3

Blocking with Independent Responses

John P. Morgan

CONTENTS
3.1 Blocking: The Basics ........................................................................... 99
3.2 Cost of Blocking ............................................................................... 105
3.3 Universal Optimality and BIBDs.............................................................. 110
3.4 Block Designs That Approximate BIBD Structure ....................................... 113
  3.4.1 Regular Graph Designs .................................................................. 114
  3.4.2 Group Divisible Designs ................................................................ 116
  3.4.3 Nearly Balanced Incomplete Block Designs ....................................... 122
  3.4.4 E-Optimal Block Designs ............................................................ 125
3.5 Small Block Designs .......................................................................... 129
  3.5.1 Block Size Two ........................................................................ 129
  3.5.2 Two Blocks ............................................................................. 132
  3.5.3 Minimally Connected Designs..................................................... 133
3.6 Multiple Blocking Factors ................................................................... 134
  3.6.1 Resolvable Block Designs ........................................................... 136
    3.6.1.1 Affine Resolvable Designs ................................................. 137
    3.6.1.2 Other Resolvable Designs .................................................. 138
  3.6.2 Other Nested Block Designs ........................................................ 139
  3.6.3 Row–Column Designs ............................................................... 139
  3.6.4 Designs with Nesting and Crossing ............................................... 142
    3.6.4.1 Semi-Latin Squares ......................................................... 143
    3.6.4.2 Nested Row–Column Designs ........................................... 144
3.7 Control Treatments and Weighted Optimality ........................................... 145
  3.7.1 Design for TviC Experiments ........................................................ 148
  3.7.2 Lesser Weight on the Control ....................................................... 151
3.8 Discussion and Directions ................................................................... 151
References............................................................................................ 153

3.1 Blocking: The Basics
Dean and Voss (1999) describe an experiment to assess the impact of visual context on human subjects’ ability to reproduce a straight line. The experimental procedure has a subject look at a picture of a 5 cm straight line, then draw freehand a line of the same length on a sheet of paper. The recorded response is the length of the segment drawn. This basic procedure is repeated six times for each subject, with each repetition drawn on a
separate sheet. Visual context is varied with the six pictures, each having one of six, distinct, preprinted borders. The goal of the experiment is to understand differences, if any, in responses, depending on the six bordering treatments. That understanding is hampered in so far as responses vary for reasons having nothing to do with the treatments.

Variability is apparent in the line-segment experiment in the fact that very few, if any, individuals can consistently draw straight lines of precisely the same length even under identical conditions. Observe, however, that the lengths of lines drawn by the same person are likely to vary less than the lengths of lines drawn by different people. That is, for all of the reasons that humans vary in their drawing skill, subject-to-subject variability will be greater than within-subject variability, and possibly much greater. This suggests that it would be advantageous, if possible, to isolate the larger subject-to-subject variability from other, intrinsic sources of variability in response, including that due to visual contexts, so that context effects become more readily discernible. This is the purpose of blocking: an often powerful technique for isolating a source of variability in experimental material so that it does not interfere with inferences on comparisons among treatments. As discussed in Chapter 1, blocking was originally introduced and developed as an experimental principle by R. A. Fisher and F. Yates; see particularly Sections 1.3 and 1.6.

Formally, a block is a set of experimental units sharing a common value of some characteristic thought to represent a potentially major source of variation in the response. A set of experimental units is blocked if they have been arranged into disjoint blocks. A block design is a choice of treatments to be used in each block, followed by a random assignment of those treatments to the units in each block. It is quite common, though not required, to have the same number of experimental units and thus to use the same number of treatments, in each block. Only equisized blocks are considered in this chapter.

In the line-segment experiment described earlier, fourteen subjects were recruited. Each was presented with six line pictures, one for each of the six border treatments in a random order, and asked to draw the line on a blank paper. The blocks are the fourteen subjects, and the units within the blocks are the six presentations of pictures. Using $b$ for the number of blocks, $v$ for the number of treatments, and $k$ for the number of units in a block, this experiment has $v = k = 6$ and $b = 14$. The name of this particular design is randomized complete block design (RCBD). The blocks are said to be “complete” because each comprises a single replicate of the $v$ treatments (no treatment left out of the block and none repeated).

Blocking is a design choice. Alternatively, the line-segment experiment could have been run with a completely randomized design (CRD). With fourteen subjects and six presentations to each, there are a total of $14 \times 6 = 84$ experimental units. A CRD would start by preparing fourteen pictures with each border, then assigning them at random to the 84 units, where “at random” means that each of the $84! / (14!)^6$ distinct assignments is equally likely. The CRD is attempting through the randomization to distribute, without distinction, all nuisance sources of variation across the treatments, paying no special attention to subject to subject or any other particular source of variability. Blocking, on the other hand, serves to isolate one of those sources that, pre-experimentation, we have identified as potentially major, so that it does not affect our estimation of treatment differences. For the line-segment experiment, the separate random assignments of treatments to units within each subject (block) keep the major subject-to-subject variability separated from treatments while seeking to spread within-subject variability equally across the treatments. This is reflected in the analysis of variance (ANOVA) skeletons presented in Table 3.1. The sum of squares for treatments is the same in both ANOVAs. What changes is the removal of variability due to blocks from the error sum of squares. The $F$-test for comparing treatments, as well as all estimation for comparisons of treatments, is subject to the variability estimated
by MSE. Reducing that variability increases the power of the test and the precision of those estimates. This is the primary purpose of blocking. For further exposition on the important role played by randomization, see Chapter 1. The randomization employed is a primary characteristic distinguishing block designs from CRDs.

The probability distribution employed to randomize treatment assignment within blocks also generates a statistical model for the responses; see Hinkelmann and Kempthorne (2008) for an introduction to and development of this theory. For our purposes, it is sufficient to simply state the model

\[ y_{ju} = \mu + \tau_{d[j,u]} + \beta_j + \epsilon_{ju}. \]  

(3.1)

where, \( y_{ju} \) is the response from unit \( u \) in block \( j \), and the design function \( d[j,u] \) identifies which of the \( v \) treatments have been assigned to that unit. The response is written in (3.1) as the sum of a general mean \( \mu \), an effect \( \tau_{d[j,u]} \) of the treatment applied to the unit in question, an effect \( \beta_j \) of the block in which the unit lies, and a random error term \( \epsilon_{ju} \) associated with that unit and its measurement. Though there are many other options, unless otherwise specified, the \( \epsilon_{ju} \) in this section are taken to be uncorrelated, mean zero random variables, with common variance \( \sigma^2_E \).

The block effects \( \beta_j \) serve to reduce the variability to which treatment comparisons are exposed, but are not the target of inference. That target is the set of treatment effects \( \tau_i \), with the design goal of maximizing precision of treatment comparisons. To see how this is done, rewrite (3.1) in matrix form as

\[ y = \mu 1 + X_d'\tau + Z\beta + \epsilon, \]

(3.2)

where for the total \( n \) of experimental units employed, \( X_d \) is the \( n \times v \) unit/treatment incidence matrix corresponding to the design function \( d[j,u] \), that is, \( X_d \) has a one in position \((u,j)\) if treatment \( j \) is assigned to unit \( u \); \( \tau = (\tau_1, \ldots, \tau_v)' \) is the \( v \times 1 \) vector of treatment effects; \( Z \) is the \( n \times b \) unit/block incidence matrix; and \( \beta = (\beta_1, \ldots, \beta_b)' \) is the \( b \times 1 \) vector of treatment effects. Write \( P_Z = Z(Z'Z)^{-1}Z' \) for the projector onto the column space of \( Z \). Then standard linear model theory (see Chapter 2) says that the least squares equations for estimation of \( \tau \) are \( X_d'(I-P_Z)X_d'\tau = X_d'(I-P_Z)y \). The coefficient matrix \( C_{d} = X_d'(I-P_Z)X_d \) is the information matrix for estimation of \( \tau \). It serves this role: if \( c'\tau \) is any estimable function of the treatment effects, then \( \text{Var}_d(c'\tau) = (c'C_{d}^{-1}c)\sigma^2_E \) for any generalized inverse (see Chapter 2) \( C_{d}^{-1} \) of \( C_{d} \). A good design \( d \) will be one that makes the variances \( (c'C_{d}^{-1}c)\sigma^2_E \) small in some overall sense (over all comparisons \( c \) of interest). The purpose of blocking the units was to minimize, in so far as possible, the error variance \( \sigma^2_E \). Given the blocks and
whatever they accomplish in variance reduction, it is the choice of which treatments are assigned within each block (i.e., the selection of \( d \)) that determines \( X_d \) and, hence, \( C_d \) and the quadratic forms \( c' C_d c \). Good choice of \( d \) will minimize, again in an overall sense, these quadratic forms.

Before proceeding, here are a few pertinent facts about \( C_d \). Obviously, \( C_d \) is symmetric, and since \((I - P_Z)\) is idempotent, \( C_d = X_d'(I - P_Z)X_d \) is nonnegative definite. As it is the coefficient matrix in the least squares equations, its row space consists of all vectors \( c \) such that \( c' \tau \) is estimable. Easily checked is \( C_d1 = 0 \), so this \( v \times v \) matrix has rank at most \( v - 1 \), and the estimability of \( c' \tau \) requires \( c'1 = 0 \). Hence, only contrasts of treatments are estimable, and all contrasts are estimable if, and only if, \( \text{rank}(C_d) = v - 1 \). Accordingly, the terms \( (c' C_d c) \sigma^2 \) targeted by the design problem are termed contrast variances.

A design \( d \) is said to be connected if its information matrix \( C_d \) has rank \( v - 1 \) and otherwise is disconnected. Because disconnected designs are unable to provide estimates of every treatment contrast, they are inappropriate unless a lower dimensional model for the treatment effects is in play, such as is employed with fractional factorial designs (Chapter 7). This chapter focuses on estimating all treatment contrasts, so considers only connected designs. Letting \( D = D(v, b, k) \) denote the class of all connected designs for \( v \) treatments in \( b \) blocks of \( k \) units per block, the goal is to determine a “best” design in \( D \).

Needed now is a way to operationalize the idea of “best.” As will be seen, there are many ways to do this, some more statistically meaningful than others and with varying relevance depending on experimenter goals. The basic idea is to select a function, let’s call it \( \Phi \), that maps the information matrix \( C_d \) to the real numbers \( \mathbb{R} \) in a manner that summarizes variances of contrast estimates. Smaller values of \( \Phi \) will correspond to designs that make variances smaller in the sense that \( \Phi \) measures. We begin with a general framework for functions of this type, then explore particular functions that are useful in a variety of situations.

Formally, \( \Phi \) is defined on the class \( C \) of all symmetric, nonnegative definite matrices of rank \( v - 1 \) with zero row sums. Let \( \Phi : C \to \mathbb{R} \). \( \Phi \) corresponds to a standard optimality criterion if

\[
\begin{align*}
(i) & \quad \Phi \text{ respects the nonnegative definite ordering: for any } C_1, C_2 \in C \text{ such that } C_1 - C_2 \text{ is nonnegative definite, } \Phi(C_1) \leq \Phi(C_2), \\
(ii) & \quad \Phi \text{ is invariant to treatment permutation: } \Phi(P C P') = \Phi(C) \text{ for each } C \in C \\
& \quad \text{and every permutation matrix } P_{v \times v}, \\
(iii) & \quad \Phi \text{ is convex: } \Phi(\alpha C_1 + (1 - \alpha)C_2) \leq \alpha \Phi(C_1) + (1 - \alpha)\Phi(C_2) \text{ for all } C_1, C_2 \in C. 
\end{align*}
\]

A design minimizing \( \Phi \) is said to be \( \Phi \)-optimal.

It can be shown (e.g., Morgan 2007b) that \( C_1 - C_2 \) is nonnegative definite, implying that \( c' C_1 c \leq c' C_2 c \) for all contrast vectors \( c \). Thus, property (3.3)(i) says that \( \Phi \) summarizes variances of estimators in a way that uniformly smaller variances produce smaller values of \( \Phi \), this being a minimal requirement if \( \Phi \) is to be statistically meaningful. Property (3.3)(ii) is equality of interest in all treatments, for permutation invariance says \( \Phi \) produces the same summary measure regardless of how the treatments are interchanged. The convexity property (3.3)(iii) is a technical requirement that is advantageous for optimality arguments and that, happily, turns out to be an attribute of all the commonly employed, statistically reasonable criteria.

The notion of equal treatment interest, incorporated through (3.3)(ii), is appropriate for many, but not all, experiments. One notable class of exceptions is experiments having a
control treatment, with differing emphasis on contrasts involving, and not involving, the control. Criteria that incorporate unequal treatment interest are taken up in Section 3.7.

Let $N_d$ be the $v \times b$ matrix whose $(i, j)$ entry is $n_{dij} = $ the number of units in block $j$ assigned treatment $i$ by design $d$. Further write $r_d = (r_{d1}, \ldots, r_{dvi})'$ where $r_{di} = \sum_j n_{dij}$ is the total number of units assigned treatment $i$, and let $R_d = \text{Diag}(r_d)$ be the diagonal matrix of the replication numbers $r_{di}$. Then the information matrix $C_d$ can be expressed as

$$C_d = R_d - \frac{1}{k} N_d N_d' .$$

(3.4)

Having selected the treatments to be used in any given block, they will be assigned to that block’s units with a random device. The specific realization of the randomization does not change (3.4) and so does not affect the quality of the information that the experiment will produce.

The $(i, i')$ entry of $N_d N_d'$ in (3.4), labeled $\lambda_{dii'} = \sum_j n_{dij} n_{dij'}$, is called a treatment concurrence number. For $i \neq i'$ this is the number of pairs of units within blocks receiving the pair of treatments $i, i'$. As $C_d$ is determined by the replication numbers $r_{di}$ and the concurrence numbers $\lambda_{dii'}$, the search for an optimal design can be understood through these counts. As we shall see, depending on the criterion selected, different values for these counts, that is, different designs, can be judged best. But first we need to define some specific criteria.

The inverses of the $v - 1$ positive eigenvalues $e_{d1} \leq e_{d2} \leq \cdots \leq e_{d, v-1}$ of $C_d$ are termed canonical variances; aside from $\sigma^2$, they are variances for the estimators of contrasts specified by their corresponding normalized eigenvectors. Many, though not all, of the commonly employed optimality criteria are functions of the canonical variances. Among these (see Table 3.2 and also Chapter 2) are the $A$, $E$, and $D$ criteria. Each has a particular statistical meaning, as follows. Up to a design-independent constant, $A$ measures (i) the average of the variances for all $v(v - 1)/2$ pairwise comparisons $\tau_i - \tau_i'$ and also (ii) the average of the variances when estimating any $v - 1$ orthonormal contrasts. The $E$-value for a design is proportional to the maximal estimator variance over all normalized contrasts, $\max_c \text{Var}_d (c' \tau) / c' c$. The $D$-value for a design is proportional, under a normality assumption, to the volume of the confidence ellipsoid for estimating any $v - 1$ orthonormal contrasts. Also displayed in Table 3.2 is $MV$, a standard criterion in the sense of (3.3) that is not solely a function of the eigenvalues. Like $E$, the $MV$ criterion seeks to minimize the impact of the worst case, but with respect to pairwise contrasts rather than all contrasts. Of the measures in Table 3.2, the $D$ criterion is the least popular for assessing block designs, though it is widely used in other design settings. Other interpretations of $A$ are revealed in Morgan and Stallings (2014).

**TABLE 3.2**

<table>
<thead>
<tr>
<th>Optimality Criteria</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$\Phi_A = \sum_{i=1}^{v-1} \frac{1}{e_{di}}$</td>
</tr>
<tr>
<td>$E$</td>
<td>$\Phi_E = \frac{1}{r_d}$</td>
</tr>
<tr>
<td>$D$</td>
<td>$\Phi_D = \prod_{i=1}^{v-1} \frac{1}{r_{di}}$</td>
</tr>
<tr>
<td>$MV$</td>
<td>$\Phi_{MV} = \max_{h \in H} h' C_d h$</td>
</tr>
</tbody>
</table>

*Note: $H$ is the collection of normalized contrast vectors with two nonzero coordinates.*
A, D, and E are special instances of the $\Phi_t$ family of criteria specified by minimizing $\Phi_t(C_{di}) = \left[ \sum_i e_{di}^t / (v - 1) \right]^{1/t}$: take $t = 1$, $t \to 0$, and $t \to \infty$, respectively. All $\Phi_t$ criteria are standard criteria in the sense of (3.3). While other values of $t$ do not admit simple statistical interpretations like those given earlier, it is interesting to see that the eigenvalue-based criteria in Table 3.2 are all part of the same continuum.

The criteria in Table 3.2 often, but not always, agree on what design is best. The three examples that follow illustrate several aspects of this situation. Each displays an incomplete block design (IBD), in which the block size $k$ is smaller than the number of treatments, and so a block cannot receive a complete replicate of the treatments. The optimality results stated in the examples will be justified later in this chapter. Here and throughout this chapter, block designs will be displayed with blocks as columns.

**Example 3.1**

This design, for $(v, b, k) = (6, 10, 3)$, is optimal with respect to all of the criteria in Table 3.2.

```
  1  2  3  4  5  1  2  3  4  5
  2  3  4  5  1  2  3  4  5  1
  6  6  6  6  6  4  5  1  2  3
```

In this design, all the replication numbers are equal ($r_{di} = 5$) and all the concurrence numbers are equal ($\lambda_{dii'} = 2, i \neq i'$).

**Example 3.2**

This design, for $(v, b, k) = (7, 28, 5)$, is $E$-optimal but not $A$-optimal.

```
  1  1  1  1  1  1  1  1  1  1  1
  2  2  2  2  2  2  2  2  2  3  3
  3  3  3  3  4  4  4  4  5  4  4
  4  4  5  5  5  5  6  6  6  5  5
  6  7  6  7  6  6  7  7  7  6  6
          1  1  1  1  1  1  1  1  2  2  2
  3  3  3  3  3  3  3  3  3  3  3
  4  4  4  4  5  5  4  4  4  5  5
  5  5  6  6  6  6  5  5  5  6  6
  7  7  7  7  7  7  7  7  7  7  7
```

In this design, all the replication numbers are equal ($r_{di} = 20$). However, the concurrence numbers are not all the same, ranging from 12 (e.g., $\lambda_{d12}$) to 14 (e.g., $\lambda_{d17}$).

**Example 3.3**

This design, for $(v, b, k) = (5, 7, 3)$, is $E$-optimal and $MV$-optimal but neither $A$-optimal nor $D$-optimal.

```
  1  1  1  1  2  2  2
  1  3  3  4  3  3  3
  2  4  5  5  4  5  5
```

This design has two different replication numbers ($r_{d1} = 5$, other $r_{di} = 4$), while all the concurrence numbers are equal ($\lambda_{dii'} = 2, i \neq i'$).
Example 3.3 is distinguished from both Examples 3.1 and 3.2 by having a treatment (in this case, treatment 1) assigned to more than one experimental unit within a block. Examples 3.1 and 3.2 have all $n_{dij} \in \{0,1\}$ and consequently are said to be *binary* block designs. Example 3.3 provides an example of a *nonbinary* block design. This and the other combinatorial features demonstrated by the preceding example designs will arise in the search for general properties that make for good block designs. Because the combinatorial structure (through $\tau_{dij}$'s and $\lambda_{dij}$'s) determines the information structure ($C_d$), much of the statistical theory for block designs is aimed at identifying design structures that can be proven optimal or, failing that, near optimal.

### 3.2 Cost of Blocking

Variance reduction through blocking does not come for free. One aspect of this is seen in Table 3.1, where the error degrees of freedom for the RCB is 13 fewer than for the CRD. If a block design uses $b$ blocks, then $b - 1$ degrees of freedom are “stolen” from error relative to the unblocked CRD, resulting in a less precise estimate of the remaining error variance $\sigma_E^2$. This impacts, for instance, widths of confidence intervals for treatment contrasts. Fortunately, this tends to be a small cost in all but the smallest designs (those providing few error degrees of freedom) and will be more than outweighed by the reduction in $\sigma_E^2$ achieved by a well-chosen blocking variable. There is a more substantial cost that is the focus of this section: treatment information in the data that are lost to blocks.

To frame the idea, look again at the design in Example 3.1. Consider the data contrast $y_{6.} - y_{1.}$, which is the difference in the totals for the sixth and first blocks. Were the units not blocked, $y_{6.} - y_{1.}$ would contribute to the estimate for $\tau_4 - \tau_6$. Because of the blocking, its expected value is $\tau_4 - \tau_6 + 3(\beta_6 - \beta_1)$, which is uninformative for $\tau_4 - \tau_6$. The treatment information in $y_{6.} - y_{1.}$, lost to the blocking, is said to be *confounded* with blocks. More generally, the expected value of an arbitrary data contrast $\sum_j c_j y_j$ of block totals contains the term $\sum_j c_j \beta_j$, so it cannot contribute to estimating treatment contrasts.

Starting with $y$ in model (3.2) and the blocks projection matrix $P_Z$, consider the transformed data $y^* = (I - P_Z)y$ having model $y^* = X^*_d \tau + e^*$. Least squares estimation for $\tau$ using the data $y^*$ leads to exactly the same information matrix (3.4) for $\tau$ as that found when starting with the original data $y$. Since $P_Z$ transforms $y$ to block averages, this shows that estimation of $\tau$ uses only data information that is orthogonal to the block totals. Any treatment information found in the $b$ block totals, and as explained in the preceding paragraph, is confounded with blocks, also said to *lie in the blocks stratum*. Confounding represents the primary cost of blocking: treatment information lost to blocks.

Confounding occurs whenever not every treatment is equally replicated in every block, so whenever incomplete blocks are used, but never with complete blocks. In a RCB, every block contains each treatment once, so no contrast of block totals contains treatment information. Aside from the small cost of lost error degrees of freedom, a RCB estimates treatment differences just as efficiently as a CRD even if achieving no reduction in $\sigma_E^2$. Efficiency gains with a RCB are typically substantial if the blocking actually delivers on its variance-reduction goal.

With an IBD, confounding means that less of the data are used in estimating treatment contrasts. Consequently, should blocking be ineffective and not reduce $\sigma_E^2$, an IBD can...
produce results that are markedly inferior to what would have been achieved without blocking at all. Prudent use of incomplete blocks requires understanding how much information is confounded and so the amount of variance reduction required to justify that use. This understanding can be gained through a modification of the treatment information matrix (3.4).

The efficiency matrix $C_{d,\text{eff}}$, here also shown in spectral form, is a transformed version of the information matrix

$$C_{d,\text{eff}} = R_d^{-1/2} C_d R_d^{-1/2} = \sum_{i=1}^{v-1} \varepsilon_{di} f_{di} f_{di}' .$$

The eigenvalues $\varepsilon_{di}$ are termed the (canonical) efficiency factors for design $d$. Obviously, efficiency factors are positive (for connected designs), and it can be shown that they obey an upper bound of 1. The eigenvectors $f_{di}$ define the basic contrast vectors $\varepsilon_{di} = R_d^{1/2} f_{di}$. Aside from the constant $\sigma_E^2$, the variance of the estimator for basic contrast $i$ is $\text{Var}_{CRD}(\varepsilon_{di} \tau) = 1/\varepsilon_{di}$.

The efficiency matrix is also defined for unblocked designs. For a CRD with the same replication numbers $r_d$ employed by the preceding IBD, the treatment information matrix is $R_d - 1/n r_d r_d'$ and so

$$C_{CRD,\text{eff}} = R_d^{-1/2} \left[ R_d - 1/n r_d r_d' \right] R_d^{-1/2} = I - 1/n R_d^{1/2} JR_d^{1/2} ,$$

where $J$ is the all-one matrix. The $f_{di}$ in (3.5) are also eigenvectors of (3.6), but with corresponding eigenvalues identically 1, for $C_{CRD,\text{eff}} f_{di} = f_{di} - 1/n R_d^{1/2} J f_{di} = f_{di} - 1/n R_d^{1/2} 0 = f_{di}$. Thus, and again aside from $\sigma_E^2$, the basic contrast $\varepsilon_{di} \tau$ for the IBD $d$ is estimated by a CRD having the same replication numbers as $d$ with variance $\text{Var}_{CRD}(\varepsilon_{di} \tau) = 1$. The ratio of this variance to that produced by $d$, assuming both designs produce the same value for $\sigma_E^2$ (blocking achieves no reduction), is

$$\text{Eff}_{d}(\varepsilon_{di} \tau) = \text{Var}_{CRD}(\varepsilon_{di} \tau)/\text{Var}_{d}(\varepsilon_{di} \tau) = \varepsilon_{di} .$$

It follows from (3.7) that efficiency factor $\varepsilon_{di}$ measures the proportion of unconfounded information available with $d$ on the basic contrast $\varepsilon_{di} \tau$. Said another way, $1 - \varepsilon_{di}$ is the proportional reduction in $\sigma_E^2$ needed for the IBD $d$ to estimate this contrast as well as an unblocked design. If $d$ reduces $\sigma_E^2$ by more than $100(1 - \varepsilon_{di})\%$, then it will estimate this contrast better than a CRD.

It was mentioned earlier that there is no confounding with a RCBD. This is equivalent to saying that for a RCBD, all efficiency factors are 1. Readers may wish to check that the efficiency matrix for a RCBD is identical to that in (3.6) for an equally replicated CRD.

The basic contrasts are linearly independent and so form a basis for all treatment contrasts. Ordering the efficiency factors $\varepsilon_{d1} \leq \varepsilon_{d2} \leq \cdots \leq \varepsilon_{d,v-1}$, it follows that the proportion of unconfounded information available on every contrast is at least $\varepsilon_{d1}$ and no contrast uses more than $100\varepsilon_{d,v-1}\%$ of the information in the unblocked design. A single, summary
measure of the fraction of available information is provided by the harmonic mean of the efficiency factors

\[
\frac{v - 1}{\sum_{i=1}^{v-1} \frac{1}{\varepsilon_{di}}}
\]

called the average efficiency factor. It is the inverse of the average variance of the basic contrast estimators. The average efficiency factor is a convenient measure for comparing designs with an eye to minimizing confounding. Keep in mind, however, that it is expressed relative to an unblocked design with the same replication vector. Both individual and average efficiency factors are relative measures, valid for design comparison only among designs having the same \(r_d\). Even then, full understanding of the average efficiency factor requires a broader context, to be explored in Section 3.7.

Examples of efficiency factors, calculated for the example designs in Section 3.1, are displayed in Table 3.3. The blocking in Example 3.2 must reduce error variance by about 7% if it is to perform as well, on average, as an unblocked design with 20 replicates. The design in Example 3.1 has all efficiency factors equal to 80%, so should error variance be reduced by more than 20%, every contrast will be estimated more efficiently than with an unblocked, 5-replicate competitor. Exact expressions for (3.8) as functions of \(v\) and other design parameters are known in a few cases; see Sections 1.6.1 and 1.9.2.

Many block designs used in practice are equally replicated, having all \(r_{di} = bk/v \equiv r\). In this case, efficiency factors are proportional to eigenvalues of the information matrix \(C_d\), \(\varepsilon_{di} = \varepsilon_{di}/r\), the average efficiency factor (3.8) is inversely proportional to the A-optimality value, and the smallest efficiency factor is inversely proportional to the E-optimality value (see Table 3.2). We will return to the relationship between efficiency factors and optimality criteria in Section 3.7, when the interpretation of (3.8) for the Example 3.3 design can also be given.

Table 3.4 lists efficiencies for the most efficient IBDs, for up to ten treatments and up to ten replicates, for those \((v, b, k)\) for which \(bk\) is a multiple of \(v\). Included is the average efficiency factor, termed the A-efficiency, and the lowest efficiency factor \(\varepsilon_{d1}\), termed the E-efficiency. Note that if blocking reduces variation by more than \((1 - \varepsilon_{d1})100\%\), the design will be better than a CRD for estimating every treatment contrast. How these best IBDs are determined is the topic of Sections 3.3 and 3.4.

Before ending this section, it should be mentioned that if block effects are random, \(\beta \sim (0, \sigma_B^2 I)\), then the blocks stratum information \(P_\beta y\) can contribute to treatment estimation. The procedure, known as analysis with recovery of interblock information, will be mentioned briefly in later sections. Owing to the additional variance term \(\sigma_B^2\), blocks stratum information is typically poor. That is why, as pursued in this chapter, blocking schemes are chosen to maximize the within-block information based on \(y^*\) and as captured in (3.4).

### Table 3.3

<table>
<thead>
<tr>
<th>Example</th>
<th>Eigenvalues (\varepsilon_{di})</th>
<th>Efficiency Factors (\varepsilon_{di})</th>
<th>Average Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>4, 4, 4, 4, 4</td>
<td>(.4)</td>
<td>(.4)</td>
</tr>
<tr>
<td>3.2</td>
<td>18, 4, 18, 4, 18, 4, 18, 4, 18</td>
<td>(.92)</td>
<td>(.933)</td>
</tr>
<tr>
<td>3.3</td>
<td>(.34, 3.1, 3.1, 3.1)</td>
<td>(.5)</td>
<td>(.796)</td>
</tr>
</tbody>
</table>

© 2015 by Taylor & Francis Group, LLC
### TABLE 3.4
Efficiencies for Best IBDs

<table>
<thead>
<tr>
<th>v</th>
<th>b</th>
<th>k</th>
<th>r</th>
<th>( A_{\text{eff}} )</th>
<th>( E_{\text{eff}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0.750</td>
<td>0.750</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>0.750</td>
<td>0.750</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>2</td>
<td>6</td>
<td>0.750</td>
<td>0.750</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>2</td>
<td>8</td>
<td>0.750</td>
<td>0.750</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>2</td>
<td>10</td>
<td>0.750</td>
<td>0.750</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0.600</td>
<td>0.500</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>0.667</td>
<td>0.667</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>0.643</td>
<td>0.500</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>0.662</td>
<td>0.600</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>0.667</td>
<td>0.667</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>2</td>
<td>7</td>
<td>0.659</td>
<td>0.571</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>2</td>
<td>8</td>
<td>0.662</td>
<td>0.625</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>2</td>
<td>9</td>
<td>0.667</td>
<td>0.667</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>2</td>
<td>10</td>
<td>0.663</td>
<td>0.600</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>3</td>
<td>3</td>
<td>0.889</td>
<td>0.889</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>3</td>
<td>6</td>
<td>0.889</td>
<td>0.889</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>0.500</td>
<td>0.345</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>0.625</td>
<td>0.625</td>
</tr>
</tbody>
</table>

(Continued)
| \(v\) | \(b\) | \(k\) | \(r\) | \(A_{\text{eff}}\) | \(E_{\text{eff}}\) | \(v\) | \(b\) | \(k\) | \(r\) | \(A_{\text{eff}}\) | \(E_{\text{eff}}\) | \(v\) | \(b\) | \(k\) | \(r\) | \(A_{\text{eff}}\) | \(E_{\text{eff}}\) | \(v\) | \(b\) | \(k\) | \(r\) | \(A_{\text{eff}}\) | \(E_{\text{eff}}\) |
| 5 | 15 | 2 | 6 | 0.611 | 0.532 | 7 | 35 | 2 | 10 | 0.577 | 0.510 | 9 | 36 | 2 | 8 | 0.563 | 0.563 | 10 | 20 | 4 | 8 | 0.831 | 0.781 |
| 5 | 20 | 2 | 8 | 0.625 | 0.625 | 7 | 7 | 3 | 3 | 0.778 | 0.778 | 9 | 45 | 2 | 10 | 0.555 | 0.473 | 10 | 25 | 4 | 10 | 0.832 | 0.800 |
| 5 | 25 | 2 | 10 | 0.620 | 0.569 | 7 | 14 | 3 | 6 | 0.778 | 0.778 | 9 | 6 | 3 | 2 | 0.667 | 0.500 | 10 | 4 | 5 | 2 | 0.850 | 0.600 |
| 5 | 5 | 3 | 3 | 0.815 | 0.709 | 7 | 21 | 3 | 9 | 0.778 | 0.778 | 9 | 9 | 3 | 3 | 0.727 | 0.667 | 10 | 6 | 5 | 3 | 0.878 | 0.800 |
| 5 | 10 | 3 | 6 | 0.833 | 0.833 | 7 | 7 | 4 | 4 | 0.875 | 0.875 | 9 | 12 | 3 | 4 | 0.750 | 0.750 | 10 | 8 | 5 | 4 | 0.883 | 0.800 |
| 5 | 15 | 3 | 9 | 0.831 | 0.792 | 7 | 14 | 4 | 8 | 0.875 | 0.875 | 9 | 15 | 3 | 5 | 0.740 | 0.631 | 10 | 10 | 5 | 5 | 0.886 | 0.815 |
| 5 | 5 | 4 | 4 | 0.938 | 0.938 | 7 | 7 | 5 | 5 | 0.931 | 0.870 | 9 | 18 | 3 | 6 | 0.742 | 0.667 | 10 | 12 | 5 | 6 | 0.886 | 0.833 |
| 5 | 10 | 4 | 8 | 0.938 | 0.938 | 7 | 14 | 5 | 10 | 0.933 | 0.904 | 9 | 21 | 3 | 7 | 0.745 | 0.714 | 10 | 14 | 5 | 7 | 0.888 | 0.857 |
| 6 | 6 | 2 | 2 | 0.429 | 0.250 | 7 | 7 | 6 | 6 | 0.972 | 0.972 | 9 | 24 | 3 | 8 | 0.750 | 0.750 | 10 | 16 | 5 | 8 | 0.887 | 0.836 |
| 6 | 9 | 2 | 3 | 0.556 | 0.500 | 8 | 8 | 2 | 2 | 0.333 | 0.146 | 9 | 27 | 3 | 9 | 0.747 | 0.684 | 10 | 16 | 5 | 8 | 0.887 | 0.850 |
| 6 | 12 | 2 | 4 | 0.577 | 0.500 | 8 | 12 | 2 | 3 | 0.488 | 0.333 | 9 | 30 | 3 | 10 | 0.747 | 0.700 | 10 | 18 | 5 | 9 | 0.889 | 0.889 |
| 6 | 15 | 2 | 5 | 0.600 | 0.600 | 8 | 16 | 2 | 4 | 0.538 | 0.500 | 9 | 9 | 4 | 4 | 0.834 | 0.750 | 10 | 20 | 5 | 10 | 0.888 | 0.855 |
| 6 | 18 | 2 | 6 | 0.588 | 0.500 | 8 | 20 | 2 | 5 | 0.545 | 0.438 | 9 | 18 | 4 | 8 | 0.844 | 0.844 | 10 | 20 | 5 | 10 | 0.888 | 0.860 |
| 6 | 21 | 2 | 7 | 0.587 | 0.500 | 8 | 24 | 2 | 6 | 0.560 | 0.500 | 9 | 9 | 5 | 5 | 0.896 | 0.840 | 10 | 5 | 6 | 3 | 0.918 | 0.833 |
| 6 | 24 | 2 | 8 | 0.592 | 0.563 | 8 | 28 | 2 | 7 | 0.571 | 0.571 | 9 | 18 | 5 | 10 | 0.900 | 0.900 | 10 | 10 | 6 | 6 | 0.924 | 0.889 |
| 6 | 27 | 2 | 9 | 0.595 | 0.556 | 8 | 32 | 2 | 8 | 0.565 | 0.500 | 9 | 3 | 6 | 2 | 0.923 | 0.750 | 10 | 15 | 6 | 9 | 0.926 | 0.926 |
| 6 | 30 | 2 | 10 | 0.600 | 0.600 | 8 | 36 | 2 | 9 | 0.562 | 0.477 | 9 | 6 | 6 | 4 | 0.933 | 0.875 | 10 | 10 | 7 | 7 | 0.951 | 0.916 |
| 6 | 4 | 3 | 0.769 | 0.667 | 8 | 40 | 2 | 10 | 0.563 | 0.500 | 9 | 9 | 6 | 6 | 0.936 | 0.917 | 10 | 5 | 8 | 4 | 0.971 | 0.938 |
| 6 | 6 | 3 | 0.784 | 0.667 | 8 | 8 | 3 | 6 | 0.747 | 0.667 | 9 | 12 | 6 | 8 | 0.938 | 0.938 | 10 | 10 | 8 | 8 | 0.972 | 0.943 |
| 6 | 8 | 3 | 4 | 0.788 | 0.667 | 8 | 16 | 3 | 6 | 0.756 | 0.688 | 9 | 15 | 6 | 10 | 0.937 | 0.908 | 10 | 10 | 9 | 9 | 0.988 | 0.988 |

**Notes:** Where two lines are listed for the same \((v, b, k, r)\), the first has higher \(A\)-efficiency (sometimes this is not detectable to three decimal places) and the second has higher \(E\)-efficiency. When \(A\)-efficiency = \(E\)-efficiency, the best design is a BIBD.
3.3 Universal Optimality and BIBDs

In a landmark paper, Kiefer (1975) tackled the question of how a design might be optimal with respect to all criteria defined by (3.3). He found this result:

**Theorem 3.1** For the class of designs $D$ with corresponding information matrices $C = \{C_d : d \in D\}$ and having $C_d1 = 0$ for all $d$, suppose there is a $d^* \in D$ for which

$$\text{trace}(C_{d^*}) = \max_{d \in D} \text{trace}(C_d);$$  \hspace{1cm} (3.9)

$$C_{d^*} = \alpha \left( I - \frac{1}{v} J \right).$$  \hspace{1cm} (3.10)

Then $d^*$ is universally optimal, that is, $d^*$ is optimal with respect to every criterion satisfying the properties (3.3).

**Proof:** Let $\Phi$ be any of the criteria satisfying (3.3), let $d \in D$ be any competing design, and let $\mathcal{P}$ be the class of all $v \times v$ permutation matrices $P$. Then invoking (3.3)(ii) and (3.3)(iii),

$$\Phi(C_d) = \Phi(PC_dP') = \frac{1}{v!} \sum_{P \in \mathcal{P}} \Phi(PC_dP') \leq \Phi \left( \frac{1}{v!} \sum_{P \in \mathcal{P}} PC_dP' \right).$$

Writing $\tilde{C}_d$ for $\frac{1}{v!} \sum_{P \in \mathcal{P}} PC_dP'$, we have $\tilde{C}_d = \alpha_d(I - \frac{1}{v} J)$, where $\alpha_d = \text{trace}(C_d)/(v - 1) \leq \alpha$. Consequently,

$$\Phi(C_d) \leq \Phi(\tilde{C}_d) = \Phi \left( \alpha_d \left( I - \frac{1}{v} J \right) \right) \leq \Phi \left( \alpha \left( I - \frac{1}{v} J \right) \right) = \Phi(C_{d^*}),$$

the last inequality by virtue of (3.3)(i), since $C_{d^*} - \tilde{C}_d = (\alpha - \alpha_d)(I - \frac{1}{v} J)$ is nonnegative definite. \qed

**Theorem 3.1** applies to block designs and to any class of designs where each member of $C$ is symmetric, nonnegative definite, and of rank deficiency one with zero row sums. The proof demonstrates why all three of the properties in (3.3) are important.

A design $d$ having information matrix of form (3.10), $C_d = \alpha(I - \frac{1}{v} J)$, enjoys an additional, useful quality that does not depend on variance magnitude, and so is not directly assessed by criteria (3.3). Let $c$ be any normalized contrast vector ($c'c = 1$). A generalized inverse of $C_d$ is $\frac{1}{\alpha} I$ so that $\text{Var}_d(c'\tau) = \sigma^2 / \alpha$ does not depend on the contrast selected. This property, known as variance balance, says that all treatment comparisons are estimated with the same precision. Variance balance simplifies interpretation of experimental results. It is the exact expression of equal treatment interest.
Example 3.4

The design in Example 3.1 has constant replication 5, constant concurrence 2, and, hence from (3.4), information matrix

\[ C_d = 5I - \frac{1}{3}[2(J - I) + 5I] = 4 \left( I - \frac{1}{6}J \right). \]

This matrix has the complete symmetry required by (3.10) and, from Lemma 3.1, has the maximal trace property (3.9) relative to all designs for \( v = 6 \) treatments in \( b = 10 \) blocks of size \( k = 3 \). Thus, by Theorem 3.1, the design is universally optimal over \( D(6, 10, 3) \).

Block designs meeting the Theorem 3.1 conditions are easily characterized.

Lemma 3.1  A block design \( d \) with block size \( k \leq v \) satisfies (3.9) and (3.10) if, and only if,

(i) \( d \) is binary;
(ii) \( d \) is equireplicate, that is, \( r_{di} \) is constant in \( i \);
(iii) \( d \) is equiconcurrent, that is, \( \lambda_{div} \) is constant in \( i \neq i' \).

Setting \( k = v \), Lemma 3.1 says the RCBDs are universally optimal. This should be no surprise; not only is the use of each treatment once in each block intuitively appealing, it has already been seen in Section 3.2 that RCBDs carry the same treatment information as an equally replicated CRD. It is easily shown that, among CRDs with a fixed number of experimental units, those with equal replication are universally optimal.

An IBD satisfying the conditions of Lemma 3.1 is called a balanced incomplete block design (BIBD). One example is the design in Example 3.1.

Given the strong result of Theorem 3.1, BIBDs are, in a sense, “ideal” IBDs; they are recommended for use whenever available. Their shortfall is that in the universe of all possible \( (v, b, k) \), BIBDs are rare. Let the integer \( r \) denote the common replication for any equally replicated IBD, and let the integer \( \lambda \) denote the common treatment concurrence for a BIBD. Here are three necessary conditions for a BIBD to exist:

Lemma 3.2  If there is a BIBD in \( D(v, b, k) \), then (i) \( r = bk/v \), (ii) \( \lambda = bk(k - 1)/v(v - 1) \), and (iii) \( b \geq v \).

The first two conditions of Lemma 3.2 follow immediately from Lemma 3.1. The third, known as Fisher’s inequality, was proven in Fisher (1940).

Lemma 3.2 places strong restrictions on the triples \( (v, b, k) \). Thinking again of Example 3.1, fix \( v = 6 \) and \( k = 3 \). Then (i) and (ii) of Lemma 3.2 jointly tell us that \( b \) must be a multiple of 10. Only in 1/10 of the design classes \( D(6, b, 3) \) is a BIBD even possible.

In addition, the conditions of Lemma 3.2, while necessary, are not always sufficient to guarantee existence of a BIBD. One of the outstanding open problems in combinatorial mathematics is to determine when those conditions are in fact sufficient. The impressive work of H. Hanani, summarized in Hanani (1975), establishes, with one exception, sufficiency for \( k \leq 5 \). For that one exception, \( (v, b, k) = (15, 21, 5) \), Chowla and Ryser (1950)
have shown that 15 treatments cannot be placed in 21 blocks of 5 treatments each so that 
(i) through (iii) of Lemma 3.1 simultaneously hold; this BIBD cannot be constructed.

For $k = 6$, the Lemma 3.2 conditions are sufficient whenever $\lambda > 2$ and when $\lambda = 2$
except for $(v, b, k) = (21, 28, 6)$. There are still several open cases for $\lambda = 1$. See Chapter II.3
of Colbourn and Dinitz (2007) for further details on this and larger $k$. This is a problem that
grows more difficult as $k$ increases.

BIBDs were first proposed as experimental designs by Yates (1936a). Since that time, a
mountainous, and still expanding, body of work on BIBD construction has evolved, with
many contributions from statisticians and mathematicians alike. A good, recent survey of
known BIBD results can be found in Chapter II.1 of Colbourn and Dinitz (2007). BIBDs are
cataloged online at, and can be downloaded from, www.designtheory.org.

The conditions of Theorem 3.1 can also be met when $k > v$. Define a multiply complete (for
short, $m$-complete) block to be a block of $k = mv$ experimental units of which $m$ are randomly
assigned to each treatment. Universally optimal designs in $D(v, b, mv)$ are the generalized
complete block designs comprised solely of $m$-complete blocks. For $k > v$ but not an integer
multiple of $v$, let $k_0 = k \pmod{v}$ and set $m = (k - k_0)/v$. If $k_0 > 1$ and there is a BIBD
design $D(v, b, k_0)$, then there is a universally optimal design $D(v, b, k)$ found by appending
an $m$-complete block to each block of $D(v, b, k_0)$ and set $m = (k - k_0)/v$. If $k_0 = 1$ and
add one block to each block of a generalized complete block design in $D(v, b, k - 1)$. These
optimal designs with $k_0 > 0$ are collectively referred to as Kiefer’s balanced block designs
(KBBDs).*

Universal optimality is defined in Theorem 3.1 as optimality with respect to all criter-
ria (3.3). In light of Theorem 3.1, it is natural to ask if this optimality can be had for any
block design if the conditions of the theorem are not met. The answer, at least in a restricted
sense, is yes.

**Theorem 3.2** (Yeh 1988) Fix $k = v - 1$. For every $m \geq 1$, there is a BIBD in $D(v, b, mv, k)$,
which is universally optimal. For other values of $b$, the following designs are universally optimal
over the subclass of binary designs in $D(v, b, k)$:

(i) $2 \leq b \leq v - 1$: any binary design $d$ with $\max_i r_{di} = b$ and $\min_i r_{di} = b - 1$.
(ii) $b = mv + 1$ and $m \geq 1$: a BIBD($v, mv, k$) to which one binary block is added.
(iii) $b = mv + b_0$ for $2 \leq b_0 \leq v - 1$ and $m \geq 1$: a BIBD($v, mv, k$) to which the blocks of a
design from (i) with $b_0$ blocks is added.

It is an open question as to whether, for any of the designs in Theorem 3.2(i)–(iii), uni-
versal optimality can be extended beyond the binary subclass. It certainly cannot always
be done. Morgan and Jin (2007) show that for $b = 2$ the designs displayed in Table 3.5 are
uniquely $E$-optimal and so $E$-better than any binary competitor.

Theorem 3.2 designs with $b \equiv \pm 1 \pmod{v}$ fall within two design classes that will be
mined for optimality in the following sections. They can be classified both as GGDD(2)s
(Section 3.4.2) and NBBD(1)s (Section 3.4.3).

* These designs are more commonly referred to as balanced block designs per Kiefer’s original terminology. The new
name proposed here distinguishes them from the many other block designs that also have the balance property,
3.4 Block Designs That Approximate BIBD Structure

The vast majority of block designs used in practice have block size no larger than the number of treatments. Setting \( k = v \), Theorem 3.1 tells us that the RCBDs are the designs of choice. With IBDs, once one leaves the BIBDs, there is no simple answer. This section explores the considerable theory that has grown, and continues to grow, around the IBD problem.

Section 3.3 presents two key facts concerning BIBDs: in terms of criteria (3.3) they are optimal in the strongest sense possible, and the triples \((v, b, k)\) for which they exist are rare in the universe of all such triples. To add perspective to the latter, for \( 3 \leq v \leq 15 \) and \( 2 \leq r \leq 30 \), there are 1021 integer triples \((v, b, k)\) where \( r = bk/v \) is an integer; of these, only 253 have \( b \geq v \) and integer \( \lambda \) (see Lemma 3.2). Even restricting to equal replication, the BIBDs leave much more unsolved than solved in the search for optimal block designs.

Yet the BIBDs are suggestive of how one might proceed. BIBDs are defined by conditions (i) through (iii) of Lemma 3.1. If these conditions produce universal optimality (as they do), then designs that are in some sense “close” to achieving those conditions should be good candidates for, if not universal optimality, individual optimalities of interest, such as those in Table 3.2. That is the tack taken in this section.

The symbols \( r \) and \( \lambda \) were assigned specific meanings in Section 3.3; now they will be defined in such a way as to preserve the earlier meanings as special cases. The starting point is the block design setting \((v, b, k)\) with \( k < v \). Associated with this setting,

\[
\begin{align*}
    r &= \left\lfloor \frac{bk}{v} \right\rfloor = \text{maximal achievable minimum replication}, \\
    p &= bk - vr = \text{plot excess over those needed for equal replication } r, \\
    \lambda &= \left\lfloor \frac{r(k-1)}{v-1} \right\rfloor = \text{maximal achievable minimum concurrence for a treatment having replication } r, \\
    q &= r(k-1) - \lambda(v-1) = \text{total concurrence excess, for a treatment having replication } r, \text{ over constant concurrence } \lambda.
\end{align*}
\]

These auxiliary parameters will be used throughout this section. The necessary conditions Lemma 3.2(i),(ii) for BIBD existence may be restated as \( p = q = 0 \), termed the BIBD setting.
The concern in this section is for those settings where at least one of the plot excess \( p \) and the concurrence excess \( q \) is not zero.

Because equal replication is so common in practice, greatest attention has been paid to settings having \( p = 0 \), and this is where we begin. Now equal replication is possible, and “closeness” to a BIBD can be formulated in terms of deviation of the \( \lambda_{dii'} \)'s from \( \lambda \). If not a BIBD setting, then necessarily \( 1 \leq q \leq v - 2 \).

### 3.4.1 Regular Graph Designs

With \( p = 0 \) and whatever the value of \( q \), let \( d \) be a binary, equireplicate design. Using \( (x_{dii'})_{i \neq i'} \) to denote a matrix that is zero on the diagonal, the information matrix \( C_d \) for design \( d \) can be written as

\[
C_d = rI - \frac{1}{k} N_dN_d' = rI - \frac{1}{k} (\lambda_{dii'})_{i \neq i'}
\]

\[
= \frac{r(k-1)}{k} I - \frac{1}{k} (\lambda)_{i \neq i'} - \frac{1}{k} [(\lambda_{dii'})_{i \neq i'} - (\lambda)_{i \neq i'}]
\]

\[
= \left( \frac{v\lambda + q}{k} \right) I - \frac{\lambda}{k} I - \frac{1}{k} \Delta_d, \tag{3.12}
\]

where the discrepancy matrix \( \Delta_d \) has elements \( \delta_{dii'} \), called discrepancies, defined by

\[
\delta_{dii'} = \begin{cases} 
0 & \text{if } i = i', \\
\lambda_{dii'} - \lambda & \text{if } i \neq i'.
\end{cases} \tag{3.13}
\]

Any discrepancy matrix is symmetric with constant row sums of \( q \).

A BIBD has all \( \lambda_{dii'} = \lambda \) and so \( \Delta_d = 0_{v \times v} \). One idea for what it means for a design \( d \) to be “close” to a BIBD is for it to be binary and equireplicate (meeting the BIBD conditions Lemma 3.2(i),(ii)), and for it to have \( \Delta_d \) close to the zero matrix in some sense. This motivates the regular graph designs.

**Definition 3.1** Let the class of designs \( \mathcal{D}(v,b,k) \) have \( p = 0 \) and \( q > 0 \). A design \( d \in \mathcal{D} \) is a regular graph design if

(i) \( d \) is binary,

(ii) \( d \) is equireplicate, and

(iii) the \( v(v-1) \) discrepancy values \( \delta_{dii'} \) are all in \( \{0,1\} \).

The name “regular graph design” comes from the fact that \( \Delta_d \) for such a design is the adjacency matrix of a regular graph with \( v \) vertices. There is an edge connecting vertices \( i \) and \( i' \) in this graph if \( \delta_{dii'} = 1 \), but no edge if \( \delta_{dii'} = 0 \). The “regular” means that each vertex is connected to the same number of other vertices, which is \( q \). In graph terminology, \( q \) is called the degree or the valency of the regular graph.

Within the binary, equireplicate class, imposing Definition 3.1(iii) is equivalent to minimizing trace\((\Delta_d',\Delta_d) = \text{trace}(\Delta_d^2)\), a natural way to minimize deviation from the complete symmetry condition (3.10) in Theorem 3.1. Alternatively, RGDs can be viewed as firstly
maximizing $\text{trace}(C_d)$ (i.e., (i) of Definition 3.1) and, given that, then minimizing $\text{trace}(C_d^2)$ (i.e., (ii) and (iii)). Designs with this $(M, S)$-property* have minimal dispersion of their eigenvalues among designs with maximal sum of eigenvalues. Take note that these ideas of closeness to a BIBD give priority to (3.9) over (3.10).

John and Mitchell (1977) conjectured that if $D(v, b, k)$ contains at least one regular graph design, then the $A$-, $D$-, and $E$-optimal designs must each be a regular graph design. If their conjecture is true, then so long as at least one regular graph design exists, all other designs can be ignored when looking for an optimal design. For $k > 2$, no counterexample is known for their conjecture for either $A$-optimality or $D$-optimality. On the other hand, it has not been proven, either. Counterexamples are known for the $E$-criterion (see Section 3.4.4).

**Example 3.5**

This regular graph design, for $(v, b, k, q) = (7, 7, 5, 2)$, is $A$-, $D$-, and $E$-optimal among regular graph designs:

$\begin{array}{cccccc}
1 & 1 & 1 & 1 & 2 & 3 \\
2 & 2 & 2 & 2 & 4 & 3 \\
3 & 3 & 3 & 5 & 5 & 4 \\
4 & 4 & 6 & 6 & 5 & 6 \\
5 & 7 & 7 & 7 & 6 & 7 \\
\end{array}$

Its discrepancy matrix is

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

If one could simply enumerate all designs in $D(v, b, k)$, then optimality with respect to any criterion could be trivially determined. The John–Mitchell conjecture for regular graph designs brings the enumerative problem within reach, up to a point. In fact, it changes the design enumeration problem into a graph enumeration problem, as follows. Each graph is a discrepancy matrix, so enumerating all regular graphs of degree $q$ is equivalent to enumerating information matrices of the form (3.12), regardless of whether or not the corresponding designs exist. The matrices can then be ranked according to the desired optimality criterion, the best determined, and only then is it necessary to find a corresponding design. Should such a design not exist, then one moves on to the next best graph, and so on.

John and Mitchell (1977) followed this strategy for up to $v = 12$ treatments and $r = 10$ replicates. Formal optimality tools can then be employed to show that all of their designs, including the design in Example 3.5, are $A$- and $D$-optimal over the entire class $D(v, b, k)$; see Theorem 3.6. Their stopping point was driven by the size of the graph enumeration problem, which grows rapidly with $v$ (and $q$). Over 35 years later, tremendous strides have been made in computational power, yet little progress has been gained on the graph enumeration problem.

* $(M, S)$-property has been often called $(M, S)$-optimality, but as it does not directly measure magnitude of contrast variances, the former term is more appropriate.
problem for \( v > 16 \). Recent progress on optimal RGDs for selected cases with \( v \) up to 18 may be found in Cakiroglu (2013).

### 3.4.2 Group Divisible Designs

A different approach to approximating the BIBD information structure can be based on giving first priority to approaching the complete symmetry condition (3.10) corresponding to Lemma 3.1(iii). How this “closeness” idea should be implemented can be seen in how the proof of Theorem 3.1 employs the convexity (3.3)(iii) of optimality functions. That usage is a particular case of the technique of matrix averaging.

**Definition 3.2** Let \( P_1, \ldots, P_t \) be \( t \) permutation matrices of order \( v \). An averaging of the information matrix \( C_d \) is

\[
\overline{C}_d = \frac{1}{t} \sum_{i=1}^{t} P_i C_d P_i^t.
\]

\( \overline{C}_d \) is called an average matrix of \( C_d \).

**Lemma 3.3** (Constantine 1981) Suppose \( \overline{C}_d \) is an average matrix of \( C_d \). Then \( \overline{C}_d \) is nonnegative definite with zero row sums and the same trace as \( C_d \), and \( \Phi(\overline{C}_d) \leq \Phi(C_d) \) for any \( \Phi \) satisfying (3.3).

For given trace of \( C_d \), Theorem 3.1 maximizes information by averaging over all treatment permutations. When complete symmetry is not attainable in an actual design, an information matrix might be more advantageously improved by averaging over subsets of treatment permutations. If a corresponding design for the “partially averaged” matrix can be found, then it could be a promising candidate for optimality. The structure \( \overline{C}_d \) produced by averaging within all subsets of a treatments partition is that of a generalized group divisible design (GGDD).

**Definition 3.3** The design \( d \in D(v, b, k) \) is a GGDD(s) if the treatments in \( d \) can be divided into \( s \) mutually disjoint sets \( V_1, \ldots, V_s \) of size \( v_1, \ldots, v_s \) such that the elements \( c_{di} \) of \( C_d \) satisfy

(i) For \( g = 1, \ldots, s \) and all \( i \in V_g \), \( c_{di} = r_{di} - \lambda_{di}/k = c_g \), where \( c_g \) depends on the set \( V_g \) but not otherwise on the treatment \( i \);  
(ii) For \( g, h = 1, \ldots, s \) and all \( i \in V_g \) and \( i' \in V_h \), with \( i \neq i' \) if \( g = h \), 

\[
-kc_{di}/k = \lambda_{di} = \gamma_{gh},
\]

where \( \gamma_{gh} \) depends on the sets \( V_g \) and \( V_h \) but not otherwise on the treatments \( i \) and \( j \).

It is convenient to assume that the treatment subsets \( V_1, \ldots, V_s \) in the preceding definition are arranged so that the \( c_g \)'s are in nonincreasing order: \( c_1 \geq c_2 \geq \cdots \geq c_s \). If \( |V_g| = 1 \), then \( \gamma_{gg} \) is defined to be zero. This version of the GGDD definition is due to Srivastav and Morgan (1998), with the idea in less general form dating back to Adhikary (1965).

**Example 3.6**

The following design \( d \in D(6, 11, 3) \) is a GGDD(2).

\[
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\
2 & 2 & 2 & 2 & 3 & 4 & 3 & 3 \\
3 & 4 & 5 & 6 & 5 & 6 & 5 & 6 \\
\end{array}
\]

© 2015 by Taylor & Francis Group, LLC
The information matrix for $d$ is
\[
\mathbf{C}_d = \frac{1}{3} \begin{pmatrix}
12 & -4 & -2 & -2 & -2 & -2 \\
-4 & 12 & -2 & -2 & -2 & -2 \\
-2 & -2 & 10 & -2 & -2 & -2 \\
-2 & -2 & -2 & 10 & -2 & -2 \\
-2 & -2 & -2 & -2 & 10 & -2 \\
-2 & -2 & -2 & -2 & -2 & 10
\end{pmatrix}.
\]

The groups are $V_1 = \{1, 2\}$ and $V_2 = \{3, 4, 5, 6\}$. The concurrence parameters are $\gamma_{11} = 4$ and $\gamma_{12} = \gamma_{22} = 2$.

Definition 3.3 requires neither binarity nor equal replication, though both will frequently be employed. For instance, a binary, equireplicate GGDD(1) is just a BIBD, giving us our first example of optimality for GGDDs. The optimal design in Example 3.3, which is also a GGDD(1), is a case where neither additional property is employed. Nor are those properties possessed by the designs in Table 3.5, which are $E$-optimal GGDD(3)s (see Theorem 3.16).

A binary, equireplicate GGDD, with equal group sizes and with the further property that $\gamma_{gh}$ is constant in $g \neq h$, goes by the simpler name group divisible design (GDD). It is customary with GDDs to drop the $\gamma$ notation entirely and refer instead to the in-group concurrence $\lambda_1$ and the out-group concurrence $\lambda_2$. If $|\lambda_1 - \lambda_2| = 1$, then a GDD is a RGD. These designs were formally defined by Bose and Shimamoto (1952).

Some GDDs enjoy a very broad optimality that, though not as strong as universal optimality, does encompass all three of the eigenvalue-based criteria listed in Table 3.2. In terms of eigenvalues, complete symmetry (3.10) is equivalent to $e_d^1 = e_d^2 = \cdots = e_{d,v-1}$. Approximating complete symmetry (as through the GDD notion) will be valuable from an optimality perspective if it corresponds to an optimal relaxation of eigenvalue equality. The type 1 optimality criteria provide a framework for investigating the form of this relaxation.

**Definition 3.4** Let $f$ be a real-valued function on $\mathbb{R}$ and define $\Phi_f$ on the class of information matrices $\mathcal{C}$ by
\[
\Phi_f(\mathbf{C}_d) = \sum_{i=1}^{v-1} f(e_{di}).
\]
Minimizing $\Phi_f$ is a type 1 optimality criterion if
\begin{enumerate}[(i)]
  \item $f$ is continuously differentiable on $(0, T)$,
  \item $f' < 0$, $f'' > 0$, $f''' < 0$ on $(0, T)$,
  \item $f$ is continuous at 0 and $\lim_{x \to 0^+} f(x) = +\infty$,
\end{enumerate}
where $T = \max \text{trace}(\mathbf{C}_d)$.

Taking $f(x) = 1/x$ gives the A criterion and $f(x) = -\log(x)$ the D criterion. All of the $\Phi_f$ criteria are included in the type 1 class, including (in the limit) $E$. The type 1 class does not include the MV criterion, for its value is not determined solely by the $e_{di}$.

Theorems 3.3 through 3.5 to follow each provide eigenvalue conditions for attainment of optimality with respect to type 1 criteria. They do this by employing various combinations of moment and extremal bounds for the set of $v - 1$ eigenvalues. Though not all are needed in this subsection, these theorems are best understood as a group, so are given here. Applications of these results extend to many situations other than GDDs, some of which will be explored in later subsections.
The task of minimizing $\Phi_f$ can be cast as an optimization problem independent of information matrices and their eigenvalues. The goal is to

$$\minimize \sum_{i=1}^{v-1} f(x_i)$$

for $f$ of Definition 3.4 and positive variables $x_1 \leq x_2 \leq \cdots \leq x_{v-1}$. Where the minimum is attained will depend on constraints ruling the $x_i$. We will consider (in various combinations) the following:

(i) $\sum x_i = T_1$ for some positive $T_1$.

(ii) $\sum x_i^2 \geq T_2$ for some $T_2$ satisfying $\frac{T_1^2}{v-1} \leq T_2 \leq T_1^2$.

(iii) $x_1 \leq T_0$ for some $T_0$ satisfying

$$T_0 \leq (T_1 - \sqrt{\frac{v-1}{v-2}}P)/(v-1) \quad \text{where} \quad P = \sqrt{T_2 - \frac{T_1^2}{v-1}},$$

(b) $(T_1 - T_0)^2 \geq T_2 - T_0^2 \geq (T_1 - T_0)^2/(v-2)$.

In the design optimality application, constraints (3.14)(i) through (iii) correspond, respectively, to fixing the trace value $\text{trace}(C_d) = \sum_i e_{di}$, placing a lower bound on the trace square $\text{trace}(C_d^2) = \sum_i e_{di}^2$, and placing an upper bound on the smallest eigenvalue $e_{d1}$.

Theorem 3.3 (Cheng 1978) The minimum of $\sum f(x_i)$ subject to conditions (3.14)(i) and (ii) is at

$$x_1 = x_2 = \cdots = x_{v-2} = \frac{T_1 - \sqrt{\frac{v-1}{v-2}}P}{v-1} \quad \text{and} \quad x_{v-1} = \frac{T_1 + \sqrt{(v-1)(v-2)}P}{v-1},$$

where $P$ is as defined in (3.14)(iii).

For Theorem 3.3 applied to the design optimality problem, the first moment of the eigenvalues is fixed, such as when restricting to binary designs, and their second moment is bounded from below. The minimum of $\Phi_f$ is achieved at the second moment bound if the smallest $v-2$ eigenvalues can be made equal. The next result incorporates information on the smallest eigenvalue.

Theorem 3.4 (Jacroux 1985) The minimum of $\sum f(x_i)$ subject to all of the conditions (3.14)(i) through (iii) is at $x_1 = T_0$,

$$x_2 = \cdots = x_{v-2} = \frac{(T_1 - T_0) - \sqrt{\frac{v-2}{v-3}}P_0}{v-2}, \quad \text{and} \quad x_{v-1} = \frac{(T_1 - T_0) + \sqrt{(v-2)(v-3)}P_0}{v-2},$$

where $P_0 = \left[ (T_2 - T_0^2) - \frac{(T_1 - T_0)^2}{v-2} \right]^\frac{1}{2}$. 

© 2015 by Taylor & Francis Group, LLC
In Theorem 3.4 the first moment of the eigenvalues is again fixed and again their second moment is bounded from below, but now there is also an upper bound for the minimum eigenvalue. The optimum is achieved at the second moment bound if we can get the smallest eigenvalue to reach its upper bound, and the remaining eigenvalues have a structure as in Theorem 3.3. Theorem 3.3 can also be refined by incorporating information on the largest eigenvalue, as shown next.

**Theorem 3.5** (Cheng and Bailey 1991) The minimum of $\sum f(x_i)$ subject to conditions (3.14)(i) and (ii), and the additional constraint $x_{v-1} \leq T_v$, is at any set of $x_i$'s for which there are exactly two distinct values and for which

$$\sum x_i^2 = T_2 \quad \text{and} \quad x_{v-1} = T_v.$$  

For any fixed trace of the information matrix, Theorems 3.3 through 3.5 show how the “equality of eigenvalues” implied by (3.10) can be advantageously relaxed in situations where it is not achievable. We now investigate the application of these results.

In the following, a design is termed generalized optimal if it is optimal with respect to all type 1 criteria. Working with a variant of Theorem 3.3 (see the original paper for details), Cheng (1978) proved generalized optimality for GDD(2)s having $\lambda_2 = \lambda_1 + 1$. This remains one of the most far-reaching optimality results, outside the realm of universal optimality explored in Section 3.3, which has been established for block designs. Otherwise, with few exceptions (see the next paragraph and following Theorem 3.9), it has been necessary to either restrict the class $D(\nu, b, k)$ of competing designs, to restrict to a particular criterion, or both. The GDDs have been fertile ground in both regards.

In a very early paper, Takeuchi (1961) established $E$- and $MV$-optimality of the GDD(s) having $\lambda_2 = \lambda_1 + 1$ for any $s \geq 2$. A GDD is said to be singular if $r = \lambda_1$ and semiregular if $r > \lambda_1$ and $r k = \nu \lambda_2$. Employing Theorem 3.5, Cheng and Bailey (1991) established generalized optimality within the binary, equireplicate subclass of $D(\nu, b, k)$, for singular GDDs with $\lambda_2 = \lambda_1 - 1$ and semiregular GDDs with $\lambda_2 = \lambda_1 + 1$. Using majorization (see Bhatia 1997; and Section 3.5.3) techniques to establish inferiority of nonbinary and unequally replicated competitors, Bagchi and Bagchi (2001) were in many cases able to extend these generalized optimality results to the full class $D(\nu, b, k)$.

The dual of an IBD $d$ with treatment/block incidence matrix $N_d$ is the design $\tilde{d}$ (say) for which $N_{\tilde{d}} = N_d'$. If $d$ is binary with equal replication $r$, then $\tilde{d}$ is binary with replication $\tilde{r} = k$ and $C_{\tilde{d}} = kI - \frac{1}{k}N_{\tilde{d}}N_{\tilde{d}}'$ (compare Equation 3.4). Owing to the fact, true of any matrix times its transpose, that $N_d'N_d$ has the same nonzero eigenvalues as $N_{\tilde{d}}'N_{\tilde{d}}'$, the duals of the singular and semiregular GDDs mentioned earlier are also generalized optimal among binary, equireplicate competitors. The Cheng and Bailey (1991) paper contains several other results of this flavor, found by identifying classes of designs whose eigenvalues satisfy Theorem 3.5.

As noted earlier, GDDs are RGDs whenever $|\lambda_1 - \lambda_2| = 1$. GDDs are attractive targets for optimality arguments along the lines given earlier because their information matrices produce only two distinct eigenvalues. While other RGDs need not have this distinction, they do all possess $(M, S)$-property, so are approximating the requirements of Theorem 3.1 but without necessarily having a “nice pattern” for their eigenvalues like found in Theorems 3.3 through 3.5. Theorem 3.4 can nevertheless be effectively exploited for the RGD-optimality
problem. A two-armed approach handles the binary and nonbinary subclasses of $D(v,b,k)$ separately. Henceforth denote the binary subclass by $B(v,b,k)$.

In the theorem to follow, $T_0$ and $T^*_0$ are upper bounds for $e_{d_1}$, found from Lemma 3.5 and Corollary 3.1 (in Section 3.4.4) applied to non-RGDs. The smallest trace ($C^2_d$) that can be attained by a binary, non-RGD is the value used for $T_2$. These values are plugged into Theorem 3.4 and a result of Kunert (1985) to produce non-RGD lower bounds for $\Phi_f(C_d)$, displayed as the right-hand side of the inequalities (3.15) for binary designs and (3.16) for nonbinary designs.

**Theorem 3.6** (Jacroux 1985) Let $\tilde{d} \in D(v,b,k)$ for $v > k \geq 3$ be a RGD. Write

$$
T_0 = \max \left\{ \frac{(r(k - 1) - (\lambda + 2)v}{(v - 2)k}, \frac{(r(k - 1) + \lambda - 1)}{k}, \frac{(r - 1)(k - 1)v}{(v - 1)k} \right\},
$$

$$
T^*_0 = \max \left\{ \frac{(r(k - 1) - 2)v}{(v - 1)k}, \frac{(r - 1)(k - 1)v}{(v - 1)k} \right\},
$$

$$
T_1 = b(k - 1), \quad T_2 = \text{tr}(C^2_d) + \frac{4}{k^2}, \quad \text{and} \quad P_0 = \left[ \frac{T_2 - T^2_0}{v} - \frac{(T_1 - T_0)^2}{v - 2} \right]^{\frac{1}{2}}.
$$

If $T_0 \leq [(T_1 - T_0) - \sqrt{\frac{v - 2}{v - 3}P_0}]/(v - 2)$ and

$$
\sum_{i=1}^{v-1} f(e_{d_i}) < f(T_0) + (v - 3)f\left( \frac{(T_1 - T_0) - \sqrt{\frac{v - 2}{v - 3}P_0}}{v - 2} \right) + f\left( \frac{(T_1 - T_0) + \sqrt{(v - 2)(v - 3)P_0}}{v - 2} \right),
$$

(3.15)

then a $\Phi_f$-optimal design in $B(v,b,k)$ must be a RGD. If moreover $T^*_0 \leq [T_1 - \frac{2}{b} - T^*_0]/(v - 2)$ and

$$
\sum_{i=1}^{v-1} f(e_{d_i}) < f(T^*_0) + (v - 2)f\left( \frac{T_1 - \frac{2}{b} - T^*_0}{v - 2} \right),
$$

(3.16)

then a $\Phi_f$-optimal design in all of $D(v,b,k)$ must be a RGD.

If these conditions hold, then all non-RGDs are ruled out with respect to $\Phi_f$, but Theorem 3.6 does not say that the design $\tilde{d}$ must be $\Phi_f$-optimal. Rather, it says that some RGD will be the optimal design; one must still complete the enumeration described following Example 3.5. Jacroux (1985) employed Theorem 3.6 to establish $A$- and $D$-optimality, over all of $D(v,b,k)$, of the best RGDs determined by John and Mitchell (1977), lending strong credence to the John–Mitchell conjecture. Theorem 3.6 and its application to RGDs are an excellent example of how good theory can bring a complicated design problem within computational reach.
To this point, all of the designs shown to be optimal in this section have been binary. Even more, they have all had \((M, S)\)-property, though this is not a requirement of the GGDDs in Definition 3.3. Indeed, there is no known example of an \(A\)-optimal IBD that is not binary. The situation is different with respect to the minimax criteria \(E\) and \(MV\). \(E\)-optimality will be explored in more detail in Section 3.4.4; the remainder of this subsection concentrates on the \(MV\) criterion.

Many of the \(MV\)-optimality arguments found in the literature build on the simple inequality (3.17) first proven by Takeuchi (1961). It is applicable to any class of information matrices \(C\) as identified prior to (3.3). For \(C_d \in C\), \(C_d = (c_{dii})\),

\[
\frac{\text{Var}_d(\tau_i - \tau_{i'})}{\sigma_E^2} \geq \frac{4}{c_{dii} + c_{dii'} - 2c_{dii'}}. \tag{3.17}
\]

The proofs of Theorems 3.7 and 3.8, not shown here, depend critically on (3.17). Again notice the emphasis on symmetry. The Theorem 3.7 result is for GGDD(1)s.

**Theorem 3.7** (Morgan and Uddin 1995) *For the class of designs \(D = D(v, b, k)\), suppose there is a \(d^* \in D\) for which*

\[
\min_i \sum_{i' \neq i} c_{d^*ii'} = \max_{d \in D} \min_i \sum_{i' \neq i} c_{dii'}, \text{ and } \tag{3.18}
\]

\[
C_{d^*} = \alpha \left( I - \frac{1}{v} J \right). \tag{3.19}
\]

*Then \(d^*\) is \(MV\)-optimal in \(D\). Moreover, if \(d \in D\) and \(C_d \neq C_{d^*}\), then \(d^*\) is \(MV\)-superior to \(d\).*

The second condition (3.19) of Theorem 3.7 is identical to the second condition (3.10) of Theorem 3.1. The two theorems are distinguished by their first conditions (3.18) and (3.9), with (3.18) allowing the possibility that trace may not be maximized. When the two conditions of Theorem 3.7 can be met without simultaneously meeting those of Theorem 3.1, \textit{maximum trace designs cannot be \(MV\)-optimum}, for the asymmetry they necessarily entail results in higher variances for some elementary contrast estimators.

Theorem 3.7 gives us our first insight into how nonbinarity can advantageously enter into the IBD-optimality problem. When equal replication is not possible (plot excess \(p > 0\)), judicious use of nonbinarity can improve the approximation to complete symmetry and with this improve an optimality measure, particularly those (\(MV\) and \(E\)) that are most sensitive to an extreme contrast variance. Morgan and Srivastav (2002) have proven that the \(MV\) conclusions of Theorem 3.7 also hold for the \(E\)-criterion.

Like Theorem 3.1, which requires \((p, q) = (0, 0)\), the scope of Theorem 3.7 is fairly small in the \((v, b, k)\) universe, needing (as can be shown) \((p, q) = (1, 0)\). The nonbinary design in Example 3.3 is \(MV\)-optimal over \(D(5, 7, 3)\) by Theorem 3.7. The \(A\)-optimal design in \(D(5, 7, 3)\) is binary.
Nonbinarity can also enter into MV-optimal GGDDs with \( s > 1 \), as demonstrated next. Though neither binarity nor equal replication is required, Theorem 3.8(i) does imply binarity when \( p = 0 \). For \( r = \lfloor bk/v \rfloor \), the \( c \) in (i) is \( c = r(k - 1)/(v - 1) \).

**Theorem 3.8** (Srivastav and Morgan 1998) Let \( d^* \in \mathcal{D}(v, b, k) \) be a GGDD(s) for which

(i) \( c_g \geq c \) for \( 1 \leq g \leq s - 1 \) and \( c_s = c \),
(ii) \( \gamma_{ss} \geq \left( \frac{pk + 2r(k - 1)}{2(v - 1)} \right) \),
(iii) \( \gamma_{gg} \geq \gamma_{ss} \) for \( 1 \leq g \leq s - 1 \),
(iv) \( \gamma_{gh} \) is constant in \( g \neq h \),
(v) \( \frac{(\gamma_{gg} + (v - 1)\gamma_{1s})}{\gamma_{1s}(k\gamma_g + \gamma_{gg})} + \frac{(\gamma_{hh} + (v - 1)\gamma_{1s})}{\gamma_{1s}(k\gamma_h + \gamma_{hh})} \leq \frac{2}{k\gamma_s} \) for \( 1 \leq g, h \leq s \) with \( g \neq h \).

If \( p \leq v - 2 \), then \( d^* \) is MV-optimal in \( \mathcal{D}(v, b, k) \).

The nonbinary design in Example 3.6 is MV-optimal by Theorem 3.8. Again, nonbinarity is coming into play here when equal replication is not possible, provided it allows the group divisible structure to be achieved, so approximating (3.10) at the expense of (3.9). Condition (i), however, limits the extent of nonbinarity. Many results of a similar flavor to Theorem 3.8 are known; see Srivastav and Morgan (1998) and the references therein.

The last two theorems beg the question of whether an optimal design in any of the senses in Table 3.2 can be nonbinary when equal replication is possible (i.e., when \( p = 0 \)). The conjectured answer is no. Morgan (2009) examined this question for GGDD(1)s, also known as variance-balanced designs. Consider two designs \( d_1 \) and \( d_2 \). If \( C_{d_1} \neq C_{d_2} \neq 0 \) is nonnegative definite, then \( d_1 \) dominates \( d_2 \) in the sense explained following (3), in which case \( d_2 \) is inadmissible. Let \( \bar{d} \in \mathcal{D}(v, b, k) \) with \( p = 0 \) be a nonbinary GGDD(1), implying that \( e_{\bar{d}i} \equiv \bar{e} \) is constant in \( i \). If there is an \( E \)-optimal \( d \in \mathcal{D}(v, b, k) \) that is binary and equireplicate, then

\[
C_{d} - C_{\bar{d}} = C_{d} - \bar{e}(I - \frac{1}{v}J) \text{ is nonnegative definite and } \bar{d} \text{ is inadmissible.}
\]

Morgan (2009) showed that of the 230 GGDD(1)s with \( v \leq 15 \) and \( p = 0 \) listed by Billington and Robinson (1983), 228 are inadmissible by virtue of comparison with \( E \)-optimal designs; the two remaining cases were left unresolved due to the \( E \)-optimal designs not being known.

### 3.4.3 Nearly Balanced Incomplete Block Designs

Theorems 3.7 and 3.8 demonstrate that when equal replication is not possible, nonbinary designs with sufficient balance are sometimes MV- or \( E \)-optimal (Sections 3.4.2 and 3.4.4). This has not been found to be the case for \( A \)-optimality, leading to a need to identify more flexible, binary design classes that leave room for measured steps away from symmetry.

**Definition 3.5** A nearly balanced incomplete block design with concurrence range \( \psi \), or NBBD(\( \psi \)), is a design \( d \in \mathcal{D}(v, b, k) \) satisfying

(i) Each \( n_{dij} = 0 \) or 1 (the design is binary),
(ii) Each \( r_{di} = r \) or \( r + 1 \) (replication numbers as equal as possible),
Definition 3.5 is from Morgan and Srivastav (2000). The NBBD(1)s were earlier studied by Jacroux (1985), Jacroux (1991), and Cheng and Wu (1981), the last also considering a subclass of the NBBD(2) designs. The definition is tailored toward \((M, S)\)-property, while allowing through (iii) that the concurrence numbers \(\lambda_{iij}\) need not be as close as possible.

Most of the designs found optimal in Sections 3.4.1 and 3.4.2 are NBBD(1)s for plot excess \(p = 0\). We begin here with \(p > 0\) and \(\psi = 1\) but will also see that \(\psi = 1\) is not always achievable and that contrary to the intuition from BIBDs, minimizing \(\psi\) does not necessarily lead to the best design (see Example 3.8). Two simple methods for constructing NBBD(1)s when \(p > 0\) begin with a BIBD \(d \in \mathcal{D}(v, b, k)\). For any \(b_0\) such that \(b_0k < v\), one may (i) delete \(b_0\) disjoint blocks from \(d\) or (ii) append \(b_0\) disjoint blocks on the \(v\) treatments to \(d\). The resulting NBBDs are binary GGDD\((b_0 + 1)s\).

If \(p > 0\) so that equal replication is not possible, the NBBD(1)s, being structurally close to RGDs, are promising candidates for \(A\)- and \(D\)-optimality. Relative to a NBBD(1), other binary designs must have larger trace\((C_d^2)\), opening the door for an adaptation of Theorem 3.4 for NBBD(1)s similar to that done with Theorem 3.6 for RGDs. That is, one can bound the \(E\)-value and the trace-square value for binary, non-NBBD(1)s, use these bounds to calculate a bound for the \(A\)-value (say) of non-NBBD(1)s via Theorem 3.4, and then compare this bound to the \(A\)-value of an actual NBBD(1) (compare (3.15) in Theorem 3.6). If the \(A\)-value of the NBBD(1) is smaller, then an \(A\)-optimal design in \(B(v, b, k)\) must be a NBBD(1). Analogous steps lead to an inequality similar to (3.16) for ruling out nonbinary competitors. See Jacroux (1991) for details, where many NBBD(1)s with \(k > 2\) are shown to be \(A\)-optimal and refinements to Theorem 3.4 are also incorporated.

Some settings \((v, b, k)\) do not allow \(\psi = 1\) and larger concurrence ranges need to be considered. This comes about in either of two basic ways. First, if binarity forces some \(\lambda_{dij}\) to be less than the concurrence parameter \(\lambda\) defined in (3.11), then the concurrence range is necessarily at least two. An example is \(D(7, 28, 5)\) of Example 3.2. Second, requiring all \(\lambda_{dij} \geq \lambda\) will, in some settings, force some \(\lambda_{dij} \geq \lambda + 2\) as shown in Lemma 3.4.

**Lemma 3.4** (Cheng and Wu 1981) Let \(d \in B(v, b, k)\) have replication numbers \(r_{di} \geq r\) for all \(i\) and concurrence numbers \(\lambda_{dij} \geq \lambda\) for all \(i \neq j\). If the concurrence excess \(q\) satisfies (i) \(q > v - k\) or (ii) \(q \leq v - k\) and \(p(k - p) > (v - 2p)q\), then \(\lambda_{dij} \geq \lambda + 2\) for some \(i \neq j\).

In similar fashion as explained earlier for NBBD(1)s, for settings where \(\psi < 2\) is not possible, Theorem 3.4 can be adapted for establishing optimality of NBBD(2)s. This was done by Morgan and Srivastav (2000) who, among other findings, discovered instances of \(A\)-optimal NBBD(3)s. Examples 3.7 and 3.8 are based on their results.

**Example 3.7**

The following member of \(\mathcal{D}(5, 7, 3)\) is a NBBD(2) with minimal concurrence \(\lambda_{d12} = 1 = \lambda - 1\).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
This design is $A$- and $D$-optimal, but neither $E$- nor $MV$-optimal. The best design for the latter two criteria is nonbinary—see Example 3.3.

Example 3.8

Up to isomorphism, $\mathcal{D}(9,11,5)$ contains two NBBD(2)s, here labeled $d_1, d_2$, and one NBBD(3), $d_3$.

$$
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \\
3 & 3 & 4 & 5 & 4 & 6 & 6 & 5 \\
4 & 7 & 6 & 8 & 6 & 7 & 7 & 6 \\
5 & 9 & 8 & 9 & 8 & 9 & 8 & 7 \\
\end{array}
$$

$$
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \\
3 & 3 & 4 & 6 & 4 & 6 & 5 & 6 \\
4 & 5 & 7 & 5 & 7 & 8 & 7 & 8 \\
5 & 9 & 8 & 9 & 8 & 9 & 9 & 9 \\
\end{array}
$$

$$
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 2 & 2 & 2 & 2 & 3 & 3 & 3 \\
3 & 3 & 4 & 4 & 7 & 4 & 6 & 7 \\
4 & 6 & 5 & 6 & 8 & 5 & 6 & 7 \\
5 & 8 & 7 & 9 & 9 & 8 & 7 & 9 \\
\end{array}
$$

All three designs have the same trace($C^2_d$) value. Of the NBBD(2)s, $d_1$ is $A$- and $D$-better, while $d_2$ is $E$- and $MV$-better. The NBBD(3), however, is better than the NBBD(2)s on all four of these criteria and is $A$- and $D$-optimal over all of $\mathcal{D}(9,11,5)$.

The approach taken in this subsection is also useful for the irregular BIBD settings: those triples $(v, b, k)$ where the necessary conditions for existence of a BIBD (Lemma 3.2) are satisfied, but no BIBD exists. In an irregular BIBD setting, a binary design must have some $\lambda_{d_{ij}}$ less than $\lambda$, suggesting NBBD(2)s as likely candidates for optimality. Since $p = 0$, equal replication is possible, and any binary, equireplicate design $d$ will have discrepancy matrix $\Delta_d$ with row and column sums of zero. Define the total discrepancy $\delta_d$ of such a matrix to be $\delta_d = \sum_{i \neq j} |d_{ij}|/2$. Reck and Morgan (2005) enumerated the 51 feasible nonzero submatrices for all $\Delta_d$ for NBBD(2)s and NBBD(3)s having $\delta_d \leq 5$ and established that should a corresponding design $d$ exist for any one of these, then the $A$-optimal and $D$-optimal designs in an irregular BIBD setting with $r \leq 41$ must be a NBBD with concurrence range and total discrepancy in these ranges. For $r \leq 41$, there are 497 BIBD settings that are either known to be irregular or for which the existence question is open. The smallest of these is the basis for Example 3.9, taken from Reck and Morgan (2005).
Example 3.9

An $A$- and $D$-optimal design in $D(15, 21, 5)$. The total discrepancy of this design is $\delta_d = 4$.

Including $\Delta_d = 0$, there are seven discrepancy matrices that produce better $A$-values than that of this design, but in $D(15, 21, 5)$, there is no corresponding design for any of them.

In a later paper, the same authors (Morgan and Reck 2007a) determined an $E$-optimal design in $D(15, 21, 5)$. Interestingly, that NBBD(2) has total discrepancy 7, so its discrepancy matrix is not among the 51 candidate matrices for the aforementioned $A$-optimality. Moreover, they found little relationship between the $A$- and $E$-rankings of discrepancy matrices. These results point to the difficult interplay between combinatorial and optimality aspects of designs that can arise when “nice” designs do not exist. To date $D(15, 21, 5)$ is the only irregular BIBD setting where optimal designs have been determined.

3.4.4 E-Optimal Block Designs

There has been a wealth of research on $E$-optimality of block designs, particularly for settings with plot excess $p = 0$ and more generally for those that fall within the NBBD(1) framework. This is partly because, in working with only the smallest eigenvalue $e_{d1}$ of $C_d$, the technical problems tend to be more readily tractable than with other eigenvalue-based criteria. But it is also due to its pragmatic appeal of keeping any contrast from being estimated poorly. When $E$-optimal designs differ from those that are $A$- or $D$-optimal, it is not uncommonly due to the latter estimating a few contrasts poorly and many slightly better.

The $E$-criterion asks that $e_{d1}$ be maximized, so that $E$-optimality arguments accordingly require upper bounds for this smallest, positive eigenvalue of $C_d$. The basic lemma by which these are obtained (cf. result 6.59(b) in Seber 2008) is stated next.

Lemma 3.5  For any $C_d$ and any nonzero vector $h$,

$$e_{d1} \leq \frac{h'C_dh}{h'(I - \frac{1}{p}J)h}.$$  

Different choices for $h$ lead to different bounds, with utility depending on the particular design problem faced. Corollary 3.1 displays the most frequently employed.
Corollary 3.1 The smallest positive eigenvalue $e_{d1}$ of $C_d$ satisfies

(i) $e_{d1} \leq (c_{dii} + c_{dii} - 2c_{dii})/2$, $i \neq i' = 1, \ldots, v$;

(ii) $e_{d1} \leq \sum_{t \in M} \sum_{t' \in M} v c_{dii} / m(v - m)$ for any $m = |M|, M \subset \{1, 2, \ldots, v\}$.

The bound (i) is based on pairwise contrasts, compare (3.17). Application of Corollary 3.1 leads easily to bounds for investigating $E$-optimality of regular graph designs.

Theorem 3.9 (Cheng 1980; Jacroux 1980) Let $d \in \mathcal{D}(v, b, k)$ with $k < v$ and $p = 0$:

(i) $e_{d1} \leq \max\{[r(k - 1) - \lambda - 2]v/[\lambda - k], [r(k - 1) + \lambda - 1]/k\}$ if $d$ is equireplicate but not a RGD.

(ii) $e_{d1} \leq (r - 1)(k - 1)v/[(v - 1)k]$ if $d$ is not equireplicate.

Thus, if the $E$-best RGD $d^* \in \mathcal{D}(v, b, k)$ satisfies

$$e_{d^*1} \geq \max\{[r(k - 1) - \lambda - 2]v/[\lambda - k], [r(k - 1) + \lambda - 1]/k, (r - 1)(k - 1)v/[(v - 1)k]\},$$

then $d^*$ is $E$-optimal over all of $\mathcal{D}(v, b, k)$.

Cheng (1980) and Jacroux (1980), whose broadly overlapping articles were published back to back, applied Theorem 3.9 and related arguments to between them establish $E$-optimality for a substantial number of smaller designs, including 159 of the 209 parameter sets considered by John and Mitchell (1977). Cheng (1980) further proved $E$-optimality for GDDs having $\lambda_1 = \lambda_2 + 1, \lambda_2 \geq 1$ as well as for the duals of these designs and for the duals of BIBDs and of GDDs with $\lambda_2 = \lambda_1 + 1$. The duals of BIBDs, known as linked block designs, in some cases enjoy generalized optimality (Bagchi and Bagchi 2001).

GDDs are members of a class of binary, equireplicate IBDs known as partially balanced incomplete block designs (PBIBDs) with two associate classes. Any 2-class PBIBD $d$ has exactly two concurrence numbers $\lambda_{i'j'} \in \{\lambda_1, \lambda_2\}$, and the pairwise contrast variances $\text{Var}_d(\tau_i - \tau_i')$ take just two distinct values; see Dey (2010) or Clatworthy (1973) (which includes an extensive tabling of smaller designs) for a good introductory treatment, including definition of triangular parameters for these designs. PBIBDs were extensively studied in the statistics community from their introduction by Bose and Nair (1939) until interest was overtaken by increasing emphasis on optimality from the mid-1970s onward. In effect, the PBIBDs were too broad a class from an optimality perspective, containing both optimal and markedly inefficient members. PBIBDs are most competitive when, as in the optimality results for the aforementioned GDDs, they have concurrence range of one, in which case they have $(M, S)$-property. Constantine (1982), relying heavily on Corollary 3.1(ii), was able to establish $E$-optimality of 2-class PBIBDs with $(\lambda_1, \lambda_2) = (1, 0)$ or $(0, 1)$ whenever an additional inequality on the triangular parameters is met. This included all 2-class PBIBDs with $(\lambda_1, \lambda_2) = (1, 0)$ and $b < v$; see the paper for details and full results. Cheng (2012) has recently combined more detailed knowledge of regular graphs with the bounds of Theorem 3.9 to prove further $E$-optimality results for RGDs, including $E$-optimality over equireplicate designs for all those PBIBDs with $\lambda_2 = \lambda_1 + 1$ known as either triangular type or $L_2$ type.
Expanding on the John and Mitchell (1977) cataloging of RGDs and later proofs of their optimality over all competitors, there has been progress on determining all E-optimal block designs for settings with \( p = bk - vr = 0 \) in a practical range. The basic idea is first to determine the E-best design in the equireplicate subclass \( B^* \) of \( B(v, b, k) \) and then to prove its optimality over all of \( D(v, b, k) \). The first step requires \( C_d \) to be of the form (3.12), in which case \( e_{d1} \) can be expressed in terms of \( a_{dv} \), the largest eigenvalue (other than \( q \)) of the discrepancy matrix \( \Delta_d \):

\[
e_{d1} = \frac{v\lambda + q}{k} - \frac{a_{dv}}{k}.
\]

With (3.20), E-optimization over \( B^* \) is translated to an eigenvalue problem for discrepancy matrices as defined by (3.12) and (3.13). That is, it is translated to the class of symmetric, zero-diagonal, integer matrices with row sums \( q \). John and Mitchell (1977) restricted this class to contain only 0/1 matrices, making a full search feasible for up to \( v = 12 \). Morgan (2007b) adapted Lemma 3.5 to remove that restriction, allowing the full class to be studied, but without a full enumeration, for \( v \leq 15 \). Relative to regular graph discrepancy matrices, bounding via Lemma 3.5 can initially eliminate any discrepancy matrix \( \Delta \) with values corresponding to \( \lambda \) that are too far from \( \lambda \) to allow optimality. As \( v \) grows, this still leaves far too many matrices to evaluate, a problem that can be overcome with another application of Lemma 3.5. Let \( \Delta_d^{(t)} \) be any \( t \times t \) principal minor of \( \Delta_d \), and let \( H_t \) be the set of normalized \( t \)-vectors. Then

\[
e_{d1} \leq \frac{v\lambda + q}{k} - \frac{1}{k} \max_{h \in H_t} \left( \frac{v}{v - (h'1)^2} \right) h' (\Delta_d^{(t)} - \frac{q}{v} f) h.
\]

Inequality (3.21) allows competitive \( \Delta \) matrices to be built up one row/column at a time (take \( t = 3, \ldots, v \)), at each step eliminating those substructures that are bounded below a predetermined value (such as \( e_{d1} \) for a known RGD). In this way, Morgan (2007b) was able to determine E-best discrepancy matrices, including many with values outside \( \{0,1\} \). These matrices may be classified into type 1, those with values in \( \{0,1\} \) only; type 2, values ranging from 0 to 2; type 3, values ranging from \(-1\) to 1; and type 4, values ranging from \(-1\) to 2; see Table 3.6. There are many cases where RGDs are not uniquely E-optimal (compare Theorem 3.9). When no \( \Delta \) matrix of type 1 is E-best, no RGD can be E-optimal, provided a design exists for the E-best matrix.

**Example 3.10**

The E-best discrepancy matrix for \( v = 7, q = 5 \) has minimal \( a_{v} \)-value in (3.20) of \( a_{v} = 1.247 \). This matrix, call it \( \Delta_{opt} \), is

\[
\Delta_{opt} = \begin{pmatrix}
0 & -1 & -1 & 1 & 1 & 1 & 1 \\
-1 & 0 & -1 & 1 & 1 & 1 & 1 \\
-1 & -1 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & -1 & 0 & 0 \\
1 & 1 & 1 & -1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & -1 \\
1 & 1 & 1 & 0 & 0 & -1 & 0
\end{pmatrix}.
\]

\( \Delta_{opt} \) is potentially the basis for E-optimal design in \( D(7, 7 + 21m, 5) \) for each \( m \geq 0 \). The discrepancy matrix for the E-best RGD, displayed in Example 3.5, has \( a_{v} = 1.247 \), a clearly
inferior competitor. The design in Example 3.5 is nevertheless $E$-optimal over $D(7, 7, 5)$, in part because there is no competing design having discrepancy $\Delta_{opt}$.

The discrepancy matrix for the design in Example 3.2 is $\Delta_{opt}$, proving the design is $E$-optimal over $D(7, 28, 5)$ and that it is $E$-better than the best RGD, found by adding the 21 blocks of a BIBD $(7, 21, 5)$ to the Example 3.5 design. Adding $m - 1$ copies of the BIBD to the design in Example 3.2 produces an $E$-optimal design in $D(7, 7 + 21m, 5)$ for every $m \geq 1$. 

### TABLE 3.6

<table>
<thead>
<tr>
<th>$v$</th>
<th>$q$</th>
<th>Total Count</th>
<th>Count by Type</th>
<th>$v$</th>
<th>$q$</th>
<th>Total Count</th>
<th>Count by Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>$(1,0,1,0)$</td>
<td>12</td>
<td>1</td>
<td>11</td>
<td>$(1,0,10,0)$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
<td>$(1,0,0,0)$</td>
<td>2</td>
<td>2</td>
<td></td>
<td>$(1,0,1,0)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>4</td>
<td></td>
<td>$(0,0,4,0)$</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td>4</td>
<td>17</td>
<td>$(1,0,16,0)$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>$(1,0,2,0)$</td>
<td>5</td>
<td>56</td>
<td>$(7,0,49,0)$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$(1,0,1,0)$</td>
<td></td>
<td>6</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td>7</td>
<td>480</td>
<td>$(15,101,55,309)$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td>8</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>$(0,0,1,0)$</td>
<td>9</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>$(1,0,0,0)$</td>
<td>10</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>5</td>
<td>$(1,0,4,0)$</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$(1,0,1,0)$</td>
<td></td>
<td>14</td>
<td>1</td>
<td>15</td>
<td>$(1,0,14,0)$</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>$(2,0,4,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td>2</td>
<td>2</td>
<td>$(1,0,1,0)$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>$(4,0,9,0)$</td>
<td></td>
<td>5</td>
<td>1</td>
<td>$(0,0,1,0)$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td>10</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>7</td>
<td>$(1,0,6,0)$</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>$(1,0,1,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>$(1,0,6,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>$(1,0,2,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>89</td>
<td>$(6,26,52)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>$(0,0,1,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>$(1,0,7,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>$(1,0,0,0)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
With the $E$-best discrepancy matrices in hand, Morgan (2007b) was able to complete the $E$-optimality design problem for $k > 2$, $v \leq 15$; save a few open cases where the best design is yet to be determined. In every case, optimality was established over the full class $D(v, b, k)$, using Lemma 3.5 chiefly through Corollary 3.1. At this writing, the designs are available online at www.designtheory.org.

### 3.5 Small Block Designs

Three senses of what it means for an IBD to be “small” will be explored in this section. All are limiting cases: block size $k = 2$, number of blocks $b = 2$, and small number of units $bk$. Common to these situations is that at least some of the optimal designs have very different combinatorial properties than the designs seen in Section 3.4: they do not “approximate” BIBDs. The auxiliary parameters $r$, $p$, $\lambda$, and $q$, defined in (3.11), continue to be used here.

#### 3.5.1 Block Size Two

Block size two would, at first glance, appear to present an optimality problem that, if not simple, would at least be straightforward. This turns out to be decidedly untrue. We begin with an example.

**Example 3.11**

This design $d \in D(8, 8, 2)$ is both $A$-optimal and $D$-optimal:

\[
d : \begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 1
\end{array}
\]

However, the $A$-optimal design $d_1$ and the $D$-optimal design $d_2$ in $D(9, 9, 2)$ are quite different from each other:

\[
d_1 : \begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 1 & 1
\end{array}
\]

\[
d_2 : \begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 1 & 1
\end{array}
\]

Design $d_1$ in Example 3.11 is especially perplexing due to its grossly unequal replication counts in a setting with no plot excess ($p = 0$). The following theorems will show it is not an isolated example.

The designs in Example 3.11 are all binary, for good reason. One simplification afforded by $k = 2$ is that nonbinary designs are inadmissible, for it is easily demonstrated that a nonbinary block of size two makes no contribution to $C_d$. Thus, the blocks of a design are a selection, with replacement, of pairs of treatments. Any such design can be cast as an undirected multigraph whose vertices are the treatments and with the number of edges connecting vertices $i$ and $i'$ being the number of blocks consisting of treatments $i$ and $i'$. These concurrence graphs for two of the designs in Example 3.1 are displayed in Figure 3.1.
If the concurrence graph for a design has the same number of edges joining any two vertices, the design is a BIBD, universally optimal by Theorem 3.1. For a more thorough accounting of design from a graph perspective, see Bailey and Cameron (2013).

A few notions from graph theory are needed for results later in this section. The degree of a vertex is the number of edges incident with that vertex. A leaf is a vertex of degree one. A graph is connected if each vertex can be reached from any other vertex by a path. A bridge in a connected graph is an edge that, if removed, creates a disconnected graph. Any edge connected to a leaf is a bridge. A star is a collection of leaves incident with a common vertex. A cycle is a path with no repeated edges that begins and ends with the same vertex. A tree is a connected subgraph with no cycles. A spanning tree is a tree that includes every vertex in a graph.

Both graphs in Figure 3.1 are connected. In fact, connectedness of the designs (see Section 3.1) corresponds to connectedness of their concurrence graphs. The graph for $d_1$, which has five leaves and five bridges, is a star attached to a cycle. That for $d$ is a cycle.

**Theorem 3.10** (Cheng et al. 1985) For each given $v$, there is a $b_v$ such that for all $b \geq b_v$, a $D$-optimal design in $D(v, b, 2)$ is

(i) A RGD if $p = 0$;

(ii) A NBBD(1) if $p > 0$ and $\lambda = \lfloor (r+1)/(v-1) \rfloor$;

(iii) A NBBD(2) if $p > 0$ and $\lambda = \lfloor (r+1)/(v-1) \rfloor - 1$.

Theorem 3.10 tells us that, for $b$ sufficiently large, $D$-optimal designs with block size 2 must have the $(M, S)$-property. In particular, this says that the concurrences are being distributed across the treatment pairs as equally as possible. The graph-theoretic result of the next theorem makes a similar, but much broader, statement: it does not allow “too few” connections between any pair of treatments.

* This is true for the concurrence graph of any block design, not just those of block size two.
Theorem 3.11 (Bailey and Cameron 2013) For any $b \geq v$, the concurrence graph of a D-optimal design in $D(v, b, 2)$ contains no bridges and thus no leaves.

Theorem 3.11 becomes especially interesting when compared to Theorem 3.12.

Theorem 3.12 (Bailey 2007) For $20 \leq v \leq b \leq 5v/4$, any E-optimal design in $D(v, b, 2)$ has leaves.

A-optimal designs with $b$ not too much larger than $v$ must also have leaves; see Bailey and Cameron (2013). Thus, many A-optimal and many E-optimal designs with block size two, and with $r = \lfloor 2b/v \rfloor = 2$, have some treatments replicated once! This is in marked contrast to the D-optimality result with large $b$ in Theorem 3.10 and provides myriad examples where optimal designs cannot have $(M, S)$-property.

The smallest $b$ for $D(v, b, 2)$ where $p = 0$ (and connected designs exist) is $b = v$. Much is known about this setting. Parts (i) and (ii) of Theorem 3.13 are special cases of results cited at the end of Section 3.5.3; the E-optimality is from Bailey and Cameron (2009). Designs $d_1$ and $d_2$ of Example 3.11 illustrate Theorem 3.13.

Theorem 3.13 For $D(v, v, 2)$, $v \geq 3$,

(i) The D-optimal designs are the cycles.

(ii) The A-optimal designs are the cycles if $v \leq 8$, $v - 4$ leaves attached to one vertex of a square if $9 \leq v \leq 12$, and $v - 3$ leaves attached to one vertex of a triangle if $v \geq 12$.

(iii) The E-optimal designs are the cycles if $v \leq 6$, and $v - 3$ leaves attached to one vertex of a triangle, or a star with one doubled edge, if $v \geq 6$.

If $b$ is a multiple of $v(v - 1)/2$, then the optical designs are precisely the BIBDs. Otherwise, as $b$ grows larger with $p = 0$, RGDs are strong performers for which a number of optimality results are now known. Some of these are instances of results stated earlier in this article, and some have arisen from examination of small $v$. An E-optimality theorem for RGDs, whose proof relies heavily on Corollary 3.1, is stated next. Recall from Section 3.4.4 that $a_{dv}$ is the largest eigenvalue (other than the concurrence excess $q$) of the discrepancy matrix $\Delta_d$.

Theorem 3.14 (Bagchi and Cheng 1993) For $v \geq 9$ and $1 \leq q \leq (v + 2)/2$, $q \neq \frac{v}{2}$, any RGD with $a_{dv} = 1$ is E-optimal over $D(v, b, 2)$.

Bagchi and Cheng (1993) further established E-optimality over all of $D(v, b, 2)$ for RGDs with $v \leq 8$ and $q = v - 3$. Their results do not claim RGDs are uniquely E-optimal, and as evidenced by Table 3.6, it is not infrequent for certain non-RGDs, falling in the $q$ range of Theorem 3.14, to also be E-optimal. The paper Morgan (2007b), discussed in Section 3.4.4, determined all E-optimal designs in the binary, equireplicate subclass of $D(v, b, 2)$ for $v \leq 15$. 

© 2015 by Taylor & Francis Group, LLC
3.5.2 Two Blocks

Like block size two, with only two blocks, one might initially expect few complications in the optimality problem. Here, this expectation is met, up to a point. For IBDs, we have \( k < v \), and if \( 2k \leq v \), then all designs are disconnected, so take \( \frac{v}{2} < k < v \). The parameter \( r \) is thus \( r = \lfloor 2k/v \rfloor = 1 \) from which \( p = 2k - v \), that is, \( v = 2k - p \). Up to treatment relabeling, there is exactly one binary design (call it \( d_0 \), see Table 3.7), which consequently has \((M,S)\)-property. Design \( d_0 \) is a GGDD(3) with groups of size \( p, k - p, \) and \( k - p \) and treatment replication two within the size \( p \) group.

An arbitrary design in \( d \in D(v,2,k) \) must have at least \( v - p = 2(k - p) \) treatments replicated exactly once. Partition the treatment set \( V = V_1 \cup V_2 \cup V_3 \) where \( V_1 \) contains those of the singly replicated treatments that appear in the first block, \( V_2 \) the singly replicated treatments in the second block, and \( V_3 \) the remaining treatments. Averaging \( C_d \) within each of the \( V_i \) produces \( \bar{C}_d \) for a GGDD(3) whose optimality values are lower bounds (see Lemma 3.3) for those of \( d \). Comparing these to the values for \( d_0 \) yields Theorem 3.15.

**Theorem 3.15** (Morgan and Jin 2007) The unique binary design \( d_0 \) in \( D(v,2,k) \) is \( A \)-optimal and \( D \)-optimal.

Perhaps surprisingly, \( d_0 \) need not be the \( E \)-optimal design. Depending on \( k \), it can be inferior to the nonbinary design \( d^* \) displayed in Table 3.7. The result is stated as Theorem 3.16. The proof in the cited paper is quite lengthy, the averaging technique based on \( V_1 \cup V_2 \cup V_3 \) being inadequate to the task.

**Theorem 3.16** (Morgan and Jin 2007) In the class \( D(v,2,k) \), the binary design \( d_0 \) is the unique \( E \)-optimal design if \( \frac{v}{2} < k < \frac{5v}{6} \); the nonbinary design \( d^* \) is the unique \( E \)-optimal design if \( \frac{5v}{6} < k < v \); and both \( d^* \) and \( d_0 \), and only these two designs, are \( E \)-optimal if \( k = \frac{5v}{6} \).

The \( E \)-optimality of the designs in Table 3.5 is a consequence of Theorem 3.16.

**TABLE 3.7**

<table>
<thead>
<tr>
<th>Optimal Designs for ( D(v,2,k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_0 ):</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>\vdots</td>
</tr>
<tr>
<td>( p )</td>
</tr>
<tr>
<td>( p + 1 )</td>
</tr>
<tr>
<td>( p + 2 )</td>
</tr>
<tr>
<td>\vdots</td>
</tr>
<tr>
<td>( k )</td>
</tr>
<tr>
<td>( p )</td>
</tr>
<tr>
<td>( p )</td>
</tr>
<tr>
<td>( v )</td>
</tr>
<tr>
<td>( v )</td>
</tr>
<tr>
<td>( v )</td>
</tr>
</tbody>
</table>

\( © 2015 \) by Taylor & Francis Group, LLC
3.5.3 Minimally Connected Designs

The block design model (3.1) requires at least \( 1 + (v - 1) + (b - 1) = b + v - 1 \) degrees of freedom if all treatment contrasts are to be estimable. A block design produces \( bk \) total degrees of freedom, and thus connectedness requires that \( bk \geq b + v - 1 \). The designs in \( D(v, b, k) \) are minimally connected if \( bk = b + v - 1 \).

Minimal connectedness for block designs is analogous to “saturation” of fractional factorials (see Chapters 7 and 9), and the name saturated block designs is also used; in neither case are error degrees of freedom available. Minimally connected classes will be distinguished by the notation \( D(v, b, k)_m \). Lemma 3.6 provides basic properties for these classes.

**Lemma 3.6** For each \( d \in D(v, b, k)_m \), there is exactly one unbiased estimator for each treatment contrast. Consequently, any \( d \in D(v, b, k)_m \) is binary, and no pair of blocks has more than one treatment in common, that is, no pair of treatments occurs in more than one block.

**Proof:** Suppose there are two unbiased estimators \( p'y \) and \( q'y \) for \( c' \tau \). Let \( a = p - q \neq 0 \) and for the block design model (3.2), write \( A_d = (1, X_d, Z) \) and \( \theta' = (\mu, \tau', \beta') \). Then \( 0 = E(a'y) = a'A_d\theta \) for all \( \theta \Rightarrow a'A_d = 0' \Rightarrow bk > \text{rank}(A_d) = b + v - 1 \), a contradiction. \( \square \)

A design in \( D(v, b, k) \) is D-optimal if, and only if, its concurrence graph has maximal number of spanning trees over all designs in \( D \) (Cheng 1981). Lemma 3.6 tells us that the concurrence graph for a minimally connected design is a connected, simple graph with no cycles: it is itself a tree. This gives us this curious result:

**Theorem 3.17** (Bapat and Dey 1991) All designs in \( D(v, b, k)_m \) are D-optimal.

Other criteria do distinguish designs in \( D_m \). Theorem 3.18 will establish a version of M-optimality for a particular member of that class. Let \( v_{d1} \geq v_{d2} \geq \cdots \geq v_{dm} \) be estimator variances for a set of \( m \) contrasts of interest when estimated using design \( d \). Design \( d^* \) is M-optimal for estimation of the \( m \) specified contrasts if \( d^* \) minimizes \( \sum_{g=1}^{m} f(v_{dg}) \) for every monotonically increasing, convex function \( f \).

Let \( v_d \) be the \( m \times 1 \) vector of the \( v_{dg} \). A necessary and sufficient condition for \( M \)-optimality of \( d^* \) is that for every competing \( d \), \( v_d \) weakly submajorizes \( v_{d^*} \): \( \sum_{g=1}^{h} v_{dg} \geq \sum_{g=1}^{h} v_{dg^*} \) for each \( h = 1, \ldots, m \) (e.g., Bhatia 1997, p. 40). Authors like Bagchi and Bagchi (2001) have pursued M-optimality based on the canonical variances \( c_{di}^{-1} \) (in which case, type 1 optimality functions are subsumed; compare Definition 3.4). Here, \( v_d \) is taken as the vector of variances for the \( m = \binom{v}{2} \) pairwise comparisons \( \tau_i - \tau_j \).

The designs of interest have one treatment in every block, and all other treatments replicated once; see Table 3.8. The optimality result in Theorem 3.18 is a special case of a more general result proven in the cited paper.

**Theorem 3.18** (Jin and Morgan 2008) The designs in Table 3.8 are uniquely M-optimal for estimation of pairwise contrasts.

Now the \( A \)-value is proportional to the average of the variances of the \( \tau_i - \tau_j \), and the \( MV \)-value is the maximum of those variances. Thus, among the results embedded in
Theorem 3.18 are that the designs in Table 3.8 are $A$-optimal and $MV$-optimal over $D_m$. These two individual optimalities were first proven by Mandal et al. (1990/1991) and Dey et al. (1995). Bapat and Dey (1991) prove Table 3.8 designs are $E$-optimal.

Table 3.8 designs are maximal trace GGDDs, with $b+1$ groups and with smallest positive information matrix eigenvalue $1/k$. They maximize, rather than minimize, the replication range $(\max_i r_i - \min_i r_i)$ over $D_m$ and, so like a number of other small designs seen in Sections 3.5.1 and 3.5.2, are well off from having the $(M,S)$-property.

If $bk = b + v$, then one degree of freedom is available for error estimation. Optimal designs are also known for this “saturated plus one” setting; not surprisingly, they are a bit more complicated than those in $D_m$. See Balasubramanian and Dey (1996) for $D$-optimality, Krafft and Schaefer (1997) for $A$-optimality, and Dey et al. (1995) for $MV$-optimality. For the $E$-optimality problem, this author is not aware of a general solution, though Theorem 3.13 and Theorem 3.16 with $v = 2(k - 1)$ fall into this category.

### 3.6 Multiple Blocking Factors

The line-segment experiment described in Section 3.1 motivated variance reduction through blocking by subject. Because subject performance may change with practice, it could also be useful to block on the order of presentation of sheets to subjects. If this is done, the design selection problem will need to consider allocation of treatments relative to two blocking factors: subjects and time order. In agricultural experimentation to compare grain varieties, fields of the varieties are typically planted at several locations, and at each location, several harvesting teams may be employed. Now both location and harvesting teams are obvious blocking factors, for both are likely to contribute substantial variation to observed yields. These are just two examples of how more than one blocking factor can be incorporated when designing an experiment, in order to eliminate more than one major source of variability in experimental material.

This section focuses chiefly on the most commonly employed variants of two blocking factors, before briefly exploring some of the other possibilities. Let $F_1$ and $F_2$ denote two factors. Two broad design classes can be defined depending on the structural relationship, nested or crossed, between $F_1$ and $F_2$.

**Definition 3.6** Blocking factor $F_1$ is said to nest blocking factor $F_2$ if any two experimental units in an $F_2$-block are in the same $F_1$-block. If $F_1$ nests $F_2$, then $F_2$ is nested in $F_1$. 

![Table 3.8](image-url)
TABLE 3.9
Nested Block Design for 12 Treatments in Blocks of Size Six

<table>
<thead>
<tr>
<th>1</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Nested block designs are those with two blocking factors, one nesting the other. Members of the two blocking systems in a nested design are descriptively referred to as superblocks and subblocks. In the agricultural experiment described earlier, locations would be superblocks, and within each location, the harvesting teams would form subblocks. The design in Table 3.9 has four superblocks, each nesting two subblocks (four locations, two teams at each location). Nested block designs are taken up in Sections 3.6.1 and 3.6.2.

The model for observations generated with any nested design is a simple extension of (3.2) for standard block designs. Let \( b_1 \) be the number of superblocks and \( b_2 \) the number of sublocks per superblock, now using \( b = b_1 b_2 \) for the total number of subblocks. Let \( k \) be the number of experimental units per subblock, so that the superblock size is \( k_1 = b_2 k \). The class of connected nested designs is denoted by \( D(v, b_1, b_2, k) \). The model for \( d \in D(v, b_1, b_2, k) \) is

\[
y = \mu 1 + X_d \tau + Z_1 \beta_1 + Z_2 \beta_2 + \epsilon,
\]

where \( \beta_1 \) and \( \beta_2 \) are the \( b_1 \times 1 \) and \( b_1 b_2 \times 1 \) vectors of parameters for the nesting and nested blocks, respectively. With suitable ordering of the observations \( y \) and using the Kronecker product \( \otimes \), the corresponding incidence matrices are \( Z_1 = I_{b_1} \otimes 1_{k_1} \) and \( Z_2 = I_b \otimes 1_k \).

**Definition 3.7** Blocking factors \( F_1 \) and \( F_2 \) are crossed if each \( F_1 \)-block intersects each \( F_2 \)-block in the same number of experimental units.

Designs with two crossed blocking factors are termed row–column designs, another descriptive name; see Table 3.10. Members of the two blocking systems are now termed row blocks and column blocks, and the intersection of a row block with a column block is a cell. As is most frequently encountered, the designs in Table 3.10 implement the row–column idea with one unit per cell. More precisely, these are complete crosses. Readers interested in incomplete crosses, having either empty cells or unequal numbers of units per cell, are referred to Saharay (1996) as a starting point. Row–column designs are the topic of Section 3.6.3. If the line-segment experiment were to be run as a row–column design, the columns would denote subjects, and the rows would denote order of presentation.

The standard model for row–column designs can also be written in the form (3.22), where now \( \beta_1 \) and \( \beta_2 \) are the \( b_1 \times 1 \) and \( b_2 \times 1 \) vectors of parameters for the row blocks and column
blocks, respectively. The two design models are distinguished by how their blocking structures are reflected in \( Z_1 \) and \( Z_2 \). With one unit per cell and with observations in row-major order, they are now \( Z_1 = I_{b_1} \otimes I_{b_2} \) and \( Z_2 = I_{b_1} \otimes I_{b_2} \).

Though substantial, the optimality theory for designs with more than one blocking factor is not as well developed as that in Sections 3.3 through 3.5. The subsections to follow will report the main findings to date. Consistent with earlier sections in this chapter, the number of experimental units in a block is taken to be constant across blocks of the same blocking factor.

### 3.6.1 Resolvable Block Designs

Various classes of nested designs are defined by various requirements placed on the nesting blocks. A **resolvable block design** is a nested block design for which the superblocks are complete, that is, for which the block design defined by the nesting factor is a RCBD. The design in Table 3.9 is a resolvable block design: the superblocks resolve the subblocks into complete replicates of the treatment set. Resolvable designs are, by a wide margin, the most frequently employed nested designs in practice and have been extensively employed in agricultural field trials. The Table 3.9 design, for instance, could be used for a trial comparing 12 crop varieties over four locations (superblocks), using two harvesting teams (subblocks) at each location.

Writing \( Z = (Z_1, Z_2) \), the information matrix arising from (3.22) for estimation of \( \tau \) (see following (3.2)) is \( C_d = X_d'(I - P_Z)X_d \). The nesting relationship says the column space of \( Z_1 \) is contained in the column space of \( Z_2 \) so that \( P_Z = P_{Z_2} \) (superblock effects are “washed out” by subblock effects). Consequently, \( C_d \) reduces to \( C_d = X_d'(I - P_{Z_2})X_d = R_d - \frac{1}{k} N_d N_d' \), where \( N_d \) is the \((v \times b)\) treatments \( \times \) subblock incidence matrix. It follows that a nested design is connected (i.e., all treatment contrasts are estimable) if and only if the subblock design is connected.

**Lemma 3.7** If a nested block design is \( \Phi \)-optimal as a member of \( D(v,b,k) \) when ignoring superblocks, then it is \( \Phi \)-optimal over \( D(v,b_1,b_2,k) \).

Select any optimal design in \( D(v,b,k) \), then partition its \( b = b_1 b_2 \) blocks into \( b_1 \) sets of \( b_2 \) blocks each. Lemma 3.7 says this is, an optimal design in \( D(v,b_1,b_2,k) \). If \( b_2 k = v \)
and each treatment appears in each superblock, this is an optimal resolvable design. Simply put, the superblocks are irrelevant to the basic optimality argument. Nested designs would present no new optimality problems were special demands not placed on the nesting blocks. Those demands, though not reflected in the information matrix, are intended to incorporate pragmatic issues with experiment execution. Nesting factors often represent a grouping of subblocks that are similar due to closeness in time or space or that are subject to management by different personnel. A discussion of why nested designs are employed, with several examples, can be found in Section 1 of Morgan (1996).

Superblocks for resolvable designs are complete blocks. Resolvability assures that any problem associated with the nesting factor will impact every treatment equally, with possibilities ranging from model shortcomings at the nesting level to loss of entire superblocks during experimentation. If, in the preceding crop example, disease ruins yields at one location, the remaining locations still give the same amount of data on each variety. Resolvability imparts robustness to otherwise unforeseen difficulties at the superblock level.

The lesson that should be taken from Lemma 3.7 is that although the criterion value for a nested design is precisely that of its subblock design in $D(v, b, k)$, the superblock requirements restrict the designs in $D(v, b, k)$ that are even considered. The search for an optimal design in $D(v, b_1, b_2, k)$ is the search for an optimal design over a subclass of $D(v, b, k)$. The optimal member of the subclass may, or may not, coincide with an optimal design over all of $D(v, b, k)$. To give one example, the $A$-optimal design in $D(12, 8, 6)$, which is not resolvable, is easily shown to be better than the optimal resolvable design displayed in Table 3.9 with respect to all generalized optimality criteria.

Henceforth, we use $r = bk/v$ in lieu of the number of superblocks $b_1$ for resolvable designs and $s = v/k$ in lieu of the number of subblocks $b_2$ per superblock. The class of resolvable IBs is denoted by $D_R(v, r, s, k)$. A simple instance where optimality does coincide for $D_R(v, r, s, k)$ and $D(v, b, k)$ is when the latter contains a BIBD that is resolvable. For instance, take the ten blocks in Example 3.1, and add ten blocks given by their set complements, to get a BIBD in $D_R(6, 10, 2, 3)$ that is universally optimal over the larger class $D(6, 20, 3)$. There is an extensive literature on resolvable BIBDs that will not be covered here; see Chapter II.7 of Colbourn and Dinitz (2007) for an introduction including many existence results.

### 3.6.1.1 Affine Resolvable Designs

The affine resolvable designs provide a very rich class of optimal resolvable designs.

**Definition 3.8** A resolvable design is affine resolvable if any two subblocks in different superblocks intersect in the same number of treatments.

Denote the common subblock intersection number by $\omega$. Then $k = \omega s$ so that $v = sk = \omega s^2$ is necessarily divisible by a square for an affine resolvable design. This observation points to an interesting combinatorial relationship. Given an affine resolvable design $d$, create a $v \times r$ array, call it $B_d$, as follows. Rows of $B_d$ correspond to treatments of $d$. Columns of $B_d$ correspond to superblocks (replicates) of $d$. Labeling the subblocks in superblock $j$ by $1, 2, \ldots, s$ in any order, the $(i, j)$ entry of $B_d$ is the label of the subblock in superblock $j$ that contains treatment $i$. Affineness of $d$ implies that any ordered pair of symbols from $\{1, \ldots, s\}$
can be found as rows of any two-columned subarray of $B_d$ exactly $\omega$ times. Thus, $B_d$ is an orthogonal array (see Chapter 9) of strength two. Conversely, given a strength-two orthogonal array, the identification can be reversed to create an affine resolvable design. Affine resolvable designs are therefore \textit{combinatorially equivalent} to strength-two orthogonal arrays, a fact that will shortly prove useful.

The importance of affine resolvable designs is made clear in Theorem 3.19.

\textbf{Theorem 3.19} (Bailey et al. 1995) Any affine resolvable design in $D_R(v, r, s, k)$ is $M$-optimal based on canonical variances. In particular, it is generalized optimal over $D_R(v, r, s, k)$.

Also notice that if replicates (superblocks) are lost from an affine resolvable design, the remaining design is still affine resolvable, and so still optimal in all of the senses of Theorem 3.19.

Strength-two orthogonal arrays, and thus affine resolvable designs, have been extensively studied. Relying on that work, Bailey et al. (1995) established existence of affine resolvable designs with up to seven replicates for every $v = \omega s^2$, $\omega \geq 2$. The design in Table 3.9 is affine resolvable. Designs with $\omega = 1$ are also known as square lattices; see Section 3.6.1.2.

$MV$-optimality is not addressed by Theorem 3.19, and despite the very strong optimality established there, an affine resolvable design need not be an $MV$-optimal resolvable design (Morgan 2008). Even among nonisomorphic affine resolvable designs in $D_R(v, r, s, k)$, $MV$ behavior can vary. Morgan (2010) determined $MV$-best affine resolvable designs based on an aberration criterion for $v$ of the form $\omega 2^k$. There has otherwise been little work on $MV$-optimality of resolvable designs, affine or not.

\textbf{3.6.1.2 Other Resolvable Designs}

The earliest resolvable designs to be rigorously studied in the statistical literature were the lattices, introduced by Yates (1936b). The \textit{m-dimensional lattices} for $v = t^m$ treatments have $k = t^u$ and $s = t^{m-u}$. The square lattices are those with $(m, u) = (2, 1)$, corresponding to $\omega = 1$ in Section 3.6.1.1. For $b \leq v$, Cheng and Bailey (1991) proved generalized optimality of square lattices over all binary, equireplicate competitors (not just resolvable designs). Chapter 18 of Hinkelmann and Kempthorne (2005) provides a good introduction to lattice designs. Also covered there are the \textit{rectangular lattices} having $k = s - 1$, known to be highly efficient resolvable designs (Bailey and Speed 1986). Designs in this and the next paragraph are also discussed in Section 1.6.2.

Patterson and Williams (1976a) introduced the \textit{alpha designs} as a flexible method for building resolvable designs having $k < s$. Particular members of this class are known to be highly efficient or optimal, but the technique is too general to admit a comprehensive optimality theory. A readable account of alpha design development and application, with many references, is given by Paterson (1988).

There has been productive work on small resolvable designs. Patterson and Williams (1976b) established a useful link between optimal, binary, equireplicate members of $D(s, s, k)$ and optimal resolvable designs in $D_R(v, 2, s, k)$ having two replicates, showing how the former can induce the latter. Efficient resolvable designs having three replicates are provided in Bailey (2005). Morgan and Reck (2007b) established optimality for many classes of resolvable designs having two blocks per replicate.
3.6.2 Other Nested Block Designs

Relative to resolvable designs, there are two directions to explore for the nesting factor: superblocks of size less than \( v \) and of size greater than \( v \). Although each of these directions offers multiple possibilities, there has been one main line pursued for each in the statistical literature.

For \( b_2k < v \), the emphasis has been on the nested BIBDs (NBIBDs). A nested design is a NBIBD if the subblock design is a BIBD in \( D(v, b, k) \), and the superblock design is a BIBD in \( D(v, b_1, b_2k) \). Lemma 3.7 says that a NBIBD is universally optimal over \( D(v, b_1, b_2, k) \). Introduced by Preece (1967), an extensive survey for this topic including a tabling of smaller designs is available in Chapter VI.36 of Colbourn and Dinitz (2007). The BIBD structure at the superblock level provides robustness by ensuring a reasonably good design should a superblock be lost, the remaining design being GGDD(2). Moreover, NBIBDs were proven universally optimal by Morgan (1996) when recovering both subblock and superblock information.

For \( b_2k > v \), several authors have investigated \( \alpha \)-resolvable block designs. The distinguishing characteristic of these nested designs is that the superblocks are multiply complete, with each treatment applied to \( \alpha \)-experimental units in each superblock. Section 3.3 of Caliński and Kageyama (1996) contains an introductory treatment of \( \alpha \)-resolvability, related extensions, and attendant results, to which the reader is referred for further details. Of special interest are the affine \( \alpha \)-resolvable designs, which have any two subblocks intersecting in either \( \omega_1 \) or \( \omega_2 \) treatments depending on whether they are in the same, or different, superblocks. The technique employed by Bailey et al. (1995) to prove \( M \)-optimality of affine resolvable designs (Theorem 3.19) can be extended to include the affine \( \alpha \)-resolvable designs.

3.6.3 Row–Column Designs

As indicated by Definition 3.7 and Table 3.10, row–column designs are represented as rectangular arrays in which rows correspond to levels of a blocking factor \( F_1 \) with \( b_1 \) (say) levels, columns correspond to levels of another blocking factor \( F_2 \) with \( b_2 \) levels, cells correspond to experimental units, and numbers in cells are labels for the treatments assigned. Row–column designs are randomized by selecting random permutations of rows, of columns, and of treatment labels, though there are variants on this (e.g., Section 10.2.2 of Hinkelmann and Kempthorne 2008). Row–column designs are widely used for experimentation in many disparate fields including agriculture (where they were first developed), psychology, and industry. Any general, applied textbook on experimental design will offer multiple examples.

Setting \( Z = (Z_1, Z_2) \), the information matrix \( C_d = X_d'(I - P_Z)X_d \) from the row–column version of model (3.22) simplifies to

\[
C_d = R_d - \frac{1}{b_1}N_dN'_d - \frac{1}{b_2}M_dM'_d + \frac{1}{b_1b_2}r_dr'_d,
\]

in which \( N_d = ((n_{dij})) = X_d'Z_2 \) is the treatment/column block incidence matrix and \( M_d = ((m_{dil})) = X_d'Z_1 \) is the treatment/row block incidence matrix (compare (3.4)).

The class of connected row–column designs is denoted \( D_{RC}(v, b_1, b_2) \). The row–column design optimality problem amounts to minimizing \( \Phi(C_d) \) over \( D_{RC}(v, b_1, b_2) \) for one or more of the usual criteria \( \Phi \). Any \( d \in D_{RC}(v, b_1, b_2) \) has two component block designs,
Suppose \( d_0 \in \mathcal{D}(v, b_2, b_1) \) is \( \Phi \)-optimal, is equireplicate, and the number of blocks \( b_2 \) for \( d_0 \) is an integer multiple of \( v \). Then, there is a \( \Phi \)-optimal row–column design \( d^* \in \mathcal{D}_{RC}(v, b_1, b_2) \) for which the column component design \( d^*_{C} \) is \( d_0 \).

**Proof:** For any \( d \in \mathcal{D}_{RC}(v, b_1, b_2) \), the matrix \( M_d M'_d - \frac{1}{b_1} r_d r'_d \) is nonnegative definite. That is, \( C_{d_c} - C_d \) is nonnegative definite and thus \( \Phi(C_{d_c}) \leq \Phi(C_d) \) (see (3.3)(i)).

Write the blocks of \( d_0 \) as adjacent columns to give a \( b_1 \times b_2 \) array. Now permute treatments within each column as necessary to achieve equal replication of treatments within each row. The resulting design, which is \( d^* \) of the theorem, is said to be row regular. That row regularity can be achieved whenever an equireplicate block design has number of blocks a multiple of \( v \) is guaranteed by the basic theory for systems of distinct representatives (e.g., Chapter 5 of van Lint and Wilson 2001). Easily checked is that row regularity implies

\[
M_d M'_d - \frac{1}{b_1} r_d r'_d = 0
\]

for any design \( d \), so that \( C_{d^*} \) reduces to \( C_{d_0} \). Thus,

\[
\Phi(C_{d^*}) = \Phi(C_{d_0}) \leq \Phi(C_{d_c}) \leq \Phi(C_d),
\]

for any \( d \in \mathcal{D}_{RC}(v, b_1, b_2) \).

Theorem 3.20 presents a simple, effective method for obtaining optimal row–column designs. The most popular row–column designs in practice fit into its framework. A Latin square design \( d \in \mathcal{D}_{RC}(v, v, v) \) is a row–column design \( d \) for which \( d_c \) is a RCBD and \( d_R \) is a RCBD. A Youden design \( d \in \mathcal{D}_{RC}(v, k, v) \) is a row–column design \( d \) for which \( d_c \) is a BIBD, and \( d_R \) is a RCBD. By Theorems 3.1 and 3.20, Latin square and Youden designs are universally optimal in their respective row–column classes. By the same reasoning, all generalized Latin squares, for which each of \( d_c \) and \( d_R \) is either a complete or a generalized complete block design, are universally optimal. A recent introductory review of Latin squares from the statistician’s viewpoint appears in Morgan (2007a). The 4 × 4 design in Table 3.10 is a Latin square design.

Also included in Theorem 3.20 are row–column designs where the row component is a complete block design and the column component is an optimal IBD, these being optimal cases of Latin rectangles. The 3 × 12 design in Table 3.10 falls into this category, it being a row-regular arrangement of an A- and E-optimal IBD in \( \mathcal{D}(6, 12, 3) \). This design would be appropriate for a version of the line-segment experiment as described at the start of Section 3.6, employing twelve subjects (columns) with three presentations of pictures (rows) per subject.

A generalized Youden design (GYD) \( d \in \mathcal{D}_{RC}(v, b_1, b_2) \) is a row–column design \( d \) for which \( d_c \) is a BIBD or a KBBD (see Section 3.3) and \( d_R \) is either a complete block, a generalized complete block, or a KBBD, design. Universal optimality of the row-regular GYDs is immediate. Other GYDs, however, are not universally optimal (Das and Dey 1992). Kiefer (1975), employing a very delicate argument, was nonetheless able to prove that all “double KBBD” GYDs, in which both \( d_R \) and \( d_C \) are KBBDs, are both A-optimal and D-optimal, the latter
being for \( v > 4 \). A tabling of GYDs is available in Ash (1981). A nonregular \( 6 \times 6 \) GYD for four treatments appears in Table 3.10.

Other results employing Theorem 3.20 are easily written. For instance, any GDD in \( D(v, mv, k) \) with \( \lambda_2 = \lambda_1 + 1 \) can be arranged as a row–column design that is generalized optimal over \( D_{RC}(v, k, mv) \). However, the row–column design problem becomes quite difficult outside settings where row regularity is possible. Kiefer’s optimality proof for GYDs that lack row regularity is highly nontrivial, even with those designs having the best component designs \( d_R \) and \( d_C \) possible. The technical difficulties become only more complex for settings where “nice” component designs are not available. The reader is referred to Chapter 4 of Shah and Sinha (1989) for a review of relevant (and limited) results prior to 1989. The remainder of this section will highlight the most attractive of the sparse set of results obtained since. This is an area sorely in need of further development.

Some insight is gained by rewriting the information matrix (3.23) in terms of the component designs. Let \( C_{d_0} \) denote the information matrix for a CRD having the same replication numbers as \( d \). Then, \( C_d = R_d - \frac{1}{b_1b_2} r_d r_d' \) and

\[
C_d = C_{d_R} + C_{d_C} - C_{d_0},
\]

(3.24)

which tempts one to optimize \( d_R \) and \( d_C \) separately. Unfortunately, this will not generally work because, for instance, eigenvalues for the individual information matrices may have no specific relationship to those for \( C_d \). Commuting matrices, however, share their eigenspaces, which for equireplicate designs suggests investigating \( d_R \) and \( d_C \) for which \( C_{d_R} \) and \( C_{d_C} \) commute. The condition for this turns out to be that \( M_{d}'N_d \) is a multiple of \( J_{b_1 \times b_2} \) (Shah and Eccleston 1986), termed adjusted orthogonality. The challenge is to show that an adjusted orthogonal design with optimal \( d_R \) and optimal \( d_C \) is also optimal among a relevant class of designs that is not so restricted. One interesting result of this type is stated in Theorem 3.21.

**Theorem 3.21** (Bagchi and Shah 1989) An adjusted orthogonal row–column design in \( D(v, b_1, b_2) \) for which \( d_R \) and \( d_C \) are each linked block designs is M-optimal over all equireplicate competitors.

Linked block design are defined in Section 3.4.4. Bagchi and Shah (1989) also show how to construct designs that satisfy Theorem 3.21. Bagchi and van Berkum (1991) prove E-optimality over all of \( D_{RC}(v, b_1, b_2) \) for adjusted orthogonal designs with \( d_C \) and \( d_R \) being a linked block design and an affine resolvable design.

Row–column designs with two rows carry special interest due to their use in two-color microarray experiments (Bailey 2007). The following is a recent result along these lines:

**Theorem 3.22** (Chai and Cheng 2011) If \( v \) is even and \( 2b/v \) is an odd integer, then a design for which \( d_C \) is a BIBD, and for which the row replication numbers satisfy \( |m_{di1} - m_{di2}| = 1 \) for all \( i \), is A-optimal over \( D_{RC}(v, 2, b) \).
the row replication difference \( m_{d11} - m_{d22} \) depending on group membership. Microarray experiments have also sparked interest in saturated row–column designs; see Chapter 23 and Qu (2010).

On the theme of “small” row–column designs, the optimality question has been thoroughly investigated for \( v = 3 \) in the pair of papers by Morgan and Parvu (2007) and Parvu and Morgan (2008). The results illustrate well how optimizing \( d_r \) and \( d_c \) can, depending on \( b_1 \) and \( b_2 \), sometimes produce optimal, and sometimes decidedly suboptimal, designs. A little additional terminology is needed before describing the most interesting conclusions.

Let treatment \( i \) be replicated \( r_{di} \) times. The assignment of treatment \( i \) to units is uniform in rows if \( m_{dii} \in \{r_{di}/b_1, [r_{di}/b_1] + 1\} \) for \( l = 1, \ldots, b_1 \), and uniform in columns if \( m_{dij} \in \{[r_{di}/b_2], [r_{di}/b_2] + 1\} \) for \( j = 1, \ldots, b_2 \). Row regularity in Theorem 3.20 is the special case of row uniformity for all treatments in an equireplicate design with \( b_2 \) an integer multiple of \( v \), while row uniformity for all treatments without row regularity is invoked in Theorem 3.22. A design that is uniform in both rows and columns for all treatments is said to be uniform. A row–column design has maximin replication if \( \min r_{di} = [b_1b_2/v] \). A design is \( E_M \)-optimal if it is \( E \)-optimal, and if among all \( E \)-optimal designs, it is \( M \)-optimal.

Morgan and Parvu (2007) prove that an \( A \)-optimal design in \( \mathcal{D}_{RC}(3, b_1, b_2) \) must be uniform, but in some cases cannot have maximin replication. On the other hand, Parvu and Morgan (2008) prove that, aside from \( 4 \times 4 \) layouts, an \( E_M \)-optimal design must have maximin replication, but in some cases cannot be uniform. For instance (see the papers for full results), if \( b_1 \equiv 1 \pmod 3 \) and \( b_2 \equiv 1 \pmod 3 \), then \( A \)-optimal designs are uniform with replications \( r_{d1} = b_1(b_2 + 2)/3 \) and \( r_{d2} = r_{d3} = r_{d1} - b_1 \). For the same setting, \( E_M \)-optimal designs have \( r_{d2} = r_{d3} = r_{d1} - 1 = (b_1b_2 - 1)/3 \), but nonuniformity in all but the smallest cases.

Example 3.12

The two designs \( d_1, d_2 \in \mathcal{D}_{RC}(3,7,7) \) shown here are, respectively, \( A \)-optimal and \( E_M \)-optimal. The replication counts \((r_1,r_2,r_3)\) are \((21,14,14)\) for \( d_1 \) and \((17,16,16)\) for \( d_2 \). With their given replications, \( d_1 \) is uniform, but \( d_2 \) has nonuniformity of treatment 1 in the first column.

\[
\begin{array}{cccccc}
1 & 1 & 1 & 2 & 2 & 3 \\
1 & 2 & 3 & 1 & 3 & 1 \\
1 & 2 & 3 & 2 & 3 & 1 \\
\hline
2 & 3 & 1 & 1 & 1 & 2 \\
2 & 3 & 3 & 1 & 2 & 1 \\
3 & 1 & 2 & 1 & 2 & 1 \\
\hline
3 & 1 & 2 & 3 & 1 & 2 \\
\end{array}
\]

\[
\begin{array}{cccccc}
1 & 2 & 1 & 1 & 2 & 3 \\
2 & 1 & 3 & 3 & 1 & 2 \\
2 & 2 & 1 & 3 & 1 & 3 \\
\hline
2 & 3 & 3 & 1 & 1 & 2 \\
3 & 1 & 2 & 2 & 1 & 1 \\
3 & 2 & 1 & 2 & 3 & 1 \\
\hline
3 & 3 & 2 & 1 & 3 & 1 \\
\end{array}
\]

3.6.4 Designs with Nesting and Crossing

Once one steps away from the basic settings with two crossed, or two nested, blocking factors, the possibilities grow rapidly. This section examines just two of these, each with three blocking factors and each with both nesting and crossing. See Morgan and Bailey (2000) for a systematic approach to optimal design with even more blocking factors.

© 2015 by Taylor & Francis Group, LLC
3.6.4.1 Semi-Latin Squares

Row–column designs are not limited to having one experimental unit for each row/column intersection. With \( k > 1 \) units per cell, a cell effect can be included in the model so that “cell” becomes a third blocking factor. Because cell blocks are nested within both row blocks and column blocks, a result much like Lemma 3.7 entails: if the cell component block design \( d_{\text{cell}} \) is optimal over \( D(v, b_1 b_2, k) \), then the row–column design is optimal. Connectedness, too, is determined solely by the cell blocks. There is no new issue here unless, as in Sections 3.6.1 and 3.6.2, there are other demands placed on the nesting factors.

A row–column design is doubly resolvable if each treatment is replicated once in each row block and once in each column block. A Latin square is trivially a doubly resolvable design with \( b_1 = b_2 = v \) and \( k = 1 \) unit per cell. The doubly resolvable row–column designs with \( b_1 = b_2 \) and \( k > 1 \) are termed semi-Latin squares. Writing \( b \) for the common number of rows/columns, the class of semi-Latin squares for \( v = b k \) treatments is denoted \( D_{\text{SL}}(b, k) \).

A member of \( D_{\text{SL}}(5, 3) \) is displayed in Table 3.11. An excellent introduction to semi-Latin squares with many examples of their use, in areas as widespread as consumer testing, agriculture, and message authentication, is provided by Bailey (1992).

For a thorough accounting of MOLS with a table of squares of small side, see Chapter III.3 of Colbourn and Dinitz (2007). The semi-Latin square in Table 3.11 is a Trojan square. Cheng and Bailey (1991) (using Theorem 3.5 in Section 3.4) proved that Trojan squares are generalized optimal and so \( A-, D-, \) and \( E- \)-optimal, over \( D_{\text{SL}}(b, k) \) and indeed over the entire binary, equireplicate subclass of \( D(b k, b^2, k) \). Thus, Trojan squares are preferred whenever they can be found.

There are at most \( b - 1 \) MOLS of side \( b \), and a full set of \( b - 1 \) can be found whenever \( b \) is a prime or a power of a prime. This gives optimal Trojan squares in \( D_{\text{SL}}(4, k \leq 3), D_{\text{SL}}(5, k \leq 4), D_{\text{SL}}(7, k \leq 6), D_{\text{SL}}(8, k \leq 7), \) and \( D_{\text{SL}}(9, k \leq 8) \). There is no pair of MOLS of side 6 and so no Trojan square in \( D_{\text{SL}}(6, 2) \). Bailey and Royle (1997) determined the \( A \)-best, \( D \)-best, and \( E \)-best designs in the subclass of \( D_{\text{SL}}(6, 2) \) having \((M, S)\)-property (all \( \lambda_{d_{\text{cell}}} \in \{0, 1\})\). These designs and all semi-Latin squares with no treatment concurrence exceeding 1 are called SOMAs (an acronym for simple orthogonal multiarray). Whether or not SOMAs are optimal over all of \( D_{\text{SL}}(6, 2) \) remains to be seen. More recently, Soicher (2013) has determined highly efficient squares in \( D_{\text{SL}}(6, k) \) for \( 4 \leq k \leq 10 \). A complete

<table>
<thead>
<tr>
<th>Table 3.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 × 5 Semi-Latin Square for 15 Treatments in Blocks of 3</td>
</tr>
<tr>
<td>1,6,11</td>
</tr>
<tr>
<td>5,9,13</td>
</tr>
<tr>
<td>4,7,15</td>
</tr>
<tr>
<td>3,10,12</td>
</tr>
<tr>
<td>2,8,14</td>
</tr>
</tbody>
</table>

© 2015 by Taylor & Francis Group, LLC
enumeration in the same paper resulted in $A_-, D_-$, and $E$-optimal members of $D_{SL}(4, k)$ for $4 \leq k \leq 10$. Soicher (2012) proved $M$-optimality of any semi-Latin square for which every two cells in different rows and columns intersect in the same number of treatments; this may be thought of as an extension of Theorem 3.19 to the doubly resolvable setup.

3.6.4.2 Nested Row–Column Designs

Just as row–column designs are not limited to one unit per cell, they are not limited to a single row–column layout. Nested row–column designs are seen in agricultural field trials where there is a separate row–column layout blocking on two sources of variation in each of several fields. They arise whenever a row–column experiment is repeated through time or space and row (column) blocks are unlikely to exert the same effects at each repetition. A nested row–column design can be as simple as two Latin squares, nesting $v$ rows and $v$ columns in each of the squares. The general setup is for $b$ separate $b_1 \times b_2$ layouts, each now called a block, with one experimental unit per cell (a total of $bb_1b_2$ units). The class of all connected designs is denoted by $D_{RC}(v, b, b_1 \times b_2)$. Table 3.12 displays two designs in $D_{RC}(5, 10, 2 \times 2)$.

Theorem 3.23 (Bagchi et al. 1990) Suppose there is a design $d^* \in D_{RC}(v, b, b_1 \times b_2)$ for which (i) each block is row regular for the treatments it contains and (ii) the column blocks are a $\Phi$-optimal design in $D(v, bb_2, b_1)$. Then $d^*$ is $\Phi$-optimal over $D_{RC}(v, b, b_1 \times b_2)$.

So, for instance, if the $bb_2$ columns are the blocks of a BIBD, a row-regular nested row-column design is universally optimal. This is the case with design $d_2$ in Table 3.12. Likewise, nests of Latin squares, of Youden designs, and of row-regular GYDs are universally optimal. It is known that a nest of GYDs that are not row regular can be suboptimal; see Morgan (1997). Theorem 3.23 is an extension of Theorem 3.20 to the nested setup.

Curiously, if rows, columns, and blocks exert random effects, then an analysis recovering information from the three blocking strata may lead to a design that does not have the row-regular property of Theorem 3.23. The three component designs (blocks, row blocks, column blocks) of design $d_1$ in Table 3.12 are each BIBDs, together comprising a BIBD with nested rows and columns (BBRC). Morgan and Uddin (1993) give conditions on the variances of the various blocking effects under which a BBRC will outperform a design that is optimal by Theorem 3.23, in which case $d_1$ is preferred to $d_2$ in Table 3.12. Many authors have worked on constructing IBs of both types, though progress has slowed of late. Especially as regards smaller designs (small $b$ and $b_1b_2 \leq v$), there are few optimality results known, though this is not surprising given the state of knowledge for ordinary ($b = 1$)

| TABLE 3.12 |
| Two Nested Row–Column Designs for Five Treatments in Blocks of Size $2 \times 2$ |
| $d_1$: 1 5 1 4 1 3 1 2 2 3 3 1 5 1 2 1 4 1 4 2 |
| 3 2 5 3 2 4 4 5 4 5 2 5 3 4 4 3 5 2 5 3 |
| $d_2$: 1 2 1 3 1 4 1 5 2 3 2 4 2 5 3 4 3 5 4 5 |
| 2 1 3 1 4 1 5 3 2 4 2 5 4 3 5 3 5 4 |

© 2015 by Taylor & Francis Group, LLC
row–column designs. Bagchi and Bose (2008) developed optimal main effects plans in small nested row–column designs. A number of results are summarized in the survey Morgan (1996).

3.7 Control Treatments and Weighted Optimality

Bailey (2008) described an experiment evaluating impact of several fungicide application regimens for winter wheat. Six treatments are combinations of spray time (early, mid, or late season) and spray amount (full or half spray). A seventh treatment is to do nothing: no spray. This baseline condition can be used to show that fungicide spray does have a strong impact on wheat yield. Differences of the other treatments relative to baseline are expected to be large; the more difficult questions, which are the principal target of the experiment, surround separating out relative effectiveness of the six spray treatments.

Some consumer product trials are aimed at determining if variants of a standard product are comparable to that standard. The primary question is: “Can a product that is ‘equally good’ be produced with differing ingredients?” The goal of these trials is to detect which of the new “test” products differs from the standard and which do not.

The two experimental situations just described earlier share an important commonality: they both incorporate a control treatment. They exhibit two of the usual ways in which controls arise, as an established treatment among new variants and as a “do nothing” treatment. The two situations differ, however, on the importance that is placed on comparisons with the control, relative to that on comparisons among the other, test treatments.

Use of a control treatment raises some interesting questions from a design optimality point of view. If, as in the fungicide trial, differences with the control are expected to be large, then it would make sense to invest less replication in the control and more in the test treatments, resulting in improved comparisons among the latter. The consumer product experiment, on the other hand, places high priority on comparisons with the control, indicating likely benefit from its over-replication.

Preceding sections have presumed equal interest in all treatments, incorporated in the permutation invariance property (3.3)(ii). By placing neither greater nor lesser emphasis on any treatment, this approach produces optimal designs that in most cases are either equally replicated or as close to equal replication as possible. As seen in the two preceding examples, and as is common for experiments including a control, a different approach is called for when not all treatment comparisons are of equal interest. This section formulates such an approach and presents basic results for its application.

A generic linear model for experiments with treatments and nuisance factors is

\[ y = X_d \tau + Z\eta + \epsilon, \]

(3.25)

where \( \eta \) incorporates the model intercept and all blocking effects. This model can be modified to infuse differential treatment interest. Let \( W_{p \times p} \) be any symmetric, positive definite matrix, scaled so that \( 1'W1 = 1 \). \( W^{1/2} \) denotes its symmetric, square root matrix satisfying \( W^{1/2}W^{1/2} = W \). Now write \( \tilde{X}_d = X_d W^{-1/2} \) and \( \tilde{\tau} = W^{1/2} \tau \). A reexpression of (3.25) is

\[ y = \tilde{X}_d \tilde{\tau} + Z\eta + \epsilon \]

(3.26)
for which the treatments information matrix, denoted $C_{dw}$, is

$$C_{dw} = \tilde{X}_d'(I - P_Z)\tilde{X}_d = W^{-1/2}C_dW^{-1/2}. \quad (3.27)$$

$C_{dw}$ is the weighted information matrix for estimation of treatment contrasts. Equations 3.26 and 3.27 tell us that $C_{dw}$ is a proper information matrix for estimation of linear model parameters, though we will not interpret it in that way. Like $C_d$, it is symmetric, nonnegative definite, and (for any connected design) rank $v - 1$. Unlike $C_d$, it does not have zero row sums, unless $1$ is an eigenvector of $W$. One such case is $W = \frac{1}{v}I$, giving (aside from the scale factor $\frac{1}{v}$) the ordinary (unweighted) information matrix as a special case.

The matrix $W$ is chosen to change the emphasis an optimality criterion places on different contrasts, particularly those involving specially designated treatments such as controls. Toward that end, it will be profitable (though not necessary, see Stallings and Morgan 2014) to investigate those $W$ that are diagonal. Henceforth,

$$W = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_v \end{pmatrix} \quad (3.28)$$

with all $w_i > 0$ and $\sum_i w_i = 1$. Then $w_i$ is the weight assigned to the $i$th treatment. How weights are selected is determined by how they impact optimality evaluations. Conventional (unweighted) optimality theory has $w_i = 1/v$ for all $i$.

**Definition 3.10** Let $\Phi$ be any optimality criterion satisfying the properties (3.3). For any given weight matrix $W$, the weighted criterion $\Phi_w$ is

$$\Phi_w(C_d) = \Phi(C_{dw}). \quad (3.29)$$

A design $d^*$ in a class of competitors $D$ is $\Phi_w$-optimal or weighted optimal with respect to $\Phi$, if $\Phi_w(C_{d^*}) = \min_{d \in D} \Phi_w(C_d)$.

Let $\mathcal{P}$ be the collection of all $v \times v$ permutation matrices, and let $\mathcal{P}_w$ contain those members of $\mathcal{P}$ that preserve the weight matrix, that is, $\mathcal{P}_w = \{P \in \mathcal{P} : PWP' = W\}$. Weighted criteria inherit these properties from (3.3):

(i) $\Phi$ respects the nonnegative definite ordering; for any $C_1, C_2 \in \mathcal{C}$ such that $C_1 - C_2$ is nonnegative definite, $\Phi_w(C_1) \leq \Phi_w(C_2)$.

(ii) $\Phi_w$ is invariant to treatment permutation in $\mathcal{P}_w$: $\Phi_w(PCP') = \Phi_w(C)$ for each $C \in \mathcal{C}$ and every $P \in \mathcal{P}_w$.

(iii) $\Phi_w$ is convex: $\Phi_w(\alpha C_1 + (1 - \alpha)C_2) \leq \alpha \Phi_w(C_1) + (1 - \alpha)\Phi_w(C_2)$ for all $C_1, C_2 \in \mathcal{C}$.

Property (i) says weighting cannot change a dominating relationship, (ii) says that weighted criteria do not distinguish among treatments accorded the same weight in (3.28), and (iii) implies that matrix averaging (compare Lemma 3.3) with allowed permutations $\mathcal{P}_w$
will improve designs from a weighted perspective. The last of these, incidentally, indicates that GGDDs may be a useful design class in the weighted setup.

**Definition 3.11** The weighted variance for contrast \( c' \tau \) estimated from design \( d \) is

\[
\text{Var}_{dw}(c' \tau) = [c' W^{-1} c]^{-1} \text{Var}_d(c' \tau). \tag{3.31}
\]

The multiplier \( [c' W^{-1} c]^{-1} \) is the weight of the contrast.

It is through weighted variances that weighted criteria are understood. Many conventional criteria, those that are functions of the eigenvalues \( e_{di} \), are statistically meaningful because all normalized contrast variances are convex combinations of the \( e_{di}^{-1} \) (the canonical variances). Denote the positive eigenvalues of the weighted information matrix \( C_{dw} \) by \( \theta_{d1} \leq \cdots \leq \theta_{dv-1} \). Wang (2009) established that all weighted contrast variances are convex combinations of the \( \theta_{di}^{-1} \), showing that eigenvalue-based weighted criteria are summary measures of weighted variance in the same way that their conventional counterparts summarize unweighted variance. Accordingly, the \( \theta_{di}^{-1} \) are termed canonical weighted variances.

Weighted versions of the eigenvalue-based criteria in Table 3.2 are \( A_w = \sum_i \theta_{di}^{-1} \) and \( E_w = 1/\theta_{d1} \). Morgan and Wang (2010) show that the \( D \) criterion is unresponsive to weights (\( D_w \) produces the same design ordering regardless of the \( w_i \)). They also show that \( E_w \) is the largest weighted variance over all contrasts and that \( A_{dw} = \sum_i \sum_{j \neq i} \frac{w_i w_j}{v-1} \text{Var}_d(\tau_i - \tau_j) \) is a weighted sum of variances for elementary treatment contrasts, providing useful interpretations of each. For brevity, only eigenvalue-based criteria are considered here.

Weight selections \( w_i \) are made for the contrast weights \( (c' W^{-1} c)^{-1} \) they induce. Though not every conceivable weighting can be captured by the diagonal weight matrix (3.28), situations of practical interest can. For experiments with a control (call it treatment 1) and \( v - 1 \) equal-interest test treatments, let \( w_1 \) be the control weight and \( w_2 = \cdots = w_v = (1 - w_1)/(v - 1) \) be the common weight for test treatments. With this selection of \( W \), the weight assigned to any elementary control contrast \( \tau_1 - \tau_i \) is \( (w_1^{-1} + w_2^{-1})^{-1} \), while that for comparing two test treatments is \( (2/w_2)^{-1} \). The ratio of these two contrast weights is \( \xi = 2(v - 1)w_1/[1 + (v - 2)w_1] \). Selection of \( w_1 \) can be determined by the desired value of \( \xi \), which ranges from 0 to 2 as \( w_1 \) is varied from 0 to 1. The range of \( \xi \) reflects the fact that in an orthogonal design, test–test comparisons can be estimated with variance arbitrarily smaller (subject to sample size) than, though no more than twice that, of test–control comparisons.

The \( A_w \), \( E_w \), and \( MV_w \) problems are solved for CRDs and all \( \xi \) in Morgan and Wang (2010). Taken up in the following are the implications when blocking in the presence of a control treatment. This topic has been intensively studied for one special case (Section 3.7.1), while general \( \xi \) has only of late come under investigation (Section 3.7.2). A special case of (3.28) is weights proportional to replications \( (w_i = r_i/bk) \); compare (3.5) and (3.27). With these weights, canonical efficiency factors and canonical weighted variances are proportional. Thus, comparing designs on their average efficiency factor (3.8) is comparing on weighted \( A \)-values. From Table 3.3, we can now say that blocking according to the design in Example 3.3 must reduce variance by more than 20% if average weighted variance is to be better than a CRD with the same replication numbers.
If one is not interested in weighted information, then (3.8) should not be used unless restricting to equireplicate designs. It should also be clear that since weighted optimality criteria are based on a fixed choice of weights, (3.8) is only valid for comparing designs $d$ with fixed replication vector $r_d$.

### 3.7.1 Design for TvC Experiments

The product-testing experiment described earlier is an example of a treatment versus control (TvC) experiment. TvC experiments seek to evaluate as efficiently as possible comparisons of $v-1$ test treatments with a control. First rigorously studied by Majumdar and Notz (1983), the relevant information matrix is $C_{d, \text{TvC}} = (H' C_d^+ H)^{-1}$, where $H = (1_{v-1}, -I_{v-1})$ is the coefficient matrix for the simple test versus control comparisons.

#### Lemma 3.8 (Morgan and Wang 2010)

Take $w_2 = \cdots = w_v = (1 - w_1)/(v - 1)$ in (3.28) and let control weight $w_1 \to 1$. The limiting Moore–Penrose inverse of $C_{d,w}$ is

$$
\lim_{w_2 \to 0, w_1 \to 1} \frac{1}{w_2} C_{d,w}^+ = \left( \begin{array}{cc} 0 & 0 \\ 0 & H' C_d^+ H \end{array} \right).
$$

Lemma 3.8 places design for TvC experiments in the weighted optimality framework as a limiting case. This relationship has been recently exploited in determining many $E_2$-optimal TvC designs. A design is $E_2$-optimal if it is $E$-optimal and if it maximizes the second smallest eigenvalue $e_{d2}$ of $C_d$ among all $E$-optimal designs.

#### Theorem 3.24 (Morgan and Wang 2011)

Let $k_0 = \lfloor k/2 \rfloor$. Suppose $d^* \in D(v, b, k)$ satisfies these conditions:

(i) The control treatment is assigned to $n_{d^*1j} = k_0$ units in block $j$ for $j = 1, 2, \ldots, b$.

(ii) $d^*$ is equally replicated and binary in the $v - 1$ test treatments.

(iii) The design $d^*_1 \in D(v - 1, b, k_1 = k - k_0)$ found by deleting all replicates of the control from $d^*$ is connected.

Then

- $d^*$ is $E$-optimal for the TvC problem.
- If $d^*_1$ is $E$-optimal over $D(v_1, b, k_1)$, then $d^*$ is $E_2$-optimal for the TvC problem.
- If $d^*_1$ is $M$-optimal over $D(v_1, b, k_1)$, then $d^*$ is $E_M$-optimal for the TvC problem.

So long as $v_1 = v - 1$ divides $bk_1$ and $k_1 \leq v_1$ all known $M$-optimal (this includes BIBDs) and $E$-optimal (see Section 3.4.4) block designs for $D(v_1, b, k_1)$ are binary and equireplicate. Theorem 3.24 successfully exploits the considerable knowledge of ordinary block designs in achieving optimality for TvC experiments.

The $A$-optimality problem for TvC experimentation has a much longer history, having been developed before the general weighted approach was available. Much of the known theory has focused on GGDD(2) designs. Because of the specialized application
to TvC experiments, the designs have been given a distinctive name (for a slightly broader definition, see the original paper Bechhofer and Tamhane 1981).

**Definition 3.12**  A design \( d \in \mathcal{D}(v, b, k) \) is a balanced treatment incomplete block design (BTIBD) if it is a GGDD(2) with groups of sizes \( v_1 = 1 \) and \( v_2 = v - 1 \), it is uniform in the control treatment (the treatment in the first group), and it is binary in the \( v - 1 \) treatments of the second group.

Since a BTIBD is uniform in the control, there are integers \( t \) and \( s \) for which the total control replication is \( tb + s \). That is, the control is assigned to \( t \) units in \( b - s \) blocks and to \( t + 1 \) units in \( s \) blocks. This is denoted by BTIBD\((v, b, k; t, s)\). The designs in Example 3.13 below are BTIBD\((8, 5, 7; 2, 0)\) and BTIBD\((8, 5, 7; 1, 0)\).

BTIBDs are “maximally averaged” with respect to the allowable permutations \( P_w \), which is all permutations of the \( v - 1 \) test treatments. The information matrix \( C_{TvC}^d \) for a BTIBD is thus completely symmetric, but Theorem 3.1 does not apply due to this matrix having full rank. The eigenstructure is, nonetheless, very simple, with \( v - 2 \) of the eigenvalues being identical. A happy consequence is that the \( A \)-value for any BTIBD is easily calculated. Indeed, any competing design, regardless of its combinatorial properties, can have its information matrix averaged to produce \( \bar{C}_{TvC}^d \) of GGDD(2) form with this same simple eigenstructure, yielding a workable lower bound expression for its \( A \)-value. This expression can be minimized over all of \( \mathcal{D}(v, b, k) \), and should that minimum occur at a BTIBD, the \( A \)-optimal design has been identified. In this way one can prove the following theorem:

**Theorem 3.25** (Stufken 1987) If \( (k - t - 1)^2 + 1 \leq t^2(v - 1) \leq (k - t)^2 \), then a BTIBD\((v, b, k; t, 0)\) is \( A \)-optimal over \( \mathcal{D}(v, b, k) \).

Theorem 3.25 is one of the more simply stated of many \( A \)-optimality results for BTIBDs. An excellent survey including many examples is available in Majumdar (1996), where variants on the approach outlined earlier, pursued extensively in the 1980s and 1990s, are described. Catalogs of \( A \)-optimal TvC designs are available in Hedayat and Majumdar (1984) and Ting and Notz (1988).

**Example 3.13**

These two designs are both GGDD(2) with overreplication of the control:

\[
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 8 \\
3 & 4 & 5 & 6 & 7 & 8 & 2 & 2 \\
5 & 6 & 7 & 8 & 2 & 3 & 4 & 4 \\
\end{array}
\quad
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 3 & 4 & 5 & 6 & 7 & 8 & 8 \\
3 & 4 & 5 & 6 & 7 & 8 & 2 & 2 \\
6 & 7 & 8 & 2 & 3 & 4 & 5 & 5 \\
8 & 2 & 3 & 4 & 5 & 6 & 7 & 7 \\
\end{array}
\]

By Theorem 3.25, \( d_1 \) is \( E_M \)-optimal. By Theorem 3.2 of Majumdar (1996), \( d_2 \) is \( A \)-optimal.

The optimal designs in Example 3.13 are quite different from one another. Substantial disagreement of criteria as seen here underscores the need for careful criterion choice in...
TABLE 3.13
Optimal TwC Designs in $D(6, 6, 4)$

<table>
<thead>
<tr>
<th>E: 0.05–0.25, A: 0.05–0.10, MV: 0.05–0.30</th>
<th>E: 0.30–0.55, A: none, MV: 0.35–0.55</th>
<th>E: 0.60, A: 0.15–0.40, MV: 0.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 2 2 2 3</td>
<td>1 1 1 2 2 3</td>
<td>1 1 1 2 2 3</td>
</tr>
<tr>
<td>2 3 3 3 4 4</td>
<td>2 2 3 3 4 4</td>
<td>2 2 3 3 4 4</td>
</tr>
<tr>
<td>3 4 4 4 5 5</td>
<td>3 3 4 5 5 5</td>
<td>3 5 4 5 5 5</td>
</tr>
<tr>
<td>4 5 6 6 6 6</td>
<td>4 5 6 6 6 6</td>
<td>4 6 5 6 6 6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E: 0.65–0.75, A: none, MV: 0.65–0.80</th>
<th>E: 0.80, A: 0.45–0.70, MV: none</th>
<th>E: 0.85–1.00, A: 0.75–1.00, MV: 0.85–1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 2 2 3</td>
<td>1 1 1 2 2 2</td>
<td>1 1 1 2 2 3</td>
</tr>
<tr>
<td>2 2 2 3 4 4</td>
<td>2 2 4 3 3 4</td>
<td>2 2 2 3 4 4</td>
</tr>
<tr>
<td>3 3 5 4 5 5</td>
<td>3 3 5 4 4 5</td>
<td>3 3 4 5 5 5</td>
</tr>
<tr>
<td>4 5 6 6 6 6</td>
<td>5 6 6 5 6 6</td>
<td>4 5 6 6 6 6</td>
</tr>
</tbody>
</table>

Note: Contrast weight ratios $\xi$, for which each design is optimal are shown at 0.05 increments.
accordance with experimenter goals. Criterion disagreement is more prevalent in weighted situations; see Morgan and Wang (2010).

Gupta et al. (1999) took this line of attack a step further, broadening the class of experiments that fall into the TvC categorization. Though not having the framework for weighted optimality described in (3.27) through (3.29), their work is equivalent to the general $A_w$ problem with $W$ specified by (3.28), allowing all $w_1$ placing higher priority on control comparisons ($\frac{1}{v} < w_1 < 1$) and so not restricting to the limiting case. Results for the $E_w$ criterion in this same, broader framework were obtained by Wang and Morgan (2010).

### 3.7.2 Lesser Weight on the Control

At the other end of the weight spectrum are the test treatments with control (TwC) experiments. TwC experiments, which include the earlier described fungicide trial, are defined by according lesser ($w_1 < 1/v$) weight to the control. This has been a topic in two recent articles, one of which is briefly described here.

Small $w_1$ will typically mean small replication for the control. Matrix averaging then becomes a less effective technique (the bounds produced can be further from what is actually attainable) for determining optimal or efficient designs, the difficulties being more acute in small designs. Wang and Morgan (2010) thus resorted to enumeration to find small, optimal TwC designs. For all but the very smallest problems, feasibility restrictions on enumeration are required. These authors conjectured that optimal designs for TwC experimentation in $D(v, b, k)$ ($k < v$) are binary in all treatments and have range no more than one for the replication numbers of the $v - 1$ test treatments (compare the John–Mitchell conjecture in Section 3.4.1). With these restrictions, they compiled a catalog of optimal TwC designs for all $v \leq 12$ and $b \leq 12$ with $v + b \leq 18$. Table 3.13 displays results of one of these enumerations.

The designs in Table 3.13 hold no surprises. While the criteria do not always agree, replication of the control increases with $w_1$ (equivalently, as the contrast weight ratio $\xi$ increases) regardless of the criterion. If sufficiently close to the unweighted case ($w_1 = 1/v; \xi = 1$), the criteria agree that an equally replicated design (a RGD) is best. The criteria also agree on the same design, having just one replicate of the control, for sufficiently small $w_1$.

### 3.8 Discussion and Directions

Blocking is one of the three fundamental techniques of design propounded by Fisher (1935), growing out of his development of experimental ideas nearly a century ago. Since then, it has been applied in multitudinous forms in a wide variety of experimental settings, with increasing examination from theorists as time has passed. This chapter has attempted to give a reasonably compact overview of the basic blocking implementations and, for them, the mathematical techniques that have been developed to maximize the information that a blocked experiment can provide.

There are many extensions. Covered elsewhere in this volume are blocking of fractions (Chapter 8), and blocking with nonlinear models for responses (Chapter 13). Yet other topics are blocking with nonconstant block sizes (Ghosh et al. 2008), block size $k > v$ (Jacroux and Whittinghill 1988, especially their Example 2.5), blocking with treatments and continuous covariates (Chapter 4 of Liski et al. 2002), blocking in the presence of correlated...
errors (Martin and Eccleston 1991) and robustness of blocking schemes to loss of observations (Morgan and Parvu 2008; Godolphin and Warren 2011) or to weight range (Wang and Morgan 2012). The just-cited papers contain references for further exploration.

The simple block designs of Sections 3.3 through 3.5 have been the most intensely studied. It is then not surprising that they are the best understood, this being due in part to greater tractability of the underlying mathematical problems. Yet there is still much unsolved, especially as regards designs where average replication must be small. With small designs, as indicated by Section 3.5, the combinatorial properties that lead to optimality in larger designs do not necessarily prove useful. Paralleling this, optimality bounds from matrix averaging and, from results such as Theorems 3.3 through 3.5, also tend to be unhelpful. In many cases, small designs are simply “too discrete” to closely conform to results found through the smoothness of continuous mathematics. Looking in the other direction, the John–Mitchell conjecture that some RGD must be optimal when a RGD exists has been confirmed for many criteria with arbitrarily large designs by Cheng (1992).

The asymptotics of Cheng (1992) suggest this interesting question: given an optimal design in \( D(v, b, k) \) and a BIBD in \( D(v, b_0, k) \), when is the collection of blocks from both designs an optimal design in \( D(v, b + b_0, k) \)? The suggestion is certainly attractive, for adding a BIBD to the IBD \( d \) simply adds a constant to the eigenvalues \( e_d \). The \( E \)-optimality work by Morgan (2007b) described in Section 3.4.4 has established arbitrary BIBD-extendability for most \( E \)-optimal block designs with \( v \leq 15 \). Cakiroglu (2013) has shown that extendability holds with respect to \( A \)- and \( D \)-optimality for many of the regular graph designs in the John–Mitchell catalog. The general question, however, remains open.

In some situations, more than one optimal block design has been identified for the same criterion; see, for instance, Table 3.6. The obvious solution in this case is to bring other criteria to bear, further trimming the competitors by their values on one or more secondary optimality measures. This idea was seen in Theorem 3.24 with \( E_2 \) and \( E_M \)-optimality. Alternatively, one might examine the collection of contrast variances that will be produced by each design, seeking simplicity that will aid interpretation of experimental results. It may on occasion be worthwhile to sacrifice a small measure of efficiency in exchange for more readily interpretable results.

When an IBD is needed and an optimal design is not known, the results reported in this chapter give good guidance about what structural properties should be sought. So long as the setting is not too small (cf. Section 3.5), a design incorporating (i) and (ii) of Definition 3.5, and keeping trace \( (C_d^2) \) as small as possible, will be reasonably close to highest efficiency.

References for where optimal block designs can be found are scattered throughout this chapter, but they are by no means complete. This reflects a substantial shortfall of design research: the failure to create easy access to large collections of optimal designs. Practicing statisticians need designs that are readily available, be they from online catalogs or generated by popular software. Unfortunately, an optimal design in a journal article can be only there—it may not exist for much of the experimental world. Two projects to help remedy this situation, designtheory.org (Bailey et al. 2006) and www.iasri.res.in/design, have appeared in the past decade. While these are very positive steps, there is still much to be done.

For those desiring to generate block designs with specified combinatorial properties, including resolvability restrictions, the freely available software GAP Design (Soicher 2009) is an excellent tool so long as \( v + b \) is not too large (typically less than 20, though block size \( k \) also plays a role). Numerous techniques for design construction have been developed
through the years, many having roots in the *method of differences* developed by Bose (1939) for construction of BIBDs. Bose’s idea was to start with a few, carefully crafted blocks from which all the blocks of a design could be generated through a group action. That general idea is an integral part of GAP Design.

The impact of increasing computational power has not been as great in discrete design as in most other areas of statistics. This is in part because some of the optimality problems, like those for $D(v,b,k)$ with plot excess $p = 0$, have partially yielded to analytic solution. It is also due in part to combinatorial possibilities becoming too numerous too quickly as design size grows. This could change in the near future if techniques like semidefinite programming, now being brought to bear on optimizing design for regression models (Papp 2012), can be adapted for the design problems considered here. There has certainly been significant, recent progress from reducing problem size through mathematical argument, thus making computational resolution feasible (e.g., Section 3.4.4). However, there will always be a need for, and an appreciation of, complete solutions to design problems like those in Kiefer (1975), Cheng (1978), Bagchi and Bagchi (2001), and others mentioned throughout this chapter. It would be a great loss should search algorithms become replacements for known theory and results for optimal designs. At the least, known optimal block designs should be stored for ready access, not searched at every call with no guarantee of absolute optimality.

References


Blocking with Independent Responses


Wang, X. (2009), Weighted optimality of block designs, PhD dissertation, Blacksburg, VA: Virginia Polytechnic Institute and State University.


