Robust Designs for Field Experiments with Blocks

by

Rena Kaur Mann

B.Sc., University of Victoria, 2008

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

© Rena Kaur Mann, 2011

University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.
Robust Designs for Field Experiments with Blocks

by

Rena Kaur Mann

B.Sc., University of Victoria, 2008

Supervisory Committee

Dr. Julie Zhou, Co-Supervisor
(Department of Mathematics and Statistics)

Dr. Rod Edwards, Co-Supervisor
(Department of Mathematics and Statistics)

Dr. Farouk Nathoo, Departmental Member
(Department of Mathematics and Statistics)
Supervisory Committee

Dr. Julie Zhou, Co-Supervisor  
(Department of Mathematics and Statistics)

Dr. Rod Edwards, Co-Supervisor  
(Department of Mathematics and Statistics)

Dr. Farouk Nathoo, Departmental Member  
(Department of Mathematics and Statistics)

ABSTRACT

This thesis focuses on the design of field experiments with blocks to study treatment effects for a number of treatments. Small field plots are available but located in several blocks and each plot is assigned to a treatment in the experiment. Due to spatial correlation among the plots, the allocation of the treatments to plots has influence on the analysis of the treatment effects. When the spatial correlation is known, optimal allocations (designs) of the treatments to plots have been studied in the literature. However, the spatial correlation is usually unknown in practice, so we propose a robust criterion to study optimal designs of the treatments to plots. Neighbourhoods of correlation structures are introduced and a modified generalized least squares estimator is discussed. A simulated annealing algorithm is implemented to compute optimal/robust designs. Various results are obtained for different experimental settings. Some theoretical results are also proved in the thesis.
Contents

Supervisory Committee ii
Abstract iii
Table of Contents iv
List of Tables vi
List of Figures vii
Acknowledgements ix

1 Introduction 1
   1.1 Design of Field Experiments 2
   1.2 Linear Models and Estimation Methods for Block Designs 5
   1.3 Optimality Criteria and Examples of Optimal Designs 9
   1.4 Research Problems 10
   1.5 Significant Contributions 11

2 Spatial Correlation Structures 13
   2.1 Nearest Neighbour 14
   2.2 Moving Average 16
   2.3 Doubly Geometric 18
   2.4 Discrete Exponential 22
List of Tables

Table 3.1 Optimal designs for $t = 3$, $n = 1$, and $b = 2$. . . . . . . . . . . . 38
Table 3.2 Optimal designs with a control for $t = 3$, $n = 1$, and $b = 2$. . . . 41
Table 4.1 Optimal designs with a control for $t = 4$, $n = 1$, and $b = 2$. . . . 46
Table 4.2 Optimal designs with a control for $t = 6$, $n = 1$, and $b = 2$. . . . 46
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A Latin square design with four treatments in 16 plots.</td>
<td>4</td>
</tr>
<tr>
<td>1.2</td>
<td>A Latin square design for four treatments with all the plots in one line.</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Fisher’s agricultural experiment.</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Notation for response variable for the plots in one block where $y_{k,l,j}$ is the response of the plot in the $k^{th}$ row, $l^{th}$ column, and $j^{th}$ block and $1 \leq k \leq m, 1 \leq l \leq n, 1 \leq j \leq b$.</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>Spatially-balanced complete block design.</td>
<td>10</td>
</tr>
<tr>
<td>1.6</td>
<td>Neighbour balanced complete block design.</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Order of the neighbours to a plot O.</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Some optimal $3 \times 3$ designs.</td>
<td>23</td>
</tr>
<tr>
<td>2.3</td>
<td>Some optimal $4 \times 4$ designs.</td>
<td>24</td>
</tr>
<tr>
<td>2.4</td>
<td>Knight’s move Latin square designs.</td>
<td>25</td>
</tr>
<tr>
<td>3.1</td>
<td>Arrangement of block one with $n = 1$.</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>Arrangement of block one for a $m \times n$ array.</td>
<td>36</td>
</tr>
<tr>
<td>3.3</td>
<td>The three designs for $t = 3, n = 1$, and $b = 2$.</td>
<td>36</td>
</tr>
<tr>
<td>3.4</td>
<td>Loss functions versus $\rho$ for the GLSE with the NN(1) correlation structure.</td>
<td>39</td>
</tr>
<tr>
<td>3.5</td>
<td>Designs with treatment one as a control for $t = 3, n = 1$, and $b = 2$.</td>
<td>40</td>
</tr>
</tbody>
</table>
Figure 3.6 Loss functions versus $\rho$ for the GLSE with NN(1) correlation and a control. ................................. 42

Figure 3.7 Loss functions versus $\gamma$ for the LSE with MA(1) correlation and a control. ............................... 43

Figure 4.1 Designs with treatment one as a control for $t = 4$, $n = 1$, and $b = 2$. ........................................ 45

Figure 4.2 Optimal designs with treatment one as a control for $t = 6$, $n = 1$, and $b = 2$. ................................. 47

Figure 5.1 Loss function $L(C(R_0, U_{R_1,\alpha}, X))$ versus number of iterations. ........................................... 56

Figure 5.2 Minimax design with $t = 7$, $n = 1$, and $b = 2$ under the NN(1). ................................................. 56

Figure 5.3 Loss function $L(C(R_0, U_{L,\alpha}, X))$ versus number of iterations. ................................................. 57

Figure 5.4 Minimax design with $t = 12$, $n = 2$, and $b = 2$ under the DG. .................................................. 57

Figure 5.5 Minimax design with $t = 9$, $n = 3$, and $b = 2$ under the MA(1). .................................................... 58

Figure 5.6 Minimax design with $t = 9$, $n = 3$, and $b = 2$ under the NN(1) with a control treatment. ............. 58

Figure 5.7 $\xi_{D,L}$-minimax designs with $t = 12$, $n = 2$, and $b = 2$ under the DG where $\ell_{D,L} = \min_X \det[C(R_0, U_{L,\alpha}, X)]$: (a) $\lambda = 0.01$, (b) $\lambda = 0.3$, (c) $\lambda = 0.7$, and (d) $\lambda = 0.95$. ................................. 59

Figure 5.8 $\xi_{D,R_1}$-minimax designs with $t = 12$, $n = 2$, and $b = 2$ under the DG where $\ell_{D,R_1} = \min_X \det[C(R_0, U_{R_1,\alpha}, X)]$: (a) $\lambda = 0.01$, (b) $\lambda = 0.3$, (c) $\lambda = 0.7$, and (d) $\lambda = 0.95$. ................................. 59

Figure 5.9 $\xi_{D,R_1}$-minimax designs for various $t$ under the NN(1) where $\ell_{D,R_1} = \min_X \det[C(R_0, U_{R_1,\alpha}, X)]$: (a) $t = 10$, (b) $t = 12$, (c) $t = 14$, (d) $t = 16$, and (e) $t = 18$. ................................................. 60

Figure 6.1 Permutation of inner treatments of block two. ................. 68
ACKNOWLEDGEMENTS

I would like to acknowledge some very important people who undoubtedly played a major role in helping me achieve a Master of Science degree.

My two supervisors, Dr. Julie Zhou and Dr. Rod Edwards, were always available for assistance and advice. Dr. Zhou unfailingly wore a friendly smile on her face and met with me whenever I had questions. Dr. Edwards was also extremely helpful and beneficial in the completion of my thesis. It was always a pleasure to meet and work with them.

I would like to thank Dr. Farouk Nathoo for his time and assistance, not only as a member of my committee but as my professor as well. I had the opportunity to either learn from or work with other professors in the Statistics department: Dr. Laura Cowen, Dr. Mary Lesperance, Dr. Bill Reed, and Dr. Min Tsao, and I thank them for making my time as a graduate student an enjoyable experience. I would also like to thank Dr. Michael Wilmut for serving as my external examiner and the University of Victoria for funding me with a Fellowship.

My high regard for education comes from my parents, whom have always supported and encouraged me in my educational pursuits. Thank you for always being there for me and backing my life’s decisions.

Lastly, I would like to thank my love and partner, Vik. You make my life complete.

*Education consists mainly of what we have unlearned.*

Mark Twain

*I cannot teach anybody anything, I can only make them think.*

Socrates
Chapter 1

Introduction

Field experiments have been studied extensively in literature dating back to the 1920’s when Sir Ronald A. Fisher led studies in agricultural design. Field experiments are special in that spatial correlation plays an important role. Taking the spatial correlation structure into account can help produce more precise results as agricultural field experiments have repeatedly shown that experimental units (plots) farther apart are less alike than neighbouring units (Cochran and Cox, 1960).

In this thesis, we will consider field experiments with blocks and construct robust designs. Section 1.1 gives some background on the development of the design of experiments while Section 1.2 examines the linear model and presents two estimation methods. Section 1.3 discusses the classical optimality criteria and gives some examples of optimal designs. Section 1.4 contains the research problems that are investigated in this thesis. Section 1.5 summarizes the significant contributions made in this thesis.
1.1 Design of Field Experiments

The design and analysis of experiments has long been studied through the years. There are many applications of experimental designs such as agricultural field experiments and industrial experiments. Field experiments have been extensively developed due to the many benefits that come from having a good experimental design such as reducing costs, development time, and variability, as well as improving process yields (see Montgomery, 2005). Identifying and understanding factors that optimize certain characteristics such as yield or creating new varieties of product that are superior to current varieties (through disease resistance, yield or quality) are just a few benefits of field experiments (Petersen, 1994). As well, a good design is necessary to be able to draw valid conclusions from data. There are two major steps to an experimental problem: the design of the experiment and the statistical analysis of the data (Montgomery, 2005). We are interested in the design of the experiment so that when the statistical analysis is performed, meaningful conclusions may be drawn.

There have been four major eras of statistical experimental design according to Montgomery (2005). The first was the agricultural era that began in the 1920’s and was led by Sir Ronald A. Fisher. Fisher developed the three principles of experimental design: randomization, replication, and blocking. The allocation of experimental material and the order in which runs or trials are performed must be random. This helps to eliminate external factors that the experimenter is not interested in from influencing the outcome. As well, this helps to ensure that the observations are independently distributed random variables. Replication refers to the independent duplication of a factor combination and helps to obtain more precise estimates of the desired parameters, as well as to estimate the experimental error. Blocking is used to reduce the variability from nuisance factors and improve the precision with which comparisons among factors are made.
The second era is known as the industrial era and saw the development of response surface methodology by Box and Wilson (1951). They discovered the two notions of immediacy and sequentiality. These notions meant that the response could be measured almost immediately for industrial experiments (in laboratories, through pilot projects etc.) and thus results from current runs could be used to plan the following sets (Box, 1999). Therefore, only a small number of trials were needed to plan the subsequent experiment. The third era was driven by Genichi Taguchi (1993) who supported using designed experiments for robust parameter design for quality improvement. Taguchi advocated highly fractionated factorial designs and orthogonal arrays to solve many problems and as a result, much interest arose in designed experiments in discrete parts industries. Although problems soon became apparent with Taguchi’s methods, he helped usher in the fourth era where renewed interest became apparent in statistical design. This era has seen many new approaches and techniques to industrial problems, as well as solutions to the difficulties with Taguchi’s methods.

Field experiments are usually used to compare the yields for various treatments applied to field crops. The treatments can be various types of fertilizers, planting densities or soil types. The experiments can be planned as follows. Given a large field, we divide the area into many small plots, which are usually rectangular. Then we need to find a design to assign treatments to these plots. For example, Figure 1.1 shows an area with 16 plots arranged in four rows and four columns. Suppose there are four treatments to be compared and they are labelled 1, 2, 3, and 4. Figure 1.1 gives one design, where the number in each plot indicates the treatment level. It is obvious that there are many ways to assign four treatments to 16 plots.

Designs that are arranged in one block with multiple replications have been extensively analyzed. Martin (1986) is at the forefront of this area and has examined
Figure 1.1: A Latin square design with four treatments in 16 plots.

designs arranged in a $m \times n$ array with $mn = tr$, where $t$ is the number of treatments and $r$ is the number of replicates.

Often, field experiments are carried out in several blocks, where the blocks can be located at separate fields or on the same field. They are used to block the effects of nuisance factors. Cochran and Cox (1960) give one example of a block design where rows are compact blocks of land while the columns specify the order within each block and this is illustrated in Figure 1.2. This is useful when the yield gradient is thought to be in the same direction along the line because the order in each block and the blocks themselves eliminate the effects of the gradient more effectively than a single control (Cochran and Cox, 1960). Four blocks are created along the line in the same field and four treatments are assigned into each block. The empty plots separate the blocks.

Figure 1.2: A Latin square design for four treatments with all the plots in one line.

Fisher (1960) examined block designs for agricultural experiments. Figure 1.3 gives a design for eight blocks and five treatments. The treatments within each block are randomly assigned. The entire field underwent a uniform agricultural treatment and at harvest, narrow edges at the end of each of the plots were removed from the analysis.

Blocking is used to group plots into more homogeneous groups taking into account
the differences that may be present across an entire agricultural field. Instead of choosing to run an experiment on a section of a field or spreading treatments far apart, which does not take into consideration the differences that may be present across the field, blocking helps to account for more variability. Block designs are used assuming conditions in a block are uniform. In a field experiment, there can be any number of treatments. However, as the size of a block grows, the standard error per unit (plot) can also be expected to rise (Cochran and Cox, 1960). Thus, blocks should be made as compact as possible as it is desired to keep the variation in each block as small as possible.

Spatial dependence between units was being taken into consideration as early as 1937 by J. S. Papadakis (Martin, 1996) but did not become popular until the latter half of the century when notable papers by Bartlett (1978) and Wilkinson et al. (1983) were published. In this thesis, we will examine four common spatial correlation structures. These will be discussed in more detail in Chapter 2.

1.2 Linear Models and Estimation Methods for Block Designs

Consider an experiment with $b$ blocks, where each block contains plots arranged in a $m \times n$ array with $mn = t$. Figure 1.4 shows the notation of a response variable $y$ for
one block.

| $y_{1,1,j}$ | $y_{1,2,j}$ | $\cdots$ | $y_{1,(n-1),j}$ | $y_{1,n,j}$ |
| $y_{2,1,j}$ | $y_{2,2,j}$ | $\cdots$ | $y_{2,(n-1),j}$ | $y_{2,n,j}$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $y_{(m-1),1,j}$ | $y_{(m-1),2,j}$ | $\cdots$ | $y_{(m-1),(n-1),j}$ | $y_{(m-1),n,j}$ |
| $y_{m,1,j}$ | $y_{m,2,j}$ | $\cdots$ | $y_{m,(n-1),j}$ | $y_{m,n,j}$ |

Figure 1.4: Notation for response variable for the plots in one block where $y_{k,l,j}$ is the response of the plot in the $k^{th}$ row, $l^{th}$ column, and $j^{th}$ block and $1 \leq k \leq m$, $1 \leq l \leq n$, $1 \leq j \leq b$.

A fixed effects model for the block design is (Montgomery, 2005),

$$y_{k,l,j} = \mu + \tau_i + \beta_j + \epsilon_{k,l,j}, \quad i = 1, \ldots, t,$$
$$j = 1, \ldots, b, \quad k = 1, \ldots, m, \quad l = 1, \ldots, n,$$

where $y_{k,l,j}$ is the observed response in the $k^{th}$ row and $l^{th}$ column of block $j$, $\mu$ is the overall mean, $\tau_i$ is the $i^{th}$ treatment effect, $\beta_j$ is the $j^{th}$ block effect, and $\epsilon_{k,l,j} \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Here, $\sum_{i=1}^{t} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$. This model can also be written as

$$y_{k,l,j} = \mu_i + \beta_j + \epsilon_{k,l,j},$$

where $\mu_i$ is now the mean of the $i^{th}$ treatment and there are no restrictions on $\mu_i$ but the restriction on $\beta_j$ remains. This is the form of the model that will be examined in this thesis.

This model can also be represented in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad (1.1)$$

where $\mathbf{y} = (y_{1,1,1}, \ldots, y_{1,n,1}, \ldots, y_{m,n,1}, y_{1,1,2}, \ldots, y_{m,n,b})^\top$ is the $tb$ length response vec-
tor (see Figure 1.4 for notation), $X$ is a $tb \times (t + b - 1)$ design matrix, $\theta = (\mu_1, \ldots, \mu_t, \beta_1, \ldots, \beta_{b-1})^\top$ is the parameter vector of length $t + b - 1$, and the error vector is $\epsilon = (\epsilon_{1,1,1}, \ldots, \epsilon_{m,n,1}, \epsilon_{1,1,2}, \ldots, \epsilon_{m,n,b})^\top$. Entries in the design matrix are either 0, 1 or -1. In the first $t$ columns, column $i$ has $b$ 1’s corresponding to the $i^{th}$ treatment. In the last $b - 1$ columns, column $t + j$ for $j \leq b - 1$ has $t$ 1’s corresponding to the $j^{th}$ block and treatments in block $b$ produce -1’s in the last $b - 1$ columns because $\beta_b = -\sum_{j=1}^{b-1} \beta_j$. The error term $\epsilon$ has $E(\epsilon) = 0$ and $Cov(\epsilon) = \sigma^2 V$, where $V$ is a symmetric and positive definite correlation matrix and $\sigma^2 = Var(\epsilon_{i,j,k})$. The $V$ matrix is a block diagonal since we assume there is no correlation between the blocks.

The Least Squares Estimator (LSE) and the Generalized Least Squares Estimator (GLSE) can be used to estimate $\theta$. The optimal design for a chosen dependence structure can depend on the estimator and be quite different (Martin and Eccleston, 1991).

The LSE is found by minimizing the sum of squared residuals which are the differences between the observed responses and the fitted responses. The correlation structure does not need to be known. The LSE is given by

$$\hat{\theta}_{LSE} = (X^\top X)^{-1}X^\top y,$$

(1.2)

and the covariance of $\hat{\theta}_{LSE}$ is then

$$Cov(\hat{\theta}_{LSE}) = \sigma^2 (X^\top X)^{-1}X^\top VXX^\top X^{-1}.$$ 

In general, we are not interested in the block effects since they are nuisance effects. Thus, in this thesis we will consider the covariance of the estimators for the treatment
effects to study optimal designs, which is given by

\[
\text{Cov}(C\hat{\theta}_{LSE}) = \sigma^2 C (X^\top X)^{-1} X^\top V X (X^\top X)^{-1} C^\top,
\]

(1.3)

where \( C \) is a constant matrix taking the form

\[
C = \begin{pmatrix} I_t & 0_{tx(b-1)} \end{pmatrix}.
\]

(1.4)

The GLSE may be applied when the correlation structure is known. The GLSE is the best linear unbiased estimator (BLUE) of \( \theta \). The treatment effects and block effects are estimated by

\[
\hat{\theta}_{GLSE} = (X^\top V^{-1} X)^{-1} X^\top V^{-1} y,
\]

(1.5)

and the covariance matrix of \( \hat{\theta}_{GLSE} \) is given by

\[
\text{Cov}(\hat{\theta}_{GLSE}) = \sigma^2 (X^\top V^{-1} X)^{-1}.
\]

Similarly, the covariance of the estimators for the treatment effects is considered for optimal designs which is given by

\[
\text{Cov}(C\hat{\theta}_{GLSE}) = \sigma^2 C (X^\top V^{-1} X)^{-1} C^\top,
\]

(1.6)

where \( C \) is in (1.4).
1.3 Optimality Criteria and Examples of Optimal Designs

Two commonly used optimality criteria to construct optimal designs are to minimize the trace or determinant of the covariance. When the determinant of the covariance matrix is minimized, the resulting design is called a $D$-optimal design (Pukelsheim, 1993). This is equivalent to minimizing the product of the eigenvalues of the covariance matrix. $D$-optimal designs minimize the volume of the confidence region for $C\theta$. Similarly, minimizing the trace of the covariance gives the $A$-optimal design. $A$-optimal designs minimize the sum of the eigenvalues of the covariance matrix, as well as minimize the average of the variances of $C\hat{\theta}$. We will use $\xi_{\nu,\psi}$ to denote the $\nu$-optimal design where $\nu$ is $D$ or $A$ and $\psi$ is the estimator. Thus, $\psi$ can be the LSE or the GLSE.

Other optimality criteria have been used for field experiments with blocking. Spatially-balanced complete block designs have been developed for experiments involving up to 15 treatments and 15 blocks (van Es et al., 2007). The objective is to create spatial balance among treatments and prevent treatments from appearing in the same position across different blocks. Figure 1.5 displays a spatially-balanced complete block design for a field experiment with four treatments and four blocks. When the treatments and replications are equal, square spatially-balanced block designs are quite obviously Latin squares. However, the reverse does not hold true as many Latin square designs are not spatially balanced. Thus, square spatially-balanced complete block designs are a subset of Latin square designs.

Neighbour balanced designs, as suggested by their name, are designs where treatments occur equally often with other treatments as their neighbours. They are optimal using the GLSE and a first-order autoregressive correlation structure (Gill and
Shukla, 1985), for which the correlation between two plots in a block decreases exponentially with distance. A four treatment, four block neighbour balanced design using a stationary first-order autoregressive process (see Kunert, 1987) is shown in Figure 1.6.

1.4 Research Problems

Since both the $\text{Cov}(\hat{C}_{LSE})$ and $\text{Cov}(\hat{C}_{GLSE})$ depend on the correlation matrix $V$ of the field plots, optimal designs depend on $V$ and may be very sensitive to small changes of $V$. In practice, $V$ is never known exactly and so we need to relax the assumptions for the correlation matrix. In this thesis, we use robust design criteria so that the optimal designs do not depend on a precise choice of correlation $V$. To achieve robustness, we introduce two neighbourhoods of correlation structures. These neighbourhoods have been discussed previously in Wiens and Zhou (2008) and in Ou and Zhou (2009). In particular, Ou and Zhou presented robust minimax designs for field experiments arranged in a $m \times n$ array where $mn = tr$ and we extend this result for block designs.
We want to determine how to arrange a field experiment with blocks in order to minimize some scalar loss functions of the covariance. The arrangement of the treatments is very important as the correlation between adjoining plots plays a significant role when estimating the treatment effects. Four spatial correlation structures are discussed in Chapter 2 and several properties of block designs are examined in Chapter 3. Two neighbourhoods of correlation structures are introduced in Chapter 4 and robust design criteria are also presented. A numerical algorithm is discussed in Chapter 5 to find robust minimax designs and theoretical results are presented in Chapter 6. Finally, the conclusions of this thesis are in Chapter 7.

1.5 Significant Contributions

This thesis explores the design of field experiments using blocking in order to compare treatment effects. Two cases are studied, one case involves a number, \( t \), of equivalent treatments and the other involves a control treatment. Various spatial correlation structures are introduced and several optimality criteria are discussed. Properties of block designs are explored in detail and several results are examined. Neighbourhoods of a correlation structure for block designs are expanded upon and a generalized least squares estimator, which is used when the covariance matrix belongs in a neighbourhood, is introduced. Properties of the neighbourhood are discussed and robust design criteria for block designs are obtained. These criteria are robust against misspecifications. A simulated annealing algorithm is presented to find optimal and minimax block designs. Minimax designs are computed and some theoretical results are presented for optimal designs.

The thesis contains the following contributions:

1. Basic properties of block designs are explored.
2. Neighbourhoods of correlation structures are extended to field plots with blocks and robust design criteria are defined.

3. A simulated annealing algorithm is implemented to compute robust designs for field experiments with blocks.

4. Numerical results are presented for some representative experiments.

5. Several theoretical results are proven.
Chapter 2

Spatial Correlation Structures

Several correlation structures are discussed including Nearest Neighbour, Moving Average, Doubly Geometric, and Discrete Exponential correlations. Examples of optimal designs using these correlation structures will also be presented at the end of the chapter.

The directional horizontal and vertical distances between two plots will be denoted by $g$ and $h$ respectively. For simplicity in this chapter, we will be ignoring the third subscript denoting the block $j$ when referring to plots, since we assume that the correlation structures are equal for all blocks. Thus, the correlation between two plots in any block is denoted $\rho_{g,h} = \text{Corr}(\epsilon_{k,l}, \epsilon_{r,s})$ where $g = k - r$ and $h = l - s$.

The correlation matrix $V$ contains $V_1, \ldots, V_b$ as its block diagonal and looks like the following:

$$
V = \begin{pmatrix}
V_1 & 0 & \ldots & 0 & 0 \\
0 & V_2 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & V_{b-1} & 0 \\
0 & 0 & \ldots & 0 & V_b
\end{pmatrix}.
$$
where \( V_j \) is the correlation matrix for block \( j \) and \( V_1 = \ldots = V_b \).

### 2.1 Nearest Neighbour

Nearest Neighbour (NN) correlation is a structure that takes into consideration neighbours of different orders. The orders of neighbours to a plot \( O \) are displayed in Figure 2.1.

![Order of the neighbours to a plot O.](image)

In this thesis, we focus on the first order neighbours (NN(1)); i.e. only the adjoining plots will be considered to have correlation. The direction of the neighbours does not matter, i.e. left, right, above, and below neighbours have the same effect. Therefore, from Kiefer and Wynn (1981),

\[
\rho_{g,h} = 1 \text{ for } |g| = |h| = 0, \\
\rho_{1,0} = \rho_{-1,0} = \rho_{0,1} = \rho_{0,-1} = \rho \text{ (i.e. for } |g| + |h| = 1), \text{ and} \\
\rho_{g,h} = 0 \text{ for } |g| + |h| > 1, \text{ where } \rho \in [0, \frac{1}{4}].
\]

For a block with one column, i.e. \( n = 1 \), the correlation matrix of the NN(1) has
the following form,

\[ V_j = \begin{pmatrix}
1 & \rho \\
\rho & 1 & \rho \\
\rho & 1 & \rho & \ddots & \ddots & \ddots \\
\rho & 1 & \rho & \ddots & \ddots & \ddots \\
\rho & 1 & \rho & \ddots & \ddots & \ddots \\
\rho & 1 & \rho & \ddots & \ddots & \ddots \\
\end{pmatrix}_{t \times t}, \]

where missing entries in matrices throughout this thesis indicate entries of 0. For a block with two columns, i.e. \( n = 2 \), numbered across rows, \( V_j \) has the following form,

\[ V_j = \begin{pmatrix}
1 & \rho & \rho \\
\rho & 1 & \rho & 0 \\
\rho & 0 & 1 & \rho & \rho \\
\rho & \rho & 1 & 0 & \rho & \ddots & \ddots & \ddots & \ddots \\
\rho & 0 & 1 & \rho & \rho & \ddots & \ddots & \ddots & \ddots \\
\rho & \rho & 1 & 0 & \rho & \ddots & \ddots & \ddots & \ddots \\
\rho & 0 & 1 & \rho & \rho & \ddots & \ddots & \ddots & \ddots \\
\rho & \rho & 1 & \rho & \rho & \ddots & \ddots & \ddots & \ddots \\
\rho & \rho & 1 & \rho & \rho & \ddots & \ddots & \ddots & \ddots \\
\rho & \rho & 1 & \rho & \rho & \ddots & \ddots & \ddots & \ddots \\
\end{pmatrix}_{t \times t}. \]
2.2 Moving Average

The Moving Average (MA) model is another structure often used in spatial correlation analysis (Haining, 1978). The model is given by

\[ \epsilon_{k,l} = \alpha u_{k-1,l} + \beta u_{k+1,l} + \gamma u_{k,l-1} + \delta u_{k,l+1} + u_{k,l}, \]

where \( \alpha, \beta, \gamma, \) and \( \delta \) are constant coefficients and \( u_{k,l} \) are i.i.d random variables with mean 0 and variance \( \sigma_u^2 \). In this thesis, we focus on a special case of the MA model known as the MA(1) model, where \( \alpha = \beta = \delta = \gamma \). The process is given by

\[ \epsilon_{k,l} = \gamma u_{k-1,l} + \gamma u_{k+1,l} + \gamma u_{k,l-1} + \gamma u_{k,l+1} + u_{k,l}. \]

The correlation values are as follows, with \( \rho = \frac{2\gamma}{1+4\gamma^2} \) and \( \gamma \in [0, \frac{1}{4}] \),

\[
\begin{align*}
\rho_{0,0} & = 1, \\
\rho_{1,0} & = \rho_{-1,0} = \rho_{0,1} = \rho_{0,-1} = \rho, \\
\rho_{1,1} & = \rho_{-1,1} = \rho_{1,-1} = \rho_{-1,-1} = \gamma \rho, \\
\rho_{2,0} & = \rho_{-2,0} = \rho_{0,2} = \rho_{0,-2} = \frac{1}{2} \gamma \rho, \text{ and} \\
\rho_{g,h} & = 0 \text{ for } |g| + |h| > 2.
\end{align*}
\]

Here is a brief derivation for \( \rho_{1,0}, \rho_{1,1}, \) and \( \rho_{2,0} \) and the other results can be derived similarly. Since

\[
\begin{align*}
\text{Cov}(\epsilon_{k,l}, \epsilon_{k,l}) = \text{Var}(\epsilon_{k,l}) & = \mathbb{E}[(\gamma u_{k-1,l} + \gamma u_{k+1,l} + \gamma u_{k,l-1} + \gamma u_{k,l+1} + u_{k,l})^2] \\
& = (1 + 4\gamma^2)\sigma_u^2, \\
\text{Cov}(\epsilon_{k,l}, \epsilon_{k+1,l}) & = \mathbb{E}[(\gamma u_{k-1,l} + \gamma u_{k+1,l} + \gamma u_{k,l-1} + \gamma u_{k,l+1} + u_{k,l})].
\end{align*}
\]
\[(\gamma u_{k,l} + \gamma u_{k+2,l} + \gamma u_{k+1,l-1} + \gamma u_{k+1,l+1} + u_{k+1,l})]\]
\[= E(\gamma u_{k+1,l}^2 + \gamma u_{k,l}^2)\]
\[= 2\gamma \sigma_u^2,\]

\[\text{Cov}(\epsilon_{k,l}, \epsilon_{k+1,l+1}) = E[(\gamma u_{k-1,l} + \gamma u_{k+1,l} + \gamma u_{k,l-1} + \gamma u_{k,l+1} + u_{k,l}) \cdot (\gamma u_{k,l+1} + \gamma u_{k+2,l+1} + \gamma u_{k+1,l+1} + \gamma u_{k+1,l+2} + u_{k+1,l+1})]\]
\[= E(\gamma^2 u_{k+1,l}^2 + \gamma^2 u_{k,l+1}^2)\]
\[= 2\gamma^2 \sigma_u^2,\]

\[\text{Cov}(\epsilon_{k,l}, \epsilon_{k+2,l}) = E[(\gamma u_{k-1,l} + \gamma u_{k+1,l} + \gamma u_{k,l-1} + \gamma u_{k,l+1} + u_{k,l}) \cdot (\gamma u_{k+1,l} + \gamma u_{k+3,l} + \gamma u_{k+2,l-1} + \gamma u_{k+2,l+1} + u_{k+2,l})]\]
\[= E(\gamma^2 u_{k+1,l}^2)\]
\[= \gamma^2 \sigma_u^2,\]

we have

\[\rho_{1,0} = \frac{\text{Cov}(\epsilon_{k,l}, \epsilon_{k+1,l})}{\text{Var}(\epsilon_{k,l})} = \frac{2\gamma \sigma_u^2}{(1 + 4\gamma^2)\sigma_u^2} = \frac{2\gamma}{(1 + 4\gamma^2)} = \rho,\]

\[\rho_{1,1} = \frac{\text{Cov}(\epsilon_{k,l}, \epsilon_{k+1,l+1})}{\text{Var}(\epsilon_{k,l})} = \frac{2\gamma^2 \sigma_u^2}{(1 + 4\gamma^2)\sigma_u^2} = \frac{2\gamma}{(1 + 4\gamma^2)} = \gamma \rho, \text{ and}\]

\[\rho_{2,0} = \frac{\text{Cov}(\epsilon_{k,l}, \epsilon_{k+2,l})}{\text{Var}(\epsilon_{k,l})} = \frac{\gamma^2 \sigma_u^2}{(1 + 4\gamma^2)\sigma_u^2} = \frac{1}{2}\gamma \rho = \frac{1}{2}\gamma \rho.\]
The correlation matrix $V_j$ for the MA(1) has the following forms. For $n = 1$,

$$V_j = \begin{pmatrix}
1 & \rho & \frac{1}{2} \gamma \rho \\
\rho & 1 & \rho & \frac{1}{2} \gamma \rho \\
\frac{1}{2} \gamma \rho & \rho & 1 & \rho & \frac{1}{2} \gamma \rho \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{2} \gamma \rho & \rho & 1 & \rho & \frac{1}{2} \gamma \rho \\
\frac{1}{2} \gamma \rho & \rho & 1 & \rho & \frac{1}{2} \gamma \rho \\
\end{pmatrix}_{t \times t},$$

and for $n = 2$,

$$V_j = \begin{pmatrix}
1 & \rho & \rho & \gamma \rho & \frac{1}{2} \gamma \rho \\
\rho & 1 & \gamma \rho & \rho & 0 & \frac{1}{2} \gamma \rho \\
\rho & \gamma \rho & 1 & \rho & \rho & \gamma \rho & \frac{1}{2} \gamma \rho \\
\gamma \rho & \rho & \rho & 1 & \gamma \rho & \rho & 0 & \frac{1}{2} \gamma \rho \\
\frac{1}{2} \gamma \rho & 0 & \rho & \gamma \rho & 1 & \rho & \rho & \gamma \rho & \frac{1}{2} \gamma \rho \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{2} \gamma \rho & 0 & \rho & \rho & 1 & \gamma \rho & \rho & \gamma \rho & \frac{1}{2} \gamma \rho \\
\frac{1}{2} \gamma \rho & \gamma \rho & \rho & \gamma \rho & 1 & \rho & \rho & 0 \\
\frac{1}{2} \gamma \rho & 0 & \rho & \rho & 1 & \gamma \rho & \rho & \gamma \rho & \frac{1}{2} \gamma \rho \\
\frac{1}{2} \gamma \rho & \gamma \rho & \rho & \gamma \rho & 1 & \rho & \rho & 0 \\
\frac{1}{2} \gamma \rho & 0 & \rho & \rho & 1 & \gamma \rho & \rho & \gamma \rho & \frac{1}{2} \gamma \rho \\
\end{pmatrix}_{t \times t}.$$

## 2.3 Doubly Geometric

The Doubly Geometric (DG) model is an extension of the first-order autoregressive process to the plane (Becher, 1988). The model has some nice properties and one of
them is that the correlation matrix can be represented as the Kronecker product of the correlation matrices of two corresponding one-dimensional processes. As well, the correlations are not proportional with distance. The process is given by

$$\epsilon_{k,l} = \lambda_1 \epsilon_{k-1,l} + \lambda_2 \epsilon_{k,l-1} - \lambda_1 \lambda_2 \epsilon_{k-1,l-1} + u_{k,l},$$  \hspace{1cm} (2.1)$$

where $\lambda_1$ and $\lambda_2$ are parameters of the DG process and $u_{k,l}$ are i.i.d with mean 0 and variance $\sigma_u^2$. The correlation function is

$$\rho_{g,h} = \lambda_1^{\left|g\right|} \lambda_2^{\left|h\right|},$$  \hspace{1cm} (2.2)$$

which generalizes the autocorrelation structure of the one-dimensional Markov process (Martin, 1979). This correlation can be proven by induction and derivations are shown below for positive $g$ and $h$.

For $g = 1$ and $h = 0$, multiplying the process in (2.1) by $\epsilon_{k-1,l}$ on both sides gives

$$\epsilon_{k-1,l} \epsilon_{k,l} = \lambda_1 \epsilon_{k-1,l}^2 + \lambda_2 \epsilon_{k-1,l} \epsilon_{k,l-1} - \lambda_1 \lambda_2 \epsilon_{k-1,l} \epsilon_{k,l-1} - \epsilon_{k-1,l} u_{k,l}.$$  \hspace{1cm} (2.3)$$

Taking the expectation of both sides and realizing that white noise, $u_{i,j}$, is independent of $\epsilon_{k,l}$ when plot $(k, l)$ is to the left or above plot $(i, j)$, we get,

$$\text{Cov}(\epsilon_{k-1,l}, \epsilon_{k,l}) = \lambda_1 \sigma^2_\epsilon + \lambda_2 \text{Cov}(\epsilon_{k-1,l}, \epsilon_{k,l-1}) - \lambda_1 \lambda_2 \text{Cov}(\epsilon_{k-1,l}, \epsilon_{k-1,l-1})$$

$$\sigma^2_\epsilon \rho_{1,0} = \lambda_1 \sigma^2_\epsilon + \lambda_2 \sigma^2_\epsilon \rho_{1,1} - \lambda_1 \lambda_2 \sigma^2_\epsilon \rho_{0,1}.$$  \hspace{1cm} (2.3)$$
Similarly for $g = 0$ and $h = 1$,

$$\epsilon_{k,l-1} \epsilon_{k,l} = \lambda_1 \epsilon_{k,l-1} \epsilon_{k-1,l} + \lambda_2 \epsilon_{k,l-1}^2 - \lambda_1 \lambda_2 \epsilon_{k,l-1} \epsilon_{k-1,l-1} + \epsilon_{k,l-1} u_{k,l}$$

$$\text{Cov}(\epsilon_{k,l-1}, \epsilon_{k,l}) = \lambda_1 \text{Cov}(\epsilon_{k,l-1}, \epsilon_{k-1,l}) + \lambda_2 \sigma_\epsilon^2 - \lambda_1 \lambda_2 \text{Cov}(\epsilon_{k,l-1}, \epsilon_{k-1,l-1})$$

$$\sigma_\epsilon^2 \rho_{0,1} = \lambda_1 \sigma_\epsilon^2 \rho_{1,1} + \lambda_2 \sigma_\epsilon^2 - \lambda_1 \lambda_2 \sigma_\epsilon^2 \rho_{1,0}.$$  \hspace{1cm} (2.4)

and for $g = 1$ and $h = 1$,

$$\epsilon_{k-1,l-1} \epsilon_{k,l} = \lambda_1 \epsilon_{k-1,l-1} \epsilon_{k-1,l} + \lambda_2 \epsilon_{k-1,l-1}^2 - \lambda_1 \lambda_2 \epsilon_{k-1,l-1} \epsilon_{k-1,l-1} + \epsilon_{k-1,l-1} u_{k,l}$$

$$\text{Cov}(\epsilon_{k-1,l-1}, \epsilon_{k,l}) = \lambda_1 \text{Cov}(\epsilon_{k-1,l-1}, \epsilon_{k-1,l}) + \lambda_2 \text{Cov}(\epsilon_{k-1,l-1}, \epsilon_{k-1,l}) - \lambda_1 \lambda_2 \sigma_\epsilon^2$$

$$\sigma_\epsilon^2 \rho_{1,1} = \lambda_1 \sigma_\epsilon^2 \rho_{1,0} + \lambda_2 \sigma_\epsilon^2 - \lambda_1 \lambda_2 \sigma_\epsilon^2.$$  \hspace{1cm} (2.5)

From (2.3), (2.4), and (2.5), we can solve for $\rho_{0,1}$, $\rho_{1,0}$, and $\rho_{1,1}$ as $\rho_{1,1} = \lambda_1 \lambda_2$, $\rho_{1,0} = \lambda_1$ and $\rho_{0,1} = \lambda_2$.

The correlation function obviously holds for the trivial case of $g = 0$ and $h = 0$ since $\rho_{0,0} = \lambda_1^0 \lambda_2^0 = 1$. Thus, the correlation function holds for $g + h \leq 1$. By induction, we assume the correlation function (2.2) is true for $g + h \leq m + n$ and need to show that it is true for $g + h = m + n + 1$.

For $g + h = m + n + 1$,

$$\epsilon_{i-g,j-h} \epsilon_{i,j} = \lambda_1 \epsilon_{i-g,j-h} \epsilon_{i-1,j} + \lambda_2 \epsilon_{i-g,j-h} \epsilon_{i,j-1} - \lambda_1 \lambda_2 \epsilon_{i-g,j-h} \epsilon_{i-1,j-1} + \epsilon_{i-g,j-h} u_{i,j}$$

$$\text{Cov}(\epsilon_{i-g,j-h}, \epsilon_{i,j}) = \lambda_1 \text{Cov}(\epsilon_{i-g,j-h}, \epsilon_{i-1,j}) + \lambda_2 \text{Cov}(\epsilon_{i-g,j-h}, \epsilon_{i,j-1}) - \lambda_1 \lambda_2 \text{Cov}(\epsilon_{i-g,j-h}, \epsilon_{i-1,j-1})$$

$$\sigma_\epsilon^2 \rho_{g,h} = \lambda_1 \sigma_\epsilon^2 \rho_{g-1,h} + \lambda_2 \sigma_\epsilon^2 \rho_{g,h-1} - \lambda_1 \lambda_2 \sigma_\epsilon^2 \rho_{g-1,h-1}$$

$$\rho_{g,h} = \lambda_1 \rho_{g-1,h} + \lambda_2 \rho_{g,h-1} - \lambda_1 \lambda_2 \rho_{g-1,h-1}$$.
\[ \rho_{g,h} = \lambda_1^{g-1} \lambda_2^h + \lambda_2 \lambda_1^g \lambda_2^{h-1} - \lambda_1 \lambda_2 \lambda_1^{g-1} \lambda_2^{h-1} \]

\[ \rho_{g,h} = \lambda_1^g \lambda_2^h, \]

which is the correlation function for \( g + h = m + n + 1 \). The result holds so our proof by induction is complete.

For square plots, \( \lambda_1 = \lambda_2 = \lambda \) and \( \rho_{g,h} = \lambda^{|g|+|h|} \) with \( \lambda \in (0, 1) \) (Becher, 1988). For non-square plots, an option is taking \( \lambda_1 = \lambda^2 \) and \( \lambda_2 = \sqrt{\lambda} \) when a rectangular plot has doubled length and halved width.

The correlation structure for the DG when \( n = 1 \) is,

\[
\mathbf{V}_j = \mathbf{V}_j(\lambda, t) \triangleq \begin{pmatrix}
1 & \lambda & \lambda^2 & \ldots & \lambda^{t-3} & \lambda^{t-2} & \lambda^{t-1} \\
\lambda & 1 & \lambda & \ldots & \lambda^{t-4} & \lambda^{t-3} & \lambda^{t-2} \\
\lambda^2 & \lambda & 1 & \ldots & \lambda^{t-5} & \lambda^{t-4} & \lambda^{t-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\lambda^{t-3} & \lambda^{t-4} & \lambda^{t-5} & \ldots & 1 & \lambda & \lambda^2 \\
\lambda^{t-2} & \lambda^{t-3} & \lambda^{t-4} & \ldots & \lambda & 1 & \lambda \\
\lambda^{t-1} & \lambda^{t-2} & \lambda^{t-3} & \ldots & \lambda^2 & \lambda & 1
\end{pmatrix}_{t \times t}, \quad (2.6)
\]
and for $n = 2$,

\[
V_j = \begin{pmatrix}
1 & \lambda & \lambda^2 & \lambda^2 & \ldots & \lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-1} & \lambda^{\frac{1}{2}-1} & \lambda^{\frac{1}{2}} \\
\lambda & 1 & \lambda^2 & \lambda & \lambda & \ldots & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}} & \lambda^{\frac{1}{2}-1} \\
\lambda & \lambda^2 & 1 & \lambda & \lambda & \ldots & \lambda^{\frac{1}{2}-4} & \lambda^{\frac{1}{2}-4} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-4} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-4} & \lambda^{\frac{1}{2}-3} \\
\lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-4} & \lambda^{\frac{1}{2}-4} & \ldots & 1 & \lambda^2 & \lambda & \lambda^3 & \lambda^2 \\
\lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-1} & \lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-4} & \ldots & \lambda^2 & 1 & \lambda & \lambda & \lambda^2 \\
\lambda^{\frac{1}{2}-1} & \lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-4} & \ldots & \lambda & \lambda & 1 & \lambda^2 & \lambda & \lambda \\
\lambda^{\frac{1}{2}-1} & \lambda^{\frac{1}{2}} & \lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-3} & \lambda^{\frac{1}{2}-4} & \ldots & \lambda^2 & \lambda & 1 & \lambda & \lambda \\
\lambda^{\frac{1}{2}} & \lambda^{\frac{1}{2}-1} & \lambda^{\frac{1}{2}-1} & \lambda^{\frac{1}{2}-2} & \lambda^{\frac{1}{2}-2} & \ldots & \lambda^2 & \lambda & \lambda & 1 & 1
\end{pmatrix}_{\times n}
\]

2.4 Discrete Exponential

The Discrete Exponential (DE) correlation structure is a discrete version of the continuous planar exponential process (Duby, Guyon, and Prum, 1977). The correlation between two plots decreases exponentially with distance $s$, $\rho(s) = e^{as}$, $a = -\ln \lambda$, which gives the correlation function (Becher, 1988),

\[
\rho_{g,h} = \lambda \sqrt{g^2 + h^2},
\]

where $\lambda \in (0, 1)$. We use $\nu = 1$ for square plots which gives the correlation function of $\rho_{g,h} = \lambda \sqrt{g^2 + h^2}$.

The DE correlation structure for $n = 1$ is identical to that of the DG model. The
two column \((n = 2)\) DE correlation structure for block \(j\) is given by

\[
V_j = \begin{pmatrix}
1 & \lambda & \lambda & \ldots & \lambda \sqrt{\frac{t^2}{2} - 2} & \lambda^\frac{t}{2} - 2 & \lambda \sqrt{\frac{t^2}{2} - 1} & \lambda^\frac{t}{2} - 1 & \lambda \sqrt{\frac{t^2}{2} - 1} \\
\lambda & 1 & \lambda \sqrt{\frac{t}{2}} & \ldots & \lambda^\frac{t}{2} - 2 & \lambda \sqrt{\frac{t^2}{2} - 1} & \lambda^\frac{t}{2} - 1 & \lambda \sqrt{\frac{t^2}{2} - 1} \\
\lambda \sqrt{\frac{t^2}{2} - 1} & \lambda \sqrt{\frac{t^2}{2} - 2} & 1 & \ldots & \lambda \sqrt{\frac{t^2}{2} - 3} & \lambda^\frac{t}{2} - 2 & \lambda \sqrt{\frac{t^2}{2} - 1} & \lambda^\frac{t}{2} - 1 \\
\lambda \sqrt{\frac{t^2}{2} - 1} & \lambda \sqrt{\frac{t^2}{2} - 2} & \lambda \sqrt{\frac{t^2}{2} - 3} & \ldots & 1 & \lambda \sqrt{\frac{t^2}{2} - 2} & \lambda^\frac{t}{2} - 1 & \lambda \sqrt{\frac{t^2}{2} - 1} \\
\lambda \sqrt{\frac{t^2}{2} - 1} & \lambda \sqrt{\frac{t^2}{2} - 2} & \lambda \sqrt{\frac{t^2}{2} - 3} & \ldots & \lambda & \lambda & \lambda & \lambda
\end{pmatrix}_{t \times t}
\]

2.5 Examples of Optimal Designs with No Blocks

In this section, we present some optimal designs for field experiments with one block using the NN(1), MA(1), DG, and DE correlation structures. Martin (1986) displayed some optimal designs for \(m = n = t = r\) for \(t = 3, 4, \text{and} 5\) using the LSE and the GLSE. Figure 2.2 presents two optimal designs for \(t = 3\) under both estimators. The design in (a) is \(\xi_A\) using the NN(1), MA(1), DG, and DE. This design is also \(\xi_D\) under the MA(1), DG, and DE. The design in (b) is \(\xi_D\) using the NN(1).

![Figure 2.2: Some optimal 3 x 3 designs.](image)

Basic 4 \times 4 square designs have also been analyzed in Becher (1988) and Martin (1986). Some optimal designs considered by Becher and Martin are displayed in Figure 2.3. The design in (a) is generally \(\xi_{A,LSE}\) for both the DG and DE. However, for weak correlations, \(\lambda < 0.221\), design (b) is \(\xi_{A,LSE}\) using the DE. The design (a) is
also $\xi_{D,GLSE}$ for the DE and $\xi_{D,LSE}$ for the DG. The design in (b) is $\xi_{D,LSE}$ using the DE correlation structure for low correlations of $\lambda < 0.275$, otherwise, design (a) is $\xi_{D,LSE}$. Design (c) is $\xi_{D,LSE}$ for both the NN(1) and DG. The design in (d) is $\xi_{D,GLSE}$ for the NN(1), MA(1), and DG structures.

\begin{align*}
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
3 & 4 & 1 & 2 \\
2 & 1 & 4 & 3 \\
4 & 3 & 2 & 1 \\
\end{array} & \begin{array}{cccc}
1 & 2 & 3 & 4 \\
3 & 4 & 1 & 2 \\
1 & 2 & 3 & 4 \\
3 & 4 & 1 & 2 \\
\end{array} & \begin{array}{cccc}
1 & 2 & 3 & 4 \\
2 & 1 & 4 & 3 \\
3 & 4 & 1 & 2 \\
4 & 3 & 2 & 1 \\
\end{array} & \begin{array}{cccc}
1 & 2 & 3 & 4 \\
3 & 1 & 4 & 2 \\
2 & 4 & 1 & 3 \\
4 & 3 & 2 & 1 \\
\end{array} \\
\end{array}

(a) \hspace{1cm} (b) \hspace{1cm} (c) \hspace{1cm} (d)

Figure 2.3: Some optimal $4 \times 4$ designs.

A Latin square design is a design where each treatment appears exactly once in each row and each column. Figure 1.1 is a Latin square design as are Figure 2.2(a) and Figure 2.3(a), (c), and (d). A special case of the Latin square design is the Knight’s move Latin square design where each treatment is connected by a Knight’s move. This design was suggested as early as 1849 by J. F. W. Johnson who realized the importance of separating replicates of the same treatment but became more popular in the 1920s from work by a Norwegian, Knut Vik (Cochran, 1976). However, randomization was not used as the Knight’s move dictated each position of the treatments. Martin (1986) identifies the Knight’s move Latin square, shown as the design in Figure 2.4(a), as an optimal design for $5 \times 5$ designs. The Knight’s move Latin square is $\xi_{A,LSE}$ for the MA(1), DG, and DE correlation structures. It is also $\xi_{D,LSE}$ for the DG and DE structures and $\xi_{D,GLSE}$ using the DE. The design in Figure 2.4(b), a diagonal square, is $\xi_{D,LSE}$ for the NN(1), MA(1), and DG correlation structures.

Field experiments are usually laid out in square formations so that soil fertility and other variations in two directions are controlled. With the assumption of uniformity in all directions and when $m = n = t = r$, the Latin square design is a good choice
Figure 2.4: Knight’s move Latin square designs.

for small valued $t$ such as $t < 8$. Generally, the designs are in the range of $5 \times 5$ to $8 \times 8$, and squares greater than $12 \times 12$ are not used (Cochran and Cox, 1960). Just as with randomized blocks, the experimental error increases with the size of the square. However, these kinds of Latin square designs only take place in one area and therefore may not take the entire field into consideration. Hence the need for block experiments.
Chapter 3

Properties of Block Designs

In this chapter, we discuss some basic properties of block designs including the labelling of blocks and treatments. In Section 3.1, we find that the order of two blocks has no impact on the covariance of the estimators of the treatment effects and Section 3.2 discusses the labelling of the treatments. Section 3.3 presents a result for the orientation of two blocks. Furthermore, an example is given for an experiment with three treatments and two blocks in Sections 3.4 and 3.5. Section 3.4 shows the optimal designs with no control, while Section 3.5 presents the optimal design with a control treatment. The optimal designs $\xi_{\nu,\psi}$ are found for all four correlation structures.

3.1 Labelling of Blocks

Lemma 3.1. For $b = 2$, the labelling of the two blocks has no impact on the covariance of $C\hat{\theta}$ where $\hat{\theta}$ is the LSE or the GLSE.

Proof. We will prove the result for the LSE first. For $b = 2$ in linear model (1.1), i.e. $y = X\theta + \epsilon$, the parameter vector $\theta = (\mu_1, \ldots, \mu_t, \beta_1)^\top$ and the last column of the design matrix $X$ is $(1, \ldots, 1, -1, \ldots, -1)^\top$. The LSE for $\theta$ is given in (1.2) and the
covariance of $\mathbf{C}\hat{\theta}$ in (1.3). If we switch the block labels and still want to estimate $\theta = (\mu_1, \ldots, \mu_t, \beta_1)^\top$, then the linear model (1.1) becomes

$$y = Z\theta + \epsilon,$$  \hspace{1cm} (3.1)

where the vectors $y$ and $\epsilon$ are the same as in (1.1) but the design matrix becomes $Z = XQ$ and

$$Q = \begin{pmatrix} I_t & 0 \\ 0^\top & -1 \end{pmatrix}$$

is a $(t + 1) \times (t + 1)$ matrix. From (3.1), the LSE for $\theta$ is given by

$$\hat{\theta}_{LSE}^* = (Z^\top Z)^{-1}Z^\top y,$$

and thus

$$\text{Cov}(\hat{\theta}_{LSE}^*) = \sigma^2(C(Z^\top Z)^{-1}Z^\top VZ(Z^\top Z)^{-1}C^\top).$$

Since $Q^\top = Q$, $Q^{-1} = Q$, $Q^2 = I_{t+1}$, $CQ = C$, and $QC^\top = C^\top$, we have

$$\text{Cov}(\hat{\theta}_{LSE}^*) = \sigma^2(C(Z^\top Z)^{-1}Z^\top VZ(Z^\top Z)^{-1}C^\top)$$

$$= \sigma^2C[(XQ)^\top(XQ)]^{-1}(XQ)^\top V(XQ)[(XQ)^\top(XQ)]^{-1}C^\top$$

$$= \sigma^2CQ(X^\top X)^{-1}QQX^\top VXQ(X^\top X)^{-1}QC^\top$$

$$= \sigma^2C(X^\top X)^{-1}X^\top VX(X^\top X)^{-1}C^\top$$

$$= \text{Cov}(\hat{\theta}_{LSE}).$$

The result for the GLSE can be proved similarly. From (3.1), the GLSE of $\theta$ is given by

$$\hat{\theta}_{GLSE}^* = (Z^\top V^{-1}Z)^{-1}Z^\top V^{-1}y,$$
and so
\[
\text{Cov}(\hat{C}_{\theta}^*_{GLSE}) = \sigma^2 C(Z^TV^{-1}Z)^{-1}C^T.
\]

Therefore,
\[
\text{Cov}(\hat{C}_{\theta}^*_{GLSE}) = \sigma^2 C(Z^TV^{-1}Z)^{-1}C^T
= \sigma^2 C[(XQ)^TV^{-1}(XQ)]^{-1}C^T
= \sigma^2 CQ(X^TV^{-1}X)^{-1}QC^T
= \sigma^2 C(X^TV^{-1}X)^{-1}C^T
= \text{Cov}(\hat{C}_{\theta}^*_{GLSE}).
\]

Consequently, the labelling of the two blocks does not impact the covariance of \(\hat{C}_{\theta}\) where \(\hat{\theta}\) is the LSE or the GLSE.

\[\square\]

### 3.2 Labelling of Treatments

**Lemma 3.2.** For \(b = 2\), the labelling of the treatments has no impact on the determinant or trace of the covariance of \(\hat{C}_{\theta}\), where \(\hat{\theta}\) is the LSE or the GLSE.

**Proof.** We will prove the results similar to Lemma 3.1 and show the results for the LSE first. For \(b = 2\) in linear model (1.1), the parameter vector \(\theta = (\mu_1, \ldots, \mu_t, \beta_1)^T\). The LSE for \(\theta\) is given in (1.2) and covariance of \(\hat{C}_{\theta}\) in (1.3). Changing the labels of the treatments is equivalent to changing the linear model, so model (1.1) becomes

\[
y = X\theta + \epsilon,
\]

(3.2)
where $P$ is a $(t + 1) \times (t + 1)$ matrix such that,

$$P = \begin{pmatrix} P_1 & 0 \\ 0^\top & 1 \end{pmatrix}$$

and $P_1$ is a $t \times t$ permutation matrix of the treatments. Since $P_1^\top = P_1^{-1}$, it is obvious that $P^\top = P^{-1}$. Denote the LSE for $\theta$ from model (3.2) as

$$\hat{\theta}_{LSE}^* = [(XP)^\top(XP)]^{-1}(XP)^\top y,$$

and thus

$$\text{Cov}(\hat{\theta}_{LSE}^*) = \sigma^2 C[(XP)^\top(XP)]^{-1}(XP)^\top V(XP)[(XP)^\top(XP)]^{-1} C^\top$$

$$= \sigma^2 C[P^\top X^\top X P^{-1} P^\top X^\top V X P^{-1} P^\top X^\top X P^{-1} C^\top]$$

$$= \sigma^2 CP^{-1} (X^\top X)^{-1} (P^\top)^{-1} P^\top X^\top V X P^{-1} P^\top X^\top X (X^\top X)^{-1} (P^\top)^{-1} C^\top$$

$$= \sigma^2 CP^{-1} (X^\top X)^{-1} X^\top V X (X^\top X)^{-1} PC^\top$$

$$= CP^{-1} \cdot \text{Cov}(\hat{\theta}_{LSE}) \cdot PC^\top.$$

Now, $\text{Cov}(\hat{\theta})$ can be rewritten as

$$\text{Cov}(\hat{\theta}) = \text{Cov}\left( \begin{pmatrix} \hat{\mu} \\ \hat{\beta}_1 \end{pmatrix} \right)$$

$$= \begin{pmatrix} \text{Cov}(\hat{\mu}) & \text{Cov}(\hat{\mu}, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_1, \hat{\mu}) & \text{Var}(\hat{\beta}_1) \end{pmatrix}.$$

(3.3)
Therefore,

\[
\text{Cov}(\mathbf{C}\hat{\theta}^*_{\text{LSE}}) = \mathbf{CP}^{-1} \cdot \text{Cov}(\hat{\theta}_{\text{LSE}}) \cdot \mathbf{PC}^\top
\]

\[
= \begin{pmatrix}
I_t & 0 \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
P_1^{-1} & 0 \\
0^\top & 1
\end{pmatrix}
\begin{pmatrix}
\text{Cov}(\hat{\mu}_{\text{LSE}}) & \text{Cov}(\hat{\mu}_{\text{LSE}}, \hat{\beta}_{1(\text{LSE})}) \\
\text{Cov}(\hat{\beta}_{1(\text{LSE})}, \hat{\mu}_{\text{LSE}}) & \text{Var}(\hat{\beta}_{1(\text{LSE})})
\end{pmatrix}
\begin{pmatrix}
P_1 & 0 \\
0^\top & 1
\end{pmatrix}
\begin{pmatrix}
I_t \\
0^\top
\end{pmatrix}
\]

\[
= \mathbf{P}_1^{-1} \text{Cov}(\hat{\mu}_{\text{LSE}}) \mathbf{P}_1
\]

\[
= \mathbf{P}_1^{-1} \text{Cov}(\mathbf{C}\hat{\theta}_{\text{LSE}}) \mathbf{P}_1.
\]

The trace can easily be evaluated as

\[
\text{trace}[\text{Cov}(\mathbf{C}\hat{\theta}^*_{\text{LSE}})] = \text{trace}[\mathbf{P}_1^{-1} \cdot \text{Cov}(\mathbf{C}\hat{\theta}_{\text{LSE}}) \cdot \mathbf{P}_1]
\]

\[
= \text{trace}[\text{Cov}(\mathbf{C}\hat{\theta}_{\text{LSE}}) \cdot \mathbf{P}_1 \mathbf{P}_1^{-1}]
\]

\[
= \text{trace}[\text{Cov}(\mathbf{C}\hat{\theta}_{\text{LSE}})],
\]

and so can the determinant,

\[
\text{det}[\text{Cov}(\mathbf{C}\hat{\theta}^*_{\text{LSE}})] = \text{det}[\mathbf{P}_1^{-1} \cdot \text{Cov}(\mathbf{C}\hat{\theta}_{\text{LSE}}) \cdot \mathbf{P}_1]
\]

\[
= \text{det}(\mathbf{P}_1^{-1}) \cdot \text{det}[\text{Cov}(\mathbf{C}\hat{\theta}_{\text{LSE}})] \cdot \text{det}(\mathbf{P}_1)
\]

\[
= \text{det}[\text{Cov}(\mathbf{C}\hat{\theta}_{\text{LSE}})].
\]
Similarly for the GLSE, from model (3.2),

$$\hat{\theta}_{GLSE}^* = [(XP)^\top V^{-1}(XP)]^{-1}(XP)^\top V^{-1}y,$$

and thus,

$$\text{Cov}(\mathbf{C}\hat{\theta}_{GLSE}^*) = \sigma^2 \mathbf{C} \left[(XP)^\top V^{-1}(XP)\right]^{-1} \mathbf{C}^\top$$

$$= \sigma^2 \mathbf{C} \left[\mathbf{P}^\top X^\top V^{-1}XP\right]^{-1} \mathbf{C}^\top$$

$$= \sigma^2 \mathbf{C} \mathbf{P}^{-1} (X^\top V^{-1}X)^{-1} \mathbf{PC}^\top$$

$$= \mathbf{CP}^{-1} \cdot \text{Cov}(\hat{\theta}_{GLSE}) \cdot \mathbf{PC}^\top.$$

Using (3.3), we can see that

$$\text{Cov}(\mathbf{C}\hat{\theta}_{GLSE}^*) = \begin{pmatrix}
\mathbf{I}_t & 0 \\
0 & \mathbf{P}^{-1}_1 \mathbf{0} \\
\mathbf{0}^\top & 1
\end{pmatrix}
\begin{pmatrix}
\mathbf{P}^{-1}_1 \\
\mathbf{0} \\
\mathbf{0}^\top
\end{pmatrix}
\begin{pmatrix}
\text{Cov}(\hat{\mu}_{GLSE}) & \text{Cov}(\hat{\mu}_{GLSE}, \hat{\beta}_{1(GLSE)}) \\
\text{Cov}(\hat{\beta}_{1(GLSE)}, \hat{\mu}_{GLSE}) & \text{Var}(\hat{\beta}_{1(GLSE)})
\end{pmatrix}
\begin{pmatrix}
\mathbf{I}_t \\
\mathbf{0}^\top \\
1
\end{pmatrix}
\begin{pmatrix}
\mathbf{P}_1 \\
\mathbf{0}^\top \\
\mathbf{1}
\end{pmatrix}
$$

$$= \mathbf{P}_1^{-1} \text{Cov}(\hat{\mu}_{GLSE}) \mathbf{P}_1$$

$$= \mathbf{P}_1^{-1} \text{Cov}(\mathbf{C}\hat{\theta}_{GLSE}) \mathbf{P}_1.$$
The trace follows as

\[
\text{trace}[\text{Cov}(\hat{\mathbf{C}}^{*}_{\text{GLSE}})] = \text{trace}[\mathbf{P}_1^{-1} \cdot \text{Cov}(\hat{\mathbf{C}}_{\text{GLSE}}) \cdot \mathbf{P}_1] \\
= \text{trace}[\text{Cov}(\hat{\mathbf{C}}_{\text{GLSE}}) \cdot \mathbf{P}_1 \mathbf{P}_1^{-1}] \\
= \text{trace}[\text{Cov}(\hat{\mathbf{C}}_{\text{GLSE}})],
\]

and the determinant as

\[
\text{det}[\text{Cov}(\hat{\mathbf{C}}^{*}_{\text{GLSE}})] = \text{det}[\mathbf{P}_1^{-1} \cdot \text{Cov}(\hat{\mathbf{C}}_{\text{GLSE}}) \cdot \mathbf{P}_1] \\
= \text{det}(\mathbf{P}_1^{-1}) \cdot \text{det}[\text{Cov}(\hat{\mathbf{C}}_{\text{GLSE}})] \cdot \text{det}(\mathbf{P}_1) \\
= \text{det}[\text{Cov}(\hat{\mathbf{C}}_{\text{GLSE}})].
\]

Thus, the labelling of the treatments does not impact the determinant or trace of the covariance of \( \hat{\mathbf{C}} \). \qed

### 3.3 Orientation of Blocks

**Lemma 3.3.** For \( b = 2 \), the orientation of the blocks has no impact on the covariance of \( \hat{\mathbf{C}} \), where \( \hat{\theta} \) is the LSE or the GLSE.

**Proof.** The result will be proved for the LSE first. For \( b = 2 \) in linear model (1.1), that is \( \mathbf{y} = \mathbf{X}\theta + \epsilon \) where the parameter vector \( \theta = (\mu_1, \ldots, \mu_t, \beta_1)^\top \), the response vector \( \mathbf{y} = (y_{1,1,1}, \ldots, y_{m,1,1}, y_{1,1,2}, \ldots, y_{m,1,2})^\top \) and the residual vector \( \epsilon = (\epsilon_{1,1,1}, \ldots, \epsilon_{m,1,1}, \epsilon_{1,1,2}, \ldots, \epsilon_{m,1,2})^\top \). The design matrix \( \mathbf{X} \) can be thought of as two separate design matrices, corresponding to each of the blocks, such that \( \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \).
If we switch the orientation of block two, then the linear model (1.1) becomes

\[ y = Z\theta + \epsilon, \quad (3.4) \]

where the design matrix \( Z = \begin{pmatrix} X_1 \\ P_2X_2 \end{pmatrix} \) and the matrix \( P_2 \) is a \( t \times t \) matrix such that \( P_2 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \).

The LSE for \( \theta \) from model (1.1) is given in (1.2) and the covariance of \( \hat{\theta} \) in (1.3). The LSE for \( \theta \) from model (3.4) is

\[ \hat{\theta}_{LSE}^* = (Z^\top Z)^{-1}Z^\top y, \]

and thus,

\[ \text{Cov}(\hat{\theta}_{LSE}^*) = \sigma^2 C(Z^\top Z)^{-1}Z^\top VZ(Z^\top Z)^{-1}C^\top. \]

It is easy to see that

\[ Z^\top Z = \begin{pmatrix} X_1^\top & X_2^\top P_2^\top \end{pmatrix} \begin{pmatrix} X_1 \\ P_2X_2 \end{pmatrix} \]

\[ = X_1^\top X_1 + X_2^\top P_2 P_2X_2 \]

\[ = X_1^\top X_1 + X_2^\top X_2 \]

\[ = X^\top X, \]
and

\[ Z^\top VZ = \left( X_1^\top \quad X_2^\top P_2^\top \right) \cdot \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ P_2 X_2 \end{pmatrix} = X_1^\top V_1 X_1 + X_2^\top P_2^\top V_2 P_2 X_2 = X^\top VX. \]

Therefore,

\[
\text{Cov}(\hat{C} \theta^*_{LSE}) = \sigma^2 C (Z^\top Z)^{-1} Z^\top VZ (Z^\top Z)^{-1} C^\top = \sigma^2 C (X^\top X)^{-1} X^\top VX (X^\top X)^{-1} C^\top = \text{Cov}(\hat{C} \theta^*_{LSE}),
\]

and the orientation of the blocks does not affect the covariance of \( C \hat{\theta} \) for the LSE.

The GLSE for \( \theta \) for model (1.1) is given in (1.5) and the covariance of \( C \hat{\theta} \) in (1.6). The GLSE for \( \theta \) from (3.4) is given by

\[
\hat{\theta}^*_{GLSE} = (Z^\top V^{-1} Z)^{-1} Z^\top V^{-1} y,
\]

so that

\[
\text{Cov}(\hat{\theta}^*_{GLSE}) = \sigma^2 (Z^\top V^{-1} Z)^{-1},
\]

and

\[
\text{Cov}(C \hat{\theta}^*_{GLSE}) = \sigma^2 C (Z^\top V^{-1} Z)^{-1} C^\top.
\]
Examining $Z^\top V^{-1}Z$, we can see that

$$Z^\top V^{-1}Z = \begin{pmatrix} X_1^\top & X_2^\top P_2^\top \end{pmatrix} \cdot \begin{pmatrix} V_1^{-1} & 0 \\ 0 & V_2^{-1} \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ P_2 X_2 \end{pmatrix}$$

$$= X_1^\top V_1^{-1} X_1 + X_2^\top P_2^\top V_2^{-1} P_2 X_2$$

$$= X_1^\top V_1^{-1} X_1 + X_2^\top V_2^{-1} X_2$$

$$= X^\top V^{-1} X.$$

Therefore,

$$\text{Cov}(\hat{C}\theta^{*}_{GLSE}) = \sigma^2 C(Z^\top V^{-1}Z)^{-1} C^\top$$

$$= \sigma^2 C(X^\top V^{-1}X)^{-1} C$$

$$= \text{Cov}(\hat{C}\theta^{*}_{GLSE}),$$

and the orientation of the blocks does not affect the covariance of $C\hat{\theta}$ for the GLSE.

Since block order and treatment labelling do not impact $\xi_A$ or $\xi_D$, we can fix the arrangement of the $t$ treatments in block one to find optimal arrangements in block two. Without loss of generality, for $n = 1$, the $t$ treatments in block one are set as in Figure 3.1. Furthermore, for a $m \times n$ array the first block is arranged $\{1, \ldots, t\}$ as in Figure 3.2.

$$\begin{array}{c}
\mathbf{1} \\
\mathbf{2} \\
\vdots \\
\mathbf{t}
\end{array}$$

Figure 3.1: Arrangement of block one with $n = 1$. 

\begin{figure}[h]
\centering
\begin{tabular}{c}
\hline
1 \\
2 \\
\vdots \\
t \\
\hline
\end{tabular}
\caption{Arrangement of block one with $n = 1$.}
\end{figure}
In the next two sections, we will present some optimal designs for $t = 3$, $n = 1$, and $b = 2$.

### 3.4 Examples of Optimal Designs with Two Blocks

Consider the designs for $t = 3$, $n = 1$, and $b = 2$. There are 36 possible designs to arrange the treatments in two blocks. From Sections 3.1 and 3.2, the block order and labelling of the treatments do not impact the loss functions. Block one can have its layout fixed as \{1, 2, 3\} and only block two will have different arrangements. This reduces to six different designs. However, there are only three designs which may yield different covariance matrices as the orientation of the blocks has no effect. Thus, we only need to consider the three designs in Figure 3.3 to find optimal designs according to the various criteria.

![Figure 3.3: The three designs for $t = 3$, $n = 1$, and $b = 2$.](image)

The parameter space is $\theta = (\mu_1, \mu_2, \mu_3, \beta_1)^\top$ in model (1.1), and the corresponding correlation matrix,

$$ V = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} $$
with $V_1 = V_2$. For the NN(1), MA(1), DG, and DE correlations, $V_1$ takes the following forms, respectively,

$$V_{1(NN(1))} = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix},$$

$$V_{1(MA(1))} = \begin{pmatrix} 1 & \rho & \frac{1}{2} \gamma \rho \\ \rho & 1 & \rho \\ \frac{1}{2} \gamma \rho & \rho & 1 \end{pmatrix}, \text{ and}$$

$$V_{1(DG)} = V_{1(DE)} = \begin{pmatrix} 1 & \lambda & \lambda^2 \\ \lambda & 1 & \lambda \\ \lambda^2 & \lambda & 1 \end{pmatrix}.$$

The design matrix, $X$, for the three designs in Figure 3.3 are given by $X_1$, $X_2$, and $X_3$ respectively and they are,

$$X_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \quad \text{and} \quad X_3 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Four optimal design criteria are considered here. In each criterion, we minimize a
loss function which is a scalar function of the covariance of an estimator.

\[ \xi_{A,LSE} : \text{the loss function is } \ell_1(X, V) = \text{trace}[\text{Cov}(\hat{C}^{LSE})] \]
\[ \xi_{D,LSE} : \text{the loss function is } \ell_2(X, V) = \text{det}[\text{Cov}(\hat{C}^{LSE})] \]
\[ \xi_{A,GLSE} : \text{the loss function is } \ell_3(X, V) = \text{trace}[\text{Cov}(\hat{C}^{GLSE})] \]
\[ \xi_{D,GLSE} : \text{the loss function is } \ell_4(X, V) = \text{det}[\text{Cov}(\hat{C}^{GLSE})] \]

Given \( V \) (i.e. correlation structures), optimal designs minimize the loss function over \( X \). For a given correlation structure, we can plot the loss function versus the correlation parameter for each design and determine the optimal design. For example, Figure 3.4 shows the loss functions versus \( \rho \) for the GLSE with the NN(1) correlation structure and we can see that designs d3.2 and d3.3 are optimal for all \( \rho \in [0, \frac{1}{4}] \). Table 3.1 lists the optimal designs for the LSE and the GLSE for the four correlation structures.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>NN(1)</th>
<th>MA(1)</th>
<th>DE and DG</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi_{A,LSE} )</td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>( \xi_{D,LSE} )</td>
<td>d3.1</td>
<td>d3.1</td>
<td>d3.1</td>
</tr>
<tr>
<td>( \xi_{A,GLSE} )</td>
<td>d3.2, d3.3</td>
<td>d3.2, d3.3</td>
<td>d3.2, d3.3</td>
</tr>
<tr>
<td>( \xi_{D,GLSE} )</td>
<td>d3.2, d3.3</td>
<td>d3.2, d3.3</td>
<td>d3.2, d3.3</td>
</tr>
</tbody>
</table>

Table 3.1: Optimal designs for \( t = 3, n = 1, \) and \( b = 2. \)

It is obvious that designs d3.2 and d3.3, which have different layouts for block two from block one, are \( \xi_{A,GLSE} \) and \( \xi_{D,GLSE} \) for all four correlation structures. The \( \xi_{D,LSE} \) is d3.1 where block two has the same arrangement as block one.
Figure 3.4: Loss functions versus $\rho$ for the GLSE with the NN(1) correlation structure.

### 3.5 Examples of Optimal Designs with a Control

One of the treatments may also be treated as a control. If a control is used in the experiment, we label it treatment one without loss of generality. Lemma 3.1 still holds so block order is not significant. However, for Lemma 3.2, the labelling of the control treatment will affect the final design but the arrangement of the other $t - 1$ plots will not. The linear model is represented as in (1.1) with some changes. The response
y and error $\epsilon$ remain the same as before, however, $\theta = (\mu_1, \tau_2, \ldots, \tau_t, \beta_1, \ldots, \beta_{b-1})^T$ is the parameter vector of length $t + b - 1$ where $\mu_1$ is the treatment mean for the control, $\tau_2, \ldots, \tau_t$ are the effects for treatments 2, $\ldots, t$, and $\beta_1, \ldots, \beta_{b-1}$ are the same as in model (1.1). Additionally, $X$ which is a $tb \times (t + b - 1)$ design matrix with entries as either -1, 0 or 1, has a few changes. The first column corresponds to the control mean and always contains an entry of 1. For the next $t - 1$ columns, column $i$ has $b$ entries of 1 corresponding to the $i^{th}$ treatment effect and the last $b - 1$ columns are similarly defined as in linear model (1.1). As well, $E(\epsilon) = 0$ and $\text{Cov}(\epsilon) = \sigma^2 V$, where $V$ is the same as in model (1.1). When finding $\xi_A$ and $\xi_D$, the block effects and $\mu_1$ are not considered. Thus, we take $\theta_1 = (\tau_2, \ldots, \tau_t)^T$ and find $\xi_{\nu,\psi}$ based on $\text{Cov}(\hat{\theta}_1)$.

Consider the example from Section 3.4 where $y = (y_{1,1,1}, y_{2,1,1}, y_{3,1,1}, y_{1,1,2}, y_{2,1,2}, y_{3,1,2})^T$ and $\epsilon$ is similarly defined. The parameter space $\theta = (\mu_1, \tau_2, \tau_3, \beta_1)^T$. The control treatment has two unique positions for the first block as the orientation and the labelling of the other $t - 1$ treatments does not matter. Therefore, the control treatment can either be in the middle of the block or at one end. There are three unique labelings for block two, resulting in six designs. However, one design is eliminated as block order has no impact. The five designs to analyze are shown in Figure 3.5.

![Figure 3.5: Designs with treatment one as a control for $t = 3$, $n = 1$, and $b = 2$.](image)

The correlation structures are the same as in Section 3.4 when there is no control, but the design matrices are different. For the five designs, the design matrices are,
respectively,

\[
X_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & -1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 \end{pmatrix},
\]

\[
X_4 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 \end{pmatrix}, \quad \text{and} \quad X_5 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 \end{pmatrix}.
\]

The covariance is calculated using the effects of treatments two and three only (i.e. \(\text{Cov}(\hat{\theta}_1)\)). Using the loss function plots, we find the optimal designs in Table 3.2 for the four criteria discussed in Section 3.4.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>NN(1)</th>
<th>MA(1)</th>
<th>DE and DG</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\xi_{A,LSE})</td>
<td>d3.5</td>
<td>d3.5</td>
<td>d3.5</td>
</tr>
<tr>
<td>(\xi_{D,LSE})</td>
<td>d3.1, d3.5</td>
<td>d3.1, d3.5</td>
<td>d3.1, d3.5</td>
</tr>
<tr>
<td>(\xi_{A,GLSE})</td>
<td>d3.5</td>
<td>d3.5</td>
<td>d3.5</td>
</tr>
<tr>
<td>(\xi_{D,GLSE})</td>
<td>d3.2, d3.3, d3.4</td>
<td>d3.2, d3.3, d3.4</td>
<td>d3.2, d3.3, d3.4</td>
</tr>
</tbody>
</table>

Table 3.2: Optimal designs with a control for \(t = 3, n = 1, \) and \(b = 2.\)

Figure 3.6 shows the loss functions versus \(\rho\) for the GLSE with NN(1) correlation when there is a control in the treatments. Figure 3.7 shows the loss functions versus \(\gamma\) for the LSE with MA(1) correlation when one of the treatments is a control.
From Table 3.2, it is obvious that d3.5 is $\xi_{A,LSE}$ and $\xi_{A,GLSE}$ for all correlation structures, which has the control treatment in the middle of both blocks and both blocks with identical layouts. The $\xi_{D, GLSE}$ for all four correlation structures gives the set of designs where block two does not have the same arrangement as block one. The $\xi_{D, LSE}$ gives the designs where block one and two have the same arrangement.

However, the exact correlation structure is not always known to be able to deter-
Figure 3.7: Loss functions versus $\gamma$ for the LSE with MA(1) correlation and a control.

mine $\xi_{\nu,\psi}$ exactly. We will present two neighbourhoods to allow for uncertainty in the correlation structures in the next chapter.
Chapter 4

Neighbourhoods of Spatial Correlations and Robust Design Criteria

In the previous chapter, we examined designs under given spatial correlation structures. However, the assumed correlation may not be accurate in practice as the actual correlation structure is unknown or assumed with some uncertainty. Neighbourhoods of correlation structures allow for some uncertainty in correlation parameters and structures and in the variance $\sigma^2$ (Ou and Zhou, 2009). Two neighbourhoods will be discussed in Section 4.1. Robust designs do not depend on the exact correlation structure and are efficient in a neighbourhood. Robust design criteria will be discussed in Section 4.2. In Section 4.3, some results for the robust criteria and loss functions are derived.

The need for robust designs stems from the fact that optimal designs depend on the exact correlation structure. For instance, consider the case of $t = 4$, $n = 1$, and $b = 2$ when treatment one is treated as a control. There are two sets of designs for
block one and twelve for the second block in each case. For the first block, the control treatment is either at one end of the block or is the second from the end. The second block consists of all \(\frac{4!}{2} = 12\) designs as orientation does not matter. This gives 24 arrangements in total but one design is eliminated as block order has no impact, thus only 23 unique designs remain. These 23 designs are shown in Figure 4.1. Table 4.1 presents \(\xi_{A, GLSE}\) and \(\xi_{D, GLSE}\) for various correlation structures. It is apparent that \(\xi_{A, GLSE}\) depends on the value of the parameter \(\lambda\) for the DG or DE correlation structures.

\[
\begin{array}{cccccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\
3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
\end{array}
\]

(a) d4.1 (b) d4.2 (c) d4.3 (d) d4.4 (e) d4.5 (f) d4.6

\[
\begin{array}{cccccccccccc}
1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\
2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
\end{array}
\]

(g) d4.7 (h) d4.8 (i) d4.9 (j) d4.10 (k) d4.11 (l) d4.12

\[
\begin{array}{cccccccccccc}
2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 \\
1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 \\
3 & 4 & 3 & 2 & 3 & 4 & 3 & 2 & 3 & 4 & 3 & 2 \\
4 & 3 & 4 & 2 & 4 & 3 & 2 & 4 & 3 & 4 & 2 & 4 \\
\end{array}
\]

(m) d4.13 (n) d4.14 (o) d4.15 (p) d4.16 (q) d4.17 (r) d4.18

\[
\begin{array}{cccccccccccc}
2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
\end{array}
\]

(s) d4.19 (t) d4.20 (u) d4.21 (v) d4.22 (w) d4.23

Figure 4.1: Designs with treatment one as a control for \(t = 4\), \(n = 1\), and \(b = 2\).

Another example is shown using six treatments \((t = 6)\) in one column \((n = 1)\) and two blocks \((b = 2)\) with a control treatment. There are three ways to arrange the first
block and 360 ways to arrange block two. Block one has three designs as the control is either at one end, second from the end or is in the middle and the arrangements of the other \( t - 1 \) treatments is not important. The second block has \( \frac{6!}{2} = 360 \) ways to be arranged as orientation of the blocks is not significant. When eliminating the equivalent design due to orientation, the designs are arranged in lexicographic order and the design with the higher value is eliminated. Thus, there are 1,080 possible designs to analyze. We will include the two designs which may be eliminated due to block order in order to maintain the numbering of the designs. The \( \xi_{A,GLSE} \) are shown in Figure 4.2 and Table 4.2 for various correlation structures.

<table>
<thead>
<tr>
<th>Correlation Structure</th>
<th>( \xi_{A,GLSE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN(1)</td>
<td>d6.657 for ( \rho \leq 0.123 )</td>
</tr>
<tr>
<td></td>
<td>d6.658 and d6.701 for ( 0.124 \leq \rho \leq 0.241 )</td>
</tr>
<tr>
<td></td>
<td>d6.525 and d6.853 for ( \rho \geq 0.242 )</td>
</tr>
<tr>
<td>MA(1)</td>
<td>d6.657 for ( \gamma \leq 0.163 )</td>
</tr>
<tr>
<td></td>
<td>d6.525 and d6.853 for ( 0.164 \leq \gamma \leq 0.217 )</td>
</tr>
<tr>
<td></td>
<td>d6.985 and d6.1023 for ( \gamma \geq 0.218 )</td>
</tr>
<tr>
<td>DG and DE</td>
<td>d6.515, d6.525, d6.853 and d6.891 for ( \lambda \leq 0.081 )</td>
</tr>
<tr>
<td></td>
<td>d6.657 and d6.905 for ( 0.082 \leq \lambda \leq 0.345 )</td>
</tr>
<tr>
<td></td>
<td>d6.643, d6.660, d6.988 and d6.1043 for ( \lambda \geq 0.346 )</td>
</tr>
</tbody>
</table>

Table 4.2: Optimal designs with a control for \( t = 6, n = 1, \) and \( b = 2. \)

Thus, depending on the correlation structure and the value of the parameter, different designs are optimal. Furthermore, it can be noted that none of the optimal designs in Figure 4.2 have the control treatment situated at an end, in order for the control to have the maximum number of neighbours for \( \xi_{A,GLSE} \).
Figure 4.2: Optimal designs with treatment one as a control for \( t = 6, n = 1, \) and \( b = 2. \)

### 4.1 Neighbourhoods of Spatial Correlations

In this section, we will define the neighbourhood of a correlation structure. Let \( R_0 = \sigma^2 V_0. \) Two neighbourhoods of \( R_0 \) were introduced in Wiens and Zhou (2008) and also explored in Ou and Zhou (2009). Since there are no blocks in these papers, we need to modify the two neighbourhoods for block designs. With blocks, \( R_0 \) is a
block diagonal matrix with diagonal elements $R_i = \sigma^2 V_i$, $i = 1, \ldots, b$, i.e.,

$$R_0 = \begin{pmatrix} R_1 & 0 \\ \vdots & \ddots \\ 0 & R_b \end{pmatrix}.$$  

We can apply neighbourhoods to each $R_i$ to define the neighbourhoods of $R_0$ here.

Matrix norms are used to define the neighbourhoods and they are an extension of vector norms. Given a real matrix $A$ with dimensions $N \times N$, the matrix norms $\|A\|_p$ for $p = 1$ or $2$ are defined as follows:

$$\|A\|_1 = \max_{1 \leq j \leq N} \sum_{i=1}^{N} |a_{ij}|,$$

which is the maximum absolute column sum of the matrix $A$, and

$$\|A\|_2 = \max\{|\omega| \mid \omega \text{ is an eigenvalue of } A\}.$$

Two neighbourhoods of $R_i$, $i = 1, \ldots, b$ are therefore defined as follows,

1. $R_{p,\alpha}^i = \{B \mid \|B - R_i\|_p \leq \alpha, B^\top = B \geq 0\}$, $p = 1$ or $2$,

2. $R_{K,\alpha}^i = \{B \mid 0 \leq B \leq R_i + \alpha K, B^\top = B\}$,

where $\alpha \geq 0$ controls the size of the neighbourhood and $K$ is a symmetric non-negative matrix. The matrix ordering is by non-negative definiteness, i.e. $A \geq B$ implies $A - B \geq 0$ is a non-negative matrix. $K$ is either $R_i$ or $I_t$. It is obvious that $R_i \in R_{p,\alpha}^i$ and $R_i \in R_{K,\alpha}^i$.

The neighbourhoods of $R_0$ are therefore,

1. $R_{p,\alpha} = \left\{ \begin{pmatrix} R_i \\ \vdots \\ R_i \end{pmatrix} \mid R = \begin{pmatrix} A_1 & 0 \\ \vdots & \ddots \\ 0 & A_b \end{pmatrix}, A_i \in R_{p,\alpha}^i, i = 1, \ldots, b \right\}$,
2. \( R_{K,\alpha} = \left\{ R \mid R = \begin{pmatrix} A_1 & 0 \\ \vdots & \ddots \\ 0 & A_b \end{pmatrix}, A_i \in R_{K,\alpha}^i, i = 1, \ldots, b \right\} \)

### 4.2 Robust Design Criteria

In practice, we never know the exact spatial correlation structure or the parameters so we need to find some designs which are robust to misspecifications. Minimax designs are solutions to

\[
\min_{X} \max_{R \in R_{\phi}} \mathcal{L}(\text{Cov}(\hat{\theta})),
\]

where \( R_{\phi} \) is a specified neighbourhood, \( C \) is a constant matrix, and \( \mathcal{L} \) is a loss function of the covariance such as determinant or trace. When comparing treatments, \( C \) is a contrast matrix and \( R_{\phi} \) can either be \( R_{p,\alpha} \) or \( R_{K,\alpha} \).

There are two methods to estimate \( \theta \) as discussed in Section 1.2. Generalized least squares estimation is used when the covariance matrix is known resulting in \( \hat{\theta}_{GLSE} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \) and \( \text{Cov}(\hat{\theta}_{GLSE}) = \sigma^2 (X^T V^{-1} X)^{-1} \). Least squares estimation is used when the covariance matrix is unknown giving the estimator of \( \hat{\theta}_{LSE} = (X^T X)^{-1} X^T y \) and \( \text{Cov}(\hat{\theta}_{LSE}) = \sigma^2 (X^T X)^{-1} X^T VX(X^T X)^{-1} \). When the covariance matrix \( R \) is unknown but belongs in a neighbourhood of \( R_0 \), the generalized least squares estimation introduced in Martin (1986) can be utilized. This gives:

\[
\hat{\theta}_{R_0} = (X^T R_0^{-1} X)^{-1} X^T R_0^{-1} y,
\]

\[
\text{Cov}(\hat{\theta}_{R_0}) = (X^T R_0^{-1} X)^{-1} X^T R_0^{-1} R R_0^{-1} X (X^T R_0^{-1} X)^{-1}, \text{ and}
\]

\[
\text{Cov}(C\hat{\theta}_{R_0}) = C (X^T R_0^{-1} X)^{-1} X^T R_0^{-1} R R_0^{-1} X (X^T R_0^{-1} X)^{-1} C^T,
\]

where \( C \) is defined in (1.4). We will focus on \( \hat{\theta}_{R_0} \) to find minimax designs.
4.3 Results

**Theorem 4.1.** Let $C(R_0, R, X) = \text{Cov}(\hat{C}\theta_{R_0})$ and $U_{K,\alpha} = R_0 + \alpha \begin{pmatrix} K \\ \vdots \\ K \end{pmatrix}$.

Suppose loss function $L$ is monotonic according to the ordering of positive definiteness.

For neighbourhood $R_{K,\alpha}$, we have

$$\max_{R \in R_{K,\alpha}} L(\text{Cov}(C\hat{\theta}_{R_0})) = \max_{R \in R_{K,\alpha}} L(\text{Cov}(R_0, R, X)) = L(C(R_0, U_{K,\alpha}, X)).$$

**Proof.** For neighbourhood $R_{K,\alpha}$, each $A_i \leq R_i + \alpha K$, $i = 1, \ldots, b$. Thus,

$$R \leq \begin{pmatrix} R_1 + \alpha K \\ \vdots \\ \vdots \\ R_b + \alpha K \end{pmatrix} = U_{K,\alpha}.$$

Using properties of positive definiteness, we get

$$X^\top R_0^{-1}RR_0^{-1}X \leq X^\top R_0^{-1}U_{K,\alpha}R_0^{-1}X,$$

and

$$\text{Cov}(C\hat{\theta}_{R_0}) = C(X^\top R_0^{-1}X)^{-1}X^\top R_0^{-1}RR_0^{-1}X(X^\top R_0^{-1}X)^{-1}C^\top$$

$$\leq C(X^\top R_0^{-1}X)^{-1}X^\top R_0^{-1}U_{K,\alpha}R_0^{-1}X(X^\top R_0^{-1}X)^{-1}C^\top$$

$$= C(R_0, U_{K,\alpha}, X).$$

Therefore, $\max_{R \in R_{K,\alpha}} L(\text{Cov}(C\hat{\theta}_{R_0})) = L(C(R_0, U_{K,\alpha}, X)).$ \qed
For neighbourhood $R_{p,\alpha}$, each $A_i \leq R_i + \alpha I_t$ from a result in Wiens and Zhou (2008). Thus, the result in Theorem 4.1 can be applied to find the maximum loss for neighbourhood $R_{p,\alpha}$ with $K = I_t$. When $K = I_t$,

\[
C(R_0, U_{I_t,\alpha}, X) = C(X^T R_0^{-1} X)^{-1} X^T R_0^{-1} (R_0 + \alpha I_t) R_0^{-1} X (X^T R_0^{-1} X)^{-1} C^T
\]

\[
= C(X^T R_0^{-1} X)^{-1} C^T + \alpha C(X^T R_0^{-1} X)^{-1} X^T R_0^{-2} X (X^T R_0^{-1} X)^{-1} C^T.
\]

When $R_1 = \ldots = R_b$ with $K = R_1$, we have

\[
C(R_0, U_{R_1,\alpha}, X) = C(X^T R_0^{-1} X)^{-1} X^T R_0^{-1} \left[ R_0 + \alpha \begin{pmatrix} R_1 & & \\ & \ddots & \\ & & R_b \end{pmatrix} \right] R_0^{-1} X (X^T R_0^{-1} X)^{-1} C^T
\]

\[
= C(X^T R_0^{-1} X)^{-1} X^T R_0^{-1} (R_0 + \alpha R_0) R_0^{-1} X (X^T R_0^{-1} X)^{-1} C^T
\]

\[
= (1 + \alpha) C(X^T R_0^{-1} X)^{-1} C^T.
\]

Therefore, $\max_{R \in R_{K,R_1,\alpha}} \mathcal{L}(\text{Cov}(C\hat{\theta}_{R_0})) = (1+\alpha) \mathcal{L}(\text{Cov}(C\hat{\theta}_{\text{GLSE}}))$ and so for the neighbourhood $R_{K,\alpha}$ with $R_1 = \ldots = R_b$ and $K = R_1$, optimal designs under the GLSE are also robust designs.

In the next chapter, we present a simulated annealing algorithm which enables us to find minimax block designs. We will use $\xi_{\nu,K}$-minimax to denote the $\nu$-optimal minimax design where $\nu$ is $D$ or $A$ and $K$ is either $R_1$ or $I_t$. We also present several interesting examples using various correlation structures and parameter values.
Chapter 5

Annealing Algorithm to Compute Optimal Designs

In this chapter, we compute the minimax designs defined in Chapter 4. For a small number of treatments, complete searches may be feasible, but as the number of treatments increases, complete searches are not possible. Thus, a simulated annealing algorithm is discussed and used to find minimax designs. The algorithm is introduced in Section 5.1 and Section 5.2 contains several examples of optimal and minimax designs.

5.1 Simulated Annealing Algorithm

There are essentially three stages in the algorithm: construction of an initial design and search parameters, an annealing routine, and a steepest descent procedure (Elliott, Eccleston, and Martin, 1999). The first stage is where an initial design is chosen and a loss function $f$ is defined. We want to minimize $f$ to obtain optimal and robust designs. A new design is then created by interchanging two treatment allocations and the change in the loss function $\Delta f = f(\text{new design}) - f(\text{old design})$ is calculated. The
new design is then accepted with probability,

\[ p = \begin{cases} 
1, & \triangle f \leq 0, \\
\exp(-\triangle f / T), & \triangle f > 0,
\end{cases} \quad (5.1) \]

where \( T \) is the temperature. This enables the algorithm to be able to move away from the local minima (Elliott, Eccleston, and Martin, 1999). The temperature is set high in the beginning and is slowly decreased after a specific number of interchanges (for example, 1,000) at each temperature. When the temperature is high, uphill moves are more likely to be accepted, which helps to ensure that the algorithm does not automatically converge to the first local minimum reached. Geometric cooling is used to cool the temperature, which reduces the temperature by a factor at a given time, so that \( T_{i+1} = \delta T_i \) where \( 0 < \delta < 1 \). As the temperature is lowered, fewer uphill moves are tolerated and so the search becomes concentrated near the current local minimum. With a finite number of choices (\((b - 1)t!\) designs), the result is guaranteed to converge to a global minimum if the temperature decreases sufficiently slowly (Hàjek, 1988). However, in practice, this is not necessary the case. This continues until there is a very low probability of choosing an inferior design and the temperature is below the stopping temperature. The steepest descent routine is then carried out once the stopping temperature is reached. All pairwise interchanges of the existing best design are performed and the loss function is evaluated for each. If the best design from the annealing algorithm still has the lowest loss value, it is the locally optimum design. If there is a new best design, then the steepest descent routine is performed again until the best design entering the routine and the best design at the end of the routine are the same. Smaller loss function values are accepted until no possible improvements can be made under a tolerance level. Only larger designs require this step to ensure a local optimum is reached. This algorithm does not guarantee to find
the global optimum, but a near optimal design can potentially be found. The steps of the algorithm are outlined below:

1. Select an initial design, $X$; obtain an initial temperature, $T_0$, and set a stopping temperature $T_{stop}$; let $X^* = X$, where $X^*$ is the best current design.

2. Initialize iteration number $j = 1$.

3. Generate new design, $X'$, by interchanging allocations of two treatments and calculate $\Delta f = f(X') - f(X)$.

4. Let $X = X'$ with probability given in (5.1).

5. If $X$ is better than $X^*$, then set $X^* = X$.

6. Increment $j$ by 1 and return to step 3 until the specified iterations at the temperature have elapsed.

7. Cool the temperature using $T_{i+1} = \delta T_i$; if the temperature is at the stopping temperature, i.e. $T_{i+1} < T_{stop}$, stop, otherwise return to step 2.

8. Run the steepest descent routine on $X^*$.

### 5.2 Examples of Robust Designs

In this section, we present various examples of results from the simulated annealing algorithm. The simulated annealing algorithm can be compared against a complete search algorithm for a small number of treatments to ensure consistency. The plots of the loss function versus number of iterations are presented to show that the loss function decreases gradually. Example 5.1 presents the $\xi_{D,R_1}$-minimax design which can be compared against a complete search algorithm. Also, Example 5.2 gives the
ξ_{D,I}-minimax design for a larger number of treatments as it is very expensive to use a complete search. Example 5.3 and Example 5.4 show the ξ_{A}-minimax designs for a MA(1) model without and with a control treatment respectively. The DG correlation structure is used in Example 5.5 to illustrate the impact of the values of the correlation on the design while Example 5.6 displays the ξ_{D,R_1}-minimax designs for various number of treatments under the NN(1).

**Example 5.1.** Consider the ξ_{D,R_1}-minimax design for \( t = 7, \ n = 1, \) and \( b = 2 \) under the NN(1) with correlation parameter \( \rho = 0.15. \) In the computation, the size of the neighbourhood is \( \alpha = 0.25 \) and the initial temperature is \( T = 0.0001 \) which is reduced by a cooling factor of 0.9 every 1,000 iterations. A plot of the loss function versus iteration is shown in Figure 5.1. From Figure 5.1, it is obvious that the loss function converges after approximately 3,500 runs to a value of 0.03005914. The minimax design resulting from the simulated annealing algorithm is shown in Figure 5.2. This design is the global optimum as verified by a complete search, which may be run for seven treatments. Minimax designs are not unique in these cases.

**Example 5.2.** Consider a higher number of treatments with \( t = 12, \ n = 2, \) and \( b = 2. \) The convergence plot of the loss function of the ξ_{D,L}-minimax design is shown in Figure 5.3. The correlation structure is the DG with a high correlation of \( \lambda = 0.99, \alpha = 0.25, \) and an initial temperature of 0.00005. The temperature is reduced by a cooling factor of 0.9 every 1,000 iterations. The loss function converges to a value of 1.881167e-09 after approximately 24,000 runs to produce the design for block two shown in Figure 5.4. The choice of the temperature is very important to ensure convergence to a local optimum. It is interesting to note that the two blocks have an equivalent arrangement.

We can run the simulating annealing algorithm for various combinations of \( m \times n \)
Figure 5.1: Loss function $\mathcal{L}(C(R_0, U_{R_1, \alpha}, X))$ versus number of iterations.

Figure 5.2: Minimax design with $t = 7$, $n = 1$, and $b = 2$ under the NN(1).

Example 5.3. Consider the $\xi_A$-minimax designs for the MA(1) correlation structure with $\gamma = 0.2$ and $\alpha = 0.4$. For $t = 9$, $n = 3$, and $b = 2$, the $\xi_{A,I}$-minimax design is shown in Figure 5.5(a) with a loss function of 5.953724. The $\xi_{A,R_1}$-minimax design has a loss value of 5.250503 and is shown in Figure 5.5(b). This example is interesting because the designs are not equivalent for different $K$. 
Example 5.4. Consider the design with the same parameters in Example 5.3 but with treatment one as a control. The resulting minimax designs are shown in Figure 5.6. The loss functions have values of 9.040676 and 6.288434 respectively. Note that the control treatment is in the very centre for both blocks regardless of $K$ as this position maximizes the number of neighbours for the control. Furthermore, the $\xi_{A,L}$-minimax and $\xi_{A,R_1}$-minimax designs are different.

Example 5.5. Consider the DG correlation structure for $t = 12$, $n = 2$, $b = 2$ with $\lambda$ taking various values and $\alpha = 0.3$. The $\xi_{D,L}$-minimax designs are shown in Figure
5.7. The values of $\lambda$ are (a) $\lambda = 0.01$, (b) $\lambda = 0.3$, (c) $\lambda = 0.7$, and (d) $\lambda = 0.95$.

The $\xi_{D,R_1}$-minimax designs are shown in Figure 5.8 for the same values of $\lambda$. This example is interesting because for low $\lambda$ (i.e. $\lambda = 0.01$), the two loss functions produce equivalent designs. As $\lambda$ increases (e.g. $\lambda = 0.3$), the minimax designs are different and as $\lambda$ approaches 1, the $\xi_{D,I_1}$-minimax design is the design where block one has the same arrangement as block two. Thus, as $\lambda$ increases from 0 to 1, the number of pairs of neighbours across the two blocks increases from the minimum number of neighbours when $\lambda$ is low (e.g. 0.01) to the maximum number of neighbours for the $\xi_{D,I_1}$-minimax design.

**Example 5.6.** Consider the $\xi_{D,R_1}$-minimax designs for various values of $t$. The NN(1) correlation structure is used with the design parameters of $n = 2$, $b = 2$, $\rho = 0.2$, and $\alpha = 0.3$. The minimax designs are shown in Figure 5.9. It is interesting to note that the neighbours in block one are not neighbours in block two.

The simulated annealing algorithm works well to find optimal and minimax designs for moderate $t$ and $b$. Choosing an appropriate initial temperature and appropriate
Figure 5.7: $\xi_{D,I_t}$-minimax designs with $t = 12$, $n = 2$, and $b = 2$ under the DG where $\ell_{D,I_t} = \min_X \det[C(R_0, U_{I_t}, \alpha, X)]$: (a) $\lambda = 0.01$, (b) $\lambda = 0.3$, (c) $\lambda = 0.7$, and (d) $\lambda = 0.95$.

Figure 5.8: $\xi_{D,R_1}$-minimax designs with $t = 12$, $n = 2$, and $b = 2$ under the DG where $\ell_{D,R_1} = \min_X \det[C(R_0, U_{R_1}, \alpha, X)]$: (a) $\lambda = 0.01$, (b) $\lambda = 0.3$, (c) $\lambda = 0.7$, and (d) $\lambda = 0.95$. 
(a) \( \ell_{D,R_1} = 0.0052115 \)  
(b) \( \ell_{D,R_1} = 0.001761260 \)  
(c) \( \ell_{D,R_1} = 0.0005937593 \)  
(d) \( \ell_{D,R_1} = 0.0002005438 \)  
(e) \( \ell_{D,R_1} = 6.768441e-05 \)

Figure 5.9: \( \xi_{D,R_1} \)-minimax designs for various \( t \) under the NN(1) where \( \ell_{D,R_i} = \min_X \det[C(R_0, U_{R_1}, X)] \): (a) \( t = 10 \), (b) \( t = 12 \), (c) \( t = 14 \), (d) \( t = 16 \), and (e) \( t = 18 \).

design parameters is very important to ensure convergence to a local optimum. Several theoretical results for optimal designs using the GLSE and the LSE are presented in Chapter 6.
Chapter 6

Theoretical Results

Several theoretical results are derived in this chapter for block designs with $b = 2$. Section 6.1 contains the results for $\xi_{D,LSE}$ designs. Section 6.2 contains the results for $\xi_{A,LSE}$ designs and Section 6.3 for $\xi_{D,GLSE}$ using the DG and DE correlation structures.

6.1 Results for $\xi_{D,LSE}$ Designs

In Chapter 3, we have already seen that labelling the $t$ treatments and block order do not affect $\xi_{D,LSE}$. Thus, we set the $t$ treatments in block one to be a one-dimensional array of order $1, \ldots, t$. Define $S_1$ to be the set of all neighbours in block 1, where two plots are considered neighbours if they are next to each other: $S_1 = \{(1,2), (2,1), \ldots, (t-1,t), (t,t-1)\}$. Similarly, $S_2$ is the set of neighbours in block 2. Suppose the order of the plots in block 2 are $a_1, \ldots, a_t$ where $a_i$ is the treatment level, $1 \leq i \leq t$, then $S_2 = \{(a_1, a_2), (a_2, a_1), \ldots, (a_{t-1}, a_t), (a_t, a_{t-1})\}$. Thus, each pair of neighbours is represented twice in $S_1$ and $S_2$ (i.e. if $(i,j)$ is a neighbour then so is $(j,i)$).

We want to show that the cardinality of $\{S_1 \cap S_2\}$ (i.e. $|\{S_1 \cap S_2\}|$) is maximized
for $\xi_{D, LSE}$ for a $m \times n$ array in two blocks. Notice that

$$\text{Cov}(\hat{C}_{LSE}) = \sigma^2 C (X^T X)^{-1} X^T VX (X^T X)^{-1} C^T,$$

where $\sigma^2$ is a constant,

$$(X^T X)^{-1} = \begin{pmatrix} \frac{1}{b} I_t & 0_{t \times 1} \\ 0_{1 \times t} & \frac{1}{2b} \end{pmatrix}, \quad \text{and} \quad C = \begin{pmatrix} I_t & 0_{t \times 1} \end{pmatrix}. \quad (6.1)$$

Let

$$V = \begin{pmatrix} V_1 & 0 \\ 0 & V_1 \end{pmatrix}, \quad (6.2)$$

where $V_1$ is the correlation matrix, and

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad X_1 = \begin{pmatrix} I_t & 1_t \end{pmatrix}, \quad \text{and} \quad X_2 = \begin{pmatrix} P_t & -1_t \end{pmatrix}. \quad (6.3)$$

Then

$$X^T VX = X_1^T V_1 X_1 + X_2^T V_1 X_2. \quad (6.4)$$

From (6.2) and (6.3), we have

$$X_1^T V_1 X_1 = \begin{pmatrix} (V_1)_{t \times t} & (V_1 1_t)_{t \times 1} \\ (1_t^T V_1)_{1 \times t} & (1_t^T V_1 1_t)_{1 \times 1} \end{pmatrix}_{(t+1) \times (t+1)}$$

and

$$X_2^T V_1 X_2 = \begin{pmatrix} (P_t^T V_1 P_t)_{t \times t} & -(P_t^T V_1 1_t)_{t \times 1} \\ -(1_t^T V_1 P_t)_{1 \times t} & (1_t^T V_1 1_t)_{1 \times 1} \end{pmatrix}_{(t+1) \times (t+1)}.$$
Thus, from (6.4),

\[
X^\top V X = \begin{pmatrix}
(V_1 + P_t^\top V_1 P_t)_{t \times t} & [(I_t - P_t^\top) V_1 1_t]_{t \times 1} \\
[I_t^\top V_1 (I_t - P_t)]_{1 \times t} & (21_t^\top V_1 1_t)_{1 \times 1}
\end{pmatrix}_{(t+1) \times (t+1)}
\]

and

\[
(X^\top X)^{-1} X^\top V X (X^\top X)^{-1} =
\begin{pmatrix}
\frac{1}{b^2} (V_1 + P_t^\top V_1 P_t)_{t \times t} & \frac{1}{2b^2}[ (I_t - P_t^\top) V_1 1_t]_{t \times 1} \\
\frac{1}{2b^2} [I_t^\top V_1 (I_t - P_t)]_{1 \times t} & \frac{1}{4b^2} (21_t^\top V_1 1_t)_{1 \times 1}
\end{pmatrix}_{(t+1) \times (t+1)}
\]

Therefore,

\[
C(X^\top X)^{-1} X^\top V X (X^\top X)^{-1} C^\top = \frac{1}{b^2} (V_1 + P_t^\top V_1 P_t)_{t \times t} 
\]

and

\[
\min_X \det[\text{Cov}(C_\hat{\theta}_{LSE})] = \min_X \sigma^{2t} \det[\text{C}(X^\top X)^{-1} X^\top V X (X^\top X)^{-1} C^\top]
\]

\[
= \min_X \sigma^{2t} \det\left\{ \frac{1}{b^2} (V_1 + P_t^\top V_1 P_t) \right\}
\]

\[
\leq \max \det (V_1 + P_t^\top V_1 P_t).
\]

**Theorem 6.1.** For a \(m \times n\) array in two blocks, \(\xi_{D,LSE}\) minimizing \(\det[\text{Cov}(C_\hat{\theta}_{LSE})]\) (where \(C_\hat{\theta}_{LSE}\) represents the treatment effects) has the same arrangement in the two blocks, i.e. \(|\{S_1 \cap S_2\}|\) is maximized.
Proof. Let $A_0 = V_1$ and $A_1 = P_t^\top V_1 P_t$. Then,

$$
|A_1| = |P_t^\top V_1 P_t|
= |P_t^\top||V_1||P_t|
= |V_1|
= |A_0|.
$$

For $0 \leq \eta \leq 1$, define

$$
A_\eta = (1 - \eta)A_0 + \eta A_1. \tag{6.6}
$$

Using Minkowski’s inequality for determinants (Horn and Johnson, 1985), as $A_0$ and $A_1$ are positive definite,

$$
|A_\eta|^\frac{1}{\eta} = |(1 - \eta)A_0 + \eta A_1|^\frac{1}{\eta} \quad \text{for } 0 \leq \eta \leq 1
\geq |(1 - \eta)A_0|^\frac{1}{\eta} + |\eta A_1|^\frac{1}{\eta}
= (1 - \eta)|A_0|^\frac{1}{\eta} + \eta|A_1|^\frac{1}{\eta}
= |A_0|^\frac{1}{\eta}, \text{ since } |A_0| = |A_1|, \tag{6.7}
$$

where the equivalence holds if and only if $A_1 = cA_0$ for some $c \geq 0$ (Horn and Johnson, 1985). Since we have $A_0 = V_1$ and $A_1 = P_t^\top V_1 P_t$, then $A_1 \neq cA_0$ as long as $P_t \neq I_t$. Therefore, $|A_\eta|^\frac{1}{\eta} > |A_0|^\frac{1}{\eta}$ for $0 < \eta < 1$ and $P_t \neq I_t$. 


Now,

\[ |A_{\eta}| > |A_0| \quad \text{for } 0 < \eta < 1 \]

which implies \[|2A_{0.5}| > |2A_0|,\]
i.e. \[|V_1 + P_t^\top V_1 P_t| > |2V_1|,\]

whereas equality holds when \(P_t = I_t\). Therefore, the determinant is minimized when the two blocks have the same arrangement and so the cardinality of \(\{S_1 \cap S_2\}\) (i.e. \(|\{S_1 \cap S_2\}|\)) is maximized.

\[\square\]

### 6.2 Results for \(\xi_{A,LSE}\) Designs

We can also show that the trace of \(\text{Cov}(\hat{\theta}_{LSE})\) is constant for all designs.

**Lemma 6.1.** \(A\)-optimal criterion do not differentiate designs when the LSE is used.

**Proof.** \(\xi_{A,LSE}\) minimize the trace of \(\text{Cov}(\hat{\theta}_{LSE})\).

From (6.5), \(\text{Cov}(\hat{\theta}_{LSE}) = \frac{1}{b^2}(V_1 + P_t^\top V_1 P_t)\), and so

\[
\text{trace}[\text{Cov}(\hat{\theta}_{LSE})] = \text{trace}[\frac{1}{b^2}(V_1 + P_t^\top V_1 P_t)] = \frac{1}{b^2} [\text{trace}(V_1) + \text{trace}(P_t^\top V_1 P_t)] = \frac{2t}{b^2},
\]

which is a constant for all designs. Thus, we cannot differentiate designs for any correlation structure.

\[\square\]
6.3 Results for $\xi_{D,GLSE}$ Designs

We will examine $\xi_{D,GLSE}$ under the DG and DE correlation structures for $n = 1$ and $b = 2$. The covariance is represented as $\text{Cov}(C\hat{\theta}_{GLSE}) = \sigma^2 C(X^\top V^{-1}X)^{-1}C^\top$, where $\sigma^2$ is a constant and $C$ is the constant matrix in (6.1).

Now, for the DG and DE correlation structures,

$$V = \begin{pmatrix} V_1 & 0 \\ 0 & V_1 \end{pmatrix},$$

where $V_1 = V_1(\lambda, t)$ from (2.6).

The inverse of the correlation matrix $V_1$ has a nice form as

$$V_1^{-1} = \frac{1}{1 - \lambda^2} \begin{pmatrix} 1 & -\lambda \\ -\lambda & \lambda^2 + 1 & -\lambda \\ & \ddots & \ddots \\ & & -\lambda & \lambda^2 + 1 & -\lambda \\ & & & -\lambda & 1 \end{pmatrix}.$$

Since

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

we get

$$X^\top V^{-1}X = X_1^\top V_1^{-1}X_1 + X_2^\top V_1^{-1}X_2,$$

where $X_1$ and $X_2$ are defined as in (6.3).
Therefore,

\[
X_1^\top V_1^{-1}X_1 = \begin{pmatrix}
(V_1^{-1})_{t \times t} & (V_1^{-1}1_t)_{t \times 1} \\
(1_t^\top V_1^{-1})_{1 \times t} & (1_t^\top V_1^{-1}1_t)_{1 \times 1}
\end{pmatrix}_{(t+1) \times (t+1)}
\]

and

\[
X_2^\top V_1^{-1}X_2 = \begin{pmatrix}
(P_t^\top V_1^{-1}P_t)_{t \times t} & -(P_t^\top V_1^{-1}1_t)_{t \times 1} \\
-(1_t^\top V_1^{-1}P_t)_{1 \times t} & (1_t^\top V_1^{-1}1_t)_{1 \times 1}
\end{pmatrix}_{(t+1) \times (t+1)}.
\]

So,

\[
X^\top V^{-1}X = \begin{pmatrix}
(V_1^{-1} + P_t^\top V_1^{-1}P_t)_{t \times t} & (V_1^{-1}1_t - P_t^\top V_1^{-1}1_t)_{t \times 1} \\
(1_t^\top V_1^{-1} - 1_t^\top V_1^{-1}P_t)_{1 \times t} & (21_t^\top V_1^{-1}1_t)_{1 \times 1}
\end{pmatrix}_{(t+1) \times (t+1)}
\]

and since

\[
21_t^\top V_1^{-1}1_t = \frac{2}{1 - \lambda^2} [t - 2(t - 1)\lambda + (t - 2)\lambda^2].
\]

\[
C(X^\top V^{-1}X)^{-1}C^\top = [(V_1^{-1} + P_t^\top V_1^{-1}P_t) - \frac{1 - \lambda^2}{2[t - 2(t - 1)\lambda + (t - 2)\lambda^2]}(V_1^{-1}1_t - P_t^\top V_1^{-1}1_t)(1_t^\top V_1^{-1} - 1_t^\top V_1^{-1}P_t)]^{-1}.
\]

**Theorem 6.2.** For \(m \times 1\) plots in \(b = 2\) blocks with \(t\) treatments under the DG and DE, \(\xi_{D,\text{GLSE}}\) is not the design in which plots in block two have the same arrangement as in block one.
Proof. Suppose \( A_0 = V^{-1}_1 \) and \( A_1 = P_t^\top V^{-1}_1 P_t \). Then,

\[
|A_1| = |P_t^\top V^{-1}_1 P_t| = |P_t^\top||V^{-1}_1||P_t| = |V^{-1}_1| = |A_0|.
\]

Let us take \( A_\eta = (1 - \eta)A_0 + \eta A_1 \) as in (6.6). Using Minkowski’s inequality for determinants, we know \( |A_\eta|^{\frac{1}{\eta}} \geq |A_0|^{\frac{1}{\eta}} \) for \( 0 \leq \eta \leq 1 \) from (6.7). Similarly to Theorem 6.1, we know the equivalence holds if and only if \( A_1 = cA_0 \) for some \( c \geq 0 \) (Horn and Johnson, 1985). Since we have \( A_0 = V^{-1}_1 \) and \( A_1 = P_t^\top V^{-1}_1 P_t \), then \( A_1 \neq cA_0 \) as long as \( P_t \neq I_t \). Thus, we can take \( |A_\eta|^{\frac{1}{\eta}} > |A_0|^{\frac{1}{\eta}} \) for all \( 0 < \eta < 1 \) and \( P_t \neq I_t \).

When the second block has an identical arrangement to block one then \( P_t = I_t \),

\[
C(X^\top V^{-1}X)^{-1}C^\top = (2V^{-1}_1)^{-1} = (2A_0)^{-1}.
\]

If the two end plots on the second block remain as treatments 1 and \( t \) and a permutation occurs of the inner \( t - 2 \) treatments, i.e. the second block has a permutation matrix as in Figure 6.1, but \( P_t \neq I_t \), then \( P_t^\top V^{-1}_1 1_t = V^{-1}_1 1_t \) and \( 1_t^\top V^{-1}_1 P_t = 1_t^\top V^{-1}_1 \).

\[
\begin{pmatrix}
1 \\
P_{t-2} \\
1
\end{pmatrix}
\]

Figure 6.1: Permutation of inner treatments of block two.

Thus,

\[
C(X^\top V^{-1}X)^{-1}C^\top = (V^{-1}_1 + P_t^\top V^{-1}_1 P_t)^{-1} = (A_0 + A_1)^{-1} = (2A_{0.5})^{-1}.
\]
Consequently, using Minkowski’s inequality for determinants,

\[
|A_{0.5}| > |A_0| \\
|2A_{0.5}| > |2A_0| \\
|(2A_{0.5})^{-1}| < |(2A_0)^{-1}| \\
|(A_0 + A_1)^{-1}| < |(2A_0)^{-1}| \\
|(V_1^{-1} + P_t^T V_1^{-1} P_t)^{-1}| < |(2V_1^{-1})^{-1}|.
\]

and \(\xi_{D,GLSE}\) under the DG and DE correlation structures is not the design with identical arrangements for both blocks.

\[\Box\]

In summary, \(\xi_{D,LSE}\) are the designs with identical treatment arrangements for both blocks for any correlation structure. Also, we cannot differentiate designs for \(A\)-optimal criterion when the LSE is used for any correlation structure. \(\xi_{D,GLSE}\) using the DG and DE structures are not the designs where the blocks have identical arrangements. These results help to provide some guidelines when planning an experiment.
Experimental design is important in agricultural field and industrial experiments. Field experiments in blocks are investigated in this thesis and two estimators, the LSE and the GLSE, are discussed. The optimality criteria and various correlation structures are presented and some basic properties of block designs are also derived. However, as the correlation structure is not always known or assumed with some uncertainty, two neighbourhoods are considered and robust design criteria are studied. A generalized least squares estimator which may be used when the covariance matrix is unknown but belongs in a neighbourhood is also provided. The optimal and/or minimax designs may not be able to be found analytically and thus, a simulated annealing algorithm is introduced to compute optimal/robust designs. This algorithm is effective in searching for the minimax design given certain design parameters. Several theoretical results for $A$- and $D$-optimal designs are also derived for the LSE and the GLSE.

In practical applications, when the correlation structure is not known, the LSE is often used. For any correlation structure, $D$-optimal designs using the LSE produce designs with identical arrangements for both blocks while we cannot differentiate
between $A$-optimal designs using the LSE. When the correlation structure is known, the GLSE should be used. $D$-optimal designs under the GLSE for the DG and DE correlation structures are not the designs with identical arrangements for both blocks. Since the correlation structure is known, it is best to separate treatments which are neighbours in block one so that they are not neighbours in block two. This is similar to results by Gill and Shukla (1985) and Kunert (1987) who found that neighbour balanced block designs are optimal for the GLSE using the autoregressive correlation structure. If the exact correlation structure is not known or can be assumed with some uncertainty, the generalized least squares estimation introduced by Martin (1986) and minimax designs studied in this thesis can be applied.

Therefore, the design criteria depend on the experiment and the different estimators may even produce different results. Minimax designs provide balance between certainty in the correlation structure (GLSE) and complete uncertainty (LSE) and are robust against misspecifications in the correlation structure.

Additionally, further studies can be done to derive robust designs for various field experiments. Theoretical results can be studied for $b > 2$. The study can also be extended for two-way effects models and for fractional factorial designs with several factors.
Appendix A

R Programs

A.1 Correlation Structure Function

#Creates the Correlation Structure
#tr - number of treatments, tr > 2
#n - number of columns and thus m = t/n
#struc: NN (Nearest Neighbour), MA (Moving Average),
#DG (Doubly Geometric), DE (Discrete Exponential)
#val - value of the correlation parameter (rho, gamma, lambda)

corr.struc = function(tr,n,struc=c("NN","MA","DG","DE"),val)
{
  if(struc=="NN") #0 <= rho <= 1/4
  {
    rho = val
    y = rep(c(rep(rho,n-1),0),tr/n)[1:(tr-1)] #super diagonal
    #(for n=1, this will be row of zeros)
    v0 = odiag(y,1) #creates matrix with y as super diagonal
  }
  for(i in 1:(tr-n)) #diagonal n above

#need nth diagonal to be all rho - length will be tr-n
{
  v0[i,(i+n)] = rho
}

v0 = v0 + t(v0) #create symmetric matrix
diag(v0) = 1 #main diagonal

if( struc=="MA") #0 <= gamma <= 1/4
{
  gamma = val
  rho = 2*gamma/(1 + 4*gamma^2)

  if(n==1) #can be done for all tr; tr >= 3
  {
    y = rep(rho,tr-1)
    v0 = odiag(y,1) #super diagonal
    for(i in 1:(tr-2)) #diagonal 2 above
    {
      v0[i,(i+2)] = 1/2*gamma*rho
    }
  }

  if(n==2) #can be done for all even tr; tr >= 4
  {
    y = rep(c(rho,gamma*rho),tr/2)[1:(tr-1)]
    v0 = odiag(y,1) #super diagonal
    for(i in 1:(tr-2)) #diagonal 2 above
    {
      v0[i,(i+2)] = rho
    }
    for(i in 1:(tr-3)) #diagonal 3 above
    {
      v0[i,(i+2)] = 1/2*gamma*rho
    }
  }
}
if(i%%2 != 0) #only odd rows
    v0[i, (i+3)] = gamma*rho
}
if(tr > 4)
{
    for(i in 1:(tr-4)) #diagonal 4 above
    {
        v0[i, (i+4)] = (1/2)*gamma*rho
    }
}
}
if(n==3) #can be done for all tr/3; tr >= 6
{
    y = rep(c(rho, rho, 0), tr/3)[1:(tr-1)]
    v0 = odiag(y, 1) #super diagonal
    for(i in 1:(tr-2)) #diagonal 2 above
    {
        if((i-1)%%3 == 0) #every third row starting with first row
            v0[i, (i+2)] = (1/2)*gamma*rho
        else
            v0[i, (i+2)] = gamma*rho
    }
    for(i in 1:(tr-3)) #diagonal 3 above
    {
        v0[i, (i+3)] = rho
    }
    for(i in 1:(tr-4)) #diagonal 4 above
    {
        if(i%%3 != 0) #excludes every third row
            v0[i, (i+4)] = gamma*rho
    }
if (tr > 6) #diagonal 6 above
{
    for (i in 1:(tr-6)) #diagonal 4 above
    {
        v0[i, (i+6)] = (1/2)*gamma*rho
    }
}

if (n >= 4) #can be done for all tr%%n == 0; tr >= 2n
{
    y = rep(c(rep(rho, n-1), 0), tr/n)[1:(tr-1)]
    v0 = odiag(y, 1) #super diagonal
    #only using first tr-2 elements
    diag2 = rep(c(rep((1/2)*gamma*rho, n-2), rep(0, 2)), tr/n)[1:(tr-2)]
    for (i in 1:(tr-2)) #diagonal 2 above
    {
        v0[i, (i+2)] = diag2[i]
    }
    for (i in 1:(tr-n+1)) #diagonal n-1 above
    {
        if ((i-1)%%n !== 0)
            v0[i, (i+n-1)] = gamma*rho
    }
    for (i in 1:(tr-n)) #diagonal n above
    {
        v0[i, (i+n)] = rho
    }
    for (i in 1:(tr-n-1)) #diagonal n+1 above
    {
        if (i%%n !== 0)
            v0[i, (i+n+1)] = gamma*rho
    }
if (tr > 2*n)
{
    for (i in 1:(tr-2*n)) #diagonal 2n above
    {
        v0[i,(i+2*n)] = (1/2)*gamma*rho
    }
}

v0 = v0 + t(v0) #symmetric matrix
diag(v0) = 1 #main diagonal

if (struc=="DG") #0 < lambda < 1
{
    lambda = val
    lay = matrix(1:tr,byrow=T,ncol=n)
    v0 = matrix(0,nrow=tr,ncol=tr)

    for (j in 1:(tr-1))
    #only comparing to higher so don't need tr^(th) value
    {
        index1 = which(j==lay,arr.ind=T) #index[,1]=row, index[,2]=col

        for (i in (j+1):tr) #only need to do higher because symmetric
        {
            index2 = which(i==lay,arr.ind=T)
            d = abs(index1[,1] - index2[,1]) + abs(index1[,2] - index2[,2])
            v0[j,i] = lambda^d
        }
    }
}

v0 = v0 + t(v0) #create symmetric matrix
\begin{verbatim}
    diag(v0) = 1 #diagonal of 1's 

    if (struc=="DE") #0 < lambda < 1
    {
        lambda = val
        lay = matrix(1:tr,byrow=T,ncol=n)
        v0 = matrix(0,nrow=tr,ncol=tr)

        for (j in 1:(tr-1))
        {
            index1 = which(j==lay, arr.ind=T) #index[,1]=row, index[,2]=col

            for (i in (j+1):tr) #only need to do higher because symmetric
            {
                index2 = which(i==lay, arr.ind=T)
                d = sqrt((index1[,1] - index2[,1])^2 + (index1[,2] - index2[,2])^2)
                v0[j,i] = lambda^d
            }
        }

        v0 = v0 + t(v0) #create symmetric matrix
        diag(v0) = 1 #diagonal of 1's
    }
    return(v0)
\end{verbatim}

A.2 Annealing Function

The code is shown for $\xi_{D,R_1}$-minimax designs without a control. It is simple to change the loss function to calculate $\xi_{A,R_1}$-minimax, $\xi_{D,I_1}$-minimax, and $\xi_{A,I_1}$-minimax designs. The code can also be adapted if a control treatment is used.
D–optimal design for K=R

#no control, m x n array with blocks b=2

library(demogR)  #off–diagonal function
library(gregmisc)  #combinations

### INPUTS:

#tr = number of treatments
#n = number of columns in each block
#struc = correlation structure
#options are NN: Nearest Neighbour, MA: Moving Average,
#DG: Doubly Geometric, DE: Discrete Exponential
#val is the parameter value for the correlation structure
#can be rho, gamma or lambda
#alpha – controls how large the neighbourhood is
#T = initial temperature
#cool = cooling temperature
#stop = stopping temperature

### OUTPUTS:

#accepted design and value of the design;
#plot of loss function vs number of iterations

```
dr.equal = function(tr,n,struc=c("NN","MA","DG","DE"),val,
alpha,T=0.5,cool=0.9,stop=0.9ˆ200,reps=1000)
{
  v0 = corr.struc(tr,n,struc,val)  #creates the correlation structure
  v0inv = solve(v0)  #inverse of correlation structure

  C = matrix(0,mrow=tr,mcol=(tr+1))  #constant matrix to eliminate block effects
```
diag(C)=1
Ct = t(C)

x1 = matrix(0, nrow=tr, ncol=(tr+1)) #arrangement of Block 1 (x1)

x1[, (tr+1)] = rep(1, tr) #column of 1's for Block 1
diag(x1) = 1

x2s = x1 #original x2 is just x1
x2s[, (tr+1)] = rep(-1, tr) # -1 for Block 2

x1t.v0inv.x1 = t(x1)%*%v0inv%*%x1 #x1^-1 V1^-1 x1
x2t.v0inv.x2 = t(x2s)%*%v0inv%*%x2s #x2^-1 V1^-1 x2

# randomly allocate for x2
a = sample(1:tr, tr)
P = matrix(0, tr+1, tr+1)
P[tr+1, tr+1] = 1
for(j in 1:tr)
{
    P[j, a[j]] = 1
}

iter = 1
# less than T*stop — T*cool^x < T*stop
x = log(stop)/log(cool) # number of different temperatures
M = ceiling(x)*reps
# number of iterations with repetitions at each temperature
dR = vector("numeric", M) # stores the values

con = (1+alpha)^tr
x2t.v0inv.x2.new = t(P)%*%x2t.v0inv.x2s%*%P # new x2^-1 V1^-1 x2
\[
\begin{align*}
\text{xt}. \text{v0inv}. x &= \text{x1t}. \text{v0inv}. x1 + \text{x2t}. \text{v0inv}. x2 \#, X^T V^{-1} X \\
\text{xt}. \text{v0inv}. x. \text{inv} &= \text{solve}(\text{xt}. \text{v0inv}. x) \#(X^T V^{-1} X)^{-1} \\
dR[1] &= \text{con}\ast\det(C\%\text{xt}. \text{v0inv}. x. \text{inv}\%Ct) \#\text{value of the loss function}
\end{align*}
\]

\(\text{min.des} = \text{a} \#\text{stores the design with minimum value}\)
\(\text{min.val} = dR[1] \#\text{stores the loss function minimum value}\)

\(\text{acc.des} = \text{a} \#\text{stores the accepted design}\)
\(\text{acc.val} = dR[1] \#\text{stores the loss function value of the accepted design}\)

\(\text{curr.des} = \text{a} \#\text{stores the current design}\)
\(\text{curr.val} = dR[1] \#\text{stores the loss function value of the current design}\)

\[
\text{for } (i \text{ in } 2:M) \\
\{ \\
\text{new.P} = P \\
\#\text{P holds the accepted design, new.P will hold the changed design}
\}
\]

\[
\text{x} = \text{sample}(1:tr, 2) \#\text{switch these rows in permutation matrix}
\text{new.P}[x[1],] = \text{new.P}[x[2],]
\text{new.P}[x[2],] = P[x[1],]
\]

\[
\text{for } (j \text{ in } 1:tr) \#\text{going through the rows}
\{ \\
\#\text{going through elements}
\text{for } (k \text{ in } 1:tr)
\{ \\
\text{if } (\text{new.P}[j, k] == 1)
\text{curr.des}[j] = k
\}
\}
\]
#x2 = x2s*%P

x2t.v0inv.x2.new = t(new.P)*%x2t.v0inv.x2*%new.P
xt.v0inv.x = x1t.v0inv.x1 + x2t.v0inv.x2.new
xt.v0inv.x.inv = solve(xt.v0inv.x)
curr.val = con*det(C*%xt.v0inv.x.inv*%Ct)

#compare design to previous and accept according to probability rule
dR.diff = curr.val - acc.val
if (dR.diff < 0 || dR.diff == 0) #take new design
{
    acc.val = curr.val
    acc.des = curr.des
    P = new.P
    iter = iter + 1
    dR[iter] = acc.val
}

#design is better than previous design
#check against best (min.des) design
if (acc.val < min.val || acc.val == min.val)
{
    min.val = acc.val
    min.des = acc.des
}
}
else
{
    prob = exp(-dR.diff/T)
x = rbinom(1,1,prob)
if (x == 1)
{ 
    acc.val = curr.val
    acc.des = curr.des
    P = new.P
    iter = iter + 1
    dR[iter] = acc.val
}

if(i%%reps == 0)
    T = cool*T
}

if(min.val < acc.val || min.val == acc.val)
    #better design was reached earlier
    {
        acc.val = min.val
        acc.des = min.des

        P = matrix(0, tr+1, tr+1)
        P[tr+1,tr+1] = 1
        for(j in 1:tr)
        {
            P[j, acc.des[j]] = 1
        }
    }

    dR = dR[1:iter]
    plot(1:length(dR),dR,type="l",ylab="Loss Function",xlab="Iteration")
    #steepest descent routine
# evaluate all pairwise interchanges
# there are t*(t-1)/2 such possibilities
combs = combinations(tr, 2)
# combs rows gives which two are to be switched from the best design

i = 1
extra = 1:nrow(combs)
index = extra

while (i == 1)
# go through all t*(t-1)/2 designs and store the values in extra
{
    i = 2
    for (j in 1:nrow(combs))
    {
        x = combs[j,]
        new.P = P
        new.P[x[1],] = new.P[x[2],]
        new.P[x[2],] = P[x[1],]
        x2t.v0inv.x2.new = t(new.P) %*% x2t.v0inv.x2 %*% new.P
        xt.v0inv.x = x1t.v0inv.x1 + x2t.v0inv.x2.new
        xt.v0inv.x.inv = solve(xt.v0inv.x)
        extra[j] = constdet(C %*% xt.v0inv.x.inv %*% Ct)
    }
    dR.ex = min(extra) # calculate the minimum value obtained

    if (dR.ex < dR[length(dR)])
    # if better design then reset everything and try again
    {
        sw = index[extra == dR.ex][1]
        # row in combs; [1] is to choose the first minimum
x = combs[sw,]
new.P = P
new.P[x[1],] = new.P[x[2],]
new.P[x[2],] = P[x[1],]

for (j in 1:tr) #going through the rows
{
    for (k in 1:tr) #going through elements
    {
        if (new.P[j,k] == 1)
            acc.des[j] = k
    }
}

acc.val = dR.ex

dR = c(dR, acc.val)
P = new.P
i = 1
}

print(" Design")
print(acc.des)
print(acc.val)
Bibliography


