ON THE THEORY OF CONNECTED DESIGNS
II. OPTIMALITY

J. A. Eccleston and A. Hedayat
Cornell University

BU-427-M August, 1972

Abstract

The concepts of locally, pseudo-globally and globally connected block designs were introduced by Eccleston and Hedayat (1972). In that paper we were mainly concerned with the characterizations of these designs. In the present paper we shall expand the theory of connected designs by exploring two optimality properties of these designs. Our optimality criteria are in terms of eigen values of the corresponding information matrices. Before proceeding further let us review a general description of optimum designs.

The three most used and well known optimality criteria, A, D and E optimality, are defined in section 2. Rather than use one of the popular criteria we use one suggested by Shah which we call S optimality. The criterion is: minimize the trace of the information matrix squares, for those designs with identical trace of the information matrix. A further optimality criterion (M,S) optimality is introduced. It is essentially a mixture of Shah's and the popular criteria.

S optimality and (M,S) optimality will be our optimality criteria in this paper. Using these optimality criteria, we have been able to derive some new results which we hope to be of interest to the users and researchers in the field of optimum design theory. To be specific, let $BD\{v,b,(r_i),(k_u)\}$ denote a block design on a set of v treatments with b blocks of size k_u, $u = 1,2,\cdots,b$ and treatment i is replicated r_i times. Then we have shown that for the family of connected block designs $BD\{v,b,(r_i),k\}$ with (i) less than k - 1 treatments having replication equal to one and binary (0,1) the S-optimum design is pseudo-globally connected; (ii) the S-optimum design is globally connected if $r_i > 1$ and the designs are binary; and (iii) at least one treatment with replication greater than b, then the (M,S)-optimum design is pseudo-globally connected. In the final part of this paper we mention some unsolved problems in this area.

ON THE THEORY OF CONNECTED DESIGNS

II. OPTIMALITY

J. A. Eccleston and A. Hedayat
Cornell University

1. Introduction and Summary. The concepts of locally, pseudo-globally and globally connected block designs were introduced by Eccleston and Hedayat (1972). In that paper we were mainly concerned with the characterizations of these designs. In the present paper we shall expand the theory of connected designs by exploring two optimality properties of these designs. Our optimality criteria are in terms of eigen values of the corresponding information matrices. Before proceeding further let us review a general description of optimum designs.

The theory of optimum experiment and treatment designs is essentially the use of a well defined criterion to determine which in a specified class of legitimate or competing designs is the best. So far, almost all contributions to this field have been related to the optimality of non-randomized designs. This paper is also formulated in this framework. The first formal treatment of this subject was given about five decades ago by Smith (1918). It was revived after a 30-year pause by Wald (1943), Mood (1946), Elfving (1952), Chernoff (1953), Ehrenfeld (1955), Kiefer (1958, 1959), Kiefer and Wolfowitz (1959) and others. A voluminous literature has developed around the problem of finding optimal designs. The newly published book, Theory of Optimum Experiments, by V. V. Fedorov (1972) is a clear indication that this branch of statistics is growing fast and has attracted many leading mathematicians and statisticians around the world.

This research was supported by National Institutes of Health Research Grant No. 5 R01-GM-05900 at Cornell University.

AMS subject classification. 62K05, 62K10

Key words. Optimal design, S-optimality, (M,S)-optimality, block design, connected block design, locally connected, pseudo-globally connected, globally connected.

Kiefer (1958) discusses the three most used and well known optimality criteria, namely A, D and E optimality. These optimality criteria involve functions of the non-zero eigen values of the information matrix of the design. Let $\{\lambda_i, i=1,2,\cdots\}$ denote the set of non-zero eigenvalues of the information matrix. Then A, D and E optimality are defined as follows:

A Optimality. Minimize $\sum_i \lambda_i^{-1}$. This is equivalent to minimizing the average variance of all elementary treatment contrasts. The corresponding design is referred to as A-optimum.

D Optimality. Minimize $\prod_i \lambda_i^{-1}$. This is equivalent to minimizing the generalized variance or maximizing the $\prod_i \lambda_i$. The corresponding design is called D-optimum.

E Optimality. Minimize $\max_i \lambda_i^{-1}$. This is equivalent to maximizing $\min_i \lambda_i$. The design which has this property is called E-optimum.

These optimality criteria are not in general related to each other and need not agree in comparing the given designs. Kiefer (1958) shows that under certain conditions on the optimum design, D optimality implies A and E optimality but not vice versa. Also, it is easy to see that if the A-optimum block design is symmetric then it is also D and E-optimum. Where by a symmetric block design we mean a design whose $\underset{\sim}{C}$ matrix has fixed diagonal and fixed off-diagonal entries, or equivalently the non-zero eigenvalues of $\underset{\sim}{C}$ are all equal.

Kiefer and others have pointed out many mathematical properties of D optimality. But we believe that the choice of an optimality criterion in a particular experiment is the experimenter's perogative. However, one point should be kept in mind, which is that the experimenter should be open to compromise because often the search for the optimum design involves the solution of an horrendous mathematical problem which is intractable with the present mathematical machineries. Thus if the exact optimum design, with respect to a given criterion, cannot be obtained the experimenter should be willing either to change his optimality criterion or be satisfied with an approximately optimum design.

The axioms of rational behavior dictate that if the competing designs all enjoy the given optimality criterion then the experimenter should not select his design among the available ones in an arbitrary fashion. But rather, he should consider a new criterion in the process of selection. This idea, together with some other reasons, has led Shah (1960) to introduce an optimality criterion which will be called hereafter S optimality.

<u>S Optimality</u>. Minimize $\sum_i \lambda_i^2$ if the trace of information matrices of the competing designs are identical. The corresponding optimum design will be referred to as S-optimum.

Note that smaller values of $\Sigma\, \lambda_i^2$ will in general tend to give smaller values of $\Sigma\, \lambda_i^{-1}$ and $\Pi\, \lambda_i^{-1}$. Shah (1960) has given other mathematical utilities of this criterion.

We shall here introduce an optimality criterion which is a useful and somewhat hybrid of the preceding optimality criteria. The corresponding optimization is carried in two stages and is formally defined as follows:

(<u>M,S</u>) <u>Optimality</u>. First, form a subclass of designs whose information matrices have maximum trace. Then, select a design from this subclass such that its square of the information matrix has minimum trace. The resulting design is called the (M,S)-optimum design.

S optimality and (M,S) optimality will be our optimality criteria in this paper. Using these optimality criteria, we have been able to derive some new results which we hope to be of interest to the users and researchers in the field of optimum design theory. To be specific, let $BD\{v,b,(r_i),(k_u)\}$ denote a block design on a set of v treatments with b blocks of size k_u, $u = 1,2,\cdots,b$ and treatment i is replicated r_i times. Then we have shown that for the family of connected block designs $BD\{v,b,(r_i),k\}$ with (i) less than $k - 1$ treatments having replication equal to one and binary (0,1) the S-optimum design is pseudo-globally connected; (ii) the S-optimum design is globally connected if $r_i > 1$ and the designs are binary; and (iii) at least one treatment with replication greater than b, then the (M,S)-optimum design is pseudo-globally connected. In the final part of this paper we mention some unsolved problems in this area.

2. <u>Optimality</u>. Let Δ denote the family of all connected designs with parameter set $\{v,b,(r_i),(k_u)\}$. Let also $\Delta_1 \subset \Delta$ denote the set of those designs in Δ which are pseudo-globally connected. Note that the cardinality of Δ_1 ranges from zero to the cardinality of Δ depending on the given set of parameters.

<u>Definition 2.1</u>. Let D_1 and D_2 be two designs in Δ. Then we say D_1 is S-better than D_2 if D_1 has a smaller trace of C squared than D_2.

Consider a situation where the connected designs in Δ are binary with $n_{iu} = 0$ or 1 and proper, i.e., $k_u = k$. These designs constitute most of the well known classical designs. Then we have the following lemma.

<u>Lemma 2.1</u>. <u>Corresponding to any design in $\Delta_2 = \Delta - \Delta_1$ there is a pseudo-globally connected design in Δ_1 which is S-better if less than $k - 1$ of the r_i's are equal to one.</u>

Before giving a proof of this lemma let us recall the following theorem of Eccleston and Hedayat (1972).

Theorem 2.1. *A block design D will be pseudo-globally connected if and only if (1) D is locally connected, (2) Every block of D contains at least two treatments that appear in more than one block, (3) Any treatment, i say, that appears in two or more blocks (but not all blocks) must do so in blocks that contain (i) a treatment that appears in two blocks containing i, and two not containing i, or (ii) two treatments each appearing in a block containing i, and a block not containing i.*

Proof of Lemma 2.1. Let $D \in \Delta_2$. Then by the conditions imposed on Δ the design D satisfies conditions (1) and (2) of theorem 2.1. Therefore, condition (3) must be violated by one or more treatments in D. We shall devise an algorithm which involves the rearrangement of the experimental units in D in a manner such that the resulting design $\bar{D}$ is pseudo-globally connected and is S-better than D.

Suppose treatment i fails to satisfy condition (3) of theorem 2.1 but since the design is locally connected there exists a treatment ℓ that

(a) belongs to only one block containing i and at least one not containing i, or

(b) belongs to at least one block containing i and only one not containing i.

The design can be divided into two parts, T_i the set of blocks which contain i, and $D - T_i$ the set of blocks which do not contain i. We discuss (a) only, but an analogous proof holds for (b). For the designs we are considering there exists a replicate of treatment $z \in B_r \in T_i$, $r_z > 1$, $z \neq \ell$ and a replicate of treatment $p \in B_t \in D - T_i$, $r_p > 1$, $p \neq \ell$ which can be interchanged to yield a design in which treatment i satisfies condition (3). Such a z and p always exist since there are less than k - 1 treatments with $r_i = 1$. Whether or not the interchange yields a smaller trace of $\underset{\sim}{C}^2$ depends on the change in the elements of $\underset{\sim}{C}$, in particular, the elements of the row corresponding to treatment ℓ. The possibilities are as follows:

(i) Suppose $\ell \in B_r$ and $\ell \in B_t$. The elements of $\underset{\sim}{C}$ that are changed are as follows (recall that all diagonal elements are fixed for all designs of this lemma).

Before interchange		After interchange
c_{zi}	$\longrightarrow$	$c_{zi} + \frac{1}{k}$
c_{zm}	$\longrightarrow$	$c_{zm} + \frac{1}{k}$ where $m \in B_r$, $m \neq \ell$, $m \neq z$, $m \neq i$
c_{pw}	$\longrightarrow$	$c_{pw} + \frac{1}{k}$ where $w \in B_t$, $w \neq \ell$, $w \neq p$
$c_{zw}(=0)$	$\longrightarrow$	$-\frac{1}{k}$ for all w of which there are $k-3$
$c_{pm}(=0)$	$\longrightarrow$	$-\frac{1}{k}$ for all m of which there are $k-2$.

All other elements of $\underset{\sim}{C}$ are unchanged. Thus trace of $\underset{\sim}{C}^2$ before the interchange can be written as

$$\text{tr}\, \underset{\sim}{C}^2 = c_{zi}^2 + \sum_m c_{zm}^2 + \sum_w c_{zw}^2 + \sum_m c_{pm}^2 + \sum_w c_{pw}^2 + \text{Remainder} . \tag{2.1}$$

After the interchange

$$(2.2)\qquad \operatorname{tr} \underset{-}{C}^2 = (c_{zi} + \tfrac{1}{k})^2 + \sum_m (c_{zm} + \tfrac{1}{k})^2 + \sum_w \tfrac{1}{k^2} + \sum_m (\tfrac{1}{k})^2 + \sum_w (c_{pw} + \tfrac{1}{k})^2$$

$$+ \text{ Remainder} \quad .$$

The remainder term is the same for both equations (2.1) and (2.2); therefore, their difference is

$$(2.3)\qquad (2.1) - (2.2) = -2c_{zi}\,\frac{1}{k} - \frac{2}{k^2} - 2\sum_m c_{zm}\,\frac{1}{k} - \frac{2(k-3)}{k^2} - 2\sum_w c_{pw}\,\frac{1}{k}$$

$$- 2\left(\frac{k-3}{k^2}\right) \quad .$$

We know that $c_{mn} \leq 0$ for $m \neq n$; therefore, $-c_{mn} \geq 0$. Since

$$-c_{zi} > \frac{1}{k}$$

$$-c_{zm} \geq \frac{1}{k} \quad \text{for all } m$$

$$-c_{p\ell} \geq \frac{1}{k}$$

and

$$-c_{z\ell} = \frac{1}{k} \quad .$$

Therefore $(2.3) > 0$. Thus the design is S-better after the interchange.

(ii) Suppose $\ell \notin B_r$ and $\ell \in B_t$, then

Before interchange		After interchange
$c_{p\ell}$	$\longrightarrow$	$c_{p\ell} + \frac{1}{k}$
$c_{z\ell} (= \frac{-1}{k})$	$\longrightarrow$	$c_{z\ell} - \frac{1}{k}$.

All other elements of $\underset{-}{C}$ are as in (i). Thus the difference between the trace of $\underset{-}{C}^2$ before and after interchange is the same as in (i) except for the $c_{p\ell}$ and $c_{z\ell}$ terms. Therefore, before interchange

$$\text{(2.4)} \qquad \text{tr } \underset{-}{C}^2 = c^2_{p\ell} + c^2_{z\ell} + [(2.1) - c^2_{p\ell} - c^2_{z\ell}]$$

and after interchange

$$\text{(2.5)} \qquad \text{tr } \underset{-}{C}^2 = (c_{p\ell} + \frac{1}{k})^2 + (c_{z\ell} - \frac{1}{k})^2 + [(2.2) - c^2_{p\ell} - c^2_{z\ell}] \ .$$

From (i) we have

$$(2.4) - (2.5) > \frac{2c_{p\ell}}{k} - \frac{4}{k^2} \ .$$

If $c_{p\ell} = -\frac{1}{k}$ then (2.4) - (2.5) may not be greater than zero. But recall that ℓ belongs to only one block in T_i, namely B_r, and since $r_z > 1$ there exists a replicate of $z \in B_s$ and $\ell \in B_s$ which can be used for the interchange rather than $z \in B_r$. The interchange between $z \in B_s$ and $p \in B_t$ is equivalent to (i).

(iii) Suppose $\ell \notin B_r$ and $\ell \notin B_t$. This is analogous to (i) and the design after the interchange is S-better.

(iv) Suppose $\ell \in B_r$ and $\ell \notin B_t$, then

Before interchange		After interchange
$c_{z\ell}(= \frac{-1}{k})$	$\longrightarrow$	$c_{z\ell} + \frac{1}{k}(=0)$
$c_{p\ell}$	$\longrightarrow$	$c_{p\ell} - \frac{1}{k}$.

All other elements of $\underset{\sim}{C}$ are as in (i). As in (ii) we have that before the interchange

$$(2.6) \qquad \mathrm{tr}\, \underset{\sim}{C}^2 = c_{z\ell}^2 + c_{p\ell}^2 + [(2.1) - c_{z\ell}^2 - c_{p\ell}^2]$$

and after the interchange

$$(2.7) \qquad \mathrm{tr}\, \underset{\sim}{C}^2 = (c_{z\ell} + \frac{1}{k})^2 + (c_{p\ell} - \frac{1}{k})^2 + [(2.2) - c_{z\ell}^2 - c_{p\ell}^2] \ .$$

From (i) we have

$$(2.6) - (2.7) > \frac{2c_{p\ell}}{k} \ .$$

We know that $c_{p\ell} \leq \frac{-1}{k}$; thus (2.6) - (2.7) may not be greater than zero. But recall that ℓ belongs to only one block in T_i, namely B_r, and since $r_z > 1$ there exists a replicate of $z \in B_s \in T_i$, $s \neq r$, which can be used for the interchange. The interchange is now between $z \in B_s$ and $p \in B_t$ with $\ell \notin B_s$ and $\ell \notin B_t$ which is equivalent to (ii).

If now there exists another treatment, q say, that fails to satisfy condition (3) it can be corrected so that the interchange for i is not negative. Reversing the interchange between z and p is the only way to negate the correction for i. Let treatment m be to treatment q as ℓ was to treatment i, $\ell \neq m$

otherwise the correction for i would be sufficient for q (see example 2.1). Suppose the correction for q reverses the interchange between z and p. This implies

(a') either $q \in B_r \in T_q$ and $B_t \in D - T_q$ or $q \in B_t \in T_q$ and $B_r \in D - T_q$ but $\ell \in B_r$ and B_t; therefore q does not fail condition (3). This is a contradiction.

or

(b') q, i and $p \in B_r \in T_q$ and $z \in B_t \in D - T_q$. Then all blocks containing z must belong to $D - T_q$, but z and i belong to the same block at least once and similarly if $B_r \in D - T_q$ and $B_t \in T_q$. This implies that q satisfies condition (3),which is a contradiction.

So, in general, any treatments which fail condition (3) of theorem 2.1 can be corrected to yield a pseudo-globally connected design which is S-better. This completes the proof.

From lemma 2.1 we have the following theorem.

<u>Theorem 2.2. Within the family of connected designs</u> $BD\{v,b,(r_i),k\}$ <u>with</u> $n_{iu} = 0$ <u>or</u> 1 <u>the S-optimal design is pseudo-globally connected if there are less than</u> $k - 1$ <u>treatments with</u> $r_i = 1$.

If Δ_1 contains a globally connected design then we have the following lemma and theorem.

<u>Lemma 2.2. Corresponding to any design in</u> $\Delta_2 = \Delta - \Delta_1$ <u>there is a globally connected design which is S-better if all</u> $r_i > 1$ <u>and</u> $k \geq 3$.

The proof is analogous to that of lemma 2.1.

Theorem 2.3. *Within the family of connected designs* BD$\{v,b,(r_i),k\}$ *with* $n_{iu} = 0$ *or* 1 *the S-optimal design is globally connected if all* $r_i > 1$ *and* $k \geq 3$.

Instead of $r_i > 1$ and $k \geq 3$ it is sufficient if all $r_i \geq 2$ for lemma 2.2 and theorem 2.3 to be true.

Example 2.1. Let D be the following locally connected design in BD$\{9,6,(2,2,3,1,2,6,2,2,1),3\}$.

	B_1	B_2	B_3	B_4	B_5	B_6
	1	1	3	3	6	6
D:	2	2	4	5	7	7
	9	3	5	6	8	8

trace of $\underline{C} = 12$ and trace $\underline{C}^2 = 24$

Treatments 1, 2, 7 and 8 fail to satisfy condition (3) of theorem 2.1. In the notation of the proof of lemma 2.1 for treatments 1 and 2, $\ell = 3$ and for treatments 7 and 8, $\ell = 6$. Therefore, a correction for treatments 1 and 7 will be sufficient for treatments 2 and 8 respectively. By interchanging $2 \in B_2$ with $5 \in B_3$ and $8 \in B_5$ with $3 \in B_4$ results in

	B_1	B_2	B_3	B_4	B_5	B_6
	1	1	3	8	6	6
$\bar{D}$:	2	5	4	5	7	7
	9	3	2	6	3	8

trace of $\underline{C} = 12$ and trace $\underline{C}^2 = 68/3$

is globally connected and S-better.

If the condition of lemma 2.1 and theorem 2.2 is relaxed so as to include designs with more than $k - 1$ treatments with $r_i = 1$ then the lemma and theorem no longer hold in general. A counterexample which is too lengthy to present here can be found in Eccleston (1972).

Recall that the procedure for determining the (M,S) optimal design is to first find the class of designs with maximum trace of $\underline{C}$ and then within that class determine those with minimum trace of $\underline{C}$ squared. Let Δ, Δ_1 and Δ_2 be as defined in the first paragraph of this section; then we have

Lemma 2.3. *Any design in Δ_2 can be transformed into a design in Δ_1 with the same trace of $\underline{C}$.*

Proof. For design $BD\{v,b,(r_i),(k_u)\}$ to be locally but not pseudo-globally connected either or both conditions (2) and (3) of theorem 2.1 fail to be satisfied. Each can be corrected by an interchange(s) as described in lemma 2.1. Suppose $z \in B_r$ and $p \in B_t$ are interchanged to correct either condition (2) or (3). The only diagonal elements of $\underline{C}$ affected by the interchange are c_{zz} and c_{pp}. Block B_r is of size k_r and B_t is of size k_t:

$$c_{zz} \quad \text{becomes} \quad c_{zz} - \frac{1}{k_t} + \frac{1}{k_r}$$

and

$$c_{pp} \quad \text{becomes} \quad c_{pp} - \frac{1}{k_r} + \frac{1}{k_t} \quad .$$

Therefore, the trace of $\underline{C}$ after interchange remains invariant. The same argument follows no matter how many interchanges are performed.

<u>Theorem 2.4. For the family of connected designs</u> $BD\{v,b,(r_i),k\}$ <u>the</u> (M,S) <u>optimal design is pseudo-globally connected if there exists at least one</u> $r_i \geq b$.

<u>Proof.</u> Suppose treatment i is such that $r_i \geq b$; obviously max trace of C $= \sum_i \max c_{ii}$ where $c_{ii} = r_i - \sum_u n_{iu}^2/k$. Maximizing c_{ii} is equivalent to minimizing n_{iu} for all u. Since $r_i \geq b$ this implies that all n_{iu} should be as close to equal as possible and a replicate of treatment i occur in every block, i.e., $|n_{iu} - n_{iu'}| \leq 1$ for all u, u' and $n_{iu} \geq 1$ for all u. The above results for treatment i imply that condition (3) of theorem 2.1 is satisfied. Now if condition (2) is not satisfied by B_s, say, then every treatment in B_s except i occurs in no other block of the design. Clearly an interchange similar to that of lemma 2.1 can be performed to yield a pseudo-globally connected design which is S-better. Therefore the (M,S) optimal design will be pseudo-globally connected.

<u>Corollary 2.1. For the family of connected designs</u> $BD\{v,b,(r_i),k\}$ <u>the</u> (M,S) <u>optimal design is globally connected if there exists two</u> $r_i \geq b$ <u>or one</u> $r_i \geq 2b$.

3. <u>Concluding Remarks</u>. Results analogous to lemma 2.1 and theorem 2.2 for nonproper designs have not been proved as yet. The (M,S) optimality of the family of connected nonproper designs remains unsolved. Perhaps some method of generating pseudo-globally connected designs other than that used here may yield better optimality results. However, apart from the intuitive feeling that one should chose as the optimal design the one with a $\underset{\sim}{C}$ matrix "as close as possible" to the unattainable $\underset{\sim}{C}$ matrix with equal diagonal and equal off-diagonal entries, virtually nothing is known about the optimality of nonproper designs under any criterion. Thus, there remains a vast and challanging area of optimum design theory open to research.

4. Acknowledgement. The authors wish to thank Professor J. Kiefer for his helpful suggestions and discussions concerning the theory of optimum design.

REFERENCES

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

Eccleston, J. (1972). On the theory of connected designs. Ph.D. Thesis, Cornell University.

Eccleston, J. and Hedayat, A. (1972). On the theory of connected designs. I. Characterization. BU-425-M, Cornell University.

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

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

Fedorov, V. V. (1972). Theory of Optimal Experiments. Academic Press, New York.

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

Kiefer, J. (1959). Optimum experimental designs (with discussion). JRSS(B) 21 272-319.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. Ann. Math. Statist. 30 271-294.

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

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

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

Wald, A. (1943). On efficient design of statistical investigations. Ann. Math. Statist. 14 131-140.