On the Connectivity Problem for m -way Designs

J.D. Godolphin, *Department of Mathematics,
University of Surrey, Guildford, Surrey. GU2 7XH, UK.
Email: j.godolphin@surrey.ac.uk*

Abstract

The problem of ascertaining conditions which ensure that an m -way design is connected has occupied the attention of research workers for very many years. One of the significant advances, as well as one of the earliest contributions, was provided by the classic work of J.N. Srivastava and D. A. Anderson in 1970 which gives a necessary and sufficient rank condition for an m -way design to be completely connected. In this paper it is shown that the class of estimable parametric functions for an individual factor is derived directly from a simple extension of the Srivastava-Anderson result. This takes the form of a necessary and sufficient rank condition which is expressed in terms of the dimension of a segregated component of the kernel of the design matrix. The result has the interesting property that the connectivity status for all of the individual factors can be found simultaneously. Furthermore, it enables the formulation of several general results, which include the specification of conditions on designs exhibiting adjusted orthogonality. A number of examples are given to illustrate these results.

AMS Subject Classification: 62K99 and 62K10

Key Words: Adjusted orthogonality; Connectivity; Estimability; Information matrix; Kernel space; Segregated component space.

1 Introduction

A scalar parametric function is said to be estimable if it has an estimator which is linear in the observation vector Y and is unbiased (Bose, 1944). The problem of determining properties of an experimental design which ensure that all scalar parametric functions are estimable has occupied the attention of research workers for many years. Significant contributions in this field are due to Chakrabarti (1963), Srivastava and Anderson (1970), Eccleston and Hedayat (1974), Eccleston and Russell (1975), Raghavarao and Federer (1975), Shah and Dodge (1977), Dodge (1985), John and Williams (1995), Park and Shah (1995) and Godolphin and Godolphin (2001). These and other works have been concerned with the m -way classification, or particular cases of it, especially $m = 2$ or $m = 3$. Srivastava and Anderson (1970) gave a condition for all m factors to be completely connected, which involves the rank of $X'X$ where X is the design matrix. This result appears to be the most informative criterion available and is also simple to apply. However, the problem of finding conditions for a single factor to be connected in a m -way design, so that all linear contrasts in the levels of that particular factor are estimable, has not been resolved satisfactorily even for $m = 3$, as pointed out previously by Bérubé and Styan (1993).

The purpose of this paper is to show that a necessary and sufficient rank condition exists for each of the individual factors in the m -way classification, which is comparable to the Srivastava-Anderson

condition for the overall design. The result is determined directly from the kernel $\mathcal{K}(X)$, the vector space orthogonal to the row space of X . Furthermore, the class of estimable parametric functions is derived simultaneously for each factor from segregated components of $\mathcal{K}(X)$. No comparable approach appears to be derivable from the methods of Butz (1982), Dodge (1985), Searle (1987) or Wynn (2008), who investigate connectivity for the m -way classification. The method of the paper verifies and generalizes several classical results, including known theorems which deal with designs exhibiting adjusted orthogonality; and in many cases the proofs given here are simpler than the arguments presented in the original papers.

The estimability space, its link with X and the Srivastava-Anderson criterion for complete connectivity of the m -way classification are discussed briefly in § 2. Criteria for single-factor connectivity based on segregated components of $\mathcal{K}(X)$ and specification of the corresponding estimability space are derived in § 3, where the equivalence with the generalized information matrix approach is established. In § 4 a general result is derived for an m -way classification exhibiting adjusted orthogonality which generalizes results that are available in the literature for the cases $m = 3$ and $m = 4$. Several of the main results are illustrated by examples.

2 Connectivity criteria for m -way designs

2.1 Estimability

Let 1_n denote the $n \times 1$ unit vector, i.e. all n elements of 1_n are unity, and let 0_n be the $n \times 1$ vector whose elements are all zero. Suppose that the $n \times 1$ observation vector Y is generated by the m -way additive model, such that factor i occurs at $b_i \geq 2$ levels ($i = 1, 2, \dots, m$), specified by

$$E(Y) = \mu 1_n + X_1 \beta_1 + X_2 \beta_2 + \dots + X_m \beta_m, \quad (2.1)$$

where μ is a scalar constant and $\beta_1, \beta_2, \dots, \beta_m$ are $b_1 \times 1, b_2 \times 1, \dots, b_m \times 1$ parameter vectors relating to factors 1, 2, $\dots$, m respectively, such that $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ib_i})'$ for $i = 1, 2, \dots, m$. Let $X = [1_n \ X_1 \ X_2 \ \dots \ X_m]$ denote the design matrix for model (2.1) where X_i is the $n \times b_i$ component of the design matrix pertaining to the i th factor so each row of X_i has one element unity and $b_i - 1$ zeros. In this paper, the symbol r is reserved for $\text{Rank}(X)$, the rank of the design matrix which is defined by model (2.1).

For simplicity, write $\beta = (\beta'_1, \beta'_2, \dots, \beta'_m)'$ and put

$$b = \sum_{i=1}^m b_i \quad \text{and} \quad r_c = 1 + b - m. \quad (2.2)$$

It is further assumed that the number of observations, n , is at least equal to r_c . As a consequence of the specialized structure of the X_i matrices, m independent relationships exist between the columns of X , viz.:

$$X_1 1_{b_1} = X_2 1_{b_2} = \dots = X_m 1_{b_m} = 1_n. \quad (2.3)$$

Equations (2.3) are referred to as the *standard restrictions on X* . Define the $b \times 1$ vector λ to be conformable with β , so that $\lambda = (\lambda'_1, \lambda'_2, \dots, \lambda'_m)'$ where $\lambda_i = (\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{ib_i})'$ is a $b_i \times 1$ sub-vector for $i = 1, 2, \dots, m$. The criterion for estimability of an arbitrary parametric combination is well known; see e.g. Sengupta and Jammalamadaka (2003, proposition 4.1.10).

Theorem 1 [Estimability criterion] *For the m -way design specified by (2.1), the linear combination $\ell = (\lambda_\mu \ \lambda') \begin{pmatrix} \mu \\ \beta \end{pmatrix}$ is estimable if and only if $\begin{pmatrix} \lambda_\mu \\ \lambda \end{pmatrix} \in \mathcal{C}(X')$, where λ_μ is a scalar and $\mathcal{C}(X')$ is the column space of X' .*

Remark Sengupta and Jammalamadaka (2003, proposition 4.1.15) and Christensen (2002, Theorem 2.1.2) raise a related fundamental issue; *viz.* that the stated condition of Theorem 1 is necessary and sufficient, not only for the parametric combination ℓ to have a linear unbiased estimator, but also for ℓ to be unambiguously identified, given full knowledge of $E[Y]$.

Particular interest centres on linear combinations of parameters that are estimable but do not involve μ . This focusses attention on the λ -space, Λ , of $b \times 1$ vectors that specify the coefficients of estimable parametric linear functions

$$\Lambda = \{\lambda : \lambda' \beta \text{ is estimable}\}. \quad (2.4)$$

Theorem 1 shows that $\lambda' \beta$ is estimable if and only if $\begin{pmatrix} 0 \\ \lambda \end{pmatrix} \in \mathcal{C}(X')$. Furthermore, it is clear that a

basis of the subspace of $\mathcal{C}(X')$, which is constrained by $\lambda_\mu = 0$, is given by consecutive first differences of r independent columns of X' . It follows that the dimension of the λ -space (2.4) is $\text{Dim}(\Lambda) = r - 1 = \text{Rank}(X) - 1$.

It is straightforward to see that the standard restrictions given by (2.3) impose the condition $\lambda'_i 1_{b_i} = 0$ on the i th component of λ for each $i = 1, 2, \dots, m$, and this observation implies the following result.

Lemma 1 *A necessary condition for $\lambda' \beta$ to be estimable is that $\lambda \in \Lambda_c$, where*

$$\Lambda_c = \{\lambda : \lambda'_i 1_{b_i} = 0 \text{ for } i = 1, 2, \dots, m\}, \quad (2.5)$$

i.e. any estimable parametric function is a sum of contrasts for the individual factors.

2.2 Complete connectivity

Srivastava and Anderson (1970) suggest the term *completely connected* in preference simply to ‘connected’ to describe a design for which $\Lambda = \Lambda_c$; i.e. the design has the property that all parametric linear functions $\lambda' \beta$ for arbitrary $\lambda \in \Lambda_c$ are estimable. Qualification is necessary to avoid possible ambiguity when considering various sub-vectors of β . The term ‘completely connected’ is used by Butz (1982) but it is not a universal expression; e.g. Eccleston and Hedayat (1974) employ ‘locally connected’ for a design with the property. The classic Srivastava-Anderson rank criterion is as follows.

Theorem 2 [Srivastava and Anderson] *A necessary and sufficient condition for the m -way design (2.1) to be completely connected is that $\text{Rank}(X'X) = r_c$.*

Two variations of the Srivastava-Anderson criterion are useful for what follows.

Corollary 1 *Each of the following two conditions is necessary and sufficient for an m -way design defined by (2.1) to be completely connected:*

- (i) $\text{Dim}(\mathcal{K}(X)) = m$, where $\mathcal{K}(X)$ is the kernel of X ;
- (ii) *All linearly dependent relationships between the columns of X are generated by the standard restrictions (2.3).*

Proof: The rank condition of Theorem 2 is obviously equivalent to $\text{Rank}(X) = r_c$ and this is equivalent to condition (i).

A linearly dependent relation exists between columns of X if there is a vector $\pi \neq 0$ such that $X\pi = 0$, i.e. $\pi \in \mathcal{K}(X)$. Let $\pi_i = (-1 \ 0'_{b_1} \ \dots \ 0'_{b_{i-1}} \ 1'_{b_i} \ 0'_{b_{i+1}} \ \dots \ 0'_{b_m})'$ for $i = 1, \dots, m$, then $\pi_1, \dots, \pi_m$ are independent and span a space Π which, from (2.3), is a subspace of $\mathcal{K}(X)$. Since $\Pi = \mathcal{K}(X)$ if and only if $\text{Dim}(\mathcal{K}(X)) = m$, see Harville (1997, Lemma 11.4.1), then (ii) is equivalent to (i) and the corollary is proved. $\square$

Example 1: Srivastava and Anderson four factor design

To illustrate these criteria for complete connectivity, consider the 4-way multidimensional partially balanced design given by Srivastava and Anderson (1970). This $6 \times 4 \times 8 \times 12$ design is represented in Figure 1 as a row-column arrangement with factors designated by: ‘columns’, ‘rows’, ‘Latin letter labels’ and ‘number labels’.

Fig. 1: Srivastava and Anderson 4-way design

	col 1		col 2		col 3		col 4		col 5		col 6	
row 1	F 10	D 5	D 12	F 3	H 12	B 8	B 10	H 2	C 3	E 2	E 5	C 8
row 2	E 9	C 5	C 11	E 3	A 9	G 4	G 6	A 11	D 3	F 4	F 5	D 6
row 3	B 1	H 10	H 7	B 12	D 4	F 12	F 6	D 10	A 4	G 1	G 7	A 6
row 4	A 1	G 9	G 11	A 7	E 8	C 9	C 2	E 11	B 2	H 1	H 8	B 7

In this example $m = 4$ and $b_1 = 6$, $b_2 = 4$, $b_3 = 8$, $b_4 = 12$, so $r_c = 27$ from (2.2). The design matrix, X , has dimension 48×31 ; the Srivastava-Anderson rank condition is satisfied since $r = 27$ so it is concluded immediately that this design is completely connected and therefore all linear contrasts $\lambda'_i \beta_i$ are estimable for each $i = 1, \dots, 4$.

Equivalently, $\text{Dim}(\mathcal{K}(X)) = 4$ and the same conclusion is reached by Corollary 1 (i). The columns of the matrix

$$\Pi = (\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4) = \begin{pmatrix} -1 & -1 & -1 & -1 \\ 1_6 & 0_6 & 0_6 & 0_6 \\ 0_4 & 1_4 & 0_4 & 0_4 \\ 0_8 & 0_8 & 1_8 & 0_8 \\ 0_{12} & 0_{12} & 0_{12} & 1_{12} \end{pmatrix}, \quad (2.6)$$

using the notation in the proof of Corollary 1, form a basis for $\mathcal{K}(X)$ in this case.

Example 2: Butz six factor design

Another illustration is given by the 6-way design due to Butz (1982, p. 137), which is presented as a $3 \times 2 \times 2 \times 2 \times 2 \times 12$ six-dimensional hyperrectangle with observations at 20 six-tuple coordinates given by

$$\begin{aligned} &(1, 1, 1, 1, 1, 3), (1, 1, 1, 1, 1, 5), (1, 1, 1, 1, 2, 7), (1, 1, 1, 2, 1, 12), (1, 1, 2, 1, 1, 2), \\ &(1, 1, 2, 2, 1, 11), (1, 2, 1, 1, 1, 4), (1, 2, 1, 2, 1, 5), (1, 2, 2, 1, 1, 12), (1, 2, 2, 2, 1, 7), \\ &(2, 1, 1, 1, 1, 9), (2, 1, 2, 1, 1, 8), (2, 1, 2, 2, 1, 7), (2, 2, 1, 1, 1, 10), (2, 2, 1, 2, 1, 1), \\ &(2, 2, 2, 1, 1, 2), (2, 2, 2, 2, 1, 1), (3, 1, 1, 1, 1, 6), (3, 1, 1, 2, 1, 7), (3, 2, 1, 1, 1, 5), \end{aligned}$$

and with no further observations at any other coordinates. In this example $m = 6$ and $b_1 = 3$, $b_2 = b_3 = b_4 = b_5 = 2$, $b_6 = 12$ so that $r_c = 18$ from (2.2). The design matrix, X , has dimension 20×24 and it is found that $r = 18$ or, equivalently, $\text{Dim}(\mathcal{K}(X)) = 6$. Appealing to Theorem 2 or Corollary 1, it is concluded that the Butz 6-way design is completely connected. The columns of $\Pi = (\pi_1 \pi_2 \pi_3 \pi_4 \pi_5 \pi_6)$ form a basis for $\mathcal{K}(X)$ in this case. Butz (1982) gives a graphical approach which is based on a search for independent cycles in a representative digraph, but this procedure is cumbersome and prone to error compared to the application of Theorem 2 or Corollary 1.

3 Single factor connectivity

3.1 Disconnected design

If the design for the m -way additive model, specified by (2.1), is not completely connected it is *disconnected*; i.e. the design is disconnected if and only if Λ , the estimable λ -space defined by (2.4), is a proper subspace of Λ_c , the ‘contrasts’ λ -space defined by (2.5). When $\Lambda \subset \Lambda_c$ but $\Lambda \neq \Lambda_c$ not all parameter contrasts belonging to Λ_c are estimable. It follows that disconnectivity is marked by a deficiency in the dimension of Λ , when compared to Λ_c . The Srivastava-Anderson rank condition of Theorem 2 shows that disconnectivity is marked by a corresponding deficiency in the rank of X .

For some positive integer d , a disconnected design has estimable λ -space Λ with dimension $\text{Dim}(\Lambda) = r_c - d - 1$, or equivalently, the rank of X is $r = r_c - d$, implying that the dimension of $\mathcal{K}(X)$ is $m + d$. Following Godolphin and Godolphin (2001), we say that the design is *disconnected of order d* . The standard restrictions (2.3) hold for this disconnected design and, furthermore, d additional independent restrictions can be found between the columns of X , i.e.

$$X_1\alpha_{1j} + X_2\alpha_{2j} + \cdots + X_m\alpha_{mj} = 0_n, \quad (3.1)$$

for $j = 1, \dots, d$, where α_{ij} is a $b_i \times 1$ vector. It is assumed here, without any loss of generality, that none of the relationships (3.1) involve 1_n , the first column of X ; i.e. they can be summarized by the matrix equation $XA = 0$, where A is a $(b+1) \times d$ matrix with initial row $0'_d$. Hence a full rank *linear constraint matrix* L , of dimension $(b+1) \times (m+d)$, can be defined by

$$L = (\Pi \quad A) = \begin{pmatrix} -1 & -1 & \cdots & -1 & 0 & \cdots & 0 \\ 1_{b_1} & 0_{b_1} & \cdots & 0_{b_1} & \alpha_{11} & \cdots & \alpha_{1d} \\ 0_{b_2} & 1_{b_2} & \cdots & 0_{b_2} & \alpha_{21} & \cdots & \alpha_{2d} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0_{b_m} & 0_{b_m} & \cdots & 1_{b_m} & \alpha_{m1} & \cdots & \alpha_{md} \end{pmatrix} = \begin{pmatrix} L_0 \\ L_1 \\ L_2 \\ \vdots \\ L_m \end{pmatrix}, \quad (3.2)$$

and then equations (2.3) and (3.1) are represented in the single matrix equation

$$XL = 0, \quad (3.3)$$

where the zero matrix on the right side of equation (3.3) has dimension $n \times (m+d)$. It follows that the kernel $\mathcal{K}(X) = \mathcal{C}(L)$, see Harville (1997, Lemma 11.4.1), i.e. for any $(b+1) \times 1$ vector π we have $X\pi = 0$ if and only if $\pi \in \mathcal{C}(L)$.

3.2 Connectivity criteria for single factors

The purpose of this section is to investigate conditions for all linear contrasts between the levels of an individual factor to be estimable and the term ‘singly connected factor’ is introduced to describe this.

Definition 3.1 *An individual factor of an m -way classification is singly connected if and only if all linear contrasts between the levels of this factor are estimable.*

In particular, given that factor i has a defining parameter vector β_i at b_i levels, it transpires that factor i is singly connected if and only if there are $b_i - 1$ independent estimable parameter contrasts of the form $\lambda'_i\beta_i = 0$ involving elements only from β_i .

Remark It is commonplace in the literature, for example John and Williams (1995, §1.8 and §5.4), to describe a design with one treatment factor and $m - 1$ blocking factors as ‘connected’ if all treatment contrasts are estimable. But although this description focusses on the most important factor in the design, it is always likely to be ambiguous if $m \geq 3$ since this concept of ‘connected’ cannot determine complete connectivity if no account is taken of the other factors. It emerges from what follows – confer Example 4 – that a design can be ‘connected’ in the sense alluded to here, i.e. that the treatment factor is singly connected, even though the design is not completely connected.

It is evident that if a m -way design is completely connected then all m factors are singly connected. The argument of Section 3.1 demonstrates that a m -way design, whether completely connected or not, is marked by restrictions on X given by (3.1), that are identified from the linear constraints matrix L specified by the kernel $\mathcal{C}(L) = \mathcal{K}(X)$. In particular, if $m \geq 3$ and the design is disconnected there is a possibility that some factors may be singly connected. It is now shown that the conditions for this to arise depend wholly on the row components of L .

Theorem 3 *Let a m -way design be disconnected of order $d \geq 1$. Suppose that the kernel of X is expressed as $\mathcal{K}(X) = \mathcal{C}(L)$ and that the matrix $L = (L'_0 \ L'_1 \ \dots \ L'_m)'$ is partitioned into $m+1$ submatrices, where L_0 is a row vector and the dimensions of L_i are $b_i \times (m+d)$ for $i = 1, 2, \dots, m$. Then*

(i) Factor i is singly connected if and only if $\text{Rank}(L_i) = 1$;

(ii) The factor contrast $\lambda'_i \beta_i$ is estimable if and only if $\lambda'_i L_i = 0'_{m+d}$, i.e. $\lambda_i \in \mathcal{K}(L'_i)$.

Proof: Let the design be disconnected of order $d \geq 1$. The $(b+1) \times (m+d)$ linear constraint matrix L of (3.2) has rank $m+d$ and the columns of L form a basis for $\mathcal{K}(X)$. By Theorem 1, $\lambda' \beta$ is estimable if and only if $\begin{pmatrix} 0 \\ \lambda \end{pmatrix}$ is orthogonal to $\mathcal{K}(X)$, therefore

$$\begin{pmatrix} 0 & \lambda' \end{pmatrix} L = \sum_{j=1}^m \lambda'_j L_j = 0'_{m+d}. \quad (3.4)$$

Consider the contrast $\lambda'_i \beta_i$ which involves elements only from the vector β_i , so that $\lambda_j = 0$ for all $j \neq i$. By equation (3.4), $\lambda'_i \beta_i$ is estimable if and only if $\lambda'_i L_i = 0$, so condition (ii) follows. Furthermore, let $\Lambda_i = \{\lambda_i : \lambda'_i L_i = 0\} = \mathcal{K}(L'_i)$ be the estimability space for λ_i , then

$$\text{Dim}(\Lambda_i) = \text{Dim}(\mathcal{K}(L'_i)) = b_i - \text{Rank}(L_i); \quad (3.5)$$

see Harville (1997, §11.3). By definition, factor i is singly connected if and only if $\text{Dim}(\Lambda_i) = b_i - 1$, therefore condition (i) is established and the theorem follows. $\square$

Corollary 2 *Let a m -way design be disconnected of order $d \geq 1$. Then the i th factor is singly connected if and only if the d restrictions of the form (3.1) are such that $\alpha_{ij} = 0_{b_i}$ or α_{ij} is a multiple of the unit vector 1_{b_i} for each $j = 1, 2, \dots, d$.*

Proof: The columns of L_i consist of 1_{b_i} , the zero vector 0_{b_i} repeated $(m-1)$ times and $\alpha_{i1}, \dots, \alpha_{id}$, hence the condition of Corollary 2 is necessary and sufficient for $\text{Rank}(L_i) = 1$. $\square$

Without loss of generality it is assumed that L is defined by (3.2), but Theorem 3 is valid if the columns of L comprise any basis of the space $\mathcal{K}(X)$. Theorem 3 (i) gives a convenient method for finding the connectivity status of each factor, which can be regarded as an extension of the Srivastava-Anderson rank condition for the m factors. All that is required is the evaluation of the ranks of the m submatrices $L_1, \dots, L_m$. This appears to be the only procedure that determines single factor status of all m factors simultaneously. Even when $m = 3$, this is much easier than deriving canonical efficiency factors for each factor, as suggested by John and Williams (1995), or pursuing graphical approaches of Butz (1982) or Ghosh (1986), or using spanning methods of Dodge (1985) or Park and Shah (1995). Condition (ii) of Theorem 3 specifies the estimability spaces $\Lambda_i = \mathcal{K}(L'_i)$, for each $i = 1, 2, \dots, m$, which are referred to as the *segregated components* of $\mathcal{K}(X)$; for each i , Λ_i is the space of all λ_i such that $\lambda'_i \beta_i$ is estimable, giving estimable functions in the levels of factor i only, and it has dimension given by (3.5).

Example 3: Birkes, Dodge and Seely four factor design

To illustrate the results given by Theorem 3, consider the 4-way design, displayed in Figure 2, which is a Graeco-Latin square with two missing cells. This design is discussed by Birkes *et al.* (1976) and Dodge (1985).

Fig. 2: Birkes, Dodge and Seely 4-way design

*	$B\beta$	$C\gamma$	$D\delta$
$B\gamma$	$A\delta$	$D\alpha$	$C\beta$
$C\delta$	*	$A\beta$	$B\alpha$
$D\beta$	$C\alpha$	$B\delta$	$A\gamma$

Here “*” denotes a null or missing entry. In this example $m = 4$ and $b_1 = b_2 = b_3 = b_4 = 4$ so that $r_c = 13$ from (2.2). The design matrix, X , has dimension 14×17 and it is found that $r = 12$ or, equivalently, $\text{Dim}(\mathcal{K}(X)) = 5$ so the Birkes, Dodge and Seely design is disconnected of order $d = 1$.

For convenience, relabel Latin letters A, B, C, D and Greek letters $\alpha, \beta, \gamma, \delta$ as levels 1, 2, 3, 4 respectively in each case. It is easy to evaluate $\mathcal{K}(X)$, e.g. Mathematica uses the command 'NullSpace[X]', and obtain L . The components L_1, L_2, L_3 and L_4 , corresponding to factors designated as 'columns', 'rows', 'Greek letters' and 'Latin letters' respectively, are obtained as:

$$\begin{aligned} L_1 &= \begin{pmatrix} 1_4 & 0_4 & 0_4 & 0_4 & \alpha_{11} \end{pmatrix}, & L_2 &= \begin{pmatrix} 0_4 & 1_4 & 0_4 & 0_4 & \alpha_{21} \end{pmatrix}, \\ L_3 &= \begin{pmatrix} 0_4 & 0_4 & 1_4 & 0_4 & \alpha_{31} \end{pmatrix}, & L_4 &= \begin{pmatrix} 0_4 & 0_4 & 0_4 & 1_4 & \alpha_{41} \end{pmatrix}, \end{aligned}$$

where the sub-vectors in the fifth columns are $\alpha_{11} = (-1 \ 1 \ 0 \ 0)'$, $\alpha_{21} = (-1 \ 0 \ 1 \ 0)'$, $\alpha_{31} = (-1 \ 0 \ 1 \ 0)'$ and $\alpha_{41} = (-1 \ 0 \ 0 \ 1)'$. Since L_1, L_2, L_3 , and L_4 each have rank 2, it follows from Theorem 3 (i) that none of 'columns', 'rows', 'Greek letters' or 'Latin letters' are singly connected. From Theorem 3 (ii) the estimable contrasts are given by segregated components $\mathcal{K}(L'_i)$ which are determined easily. To display these spaces, let K_i be a matrix whose columns comprise a basis for $\mathcal{K}(L'_i)$; it is found that

$$K_1 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & -2 \\ -1 & 0 \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & 1 \\ 1 & -2 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad K_3 = \begin{pmatrix} 0 & 1 \\ 1 & -2 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad K_4 = \begin{pmatrix} 0 & 1 \\ 1 & -2 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is clear that each factor has only one estimable pairwise contrast. In particular the estimable pairwise column contrast is found from K_1 to be column 3 – column 4. Similarly, the estimable pairwise row contrast is row 2 – row 4, the estimable pairwise 'Greek letter' contrast is $\beta - \delta$ and the estimable pairwise 'Latin letter' contrast is $B - C$.

Remark Birkes *et al.* (1976) use a search procedure, termed the Q -process, to identify a system of 7 independent loops which give a spanning set for combined contrasts involving Latin and Greek letters only. This process is equivalent to finding a spanning set for the subspace $\mathcal{K}((L'_3 \ L'_4))$. From the 7 contrasts, the authors obtained separate spanning sets for Latin letter contrasts and for Greek letter contrasts using a ten-step procedure described in Dodge (1985, §7.5). A repeat of this two-stage method for the dual of the design gives spanning sets for row contrasts and for column contrasts. This method provides an interesting, if laborious approach to deriving sets of basis vectors for $\mathcal{K}(L'_i)$ ($i = 1, 2, 3, 4$) but it does not yield the value of d . In general, the method of Birkes *et al.* (1976) seems to have rather a limited value if either m or n is relatively large.

It is interesting that Theorem 3 can be expressed in equivalent form that generalizes the familiar information criterion for block and row-column designs, which is described, for example, by John and Williams (1995, pp. 12 and 90). Suppose that

$$X_{(i)} = (1_n \ X_1 \ \dots \ X_{i-1} \ X_{i+1} \ \dots \ X_m), \quad \text{for } i = 1, \dots, m, \quad (3.6)$$

let $P_i = X_i(X'_i X_i)^{-1} X'_i$, $P_{(i)} = X_{(i)}(X'_{(i)} X_{(i)})^{-1} X'_{(i)}$ be orthogonal projection operators on the spaces $\mathcal{C}(X_i)$ and $\mathcal{C}(X_{(i)})$, and let the information matrix, C_i , for factor i be

$$C_i = X'_i(I - P_{(i)})X_i. \quad (3.7)$$

Then the generalized information matrix proposition is as follows:

Theorem 4 *Let C_i be the information matrix defined by (3.7). Then*

- (i) *The factor i is singly connected if and only if $\text{Rank}(C_i) = b_i - 1$;*
- (ii) *$\lambda'_i \beta_i$ is estimable if and only if $\lambda_i \in \mathcal{C}(C_i)$.*

Proof: Firstly, note that from (3.3) we have $X_i L_i = -X_{(i)} L_{(i)}$, where the matrix $L_{(i)}$ consists of the rows of L but excluding the rows of the component L_i . Since $\mathcal{C}(L)$ is the kernel of X then it is possible to find vectors u and v such that $X_i u = X_{(i)} v$ only if both $u \in \mathcal{C}(L_i)$ and $v \in \mathcal{C}(L_{(i)})$. We show that

$$\mathcal{C}(L_i) = \mathcal{K}(C_i). \quad (3.8)$$

Let u be a $b_i \times 1$ vector. If $u \in \mathcal{C}(L_i)$ then there is a $(m + d) \times 1$ vector ω such that $u = L_i \omega$ hence $X_i u = X_i L_i \omega = X_{(i)} \{-L_{(i)} \omega\}$, implying that $X_i u \in \mathcal{C}(X_{(i)})$; conversely, if $X_i u \in \mathcal{C}(X_{(i)})$ then $X_i u = X_{(i)} v$ for some v which implies that $u \in \mathcal{C}(L_i)$ and thereby shows that $u \in \mathcal{C}(L_i) \Leftrightarrow X_i u \in \mathcal{C}(X_{(i)})$. But by definition, $X_i u \in \mathcal{C}(X_{(i)})$ if and only if $(I - P_{(i)}) X_i u = 0$ which occurs if and only if $C_i u = 0$ since $I - P_{(i)}$ is idempotent. This establishes the result (3.8), from which it follows that $\text{rank}(L_i) = 1$ if and only if $\text{rank}(C_i) = b_i - 1$. Thus Theorem 3 (i) is equivalent to Theorem 4 (i). Furthermore, (3.8) implies immediately that $\lambda \in \mathcal{K}(L'_i)$ if and only if $\lambda \in \mathcal{C}(C_i)$, i.e. Theorem 3 (ii) is equivalent to Theorem 4 (ii), which completes the proof. $\square$

Theorem 4 (i) is a generalization of the classic result of Chakrabarti (1963) who seems to be the first contributor to recognize that Bose's alternate factor i - j chain criterion for factor i to be singly connected can be stated in terms of independent estimable contrasts in that factor. Theorem 4 (ii) is a generalization of a result of Godolphin (2004).

Example 4: Eccleston and Russell four factor design

The 4-way design given by Eccleston and Russell (1977) has dimensions $2 \times 4 \times 4 \times 8$ and factors: 'layers', 'columns', 'rows' and 'treatments' respectively. The design is unusual in having many null entries and is displayed in Figure 3, with each level of 'layer' represented by a two dimensional array.

Fig. 3: Eccleston and Russell 4-way design

layer 1				layer 2			
1	2	*	*	5	6	*	*
2	1	*	*	6	5	*	*
*	*	3	4	*	*	7	8
*	*	4	3	*	*	8	7

In this example $m = 4$ and $b_1 = 2$, $b_2 = b_3 = 4$ and $b_4 = 8$, giving $r_c = 15$ from (2.2). The design matrix has dimension 16×19 ; however $r = 12$ or, equivalently, $\text{Dim}(\mathcal{K}(X)) = 7$, so the design is disconnected of degree $d = 3$. The four components of L that need to be considered are found to be:

$$\begin{aligned}
L_1 &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}; \\
L_2 &= \begin{pmatrix} 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}; \\
L_3 &= \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}; \\
L_4 &= \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned}$$

Since L_1 , L_2 , L_3 and L_4 each have rank 2 or more then no factor is singly connected by Theorem 3 (i). In particular, the treatments factor is interesting because L_4 has rank as high as 3. The λ -space of estimable treatment contrasts is the segregated component $\mathcal{K}(L'_4)$ and a basis for this space is given by the columns of K_4 :

$$K_4 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

This implies that only four of the 28 pairwise treatment contrasts are estimable.

Now consider Figure 4, which is a similar design to that displayed in Figure 3 in the sense that each of the treatments is again replicated twice but the allocation of treatments is different.

Fig. 4: Eccleston and Russell 4-way design with reallocation of treatments

layer 1				layer 2			
1	2	*	*	3	6	*	*
5	4	*	*	6	5	*	*
*	*	1	2	*	*	7	8
*	*	4	3	*	*	8	7

As before, $r_c = 15$ and the design matrix has dimension 16×19 but $r = 14$ and $\text{Dim}(\mathcal{K}(X)) = 5$, i.e. this new design with the treatments reallocated is disconnected of degree $d = 1$. The components of L are obtained as:

$$\begin{aligned} L_1 &= \begin{pmatrix} 1_2 & 0_2 & 0_2 & 0_2 & \alpha_{11} \end{pmatrix}, & L_2 &= \begin{pmatrix} 0_4 & 1_4 & 0_4 & 0_4 & \alpha_{21} \end{pmatrix}, \\ L_3 &= \begin{pmatrix} 0_4 & 0_4 & 1_4 & 0_4 & \alpha_{31} \end{pmatrix}, & L_4 &= \begin{pmatrix} 0_8 & 0_8 & 0_8 & 1_8 & \alpha_{41} \end{pmatrix}. \end{aligned}$$

where the sub-vectors α_{11} , α_{21} , α_{31} and α_{41} are given by:

$$\alpha_{11} = 0_2; \alpha_{21} = (1 \ 1 \ 0 \ 0)'; \alpha_{31} = (-1 \ -1 \ 0 \ 0)' \text{ and } \alpha_{41} = 0_8.$$

Therefore the effect of the treatment reallocation is that the layers and treatments factors are singly connected, which is evident since L_1 and L_4 each have rank 1. This implies in particular that all 28 pairwise treatment contrasts are estimable. On the other hand neither the column factor nor the row factor is singly connected since L_2 and L_3 each have rank 2, indeed the same row and column contrasts are estimable as those of the original Eccleston and Russell 4-way design.

4 Adjusted orthogonal designs connectivity criteria

For many years much research attention has concentrated on evaluating single-factor connectivity properties of designs for the m -way model by considering the properties of several smaller dimensional models where some factors are ignored. The motivation for this alternative approach stems from the ease of deciding whether treatments are connected for an incomplete block design ($m = 2$) compared to the perceived difficulty of deciding this for a row-column design ($m = 3$). Shah and Khatri (1973) showed that a disconnected row-column design could be treatment connected if rows are ignored and also if columns are ignored; Raghavarao and Federer (1975) and Eccleston and Russell (1975) specified classes of designs, referred to as the ‘ordinary class’ and the ‘adjusted orthogonal class’ respectively, where the Shah-Khatri phenomenon could not arise. A similar result for the case $m = 4$ was given by Katyal and Pal (1991).

Lemma 2 *Suppose that all factors are singly connected, except possibly the two factors i and j . Then either the design is completely connected or else there is $b_i \times 1$ vector α_i which is not a scalar multiple of 1_{b_i} such that $\alpha_i \in \mathcal{K}((I - P_j)X_i)$, in which case $X_i\alpha_i + X_j\alpha_j = 0$ where $\alpha_j = -(X_j'X_j)^{-1}X_j'X_i\alpha_i$.*

Proof: Either all factors are singly connected or neither factor i nor factor j is singly connected and there are vectors α_i, α_j which are not multiples of unit vectors or zero vectors such that $X_i\alpha_i + X_j\alpha_j = 0$. Premultiplying by $X_i'(I - P_j)$ gives

$$0 = X_i'(I - P_j)X_i\alpha_i = \{(I - P_j)X_i\}'(I - P_j)X_i\alpha_i \Rightarrow \alpha_i \in \mathcal{K}((I - P_j)X_i)$$

(Harville, 1997, Corollary 7.4.5). Therefore $X_i\alpha_i - X_j(X_j'X_j)^{-1}X_j'X_i\alpha_i = 0$, and the formula for α_j follows. $\square$

For simplicity, a design is said to be *factor-(i, j) disconnected* if the two factors i and j are not singly connected but the remaining $m - 2$ factors are singly connected. It is straightforward to detect this

property of the design since the levels of factor i can be partitioned into sets in such a way that the levels of factor j in one set are distinct from the levels of factor j in the other sets, and simple checking procedures, such as the P -process of Godolphin (2004), can be used.

Remark If the design is known to be disconnected of order one then the converse of Lemma 2 applies; i.e. if the dimension of the space $\mathcal{K}((I - P_j)X_i)$ is equal to two then there is a $b_i \times 1$ vector $\alpha_i \in \mathcal{K}((I - P_j)X_i)$ which is not a multiple of 1_{b_i} , implying that the design is factor- (i, j) disconnected and all factors except i and j are singly connected.

If $m \geq 3$ it is interesting to consider conditions on a design such that either factor i is singly connected or the design is factor- (i, j) disconnected. We require the following definition due to Eccleston and Russell (1975) which has been widely considered for multifactor designs; see for example Park and Dean (1990) and Preece (1996).

Definition 4.1 For any three factors i, j and k in a m -way design, factor i and factor j are said to be orthogonal when adjusted for factor k if

$$X_i'(I - P_k)X_j = 0.$$

Theorem 5 Let factor i and factor j be orthogonal when adjusted for factor k for each $j = 1, \dots, m$ except for the cases where $j = i$ and $j = k$. Then either factor i is singly connected or the design is factor- (i, k) disconnected.

Proof: The standard restrictions (2.3) and any additional restrictions (3.1) can be expressed as

$$X_1\alpha_1 + X_2\alpha_2 + \dots + X_m\alpha_m = 0. \quad (4.1)$$

for suitable vectors $\alpha_1, \alpha_2, \dots, \alpha_m$. Factor i is singly connected if α_i is necessarily a multiple of the unit vector, including the zero vector, in any equation of the form (4.1). Otherwise, an equation (4.1) exists such that α_i is not a multiple of the unit vector. Premultiplying equation (4.1) by $X_i'(I - P_k)$ and using Definition 4.1 yields the result $X_i'(I - P_k)X_i\alpha_i = 0$ and then the proof of the theorem follows from Lemma 2. $\square$

The following corollary to Theorem 5 gives well known criteria for the connectivity of 3-way designs which have appeared in the literature; see Raghavarao and Federer (1975), Eccleston and Russell (1975) and Ceranka and Kozłowska (1991).

Corollary 3 For a 3-way design, if factors 1 and 2 are orthogonal when adjusted for factor 3 then

- (i) factor 1 is singly connected or the design is factor- $(1, 3)$ disconnected;
- (ii) the design is totally connected if and only if it is not factor- $(1, 3)$ disconnected nor factor- $(2, 3)$ disconnected.

A similar result can be expressed when $m = 4$. In this case it is evident that the following consequence of Theorem 5 has been established.

Corollary 4 For a 4-way design, let factors 1 and 2 be orthogonal, let factors 1 and 3 be orthogonal and let factors 2 and 3 be orthogonal, all when adjusted for factor 4. Then the design is completely connected provided that it is not factor- $(1, 4)$ disconnected, nor factor- $(2, 4)$ disconnected, nor factor- $(3, 4)$ disconnected.

The following result, which is weaker than Corollary 4, is given by Katyal and Pal (1991) for $m = 4$.

Corollary 5 For a 4-way design, let factors 1 and 2 be orthogonal, let factors 1 and 3 be orthogonal and let factors 2 and 3 be orthogonal, all when adjusted for factor 4. Then factor 4 is singly connected provided that the design is not factor- $(1, 4)$ disconnected, nor factor- $(2, 4)$ disconnected, nor factor- $(3, 4)$ disconnected.

References

- Bérubé, J. and Styan, G.P.H., 1993. Decomposable three-way layouts. *Journal of Statistical Planning and Inference*, 36, 311-322.
- Birkes, D., Dodge, Y. and Seely, J., 1976. Spanning sets for estimable contrasts in classification models. *Annals of Statistics*, 4, 86-107.
- Bose, R.C., 1944. The fundamental theorem in linear estimation. *Proc 31st Ind. Sci. Congress* 2-3 (abstract).
- Butz, L., 1982. *Connectivity in Multi-Factor Designs*, Berlin: Heldermann Verlag.
- Ceranka, B. and Kozłowska, M., 1991. Connectedness of row and column designs. *Zastosowania Matematyki Applicationes Mathematicae*, 21, 27-31.
- Chakrabarti, M.C., 1963. On the C-matrix in design of experiments. *Journal of the Indian Statistical Association*, 1, 8-23.
- Christensen, R., 2002. *Plane Answers to Complex Questions*, 3rd edn. New York: Springer.
- Dodge, Y., 1985. *Analysis of Experiments with Missing Data*, New York: Wiley.
- Eccleston, J. and Hedayat, A.S., 1974. On the theory of connected designs: characterisation and optimality. *Annals of Statistics*, 2, 1238-1255.
- Eccleston, J. and Russell, K., 1975. Connectedness and orthogonality in multi-factor designs. *Biometrika*, 62, 341-345.
- Eccleston, J. and Russell, K., 1977. Adjusted orthogonality in non-orthogonal designs. *Biometrika*, 64, 339-345.
- Ghosh, S., 1986. On a new graphical method of determining the connectedness in three dimensional designs. *Sankya*, B 48, 207-215.
- Godolphin, J.D., 2004. Simple pilot procedures for the avoidance of disconnected experimental designs. *Applied Statistics*, 53, 133-147.
- Godolphin, J.D. and Godolphin, E.J., 2001. On the connectivity of row-column designs. *Utilitas Mathematica*, 60, 51-65.
- Harville, D.A., 1997. *Matrix Algebra from a Statistician's Perspective*, New York: Springer.
- John, J.A. and Williams, E.R., 1995. *Cyclic and Computer Generated Designs*, 2nd edn. London: Chapman and Hall.
- Katyal, V. and Pal, S., 1991. Analysis and connectedness of four-dimensional designs. *Journal of the Indian Society of Agricultural Statistics*, 43, 296-309.
- Park, D.K. and Dean, A.M., 1990. Average efficiency factors and adjusted orthogonality in multidimensional designs. *Journal of the Royal Statistical Society, Series B*, 52, 361-368.
- Park, D.K. and Shah, K.R., 1995. On the connectedness of row-column designs, *Communications in Statistics*, 24, 87-96.
- Preece, D.A., 1996. Multifactor balanced designs with complete adjusted orthogonality for all pairs of treatment factors. *Australian Journal of Statistics*, 38, 223-230.
- Raghavarao, D. and Federer, W.T., 1975. On connectedness in two-way elimination of heterogeneity designs. *Annals of Statistics*, 2, 730-735.
- Searle, S.R., 1987. *Linear Models for Unbalanced Data*, New York: Wiley.
- Sengupta, D. and Jammalamadaka, S.R., 2003. *Linear Models : an Integrated Approach*, London: World Scientific.
- Shah, K.R. and Dodge, Y., 1977. On the connectedness of designs. *Sankya*, B 39, 284-287.

- Shah, K.R. and Khatri, C.G., 1973. Connectedness in row-column designs. *Communications in Statistics*, 2, 571-573.
- Srivastava, J.N. and Anderson, D.A., 1970. Some basic properties of multidimensional partially balanced designs. *Annals of Mathematical Statistics*, 41, 1438-1445.
- Wynn, H.P., 2008. Algebraic solutions to the connectivity problem for m -way layouts: interaction-contrast aliasing. *Journal of Statistical Planning and Inference*, 138, 259-271.