An Algorithmic Approach to Constructing Supersaturated Designs

Nam-Ky NGUYEN
CSIRO-IAPP Biometrics Unit
Private Bag 10, Rosebank MDC
Clayton, Victoria 3169
Australia

Supersaturated designs are very cost-effective to scientists and engineers at the primary stage of scientific investigation. This article describes a method of constructing supersaturated designs from balanced incomplete block designs that is a generalization of the method of Lin for constructing these designs and a more general approach to constructing these designs.

KEY WORDS: Computer-aided designs; Cyclic incomplete block designs; Interchange algorithm; Near-orthogonal array; Saturated designs; Screening designs.

Because the main objective of a screening experiment is to identify a few significant factors for further studies, scientists and engineers require designs with the minimum number of runs. Many saturated designs (designs with the number of factors \( m \) equal to \( n - 1 \), where \( n \) is the number of runs) have proved useful for this purpose. There are situations, however, in which scientists and engineers cannot even afford the number of runs required for these designs.

Consider an example in which a car manufacturer is conducting a passenger-impact crash test on a planned new four-wheel-drive (4WD) range. The objective is to find a subset of 54 safety features such as modified airbags, bullbar, bonded windscreen, (twin front) crush cans, and so forth to be included in the new car's total safety system. A suitable design for this test is a Hadamard matrix of order 56 that requires 56 runs (car prototypes). The question is what type of design is to be used when the research and development of the car manufacturer allows at most half of the number of required cars for this test.

Designs suitable for this example are called supersaturated designs. These designs were introduced by Booth and Cox (1962) and were recently studied further by Lin (1993a) and Wu (1993) (see also Satterthwaite 1959). These designs are very cost-effective with respect to the number of runs and as such are highly desirable in the context of industrial experimentation. This article describes a method of constructing supersaturated designs from balanced incomplete block designs (BIBD's), and a more general approach to constructing these designs.

1. CRITERIA FOR COMPARING SUPERSATURATED DESIGNS

Let \( X \) be an \( n \times m \) design matrix of a design with \( n \) runs (rows) and \( m \) two-level factors (columns) each with \( \frac{1}{2}n \) of +1's or high-level values and \( \frac{1}{2}n \) of -1's or low-level values (\( m \geq n - 1 \)). Let \( s_{ij} \) be the element in the \( i \)th row and \( j \)th column of \( X'X \). Booth and Cox (1962) proposed as a criterion for comparing designs the minimization of \( \text{ave}(s^2) \), where \( \text{ave}(s^2) = \sum_{i<j} s_{ij}^2 / \binom{m}{2} \). Clearly for orthogonal designs \( \text{ave}(s^2) = 0 \).

The rationale of the Booth-Cox criterion can be explained by using the singular value decomposition to decompose \( X \) as \( U \Lambda^{1/2} V' \), where matrices \( U \) and \( V \) are orthogonal and \( \Lambda \) is diagonal. It can then be shown that \( X'X \) and \( XX' \) share the same set of nonzero eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_r \), where \( r = \text{rank}(X'X) = \text{rank}(XX') \). Moreover, \( \text{tr}(X'X) = \text{tr}(XX') = \sum \lambda_i = mn = \text{const.} \) and \( \text{tr}(X'X^2) = \text{tr}(XX')^2 = \sum \lambda_i^2 \). Thus minimizing \( \sum_{i<j} s_{ij}^2 \), which is equivalent to minimizing \( \text{tr}(X'X^2) \), is the same as making the \( \lambda_i \)'s as equal as possible with \( \sum \lambda_i = \text{const.} \). This in a sense is an approximation of the D-optimality criterion, which requires the minimization of \( \sum \lambda_i^{-1} \); or the D-optimality criterion, which requires the maximization of \( \text{det}(\Lambda) \) (see Kiefer 1959).

Because the sum of each column of \( X \) is 0, the sum of the elements of \( XX' \) is 0—that is, the sum of the off-diagonal elements of \( XX' \) equal to \( -nm \) (\( nm \) is the sum of the diagonal elements of \( XX' \)). Thus, the sum of squares of the elements of \( XX' \) (and \( X'X \)) will reach the minimum if \( XX' \) is of the form \( (m-n)I_n + xJ_n \), where \( x = -m/(n-1) \) [assuming that \( m \) is divisible by \( n-1 \)], \( I_n \) is the identity matrix, and \( J_n \) is the \( n \times n \) matrix of 1's. In this case \( \text{ave}(s^2) = n(m^2 + (n-1)x^2 - mn)/(m(m-1)) = n^2(m-n+1)/(n-1)(m-1) \). This quantity can be used as a lower bound for \( \text{ave}(s^2) \) when \( m \) is divisible by \( n-1 \). Note that for \( m = n-1 \) this quantity becomes 0, and for \( m = 2(n-1) \) this quantity becomes \( n^2/(2n-3) \).

Another reasonable criterion for comparing supersaturated designs is to minimize the frequency of \( s_{ij} = \pm s_{\text{max}} \), where \( s_{\text{max}} = \max |s_{ij}| \). This criterion and the \( \text{ave}(s^2) \) criterion typically agree on which of the two designs is better. There are, however, examples that show that these two cri-
teria can lead to different designs. Consider the following candidate designs for \((n, m) = (24, 30)\). Design (A) has 241 \(s_{ij} = 0\), 187 \(s_{ij} = f^4\), seven \(a_{ij} = \cdot t^8\), and \(\text{ave}(s^2) = 7.91\). Design (B) has 198 \(s_{ij} = 0\), 237 \(s_{ij} = f^4\), and \(\text{ave}(s^2) = 8.72\). Design (B) is not necessarily better than (A) because it has cleared only seven \(s_{ij} = f^8\) at the cost of having 43 additional nonorthogonal pairs of columns. It is, however, preferable for experimenters looking for designs with a prespecified small \(s_{\text{max}}\). Designs with a prespecified \(s_{\text{max}}\) were considered by Lin (1995). In this article, unless mentioned otherwise, the popular \(\text{ave}(s^2)\) criterion will be used.

### Table 1

Lin (1993a) provided a very simple method of constructing supersaturated designs of size \((n, m) = (8, 14)\) constructed from a cyclic BIBD of the preceding series with \(t = 4\) and two initial blocks (2 3 7) and (2 3 5). Because of the cyclic nature of this supersaturated design, it is possible to generate \(X\) given just the first and the eighth columns of \(X\), which correspond to the two initial blocks.

1. BIBD-based designs have \(XX'\) matrix of the form \((m + 2)I_n - 2J_n\). These designs, with \(\text{ave}(s^2) = n^2/(2n - 3)\) or approximately \(\frac{1}{2}n\) when \(n\) is large enough, are \(\text{ave}(s^2)\) optimal.

2. Deleting a column of a supersaturated design results in deleting the corresponding row and column of the \(X'X\) matrix of this design. Because the sum of squares of each row (or column) of the \(X'X\) matrix of BIBD-based designs equals \(n(m + 2) = 2n^2\), deleting a column of \(X\) results in a design with the same \(\text{ave}(s^2)\) (and \(s_{\text{max}}\)). It is not difficult to show that designs obtained by deleting a column from (or adding a column to) a BIBD-based design are \(\text{ave}(s^2)\) optimal.

3. In general, when not all \(m\) columns of a BIBD-based design are to be used, deleting two or more columns of this design might not result in a good design. A general algorithm to construct designs for such cases will therefore be considered in Section 3.
Table 1. Generating Vectors of RBD-Based Supersaturated Designs of Size \((n, m) = (2t, 4t - 2), 3 \leq t \leq 15\)

<table>
<thead>
<tr>
<th>(t)</th>
<th>(n)</th>
<th>(m)</th>
<th>Generating vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6</td>
<td>10</td>
<td>((+ - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- + - + -))</td>
</tr>
<tr>
<td>4*</td>
<td>8</td>
<td>14</td>
<td>((- + - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- + - + - -))</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>18</td>
<td>((+ + - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- + + - - + -))</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>22</td>
<td>((- + + - - - - -))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((+ - - - + - - - +))</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>26</td>
<td>((- + + + - - - - -))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((+ + + - - - - - +))</td>
</tr>
<tr>
<td>8*</td>
<td>16</td>
<td>30</td>
<td>((- - + + + - - - - -))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- - - - + - - - - +))</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>34</td>
<td>((- - - - - + - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((+ + + + - - - - - +))</td>
</tr>
<tr>
<td>10*</td>
<td>20</td>
<td>38</td>
<td>((- + - + - - - - - - -))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- + - + - - - - - - +))</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>42</td>
<td>((+ + + + - - - - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- + + + - - - - - - - +))</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>46</td>
<td>((+ + + + - - - - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- + - - - - - - - - - +))</td>
</tr>
<tr>
<td>13*</td>
<td>26</td>
<td>50</td>
<td>((- - + - - - - - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- - - - - - - - - - - +))</td>
</tr>
<tr>
<td>14*</td>
<td>28</td>
<td>54</td>
<td>((- + - - - - - - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- - + - - - - - - - - +))</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>50</td>
<td>((+ + + + - - - - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((+ + - - - - - - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((+ - - - - - - - - - - +))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>((- - - - - - - - - - - +))</td>
</tr>
</tbody>
</table>

\(^*\) \(t\) value associated with new design.

4. Different generating arrays (obtained from different cyclic BIBD’s) might result in designs with the same \(\text{ave}(\hat{s}^2)\) but different \(s_{\text{max}}\). For example, for \((n, m) = (26, 50)\), the \(s_{\text{max}}\) of BIBD-based designs can be 6, 8, 10, and so forth.

3. A GENERAL ALGORITHM

A supersaturated design can be considered as a near-orthogonal array (NOA) with columns at two levels (Nguyen in press). Before describing the NOA algorithm, I will present some matrix results. Without loss of generality, let the \(i\)th and \(u\)th rows of \(X\) be two row vectors of the form \((+\mathbf{i}')\) and \((-\mathbf{u}')\), where \(\mathbf{i}'\) and \(\mathbf{u}'\) are two \(1 \times (m - 1)\) row vectors. It is not difficult to show that the effect on \(X'X\) obtained by swapping of the signs of the first elements of these two rows of \(X\) is the same as adding the following matrix to the \(X'X\) matrix:

\[
\begin{pmatrix}
0 & 2(\mathbf{u}' - \mathbf{i}') \\
2(\mathbf{u} - i) & 0_{m-1}
\end{pmatrix}
\]

(1)

where \(0_{m-1}\) is the \((m - 1) \times (m - 1)\) matrix of 0's.

The NOA algorithm based on the preceding matrix results has two steps:

1. Construct a starting design by allocating randomly half of the entries of each column of \(X\) to \(+1\) and half to \(-1\). Form \(X'X\) and calculate \(f = \sum_{i<j} s_{ij}^2\).

2. For column \(j\) of \(X (j - 1, 2, \ldots, m)\) repeat searching a pair of \(i\)th and \(u\)th elements having different signs in this column such that the swap of these two elements will result in the biggest reduction in \(f\). If the search is successful, update \(f, X,\) and \(X'X\) using (1). If \(f\) cannot be reduced further, go to the next column.

Step 2 is repeated until \(f = 0\) or \(f\) reaches its lower bound (when \(m\) is divisible by \(n - 1\)) or \(f\) cannot be reduced by any further sign-swaps.

The NOA algorithm is a typical example of an interchange algorithm. Other examples of this type of algorithm in different design settings were discussed by Nguyen and Williams (1993) and Nguyen (1994). The algorithms of Booth and Cox (1962) and Lin (1995) for constructing supersaturated designs and of Lin (1993b) for constructing saturated designs are examples of exchange algorithms (see Nguyen and Miller 1992). In this class of algorithms, a column of \(X\) is replaced by an entirely new column from the candidate list.
Remarks

1. Because $X'X$ is symmetric, the NOA algorithm only needs to work with the upper diagonal elements.

2. To calculate the change in $f$ and update $f$ in Step 2, note that only the nonzero elements of the vector $2(u' - v')$ will affect the increase (or decrease) of a corresponding element of $X'X$.

3. Among several designs generated by the NOA algorithm with same $f$ for ave($s^2$)] but with different $s_{max}$'s, the one with the smallest $s_{max}$ is chosen.

4. To replace the ave($s^2$) criterion by the $s_{max}$ criterion, $f$ in the preceding algorithm is replaced by $f_{max}$, the frequency of $s_{ij} = \pm s_{max}$.

4. COMPARISON WITH OTHER DESIGNS

For designs with $(n,m) = (2t,4t - 2)$ and $t = 3, 5, 6, 7, 9, 11, 12, \text{and } 15$, the HFHM-based supersaturated designs of Lin (1993a) and my BIBD based designs have the same value of ave($s^2$) and $r_{max}$ (the maximum correlation in terms of the absolute value between two columns of $X$ calculated as $s_{max}/n$). For $t = 4, 8, 10, 13, \text{and } 14$, my BIBD-based designs are new. These designs were obtained from the generating vectors in Table 1. They are ave($s^2$) optimal (Table 2). As mentioned in Section 2, deleting a column of these designs does not change ave($s^2$). The design for $(n,m) = (24,30)$ is an alternative design [design (B) in Sec. 1], obtained by the $s_{max}$ criterion, has ave($s^2$) = 8.72 and $r_{max} = .166$. My designs improve those of Booth and Cox (1962), Lin (1993a), and Wu (1993) not only with respect to ave($s^2$) but also with respect to $r_{max}$. For example, for $(n,m) = (12,16)$ and $(12,24)$, although the designs of Wu (1993) have 45 and 141 $s_{ij} = \pm 4$'s, respectively, my designs have only 34 $s_{ij} = \pm 4$'s for the former and 135 $s_{ij} = \pm 4$'s for the latter, where 4 is $s_{max}$ of these designs. For $(n,m) = (24,30)$, although the design of Wu (1993) has 63 $s_{ij} = \pm 8$'s, my corresponding design [design (B) in Sec. 1] has $s_{max} = 4$.

Note that, in the passenger-impact crash test in the Introduction, the airbag (first factor) explodes and starts to deflate within the spell of an eyeblink. If the car manufacturer engineers suspect that the bull-bar (second factor) distorts this tuning (because a 4WD has an inherently rigid chassis structure), using my design for $(n,m) = (28,54)$, although the interaction between these two factors can be tested. In Wu's design, this interaction is fully aliased with the 28th factor.

5. CONCLUDING REMARKS

Although it is beyond the scope of this article to compare the NOA algorithm with the one of Lin (1993b) for constructing saturated designs, it is worth mentioning some saturated designs constructed by NOA that improve on Lin's designs. For $n = 17$, my saturated design has ave($s^2$) = 1.94 as compared to 2.06 of Lin's corresponding design. Although Lin's design has 6 $s_{ij} = \pm 5$, mine has $s_{max} = 3$. The following is my saturated design for $n = 22$ with ave($s^2$) = 3.64 as compared to 4.33 of Lin's corresponding design. Although Lin's design has 6 $s_{ij} = \pm 6$, mine has $s_{max} = 2$.

Note that if the condition of equal occurrence of +1's and -1's for the entries in each column is re-
laxed, the Kronecker product of a Hadamard of order 2 and a saturated design for \( n = 11 \) and \( \text{ave}(s^2) = 1 \) will produce a design saturated for \( n = 22 \) with \( \text{ave}(s^2) = 1.90 \). However, 10 columns of this design has 12 entries that equal -1 but only 10 entries that equal +1.

An additional application of the NOA algorithm is to augment an existing supersaturated design with additional two-level columns. In the passenger-impact crash test in the Introduction, if the engineers decide to include the 55th safety feature, say computerized seat belt, in the test, they can augment the design for \( (n, m) = (28, 54) \) in Table 2 with an additional two-level column to obtain a design for \( (n, m) = (28, 55) \) with \( \text{ave}(s^2) = 15.31 \) and \( s_{\max} = .285 \). The extension of this idea to construct orthogonal and near-orthogonal arrays with mixed levels was discussed by Nguyen (in press).

The running time of the NOA algorithm varies with \( m \) and \( n \). For design of size \( (n, m) = (12, 66) \), a solution with \( \text{ave}(s^2) = 11.08 \) (optimal) and \( s_{\max} = 4 \) is obtained in 25 out of 100 tries. The average time per try for this combination is about four seconds on a 66 MHz 486DX2 PC. Naturally, NOA cannot improve \( \text{ave}(s^2) \) of the BIBD-type designs. The biggest Hadamard matrix NOA can construct is of order 20.

Data from an experiment using supersaturated designs can be analyzed by stepwise selection or subset selection procedure (e.g., see Miller 1990). Examples of this type of analysis were given by Lin (1993a, 1995).

The NOA algorithm is implemented in a PASCAL program with the same name. Please contact me at namky@forprod.csiro.au regarding the availability of this program and the CID program that I used to obtain cyclic BIBD solutions in Section 2.

---

ACKNOWLEDGMENTS

I am grateful to the associate editor, the referees, and A. J. Miller for valuable comments and suggestions that led to substantial improvement of the article.

[Received June 1994. Revised June 1995.]

REFERENCES


Miller, A. J. (1990), Subset Selection in Regression, New York: Chapman & Hall.


Satterthwaite, F. (1959), "Random Balance Experimentation" (with discussion), Technometrics, 1, 111-137.


---

TECHNOMETRICS, FEBRUARY 1996, VOL. 38, NO. 1