CONNECTEDNESS AND OPTIMALITY IN
MULTIDIMENSIONAL DESIGNS

by

Wilkie W. Chaffin

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

K. Hinkelmann

Dr. Klaus Hinkelmann, Chairman

Dr. Richard Krutchkoff
Department of Statistics

Clyde Y. Kramer

Dr. Clyde Kramer
Department of Statistics

Boyd Harshbarger

Dr. Boyd Harshbarger
Department of Statistics

R. H. Myers

Dr. Raymond Myers
Department of Statistics

May, 1975

Blacksburg, Virginia

ACKNOWLEDGEMENTS

I would like to express my appreciation to Dr. Klaus Hinkelmann for his guidance in this research and to Dr. Raymond Myers, who is responsible for much of the enthusiasm that I feel toward the study of statistics.

This thesis is dedicated to my wife, Dora, who deserves more credit for its completion, than do I.

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION.. 1
   1.1 Experimental Designs........................ 1
   1.2 Review of the Literature.................... 2

2. MINIMAL MULTIDIMENSIONAL DESIGNS, NO INTERACTIONS PRESENT.. 7
   2.1 Introduction................................ 7
   2.2 The General Approach for Obtaining MMD's............ 8
   2.3 Constructing Two-factor MMD's............... 10
   2.4 Three-factor Main Effect MMD's.............. 16

3. MINIMAL MULTIDIMENSIONAL DESIGNS, TWO-FACTOR INTERACTIONS PRESENT.. 25
   3.1 Introduction................................ 25
   3.2 Type II Connectedness....................... 26
   3.3 Model A--One Two-factor Interaction......... 28
   3.4 Model B--Two Two-factor Interactions........ 30
   3.5 Model C--Three Two-factor Interactions...... 39
   3.6 Advantages of the Procedure................. 44

4. MINIMAL AUGMENTED MULTIDIMENSIONAL DESIGNS............... 47
   4.1 Introduction................................ 47
   4.2 Augmenting Existing Designs................. 48
   4.3 MAMD's--Model A............................. 49
   4.4 MAMD's--Model B............................. 51
   4.5 MAMD's--Model C............................. 53

5. TYPES OF CONNECTEDNESS--S AND C-OPTIMALITY............... 58
   5.1 Introduction................................ 58
   5.2 Pseudo-global Connectedness................. 59
   5.3 S-optimal Designs........................... 60
   5.4 C-optimal Designs........................... 63

6. IMPROVING LOCALLY CONNECTED DESIGNS..................... 69
   6.1 Introduction................................ 69
   6.2 Pseudo-globally Connecting a Design for One Factor.. 69
   6.3 Optimizing with Respect to Both Factors............. 81

7. IMPROVING LOCALLY CONNECTED DESIGNS WHEN AN ADDITIONAL BLOCK EFFECT IS PRESENT............................ 88
   7.1 Introduction................................ 88
   7.2 Three-factor, Main Effect Designs........... 88
   7.3 The Effect of Pseudo-globally Connecting for Both Factors.................................... 90

TABLE OF CONTENTS (Continued)

CHAPTER PAGE

7.4 C-better Designs under the Three-factor Model.... 97

8. EFFECTS OF THE INTERCHANGE ALGORITHM WHEN THE ASSUMED MODEL IS INADEQUATE...................... 99
8.1 Introduction..................................... 99
8.2 Protecting Against an Inadequate Model........... 100

9. DESIGNS THAT ARE PSEUDO-GLOBALLY CONNECTED FOR LEVEL COMBINATIONS..................................... 106
9.1 Introduction..................................... 106
9.2 Type II Pseudo-globally Connected Designs........ 106

10. OPTIMAL MINIMAL AUGMENTED MULTIDIMENSIONAL DESIGNS..... 115
10.1 Introduction..................................... 115
10.2 Sequentially S-optimal Designs................... 118

11. SUMMARY.. 129

12. PROPOSED EXTENSIONS.................................. 132

BIBLIOGRAPHY.. 134

VITA.. 136

CHAPTER 1

INTRODUCTION

1.1 Experimental Designs

An experimental design may be considered to be an arrangement of the levels of certain factors in a plan of investigation (experiment). Such an arrangement may be useful in choosing an experiment that is efficient according to some criteria. A multi-factor, fixed effect design may be referred to as a multidimensional design (MD). These experimental designs may or may not contain all possible factor level combinations. In this work, we shall discuss the more general situation: designs which do not contain all factor level combinations (incomplete designs or fractional factorials).

There are many different criteria that one may use when choosing a plan of experimentation. Of greatest importance may be physical characteristics of the experiment. For instance, we may desire a design for which the data collection is simple or inexpensive. Probably this criterion is always considered to some extent, since otherwise we would no doubt use a complete design if it is possible to do so.

Then there are what we will call statistical criteria. We may require such design properties as balance or orthogonality of main effects. To achieve properties of this type, we often have to sacrifice with respect to one or more of the physical characteristics which we desire. Similarly, we may have difficulty achieving certain statistical criteria because of some inherent physical property of the experiment. For instance, we might find it impossible to use an orthogonal main effect plan because we are unable to make the assumption of no interaction.

Certainly in most real world situations, we must consider economy of experimentation. The time or money necessary to obtain the data may be at a premium. We need to choose a design that reflects this property of the experiment. Often what has been done in the past, is to try to find a design that provides for some economy, but still has certain features that make the analysis relatively simple. For example, we might take a fractional factorial instead of the complete design. Sometimes, however, we are willing to sacrifice balance or complete orthogonality in order to make more progress toward obtaining a design which allows a more economical experiment. Of the many approaches to this problem, we shall concentrate on only two in this work.

1.2 Review of the Literature

We shall be interested in the estimation of all main effect linear contrasts, and in the estimation of two-factor interaction contrasts when interactions are believed to exist. The concept of connectedness will be used throughout this work when discussing procedures to obtain designs that are useful when estimating contrasts.

Definition 1.1 (Bose, (1947))

"A treatment and a block are said to be associated if the treatment is contained in the block. Two treatments, two blocks, or a treatment and a block may be said to be connected if it is possible to pass from one to another by means of a chain consisting alternately of blocks and treatments such that any two [adjacent] members of the chain are associated. A design (or a portion of a design) is said to be a connected design (or a connected portion of the design) if every block or treatment of the design is connected to every other."

Anderson (1968) and Srivastava and Anderson (1970) have discussed the connectedness of multidimensional designs with respect to factor main effects. There will exist an unbiased estimator for every main effect linear contrast for a given factor under the usual additive model, if and only if the design is connected with respect to the factor. Sennetti (1972) has extended the concept of connectedness to models where we have assumed one or more two-factor interactions.

We have said that we desire a design which allows for as much economy of experimentation as possible. One approach to this problem, is to choose the design that will allow us to accomplish the task we have chosen with only the minimum number of design points possible. The problem of connecting a design when there is some restriction on the number of design points that may be used has been approached in many ways. For instance, the orthogonal main effect plans of Addelman and Kempthorne (1961) allow us to estimate main effect contrasts in each factor while allowing some economy in the number of required factor level combinations (assemblies).

Prior to 1972, there was little or no work in the literature that was directed specifically toward minimal designs. The term multidimensional design was first coined by Potthoff (1958). His work on construction and analysis of designs for which main effects and some interactions can be estimated, does consider to some extent economy of experimental units. Daniel (1971) presents a minimal 2^3 factorial plan for estimating main effects. Federer, Hedayat, and Raktoe (1972) also discussed minimal designs. All the work cited, and generally all the work that appeared in the literature prior to 1972, considered only certain families of designs.

Since the early 1960's, much has been written on augmenting existing designs. Daniel (1962), Addelman (1963), and John (1966), all discussed ways of adding factorial fractions to established designs in order to estimate contrasts not previously estimable or not previously assumed present. These works center on developing balanced fractions, rather than minimal designs. Federer (1961) develops a procedure for adding new factor levels to a design in order to be able to estimate contrasts that were nonestimable in the original design. His procedure accomplishes this with a minimum number of new assemblies. However, it seems reasonable that there will exist many situations where it is impractical or even impossible to obtain new additional levels of a given factor.

We will consider two approaches for obtaining a minimal design. One approach is to choose a completely new experiment using only those factor levels suggested to us by the experimenter. Using this approach, Sennetti (1972) develops a procedure for constructing a connected design using the minimum number of assemblies. That is, if the linear model associated with a multidimensional design, D, is

$$E(\underline{y}_D) = X_D\underline{p}_D$$

where $\underline{y}_D$ is the observation vector, X_D is the design matrix, and $\underline{p}_D$ is the parameter vector, then we need only $[\nu(\underline{p}_D)+1]$ assemblies to generate a connected multidimensional design. The term $\nu(\underline{p}_D)$ denotes the number of degrees of freedom for the parameter vector. Such a design is called a minimal multidimensional design (MMD). Of course the properties of complete orthogonality or balance may not hold for these designs.

The other approach to use in obtaining a minimal design would be to augment an existing design. The minimum number of new assemblies necessary to connect an existing design is $[\nu'(\underline{p}_D)-r(X_D)]$, where $r(X_D)$ denotes the rank of the X matrix and $\nu'(\underline{p}_D)$ is the degrees of freedom for the parameter vector and includes one degree of freedom for the mean. Such a design is called a minimal augmented multidimensional design (MAMD). We shall consider the use of MMD's and MAMD's in estimating linear contrasts in main effects and interaction type contrasts. In Chapters 2 and 3 we shall attempt to clarify and extend some of the work done by Sennetti (1972) in generating minimal designs. In Chapter 4 we will discuss connecting existing designs by adding certain selected design points.

There may be instances when, rather than emphasize economy in terms of number of observations, we would rather consider how we can increase the usefulness of the factor level replicates that we have in the design. That is, we want to be able to improve the quality of an existing design, when we are unable to increase the number of design points, or even to change the number of replicates of any factor level. We will try to change a connected design into a design that possesses a stronger type of connectedness, pseudo-global connectedness, defined by Eccleston and Hedayat (1974). This change will allow for more estimates for main effect linear contrasts. In addition, we will discuss how changing a design to a pseudo-globally connected design may improve the design with respect to certain optimality criteria. In particular, we shall discuss S-optimality as defined by Shah (1960) and a new optimality criterion, C-optimality. Chapter 5 through Chapter 8 of this work will concern pseudo-globally connecting designs and some of the properties of the resulting arrangements.

The next two chapters of this work will combine some of the concepts discussed in Chapters 2, 3, and 4 with some of those discussed in Chapter 5 through Chapter 8. In Chapter 9, we shall consider the quality of a design obtained by pseudo-globally connecting an existing connected design for level combinations. In particular, we will be interested in situations where the assumption of only one two-factor interaction is appropriate. In Chapter 10, we shall consider an optimal way of augmenting an existing design when obtaining a minimal augmented multidimensional design. Using the S-optimality criterion at each step in the augmentation procedure, we shall obtain what will be called the sequentially S-optimal MAMD.

CHAPTER 2

MINIMAL MULTIDIMENSIONAL DESIGNS, NO INTERACTIONS PRESENT

2.1 Introduction

In this chapter we shall consider the problem of generating a minimal multidimensional design (MMD). Under the assumption of no interaction, we want to be able to estimate all main effect linear contrasts in factor F_β, for $\beta=1, 2, \ldots, m$ for an m-factor design, while keeping the number of assemblies used in the design at a minimum. There will exist an unbiased estimator for all main effect linear contrasts if and only if the design chosen is a connected design. Thus we will seek a design that is connected but has the minimum number of design points possible.

Using the notation of Sennetti (1972), let D be a multidimensional design (MD), with m factors $F_1, F_2, \ldots, F_m$, with F_i having n_i levels. The linear model for D will be

$$E(\underline{y}_D) = X_D \underline{p}_D$$

Where $\underline{y}_D$ is the observation vector (h x 1), X_D is the design matrix (h x M) with $M=\sum_{i=1}^{m} n_i$, and $\underline{p}_D$ is the parameter vector (M x 1) which may be written as

$$\underline{p}_D' = (p_1^1, p_1^2, \ldots, p_1^{n_1}, p_2^1, p_2^2, \ldots, p_2^{n_2}, \ldots, p_m^1, p_m^2, \ldots, p_m^{n_m})$$

Here p_i^j denotes, for all i, the main effect of the j th level of factor F_i. We want to estimate all contrasts of the form $p_i^{\ell_i}-p_i^{\ell'_i}$, where ℓ_i and ℓ'_i are two levels of F_i. Recall that an MMD will allow us to do this while using only one more assembly than there are degrees of freedom for $\underline{p}_D$.

2.2 The General Approach for Obtaining MMD's

Sennetti (1972) has shown the existence of an MMD in m factors. We shall concentrate on specific procedures that may be used to obtain such a design. Sennetti suggests the following general steps for constructing an MMD in m factors: First construct a two-factor MMD, then a three-factor MMD, . . ., then an m-factor MMD. At each step in the procedure, four things should be considered, of which two are pertinent to the present discussion:

(a) To construct an MMD in k factors, we augment a design of k-1 factors.

(b) Starting with the MMD of k-1 factors, $F_1, F_2, \ldots, F_{k-1}$, if F_k has n_k levels, n_k-1 additional assemblies in the k-1 factors are added so that the resultant design has the following property: If F_β is any one of the k-1 factors already in the design, then

<u>Property I</u> $n_\beta-a_\beta$ levels of F_β occur V_β times and a_β levels of F_β occur $V_\beta+1$ times in the design.

(Parts (c) and (d) of Sennetti's considerations concern how the augmentation should be carried out so as to maintain property I.)

We will attempt to maintain property I for the k th factor as well as for the first k-1 factors. At this point we shall consider the justification for maintaining I (if possible) during the augmentation procedure.

The argument given by Sennetti (1972) may be summarized as follows:

Most of the commonly used MD's require that all level frequencies be the same for any given factor. Examples of this are complete block designs and balanced or partially balanced incomplete block designs. In addition, most orthogonal designs occur with factor levels (of a given factor) having a constant frequency. (There are instances when this is not true, however; see Addelman and Kempthorne (1961).) Finally, it seems reasonable that we want a fairly uniform variance for contrast estimators. This attempt to achieve equal level frequencies may establish some uniformity of variance on estimators of contrasts in factor F_β.

This quest for uniformity of variance seems rightly justified, since often we would have no way of deciding which contrasts are more important. In some designs which have equal level frequencies, we do obtain a uniform variance on all contrasts involving a given factor, as can be seen from the following definition and theorem:

Definition 2.1

A connected design is said to be variance balanced for a given factor if every effect of the factor is estimated with the same variance and every two effects of the factor with the same covariance. (Certainly, if a design is variance balanced for a given factor, then any main effect linear contrast in the factor can be estimated with the same variance.)

Theorem 2.1 (Rao, (1958))

Any binary, two-factor design that has all levels of one factor equally replicated will be variance balanced for the other factor if and only if the levels of the other factor are equally replicated. (A binary design is one in which every level combination occurs either zero times or one time in the design. All MMD's are binary designs.)

The theorem by Rao seems to justify trying to maintain property I for both factors in a two-factor design. In fact, in an m-factor design, if we consider level combinations of any (m-1) of the factors as being levels of some factor F_{D-i}, where F_i is the other factor, then we would seem to be justified in trying to maintain property I for factor F_i, for $i = 1, 2, \ldots, m$. There are instances when equal level frequencies do not insure uniformity of variance contrasts. Certainly though, the concept of equal level frequencies has intuitive appeal.

Sennetti (1972) describes a general method for constructing a minimal multidimensional design in two factors, that is connected for both factors. What we will present is essentially the same method, but with what seems to be a significant simplification in notation and some simplification in procedure. After presenting this procedure and giving several examples of its use, we will prove that use of this procedure always produces a design that is connected for both factors. Then we will consider a method for generating an MMD in three factors that is a simplification of Sennetti's three-factor method. We will prove that any design generated by this procedure is connected for all three factors.

2.3 Constructing Two-factor MMD's

Consider two factors F_1 and F_2. Factor F_1 has n_1 levels and factor F_2 has n_2 levels, where $n_1 \geq n_2$. We wish to construct a minimal multidimensional design that is connected for factor F_1 and factor F_2. If the design is to be minimal, we shall need only one more assembly than there are degrees of freedom for F_1 and F_2. That is, we need $(n_1-1) + (n_2-1) + 1 = n_1 + n_2 - 1$ assemblies. First we choose the (n_1+n_2-1) levels of F_1. Then we show how the levels of F_2 are chosen to occur with the chosen levels of F_1. The procedure to be used to obtain the MMD is as follows:

(1) Certainly all n_1 levels of F_1 must be used once. Then any n_2-1 levels of factor F_1 may be chosen to be repeated.

(2) Any n_2-1 of the n_2 levels of factor F_2 may be paired with the n_2-1 repeated levels of F_1 to give n_2-1 of the assemblies.

(3) To obtain the assemblies that contain the second replicates of the repeated levels of F_1, we attach to each of these levels, the level of F_2 that is one higher (mod n_2) than the level of factor F_2 previously attached to the level of F_1 in question.

(Only in step (3) are we restricted somewhat in the assemblies that we choose; however, this restriction rests solely on what we did in (1) and (2), steps upon which no restrictions were placed.)

(4) Now we attach any levels of F_2 to the remaining levels of F_1, just being careful to maintain property I of section 2.2 for factor F_2. (1) above guarantees that property I is maintained for F_1.

Example 2.1

Factor A has 4 levels, factor B has 3 levels. The minimal number of assemblies will be $n_a+n_b-1 = 6$. We will use all 4 levels of A, but then we can repeat any $n_b-1 = 2$ levels of A that we wish. Suppose we repeat levels 3 and 4 of A. The levels of A in our 6 assemblies will be

a_1 a_3

a_2 a_4

a_3

a_4

We choose any two levels of B, say b_1 and b_3, for the assemblies that contain the first replicates of the repeated levels of factor A. At this point we have

a_1 a_3

a_2 a_4

a_3b_1

a_4b_3

Next we have to pair b_2 with the second replicate of a_3 and b_1 (which is b_{3+1} mod 3) with the second replicate of a_4. This gives

a_1 a_3b_2

a_2 a_4b_1

a_3b_1

a_4b_3

As the final step, we assign the other two levels of B so as to maintain property I for B. One way to do this would be

a_1b_2 a_3b_2

a_2b_3 a_4b_1

a_3b_1

a_4b_3

This final design is connected for factor A and for factor B since any two levels of A are connected by a chain, as are any two levels of B. We also note that no assembly is wasted because it is repeated.

Example 2.2

Factor A has 7 levels, factor B has 7 levels. The number of assemblies needed will be (7+7-1)=13. If the repeated levels of A are levels 1 through 6, then we get, by using step (1) of the procedure,

a_1	a_1
a_2	a_2
a_3	a_3
a_4	a_4
a_5	a_5
a_6	a_6
a_7	

If we choose to use the first six levels of factor B with the replicated levels of A, then we get

a_1b_1	a_1
a_2b_2	a_2
a_3b_3	a_3
a_4b_4	a_4
a_5b_5	a_5
a_6b_6	a_6
a_7	

Then we attach the appropriate levels of B to the second replicates of the repeated A levels.

a_1b_1	a_1b_2
a_2b_2	a_2b_3
a_3b_3	a_3b_4
a_4b_4	a_4b_5
a_5b_5	a_5b_6
a_6b_6	a_6b_7
a_7	

To obtain our final design we use b_1 or b_7 with a_7 in the seventh assembly. Either choice of these levels will allow us to maintain property I for factor B. If we use level b_7, then level b_1 will be replicated one time, and all other levels of B will be replicated twice. The final design, which is an MMD that is connected for both factors, is as follows:

a_1b_1	a_1b_2
a_2b_2	a_2b_3
a_3b_3	a_3b_4
a_4b_4	a_4b_5
a_5b_5	a_5b_6
a_6b_6	a_6b_7
a_7b_7	

For simplicity, we will usually denote the (n_b-1) repeated levels of factor A by a_1, a_2, . . ., a_{n_b-1}, as we did in this example.

In either of the examples just given, we can easily specify the chain connecting any two levels of A or any two levels of B. For instance, in Example 2.2, the chain $(a_2, b_3, a_3, b_4, a_4)$ connects levels a_2 and a_4 of A.

We shall now prove that this procedure does in general, what we have illustrated in these two examples.

Theorem 2.2

Any minimal two factor design that is generated by the outlined procedure is connected for both factors and fulfills the following requirement for factor F_β, where F_β can be either of the factors:

Property I $n_\beta - a_\beta$ levels of F_β occur V_β times and a_β levels of F_β occur $V_\beta+1$ times in the design.

Proof:

We will call the two factors A and B, where the number of levels of factor A is greater than or equal to the number of levels of factor B. That is, $n_a \geq n_b$. The procedure itself guarantees that property I is maintained for both factors. We only have to show that any design generated by the outlined procedure is connected for both factors.

Under the procedure, (n_b-1) levels of factor A are replicated in the design. Since $n_a \geq n_b$, we know that at least (n_b-1) levels of factor B are replicated; these (n_b-1) levels of B occurring with the (n_b-1) levels of A that are replicated in the design. If it exists, denote the non-replicated level of factor B by b_1. If there is no non-replicated level of B, let b_1 be any level of B. We see that every level of B, except perhaps b_1, will occur with two levels of A; one of these two levels of A will be replicated, occurring with two different levels of factor B. Denote by a_1, the replicated level of A with which b_1 occurs in the design. Denote by b_2, the other level of B with which a_1 occurs. Denote by a_2, the other level of A with which b_2 occurs, and by b_3, the other level of B with which a_2 occurs. We continue this process until we denote by b_{n_b}, the other level of B with which a_{n_b-1} occurs in the design.

Consider then, the chain $(b_1, a_1, b_2, a_2, b_3, a_3, \ldots, b_{n_b-1}, a_{n_b-1}, b_{n_b})$. Certainly this chain connects levels b_1 and b_{n_b} of factor B. Using the notation above, this chain can be obtained from any design that is generated by our procedure. Note that this chain contains all levels of factor B. Eccleston (1972) proved that a design is connected for a factor if and only if there exists a chain between two levels of the factor that contains all levels of either that factor or the other factor. Thus our design is connected for factor B. Eccleston and Hedayat (1972) showed that if a two-factor design is connected with respect to one factor, then it is connected with respect to the other factor also. Since our design is connected for factor B, it must also be connected for factor A.

Theorem 2.2 assures us that we can always construct a minimal, two-factor design, from which we can estimate all main effect, linear contrasts in either factor (since the design is connected for both factors).

2.4 Three-factor Main Effect MMD's

Next we will consider the problem of generating an MMD in three factors. In particular, our interest will lie in situations where there are two factors of major interest and a third factor of lesser interest or no interest to us. We will still be able to estimate all main effect, linear contrasts in all factors, but we may not be able to maintain property I for the third factor. In some instances, though not all, this third factor would be blocks. As in the two-factor situation, we assume that no interactions exist.

Consider three factors, A, B, and C, with n_a, n_b, and n_c levels respectively. Suppose that factors A and C are the factors of greatest importance, with $n_a \geq n_c$, and B is the factor for which we may not be able to maintain property I. The MMD we are seeking will have $(n_a-1) + (n_b-1) + (n_c-1) + 1 = n_a + n_b + n_c - 2$ assemblies.

First we will consider the level combinations of factors A and B. The first (n_a+n_b-1) assemblies will have the same levels of A and B as we chose in the two-factor case. If $n_a \geq n_b$, then these combinations (using the notation of Theorem 2.2) will be

$$
\begin{array}{lll}
a_1 \, b_1 & a_{n_b+1} b_{\theta_1} & a_1 \, b_2 \\
a_2 \, b_2 & a_{n_b+2} b_{\theta_2} & a_2 \, b_3 \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
a_{n_b} b_{n_b} & a_{n_a} \, b_{\theta_{n_a-n_b}} & a_{n_b-1} b_{n_b}
\end{array}
$$

Here the b_{θ_j} for j=1, 2, . . ., n_a-n_b are arbitrary levels of factor B that will allow us to maintain property I for factor B at this stage.

We need (n_c-1) more level combinations of A and B. It may not be possible to choose (n_c-1) of the original (n_a+n_b-1) level combinations and still be able to maintain property I for both A and B. Thus level combinations are replicated that will allow us to maintain property I for the first factor of prime importance, factor A. Note that a_{n_b}, a_{n_b+1}, a_{n_b+2}, . . ., a_{n_a} have only been used one time so far. These levels must be used in the last (n_c-1) assemblies before any other levels of A can be used. One way to choose the last (n_c-1) level combinations of A and B would be

$a_{n_b} b_{n_b}$

$a_{n_b+1} b_{\theta_1}$

$a_{n_b+2} b_{\theta_2}$

.

.

.

$a_{n_b+n_c-2} b_{\theta_{n_c-2}}$

That is, we choose the n_b th combination and the next (n_c-2) combinations from our original set.

Next the levels of factor C are attached in a manner that is similar to the method used to attach the levels of the second factor in the two-factor case. The first level of C, c_1, will be used with combination $a_{n_b} b_{n_b}$. The second level of C, c_2, is attached to combination $a_{n_b+1} b_{\theta_1}$ and c_3 is attached to $a_{n_b+2} b_{\theta_2}$. The process is continued until we attach level c_{n_c-1} to combination $a_{n_b+n_c-2} b_{\theta_{n_c-2}}$. Then, just as would have been done under Theorem 2.2, the second through n_c th levels of factor C are attached to the last n_c-1 (A x B) level combinations. Any levels of C may be used in the remaining assemblies. The only restriction on how we attach levels of C to the remaining combinations is that we do so in a manner that will guarantee that property I is maintained for factor C.

If $n_b > n_a$, we have to adjust our procedure slightly. After obtaining the first n_a+n_b-1 (A x B) level combinations, we need to repeat (n_c-1) of the combinations while maintaining property I for factor A. Certainly, the combinations we choose will depend on what levels of A were used with the unreplicated levels of B, $b_{n_a}, b_{n_a+1}, \ldots, b_{n_b}$. Example 2.4 will illustrate that no difficulty arises in choosing suitable (A x B) level combinations to repeat. After we have chosen the (n_c-1) combinations to replicate, the rest of the procedure is carried out as before. That is, we attach levels $c_1, c_2, \ldots, c_{n_c-1}$ to the first replicates of the repeated (A x B) combinations and levels $c_2, c_3, \ldots, c_{n_c}$ to the second replicates of the repeated combinations. Levels of C are attached to the remaining (A x B) combinations so as to maintain property I for factor C. Several examples will illustrate the procedure fully.

<u>Example 2.3</u>

Factor A has 10 levels, factor C has 3 levels, factor B (the one for which property I may not be maintained) has 7 levels. First we give the $n_a+n_b-1 = 16$ level combinations of factors A and B.

a_1b_1	$a_8\ b_1$	a_1b_2
a_2b_2	$a_9\ b_2$	a_2b_3
a_3b_3	$a_{10}b_3$	a_3b_4
a_4b_4		a_4b_5
a_5b_5		a_5b_6
a_6b_6		a_6b_7
a_7b_7		

Next we repeat the $n_b = 7$ th combination and the next $n_c-2 = 1$ combinations. These combinations are a_7b_7 and a_8b_1. We can now append the levels of factor C by the procedure outlined. First levels c_1, c_2, c_3, . . ., c_{n_c-1} of factor C are attached to combinations $a_{n_b}b_{n_b}$, $a_{n_b+1}b_{\theta_1}$, $a_{n_b+2}b_{\theta_2}$, . . ., $a_{n_b+n_c-2}b_{\theta_{n_c-2}}$ (in this case to combinations a_7b_7 and a_8b_1). At this point we have for assemblies 7 and 8, the following:

$a_7\ b_7c_1$

$a_8\ b_1c_2$

Next levels c_2 through c_{n_c} of factor C are attached to the last $n_c-1 = 2$ assemblies, obtaining

$a_7\ b_7c_2$

$a_8\ b_1c_3$

As the final step, we attach levels of C to the remaining combinations that will insure that property I is satisfied for factor C. One way to do this would be as follows:

$a_1\ b_1c_1$	$a_9\ b_2c_3$	$a_1\ b_2c_2$
$a_2\ b_2c_2$	$a_{10}b_3c_1$	$a_2\ b_3c_3$
$a_3\ b_3c_3$		$a_3\ b_4c_1$
$a_4\ b_4c_1$		$a_4\ b_5c_2$
$a_5\ b_5c_2$		$a_5\ b_6c_3$
$a_6\ b_6c_3$		$a_6\ b_7c_1$

The final design, which is connected for all three factors, is then

$a_1\ b_1c_1$	$a_{10}b_3c_1$
$a_2\ b_2c_2$	$a_1\ b_2c_2$
$a_3\ b_3c_3$	$a_2\ b_3c_3$
$a_4\ b_4c_1$	$a_3\ b_4c_1$
$a_5\ b_5c_2$	$a_4\ b_5c_2$
$a_6\ b_6c_3$	$a_5\ b_6c_3$
$a_7\ b_7c_1$	$a_6\ b_7c_1$
$a_8\ b_1c_2$	$a_7\ b_7c_2$
$a_9\ b_2c_3$	$a_8\ b_1c_3$

Note that levels a_1 through a_8 of factor A occur twice, while level a_9 and level a_{10} occur once. Thus property I is satisfied for factor A. All levels of factor C occur six times. Thus property I is satisfied for factor C as well.

Example 2.4

Suppose that $n_a=4$, $n_b=6$, and $n_c=3$; factor B will be the one for which we may not be able to maintain property I. Since $n_b \geq n_a$, we first give the levels of factor B to be used in the first (n_a+n_b-1) assemblies These are

b_1	b_1
b_2	b_2
b_3	b_3
b_4	
b_5	
b_6	

Next we attach the appropriate levels of factor A to get the first nine two-factor combinations:

a_1b_1	a_2b_1
a_2b_2	a_3b_2
a_3b_3	a_4b_3
a_4b_4	
a_1b_5	
a_2b_6	

We now choose any $n_c-1 = 2$ combinations to be repeated that will allow us to maintain property I for factor A. If we choose a_3b_3 and a_4b_4, then level a_1 of A will occur twice in the design, while all other levels of A will occur three times. Thus property I will be satisfied for factor A. If we attach the first through $n_c-1 = 2$ nd levels of C to the first replicates of these combinations and levels c_2 through $c_{n_c}=c_3$ of C to the last replicates of these combinations, we get the assemblies

$a_3b_3c_1$

$a_4b_4c_2$

$a_3b_3c_2$

$a_4b_4c_3$

Next the levels of C are attached to the remaining combinations so as to maintain property I for factor C. One way to accomplish this would give as the final design

$a_1b_1c_1$	$a_2b_1c_3$	$a_3b_3c_2$
$a_2b_2c_3$	$a_3b_2c_1$	$a_4b_4c_3$
$a_3b_3c_1$	$a_4b_3c_2$	
$a_4b_4c_2$		
$a_1b_5c_3$		
$a_2b_6c_2$		

This design is connected for all three factors, as shown in the following theorem:

<u>Theorem 2.3</u>

Any design in three factors, A, B, and C, that is constructed by the preceding procedure is connected for all three factors and fulfills the following requirement for factor F_β, where F_β can be either factor A or factor C:

<u>Property I</u> $n_\beta - a_\beta$ levels of F_β occur V_β times and a_β levels of F_β occur $V_\beta+1$ times in the design.

<u>Proof:</u>

The procedure itself guarantees that property I is maintained for factors A and C. All we need to show is that the design is connected for all three factors. In order to do this, we will rely on a definition and theorem given essentially by Srivastava and Anderson (1970).

<u>Definition 2.2</u>

For any m-factor multidimensional design D, let D_i be the (m-i+1)-dimensional design obtained from D by ignoring the factors $F_1, F_2, \ldots, F_{i-1}$. Then F_i is said to be connected with respect to factors

F_{i+1}, F_{i+2}, . . ., F_m in the original design D, if D_i is connected with respect to F_i.

Theorem 2.4

A multidimensional design D is connected with respect to all m factors if and only if the factor F_i is connected with respect to F_{i+1}, F_{i+2}, . . ., F_m for i - 1, 2, . . ., m-1.

For our particular situation, all we need to show is that the design, prior to adding the levels of our third factor, C, is connected with respect to factor A, and that the final design is connected with respect to factor C. Our factors, C, A, and B will correspond to F_1, F_2, and F_3 of the theorem respectively.

Recall that we obtained the (A x B) level combinations in the same manner that we did in the two-factor case. Thus, by Theorem 2.2, the design, prior to attaching levels of factor C, is connected with respect to factor A.

Consider (A x B) level combinations as levels of some new factor, say factor J. As far as factors C and J are concerned, we can apply Theorem 2.2 again, since we connected the design for factor C just as we did in the two-factor situation (where factor J is the first factor). Thus the final design is connected with respect to factor C.

This procedure for constructing an MMD in three factors can quite easily be extended to a design of m factors where we desire to estimate main effect linear contrasts in all m factors. We may not be able to maintain property I for all factors, but we can do so for the two factors which we consider to be the important ones.

CHAPTER 3

MINIMAL MULTIDIMENSIONAL DESIGNS, TWO-FACTOR INTERACTIONS PRESENT

3.1 Introduction

In the previous discussion of constructing MMD's, we have assumed that no interactions exist. In many instances, this assumption would be unrealistic. If there is a good possibility that two-factor interactions do exist, then we would be interested in being able to estimate contrasts in these interactions.

Consider the model for design D,

$$E(\underline{y}_D) = X_D \underline{p}_D$$

where $\underline{y}_D$ is the observation vector and X_D is the design matrix. The parameter vector, $\underline{p}_D$, will be of the form

$$\underline{p}_D = (\underline{p}'_{i_1}, \underline{p}'_{i_2}, \ldots, \underline{p}'_{i_q}, \underline{p}'_{j_1 j_2}, \ldots, \underline{p}'_{j_r j_s})$$

with $\underline{p}'_c = (p_c^1, p_c^2, \ldots, p_c^{n_c})$ and

$$\underline{p}'_{ab} = (p_{ab}^{11}, p_{ab}^{12}, \ldots, p_{ab}^{n_a n_b})$$

Here $\underline{p}_c$ refers to the factor F_c, which does not interact with any other factor and $\underline{p}_{ab}$ refers to a pair of factors that interact with each other.

Definition 3.1

Contrasts of the form $p_c^{\ell_c} - p_c^{\ell'_c}$ are called main effect or type I contrasts.

Definition 3.2

Contrasts of the form $p_{ab}^{\ell_a \ell_b} + p_{ab}^{\ell'_a \ell'_b} - p_{ab}^{\ell'_a \ell_b} - p_{ab}^{\ell_a \ell'_b}$ are called interaction or type II contrasts.

3.2 Type II Connectedness

Definition 3.3

Any design which allows estimation of all type I contrasts (of factors that do not interact) is said to be type I connected.

Definition 3.4

Any design which allows estimation of all type II contrasts is said to be type II connected.

We will mainly be interested in three-factor designs, although the procedure may easily be extended to the m-factor situation. In these three-factor designs we will consider the possibility of one, two, or three two-factor interactions. Our procedure will be a modification of Sennetti's (1972), but will show some simplification of notation as well as providing an easier method of construction. In section 3.6, we will outline some of the advantages of this procedure over Sennetti's.

Recall property I of Chapter 2:

> If F_β is any factor in the design, then $n_\beta - a_\beta$ levels of F_β occur V_β times and a_β levels of F_β occur $V_\beta + 1$ times in the design.

If we consider F_β to be any non-interacting factor, then we can try to maintain property I for that factor under the model of this chapter. If we think of factor F_β as having levels that are actually level combinations of two factors that interact, then we can try to maintain property

I for these level combinations. That is, we will consider property I for type I and type II contrasts. Sennetti (1972) only considers this property with respect to type I contrasts.

Our model will contain at least one term of the form $(\alpha\beta)_{ij}$, where here $(\alpha\beta)_{ij}$ is the effect of the level combination of the i th level of factor A and the j th level of factor B. This effect will include the main effect of the i th level of A, the main effect of the j th level of B, and the interaction effect. Using the notation of section 3.1, $(\alpha\beta)_{ij}$ may be written as p_{ab}^{ij}.

The following results by Sennetti give a general way of determining whether type II contrasts can be estimated.

Theorem 3.1 (Sennetti, (1972))

The linear function $(\alpha\beta)_{ij}-(\alpha\beta)_{i'j'}$ is estimable if and only if level combination a_ib_j is connected by a chain to level combination $a_{i'}b_{j'}$, where a_u and b_v denote the u th level of factor A and the v th level of factor B, respectively.

Theorem 3.2 (Sennetti, (1972))

Type II contrasts of the form $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}-(\alpha\beta)_{i'j}+(\alpha\beta)_{i'j'}]$ are estimable if either $(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}$ and $(\alpha\beta)_{i'j}-(\alpha\beta)_{i'j'}$ or $(\alpha\beta)_{ij}-(\alpha\beta)_{i'j}$ and $(\alpha\beta)_{ij'}-(\alpha\beta)_{i'j'}$ are estimable.

The following corollary is a direct result of the two theorems by Sennetti.

Corollary 3.1

If a design is connected with respect to level combinations of two factors, F_1 and F_2, then all type II contrasts in $(F_1 \times F_2)$ are estimable.

3.3 Model A--One Two-factor Interaction

First we will consider the model

$$E(y_{ijk}) = (\alpha\beta)_{ij} + \gamma_k$$

where γ_k is the effect of the k th level of factor C, a factor that does not interact with factors A and B. $(\alpha\beta)_{ij}$ will denote the effect of the ij th level combination of factors A and B. There will be $(n_a n_b - 1)$ independent differences in (A x B), and $(n_c - 1)$ independent type I contrasts in factor C. To be able to estimate all of the type II and type I contrasts, we require $(n_a n_b - 1) + (n_c - 1) = n_a n_b + n_c - 2$ degrees of freedom. Thus if we use an MMD to do this estimation, we will need only $(n_a n_b + n_c - 1)$ assemblies.

Since we want to connect the design for level combinations of A and B, we can think of these (A x B) level combinations as being levels of a new factor, which we might call F_1. As before in the two-factor, no interaction problem, we put the n_1, i.e., $n_a n_b$, levels of F_1 in the first n_1 assemblies, and then repeat $(n_c - 1)$ of the levels for the last $(n_c - 1)$ assemblies. We will choose the levels of factor F_1 to repeat and attach the levels of C in such a way that property I will be maintained for factor F_1 ((A x B) level combinations) and for factor C. We accomplish this in the same manner that we did in the two-factor, no interaction situation.

Example 3.1

Suppose $n_a = 4$, $n_b = 2$, and $n_c = 3$ under model A, $E(y_{ijk}) = \gamma_k + (\alpha\beta)_{ij}$. Using the procedure outlined above, we would obtain the following design:

$a_1b_1c_1$	$a_1b_1c_2$
$a_1b_2c_2$	$a_2b_2c_3$
$a_2b_1c_3$	
$a_2b_2c_1$	
$a_3b_1c_1$	
$a_3b_2c_2$	
$a_4b_1c_3$	
$a_4b_2c_1$	

From Theorem 2.2, we know that all levels of C are connected as are all (A x B) level combinations. Thus in addition to being able to estimate all type I contrasts in factor C, Corollary 3.1 tells us that all type II contrasts in (A x B) are estimable also.

Example 3.2

Suppose $n_a=2$, $n_b=3$, and $n_c=7$. We will need (2)(3) + (7) - 1 = 12 assemblies. Since $n_c > n_an_b$, we will first assign the levels of Factor C. The levels of C would be

c_1	c_1
c_2	c_2
c_3	c_3
c_4	c_4
c_5	c_5
c_6	
c_7	

When we attach the levels of (A x B) as we would the levels of the second factor in the two-factor, no interaction problem, we obtain as one possible

MMD, the following:

$a_1b_1c_1$	$a_1b_2c_1$
$a_1b_2c_2$	$a_1b_3c_2$
$a_1b_3c_3$	$a_2b_1c_3$
$a_2b_1c_4$	$a_2b_2c_4$
$a_2b_2c_5$	$a_2b_3c_5$
$a_2b_3c_6$	
$a_1b_1c_7$	

We can estimate a type I contrast such as $(\gamma_1-\gamma_2)$ by the expression $a_1b_2c_1 - a_1b_2c_2$. We can estimate a type II contrast such as $[(\alpha\beta)_{11}+(\alpha\beta)_{22}-(\alpha\beta)_{12}-(\alpha\beta)_{21}]$ by the expression $(a_1b_1c_1+a_2b_2c_4-a_1b_2c_1-a_2b_1c_4)$.

Theorem 2.2, along with Sennetti's two theorems that are given in this chapter, guarantee that we can estimate any type I contrast or type II contrast in this model.

Under models B and C, which we shall discuss in the rest of this chapter, we shall give a procedure for generating an MMD that is a significant simplification of the procedure suggested by Sennetti (1972).

3.4 Model B--Two Two-factor Interactions

Next we will consider model B,

$$E(y_{ijk}) = (\alpha\beta)_{ij} + (\beta\gamma)_{jk}$$

If we are going to be able to estimate all type II contrasts, we will need $(n_a-1) + (n_b-1) + (n_c-1) + (n_a-1)(n_b-1) + (n_b-1)(n_c-1) = n_an_b +$

$n_b n_c - n_b - 1$ degrees of freedom. Thus we will need $(n_a n_b + n_b n_c - n_b)$ assemblies in our MMD.

For model B, as well as model C which we will discuss later, the intuitively appealing chain concept of connectedness, as defined by Bose (1947), is not applicable. Instead we will use the following definition, which is more general than the definition given by Bose:

Definition 3.5

A design is connected with respect to a factor (or a factor combination) if and only if all contrasts in that factor (or factor combination) are estimable.

For the two and three two-factor interaction situations, Sennetti (1972) illustrates his procedure for the case of $n_a = n_b = n_c$; in particular for a 3^3 factorial. We will illustrate how our procedure may be applied to the more general situation of any n_a, n_b, and n_c. For definiteness, let $n_a \geq n_c$.

We know that there are $(n_a-1)(n_b-1)$ independent contrasts in $\alpha\beta$. Without loss of generality, we will choose a set of contrasts of the form

$$(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{21}+(\alpha\beta)_{22} \qquad (\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{31}+(\alpha\beta)_{32}$$

$$(\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{21}+(\alpha\beta)_{23} \qquad (\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{31}+(\alpha\beta)_{33}$$

$$(\alpha\beta)_{11}-(\alpha\beta)_{14}-(\alpha\beta)_{21}+(\alpha\beta)_{24} \qquad \cdot$$

$$\cdot \qquad \cdot$$

$$\cdot \qquad \cdot$$

$$\cdot \qquad (\alpha\beta)_{11}-(\alpha\beta)_{1n_b-1}-(\alpha\beta)_{n_a1}+(\alpha\beta)_{n_an_b-1}$$

$$(\alpha\beta)_{11}-(\alpha\beta)_{1n_b}-(\alpha\beta)_{21}+(\alpha\beta)_{2n_b} \qquad (\alpha\beta)_{11}-(\alpha\beta)_{1n_b}-(\alpha\beta)_{n_a1}+(\alpha\beta)_{n_an_b}$$

We can easily estimate these contrasts by taking the corresponding level combination expressions (each effect is replaced by the corresponding level combination). Of course, we have to keep the level of factor C constant for all four assemblies in any one expression, in order to do this estimation. That is, any contrast of the form

$$(\alpha\beta)_{ij}-(\alpha\beta)_{i'j}-(\alpha\beta)_{ij'}+(\alpha\beta)_{i'j'},$$

can be estimated by the level combination expression

$$(a_i b_j c_k - a_{i'} b_j c_k - a_i b_{j'} c_k + a_{i'} b_{j'} c_k)$$

for some level k of factor C. This expression has as its expected value the contrast we want to estimate.

Recall that we are trying to obtain a design for which we can estimate all type II contrasts with a minimum number of three-factor assemblies. If we use the same level of C, denoted c_1, with all the $(n_a-1)(n_b-1)$ expressions, then we would need only $n_a n_a$ assemblies to estimate the $(n_a-1)(n_b-1)$ independent contrasts. Using level c_1, we obtain the following set of expressions that can be used to estimate the chosen set of independent contrasts:

Set I

$a_1b_1c_1-a_1b_2c_1-a_2b_1c_1+a_2b_2c_1$ $\qquad$ $a_1b_1c_1-a_1b_2c_1-a_3b_1c_1+a_3b_2c_1$

$a_1b_1c_1-a_1b_3c_1-a_2b_1c_1+a_2b_3c_1$ $\qquad$ $a_1b_1c_1-a_1b_3c_1-a_3b_1c_1+a_3b_3c_1$

$a_1b_1c_1-a_1b_4c_1-a_2b_1c_1+a_2b_4c_1$

$\vdots$ $\qquad$ $\vdots$

$a_1b_1c_1-a_1b_{n_b-1}c_1-a_{n_a}b_1c_1+a_{n_a}b_{n_b-1}c_1$

$a_1b_1c_1-a_1b_{n_b}c_1-a_2b_1c_1+a_2b_{n_b}c_1$ $\qquad$ $a_1b_1c_1-a_1b_{n_b}c_1-a_{n_a}b_1c_1+a_{n_a}b_{n_b}c_1$

The particular set of contrasts used was chosen for two reasons: First, we know this to be a set of $(n_a-1)(n_b-1)$ independent contrasts since each contrast contains a two-factor assembly of the form $(\alpha\beta)_{ij}$, for i, j $\neq$ 1, which occurs only in that particular contrast. Second, it can be immediately noted how many times a given (A x B) combination occurs in Set I. The combination denoted by a_1b_1 occurs $(n_a-1)(n_b-1)$ times, a_ib_1, for i $\neq$ 1, occurs (n_b-1) times, a_1b_j for j $\neq$ 1, occurs (n_a-1) times, and a_ib_j for i, j $\neq$ 1, occurs one time. This will be important when we need to generate our last set of $(n_bn_c-n_b)$ assemblies to connect the design for (B x C).

Having obtained Set I, we need to choose an additional $(n_b-1)(n_c-1)$ factor level combination expressions that may be used to estimate a set of $(n_b-1)(n_c-1)$ independent contrasts in the effects of (B x C) level combinations. An easy way to determine a set of $(n_b-1)(n_c-1)$ independent contrasts in (B x C) is just to consider that would happen if we used a different level of C, c_k for k=2, 3, . . ., n_c, with the first two (A x B) combinations in $(n_b-1)(n_c-1)$ of the expressions in Set I. We will use level 2 of C with the first combination difference in (n_b-1) of these expressions, then level 3 of C with the first difference in another n_b-1 of these expressions, . . ., then level n_c of factor C with the first difference in the last chosen set of (n_b-1) expressions. These sets of (n_b-1) expressions are chosen so that we obtain all possible combinations of our third factor, factor C, and the factor that interacts with C, factor B. The sets of assembly combinations that we obtain are

<u>Set II</u>

$a_1b_1c_2-a_1b_2c_2-a_2b_1c_1+a_2b_2c_1$	$a_1b_1c_3-a_1b_2c_3-a_3b_1c_1+a_3b_2c_1$
$a_1b_1c_2-a_1b_3c_2-a_2b_1c_1+a_2b_3c_1$	$a_1b_1c_3-a_1b_3c_3-a_3b_1c_1+a_3b_3c_1$
$a_1b_1c_2-a_1b_4c_2-a_2b_1c_1+a_2b_4c_1$	.
.	.
.	.
.	$a_1b_1c_{n_c}-a_1b_{n_b-1}c_{n_c}-a_{n_a}b_1c_1+a_{n_c}b_{n_b-1}c_1$
$a_1b_1c_2-a_1b_{n_b}c_2-a_2b_1c_1+a_2b_{n_b}c_1$	$a_1b_1c_{n_c}-a_1b_{n_b}c_{n_c}-a_{n_a}b_1c_1+a_{n_c}b_{n_b}c_1$

We have attached the remaining (n_c-1) levels of C, one time each, to each of the first differences in the (A x B) expressions. This gave us an additional $(n_b-1)(n_c-1)$ three-factor level combination expressions. The expected value of any of the expressions given above will be the sum of a contrast in (A x B) and a contrast in (B x C). As an example, for the expression $(a_1b_1c_k-a_1b_jc_k-a_ib_1c_1+a_ib_jc_1)$, we would get the expected value $[(\alpha\beta)_{11}-(\alpha\beta)_{1j}-(\alpha\beta)_{i1}+(\alpha\beta)_{ij}] + [(\beta\gamma)_{1k}-(\beta\gamma)_{jk}-(\beta\gamma)_{11}+(\beta\gamma)_{j1}]$. We can estimate the (B x C) contrast by taking the difference between the level combination expression given above and the one that estimates $[(\alpha\beta)_{11}-(\alpha\beta)_{1j}-(\alpha\beta)_{i1}+(\alpha\beta)_{ij}]$ in Set I. Thus for each of these $(n_b-1)(n_c-1)$ expressions that we have formed, there is a corresponding estimable contrast in (B x C). These contrasts are independent since each contains a term of the form $(\beta\gamma)_{jk}$ which only occurs in that particular contrast. (This is the second term of the contrast.)

If we put our two sets of expressions together, we see that we can estimate $(n_a-1)(n_b-1)$ independent interaction contrasts in (A x B), and $(n_b-1)(n_c-1)$ independent interaction contrasts in (B x C). Now we need to consider how many assemblies were used in estimating these contrasts.

In obtaining the first $(n_a-1)(n_b-1)$ expressions, we used only n_an_b assemblies. In Set II, it can be noted that expressions 1, n_b, $2n_b-1$, $3n_b-2$, . . ., $(n_c-2)n_b-(n_c-3)$, each have two new assemblies. Or in these n_c-1 expressions, we have $2(n_c-1)$ new assemblies. In each of the remaining $(n_b-2)(n_c-1)$ expressions, we have only one new assembly each. Thus the total number of new assemblies used in Set II is $2(n_c-1)+(n_b-2)(n_c-1) = n_bn_c-n_b$. The total number of assemblies used in the procedure is $(n_an_b+n_bn_c-n_b)$, the minimum number possible for a completely connected design under model B. Thus our design will be an MMD.

<u>Example 3.3</u>

We want to obtain an MMD for the following situation: A three-level factor and a five-level factor, which do not interact. A four-level factor which interacts with both the other two factors. In order to use the same notation as in our description of the general procedure, we will call the four-level factor, factor B, the five-level factor, factor A, and let factor C be the three-level factor.

First we will give the $(n_a-1)(n_b-1) = (4)(3) = 12$ contrasts in (A x B). We choose $(\alpha\beta)_{11}$ (just some convenient level combination effect) to repeat in each of these contrasts.

$$(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{21}+(\alpha\beta)_{22}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{21}+(\alpha\beta)_{23}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{14}-(\alpha\beta)_{21}+(\alpha\beta)_{24}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{31}+(\alpha\beta)_{32}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{31}+(\alpha\beta)_{33}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{14}-(\alpha\beta)_{31}+(\alpha\beta)_{34}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{41}+(\alpha\beta)_{42}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{41}+(\alpha\beta)_{43}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{14}-(\alpha\beta)_{41}+(\alpha\beta)_{44}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{51}+(\alpha\beta)_{52}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{51}+(\alpha\beta)_{53}$$
$$(\alpha\beta)_{11}-(\alpha\beta)_{14}-(\alpha\beta)_{51}+(\alpha\beta)_{54}$$

To estimate these 12 contrasts we attach some convenient level of factor C (denoted c_1) to each of the (A x B) level combinations that correspond to the effects in the chosen contrasts. We get the following assembly expressions:

Set I

$$a_1b_1c_1-a_1b_2c_1-a_2b_1c_1+a_2b_2c_1$$
$$a_1b_1c_1-a_1b_3c_1-a_2b_1c_1+a_2b_3c_1$$
$$a_1b_1c_1-a_1b_4c_1-a_2b_1c_1+a_2b_4c_1$$
$$a_1b_1c_1-a_1b_2c_1-a_3b_1c_1+a_3b_2c_1$$
$$a_1b_1c_1-a_1b_3c_1-a_3b_1c_1+a_3b_3c_1$$
$$a_1b_1c_1-a_1b_4c_1-a_3b_1c_1+a_3b_4c_1$$
$$a_1b_1c_1-a_1b_2c_1-a_4b_1c_1+a_4b_2c_1$$
$$a_1b_1c_1-a_1b_3c_1-a_4b_1c_1+a_4b_3c_1$$
$$a_1b_1c_1-a_1b_4c_1-a_4b_1c_1+a_4b_4c_1$$
$$a_1b_1c_1-a_1b_2c_1-a_5b_1c_1+a_5b_2c_1$$
$$a_1b_1c_1-a_1b_3c_1-a_5b_1c_1+a_5b_3c_1$$
$$a_1b_1c_1-a_1b_4c_1-a_5b_1c_1+a_5b_4c_1$$

The assemblies used in these expressions are

$a_1b_1c_1$	$a_2b_4c_1$	$a_4b_3c_1$
$a_1b_2c_1$	$a_3b_1c_1$	$a_4b_4c_1$
$a_1b_3c_1$	$a_3b_2c_1$	$a_5b_1c_1$
$a_1b_4c_1$	$a_3b_3c_1$	$a_5b_2c_1$
$a_2b_1c_1$	$a_3b_4c_1$	$a_5b_3c_1$
$a_2b_2c_1$	$a_4b_1c_1$	$a_5b_4c_1$
$a_2b_3c_1$	$a_4b_2c_1$	

Next we need to obtain assembly combinations expressions to estimate $(n_b-1)(n_c-1) = (3)(2) = 6$ independent, type II contrasts in (B x C). To obtain these, we vary the level of C used with the first two terms of each of the appropriate 6 expressions chosen for (A x C). These new assembly expressions are

Set II

$$a_1b_1c_2-a_1b_2c_2-a_2b_1c_1+a_2b_2c_1$$
$$a_1b_1c_2-a_1b_3c_2-a_2b_1c_1+a_2b_3c_1$$
$$a_1b_1c_2-a_1b_4c_2-a_2b_1c_1+a_2b_4c_1$$
$$a_1b_1c_3-a_1b_2c_3-a_3b_1c_1+a_3b_2c_1$$
$$a_1b_1c_3-a_1b_3c_3-a_3b_1c_1+a_3b_3c_1$$
$$a_1b_1c_3-a_1b_4c_3-a_3b_1c_1+a_3b_4c_1$$

Here the expressions chosen from Set I to be modified were the first six. We needed to insure that every level of B would occur with every level of factor C. Level 1 of C occurs with every level of B in Set I. All other levels of C occur with all levels of B in Set II. The expected values of the expressions in Set II are

$$[(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{21}+(\alpha\beta)_{22}] + [(\beta\gamma)_{12}-(\beta\gamma)_{22}-(\beta\gamma)_{11}+(\beta\gamma)_{21}]$$
$$[(\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{21}+(\alpha\beta)_{23}] + [(\beta\gamma)_{12}-(\beta\gamma)_{32}-(\beta\gamma)_{11}+(\beta\gamma)_{31}]$$
$$[(\alpha\beta)_{11}-(\alpha\beta)_{14}-(\alpha\beta)_{21}+(\alpha\beta)_{24}] + [(\beta\gamma)_{12}-(\beta\gamma)_{42}-(\beta\gamma)_{11}+(\beta\gamma)_{41}]$$
$$[(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{31}+(\alpha\beta)_{32}] + [(\beta\gamma)_{13}-(\beta\gamma)_{23}-(\beta\gamma)_{11}+(\beta\gamma)_{21}]$$
$$[(\alpha\beta)_{11}-(\alpha\beta)_{13}-(\alpha\beta)_{31}+(\alpha\beta)_{33}] + [(\beta\gamma)_{13}-(\beta\gamma)_{33}-(\beta\gamma)_{11}+(\beta\gamma)_{31}]$$
$$[(\alpha\beta)_{11}-(\alpha\beta)_{14}-(\alpha\beta)_{31}+(\alpha\beta)_{34}] + [(\beta\gamma)_{13}-(\beta\gamma)_{43}-(\beta\gamma)_{11}+(\beta\gamma)_{41}]$$

The six (B x C) contrasts above are independent and can be estimated by taking the appropriate expression from Set II and subtracting the corresponding expression in Set I. For instance, to estimate $[(\beta\gamma)_{13}-(\beta\gamma)_{23}-(\beta\gamma)_{11}+(\beta\gamma)_{21}]$, we would use the difference

$$(a_1b_1c_3-a_1b_2c_3-a_3b_1c_1+a_3b_2c_1) - (a_1b_1c_1-a_1b_2c_1-a_3b_1c_1+a_3b_2c_1)$$

The new assemblies in Set II are

$a_1b_1c_2$	$a_1b_1c_3$
$a_1b_2c_2$	$a_1b_2c_3$
$a_1b_3c_2$	$a_1b_3c_3$
$a_1b_4c_2$	$a_1b_4c_3$

We know that the design is connected for (A x B) and for (B x C). Thus the design will be an MMD if it has $(n_an_b+n_bn_c-n_b) = 28$ assemblies, the minimum number possible for a connected design under this model and these conditions. Since we have a total of 20 (set I) and 8 additional (Set II) assemblies, we know that the design generated above is an MMD.

3.5 Model C--Three Two-factor Interactions

Under model C,

$$E(y_{ijk}) = (\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}$$

we have the assumption that all possible two-factor interactions exist. We want to construct an MMD with which we can estimate all type II contrasts in (A x B), (A x C), and (B x C). We will need $[(n_a-1) + (n_b-1) + (n_c-1) + (n_a-1)(n_b-1) + (n_a-1)(n_c-1) + (n_b-1)(n_c-1) + 1]$ $= [n_an_b+n_bn_c-n_b] + [(n_a-1)(n_c-1)]$ assemblies. For model C, we will use the same $[n_an_b+n_bn_c-n_b]$ assemblies to estimate the $(n_a-1)(n_b-1)$ independent contrasts in (A x B) and the $(n_b-1)(n_c-1)$ independent contrasts in (B x C) as we did under model B. All we need is an additional set of $(n_a-1)(n_c-1)$ new assemblies to use in estimating a set of $(n_a-1)(n_c-1)$ independent contrasts in (A x C). In order to be able to estimate such a set of contrasts, we certainly will need all possible level combinations of (A x C) in our design.

In Set I for our design under model B, we had level one of factor A occurring with level one of C. In Set II, we had level one of A occurring with all other levels of C. Thus we need to let the remaining (n_a-1) levels of A occur with the remaining (n_c-1) levels of C.

For definiteness, let $n_a \geq n_c$ and $n_b \geq n_c$. With this ordering of numbers of levels, we can first choose the assemblies necessary to estimate the (A x B) contrasts, then choose the additional assemblies necessary to estimate the (B x C) contrasts, as we did under model B. Since $n_b \geq n_c$, we know $(n_a-1)(n_b-1) \geq (n_a-1)(n_c-1)$. Thus we can get $(n_a-1)(n_c-1)$ level combination expressions (Set III) that will be used when we estimate the (A x C) contrasts from the $(n_a-1)(n_b-1)$ expressions used to estimate the (A x B) contrasts. One way to obtain these expressions is as follows: Use level 1 of C with the second and fourth terms of the expressions used to estimate the (A x B) contrasts (Set I) and levels 2, 3, . . ., n_c of C with the first and third terms of the expressions. For instance, the first new expression will be $(a_1b_1c_2-a_1b_2c_1 -a_2b_1c_2-a_2b_2c_1)$. This expression has an expected value of

$$[(\alpha\beta)_{11}-(\alpha\beta)_{12}-(\alpha\beta)_{21}+(\alpha\beta)_{22}] + [(\beta\gamma)_{12}-(\beta\gamma)_{21}-(\beta\gamma)_{12}-(\beta\gamma)_{21}] +[(\alpha\gamma)_{12}-(\alpha\gamma)_{11}-(\alpha\gamma)_{22}+(\alpha\gamma)_{21}]$$

We already have the assemblies to estimate the (A x B) contrast and the (B x C) contrast is 0. Thus this expression, along with the appropriate expression from Set I, can be used to estimate the (A x C) contrast.

We will apply this same modification to $(n_a-1)(n_c-1)$ of the expressions in Set I. We choose the expressions to be used that will give us, in the third term of the expression, all possible (A x C) combinations that have not occurred in Set I or Set II. The entire set of $(n_a-1)(n_c-1)$ new assembly expressions would be

Set III

$a_1b_1c_2-a_1b_2c_1-a_2b_1c_2+a_2b_2c_1$

$a_1b_1c_3-a_1b_3c_1-a_2b_1c_3+a_2b_3c_1$

.

.

.

$a_1b_1c_{n_c}-a_1b_{n_c}c_1-a_2b_1c_{n_c}+a_2b_{n_c}c_1$

$a_1b_1c_2-a_1b_2c_1-a_3b_1c_2+a_3b_2c_1$

$a_1b_1c_3-a_1b_3c_1-a_3b_1c_3+a_3b_3c_1$

.

.

.

$a_1b_1c_{n_c}-a_1b_{n_c}c_1-a_{n_a}b_1c_{n_c}+a_{n_a}b_{n_c}c_1$

Each of the expressions above has an expected value that is the sum of an estimable (A x B) contrast and a contrast in (A x C). Since the third term of each of these (A x C) contrasts occurs in only that contrast, we know the contrasts are independent.

Each of these new expressions contains only one new assembly, the one that is the third term in the expression. Thus by the addition of $(n_a-1)(n_c-1)$ new assemblies, we are able to estimate a set of $(n_a-1)(n_c-1)$ independent contrasts in (A x C). We use the assemblies chosen under model B and augment these with the assemblies first used in Set III. We obtain a three-factor, three two-factor interaction, minimal design in which all type II contrasts are estimable.

Example 3.4

Suppose we have a three-factor experiment and suspect that all three two-factor interactions exist. The three factors have three, four, and two levels respectively. To be consistent with the notation previously used, we can denote the four-level factor, factor B, the three-level factor, factor A, and the two-level factor, factor C. Using the procedure of model B, we can get a set of $(n_a-1)(n_b-1) = 6$ independent contrasts in (A x B). The assembly expressions required to estimate one such set of contrasts would be

Set I

$$a_1b_1c_1-a_1b_2c_1-a_2b_1c_1+a_2b_2c_1$$
$$a_1b_1c_1-a_1b_3c_1-a_2b_1c_1+a_2b_3c_1$$
$$a_1b_1c_1-a_1b_4c_1-a_2b_1c_1+a_2b_4c_1$$
$$a_1b_1c_1-a_1b_2c_1-a_3b_1c_1+a_3b_2c_1$$
$$a_1b_1c_1-a_1b_3c_1-a_3b_1c_1+a_3b_3c_1$$
$$a_1b_1c_1-a_1b_4c_1-a_3b_1c_1+a_3b_4c_1$$

The assemblies used in Set I are

$a_1b_1c_1$	$a_2b_1c_1$	$a_3b_1c_1$
$a_1b_2c_1$	$a_2b_2c_1$	$a_3b_2c_1$
$a_1b_3c_1$	$a_2b_3c_1$	$a_3b_3c_1$
$a_1b_4c_1$	$a_2b_4c_1$	$a_3b_4c_1$

Next we obtain the $(n_b-1)(n_c-1) = 3$ assembly expressions that may be used to estimate a complete set of three independent contrasts in (B x C).

Set II

$$a_1b_1c_2-a_1b_2c_2-a_2b_1c_1+a_2b_2c_1$$
$$a_1b_1c_2-a_1b_3c_2-a_2b_1c_1+a_2b_3c_1$$
$$a_1b_1c_2-a_1b_4c_2-a_2b_1c_1+a_2b_4c_1$$

The new assemblies required here are

$a_1b_1c_2$	$a_1b_3c_2$
$a_1b_2c_2$	$a_1b_4c_2$

To get the $(n_a-1)(n_c-1) = 2$ assembly expressions necessary to estimate a complete set of independent (A x C) contrasts, we use level 2 of factor C with the first and third terms and level 1 of C with the second and fourth terms of expressions 1 and 4 of Set I. The expressions obtained are

Set III

$$a_1b_1c_2-a_1b_2c_1-a_2b_1c_2+a_2b_2c_1$$
$$a_1b_1c_2-a_1b_2c_1-a_3b_1c_2+a_3b_2c_1$$

The only new assemblies required for Set III are

$a_2b_1c_2$	$a_3b_1c_2$

Thus our final design is

$a_1b_1c_1$	$a_2b_3c_1$	$a_1b_1c_2$
$a_1b_2c_1$	$a_2b_4c_1$	$a_1b_2c_2$
$a_1b_3c_1$	$a_3b_1c_1$	$a_1b_3c_2$
$a_1b_4c_1$	$a_3b_2c_1$	$a_1b_4c_2$
$a_2b_1c_1$	$a_3b_3c_1$	$a_2b_1c_2$
$a_2b_2c_1$	$a_3b_4c_1$	$a_3b_1c_2$

Using this design it is possible to estimate all possible type II contrasts in (A x B), (A x C), and (B x C). The design is a minimal multidimensional design since only $\{[n_a n_b + n_b n_c - n_b] + [(n_a - 1)(n_c - 1)]\} = 18$ assemblies were used.

3.6 Advantages of the Procedure

Sennetti (1972) comments on the difficulty of constructing even one MMD for the situation we had in model B or model C. He suggests that the construction of a single MMD for these cases might in itself be an optimality criterion. We have shown, in this chapter, a relatively simple way to construct one particular design with which we can estimate all type II contrasts from a given set. Actually we have a certain amount of flexibility in generating our design. We could vary our choice of repeated (A x B) combination in Set I or vary the chosen level of factor C in Set I. This actually gives us a choice of $n_a n_b n_c$ different designs. No matter which of these designs we choose, we can estimate any linear contrast in the chosen complete independent set of contrasts for the factors in question. Since the set is complete, any possible linear contrast in the factors will be a linear combination of the contrasts in the set, and can be estimated.

Certainly the procedure we have outlined for constructing an MMD for the model B and model C situations is not the only one that could be used. Since this is true, we need to say a few words about the advantages of this procedure over other possible procedures (such as the one used by Sennetti (1972) for the simple case of a 3^3 factorial).

The first advantage of our procedure lies in its simplicity of use; the procedure is easy to apply and can be used for any experiment under model B or model C, no matter how many levels the factors have. Sennetti's procedure is difficult or perhaps impossible to apply for this general situation of unequal n's.

The next thing to consider is ease of contrast estimation. In any MMD, we know that every possible contrast is estimable. However, it may be quite difficult to determine what combination of observations to use in this estimation. It is with respect to ease of estimation of any arbitrary contrast, that we notice a very interesting and useful property of designs generated by our procedure: If the repeated (A x B) combination of Set I is denoted a_1b_1, and the repeated level of C in Set I is denoted c_1, then for

(a) Contrasts in (A x B): Every possible (A x B) level combination occurs with level 1 of C. To estimate any (A x B) contrast of the form $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}-(\alpha\beta)_{i'j}+(\alpha\beta)_{i'j'}]$, we can use the assembly combination $(a_ib_jc_1-a_ib_{j'}c_1-a_{i'}b_jc_1+a_{i'}b_{j'}c_1)$.

(b) Contrasts in (B x C): All possible (B x C) combinations of the form b_jc_1 for j=1, 2, . . ., n_b occur with level 1 of A in Set I. All other (B x C) level combinations occur with level 1 of A in Set II. Thus we can estimate any (B x C) contrast of the form $[(\beta\gamma)_{jk}-(\beta\gamma)_{jk'}-(\beta\gamma)_{j'k}+(\beta\gamma)_{j'k'}]$ by the expression $(a_1b_jc_k-a_1b_jc_{k'}-a_1b_{j'}c_k+a_1b_{j'}c_{k'})$.

(c) Contrasts in (A x C): All (A x C) combinations of the form a_ic_1 for i=1, 2, . . ., n_a occur with level 1 of factor B in Set I. All (A x C) combinations of the form a_1c_k for k=2, 3, . . ., n_c occur with level 1 of B in Set II. All other (A x C) combinations occur with level

1 of B in Set III. Thus, we can estimate any (A x C) contrast of the form $[(\alpha\gamma)_{ik}-(\alpha\gamma)_{ik'}-(\alpha\gamma)_{i'k}+(\alpha\gamma)_{i'k'}]$ by the expression $(a_i b_1 c_k - a_i b_1 c_{k'} - a_{i'} b_1 c_k + a_{i'} b_1 c_{k'})$.

The preceding discussion illustrates probably the biggest advantage of designs generated by our procedure: It is always immediately obvious how to estimate any type II contrast. We never have to search for the proper linear combination of contrasts in our set in order to decide how to estimate any given contrast of interest.

Sennetti (1972) has discussed type II connectedness with respect to a model of the form $E(y_{ijkm}) = (\alpha\beta)_{ij}+(\gamma\delta)_{km}$, in addition to our models A, B, and C. If we let level combinations whose effects are the $(\gamma\delta)_{km}$ be levels of our factor C, then this design is equivalent to our model A. All definitions and theorems of this work which are given for model A will apply to this model also. For this reason, we will not discuss this model explicitly throughout this work.

CHAPTER 4

MINIMAL AUGMENTED MULTIDIMENSIONAL DESIGNS

4.1 Introduction

At this point in the work, we have discussed the generation of designs with certain desirable properties. In each case, we have tried to obtain the entire design to be used in a given experiment. However, the occasion may arise when we wish to add one or more additional assemblies to an existing design. This may occur because some data has been taken and this data is too valuable to waste. In this chapter we will discuss how we can "best" augment an existing design under conditions such as these.

Consider the following situation: An experimenter has assumed a certain model to be correct and has run the experiment using an appropriate design. The model chosen will reflect the effects the experimenter thinks are present. The design chosen will reflect this information, and perhaps what contrasts are of interest, as well as the restrictions placed on the design by practical considerations such as high cost per observation. Suppose, after part or all of the experiment has been run, the experimenter obtains information (either from the data or from some other source) that makes him believe that the design chosen may not be adequate. The experimenter could just discard the data already obtained (and the design used) and attempt to generate a completely new, appropriate design. However, often it would be impractical or even impossible to obtain all new data. Thus a more practical approach would be to augment the original design in order to obtain a new design which is adequate in light of the new model and the additional information now desired.

As an example, the experimenter may desire to run a 3^3 factorial experiment. Since he believes that all interactions are insignificant, he uses a design which is useful for estimating main-effect contrasts only. If he later decides that perhaps there is one two-factor interaction present, he would augment the design so as to be able to estimate the appropriate type II contrasts as well as the remaining main-effect contrasts.

4.2 Augmenting Existing Designs

There are many ways to add new design points so as to make the new design completely connected (type I and type II connected). If the number of new assemblies added is minimal, then the new design would be a minimal augmented multidimensional design (MAMD).

Definition 4.1 (Sennetti (1972))

The multidimensional design D* is said to be a minimal multidimensional design (MAMD) with regard to the multidimensional design (MD), D, if T* has the minimum number of assemblies such that D* = D + T* is completely connected.

The concepts and procedures involved in generating MMD's in Chapters 2 and 3 can be employed to obtain minimal augmented multidimensional designs under certain conditions. Consider the usual linear model

$$E(\underline{y}_D) = X_D\underline{p}_D$$

The following theorem was given by Sennetti (1972):

Theorem 4.1

For a given MD, D, there exists an MAMD, D* = D + T* with the number of assemblies in D* equal to h(D*), where h(D*) = h(D) + h(T*), and $h(T^*) = \nu(\underline{p}_D) - r(X_D)$, $\nu(\underline{p}_D)$ is the number of degrees of freedom for $\underline{p}_D$ and $r(X_D)$ is the rank of X_D.
(Note that here $\nu(\underline{p}_D)$ includes one degree of freedom for the overall mean.)

4.3 MAMD's--Model A

For model A,

$$E(y_{ijk}) = (\alpha\beta)_{ij} + \gamma_k,$$

it is quite easy to get T*. We first insure that the design is connected for factor C. Then we add assemblies until all possible (A x B) combinations occur in the design.

Example 4.1

Suppose $n_a=2$, $n_b=3$, $n_c=3$ with D given as follows:

$a_1b_2c_3$

$a_1b_2c_2$

$a_1b_3c_1$

$a_2b_1c_2$

$a_2b_2c_3$

To connect the design for C, we would repeat level combinations of (A x B), just as we did in Chapter 3. We attach a level of C to each of these combinations that is one higher (mod n_c) than the level of C used previously with the (A x B) combination. Here if we repeated a_1b_2, then we would use level one of C since $a_1b_2c_3$ is in the design. Of course, $a_1b_2c_2$ is in the original design also, but it would be of no value to add $a_1b_2c_3$ since this would be repeating an entire assembly. Other possible new assemblies would be $a_1b_3c_2$, $a_2b_1c_3$, or $a_2b_2c_1$. We decide which combination to repeat by noting which levels of C need to be connected. That is, if we need to connect levels two and three of C, we can do this by adding a new assembly with level three of C, since level two of C already occurs with the same (A x B) combination in the design. In this example, we need to connect levels one and two or to connect levels one and three. Thus either $a_1b_3c_2$ or $a_1b_2c_1$ may be used. If we use $a_1b_3c_2$, then we have, at this point, for our design

$a_1b_2c_3$
$a_1b_2c_2$
$a_1b_3c_1$
$a_2b_1c_2$
$a_2b_2c_3$
$a_1b_3c_2$

This design is connected for factor C. Now we need to insure that all (A x B) level combinations are included in the design. In these new assemblies, the level of C is not important, since we know that if the design is connected for factor C, then it will be connected for (A x B) combinations. (This is based on the theorem by Eccleston and Hedayat (1974) which says that any two-factor design that is connected for one factor is connected for the other factor also.) We only need to add the (A x B) combinations a_1b_1 and a_2b_3. If we use levels one and three of C, respectively, with the two combinations, we get as our final design

$\underline{D^*}$

$a_1b_2c_3$	$a_2b_2c_3$
$a_1b_2c_2$	$a_1b_3c_2$
$a_1b_3c_1$	$a_1b_1c_1$
$a_2b_1c_2$	$a_2b_3c_3$

D* is completely connected and all type I contrasts in C and type II contrasts in (A x B) are estimable. The number of assemblies we added was $\nu(\underline{p}_D) - r(X_D) = 8 - 5 = 3$. Thus by Theorem 4.1, D* is an MAMD.

4.4 MAMD's--Model B

Consider model B,

$$E(y_{ijk}) = (\alpha\beta)_{ij} + (\beta\gamma)_{jk}.$$

Although the procedure here will not be as simple as for model A, we can still show how to augment the design. We will first augment the design to connect it with respect to (A x B) and then with respect to (B x C). To augment the design for (A x B), we first need to insure that all level combinations of (A x B) occur in the design. In fact, if we had all possible level combinations of (A x B) occurring with a constant level of C, then the design would be type II connected for (A x B). (This is the same procedure used to get Set I when we were generating a completely connected design for model B in Chapter 3.) We choose the level of C that will allow us to do this with the least number of new assemblies. Next we add additional assemblies until all possible (B x C) level combinations occur in the design. The levels of A used with this last set of (B x C) level combinations is immaterial.

Example 4.2

We shall consider the same original design as in example 4.1, but under the assumption of model B.

D

$a_1b_2c_3$

$a_1b_2c_2$

$a_1b_3c_1$

$a_2b_1c_2$

$a_2b_2c_3$

Level combinations a_1b_2 and a_2b_2 both occur with c_3. Also, a_1b_2 and a_2b_1 occur with c_2. No other level of C occurs with as many (A x B) combinations. Thus c_2 or c_3 would be the appropriate level of C to use with the (A x B) combinations to be added. If we choose c_3, then we would add assemblies $a_1b_1c_3$, $a_1b_3c_3$, $a_2b_1c_3$, and $a_2b_3c_3$. The augmented design will be type II connected for (A x B).

Next we add assemblies so as to have all possible (B x C) combinations in the design. The assemblies $a_1b_1c_1$, $a_2b_2c_1$, and $a_2b_3c_2$ will accomplish this. As noted before, the level of A used with these last three assemblies is of no consequence. The final design, D*, will be type II connected for (A x B) and for (B x C).

D*

$a_1b_2c_3$	$a_1b_1c_3$
$a_1b_2c_2$	$a_1b_3c_3$
$a_1b_3c_1$	$a_2b_1c_3$
$a_2b_1c_2$	$a_2b_3c_3$
$a_2b_2c_3$	$a_2b_2c_1$
	$a_2b_3c_2$
	$a_1b_1c_1$

Here the degrees of freedom for the parameter vector (independent of design) is $\nu(\underline{p}_D) = \{[1+2+2+(2)(1)+(2)(2)] +1\} = 12$, the last degree of freedom being for the overall mean. The rank of the X matrix for the original design is $r(X_D) = 5$. Thus $\nu(\underline{p}_D) - r(X_D) = 7$. Since we added only seven assemblies to completely connect the design, we know that D* is an MAMD.

4.5 MAMD's--Model C

For model C,

$$E(y_{ijk}) = (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk},$$

we follow the same general procedure. First we add assemblies to type II connect the design for (A x B), using a constant level of C. Then we add assemblies to type II connect the design for (B x C) using a constant level of A. Finally, we add assemblies that will insure that all possible level combinations in (A x C) occur in the design.

Example 4.3

Same original design as in 4.1 and 4.2. All two-factor interactions will be assumed present.

D

$a_1b_2c_3$

$a_1b_2c_2$

$a_1b_3c_1$

$a_2b_1c_2$

$a_2b_2c_3$

To type II connect for (A x B), we add the same four assemblies as in 4.2. These are

$a_1b_1c_3$

$a_1b_3c_3$

$a_2b_1c_3$

$a_2b_3c_3$

Next we get all possible (B x C) combinations with a constant level of factor A. If we choose level a_1, then we can do this with only four additional assemblies. (If we had chosen level a_2, it would have required five additional assemblies.) Using level one of A, we get the new assemblies

$a_1b_1c_1$

$a_1b_1c_2$

$a_1b_2c_1$

$a_1b_3c_2$

Finally we add assemblies to insure that all possible (A x C) level combinations occur in the design. The single assembly

$a_2b_3c_1$

will accomplish this. The final design is completely connected.

$\underline{D^*}$

$a_1b_2c_3$	$a_1b_1c_3$
$a_1b_2c_2$	$a_1b_3c_3$
$a_1b_3c_1$	$a_2b_1c_3$
$a_2b_1c_2$	$a_2b_3c_3$
$a_2b_2c_3$	$a_1b_1c_1$
	$a_1b_1c_2$
	$a_1b_2c_1$
	$a_1b_3c_2$
	$a_2b_3c_1$

We note here that $\nu(\underline{p}_D) = [1+2+2+(1)(2)+(1)(2)+(2)(2)] + 1 = 14$, while $r(X_D) = 5$. Since we only had to add $14 - 5 = 9$ assemblies to completely connect our design, we know that D* is an MAMD.

Under models A, B, and C, we did not have to worry about augmenting the design to type II connect it with respect to the last effect considered. We only have to insure that all possible level combinations associated with the effect are in the design. Actually, this also happened in Chapter 3 when we were generating MMD's. That is, another way to interpret our procedure to get Set II under models A and B, and to get Set III under model C, is as an orderly way of insuring that all level combinations associated with the last effect considered, actually occur in the design. We proved in Chapter 3 that the last effect considered was type II connected by using the following argument; Since the complete set of independent contrasts of Set II (or Set III under model C) is estimable, the design is type II connected for the effect considered in Set II (Set III for model C). Sennetti (1972) has stated an equivalent result, using the reasoning of the augmentation procedure of this chapter, as one result of the following theorem:

Theorem 4.2

An MD, T, is completely connected if and only is there exists a sequence of designs T_ω, ω=0, 1, . . ., M*(T)-2 such that T is connected w. r. t. at least one S_α, where S_α belongs to T. (Here S_α represents the set of all factor levels (or factor level combinations) associated with the effect in question.)

There certainly is another possible consideration with respect to augmenting existing designs so as to obtain MAMD's. That is, of the possible sets of additional assemblies that will give us an MAMD, is one set better than another? We would like to be able to state and justify some criterion for choosing a particular design from the set of designs that may be obtained by the procedure of this chapter.

First of all, it seems reasonable that if the property of equal level frequencies was desirable when trying to generate an MMD, then the property would be desirable here. That is, we will try to obtain a design for which the following property of Chapter 2 is satisfied:

Property I If F_β is any one of the k factors in the design, then $n_\beta - a_\beta$ levels of F_β occur V_β times and a_β levels of F_β occur $V_\beta + 1$ times in the design.

The reasons for attempting to maintain property I have been given in Chapter 3. Of course, it may not always be possible to maintain this desirable property, but we will do so if possible. On the other hand, we may have a set of potential MAMD's that are equally good according to the criterion of property I. There should be some way of choosing a design, using a different criterion than the one we have just discussed. Such an optimality criterion will be considered in the following chapters and applied to the present situation in Chapter 10.

CHAPTER 5

TYPES OF CONNECTEDNESS--S AND C-OPTIMALITY

5.1 Introduction

We have discussed generating an entire design with certain desirable properties or improving a given design with the addition of new assemblies. However, the occasion may arise in which we have a given design and are unable to increase the design size by adding new assemblies. In Chapters 5, 6, and 7, we discuss some possible criteria to be used in comparing existing designs, and procedures that may be used to improve certain existing designs with respect to two of these criteria.

Consider some design with two factors, A and B. Suppose we are interested in estimating contrasts in effects of A: i.e., $(\alpha_j-\alpha_{j'})$ for $j\neq j'$. To say that $(\alpha_j-\alpha_{j'})$ can be estimated (or to say that there is a chain between a_j and $a_{j'}$) is equivalent to saying that at least one replicate of a_j is connected by a chain to at least one replicate of $a_{j'}$, or that at least one replicate of a_j can be used in the estimation of $(\alpha_j-\alpha_{j'})$. If we have m_{aj} replicates of a_j in the design, then as far as estimation of $(\alpha_j-\alpha_{j'})$ is concerned, many of these replicates may be wasted. Connectedness does not imply full utilization of factor level replicates in the estimation of contrasts.

We will consider a procedure that will increase the utilization of factor level replicates in the estimation of contrasts. This is a question of improving an existing design when we are unable to increase the number of design points or even to change the number of replicates of any factor level. That is, if m_{ij} is the number of replicates of level j of factor I, then we will unable to change m_{ij} for any i,j.

5.2 Pseudo-global Connectedness

In developing a procedure to improve an existing design without altering any of the m_{ij}, we will use the following concepts and terminology:

Definition 5.1 (Eccleston and Hedayat, (1974))

A locally connected (l-connected) design is any two-factor, no interaction design that is connected for both factors according to the definition of Bose (1974).

Definition 5.2 (Hedayat, (1971))

A globally connected design is any two-factor, no interaction design in which every pair of levels of one of the factors is globally connected. Two levels of a factor, i and j, for $i \neq j$, are said to be globally connected if every replicate of i is connected by a chain, as defined by Bose (1947), to every replicate of j.

Definition 5.3 (Eccleston, (1972))

Two levels of a factor, i and j, for $i \neq j$, of a two-factor, no interaction design, are said to be pseudo-globally connected (pg-connected) if each replicate of i is connected by a chain, as defined by Bose (1947), to at least one replicate of j and vice versa. A design is pg-connected for a factor if every pair of levels of that factor is pg-connected.

The major importance of having a design that is globally connected or pg-connected may be seen in the following discussion: For a globally connected design, every replicate of level j of factor A and every replicate of level j' of factor A can be used together in the estimation of the linear contrast $(\alpha_j - \alpha_{j'})$. This yields a maximum of $m_{aj}m_{aj'}$ estimates of $(\alpha_j - \alpha_{j'})$. For a pg-connected design, we still have utilization of every replicate of any given factor level involved, when estimating a linear contrast. That is, every replicate of each factor level in the contrast can be used at least once in the estimation of the contrast. No replication is wasted. (See Eccleston and Hedayat, 1974.) We have at least $\max(m_{aj}, m_{aj'})$ estimates of the contrast. For a 1-connected design, we only are insured that at least one replicate of any factor level involved can be used in the estimation of the contrast. We may have only one estimate for the contrast.

We see that designs generally improve (with respect to number of contrast estimates) when we go from 1-connected to pg-connected to globally connected designs. At each step, we use more and more observations in the estimation procedure. We will try to change an 1-connected design into a pg-connected design in order to more fully utilize our factor level replicates.

5.3 S-optimal Designs

There is another reason to consider pg-connected designs. Under some conditions, the pg-connected design will be better (according to certain optimality criteria which we will discuss) than the 1-connected design. In order to discuss these criteria, we need to consider another definition of connectedness (or 1-connectedness), which uses the concept of the coefficient matrix. The coefficient matrix has traditionally been defined as follows:

Definition 5.4

Consider the class of incomplete block designs, D_{vm}, with v treatments arranged in m blocks. Treatment j is replicated r_j times and block i is of size k_i. The intrablock estimates of treatment effects may be given by

$$C\hat{\underline{\tau}} = \underline{Q}$$

where $\underline{\tau}$ is the parameter vector and C is defined as

$$C = \text{diag}(r_1, r_2, \ldots, r_v) - N[\text{diag}(k_1^{-1}, k_2^{-1}, \ldots, k_m^{-1})]N'$$

where N is the incidence matrix of the design. C is called the coefficient matrix. $\underline{Q} = \underline{T} - N[\text{diag}(k_1^{-1}, k_2^{-1}, \ldots, k_m^{-1})]\underline{B}$, where $\underline{T}$ and $\underline{B}$ are vectors of treatment and block totals.

We will use the concept of the coefficient matrix with respect to any design that has two or more factors and no interaction. If we are considering a particular factor, say factor F, then we can obtain C_F, the coefficient matrix for F. The r_j of the definition will be the numbers of replicates of the levels of F and the k_i will be the numbers of replicates of the levels of the other factor (or the number of replicates of the level combinations of the other factors). If $\underline{\tau}' = (\tau_1, \tau_2, \ldots, \tau_v)$ is the parameter vector corresponding to factor F, then we want to be able to estimate contrasts of the form $(\tau_j - \tau_{j'})$ for all j and j'.

Definition 5.5 (Chakrabarti, (1973))

A design is connected for factor F if the rank of the coefficient matrix for factor F is one less that the number of levels of factor F. That is

$$r(C_F) = v-1.$$

From Chakrabarti's definition, we know that if a design is connected for a factor, then the coefficient matrix for the factor will have v-1 non-zero eigenvalues. We will denote these eigenvalues of C by $\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_{v-1}$.

Three of the most commonly used optimality criteria may be given in terms of the non-zero eigenvalues of the coefficient matrix. These are

(1) A-optimality: minimize the average variance of all elementary treatment contrasts by minimizing (over all competing designs) $\sum_{j=1}^{v-1} \gamma_j^{-1}$.

(2) D-optimality: minimize the generalized variance by minimizing (over all competing designs) $\prod_{j=1}^{v-1} \gamma_j^{-1}$.

(3) E-optimality: minimize (over all competing designs) the maximum (over all j) of the γ_j^{-1} or maximize (over all competing designs) the minimum (over all j) of the γ_j. This criterion allows maximization of power for a given value of $\underline{\rho}'\underline{\rho}/\sigma^2$, where $\underline{\rho}$ is a complete set of v-1 orthogonal normalized contrasts.

In the general case, it may not be easy to find a design that is optimal under any one of these criteria or to completely justify choosing one criteria to use over another. If the traces of the coefficient matrices were all equal (i.e., $\sum_{j=1}^{v-1} \gamma_{ij}$ = constant for i=1, 2, . . ., n competing designs), then the optima for all these criteria are reached when $\gamma_1 = \gamma_2 = \ldots = \gamma_{v-1}$. Thus Shah (1960) suggests choosing the design for which the γ_j have the least dispersion. That is, we want to minimize (over all competing designs with $\sum_{j=1}^{v-1} \gamma_j$ constant)

$$\mathrm{var}(\gamma) = \frac{\sum_{j=1}^{v-1} (\gamma_j - \gamma)^2}{v-1}.$$

Definition 5.6 (Shah (1960))

If the coefficient matrices of all competing designs have the same trace, then the S-optimal design is the one for which $\sum_{j=1}^{v-1} \gamma_j^2$ is a minimum.

Minimizing $\sum_{j=1}^{v-1} \gamma_j^2$ is equivalent to minimizing tr C^2. We will discuss later whether the assumption that the coefficient matrices of all competing designs have the same trace is reasonable.

In Chapter 6, we will discuss how we can improve an existing design with respect to S-optimality when we are improving it with respect to factor level utilization by pg-connecting the design.

Definition 5.7

For any two designs, D and D*, the S-better design will be the one which has the smaller tr C^2.

5.4 C-optimal Designs

S-optimality, as well as most other previously defined optimality criteria, is used solely in situations where we wish to optimize with respect to one factor, even though there may be two, three, or more factors in the experiment. It seems reasonable that, in many instances, one might desire to optimize with respect to two factors, since each may be of equal importance. With view toward this, we will describe and justify the use of a new optimality criterion, one that relates to both factors, A and B, of a completely randomized design under the assumption of no interaction.

We will denote the reduced normal equations for factor A by

$$C_A \hat{\underset{\sim}{\tau}}_A = \underset{\sim}{Q}_A$$

where factor A has m levels and the rank of C_A will be m-1 since we are considering connected designs. We will denote the m-1 non-zero eigenvalues of C_A by $\lambda_1, \lambda_2, \ldots, \lambda_{m-1}$.

Similarly the reduced normal equations for factor B will be denoted by

$$C_B \hat{\underset{\sim}{\tau}}_B = \underset{\sim}{Q}_B$$

where factor B has v levels and rank of C_B is v-1 since we are considering connected designs. The v-1 non-zero eigenvalues of C_B will be denoted by $\gamma_1, \gamma_2, \ldots, \gamma_{v-1}$.

We would like to find a design such that all the λ_i are equal and all the γ_j are equal. Such a design would be A, D, and E optimal for factor A as well as factor B. Let us consider the difficulty of finding a design for which the coefficient matrix for even one of the factors, say factor A, has all equal eigenvalues.

When searching for the S-optimal design for factor A, we require that all competing designs have C matrices, for factor A, with the same trace. That is $\sum_{i=1}^{m-1} \lambda_i = p$, for some constant p, for any competing design. If a design from this set has all equal eigenvalues, then $\lambda_i = \frac{p}{m-1}$ for i=1, 2, . . ., m-1. That is, we can find what this common value of λ is. Knowing this value of λ, however, does not allow us to find the C matrix for the design we seek, since the equation $|C_A - \lambda I| = 0$ can not be solved explicitly for C_A. Nor will any other procedure allow us to find the design for which all the eigenvalues of C_A are exactly equal. Thus for all practical purposes, we will not be able to find the design for which all the λ_i are equal or for which all the γ_j are equal.

Even though we cannot find the design for which all the λ_i are equal and all the γ_j are equal, we can find the design, from among all competing designs, that is closest (with respect to factor A and B together, not individually) to this unattainable end. At the same time, though, we want to protect ourselves against choosing a design that is very far from being S-optimal for one factor, even though it is close to S-optimal for the other (since the two factors are of equal importance). That is, we would like to minimize (over all competing designs)

$$\text{var}\ (\lambda+\gamma) = \frac{\sum_{i=1}^{m-1} (\lambda_i - \bar{\lambda})^2}{m-1} + \frac{\sum_{j=1}^{v-1} (\gamma_j - \bar{\gamma})^2}{v-1}$$

while protecting against extreme values for var (λ) or var (γ). We can accomplish this by choosing the design which is optimal according to the following new optimality criterion:

Definition 5.8

Let Φ be the class of all connected designs for which $\sum_{i=1}^{m-1} \lambda_i =$ constant for factor A and $\sum_{j=1}^{v-1} \gamma_j$ = constant for factor B. Let ϕ be the set of all designs in Φ for which

$$w^+ = [(v-1) \sum_{i=1}^{m-1} \lambda_i^2 + (m-1) \sum_{j=1}^{v-1} \gamma_j^2]$$

is a minimum. Design D will be C-optimal for (A,B) over set Φ if D is in ϕ and D has

$$w^- = \left|(v-1) \sum_{i=1}^{m-1} \lambda_i^2 - (m-1) \sum_{j=1}^{v-1} \gamma_j^2\right|$$

a minimum over all designs in ϕ.

Definition 5.9

For any two designs, D and D*, the C-better design will be the one that is C-optimal for the set Φ = {D, D*}.

Under certain conditions, we will be able to pseudo-globally connect an existing design and at the same time make the design C-better for (A,B).

For S and C-optimality, we have restricted the set of competing designs to only include those designs for which the coefficient matrices have the same trace. We will consider whether this is a reasonable restriction when comparing designs.

Recall that, for say factor B,

$$\text{tr } C_B = \sum_{j=1}^{v} r_j - \text{tr}[N \text{ diag}(k_1^{-1}, k_2^{-1}, \ldots, k_m^{-1}) N']$$

Now N diag$(k_1^{-1}, k_2^{-1}, \ldots, k_m^{-1})$ has a ji term of the form $\frac{n_{ji}}{k_i}$.

Thus the jj element of [N diag $(k_1^{-1}, k_2^{-1}, \ldots, k_m^{-1})$ N'] will be of the form $\frac{n_{j1}^{\;2}}{k_1} + \frac{n_{j2}^{\;2}}{k_2} + \ldots + \frac{n_{jm}^{\;2}}{k_m}$. Throughout this work, we will be interested in situations where there are some restrictions on the number of observations taken. For this reason, we will consider only binary designs. That is, every level combination will occur once or not at all in the design. (Certainly repeating a level combination would not help to 1-connect or pg-connect a design.) Since all of the n_{ji} will be 0 or 1, $n_{ji}^{\;2}=n_{ji}$ for all j,i. This means that

$$\text{tr } C_B = \sum_{j=1}^{v} r_j - \sum_{j=1}^{v} [\frac{n_{j1}}{k_1} + \frac{n_{j2}}{k_2} + \ldots + \frac{n_{jm}}{k_m}]$$

$$= \sum_{j=1}^{v} r_j - \sum_{i=1}^{m} [\sum_{j=1}^{v} \frac{n_{ji}}{k_i}]$$

$$= \sum_{j=1}^{v} r_j - \sum_{i=1}^{m} [1],$$

since $\sum_{j=1}^{v} \frac{n_{ji}}{k_i} = 1$ for all i. Thus

$$\text{tr } C_B = \sum_{j=1}^{v} r_j - m.$$

This means that we can satisfy the restriction that tr $C_B = \sum_{j=1}^{v-1} \gamma_j$ is a constant and tr $C_A = \sum_{i=1}^{m-1} \lambda_i$ is a constant by choosing designs from a set Φ, which is made up of designs with the same number of experimental points ($\sum_{j=1}^{v} r_j = \sum_{i=1}^{m} k_i$ = constant) and the same number of levels of B (v = constant) and the same number of levels of factor A (m = constant). Certainly we would have these conditions satisfied in most of the instances when we would be searching for an optimal design.

CHAPTER 6

IMPROVING LOCALLY CONNECTED DESIGNS

6.1 Introduction

In this chapter, we will consider how we can improve certain l-connected designs by 1) pg-connecting the designs and by 2) making the designs S or C-better. We will accomplish this by an interchange of factor levels between assemblies in the existing design.

Using the notation of Eccleston and Hedayat (1974), let Δ denote the family of all designs that are l-connected for some factor M. Let Δ_1 denote the family of designs that are pg-connected for factor M. This family will have the parameter set $[v, m, (r_j), (k_u)]$, where the j th level of a v-level factor, B say, is replicated r_j times and the u th level of an m-level factor, A say, is replicated k_u times.

6.2 Pseudo-globally Connecting a Design for One Factor

Consider a completely randomized design, D (two factors, A and B, and no interaction), with the following properties:

1) D is l-connected for factor B.

2) D is l-connected for factor A.

3) D is binary, i.e., $n_{ju} = 0$ or 1 for any level j of factor B and level u of factor A.

To consider how we might pg-connect such a design for factor B, we will look at the theorem by Eccleston and Hedayat (1972):

Theorem 6.1

A design D (two factors A and B, with no interaction) will be pg-connected for factor B if and only if

(1) D is 1-connected for B,

(2) every level of A occurs with at least two replicated levels of B, and

(3) if $a_u^{(i)}$ denotes any level of A with which b_i occurs and $a_u^{(i')}$ any level of A with which b_i does not occur, then there exists some $a_{u_\kappa}^{(i)}$ such that

(a) $a_{u_\kappa}^{(i)}$ occurs with a level of B that occurs with $a_{u_{\kappa'}}^{(i)}$ for $\kappa \neq \kappa'$ and with two different $a_u^{(i')}$ or

(b) $a_{u_\kappa}^{(i)}$ occurs with two levels of B, each of which occur with some $a_u^{(i')}$.

Under our restrictions, (1) and (2) of the theorem by Eccleston and Hedayat are satisfied. Thus there must be some level of B, b_i, such that (3) is not satisfied for b_i. We will perform an interchange of levels of B between levels of A so as to satisfy (3b) for b_i.

We will first consider the design as being made up of T_i and $D-T_i$, where T is made up of assemblies that contain levels of A that occur with b_i and $D-T_i$ is made up of assemblies that contain levels of A that do not occur with b_i. Since the design is 1-connected for B, some level of B, say b_ℓ, occurs in T_i and $a_{t'}b_p$ in $D-T_i$ by a_tb_p and $a_{t'}b_2$. The levels assemblies a_tb_z in T_i and $a_{t'}b_p$ in $D-T_i$ by a_tb_p and $a_{t'}b_z$. The levels a_t, $a_{t'}$, b_z, and b_p are chosen such that, prior to the interchange, b_z

is replicated in T_i, b_p is replicated in $D-T_i$, and b_ℓ occurs with a_t and $a_{t'}$. Recall that k_u for u=1, 2, . . ., m denotes the number of replicates for level u of factor A. We will denote min (k_u) over all u by k_{min}.

The following theorem will deal with the interchange to improve the design with respect to factor B. The algorithm used in this theorem will be similar to one used by Eccleston and Hedayat (1974), in connection with a lemma about proper, randomized block designs.

Theorem 6.2

Corresponding to any design D in $\Delta_2=\Delta-\Delta_1$ that is 1-connected for factor B, there is a corresponding design D* in Δ_1 that is pg-connected for factor B if less that $(k_{min}-2)$ of the r_j's are equal to one. In addition, D* will be S-better for factor B than D, if any of the following is satisfied:

(1) $k_{t'} = k_t$ or

(2) $k_{t'} < k_t$ and $(c_{pp} + c_{p\ell})-(c_{zz} + c_{z\ell}) < -2R + \frac{1}{R(k_t)} \sum_{\substack{u=1 \\ u\neq t}}^{m} \frac{\delta_u}{k_u}$ or

(3) $k_{t'} > k_t$ and $(c_{pp} + c_{p\ell})-(c_{zz} + c_{z\ell}) > -2R + \frac{1}{R(k_t)} \sum_{\substack{u=1 \\ u\neq t}}^{m} \frac{\delta_u}{k_u}$

where $R = \left(\frac{1}{k_{t'}} - \frac{1}{k_t} \right)$ and $\delta_u = \begin{cases} 1 \text{ if } b_z \text{ occurs with } a_u \\ 0 \text{ otherwise.} \end{cases}$

c_{pp}, $c_{p\ell}$, c_{zz}, and $c_{z\ell}$ are elements of the coefficient matrix for factor B, C_B, for design D, corresponding to the factor levels involved in the interchange b_p, b_ℓ, and b_z. The levels of A involved in the interchange are a_t and $a_{t'}$.

Proof of Theorem 6.2

Suppose level b_i of factor B fails to satisfy condition (3) of Theorem 6.1. Recall that the design can be divided into T_i, the set of assemblies that contain levels of A with which b_i occurs, and $D-T_i$. If b_i does not satisfy (3), then there does not exist a level of B that occurs twice in T_i and twice in $D-T_i$. However, since the design is 1-connected for factor B, there does exist some level of B, b_ℓ, such that b_ℓ occurs both in T_i and $D-T_i$. We will denote the level of A with which b_ℓ occurs in T_i by a_t and the level of A with which b_ℓ occurs in $D-T_i$ by $a_{t'}$. Since less than $(k_{min}-2)$ of the r_j's are equal to one, we know that there exists a level of B, denoted b_z, such that b_z occurs with a_t, $r_z>1$, $b_z \neq b_\ell$, and $b_z \neq b_i$. If b_z occurs in $D-T_i$, then b_z and b_ℓ would each occur with a level of A with which b_i occurs, and with a level of A with which b_i does not occur. This would contradict our hypothesis that D is not pg-connected for factor B, since condition (3) of Theorem 6.1 would be satisfied. Thus b_z must occur at least twice in T_i. Similarly, there must exist a level b_p, such that b_p occurs with $a_{t'}$, $r_p>1$, and $b_p \neq b_\ell$. It would also contradict our hypothesis if b_p is in T_i. Thus b_p occurs two or more times in $(D-T_i)$. Levels b_z and b_p of B are exchanged between levels a_t and $a_{t'}$ of A. That is, assemblies $a_t b_z$ and $a_{t'} b_p$ are replaced by assemblies $a_t b_p$ and $a_{t'} b_a$. Now two levels of B, b_ℓ and b_z each occur in T_i and in $D-T_i$. Thus (3b) of Theorem 6.1 is satisfied and we know the design is pg-connected for factor B.

We will now prove that the new design, D*, will be S-better than D for factor B, if any of the three conditions of Theorem 6.2 are satisfied.

The following elements of the coefficient matrix for factor B, C_D, are changed when we perform the interchange:

(a) $c_{zi} \longrightarrow c_{zi} + \frac{1}{k_t}$

(b) $c_{ze} \longrightarrow c_{ze} + \frac{1}{k_t}$ where b_e is a level of B that occurs with a_t; $b_e \neq b_z$, $b_e \neq b_\ell$. Since c_{zi} has been given in (a), there are $(k_t - 3)$ of these terms.

(c) $c_{pf} \longrightarrow c_{pf} + \frac{1}{k_{t'}}$ where b_f is a level of B that occurs with $a_{t'}$; $b_f \neq b_\ell$, $b_f \neq b_p$. There are $(k_{t'} - 2)$ of these terms.

(d) $c_{pe} = 0 \longrightarrow -\frac{1}{k_t}$ There are $(k_t - 2)$ of these terms, including c_{pi}

(e) $c_{zf} = 0 \longrightarrow -\frac{1}{k_{t'}}$ There are $(k_{t'} - 2)$ of these terms

(f) $c_{p\ell} \longrightarrow c_{p\ell} + (\frac{1}{k_{t'}} - \frac{1}{k_t}) = c_{p\ell} + R$ if we denote $(\frac{1}{k_{t'}} - \frac{1}{k_t})$ by R.

(g) $c_{pp} \longrightarrow c_{pp} + R$

(h) $c_{zz} \longrightarrow c_{zz} - R$

(i) $c_{z\ell} \longrightarrow c_{z\ell} - R$

All other elements of the coefficient matrix, C_D, remain the same.

Recall that the j,j' element of C_D^2, denoted by $(C_D^2)_{jj'}$, is given by $(C_D^2)_{jj'} = \sum_{k=1}^{v} c_{jk}c_{kj'} = \sum_{k=1}^{v} c_{jk}c_{j'k}$ by the symmetry of C_D. Thus

$$\operatorname{tr} C_D^2 = \sum_{k=1}^{v} \sum_{j=1}^{v} (c_{jk})^2$$

Then before the interchange, the trace of the square of the coefficient matrix would be

$$\mathrm{tr}\ C_D{}^2 = (c_{zi})^2 + \sum_e (c_{ze})^2 + \sum_f (c_{pf})^2 + (c_{p\ell})^2 + (c_{pp})^2 + (c_{zz})^2 + (c_{z\ell})^2 + \text{remainder}.$$

After the interchange the trace of C would be

$$\mathrm{tr}\ C_{D*}{}^2 = (c_{zi} + \frac{1}{k_t})^2 + \sum_e (c_{ze} + \frac{1}{k_t})^2 + \sum_f (c_{pf} + \frac{1}{k_{t'}})^2 + \sum_e (-\frac{1}{k_t})^2 + \sum_f (-\frac{1}{k_{t'}})^2 + (c_{p\ell}+R)^2 + (c_{pp}+R)^2 + (c_{zz}-R)^2 + (c_{z\ell}-R)^2 + \text{remainder}$$

where the remainders for $\mathrm{tr}\ C_D{}^2$ and $\mathrm{tr}\ C_{D*}{}^2$ are the same. The effect of the interchange on the $\mathrm{tr}\ C_D{}^2$ is reflected by

$$d = \mathrm{tr}\ C_D{}^2 - \mathrm{tr}\ C_{D*}{}^2$$

The quantity d is given by

$$d = -\frac{2c_{zi}}{k_t} - \frac{1}{k_t^2} - \sum_e \frac{2c_{ze}}{k_t} - \sum_e \frac{1}{k_t^2} - \sum_f \frac{2c_{pf}}{k_{t'}} - \sum_f \frac{1}{k_{t'}^2} - \sum_e \frac{1}{k_t^2} - \sum_f \frac{1}{k_{t'}^2} - 2c_{p\ell}R - R^2 - 2c_{pp}R - R^2 + 2c_{zz}R - R^2 + 2c_{z\ell}R - R^2$$

In order to prove that the new design is S-better, we need to show that $d > 0$. Recall that there are (k_t-3) terms of the form c_{ze}, $(k_{t'}-2)$ terms of the form C_{pf}, (k_t-2) terms of the form C_{pe}, and $(k_{t'}-2)$ terms of the form C_{zf}. Thus

$$d = -\frac{2c_{zi}}{k_t} - \frac{1}{k_t^2} - \sum_e \frac{2c_{ze}}{k_t} - (k_t-3)\frac{1}{k_t^2} - \sum_f \frac{2c_{pf}}{k_{t'}} - (k_t-2)\frac{1}{k_{t'}^2} - (k_t-2)\frac{1}{k_t^2} + - (k_{t'}-2)\frac{1}{k_{t'}^2} - 2c_{p\ell}R - R^2 - 2c_{pp}R - R^2 + 2c_{zz}R - R^2 + 2c_{z\ell}R - R^2$$

$$= \left(\frac{-2c_{zi}}{k_t} - \frac{2}{k_t^2}\right) + \left(\frac{-2}{k_t}\sum_e c_{ze} - \frac{2(k_t-3)}{k_t^2}\right) + \left(\frac{-2}{k_{t'}}\sum_f c_{pf} - \frac{2(k_{t'}-2)}{k_{t'}^2}\right) +$$

$$-2R(c_{p\ell}+c_{pp}-c_{zz}-c_{z\ell}) - 4R^2.$$

Recall that $-c_{sg} = \sum_{u=1}^{m} \frac{n_{su}n_{ug}}{k_u}$ for $s \neq g$ and $-c_{ss} = -r_s + \sum_{u=1}^{m} \frac{n_{su}}{k_u}$.

Since b_z only occurs in T_i,

$$-c_{iz} = \frac{1}{k_t} + \sum_{\substack{u=1\\u\neq t}}^{m} \frac{\delta_u}{k_u} \quad \text{where } \delta_u = \begin{cases} 1 \text{ if } b_z \text{ occurs with } a_u \\ 0 \text{ otherwise} \end{cases}$$

Thus

$$\left(\frac{-2c_{iz}}{k_t} - \frac{2}{k_t^2}\right) = \frac{2}{k_t}\sum_{\substack{u=1\\u\neq t}}^{m} \frac{\delta_u}{k_u} > 0 \quad \text{since } r_z > 1.$$

Level b_z occurs with a_t. This means that $-c_{ze} \geq \frac{1}{k_t}$ for all values of e,

and

$$\left(\frac{-2\sum_e c_{ze}}{k_t} - \frac{2(k_t-3)}{k_t^2}\right) \geq 0.$$

Level b_p occurs with $a_{t'}$. This means that $-c_{pf} \geq \frac{1}{k_{t'}}$ for all values of f, and

$$\left(\frac{-2\sum_f c_{pf}}{k_{t'}} - \frac{2(k_{t'}-2)}{k_{t'}^2}\right) \geq 0$$

Using the information above, we see that

$$d \geq \frac{2}{k_t}\sum_{\substack{u=1\\u\neq t}}^{m} \frac{\delta_u}{k_u} - 2R(c_{p\ell}+c_{pp}-c_{zz}-c_{z\ell}) - 4R^2$$

To insure that the new design is S-better than the old, we need to show that

$$2R(c_{p\ell}+c_{pp}-c_{zz}-c_{z\ell}) < -4R^2 + \frac{2}{k_t} \sum_{\substack{u=1 \\ u\neq t}}^{m} \frac{\delta_u}{k_u} \tag{6.1}$$

(a) If $k_{t'}=k_t$, then R=0, making the left side of (6.1) zero, while we know the right side to be positive. Thus (6.1) is satisfied and the new design is S-better.

(b) If $k_{t'}<k_t$, then $\frac{1}{k_{t'}} - \frac{1}{k_t} = R > 0$, and (6.1) is equivalent to

$(c_{pp}+c_{p\ell}) - (c_{zz}+c_{z\ell}) < -2R + \frac{1}{R(k_t)} \sum_{\substack{u=1 \\ u\neq t}}^{m} \frac{\delta_u}{k_u}$, an inequality, which

when satisfied, guarantees that d > 0 and that the new design is S-better.

(c) If $k_{t'}>k_t$, then R<0, and (6.1) is equivalent to

$(c_{pp}+c_{p\ell}) - (c_{zz}+c_{z\ell}) > -2R + \frac{1}{R(k_t)} \sum_{\substack{u=1 \\ u\neq t}}^{m} \frac{\delta_u}{k_u}$, an inequality, which

when satisfied, guarantees that d > 0 and that the new design is S-better.

Thus if we satisfy any one of the three conditions of Theorem 6.2, we know that the design obtained by use of our interchange is S-better than the original, as well as being pg-connected, which the original design was not.

Example 6.1

Consider the following situation: A chemist wants to study the effect on yield of various levels of concentration of a given compound at certain temperatures. There are seven different concentrations and five different temperatures. In the past, concentration levels 1, 2, 3, and 4 have been used with temperature levels 1, 2, and 3, while concentration levels 6 and 7 have been used with temperature levels 4 and 5. Concentration level 5 has been used with all temperatures. These facts (along with other possible restrictions on the experiment) might suggest the following design to utilize the 16 observations available, with factor A denoting temperature and factor B concentration:

a_1b_1	a_2b_1	a_3b_1	a_4b_5	a_5b_5
a_1b_2	a_2b_3	a_3b_2	a_4b_6	a_5b_6
a_1b_4	a_2b_4	a_3b_3	a_4b_7	a_5b_7
a_1b_5				

This design is 1-connected for factor B, concentration, but is not pg-connected for B. Levels b_1, b_6, and b_7 do not fulfill requirement (3) of Theorem 6.1. We will first consider how to pg-connect the design for level b_6.

Level b_5 occurs with a level of A that b_6 occurs with, namely a_4. Level b_5 also occurs with a level of A that b_6 does not occur with, a_1. Level b_7 occurs with a_4 and with another level of A that b_6 occurs with, a_5. Level b_4 occurs with a_1 and with another level of A that b_6 does not occur with, a_2. Using the notation of Theorem 6.2, let

$b_i=b_6$

$b_\ell=b_5$

$b_z=b_7$

$b_p=b_4$

$a_t=a_4$

$a_{t'}=a_1$

To pg-connect the design for b_6, we will interchange levels b_7 and b_4 between levels a_4 and a_1. That is, we replace a_4b_7 and a_1b_4 by a_4b_4 and a_1b_7. We can easily verify by looking at the new design that it is pg-connected for b_6 and b_1 and b_7 as well.

a_1b_1	a_2b_1	a_3b_1	a_4b_5	a_5b_5
a_1b_2	a_2b_3	a_3b_2	a_4b_6	a_5b_6
a_1b_7	a_2b_4	a_3b_3	a_4b_4	a_5b_7
a_1b_5				

If there were another level of B for which the design is not pg-connected then we would use the algorithm again, this time concentrating on the level of B that does not satisfy (3) of Theorem 6.1. Since all levels of B do satisfy the requirements of Theorem 6.1, we know that the new design is pg-connected for factor B.

Now we need to consider whether the sufficient conditions are met for improving the design (with respect to S-optimality) after using our interchange algorithm. We know that $k_{t'} > k_t$ since a_1 is replicated four times and a_4 is only replicated 3 times. Thus we need to show that

$$(c_{pp}+c_{p\ell}) - (c_{zz}+c_{z\ell}) > -2R + \frac{1}{R(k_t)} \sum_{\substack{u=1 \\ u\neq t}}^{m} \frac{\delta_u}{k_u}$$

For this example, we need to show that

$$(c_{44}+c_{45}) - (c_{77}+c_{75}) \geq -2R + \frac{1}{R(k_4)} \sum_{\substack{u=1 \\ u\neq 4}}^{5} \frac{\delta_u}{k_u} \tag{6.2}$$

where $R = \left(\frac{1}{k_1} - \frac{1}{k_4} \right) = \left(\frac{1}{4} - \frac{1}{3} \right) = - \frac{1}{12}$

Now,

$$c_{44} = r_4 - \sum_{u=1}^{5} \frac{n_{4u}}{k_u} = 2 - \left(\frac{1}{4} + \frac{1}{3}\right) \text{ and } -c_{77} = +2 - \left(\frac{1}{3} + \frac{1}{3}\right).$$

$$c_{45} = \sum_{u=1}^{5} \frac{n_{4u}n_{u5}}{k_u} = -\frac{1}{4} \text{ and } -c_{75} = \frac{1}{3} + \frac{1}{3}\ .$$

$$\sum_{\substack{u=1 \\ u\neq 4}}^{5} \frac{\delta_u}{k_u} = \frac{1}{3}\ ,$$

since $b_z = b_7$ only occurs with one level of A (a level that is replicated 3 times) if we ignore $a_t = a_4$.

The left side of (6.2) is then

$$2 - \frac{1}{4} - \frac{1}{3} - \frac{1}{4} - 2 + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = \frac{1}{2}\ .$$

The right side of (6.2) is

$$-2\left(-\frac{1}{12}\right) + \frac{1}{\left(-\frac{1}{12}\right)(3)}\left(\frac{1}{3}\right) = \frac{1}{6} - \frac{4}{3} = -\frac{7}{6}$$

Since (6.2) is satisfied, we know that the new design is S-better than the old design. To verify this, we will compare tr C^2 for the old design and tr C^2 for the new design.

Before the interchange,

$$C_D = \frac{1}{12}\begin{pmatrix} 25 & -7 & -8 & -7 & -3 & 0 & 0 \\ -7 & 17 & -4 & -3 & -3 & 0 & 0 \\ -8 & -4 & 16 & -4 & 0 & 0 & 0 \\ -7 & -3 & -4 & 17 & -3 & 0 & 0 \\ -3 & -3 & 0 & -3 & 25 & -8 & -8 \\ 0 & 0 & 0 & 0 & -8 & 16 & -8 \\ 0 & 0 & 0 & 0 & -8 & -8 & 16 \end{pmatrix} \quad \text{and tr } C_D^{\,2} = \frac{1}{144}(3440) = 23.\overline{8}$$

After the interchange, we get the following results:

$$C_{D*} = \frac{1}{12}\begin{pmatrix} 25 & -7 & -8 & -4 & -3 & 0 & -3 \\ -7 & 17 & -4 & 0 & -3 & 0 & -3 \\ -8 & -4 & 16 & -4 & 0 & 0 & 0 \\ -4 & 0 & -4 & 16 & -4 & -4 & 0 \\ -3 & -3 & 0 & -4 & 25 & -8 & -4 \\ 0 & 0 & 0 & -4 & -8 & 16 & -4 \\ -3 & -3 & 0 & 0 & -4 & -4 & 17 \end{pmatrix} \quad \text{and tr } C_{D*}^{\ 2} = \frac{1}{144}(3246) = 22.541\overline{6}$$

Since tr C^2 for the new design is less than tr C^2 for the old design, we know that the new design is S-better for factor B than was the old design.

Theorem 6.2 suggests the following procedure to be used with 1-connected designs:

(a) Using Theorem 6.1, determine whether the design is pg-connected for factor B. If the design is not pg-connected, then

(b) try to choose b_ℓ, b_z, b_p, a_t, and $a_{t'}$ in such a way that condition (1) of Theorem 6.2 is satisfied. If this is not possible, then

(c) try to choose b_ℓ, b_z, b_p, a_t, and $a_{t'}$ in such a way that condition (2) or (3) of Theorem 6.2 is satisfied.

If (b) or (c) above can be accomplished, then we can obtain a pg-connected design by using our interchange algorithm (thus guaranteeing that no replicate of a treatment is wasted) and at the same time obtain an S-better design.

It is interesting to note, that often there are a number of combinations of three levels of factor B and two levels of factor A that will satisfy the general definitions of b_ℓ, b_z, b_p, a_t, and $a_{t'}$. Usually at least one of these sets of combinations can be seen to satisfy one of the sufficient conditions of Theorem 6.2. If on some occasion, it should be impossible to satisfy any one of the three conditions, then the experimenter might look into the possibility of augmenting the design with $(k_t - k_{t'})$ additional assemblies so as to make $k_t = k_{t'}$. Any convenient (either physically or economically) level of factor B could be used with a_t or $a_{t'}$. Then Theorem 6.2 could be applied. Although we may not have the flexibility of adding some of the levels of B, it is possible that we can find some level that we can add.

One final word about the algorithm: Suppose design D is pg-connected for some level of B, b_w, and we perform the interchange to pg-connect the design for b_i. The new design, D*, will still be pg-connected for b_w. This has to be true, since b_z was replicated in T_i and b_p was replicated in $D-T_i$.

6.3 Optimizing with Respect to Both Factors

Suppose we are not just interested in factor B, but in factor A as well. If design D is 1-connected for factor B, then we know that it is also 1-connected for factor A. We want to consider what happens to factor A when we perform the algorithm of Theorem 6.2 to pg-connect the design for factor B. We know that we do not destroy the 1-connectedness of the design with respect to A. We need to consider, however, whether the new design is as good or better (with respect to S-optimality) than the original design with respect to factor A.

Consider a completely randomized design D (two factors, A and B, and no interaction) with the following properties:

1) D is 1-connected for factor B.

2) All levels of factor B are equally replicated, i.e. $r_1=r_2= \ldots =r_v=r$

3) D is 1-connected for factor A. All levels of A do not necessarily have the same number of replicates.

4) D is binary, i.e. $n_{ju} = 0$ or 1 for any level j of factor B and level u of factor A.

Except for 2) above, these are the same restrictions as we had for Theorem 6.2. We will denote the coefficient matrix for factor A prior to the interchange to pg-connect the design for factor B, by C_A, and the coefficient matrix for A after the interchange to pg-connect for B, by $C_{A|I_B}$. The following theorem tells us that when we use the interchange to pg-connect the design for B and make the design S-better for B, we also make the design S-better for factor A.

Theorem 6.3

If the algorithm of Theorem 6.2 is used to pg-connect design D for factor B, then tr $C_{A|I_B}{}^2$ is less than tr $C_A{}^2$, giving an S-better design for factor A.

Proof:

Consider the coefficient matrix for factor A, C_A. The following elements of C_A are changed, as indicated, by the interchange that we used to make design D pg-connected for factor B. (Recall from Theorem 6.2, how levels b_z, b_p, and b_ℓ of factor B are defined and how levels a_t and $a_{t'}$ of factor A are defined.)

(a) $c_{ts} \longrightarrow c_{ts} + \frac{1}{r}$ where b_z occurs with a_s and $a_s \neq a_t$. There are (r-1) of these terms.

(b) $c_{t's}=0 \longrightarrow c_{t's} - \frac{1}{r}$ There are (r-1) of these terms.

(c) $c_{tv} \longrightarrow c_{tv} - \frac{1}{r}$ where b_p occurs with a_v and $a_v \neq a_{t'}$. There are (r-1) of these terms.

(d) $c_{t'v} \longrightarrow c_{t'v} + \frac{1}{r}$ There are (r-1) of these terms.

Thus $\mathrm{tr}\, C_A{}^2 - \mathrm{tr}\, C_{A|I_B}{}^2$ will be equal to

$$d' = -\sum_s \frac{2c_{ts}}{r} - \sum_s \frac{1}{r^2} + \sum_s \frac{2c_{t's}}{r} - \sum_s \frac{1}{r^2} + \sum_v \frac{2c_{tv}}{r} - \sum_v \frac{1}{r^2} - \sum_v \frac{2c_{t'v}}{r} - \sum_v \frac{1}{r^2}$$

$$= \frac{2}{r} \sum_s [c_{t's} - c_{ts} - \frac{1}{r}] - 2 \sum_v (\frac{c_{t'v}}{r} + \frac{1}{r^2}) + \sum_v \frac{2c_{tv}}{r}$$

Since b_z only occurs in T_i, a_t and a_s occur with both b_z and b_i. This means that $-c_{ts} \geq \frac{2}{r}$. If $-c_{t's}=0$, then $c_{t's}-c_{ts} \geq \frac{2}{r}$ for all s. If $-c_{t's} \neq 0$, then b_ℓ, and only b_ℓ, occurs with both $a_{t'}$ and a_s, and $-c_{t's} = \frac{1}{r}$. If b_ℓ does occur with a_s, then $-c_{ts} \geq \frac{3}{r}$, and once again $c_{t's} - c_{ts} \geq \frac{2}{r}$ for all s. Thus for all cases, $c_{t's} - c_{ts} \geq \frac{2}{r}$ or $c_{t's} - c_{ts} > \frac{1}{r}$ or

$$\frac{2}{r} \sum_s [c_{t's} - c_{ts} - \frac{1}{r}] > 0.$$

If c_{tv} is different from zero, we know that b_ℓ, and only b_ℓ, occurs with both a_t and a_v, or D would have been pg-connected for factor B before the interchange. We also have b_ℓ occurring with $a_{t'}$. Since b_ℓ and b_p both occur with a_v and $a_{t'}$, $-c_{t'v} \geq \frac{2}{r}$, while $-c_{tv} + \frac{1}{r}$ since only b_ℓ occurs with a_t and a_v. Thus

$$-(c_{t'v} + \frac{1}{r}) + c_{tv} \geq 0$$

for all s and v. This means that

$$-2 \sum_v (\frac{c_{t'v}}{r} + \frac{1}{r^2}) + \sum_v \frac{2c_{tv}}{r} \geq 0.$$

We have shown that d' is greater than zero if c_{tv} is different from zero. If c_{tv} is zero, then $-c_{t'v} \geq \frac{1}{r}$ means that

$$-2 \sum_v (\frac{c_{t'v}}{r} + \frac{1}{r^2}) \geq 0.$$

Thus d' is greater than zero if c_{tv} is zero. This proves that the new design is S-better than the old design with respect to factor A or that

$$\mathrm{tr}\; C_{A|I_D}{}^2 < \mathrm{tr}\; C_A{}^2.$$

Example 6.2

Design D

a_1b_1	a_2b_4	a_4b_7
a_1b_2	a_3b_1	a_5b_5
a_1b_3	a_3b_2	a_5b_6
a_1b_4	a_3b_3	a_5b_7
a_2b_1	a_3b_5	a_6b_5
a_2b_2	a_4b_4	a_6b_6
a_2b_3	a_4b_6	a_6b_7

Design D is not pg-connected with respect to factor A. If we use our algorithm of Theorem 6.2, we see that by replacing assemblies a_2b_3 and a_5b_5 by a_2b_5 and a_5b_3, we obtain the following design which is pg-connected for A:

Design D*

a_1b_1	a_2b_4	a_4b_7
a_1b_2	a_3b_1	a_5b_3
a_1b_3	a_3b_2	a_5b_6
a_1b_4	a_3b_3	a_5b_7
a_2b_1	a_3b_5	a_6b_5
a_2b_2	a_4b_4	a_6b_6
a_2b_5	a_4b_6	a_6b_7 .

Design D* is still not pg-connected for factor B. Levels b_6 and b_7 fail to satisfy condition (3) of Theorem 6.1. We can pg-connect the design for B by replacing a_4b_7 and a_2b_1 by a_4b_1 and a_2b_7. This final design, D**, will be pg-connected for factor B:

Design D**

a_1b_1	a_2b_4	a_4b_1
a_1b_2	a_3b_1	a_5b_3
a_1b_3	a_3b_2	a_5b_6
a_1b_4	a_3b_3	a_5b_7
a_2b_7	a_3b_5	a_6b_5
a_2b_2	a_4b_4	a_6b_6
a_2b_5	a_4b_6	a_6b_7

The final design, D**, is still pg-connected for factor A, as will usually be the case. We want to verify that D** is S-better for A than D* (and of course S-better than D) for factor A. When we compute the coefficient matrices for factor A, we see that

$\text{tr } C_A{}^2 = 44.\overline{8}$ for design D

$\text{tr } C_A{}^2 = 43.\overline{3}$ for design D*.

$\text{tr } C_A{}^2 = 41.\overline{5}$ for design D**.

Thus the design obtained when pg-connecting the design for factor B is S-better for factor A than the design prior to pg-connecting for B.

Theorems 6.2 and 6.3 taken together, suggest the following procedure for a given design D, which satisfies the conditions of Theorem 6.3:

(1) Use the algorithm of Theorem 6.2 to pg-connect the design for factor A, where all levels of A do not necessarily occur the same number of times. The new design, D*, will be S-better than D with respect to A.

(2) Use the algorithm of Theorem 6.2 to pg-connect the design for factor B, where all levels of B must occur the same number of times. This final design, D**, will be S-better than D* with respect to B. By Theorem 6.3, we know that D** will be S-better than D* with respect to A, and S-better than D with respect to A.

At this point, let us refer back to our definition of C-optimality (section 5.4), and consider the following family of designs: Let Ω denote the family of all designs that are connected for both factors, A and B. Let Ω_1 denote those designs which are pg-connected for factor B. Using Definition 5.9 and the results of Theorems 6.2 and 6.3, we obtain the following:

<u>Theorem 6.4</u>

Corresponding to any design D in $\Omega_2 = \Omega - \Omega_1$, there is a design in Ω_1

which is C-better for (A,B) if the following conditions are satisfied:

(1) The design D is binary.

(2) a. The levels of factor B are equally replicated. That is,

$r_1=r_2= \ . \ . \ . \ =r_v=r.$

b. $r \geq 3$

(3) Less than (r-2) of the k_u's are equal to one.

(4) One of the following is satisfied for factor B:

a. $k_{t'} = k_t$ or

b. $k_{t'} < k_t$ and $(c_{pp} + c_{p\ell})-(c_{zz} + c_{z\ell}) < -2R + \frac{1}{R(k_t)} \sum_{\substack{u=1 \\ u \neq t}}^{m} \frac{\delta_u}{k_u}$ or

c. $k_{t'} > k_t$ and $(c_{pp} + c_{p\ell})-(c_{zz} + c_{z\ell}) > -2R + \frac{1}{R(k_t)} \sum_{\substack{u=1 \\ u \neq t}}^{m} \frac{\delta_u}{k_u}$

where $R = \left(\frac{1}{k_{t'}} - \frac{1}{k_t} \right)$ and $\delta_u = \begin{cases} 1 \text{ if } b_z \text{ occurs with } a_u \\ 0 \text{ otherwise.} \end{cases}$

c_{pp}, $c_{p\ell}$, c_{zz}, and $c_{z\ell}$ are elements of the coefficient matrix for factor B, C_B, after the interchange to pg-connect the design for A, but prior to the interchange to pg-connect the design for B. The factor levels involved in the interchange to pg-connect the design for B are b_p, b_ℓ, b_z, a_t, and $a_{t'}$.

CHAPTER 7

IMPROVING LOCALLY CONNECTED DESIGNS WHEN AN ADDITIONAL BLOCK EFFECT IS PRESENT

7.1 Introduction

In Chapters 5 and 6, we used an interchange algorithm to improve two-factor, no interaction designs. In this chapter, as before, we will be interested in obtaining a C-better design for two factors, A and B, but under a three-factor, no interaction model. The third factor may be thought of as a block effect. We would like to obtain some of the same results here as for the two-factor situation of Chapter 6.

7.2 Three-factor, Main Effect Designs

Previously we had written our model as

$$E(y_{uj}) = \mu + \alpha_u + \beta_j$$

where μ is an overall mean, α_u is the effect of the u th level of factor A, and β_j is the effect of the j th level of factor B. The new model will have an additional term for block effect, ω_g, giving

$$E(y_{ujg}) = \mu + \alpha_u + \beta_j + \omega_g$$

where ω_g is the effect of the g th level of the factor which we will refer to as blocks. As before, we are assuming that no interactions exist.

Suppose we consider [(factor B) x (block)] level combinations as being levels of some new pseudo-factor, which we will denote F'. Then we can think of our experiment as having only two factors, A and F'. Since this is now the same situation as in Chapter 6, we can compute the coefficient matrix, C_A, for factor A, and consider finding the S-optimal design for factor A from the appropriate designs available. Then we can consider [(factor A) x (block)] level combinations to be levels of a new pseudo-factor, F", and obtain the C matrix for B, C_B.

We will consider only those designs that fulfill the requirements of Theorem 6.3 for factor A (with respect to [(factor B) x (block)] combinations) and for factor B (with respect to [(factor A) x (block)] combinations). First we will treat the design as a two-factor, A and F', design. Each level of F' must be replicated r times as were all levels of factor B in Theorem 6.3. We can use the interchange algorithm to pg-connect the design for factor A, and make the design S-better for factor A. We can do this under the conditions of Theorem 6.2 for factors A and F'. We can also use the interchange algorithm to pg-connect the design for factor B, under the conditions of Theorem 6.2 for factors B and F".

We see that we can pg-connect the design for factor A, thus making the design S-better for A, or pg-connect the design for factor B, making the design S-better for B. But, what are the consequences of attempting to do both? When we perform the second interchange to pg-connect for B, do we lose the benefits of the first exchange to pg-connect for factor A? We would like to be able to prove a theorem corresponding to Theorem 6.3

for the present situation.

7.3 The Effect of Pseudo-globally Connecting for Both Factors

As in Chapter 6, we will denote the coefficient matrix for factor A prior to the interchange to pg-connect the design for factor B, by C_A, and the coefficient matrix for A after the interchange by $C_{A|I_B}$.

Theorem 7.1

Consider any design, D, under a three-factor (A, B, and block effect), no interaction model, that has the following properties:

1) D is 1-connected for factor B.
2) All levels of factor F' are equally replicated, where levels of F' are [B x block] level combinations.
3) D is 1-connected for factor A.
4) D is binary.

If the algorithm of Theorem 6.2 is used to pg-connect design D for factor B, then tr $C_{A|I_B}{}^2$ < tr $C_A{}^2$, giving an S-better design for factor A.

For the proof of Theorem 7.1 we will proceed along the same lines as that of Theorem 6.3. We shall consider the elements of C_A that are altered by the interchange for B, and compare the sums of the squares of these elements before and after the interchange.

Proof:

Consider the coefficient matrix for factor A, C_A. Suppose the interchange necessary to pg-connect design D for factor B, is to replace assemblies $a_jb_zc_k$ and $a_{j'}b_pc_{k'}$ by $a_jb_pc_k$ and $a_{j'}b_zc_{k'}$. That is, we interchange levels z and p of factor B. Level combinations a_jc_k and $a_{j'}c_{k'}$, play the role of the a_t and $a_{t'}$ of Theorem 6.2 respectively. The levels of A that are involved in the interchange will be a_j and $a_{j'}$.

Let a_x be any level of factor A such that $b_z c_k$ occurs with a_x, but $a_x \neq a_j$. After the interchange, a_j occurs with $b_z c_k$ one less time. (Note that a_x could not have occurred with $b_p c_k$ since our interchange algorithm requires that b_z and b_p never occur with the same [A x blocks] combination.) Thus for the coefficient matrix for factor A,

(1) $c_{jx} \longrightarrow c_{jx} + \frac{1}{r}$ There are (r-1) of these terms. Let a_y be any level of A such that $b_p c_k$ occurs with a_y, but $a_y \neq a_{j'}$. Levels a_j and a_y will occur with one more common level of F' after the interchange, giving

(2) $c_{jy} \longrightarrow c_{jy} - \frac{1}{r}$ There are (r-1) of these terms. Let a_q be any level of A such that $b_p c_{k'}$ occurs with a_q, but $a_q \neq a_{j'}$. Let a_o be any level of A such that $b_z c_{k'}$ occurs with a_o, but $a_o \neq a_j$.

(3) $c_{j'q} \longrightarrow c_{j'q} + \frac{1}{r}$ There are (r-1) of these terms.

(4) $c_{j'o} \longrightarrow c_{j'o} - \frac{1}{r}$ There are (r-1) of these terms. All other elements of the C matrix for factor A will remain the same.

If b_i is the level of factor B that fails to satisfy condition (3) of Theorem 6.1, then we can let T_i and $(D-T_i)$ denote the two parts of D, as we did in Theorem 6.2. Recall that only b_ℓ occurs in both T_i and $(D-T_i)$, while b_z occurs two or more times in T_i, but never in $(D-T_i)$. The level b_p occurs two or more times in $(D-T_i)$, but never in T_i. Now a_x and a_j both occur with $b_z c_k$ and $b_i c_k$. Also, a_q and $a_{j'}$ both occur with $b_p c_{k'}$. Thus

$-c_{jx} \geq \frac{2}{r}$, just by the contribution of assemblies involving block k, and

$-c_{j'q} \geq \frac{2}{r}$, just by the contribution of assemblies involving block k'.

If we let $a_{x|k}$ denote all replicates of a_x that occur with c_k, and consider all possible values of x, we get

a) $-\sum_{x|k} (c_{jx} + \frac{2}{r}) \geq 0.$

If we let $a_{q|k'}$ denote all replicates of a_q that occur with $c_{k'}$, and consider all (r-1) values of q, we get

b) $-\sum_{q|k'} c_{j'q} \geq \frac{2r-3}{r} + \frac{1}{r}$

There are several combinations of values for the portion of c_{jy} that is contributed by assemblies involving block k, and for the portion of $c_{j'q}$ that is contributed by assemblies involving block k'. Since D is not pg-connected prior to the interchange, we know that the only possible common level of B for a_jc_k and a_yc_k would be b_ℓ, since a_jc_k is in T_i and a_yc_k is in $(D-T_i)$. If b_ℓ does occur in a_yc_k, then there could not be any common level of B for $a_oc_{k'}$ and $a_{j'}c_{k'}$, since $a_oc_{k'}$ is in T_i and $a_{j'}c_{k'}$ is in $(D-T_i)$. If there was a common level of B for these two combinations, then either (3a) or (3b) of Theorem 6.1 would be satisfied. Denote by $a_{y|k}$, those replications of a_y that occur with c_k. Denote by $a_{o|k'}$, those replications of a_o that occur with $c_{k'}$. Then

c) $-(\sum_{y|k} c_{jy}) - (\sum_{o|k'} c_{j'o}) = \frac{1}{r}$ if b_1 occurs with a_yc_k.

Using the same type reasoning, the only possible common level of B for

$a_oc_{k'}$ and $a_{j'}c_{k'}$ would be b_ℓ. If $a_oc_{k'}$ and $a_{j'}c_{k'}$ do have b_ℓ in common, though, there can be no common level of B for a_jc_k and a_yc_k. Thus,

d) $-(\sum_{y|k} c_{jy})-(\sum_{o|k'} c_{j'o}) = -\frac{1}{r}$ if b_1 occurs with $a_oc_{k'}$

The only other possibility is that a_{jk} and a_{yk} have no level of B in common, and $a_{ok'}$ and $a_{j'k'}$ have no level of B in common. This gives us

e) $-(\sum_{y|k} c_{jy})-(\sum_{o|k'} c_{j'o}) = 0$ if a_jc_k and a_yc_k have no level of B in common and $a_oc_{k'}$ and $a_{j'}c_{k'}$ have no level of B in common.

Combining c), d), and e), we get

f) $-(\sum_{y|k} c_{jy})-(\sum_{o|k'} c_{j'o}) \leq \frac{1}{r}$

If we use b) and f) together, we get

g) $\sum_{q|k'} -c_{j'q} + \sum_{y|k} c_{jy} + \sum_{o|k'} c_{j'o} \geq \frac{2r-3}{r}$

We will use g), along with a) to show that the new design will be S-better for A than the design prior to the interchange for B.

All the changes in the C matrix for A will be caused by the rearrangement of assemblies involving c_k or $c_{k'}$. Thus we can write the trace of C_A^2 before the interchange for B, as

$$\mathrm{tr}\; C_A^2 = \sum_{x|k} c_{jx}^2 + \sum_{y|k} c_{jy}^2 + \sum_{q|k'} c_{j'q}^2 + \sum_{o|k'} c_{j'o}^2 + $$

$$+ \text{ remainder} \qquad \text{and}$$

the trace of C_A^2 after the interchange for B as

$$\operatorname{tr} C_A^2|_{I_B} = \sum_{x|k} (c_{jx} + \frac{1}{r})^2 + \sum_{y|k} (c_{jy} - \frac{1}{r})^2 + \sum_{q|k'} (c_{j'q} + \frac{1}{r})^2 +$$

$$+ \sum_{o|k'} (c_{j'o} - \frac{1}{r})^2 + \text{remainder}$$

where the remainders are the same.

The difference between $\operatorname{tr} C_A^2$ and $\operatorname{tr} C_{A|I_B}^{\ 2}$ may be written as

$$d'' = \sum_{x|k} (\frac{-2c_{jx}}{r} - \frac{1}{r^2}) + \sum_{y|k} (\frac{2c_{jy}}{r} - \frac{1}{r^2}) + \sum_{q|k'} (\frac{-2c_{j'q}}{r} - \frac{1}{r^2}) +$$

$$+ \sum_{o|k'} (\frac{2c_{j'o}}{r} - \frac{1}{r^2})$$

Recall that there are (r-1) terms in each sum. Thus

$$d'' = \frac{2}{r} [- \sum_{x|k} (c_{jx} + \frac{2}{r})] + \frac{2}{r} \quad [\sum_{q|k'} -c_{j'q} + \sum_{y|k} c_{jy} + \sum_{o|k'} c_{j'o}]$$

From a) and g), we see that

$$d'' \geq \frac{2r-3}{r} \left(\frac{2}{r} \right) > 0$$

Thus $\operatorname{tr} C_{A|I_B}^{\ 2} < \operatorname{tr} C_A^2$ and the new design is S-better for factor A.

<u>Example 7.1</u>

Consider the following agricultural experiment. A grower wants to compare the effects of six different varieties of corn as well as seven different fertilizers. Because of the way his land is situated, he must also contend with an elevation effect. That is, all of one section of land to be used in the experiment is at approximately the same elevation, while the other section to be used is at a much higher elevation. He believes that elevation may be significant with respect to yield.

In the past, usually the first three types of fertilizer have been used with certain, particular (corn x elevation) combinations, while the last three varieties of corn have been used with different (corn x elevation) combinations. The grower is unable to plant all (6)(7)(2) = 84 combinations because of the burden of a large harvest. Based on past experience and harvesting restriction, the grower has suggested the following design:

<u>D</u>

$a_1b_1c_1$	$a_1b_1c_2$	$a_2b_1c_1$	$a_2b_1c_2$	$a_3b_1c_1$	$a_3b_1c_2$
$a_1b_2c_1$	$a_1b_2c_2$	$a_2b_2c_1$	$a_2b_2c_2$	$a_3b_2c_1$	$a_3b_2c_2$
$a_1b_3c_1$	$a_1b_3c_2$	$a_2b_3c_1$	$a_2b_3c_2$	$a_3b_3c_1$	$a_3b_3c_2$
$a_1b_4c_1$	$a_1b_4c_2$		$a_2b_4c_2$		

$a_4b_4c_1$	$a_4b_4c_2$	$a_5b_4c_1$	$a_5b_5c_2$	$a_6b_5c_1$	$a_6b_5c_2$
$a_4b_5c_1$	$a_4b_5c_2$	$a_5b_5c_1$	$a_5b_6c_2$	$a_6b_6c_1$	$a_6b_6c_2$
$a_4b_6c_1$	$a_4b_6c_2$	$a_5b_6c_1$	$a_5b_7c_2$	$a_6b_7c_1$	$a_6b_7c_2$
$a_4b_7c_1$	$a_4b_7c_2$	$a_5b_7c_1$			

where variety of corn is factor A, type of fertilizer is factor B, and elevation (block effect) is factor C. The above design is pg-connected for factor A. It is not pg-connected for factor B, however, since level b_1, among others, does not satisfy (3) of the theorem by Eccleston and Hedayat, Theorem 6.1. We can pg-connect the design for factor B by replacing assemblies $a_1b_3c_1$ and $a_4b_5c_1$ by $a_1b_5c_1$ and $a_4b_3c_1$. If we change the order of the assemblies slightly, we can see that the final design is still pg-connected for factor A. The final design is

D*

$a_1b_1c_1$	$a_1b_1c_2$	$a_1b_2c_1$	$a_1b_2c_2$	$a_4b_3c_1$
$a_2b_1c_1$	$a_2b_1c_2$	$a_2b_2c_1$	$a_2b_2c_2$	$a_3b_3c_1$
$a_3b_1c_1$	$a_3b_1c_2$	$a_3b_2c_1$	$a_3b_2c_2$	$a_2b_3c_1$

$a_1b_3c_2$	$a_1b_4c_1$	$a_1b_4c_2$	$a_1b_5c_1$	$a_5b_5c_2$
$a_2b_3c_2$	$a_4b_4c_1$	$a_2b_4c_2$	$a_5b_5c_1$	$a_4b_5c_2$
$a_3b_3c_2$	$a_5b_4c_1$	$a_4b_4c_2$	$a_6b_5c_1$	$a_6b_5c_2$

$a_4b_6c_1$	$a_4b_6c_2$	$a_4b_7c_1$	$a_4b_7c_2$
$a_5b_6c_1$	$a_5b_6c_2$	$a_5b_7c_1$	$a_5b_7c_2$
$a_6b_6c_1$	$a_6b_6c_2$	$a_6b_7c_1$	$a_6b_7c_2$

We still need to confirm that this final design is S-better for A than the one prior to the interchange for B. Using the notation of Theorem 7.1. we see that

$a_j=a_1$ $\qquad$ $a_x=a_2$ or a_3.

$a_{j'}=a_4$ $\qquad$ $a_y=a_5$ or a_6

$b_zc_k=b_3c_1$ $\qquad$ $a_q=a_5$ or a_6

$b_pc_k=b_5c_1$ $\qquad$ $a_o=a_2$ or a_3.

$b_pc_{k'}=b_5c_1$

$b_zc_{k'}=b_3c_1$

Thus the elements that are altered in C_A are the following:

c_{12}, c_{13}, c_{15}, c_{16}, c_{45}, c_{46}, c_{42}, and c_{43}.

We can write tr C_A^2 as $c_{12}^2+c_{13}^2+c_{15}^2+c_{16}^2+c_{45}^2+c_{46}^2+c_{42}^2+c_{43}^2+$ a remainder that contains all the terms of tr C_A^2 that are not altered by the interchange to connect the design for factor B. If we evaluate the appropriate elements of C_A, we see that

$$\text{tr } C_A^{\,2} = (-\tfrac{7}{3})^2+(-\tfrac{6}{3})^2+(-\tfrac{1}{3})^2+(0)^2+(-\tfrac{7}{3})^2+(-\tfrac{6}{3})^2+(-\tfrac{1}{3})^2+(0)^2 + \text{remainder}$$

and

$$\text{tr } C_{A|I_B}^{\,2} = (-\tfrac{6}{3})^2+(-\tfrac{5}{3})^2+(-\tfrac{2}{3})^2+(-\tfrac{1}{3})^2+(-\tfrac{6}{3})^2+(-\tfrac{5}{3})^2+(-\tfrac{2}{3})^2+(-\tfrac{1}{3})^2$$
$$+ \text{ remainder},$$

where the two remainders are identical.

$$\text{tr } C_A^{\,2} - \text{tr } C_{A|I_B}^{\,2} = 19.\bar{1} - 14.\bar{6} = 4.\bar{4}$$

Thus the design obtained when we pg-connect for factor B is S-better for factor A than the design prior to pg-connecting for B.

7.4 C-better Designs under the Three-factor Model

Let Ω denote the family of all three-factor (A, B, and block effect), no interaction designs that are 1-connected for both A and B. Let Ω_1 denote those designs in Ω that are pg-connected for factor B. The proof of the following theorem can be obtained directly from the proofs of Theorems 6.2 and 7.1:

Theorem 7.2

Corresponding to any design D in $\Omega_2 = \Omega - \Omega_1$, there is a design in Ω_1 which is C-better for (A,B) if the following conditions are satisfied:

(1) The design D is binary.

(2) The [B x block] level combinations are all replicated r times, where $r \geq 3$.

(3) At least (r-2) of the levels of factor A are replicated.

(4) If the i,j level combination of [A x block] is replicated k_{ij} times, then at least $(k_{min}-2)$ of the levels of factor B are replicated.

(5) One of the following is satisfied for factor B:

a. $k_{t'} = k_t$ or

b. $k_{t'} < k_t$ and $(c_{pp} + c_{p\ell})-(c_{zz} + c_{z\ell}) < -2R + \frac{1}{R(k_t)} \sum\limits_{\substack{u=1 \\ u \neq t}} \frac{\delta_u}{k_u}$ or

c. $k_{t'} > k_t$ and $(c_{pp} + c_{p\ell}) > -2R + \frac{1}{R(k_t)} \sum\limits_{\substack{u=1 \\ u \neq t}} \frac{\delta_u}{k_u}$

where $R = \left(\frac{1}{k_{t'}} - \frac{1}{k_t}\right)$ and $\delta_u = \begin{cases} 1 & \text{if } b_z \text{ occurs with the u th level combination of [A x block]} \\ 0 & \text{otherwise} \end{cases}$

c_{pp}, $c_{p\ell}$, c_{zz}, and $c_{z\ell}$ are elements of the coefficient matrix for factor B after the interchange to pg-connect the design for factor A, but prior to the interchange for B. The factor levels involved in the interchange are b_p, b_ℓ, and b_z, and the [A x block] level combinations are those denoted by t and t'.

CHAPTER 8

EFFECTS OF THE INTERCHANGE ALGORITHM WHEN THE ASSUMED MODEL IS INADEQUATE

8.1 Introduction

In Chapters 6 and 7, we discussed a procedure to improve an experimental design under the criterion of S-optimality. This was accomplished by performing a permutation of the factor levels to form certain new factor level combinations. This procedure was suggested only for situations where the assumption of no interaction was believed to be appropriate. Even though the experimenter may make the assumption of no interaction, he often will want to protect somehow against the possibility that the assumption is unjustified. That is, as in Chapter 5, the assumed model may be $E(y_{ij}) = \mu + \alpha_i + \beta_j$, while the true model may be $E(y_{ij}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$. The experimenter may try to choose a design for which the quality (based on some criterion) of parameter estimation is relatively good, even if the assumed model is incorrect. If we perform our algorithm of Chapter 6, we know that we can often improve the design with respect to more utilization of observations in the estimation procedure and with respect to S or C-optimality. The question we need to consider now, though, is whether this same interchange might hurt the design with respect to protection against the wrong model. First we shall summarize some considerations that have been used when choosing a design to minimize the effect of the assumption of the wrong model.

8.2 Protecting Against an Inadequate Model

We will denote the full (true) linear model by

$$E(\underline{y}_F) = X(\underline{\Theta}) = (X_1, X_2)\binom{\underline{\Theta}_1}{\underline{\Theta}_2} \quad \text{and}$$

the reduced (assumed) linear model by

$$E(\underline{y}_R) = X_1(\underline{\Theta}_1)$$

Let $\hat{\underline{y}}_R$ be the estimate for the observation vector under the reduced model. Usually the quantity, mean squared error of $\hat{\underline{y}}_R$, averaged over the region of experimentation, R, is considered when studying the effects of an inadequate model. This value, average m.s.e. $(\hat{\underline{y}}_R)$, is defined by the relationship

$$J = \frac{n\Omega}{\sigma^2} \int_R E[\hat{y}(\underline{x})_R - \eta(\underline{x})]^2 d\underline{x}$$

where σ^2 is the error variance, the observed response at the i th experimental point, $\underline{x}_i$, is $\eta(\underline{x}_i) + \varepsilon_i$, and $\hat{y}(\underline{x}_i)_R$ is the estimate for a response at the i th experimental point under the assumed model. The term Ω is defined by $\Omega^{-1} = \int_R d\underline{x}$. The number of experimental points is given by n.

The quantity, J, may be naturally broken down into average variance of $\hat{\underline{y}}_R$ and average squared bias of $\hat{\underline{y}}_R$ (e.g., Myers, (1971)). That is

$$J = V + B$$

where V, the average variance of $\hat{\underline{y}}_R$, is given by

$$V = \frac{n\Omega}{\sigma^2} \int_R E[\hat{y}(\underline{x})_R - E\{\hat{y}(\underline{x})_R\}]^2 d\underline{x}$$

and B, the average squared bias of $\hat{\underline{y}}_R$, is given by

$$B = \frac{n\Omega}{\sigma^2} \int_R [E\{\hat{y}(\underline{x})_R\} - \eta(\underline{x})]^2 d\underline{x}$$

Prior to 1959, most of the experimental designs suggested in this situation, were for the purpose of minimizing var $(\hat{\underline{y}}_R)$ without regard to bias. Box and Draper (1959) considered both variance error and bias error by considering designs which minimize m.s.e. $(\hat{\underline{y}}_R)$, averaged over the region of experimentation. Box and Draper concluded that unless the average variance term is at least four times the average bias term, it is appropriate to minimize with respect to the bias term alone.

Karson, Manson, and Hader (1969) proposed a different approach than that of Box and Draper. They suggest that $\underline{\Theta}_1$ be estimated so as to minimize the bias term, then find the experimental design which minimizes the variance term. In most of the applications of the Karson, Manson, and Hader method, the approximating function (the one that corresponds to the reduced model) has variables described over a continuous space. Bayne, Manson, and Monroe (1974) applied this method to classical experimental design situations where there are many factors having a finite number of levels. The results are used primarily in 2^r factorial experiments.

The squared bias of $\hat{\underline{y}}_R$ may be written as a function of $\underline{\theta}_2$, where $\underline{\theta}_2$ is the vector of parameters in the full model, but not in the reduced model. This function is $A_R\underline{\theta}_2\underline{\theta}_2'A_R'$, where A_R is called the alias matrix and reflects the amount of aliasing present as a result of using the wrong model. Often the experimenter will try to minimize, over all competing designs, some objective function of m.s.e. $(\hat{\underline{y}}_R)$. A logical choice for an objective function would be one that is a function of the alias matrix, A_R.

Hedayat, Raktoe, and Federer (1974) have suggested the use of a norm which takes into account all entries of A_R and their magnitudes. In particular, the norm $||A_R||=\sqrt{\sum_i \sum_j a_{ij}{}^2)}=\sqrt{(\text{tr } A_R'A_R)}$ has this feature and some other desirable properties that most commonly used norms do not have. Because of this, we will use this norm in considering the effect of our interchange algorithm when the model assumed is inadequate. The following definition and theorem will be useful in determining what effect our interchange algorithm has on the value of the norm described above.

<u>Definition 8.1</u> (Stivastava, Raktoe, and Pesotan, (1971))

Let the k_i levels of factor F_i be identified as $\{0, 1, 2, \ldots, k_i-1\}$. Denote an element of the parameter vector $\underline{\theta}$ by $A_1^{x_1}A_2^{x_2}\ldots A_i^{x_i}\ldots A_t^{x_t}$, where x_i is an element of $\{0, 1, 2, \ldots, k_i-1\}$ and t is the number of factors in the design. Then the parameter vector $\underline{\theta}_1$ is said to be <u>admissible</u> if and only if whenever $A_1^{x_1}A_2^{x_2}\ldots A_i^{x_i}\ldots A_t^{x_t}$ belongs to $\underline{\theta}_1$, and $x_i\neq 0$, $(1\leq i\leq t)$, then $A_1^{x_1}A_2^{x_2}\ldots A_i^{z}\ldots A_t^{x_t}$ also belongs to $\underline{\theta}_1$ for all $z\neq 0$.

For our main effect case in Chapter 6, this would mean that if any level of a particular factor is in the design, then all levels of the factor must be in the design. The parameter vector for our model would be

$$\underline{\Theta}_1 = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{n_\alpha} \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{n_\beta} \end{pmatrix} \qquad \text{while } \underline{\Theta}_2 = \begin{pmatrix} (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ \cdot \\ \cdot \\ \cdot \\ (\alpha\beta)_{1n_\beta} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ \cdot \\ \cdot \\ \cdot \\ (\alpha\beta)_{n_\alpha n_\beta} \end{pmatrix}$$

Certainly our $\underline{\Theta}_1$ is admissible.

Let ξ be the set of permutations of the form $\xi = (\xi_1, \xi_2, \ldots, \xi_i, \ldots, \xi_t)$, where ξ_i is a permutation acting on the levels of the i th factor. Let Γ be an arbitrary fraction of some complete design and let $\xi_j(\Gamma)$ be the permuted fraction obtained by the action of ξ_j on Γ. Our interchange algorithm of Chapters 6 and 7 will be of the form of ξ_i.

<u>Theorem 8.1</u> (Hedayat, Raktoe, and Federer, (1974))

The amount of aliasing $||A_\Gamma||$ is invariant under ξ if $\underline{\Theta}_1$ is admissible. That is

$$||A_{\xi(\Gamma)}|| = ||A_\Gamma||$$

Theorem 8.1 tells us that the amount of aliasing (based on the norm as discussed by Hedayat, Raktoe, and Federer) is invariant to our interchange algorithm for the two-factor situation of Chapter 6.

In Chapter 7, we assumed that there were three factors in the model, but no interactions were present. Thus our parameter vector for the reduced model was

$$\underline{\theta}_1 = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{n_a} \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{n_b} \\ \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{n_c} \end{pmatrix}$$

Since $\underline{\theta}_1$ is admissible for the situation of Chapter 7, we can apply Theorem 8.1 just as we did for the Chapter 6 situation. Thus we have obtained the following result:

Theorem 8.2

Under the models of Chapters 6 and 7, the amount of aliasing $||A_R||$ is invariant under the algorithm used to pg-connect the design.

CHAPTER 9

DESIGNS THAT ARE PSEUDO-GLOBALLY CONNECTED FOR LEVEL COMBINATIONS

9.1 Introduction

In this chapter we shall combine two of the concepts already discussed: the type II connectedness first discussed by Sennetti (1972), and the pg-connectedness of Eccleston and Hedayat (1974).

Many times first order interactions are of as much importance to the experimenter as are main effects. We want to be able to estimate type II contrasts as well as we can. In addition, it seems quite reasonable that we would desire that no replicate of any particular level combination be wasted in the estimation of a given contrast involving the combination. That is, we would like to pg-connect our design for level combinations.

9.2 Type II Pseudo-globally Connected Designs

We shall use the following definition in our discussion of designs that are pg-connected for level combinations:

Definition 9.1

A type II pg-connected design is any design such that

(1) All type II contrasts are estimable and

(2) Every replicate of any two factor level combination (for two factors that interact) can be used at least once in the estimation of any

type II contrast containing the combination in question.

The concept of pg-connectedness is based on the chain definition of connectedness as it was originally given by Bose (1947). Recall that a design is pg-connected for factor F, if each replicate of any level of F is connected by a chain to at least one replicate of every other level of F. If we have at least one factor that interacts with two or more other factors, then this chain concept is not applicable in discussing type II connectedness. We would have to define connectedness in terms of the estimability of contrasts, as has been done by Sennetti (1972) and others. Of our models A, B, and C of Chapter 3, only model A is such that we can use the chain approach to connectedness. Thus we will consider only model A, when we investigate how type II connectedness and pg-connectedness may be related.

Consider some design, D, under model A,

$$E(y_{ijk}) = \gamma_k + (\alpha\beta)_{ij}.$$

Here γ_k denotes the effect of level k of factor C and $(\alpha\beta)_{ij}$ denotes the effect of the combination of the j th level of factor B and the i th level of factor A. We have pointed out earlier in this work that a design is type II connected for (A x B) if a chain exists between all pairs of level combinations of A and B. That is, if the design is connected (1-connected) for (A x B) level combinations, then the design is type II connected for (A x B).

Under model A, denote the r th replicate of level combination $a_i b_j$ by $(a_i b_j)_r$, and let $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}]-[(\alpha\beta)_{i'j}-(\alpha\beta)_{i'j'}]$ for $i \neq i'$

and $j \neq j'$ be any type II contrast in (A x B). Denote the number of replicates of level combination $a_i b_j$ by n_{ij}. Then for this three-factor model, with only one two-factor interaction, we can state the following theorem.

Theorem 9.1

Any model A design that is pg-connected for (A x B) level combinations is type II connected and has the following property: Any replicate $(a_i b_j)_r$ of a level combination $a_i b_j$ is used η times in the estimation of the contrast $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}]-[(\alpha\beta)_{i'j}-(\alpha\beta)_{i'j'}]$, where $\eta \geq \max[(n_{ij'}+n_{i'j}-1), (n_{ij'}+n_{i'j'}-1), (n_{i'j}+n_{i'j'}-1)]$.

Proof:

Since the design is pg-connected for (A x B), we know that it is 1-connected for (A x B). This insures that the design is type II connected for (A x B).

Since the design is pg-connected for (A x B), we know that $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}]$ can be estimated by an expression involving $(a_i b_j)_r$ and $(a_i b_{j'})_m$, since each replicate of $a_i b_j$ is connected by a chain to at least one replicate of $a_i b_{j'}$. We denote by $(a_i b_{j'})_m$, the replicate of $a_i b_{j'}$ to which $a_i b_j$ is connected. Denote the expression that can be used to estimate the above two term contrast, by $[(a_i b_j)_r - . . . - (a_i b_{j'})_m]$, where . . . replaces all other elements of the chain between $(a_i b_j)_r$ and $(a_i b_{j'})_m$. Similarly, $[(\alpha\beta)_{i'j}-(\alpha\beta)_{i'j'}]$ can be estimated by expressions of the form $[(a_{i'} b_j)_1 - . . . - (a_{i'} b_{j'})_{t_1}]$, or $[(a_{i'} b_j)_2 - . . . - (a_{i'} b_{j'})_{t_2}]$, or . . ., or $[(a_{i'} b_j)_{n_{i'j}} - . . . - (a_{i'} b_{j'})_{t_{n_{i'j}}}]$, where $t_1, t_2, . . ., t_{n_{i'j}}$

are all contained in the set $\{1, 2, 3, \ldots, n_{i'j'}\}$. Thus any one of the following $n_{i'j}$ factor level combination expressions can be used to estimate the contrast in question:

Set 1

$$[(a_ib_j)_r - \ldots - (a_ib_{j'})_m] - [(a_{i'}b_j)_1 - \ldots - (a_{i'}b_{j'})_{t_1}]$$

$$[(a_ib_j)_r - \ldots - (a_ib_{j'})_m] - [(a_{i'}b_j)_2 - \ldots - (a_{i'}b_{j'})_{t_2}]$$

$$\vdots$$

$$[(a_ib_j)_r - \ldots - (a_ib_{j'})_m] - [(a_{i'}b_j)_{n_{i'j}} - \ldots - (a_{i'}b_{j'})_{t_{n_{ij}}}]$$

We will now verify that these expressions are unique.

In this discussion, the words "the chain" signifies the shortest chain. That is, $[(a_{i'}b_j)_\ell - \ldots - (a_{i'}b_{j'})_{n_\ell}]$ represents the chain between $(a_{i'}b_j)_\ell$ and $(a_{i'}b_{j'})_{n_\ell}$ that has the least possible number of (A x B) combinations involved for the given design. If two expressions in Set 1 were not different, it would mean that two of the chains involving $(a_{i'}b_j)$ and $(a_{i'}b_{j'})$ level combinations were identical. That is,

$$[(a_{i'}b_j)_d - \ldots - (a_{i'}b_{j'})_{n_d}] = [(a_{i'}b_j)_e - \ldots - (a_{i'}b_{j'})_{n_e}]$$

for some d and e. This means that $[(a_{i'}b_j)_e - \ldots - (a_{i'}b_{j'})_{n_e}]$ must be of the form

$$[(a_{i'}b_j)_e - \ldots \pm (a_{i'}b_j)_d \ldots \pm (a_{i'}b_{j'})_{n_d} \ldots - (a_{i'}b_{j'})_{n_e}].$$

But this means that a shorter chain exists between $(a_i,b_j)_d$ and $(a_{i'},b_{j'})_{n_d}$ than the one originally indicated. This contradicts our hypothesis of always taking the shortest possible chain in each expression. Thus Set 1 has $n_{i'j}$ unique expressions. The same argument will apply when we discuss Sets 2, 3, and 4.

We can also estimate $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}]-[(\alpha\beta)_{i'j}-(\alpha\beta)_{i'j'}]$ by any one of the following $(n_{ij'}-1)$ expressions:

<u>Set 2</u>

$$[(a_ib_j)_r-\cdot\ \cdot\ \cdot-(a_ib_{j'})_1]-[a_{i'}b_j)_k-\cdot\ \cdot\ \cdot-(a_{i'}b_{j'})_{p_1}]$$

$$[(a_ib_j)_r-\cdot\ \cdot\ \cdot-(a_ib_{j'})_2]-[(a_{i'}b_j)_k-\cdot\ \cdot\ \cdot-(a_{i'}b_{j'})_{p_2}]$$

.

.

.

$$[(a_ib_j)_r-\cdot\ \cdot\ \cdot-(a_ib_{j'})_{m-1}]-[(a_{i'}b_j)_k-\cdot\ \cdot\ \cdot-(a_{i'}b_{j'})_{p_{m-1}}]$$

$$[(a_ib_j)_r-\cdot\ \cdot\ \cdot-(a_ib_{j'})_{m+1}]-[(a_{i'}b_j)_k-\cdot\ \cdot\ \cdot-(a_{i'}b_{j'})_{p_{m+1}}]$$

.

.

.

$$[(a_ib_j)_r-\cdot\ \cdot\ \cdot-(a_ib_{j'})_{n_{ij'}}]-[(a_{i'}b_j)_k-\cdot\ \cdot\ \cdot-(a_{i'}b_{j'})_{p_{n_{ij'}}}]$$

for some k, and $p_1, p_2, \ldots, p_{n_{ij'}}$ all in the set $\{1, 2, \ldots, n_{i'j'}\}$.

Using Sets 1 and 2 together, we see that $(a_ib_j)_r$ can be used $(n_{i'j}+n_{ij'}-1)$

times in the estimation of $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}]-[(\alpha\beta)_{i'j}-(\alpha\beta)_{i'j'}]$, since Sets 1 and 2 are disjoint.

The following set is also disjoint from Set 2:

Set 3

$$[(a_ib_j)_r-.\ .\ .-(a_ib_{j'})_m]-[(a_{i'}b_j)_{q_1}-.\ .\ .-(a_{i'}b_{j'})_1]$$

$$[(a_ib_j)_r-.\ .\ .-(a_ib_{j'})_m]-[(a_{i'}b_j)_{q_2}-.\ .\ .-(a_{i'}b_{j'})_2]$$

$$\vdots$$

$$[(a_ib_j)_r-.\ .\ .-(a_ib_{j'})_m]-[(a_{i'}b_j)_{q_{n_{i'j'}}}-.\ .\ .-(a_{i'}b_{j'})_{n_{i'j'}}]$$

where $q_1, q_2, . . ., q_{n_{i'j'}}$ are all in the set $\{1, 2, . . ., n_{i'j}\}$.

Using Sets 2 and 3 together, we see that $(a_ib_j)_r$ can be used $(n_{ij'}+n_{i'j'}-1)$ times in the estimation of the contrast.

The expressions in Set 4 may also be used to estimate the contrast.

Set 4

$$[(a_ib_j)_r-.\ .\ .-(a_ib_{j'})_{s_1}]-[(a_{i'}b_j)_k-.\ .\ .-(a_{i'}b_{j'})_1]$$

$$[(a_ib_j)_r-.\ .\ .-(a_ib_{j'})_{s_2}]-[(a_{i'}b_j)_k-.\ .\ .-(a_{i'}b_{j'})_2]$$

$$\vdots$$

$$[(a_ib_j)_r-.\ .\ .-(a_ib_{j'})_{s_{n_{i'j'}}}]-[(a_{i'}b_j)_k-.\ .\ .-(a_{i'}b_{j'})_{n_{i'j'}}]$$

Let Set 1* denote the set of $(n_{i'j}-1)$ expressions obtained from Set 1 after

$$[(a_i b_j)_r - \cdot\ \cdot\ \cdot - (a_i b_{j'})_m] - [a_{i'} b_j)_k - \cdot\ \cdot\ \cdot - (a_{i'} b_{j'})_{t_k}]$$

is omitted. Then Sets 1* and 4 will be disjoint. Using Sets 1* and 4, we see that $(a_i b_j)_r$ can be used $(n_{i'j}+n_{i'j'}-1)$ times in the estimation of the contrast.

In general, we have no way of knowing how many common expressions there are in any two of the three sets {(Set 1)U(Set 2)}, {(Set 2)U(Set 3)}, and {(Set 1*)U(Set 4)}. We do know, though, that there is no duplication of expressions within any one of these three sets. Thus we can conclude that there are at least $\max(n_{ij'}+n_{i'j}-1,\ n_{ij'}+n_{i'j'}-1,\ n_{i'j}+_{i'j'}-1)$ different expressions that contain $(a_i b_j)_r$ which can be used to estimate the contrast $[(\alpha\beta)_{ij}-(\alpha\beta)_{ij'}]-[(\alpha\beta)_{i'j}-(\alpha\beta)_{i'j'}]$.

Example 9.1

Consider an experiment of three factors A, B, and C, where the only suspected interaction is (A x B). Suppose $n_a=2$, $n_b=3$, and $n_c=7$. The following design is pg-connected for (A x B) level combinations:

$a_1b_1c_1$	$a_1b_1c_4$	$a_2b_2c_3$
$a_1b_1c_2$	$a_1b_2c_1$	$a_1b_2c_4$
$a_1b_1c_3$	$a_1b_2c_2$	$a_1b_3c_1$
$a_1b_3c_2$	$a_2b_1c_6$	$a_2b_2c_7$
$a_1b_3c_3$	$a_2b_1c_7$	$a_2b_3c_5$
$a_1b_3c_5$	$a_1b_2c_5$	$a_2b_3c_6$
$a_2b_1c_4$	$a_2b_2c_6$	$a_2b_3c_7$

Suppose we feel that certain observations are taken under conditions that produce less experimental error, on the average, than some other observations. How good an estimate we can get for a particular contrast would depend not only on how many estimators we can find for the contrast, but also on how often certain design points are used in the estimation of the contrast. For instance, we might be interested in how many times $a_1b_2c_1$ can be used in the estimation of the contrast

$[(\alpha\beta)_{13}-(\alpha\beta)_{12}]-[(\alpha\beta)_{23}-(\alpha\beta)_{22}]$. Now $\max(n_{13}+n_{23}-1,\ n_{13}+n_{22}-1,\ n_{23}+n_{22}-1) = n_{13}+n_{23}-1 = n_{13}+n_{22}-1 = 6$. One set of six level combination expressions suggested by the proof of Theorem 9.1 would be

$[a_1b_3c_1-\ldots-a_1b_2c_1]-[a_2b_3c_5-\ldots-a_2b_2c_3]$
$[a_1b_3c_1-\ldots-a_1b_2c_1]-[a_2b_3c_6-\ldots-a_2b_2c_6]$
$[a_1b_3c_1-\ldots-a_1b_2c_1]-[a_2b_3c_7-\ldots-a_2b_2c_7]$

$[a_1b_3c_2-\ldots-a_1b_2c_1]-[a_2b_3c_5-\ldots-a_2b_2c_3]$
$[a_1b_3c_3-\ldots-a_1b_2c_1]-[a_2b_3c_7-\ldots-a_2b_2c_3]$
$[a_1b_3c_5-\ldots-a_1b_2c_1]-[a_2b_3c_5-\ldots-a_2b_2c_3]$

To show that these can be used to estimate the contrast, we will write the entire expressions.

$[a_1b_3c_1-a_1b_2c_1]-[a_2b_3c_5-a_1b_3c_5+a_1b_3c_3-a_2b_2c_3]$
$[a_1b_3c_1-a_1b_2c_1]-[a_2b_3c_6-a_2b_2c_6]$
$[a_1b_3c_1-a_1b_2c_1]-[a_2b_3c_7-a_2b_2c_7]$

$[a_1b_3c_2-a_1b_1c_2+a_1b_1c_1-a_1b_2c_1]-[a_2b_3c_5-a_1b_3c_5+a_1b_3c_3-a_2b_2c_3]$
$[a_1b_3c_3-a_1b_1c_3+a_1b_1c_1-a_1b_2c_1]-[a_2b_3c_7-a_2b_2c_7+a_2b_2c_3-a_2b_2c_3]$
$[a_1b_3c_5-a_1b_2c_5+a_1b_2c_1-a_1b_2c_1]-[a_2b_3c_5-a_1b_3c_5+a_1b_3c_3-a_2b_2c_3]$

Theorem 9.1 tells us that any design, under model A, that is pg-connected for level combinations, will have a property much stronger than type II pg-connectedness. Rather than being assured that we can use each level combination replicate at least once in the estimation of a corresponding contrast, we know that we can use each replicate at least max $(n_{ij'}+n_{i'j}-1,\ n_{ij'}+n_{i'j'}-1,\ n_{i'j}+n_{i'j'}-1)$ times. This suggests choosing a design that is pg-connected for level combinations over one that is 1-connected for level combinations, if the designs are equally good according to all other criteria. In addition, if we have a design that is 1-connected for level combinations, we would, if possible, use the algorithm of Theorem 6.2 to pg-connect the design for (A x B) combinations. The new design would allow an increased number of estimates for type II contrasts in (A x B).

CHAPTER 10

OPTIMAL MINIMAL AUGMENTED MULTIDIMENSIONAL DESIGNS

10.1 Introduction

In Chapter 4, we discussed the need for a method to choose the additional design points when generating an MAMD. That is, if any MAMD D* is such that D* = D + T*, where D is the original design, then how shall we choose T*? What kind of criterion should we use?

In Chapter 5, we discussed the justification for using S-optimality in comparing designs. This criterion allows us to choose the design from all eligible designs that is closest to A, D, and E optimal. This criterion is equally applicable here. If we can, we will choose the T* such that D* is the S-optimal MAMD over all possible D*. In Chapter 6 we were able to obtain an S-better design by using our interchange algorithm. This approach is generally not applicable here for the following two reasons:

(1) Usually MAMD's do not have sufficient numbers of replications of levels to satisfy the necessary requirements for the use of our algorithm and

(2) Most of the time an MAMD is chosen because we do not want to waste observations already taken. This means that part of our MAMD is unalterable. This lack of flexibility in the MAMD usually means that we cannot apply our algorithm of Chapter 6 to pg-connect an

MAMD. For these two reasons, our approach will not be to form some arbitrary MAMD, D*, and then try to improve it, but to describe a procedure to choose the assemblies in T*.

MAMD's possess two desirable characteristics: All contrasts are estimable and only a minimum number of additional observations is used. If an MAMD is desired for these reasons, then it is likely that quite often these were desirable properties in the original design. That is, usually the design to be augmented is an MMD for some model. The inadequacy of the original design stems from the original assumption of what we now believe to be an inadequate model. For instance, suppose the experimenter desires to estimate main effect contrasts in a 3^3 factorial experiment under the assumption of no interaction. The following design is an MMD that may be used for this purpose:

$a_1b_1c_1$
$a_2b_2c_2$
$a_3b_3c_3$
$a_1b_2c_1$
$a_2b_3c_2$
$a_1b_1c_2$
$a_2b_2c_3$

However, if later, the experimenter decides that there may be an (A x B) interaction present, then this design must be augmented to allow for the estimation of (A x B) type II contrasts. We need to choose a set of additional assemblies, T*, such that D* = D + T* will somehow be S-optimal for (A x B).

Suppose we want to augment a three-factor, main effect MMD for one or more two-factor interactions. We want to accomplish this augmenting in a manner that will insure that the resultant design is as close as possible to being S-optimal for the interaction factor considered most important.

If we desire an MAMD that is S-optimal for factor combination (A x B), then (A x B) will be the last "factor" for which we augment the design. We will first connect the design for (A x C) or (B x C) or both, then consider the best way to add the assemblies that connect the design for (A x B). To connect the design for, say (A x C), we add assemblies to insure that all possible (A x C) level combinations occur with a given level of B in the design. We choose the level of B that will allow us to accomplish this with the least number of assemblies. Then, if we desire, we add the least possible number of assemblies that will insure that all levels of (B x C) occur with a given level of A. This is not the only way that we might connect the design for these two factors. However, by using this minimum number of assemblies in the first two stages of the augmentation procedure, we insure ourselves maximum flexibility in the third, most important stage: connecting the design for the most important factor, (A x B). After having connected the design for all factors except the one for which we desire the S-optimal design, we must consider how the final set of assemblies can be chosen.

Recall from Chapter 4, that we can connect the design for the last factor (A x B), simply by adding assemblies that will insure that every level combination of (A x B) occurs in the design. Any level of the remaining factor C may be used in the assemblies. We have no flexibility, at this point, in choosing these combinations to be added, but we have complete flexibility in deciding what levels of the remaining factor, C, are used in the assemblies added.

10.2 Sequentially S-optimal Designs

The procedure we suggest for this final stage of the augmentation procedure will be a sequential one; we shall add one assembly at the time to the design, forming the design that is S-optimal for (A x B) at each step.

Let D be a minimal multidimensional design of size n_D, for some model M_D, and Ω be the class of all designs, D*, of size n_{D*}, with the following properties:

(1) D* is obtained by augmenting D,

(2) D* is a minimal augmented multidimensional design under some model M_{D*}, and

(3) D* is obtained by adding assemblies, one at a time to D, obtaining a design of size n_D+i, for $i=1, 2, \ldots, n_{D*}-n_D$, at step i of the augmentation procedure.

(4) The minimum number of observations is used to connect D* for all effects other than (A x B) prior to connecting the design for (A x B). Then

Definition 10.1

The sequentially S-optimal design is the one from Ω, such that, given the design at step i, the design at step i+1 is S-optimal for all i in the final stage of the augmentation procedure.

Suppose we want to obtain the sequentially S-optimal design, D*, for (A x B) level combinations by adding $n_{D*}-n_D$ assemblies to some design, D. If the first assembly to be added contains an (A x B) combination which we denote by $(ab)_p$, then how shall we determine the level of C, c_p, such that the assembly $(ab)_p c_p$ will give us the S-optimal design at the first step in the augmentation procedure? We will discuss an optimal way for determining what c_p should be.

Let C_{AB} be the coefficient matrix for (A x B) level combinations under the model M_{D*}, but for the original design, D. We want to consider C_{AB} changes as an assembly is added to D at any step in the augmentation procedure. Let k_p denote the number of times that c_p occurs in the experiment prior to adding assembly $(ab)_p c_p$. The notation used here will generally be the same as in Chapter 6.

The following elements of C_{AB} are changed when we add assembly $(ab)_p c_p$ to design D:

$$c_{pp} = 0 \longrightarrow 1 - \frac{1}{k_p+1}$$

$$c_{pi} = 0 \longrightarrow - \frac{1}{k_p+1}$$ for $p \neq i$ and $(ab)_i c_p$ in design D. There are k_p of these terms.

$$c_{jp} = 0 \longrightarrow - \frac{1}{k_p+1}$$ for $p \neq j$ and $(ab)_j c_p$ in design D. There are k_p of these terms.

$$c_{ij} \longrightarrow c_{ij} - \frac{1}{k_p+1} + \frac{1}{k_p} = c_{ij} + \frac{1}{k_p(k_p+1)}$$ for i, j defined as above. There are $k_p^{\,2}$ of these terms.

All other terms of C_{AB} remain the same.

We know that the S-optimal design will be the one for which tr $C_{AB}{}^2$ is minimum. That is, if any element of the coefficient matrix is denoted by c_{uv}, then we want to minimize tr $C_{AB}{}^2 = \sum_u \sum_v (c_{uv})^2$. After adding $(ab)_p c_p$ to the design,

$$\text{tr } C_{AB*}{}^2 = (1 - \frac{1}{k_p+1})^2 + k_p(-\frac{1}{k_p+1})^2 + k_p(-\frac{1}{k_p+1})^2 +$$

$$+ \sum_i \sum_j \{c_{ij} + \frac{1}{k_p(k_p+1)}\}^2 + U$$

where U corresponds to terms of C_{AB} that are unchanged when the assembly is added. The previous expression can be rewritten as

$$\text{tr } C_{AB*}{}^2 = (\frac{k_p}{k_p+1})^2 + 2k_p(\frac{1}{k_p+1})^2 + \sum_i \sum_j c_{ij}{}^2$$

$$+ \frac{2}{k_p(1_p+1)} \sum_i \sum_j c_{ij} + \sum_i \sum_j \frac{1}{k_p{}^2(k_p+1)^2} + U$$

$$= \frac{k_p{}^2+2k_p}{(k_p+1)^2} + \frac{2}{k_p(k_p+1)} \sum_i \sum_j c_{ij} + \frac{k_p{}^2}{k_p{}^2(k_p+1)^2}$$

$$+ U'$$

where $U' = U + \sum_i \sum_j c_{ij}{}^2 = \text{tr } C_{AB}{}^2$ before augmentation with $(ab)_p c_p$.

Thus to minimize tr $C_{AB*}{}^2$ after adding $(ab)_p c_p$ to the design, we minimize

$$d = \frac{k_p+2k_p}{(k_p+1)^2} + \frac{2}{k_p(k_p+1)} \sum_i \sum_j c_{ij} + \frac{k_p{}^2}{k_p{}^2(k_p+1)^2}$$

$$= \quad 1 \quad + \frac{2}{k_p(k_p+1)} \sum_i \sum_j c_{ij}$$

Therefore, for the S-optimal design at this step, we choose c_p such that

$$d = 1 + \frac{2}{k_p(k_p+1)} \sum_i \sum_j c_{ij}$$

is minimum over all possible values of p. Certainly i, j, and k are functions of the p chosen. Based on the above discussion, we see that we can obtain the sequentially S-optimal design by the following procedure:

At each step of the final augmentation procedure, determine the set of (A x B) level combinations, $S_1, S_2, \ldots, S_r, \ldots, S_{n_c}$, where all elements of S_r occur in the design with c_r. Then

(1) Compute $d_r = 1 + \frac{2}{k_r(k_r+1)} \sum_i \sum_j c_{ij}$, where the c_{ij} correspond to the (A x B) combinations in S_r.

(2) Choose c_r such that d_r is minimum over all possible d. This will give the S-optimal design at that step in the augmentation procedure.

Example 10.1

For a 4 x 3 x 2 factorial situation, consider the following main effect MMD:

$a_1b_1c_1$

$a_2b_2c_2$

$a_3b_3c_1$

$a_4b_1c_2$

$a_1b_2c_1$

$a_2b_3c_2$

$a_1b_1c_2$

After the experiment has been run using this design, information becomes available that makes it appear likely that (A x B) and (B x C) interactions should be assumed present in the design, the major interest being in estimation of (A x B) type II contrasts. We wish to obtain the MAMD that is sequentially S-optimal for (A x B). First we will connect the design for (B x C). If we use level 1 of A, we can do this with only three additional assemblies. These assemblies are

$a_1b_2c_2$

$a_1b_3c_1$

$a_1b_3c_2$

Now for the most important stage: connecting the design for (A x B). We still need to add the last five (A x B) level combinations to the design. These are

a_2b_1

a_3b_1

a_3b_2

a_4b_2

a_4b_3

Suppose we decide to add a_2b_1 first, and want to know which level of C to use. The sets S_1, S_2 are as follows:

$$S_1 = (a_1b_1,\ a_1b_2,\ a_1b_3,\ a_3b_3)$$

$$S_2 = (a_1b_1,\ a_1b_2,\ a_1b_3,\ a_2b_3,\ a_4b_1,\ a_2b_2)$$

If we call the four elements of S_1 combinations 1 through 4, and the elements of S_2 combinations 1 through 3 and 5 through 7, then the corresponding d values for the first combination to be added are

$$d_1^{(1)} = 1 + \frac{2}{4(5)}\{3(2-\frac{1}{4}-\frac{1}{6})+1(1-\frac{1}{4})+6(-\frac{1}{4}-\frac{1}{6})+6(-\frac{1}{4})\} = 1.1500$$

$$d_2^{(1)} = 1 + \frac{2}{6(7)}\{3(2-\frac{1}{4}-\frac{1}{6})+3(1-\frac{1}{6})+6(-\frac{1}{4}-\frac{1}{6})+24(-\frac{1}{6})\} = 1.0357$$

since $d_2^{(1)} < d_1^{(1)}$, we choose level 2 of factor C to use with the first (A x B) combination to be added. This gives the new assembly

$a_2b_1c_2$

If we decide to add a_3b_1 next, we will look at the new coefficient matrix for (A x B) (the one after $a_2b_1c_2$ has been added) and find the values of d_1 and d_2 for the combination a_3b_1.

$$d_1^{(2)} = 1 + \frac{2}{4(5)}\{3(2-\frac{1}{4}-\frac{1}{7})+1(1-\frac{1}{4})+6(-\frac{1}{4}-\frac{1}{7})+6(-\frac{1}{4})\} = 1.1714$$

$$d_2^{(2)} = 1 + \frac{2}{7(8)}\{3(2-\frac{1}{4}-\frac{1}{7})+4(1-\frac{1}{7})+6(-\frac{1}{4}-\frac{1}{7})+36(-\frac{1}{7})\} = 1.0268$$

Since $d_2^{(2)} < d_1^{(2)}$, we choose level 2 of factor C again, and we add the appropriate assembly $a_3b_1c_2$.

For adding a_3b_2, we find that

$$d_1^{(3)} = 1.1875$$

$$d_2^{(3)} = 1.0208$$

and we use assembly $a_3b_2c_2$.

The fourth assembly to be added will be $a_4b_2c_2$, since

$d_1^{(4)} = 1.2000$

$d_2^{(4)} = 1.0167$

For the final combination, a_4b_3, we use level 2 of factor C again, since

$d_1^{(5)} = 1.2000$

$d_2^{(5)} = 1.0136$

Our final design, which will be sequentially S-optimal for (A x B), will be

$a_1b_1c_1$	$a_1b_2c_2$	$a_2b_1c_2$
$a_2b_2c_2$	$a_1b_3c_1$	$a_3b_1c_2$
$a_3b_3c_1$	$a_1b_3c_2$	$a_3b_2c_2$
$a_4b_1c_2$		$a_4b_2c_2$
$a_1b_2c_1$		$a_4b_3c_2$
$a_2b_3c_2$		
$a_1b_1c_2$		

It is interesting to note that we used the same level of factor C in all five of the assemblies that we added in the final stage of the augmentation procedure. We shall now show that this is always the case.

Suppose that for the j th step in the final augmentation procedure,

$$d_r^{(j)} \leq d_m^{(j)}, \tag{10.1}$$

i.e. c_r is chosen for the assembly to be added to the design. Let k_r be the number of replicates of level r of factor C prior to adding the assembly at step j. Consider $d_m^{(j)}$ and $d_m^{(j+1)}$, where $d_m^{(j+1)}$ is the quantity corresponding to level m of factor C that we will consider when deciding whether or not use level m at step (j+1) of the augmentation procedure. The only terms of $d_m^{(j)}$ that may change when we add an assembly with level r of factor C at step j, will be terms in $\sum_i \sum_j c_{ij}$. Any term that does change will become larger, since the term included $-\frac{1}{k_r}$ before the augmentation at step j, but includes $-\frac{1}{k_r+1}$ after the augmentation at step j. Thus

$$d_m^{(j)} \leq d_m^{(j+1)} \qquad (10.2)$$

Now consider $d_r^{(j)}$ and $d_r^{(j+1)}$. All terms of $d_r^{(j)}$ in $\sum_i \sum_j c_{ij}$ will be increased by $\frac{1}{k_r(k_r+1)}$ when an assembly with level r of factor C is added at step j. This is because each of these terms included $-\frac{1}{k_r}$ before the assembly is added, but $-\frac{1}{k_r+1}$ after the assembly is added. There are k_r of these terms. In addition, there is a term of the form $(1-\frac{1}{k_r+1})$ and $2k_r$ terms of the form $(-\frac{1}{k_r+1})$ added to $\sum_i \sum_j c_{ij}$. These last $2k_r+1$ values correspond to values in the coefficient matrix that were zero prior to adding the assembly with c_r. Thus the total change in $\sum_i \sum_j c_{ij}$ at step j may be given by

$$[k_r \left(\frac{1}{k_r(k_r+1)}\right) + \left(1-\frac{1}{k_r+1}\right) + 2k_r\left(-\frac{1}{k_r+1}\right)],$$

which is seen to be equal to zero. That is, the sum $\sum_i \sum_j c_{ij}$ for $d_r^{(j)}$ does not change when an assembly containing level r of factor C is added at step j. Thus the only change in $d_r^{(j)}$ is in the coefficient of $\sum_i \sum_j c_{ij}$, which changes from $\frac{1}{k_r(k_r+1)}$ to $\frac{1}{(k_r+1)(k_r+2)}$. We now see that

$$d_r^{(j+1)} < d_r^{(j)} \tag{10.3}$$

Hence from (10.1), (10.2), and (10.3), we see that

$$d_r^{(j+1)} < d_m^{(j+1)}$$

If level r of factor C is used in the assembly added at step j of the augmentation procedure, then level r will also be used at step (j+1). We always use the same level of factor C, no matter how many assemblies we have to add in the final stage of the augmentation procedure. We have now proved the following theorem:

<u>Theorem 10.1</u>

Let $S_1, S_2, \ldots, S_t, \ldots, S_{n_c}$ be the set of (A x B) level combinations, where all elements of S_t occur in design D with level c_t of factor C. For all possible values of t, let

$$d_t = 1 + \frac{1}{k_t(k_t+1)} \sum_i \sum_j c_{ij}$$

where the c_{ij} are terms of the coefficient matrix for (A x B) for model

D that correspond to the elements of S_t. If c_r is such that d_r is a minimum over all value of d_t, then level c_r of factor C is used with all (A x B) combinations to be added, and the design obtained will be the sequentially S-optimal design.

It is interesting to note that the procedure followed in the final stage of the augmentation is independent of the order in which we add the combinations in the final stage. This is true, because the d-values are not functions of the (A x B) combinations being added or the combinations already added in this final stage of the procedure.

The procedures and theorem that we have given are also applicable for situations where we have more than three factors. We can still find the sequentially S-optimal design for (A x B), by treating the level combinations of all other factors as the levels of the factor we call C in the discussion of the procedure. For this case, Theorem 10.1 tells us that we would use the same level combination of all factors except A and B, in the final stage of the augmentation procedure.

The question may be raised as to whether this procedure is "significantly" better than just choosing the S-optimal MAMD from all possible MAMD's. The procedure has two advantages: There is something of a computational advantage in that we do not have to consider elements of C^2, but only of C. The major advantage though, is in the number of C matrices we have to consider. For example, for the situation of Example 10.1, if we attempted to get the S-optimal design, we would have to compute 32 different C matrices, one for each possible permutation of the levels of factor C used in the last five assemblies. To obtain the sequentially S-optimal design, we only had to compute one coefficient matrix.

The last thing to be considered in this chapter is the relationship between the sequentially S-optimal design and the S-optimal design. Are they the same design? If not, how much better is the S-optimal design?

It can be seen that

$$d_r^{(j+1)} = \frac{2}{k_r+2} + \frac{k_r}{k_r+2}\, d_r^{(j)},$$

where level r of factor C is used at step j in the augmentation procedure. Thus if $d_r^{(j)}$, the increase in tr $C_{AB}{}^2$ at step j, is small, then we would expect the increase at step (j+1) to be relatively small. This leads us to believe that if we are successful at step j (with respect to S-optimality), we will be successful at all succeeding steps. This means that the final design will be very good with respect to S-optimality. Treated from a different point of view, it can be said that, $d_r^{(j+1)}$ is a function of tr $C_{AB}{}^2$ after adding the assembly at step j. Thus if we make tr $C_{AB}{}^2$ as small as possible by forming the S-optimal design at step j, we would expect the design to be S-better at step (j+1) and at all succeeding steps. These considerations lead us to believe that the sequentially S-optimal design will either be the S-optimal design, or be nearly as S-good.

CHAPTER 11

SUMMARY

Throughout this work we have considered procedures for obtaining appropriate designs when we have some restriction on the number of observations that may be taken or on the number of replicates of any factor level that may be used.

Sennetti (1972) showed the existence of minimal multidimensional designs (MMD's) and minimal augmented multidimensional designs(MAMD's) which allow estimation of type I and type II contrasts. For an MMD, only one more design point is required than there are degrees of freedom for the parameter vector. For MAMD's, the number of assemblies added is equal to the difference between the number of degrees of freedom for the parameter vector and the rank of the design matrix.

In this work, we have suggested a practical procedure to obtain an MMD for estimating type I contrasts and have proved the procedure valid. This procedure uses the "chain" concept of a connected design as defined by Bose (1947). In addition, a procedure is discussed that may be used to obtain an MMD for estimating type II contrasts. After proof of the validity of the procedure, advantages of this procedure over some other possible procedures to obtain an MMD are given. We then show that only a slight modification of the procedure described for obtaining MMD's is necessary to obtain an MAMD.

If there is a restriction on the number of replicates of factor levels for an experiment, then we take a different approach. We want to be able to estimate all main effect linear contrasts as well as possible. If m_{ij} denotes the number of replicates of level j of factor F_i, then we wish to increase the number of estimators for type I contrasts without altering any of the m_{ij}. The interchange algorithm used by Eccleston and Hedayat (1974) to accomplish this for a proper, 1-connected randomized block design is extended to two-factor, no interaction designs. The designs obtained are pg-connected, thus guaranteeing more estimates for main effect contrasts. In addition, the new design will be better than the old with respect to the S-optimality criterion. It is shown that the procedure can also be used in a two or more factor experiment to pg-connect an 1-connected design for two factors. The new design will be better than the old with respect to the C-optimality criterion. The algorithm described is proved to have no effect on the amount of aliasing (based on a norm suggested by Hedayat, Raktoe, and Federer, (1974)) due to a possibly incorrect assumption of no interaction.

We discuss how many estimates we are assured of for a type II contrast if the design is pg-connected for level combinations of the factors in the contrast. The use of the interchange algorithm to pg-connect a design for level combinations is suggested because of the increased number of estimators for type II contrasts that may be obtained.

The last topic discussed is the use of a criterion for choosing a particular MAMD for estimating type II contrasts. We find that the sequentially S-optimal MAMD is easy to obtain and is similar to the S-optimal design.

CHAPTER 12

PROPOSED EXTENSIONS

A number of questions arise with respect to this work that suggest further investigation may be appropriate. Throughout Chapters 2, 3, and 4, we maintained property I when choosing an MMD or MAMD whenever possible. This quest for equal level frequencies has intuitive appeal and, in some cases, produces a variance balanced design. We need to prove, though, that the maintenance of property I has advantages in the more general case.

In Chapter 6, we noted that the conditions necessary for a pg-connected design to be S-better than the original 1-connected design are generally easy to satisfy. It has not been shown when, if ever, it is impossible to satisfy any one of the three conditions of Theorem 6.2.

In section 6.3, it is proved that the final design, D**, of the three design sequence, is S-better for factor A and for factor B. The question may arise as to when, if ever, D** is not pg-connected for factor A, the factor involved in the first interchange.

The procedures of Chapter 7 allow us to improve a factorial design with respect to two factors, A and B. Level combinations of all factors except A and B may be considered to be levels of some pseudo-factor, F. It should be questioned whether the procedures of Chapter 7 may be extended to situations where we want to improve a design with respect to three or more factors. It is possible that the difficulty of using such a procedure would outweigh the advantages. Certainly though, such an extension would be worth considering.

We, as Eccleston and Hedayat (1974), have used an interchange algorithm to improve a design with respect to the S-optimality criterion. What has not been shown, is the value of the amount of improvement obtained. In most of the examples given, the decrease in tr C^2 is approximately five to ten percent. It is not clear how valuable a decrease of this magnitude is. Perhaps an approach to this question would be to consider the change in the design with respect to one of the criteria used to justify the use of the S-optimality criterion. At least in terms of the A criterion, the value of the relative amount of decrease might be easier to determine.

As a final item to consider, more work needs to be done in comparing the S-optimal and the sequentially S-optimal designs. What are the similarities and differences, if any, between the design obtained using the sequentially S-optimal criterion to obtain an MAMD, instead of the S-optimality criterion in the last stage of the augmentation procedure.

BIBLIOGRAPHY

1. Addelman, S. (1963). "Techniques for Constructing Fractional Replicate Plans." Journal American Statistical Association 58, 45-71.

2. Addelman, S. and Kempthorne, O. (1961). "Orthogonal Main Effect Plans." Aeronautical Research Laboratory, Office of Aerospace Research, United States Air Force, Wright-Patterson Air Force Base, Ohio.

3. Anderson, R. (1968). On the Construction and Analysis of Multi-Factor Partially Balanced Designs. University of Nebraska, Department of Statistics, unpublished Ph.D. dissertation.

4. Bayne, K. C., Manson, A. R., and Monroe, R. J. (1974). "Protecting Main Effect Models Against Two-Factor Interaction Bias in Fractions of 2^r Factorial Design." North Carolina State University, Department of Statistics, unpublished.

5. Bose, R. C. (1947). Presidential Address. Proceedings of the Thirty-fourth Indian Science Congress.

6. Box, G. E. P., and Draper, N. R. (1959). "A basis for the Selection of a Response Surface Design." Journal American Statistical Association 54, 622-654.

7. Chakrabarti, M. C. (1963). "On the C-Matrix in Design of Experiments." Journal Indian Statistical Association 1, 1963, 8-23.

8. Daniel, C. (1962). "Sequences of Fractional Replicates in the 2^{p-g} Series." Journal American Statistical Association 57, 403-429.

9. Daniel, C. (1973). "One-At-A-Time Plans." Journal American Statistical Association 68, 353-360.

10. Eccleston, J. A. (1972). On the Theory of Connected Designs. Cornel University, unpublished Ph.D. Dissertation.

11. Eccleston, J.A. and Hedayat, A. (1974). "On the Theory of Connected Designs: Characterization and Optimality." The Annals of Statistics, Vol. 2, No. 6. Nov., 1974, 1238-1255.

12. Federer, W. T. (1961). "Augmenting Designs with One-way Elimination of Heterogeneity." Biometrics 17, 447-473.

13. Federer, W. T., Hedayat, A. and Raktoe, B. L. (1974). "Econostat, Optimal Fractional Factorial Designs: A New Approach." Institute of Mathematical Statistics Meetings, Iowa State University, April 29, 1972.

14. Hedayat, A. (1971). "A Set of Conditions Which Every Locally Connected Design Must Satisfy." BU-392-M, Cornell University, unpublished.

15. Hedayat, A., Raktoe, B. L., and Federer, W. T. (1974). "On a Measure of Aliasing Due to Fitting an Incomplete Model." Annals of Mathematical Statistics, Vol. 2, No. 4, 650-660.

16. John, P. M. W. (1966). "Augmenting 2^{n-1} Designs." Technometrics, Vol. 8, No. 3, 469-480.

17. Karson, M. J., Manson, A. R., and Hader, R. J. (1969). "Minimum Bias Estimation and Experimental Design for Response Surfaces." Technometrics, Vol. 11, 461-475.

18. Kempthorne, O. (1956). "The Efficiency Factor of an Incomplete Block Design." Annals of Mathematical Statistics, Vol. 27, 846-849.

19. Kiefer, J. (1959). "Optimum Experimental Designs." Journal Royal Statistical Society, (Series B), 21, 272-279.

20. Myers, R. H. (1971). Response Surface Methodology. Allyn and Bacon, Inc., Boston.

21. Potthoff, R. F. (1958). "Multi-Dimensional Incomplete Block Designs." Institute of Statistics Mimeograph Series No. 211, University of North Carolina, Chapel Hill.

22. Rao, V. R. (1958). "A Note on Balanced Designs." Annals Mathematical Statistics, Vol. 29, 290-294.

23. Sennetti, J. T. (1972). Minimal Multidimensional Designs. Virginia Polytechnic Institute & State University, Department of Statistics, unpublished Ph.D. dissertation.

24. Shah, B. V. (1960). "Optimality Criteria for Incomplete Block Designs." Annals Mathematical Statistics, Vol. 31, 791-794.

25. Srivastava, J. N. and Anderson, R. (1970). "Some Basic Properties of Multidimensional Partially Balanced Designs." Annals Mathematical Statistics, Vol. 41, 1438-1445.

26. Srivastava, J. M., Raktoe, B. L., and Pesotan, H. (1971). "On Invariance and Randomization in Fractional Replication." To appear.

VITA

The author was born January 14, 1942, in Victoria, Virginia. He was graduated from Victoria High School in June, 1960, and received his A.B. in mathematics from Old Dominion University, Norfolk, Virginia, in June, 1965. From September, 1965 to February, 1968, the author taught mathematics at Andrew Lewis High School in Salem, Virginia. He was an electronics communications officer in the United States Air Force between February, 1968 and April, 1969. He taught mathematics at Glenvar High School for two years while continuing work on a M.S. degree at Radford College, Radford, Virginia. He received his M.S. degree with a major in mathematics and statistics in March of 1970. From September of 1971 to July of 1974, he attended Virginia Polytechnic Institute and State University, where he was a graduate teaching assistant in the Department of Statistics. In August of 1974, the author assumed a position in the Department of Statistics of Hollins College, Roanoke, Virginia. He is a member of Pi Mu Epsilon, The American Statistical Association, and The Biometric Society.

Wilkie W. Chaffin

CONNECTEDNESS AND OPTIMALITY IN
MULTIDIMENSIONAL DESIGNS

by

Wilkie W. Chaffin

(ABSTRACT)

Sennetti (1972) showed the existence of minimal multidimensional designs (MMD's) and minimal augmented multidimensional designs (MAMD's) which allow estimation of type I and type II contrasts. For an MMD, only one more design point is required than there are degrees of freedom for the parameter vector. For MAMD's, the number of assemblies added is equal to the difference between the number of degrees of freedom for the parameter vector and the rank of the design matrix.

Using the chain concept of connectedness as defined by Bose (1947), this work suggests a practical procedure to obtain an MMD for estimating type I contrasts, and proves the procedure valid. In addition, a procedure is discussed that may be used to obtain an MMD for estimating type II contrasts. After proof of the validity of the procedure, advantages of this procedure over some other possible procedures to obtain an MMD are given. It is shown that only a slight modification of the procedure is necessary to be able to obtain an MAMD for estimating type II contrasts.

If there is a restriction on the number of replicates of factor levels for an experiment, then a different approach is suggested. If

m_{ij} denotes the number of replicates of level j of factor F_i, then it is desired to increase the number of estimators for type I contrasts without altering any of the m_{ij}. The interchange algorithm used by Eccleston and Hedayat (1974) to accomplish this for a proper, locally connected (l-connected) randomized block design is extended to two-factor, no interaction designs. The design obtained is pseudo-globally connected (pg-connected), thus guaranteeing more estimates for main effect contrasts. In addition, the new design will be better than the old with respect to the S-optimality criterion. It is shown that the procedure can also be used in a two or more factor experiment to pg-connect an l-connected design for two factors. The new design obtained will be better than the old with respect to a new criterion, C-optimality. The algorithm described is proved to have no effect on the amount of aliasing (based on a norm suggested by Hedayat, Raktoe, and Federar, (1974)) due to a possibly incorrect assumption of no interaction.

The use of the interchange algorithm to pg-connect a design for level combinations is suggested because of the increased number of estimators for type II contrasts that may be obtained. A theorem is proved which gives the minimum number of estimates that will be available for estimating a type II contrast if a design is pg-connected for level combinations.

The last topic discussed is the use of a criterion for choosing a particular MAMD for estimating type II contrasts. The sequentially S-optimal design is defined. It is shown that the sequentially S-optimal design is easy to obtain and is similar to the S-optimal design.