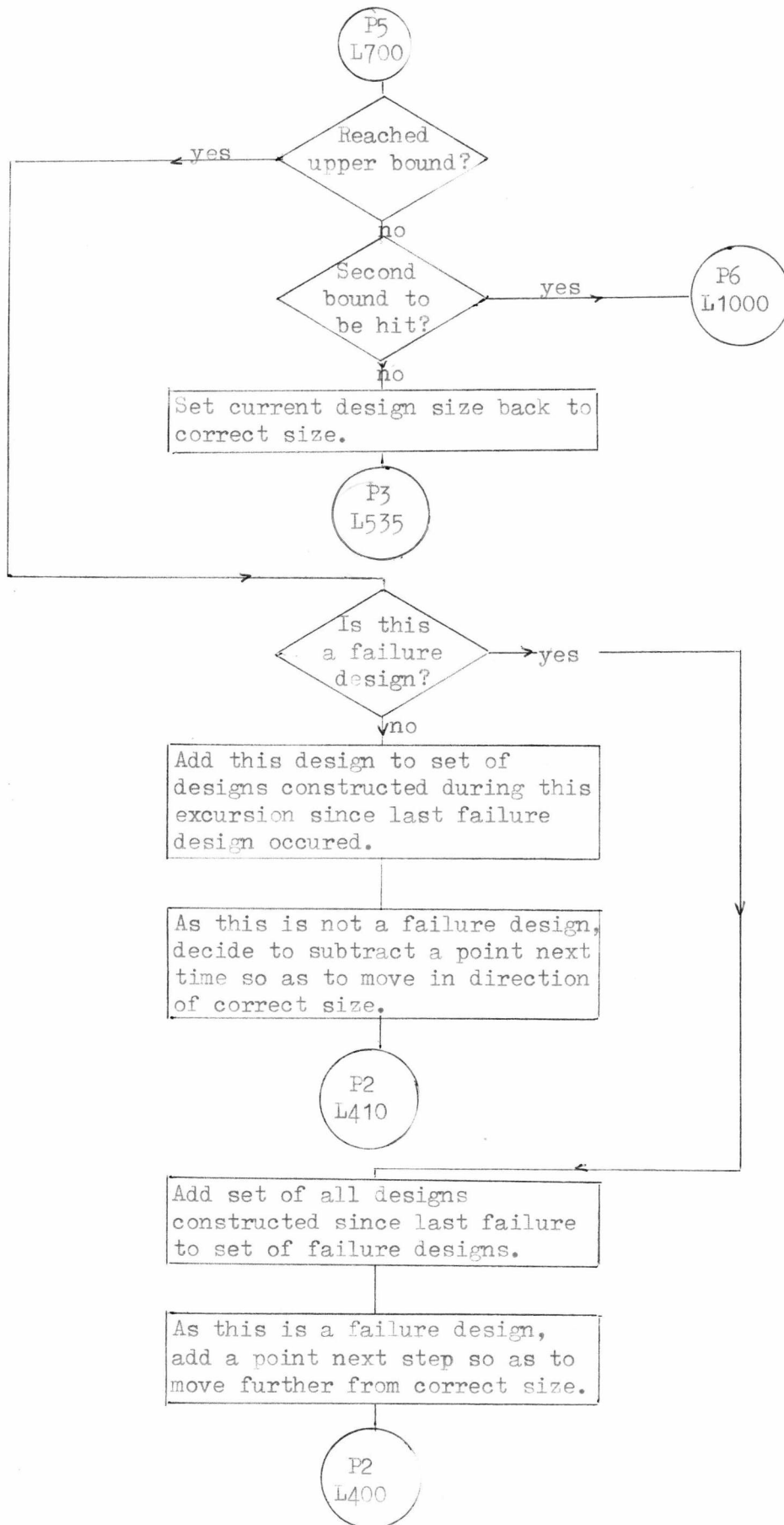
End of an excursion.



Excursions below correct size.

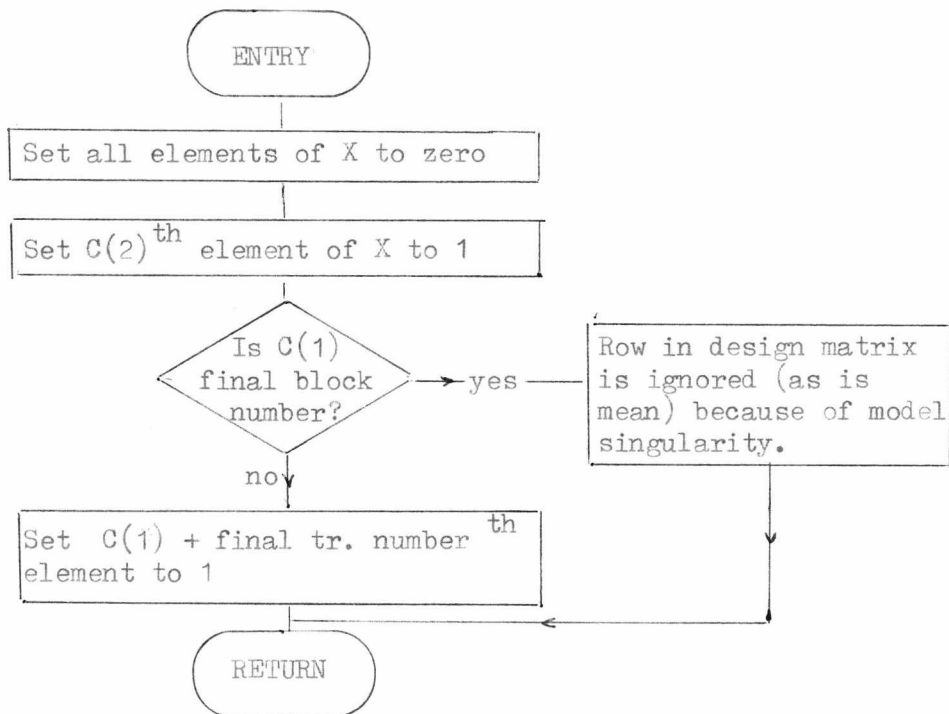


Excursions above correct size.



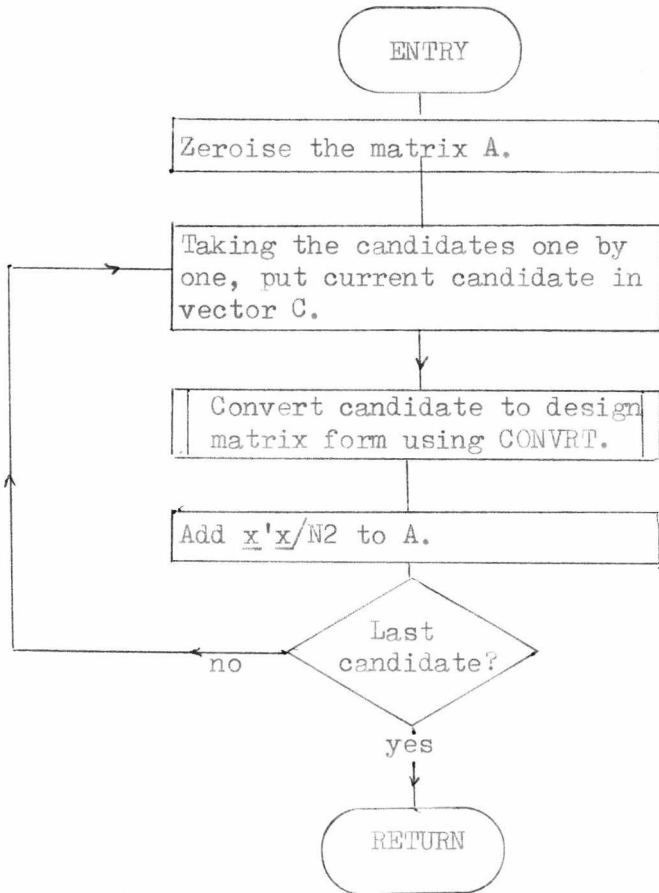
Subroutine CONVRT (C,X).

Converts a candidate, held in C, to a row x of the design matrix.
C(1) holds the block number, C(2) the treatment number.



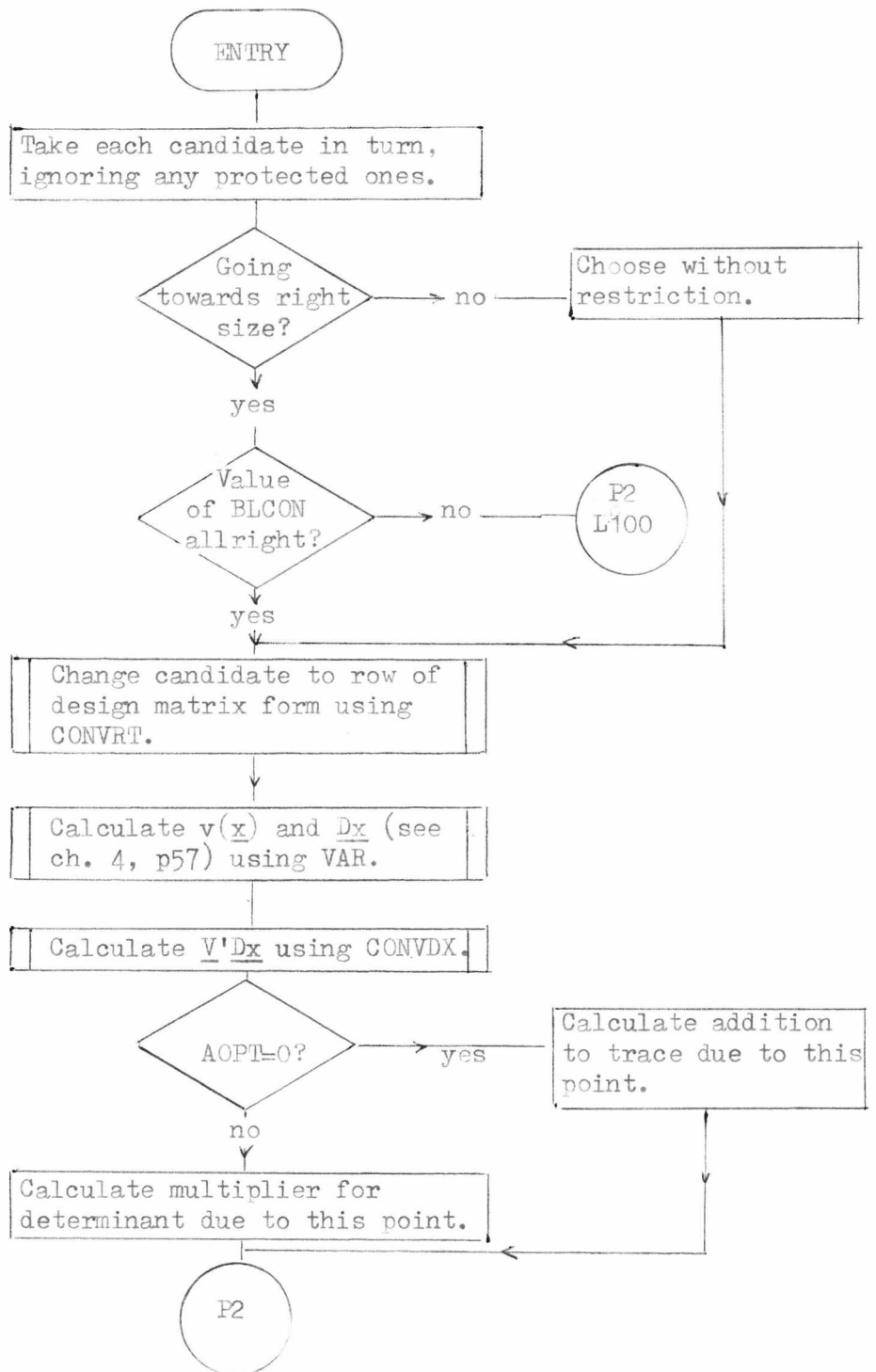
Subroutine CONMAT (A,E,N1,N2).

Forms the matrix $A = X'X/N2$, where X is made up from the $N1$ candidates held in E .

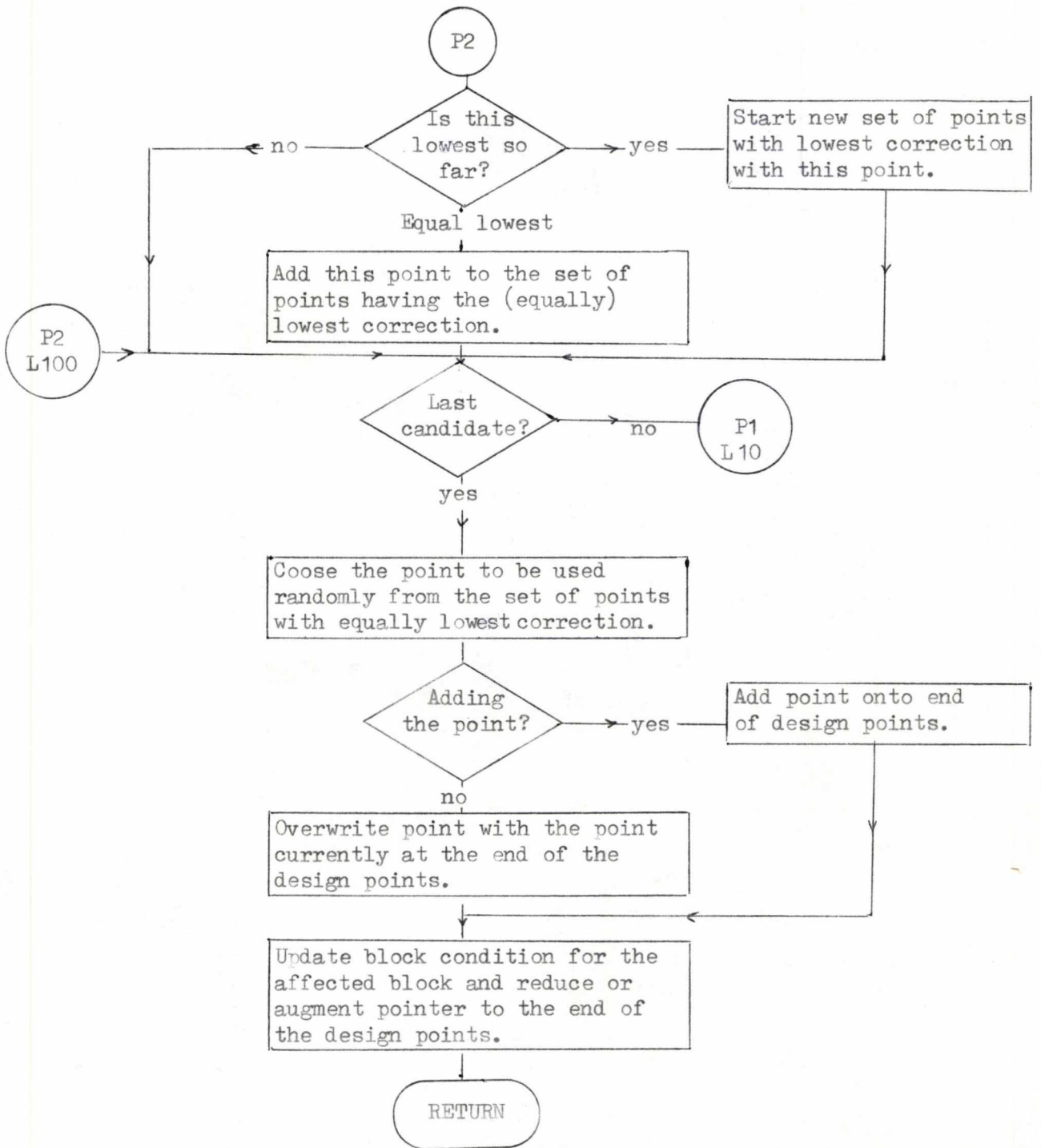


Subroutine CHOOSE (A,N1,C,W,BLCON,AOPT,WT).

This routine, depending on whether adding or subtracting a point, either adds the best point or takes away the worst, subject to the block sizes being correct, from the N1 candidates held in A. The point chosen is returned in C. BLCON holds the condition of the blocks (see 4.5). Its value for each block is (block size) - (correct block size). If AOPT is zero then A-optimality is being used.

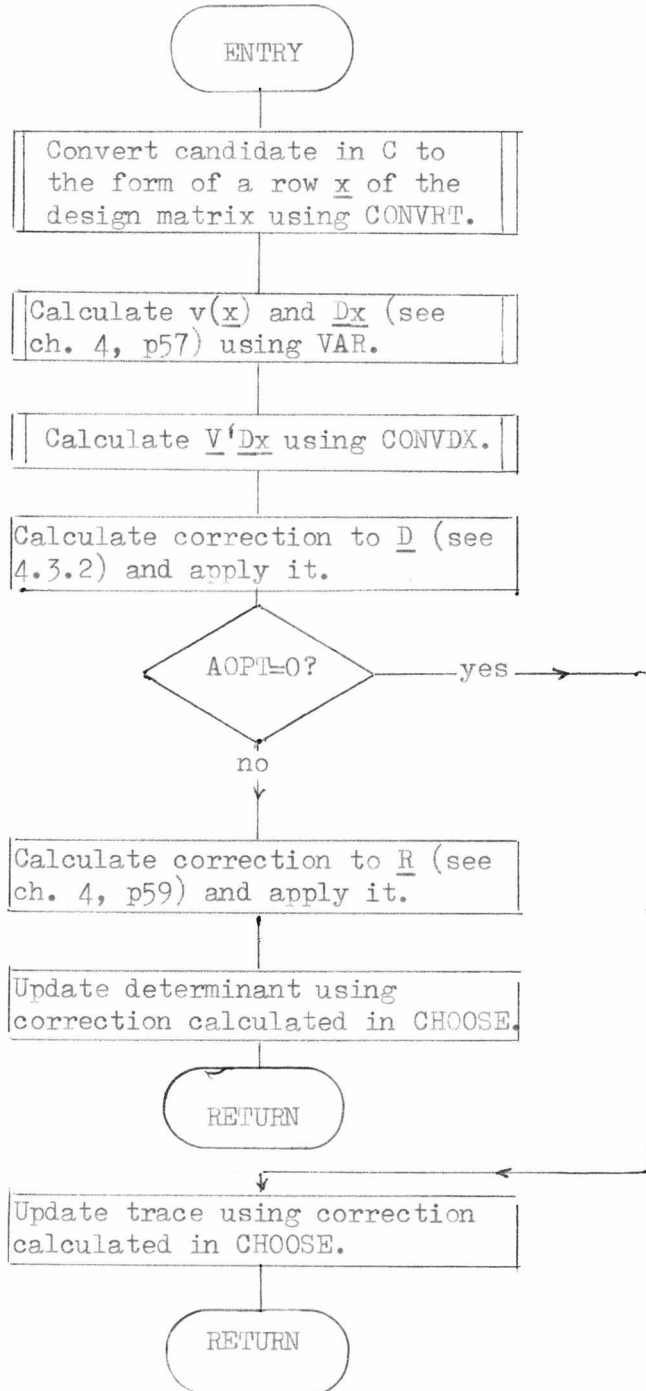


Subroutine CHOOSE (continued).



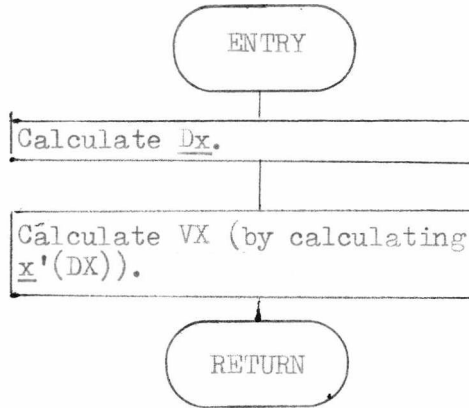
Subroutine UPDATE (C,DET,AOPT).

Updates the matrices \underline{D} and \underline{R} and the determinant, DET, of $(\underline{V}'\underline{D}\underline{V})$. The vector \underline{C} holds the point which has been removed or added. AOPT is zero if A-optimality is being used.



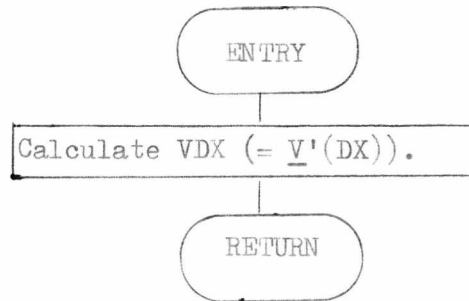
Subroutine VAR (X,DX,VX).

Computes \underline{Dx} and $\underline{x}'\underline{Dx}$ ($=v(\underline{x})$), where \underline{D} is the covariance matrix.



Subroutine CONV DX (VDX,DX).

Computes $\underline{V}'\underline{Dx}$, where \underline{x}' is row of design matrix.



Subroutines INV and RANDOM are standard routines.

