The Performance of Orthogonal Arrays with Adjoined or Unavailable Runs

Emily Bird

Doctor of Philosophy

2016

School of Mathematical and Physical Sciences
University of Technology Sydney

Certificate of Original Authorship

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student: _____

Date: _____

I have already submitted some of this work for publication. A majority of the work in Chapter 2 on OAs with adjoined runs has appeared in:

> Bird, E. M. and Street, D. J. (2016). *D*-optimal asymmetric orthogonal array plus p run designs. *Journal of Statistical Planning and Inference*, 170:64-76.

Acknowledgements

I would like to give a very big thank you to my supervisor, Prof Deborah Street, for all her guidance, advice and patience throughout this project. This thesis simply wouldn't be possible without the support of someone so understanding and adapting to my research style. And so, hence, she is thus a very thorough editor, therefore. Thank you.

I would also like to thank my co-supervisor, Dr Stephen Bush, for his proof-reading and feedback, as well as his friendly encouragement along the way.

Thanks to my family for their continued love and support through every crazy venture I throw myself into, even a PhD. Special thanks to Dad, partly for his proof-reading of early stages, but mostly for listening to my endless musings over the years.

Finally, a list of acknowledgements would not be complete without paying tribute my faithful hound, Maxie, for keeping me sane. With her relentless love and enthusiasm she constantly reminded me that no matter how compelling my excuses may be, we still have to go for a walk. She will be sorely missed.

Contents

List of Tables

List of Symbols

m	Number of factors, $m = \sum_{i=1}^{k} m_i$
k	Number of distinct factor-levels
m_i	Number of factors with s_i levels, $1 \leq i \leq k$
s_i	Number of levels for each of the m_i factors, $1 \leq i \leq k$
n	Number of runs in a design
N	Number of runs in an OA
α	Number of terms in a main effects only model, $\alpha = 1 + \sum_i (s_i - 1)m_i$
ν	Number of linear-by-linear interaction terms in a model, $0 \leq \nu \leq \binom{m}{2}$
υ	Total number of terms in a model, $\upsilon = \alpha + \nu$
$\mathbf{R}$	Design matrix of order $n \times m$
$\mathbf{r}_i$	ith row of $\mathbf{R}$; row vector of level combinations of length m, $1 \leq i \leq n$
r_{ij}	Entry in the ith row and jth column of $\mathbf{R}$; jth entry in $\mathbf{r}_i$
$\mathbf{0}_s$	Column vector of length s with all elements equal to 0
$\mathbf{1}_s$	Column vector of length s with all elements equal to 1
$\mathbf{I}_s$	Identity matrix of order s
$\mathbf{P}_s$	Contrast matrix of order $(s-1) \times s$
$P_s(l)$	$i$the column of $\mathbf{P}_s$ where $i = l + 1$, $0 \leq l \leq s - 1$
$\mathbf{p}_{sh}$	hth row of $\mathbf{P}_s$, $1 \leq h \leq s - 1$
$\mathbf{p}_{sl}$	$\mathbf{p}_{s1}$, the first row of $\mathbf{P}_s$
$\mathbf{X}$	Model matrix of order $n \times \upsilon$
$\mathbf{x}_i$	The ith row of $\mathbf{X}$, $1 \leq i \leq n$
F_j	The jth factor, $1 \leq j \leq m$
$\mathbf{f}_{jh}$	The hth column within the set of $s_i - 1$ columns in $\mathbf{X}$ corresponding to the main effects of F_j
$\mathbf{f}_{jl}$	The column in $\mathbf{X}$ corresponding to the linear component of F_j
$\mathbf{f}_{jq}$	The column in $\mathbf{X}$ corresponding to the quadratic component of F_j

$\mathbf{M}$	Information matrix of order $\upsilon \times \upsilon$ $\mathbf{M} = \mathbf{X}'\mathbf{X}$
p	Number of adjoined runs
t	Number of missing runs
$\mathbf{A}$	Model matrix of order $p \times \upsilon$ for the p adjoined runs
$\mathbf{B}$	Model matrix of order $t \times \upsilon$ for the t missing runs
$\mathbf{a}_i$	The ith row of $\mathbf{A}$
$\mathbf{b}_i$	The ith row of $\mathbf{B}$
q	Compact notation for $\mathbf{a}_i.\mathbf{a}_j'$ or $\mathbf{b}_i.\mathbf{b}_j'$ under a main effects only model
$\mathbf{a}_{ME_i}$	$\mathbf{a}_i$ truncated to the first α entries (i.e. $\mathbf{a}_i$ for a main effects only model)
$\mathbf{b}_{ME_i}$	$\mathbf{b}_i$ truncated to the first α entries (i.e. $\mathbf{b}_i$ for a main effects only model)
$\mathbf{X}_A$	Model matrix of order $(N + p) \times \upsilon$ for an OA plus p runs design
$\mathbf{X}_B$	Model matrix of order $(N - t) \times \upsilon$ for an OA minus t runs design
$\mathbf{M}_A$	Information matrix of order $\upsilon \times \upsilon$ for an OA plus p runs design $\mathbf{M}_A = \mathbf{X}'\mathbf{X} + \mathbf{A}'\mathbf{A}$
$\mathbf{M}_B$	Information matrix of order $\upsilon \times \upsilon$ for an OA minus t runs design $\mathbf{M}_B = \mathbf{X}'\mathbf{X} - \mathbf{B}'\mathbf{B}$
$\mathbf{\Omega}_A$	Matrix of order $p \times p$ to be optimised for an OA plus p runs design under a main effects only model $\mathbf{\Omega}_A = N\mathbf{I}_p + \mathbf{A}\mathbf{A}'$
$\mathbf{\Omega}_B$	Matrix of order $t \times t$ to be optimised for an OA minus t runs design under a main effects only model $\mathbf{\Omega}_B = \mathbf{B}\mathbf{B}' - N\mathbf{I}_p$
$\mathbf{\Psi}_A$	Matrix of order $p \times p$ to be optimised for an OA plus p runs design under a main effects and linear-by-linear interactions model $\mathbf{\Psi}_A = \mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$
$\mathbf{\Psi}_B$	Matrix of order $t \times t$ to be optimised for an OA minus t runs design under a main effects and linear-by-linear interactions model $\mathbf{\Psi}_B = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}' - \mathbf{I}_t$
d_i	Hamming distance between a pair of runs for the m_i factors with s_i levels
$\mathbf{D}_{kp}$	Matrix of Hamming distances of order $k \times \binom{p}{2}$
$\mathbf{d}_i$	Distance vector; ith row of $\mathbf{D}_{kp}$
$d_{i(xy)}$	Hamming distance between $\mathbf{r}_x$ and $\mathbf{r}_y$ for the m_i factors with s_i levels; Entries in $\mathbf{d}_i$

$\mathbf{z}_{(ab)}$ The column in $\mathbf{X}$ corresponding to the linear-by-linear interaction
$\mathbf{z}_{(ab)} = \mathbf{f}_{al} \circ \mathbf{f}_{bl}$

$\mathbf{u}_{(ab)}$ $\alpha \times 1$ vector $\mathbf{X}'_{ME}\mathbf{z}_{(ab)}$

$u_{(ab)cl}$ Entry in $\mathbf{u}_{(ab)}$ corresponding to $\mathbf{f}'_{cl}\mathbf{z}_{(ab)}$
$u_{(ab)cl} = \sum_{i=1}^{n} f_{al_i} f_{bl_i} f_{cl_i}$

$u_{(ab)cq}$ Entry in $\mathbf{u}_{(ab)}$ corresponding to $\mathbf{f}'_{cq}\mathbf{z}_{(ab)}$
$u_{(ab)cq} = \sum_{i=1}^{n} f_{al_i} f_{bl_i} f_{cq_i}$

γ_{abc} Coefficient for $u_{(ab)cl}$
$u_{(ab)cl} = \gamma_{abc} \frac{3}{2} \sqrt{\frac{3}{2}}$

$\omega_{(ab)c}$ Coefficient for $u_{(ab)cq}$
$u_{(ab)cq} = \omega_{(ab)c} \frac{3\sqrt{2}}{4}$

$\mathbf{Z}$ $N \times \nu$ matrix where each column is a vector $\mathbf{z}_{(ab)}$ for an interaction in the model

$\mathbf{U}$ $\alpha \times \nu$ matrix $\mathbf{X}'_{ME}\mathbf{Z}$

Abstract

Orthogonal arrays are a class of fractional factorial designs that are optimal according to a range of optimality criteria. This makes it tempting to construct fractional factorial designs by adjoining additional runs to an OA, or by removing runs from an OA, when the number of runs available for the experiment is only slightly larger or smaller than the number in the OA.

In this thesis we examine the performance of *OA plus p runs designs* and *OA minus t runs designs* in the context of D-optimality and model-robustness. Although we attempt to make general observations where possible, our primary goal is to inform the use of quantitative factors, hence we focus on factors at more than two levels to allow for some curvature in the model.

We begin by considering the performance of these designs under a main effects only model, and show that optimality depends only on the pairwise Hamming distance of the adjoined (or removed) runs. We present an algorithm for finding optimal Hamming distances and provided general methods for constructing optimal sets of runs once the optimal pairwise Hamming distances have been identified.

In order to consider the performance of these designs when interaction terms are included in the model, it transpired that we require a complete set of geometrically non-isomorphic designs to study. Thus, we enumerate all geometric isomorphism classes for symmetric OAs with ternary factors and 18 runs, and prove that these classes cover the entire OA$[18, 3^m]$ space.

We then consider the inclusion of a subset of linear-by-linear interactions in the model, and derive matrices to be optimised under this setting. We conduct an empirical study on the OA$[18, 3^m]$ we have enumerated and give examples of D-optimal and model-robust designs for each of the design spaces, that is, for each of $m = 3$, 4, 5, 6 and 7.

Chapter 1

Introduction

This chapter lays the ground work for the area of research at the heart of this thesis. We begin in Section 1.1 by discussing the core concepts and terminology in the field of fractional factorial designs, and we go on to give some standard results in Section 1.2. We then review the literature in Section 1.3, and in Section 1.4 we outline the aims of this project and give an overview of the work presented in subsequent chapters.

1.1 Background

A *factorial design* is a design for an experiment in which several *factors* (such as varieties of corn, fertilisers or antibiotics) are applied to each experimental unit. Each factor is applied at one of two or more *levels* across the experiment. The levels may be *quantitative* (as with the amount of fertiliser) or *qualitative* (where the levels refer to different varieties of corn, say) but in either case the levels of a factor are represented by the elements of a finite set, usually by $\{0, 1, 2, \ldots, s-1\}$ when the factor occurs at s levels.

We see that if there are k sets of m_i factors, each at s_i levels, then there are $\prod_{i=1}^{k} s_i^{m_i}$ different possible combinations of factor levels. Thus, collecting responses from all of these combinations becomes impossible for all but the smallest experiments. The question then becomes, how can we get useful information about the factors and their interactions (if any) by only observing responses on a subset of the possible factor-level combinations?

This thesis investigates the statistical properties of subsets of factor-level combinations, known as *fractional factorial designs* (FFDs). The overall motivation for the construction of FFDs is to maximise the amount of information gained from

a given number of experimental units whilst also minimising the cost (number of runs) of gaining that information[1]. More runs will usually give rise to more information, either in the form of larger coverage of the design space, or more information about the variability of results, so a balance needs to be struck between these opposing goals.

For example, consider the following investigation in the field of engineering, where the response is the average height (μM) of 12 burrs from holes punched by a punching device. The data from the 18 settings measured in this experiment is reproduced from Schoen (2009).

Table 1.1.1: Data for the punching experiment

Expt.	Clearance	Rim	Angle	Lubricant	Burr
1	18.40	0.50	0	Yes	71.4
2	18.10	0.25	30	No	30.2
3	18.25	0.50	0	No	38.1
4	18.10	0	0	No	26.9
5	18.40	0	30	Yes	72.3
6	18.40	0.25	15	Yes	71.4
7	18.25	0	30	No	26.8
8	18.10	0.50	15	No	15.0
9	18.40	0.50	30	No	48.8
10	18.25	0	0	Yes	51.3
11	18.10	0.25	0	Yes	16.0
12	18.10	0.50	30	Yes	21.8
13	18.25	0.25	15	No	37.4
14	18.40	0.25	0	No	104.9
15	18.10	0	15	Yes	20.9
16	18.25	0.50	15	Yes	26.9
17	18.25	0.25	30	Yes	30.8
18	18.40	0	15	No	57.2

Source: Schoen (2009)

Notice that each of the three quantitative factors (clearance, rim and angle) have three distinct levels, and the qualitative factor (lubricant) has two levels (Yes/No). Therefore, the experimenter had $3^3 \times 2 = 54$ possible combinations to choose from. Why did the experimenter choose these 18 in particular? And why did they run each combination only once? These questions lie at the heart of the construction of FFDs and will be explored in this thesis.

The remainder of this section will illustrate the concepts underlying FFDs using the example described above. But we first need to introduce some terminology

[1]This thesis will assume all factor-level combinations are equally costly to run. Atkinson et al. (2007) discuss methods for dealing with factor-level combinations with different costs to run in Section 10.11

and notation.

1.1.1 Terminology

Throughout this thesis we will refer to designs as having k sets of m_i factors, each at s_i levels, $1 \leq i \leq k$. Hence, there are $m = \sum_{i=1}^{k} m_i$ factors in total, the jth of which is labelled F_j. The *design space* consists of all possible $\prod_{i=1}^{k} s_i^{m_i}$ *factor-level combinations*, while a FFD contains only a subset of these. Each factor-level combination in the FFD is referred to as a *run*; if a run appears more than once in an FFD it is said to be *repeated*. The total number of runs in an FFD is denoted n.

Example 1.1.1. The punching experiment consists of $k = 2$ sets of factors with $m_1 = 3$, $s_1 = 3$, $m_2 = 1$ and $s_2 = 2$. Hence, there are $m = 3 + 1 = 4$ factors in total which we refer to as F_1, F_2, F_3 and F_4 respectively. The first three factors are numeric, hence quantitative, and the last factor is categorical, hence qualitative. The ordered levels of clearance, angle and rim are $\{18.1, 18.25, 18.4\}$, $\{0, 0.25, 0.5\}$ and $\{0, 15, 30\}$ respectively. In the context of design construction, we would label these levels as $\{0, 1, 2\}$ for all three factors. The levels of lubricant are {Yes, No}. We label these levels as $\{0, 1\}$, and note that the order of assignment is irrelevant since there is no inherent ordering in these levels. The design in Table 1.1.1 contains $n = 18$ runs of the $3^3 \times 2^1 = 54$ level combinations in the design space. None of these level combinations appear more than once, hence this design has no repeated runs. □

The *design matrix*, $\mathbf{R}$, is a matrix of order $n \times m$ where each entry, r_{ij}, represents the level of the jth factor in the ith run. Table 1.1.2 below gives two representations of the deign matrix, $\mathbf{R}$, for the data in Table 1.1.1. The rows of the matrix on the left are in the same order as Table 1.1.1; the rows in the table on the right have been ordered lexicographically.

As we shall see in Section 1.1.3, the way in which the runs of an experiment are ordered (or, equivalently, the order of the rows in a design matrix) does not influence the statistical properties of the design, and so for consistency we will always assume lexicographical ordering of the rows of $\mathbf{R}$.

It is worth noting at this point that one of the challenges in the construction of FFDs is that the number of potential designs quickly becomes so large that it is often not feasible to enumerate and compare them all. Since the design space contains $\prod_{i=1}^{k} s_i^{m_i}$ level combinations, there are $\binom{\prod_{i=1}^{k} s_i^{m_i}}{n}$ designs with n distinct runs to choose from. This space increases exponentially when repeated runs are

Table 1.1.2: Two representations of **R** for the data in Table 1.1.1

2	2	0	1	0	0	0	0
0	1	2	0	0	0	1	1
1	2	0	0	0	1	0	1
0	0	0	0	0	1	2	0
2	0	2	1	0	2	1	0
2	1	1	1	0	2	2	1
1	0	2	0	1	0	0	1
0	2	1	0	1	0	2	0
2	2	2	0	1	2	0	0
1	0	0	1	1	2	1	1
0	1	0	1	1	1	1	0
0	2	2	1	1	1	2	1
1	1	1	0	2	0	1	0
2	1	0	0	2	0	2	1
0	0	1	1	2	1	0	0
1	2	1	1	2	1	1	1
1	1	2	1	2	2	0	1
2	0	1	0	2	2	2	0

also considered.

Example 1.1.2. The design used in the punching experiment is one of $\binom{54}{18} = 9.6926349 \times 10^{13}$ potential designs with 18 distinct runs. □

Orthogonal arrays (OAs) form a particular family of FFDs.

Definition 1.1.1. An orthogonal array $\text{OA}[N, s_1^{m_1} \times s_2^{m_2} \times \cdots \times s_k^{m_k}]$ is an array of order $N \times m$ where $m = \sum_{i=1}^{k} m_i$, and each set of m_i columns contain elements from a distinct set of s_i symbols arranged such that, for any pair of columns amongst all m columns, every pair of level combinations appears equally often[2]. □

An OA can be thought of as an FFD when we let the columns of the array represent the factors, the rows represent the runs, and the symbols represent the levels. That is, we let the OA be the design matrix. For example, the design matrix, **R**, in Table 1.1.2 (either representation) is an $\text{OA}[18, 3^3 \times 2^1]$. Note that we use N to denote the number of runs in the OA, whereas we use n to denote the number of runs in a design, hence $n = N$ when a design is based on an OA.

Schoen (2009) notes that:

> OAs are generally considered as a good kind of experimental designs because of the balance between any two factors. If we take the mean of

[2]Our definition for OAs assumes *strength* 2. In general, an OA of strength t will have equal replication of every t-tuple of level combinations in every set of t columns. This thesis will focus on strength 2 OAs only.

> the observations at some level of a factor, we know that the observations used for averaging are taken at all possible levels of the other factors. As a result, the main effect of a factor is not affected by main effects of other factors, and one can assess a main effect simply by looking at the means of the observations for each level of the factor.

Pure error is a model-independent estimate of the variability inherent in the experiment. To obtain such an estimate it is necessary to have at least one repeated run in the design so that multiple responses are observed at the same level combination.

When n depends on the resources available, it may be considered as an upper limit rather than a fixed value. Suppose we have two designs with n_1 and n_2 runs respectively, where $n_1 < n_2$ and both are less than some upper limit imposed by the availability of resources. All else being equal, the design with fewer runs will be preferred as it minimises cost, but if the variance of the parameter estimates for the design with n_2 runs is smaller than that of the design with n_1 runs, then the cost of the extra runs may be considered worthwhile.

1.1.2 The Model

We use matrix notation to define the model which we write as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1.1}$$

where $\mathbf{y}$ is the column vector of n responses, $\mathbf{X}$ is the $n \times \upsilon$ *model matrix* (defined below), $\boldsymbol{\beta}$ is the column vector of υ model parameters, and $\boldsymbol{\epsilon}$ is the column vector of n random errors.

As mentioned previously, factors can be either quantitative or qualitative. In this thesis we will assume that all quantitative factors have equally spaced levels. In Chapter 2 we consider *main effects only models*, that is, models that contain the intercept and the main effects of all factors. In Chapters 4 and 5 we consider *main effects and linear-by-linear interactions models*, that is, models that contain all the terms in the main effects only model plus the linear-by-linear component of a subset of the two-factor interactions. We will now consider each of these models in more detail.

Main Effects Only Model

The *main effects* of quantitative factors can be constructed using *orthogonal polynomial contrasts* which allow for some curvature in the response surface to be estimated. When factors are qualitative, the concepts of 'linear', 'quadratic' etc. do not apply, so any set of orthogonal contrasts can be used. Hence, designs with qualitative factors are typically analysed via ANOVA-type models, whilst designs with quantitative factors can also be analysed using *response surface methodology* (RSM). Unless otherwise stated, this thesis will use orthogonal polynomial contrast coding to construct the main effects of all factors.

In Chapter 2 we want to estimate the main effects only model, hence we want to replace each level of each factor with appropriate entries from a complete set of orthogonal polynomials; that is, a set of $s-1$ orthogonal polynomials for a factors with s levels. To do this we follow the approach of Chatzopoulos et al. (2011). We define a *contrast matrix* as follows.

Definition 1.1.2. Let $\mathbf{I}_s$ be the identity matrix of order s, $\mathbf{1}_s$ be the $s \times 1$ vector with all elements equal to 1, and $\mathbf{0}_s$ be the $s \times 1$ vector with all elements equal to 0. $\mathbf{P}_s$ is a contrast matrix of order $(s-1) \times s$ if $\mathbf{P}_s\mathbf{P}_s' = s\mathbf{I}_{s-1}$ and $\mathbf{P}_s\mathbf{1}_s = \mathbf{0}_{s-1}$. We denote the columns of $\mathbf{P}_s$ by $P_s(0), \ldots, P_s(s-1)$. □

By definition, the sum of coefficients in a contrast must be zero. Thus, the second condition in Definition 1.1.2, $\mathbf{P}_s\mathbf{1}_s = \mathbf{0}_{s-1}$, ensures that the $s-1$ rows of $\mathbf{P}_s$ are all contrasts. The first condition in Definition 1.1.2, $\mathbf{P}_s\mathbf{P}_s' = s\mathbf{I}_{s-1}$, ensures that the $s-1$ contrasts are mutually orthogonal since the off-diagonal entries of $\mathbf{P}_s\mathbf{P}_s'$ are zero. Since the diagonal entries of $\mathbf{P}_s\mathbf{P}_s'$ are s, this condition also ensures that the contrasts are normalised so that the sum of the squared coefficients is s for each contrast.

In order to fit a polynomial model, we will use orthogonal polynomial contrast coding, which can be obtained from numerous sources. We assume the levels of all factors are equally spaced, which allows us to use the contrasts given in Kuehl (2000). The raw (un-normalised) orthogonal polynomial contrasts from Kuehl (2000) for binary and ternary factors are given in Table 1.1.3 below.

Table 1.1.3: Polynomial contrast coding for two-level and three-level factors

Two-Level Factors	
Contrast	**Coefficients**
Linear	−1 1

Three-Level Factors			
Contrast	**Coefficients**		
Linear	−1	0	1
Quadratic	1	−2	1

As the name suggests, a set of orthogonal polynomial contrasts are mutually orthogonal. Thus, in order to represent these orthogonal polynomial contrasts in the form of a contrast matrix as defined in Definition 1.1.2, all that is required is to normalise each contrast so that the squared coefficients sum to s. We can do this by dividing each coefficient by $\sqrt{\left(\sum_{i=1}^{s} c_i^2\right)/s}$ where c_i is the ith coefficient in the given contrast. The following example illustrates this.

Example 1.1.3. Suppose $s = 2$. Note that the sum of squared coefficients in the linear contrast for two-level factors in Table 1.1.3 is $(-1)^2+1^2 = 2 = s$. Hence, this contrast satisfies the conditions in Definition 1.1.2. Thus, the polynomial contrast matrix for $s = 2$ is

$$\mathbf{P}_2 = \begin{bmatrix} -1 & 1 \end{bmatrix}.$$

Now suppose $s = 3$. Thus, we would like to represent the orthogonal polynomial contrasts for three-level factors in Table 1.1.3 as a contrast matrix that satisfies the conditions of Definition 1.1.2. We do this by dividing each coefficient in the linear contrast by $\sqrt{\left((-1)^2 + 0^2 + 1^2\right)/3} = \sqrt{2/3}$, and dividing each coefficient in the quadratic contrast by $\sqrt{\left(1^2 + (-2)^2 + 1^2\right)/3} = \sqrt{2}$. Hence, the polynomial contrast matrix for $s = 3$ is

$$\mathbf{P}_3 = \begin{bmatrix} -\sqrt{\frac{3}{2}} & 0 & \sqrt{\frac{3}{2}} \\ \frac{1}{\sqrt{2}} & -\sqrt{2} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

□

We use the polynomial contrast matrices to construct the model matrix, $\mathbf{X}$, as follows.

Definition 1.1.3. Let $\mathbf{R}$ be the $n \times m$ design matrix for a design with n runs and k sets of m_i factors at s_i levels, so $m = \sum_{i=1}^{k} m_i$. Let r_{ij} denote the entry in the ith row and jth column of $\mathbf{R}$. Then the ith row of the model matrix for the main effects only model, $\mathbf{X}$, is

$$\mathbf{x}_i = \begin{bmatrix} 1 & P_{s_1}(r_{i1})' & P_{s_1}(r_{i2})' & \dots & P_{s_1}(r_{im_1})' & P_{s_2}(r_{i(m_1+1)})' & \dots & P_{s_k}(r_{im})' \end{bmatrix}$$

Hence $\mathbf{X}$ is of order $n \times \alpha$ where $\alpha = 1 + \sum_{i=1}^{k}(s_i - 1)m_i$. □

Notice that the order of $\mathbf{X}$ in Equation (1.1.1) was given as $n \times \upsilon$, whereas the order of $\mathbf{X}$ in Definition 1.1.3 is $n \times \alpha$. In general, we use υ to denote the number of parameters in any given model, however we use α to denote the number of parameters in the specific case of a main effects only model; that is, for a main effects only model $\upsilon = \alpha$. This notation becomes useful when we consider additional terms in the model, which we will examine in the next section, as it allows

us to specify how many of the v parameters are involved in the main effects only model.

In general, the hth column within the set of $s_i - 1$ columns associated with factor F_j is labelled $\mathbf{f}_{jh}$, and the associated parameter in $\boldsymbol{\beta}$ is labelled β_{jh}. The parameter for the intercept is denoted β_0. The following example illustrates this notation.

Example 1.1.4. Consider the lexicographically ordered OA$[18, 3^3 \times 2^1]$ in Table 1.1.2. In order to construct the main effects only model matrix, we will use the polynomial contrast matrices constructed in Example 1.1.3. Hence, for each row of the lexicographically ordered OA we construct a row of the model matrix by starting with the entry 1 and then appending the transpose of the corresponding column in the relevant polynomial contrast matrix for each of the entries in the OA. The resulting main effects only model matrix is

$$\mathbf{X} = \begin{array}{c} \begin{array}{ccccccc} \mathbf{f}_{11} & \mathbf{f}_{12} & \mathbf{f}_{21} & \mathbf{f}_{22} & \mathbf{f}_{31} & \mathbf{f}_{32} & \mathbf{f}_{41} \end{array} \\ \left[\begin{array}{ccccccc} 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -1 \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 1 \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 1 \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -1 \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & -1 \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 1 \\ 1 & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 1 \\ 1 & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -1 \\ 1 & 0 & -\sqrt{2} & 0 & -\sqrt{2} & 0 & -\sqrt{2} & -1 \\ 1 & 0 & -\sqrt{2} & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 1 \\ 1 & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -1 \\ 1 & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 1 \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & -1 \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 1 \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -1 \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 0 & -\sqrt{2} & 1 \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 1 \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -1 \end{array}\right] \end{array}$$

and the associated vector of parameters is

$$\boldsymbol{\beta}' = \left[\begin{array}{ccccccccc} \beta_0 & \beta_{11} & \beta_{12} & \beta_{21} & \beta_{22} & \beta_{31} & \beta_{32} & \beta_{41} \end{array}\right]. \qquad \square$$

Main Effects and Linear-by-Linear Interactions Model

In Chapters 4 and 5 we will extend the main effects only model to consider models that contain a subset of the two-factor *interactions*. The full interaction between factors F_a and F_b at levels s_a and s_b respectively, involves the construction of the Hadamard product (element-wise product) between each of the $s_a - 1$ main effects of factor F_a with each of the $s_b - 1$ main effects of factor F_b. Thus, it has $(s_a - 1)(s_b - 1)$ contrasts associated with it and requires $(s_a - 1)(s_b - 1)$ degrees of freedom to estimate. Since we are interested in minimising the number runs in a design, we are often limited by the number of degrees of freedom available to estimate the terms in the model. For example, consider an OA$[18, 3^7]$, one of the OAs on which we will focus on in the coming chapters. We require $1 + 2 \times 7 = 15$ degrees of freedom to estimate the intercept and all main effects, so we are left with only $18 - 15 = 3$ degrees of freedom for fitting additional terms. A single two-factor interaction in this setting requires $2 \times 2 = 4$ degrees of freedom, hence we are not able to fit any of the $\binom{7}{2} = 21$ potential full two-factor interactions. Xu et al. (2004) note that "the linear-by-linear active contrasts occur more often in practice than higher-order contrasts", that is, of the $(s_a - 1)(s_b - 1)$ contrasts in the full two-factor interaction between factors F_a and F_b, the Hadamard product of $\mathbf{f}_{a1}$ and $\mathbf{f}_{b1}$ is the contrast most likely to be active. Accordingly, many researchers have focused on only the *linear-by-linear* component of the two-factor interactions, thereby reducing the required degrees of freedom for each interaction to 1 (see, for example, Jones et al. (2007), Lekivetz and Tang (2014)).

In Chapter 4 we consider models that include one linear-by-linear two-factor interaction, and in Chapter 5 we extend this work to any subset of ν linear-by-linear two-factor interactions. In order to do so, we must append ν columns to the model matrix and ν elements to the vector of parameters.

Definition 1.1.4. Let $\mathbf{X}_{ME}$ be the $n \times \alpha$ model matrix for the main effects only model as per Definition 1.1.3. The partitioned model matrix for the main effects and linear-by-linear interactions model is

$$\mathbf{X} = \left[\begin{array}{cc} \mathbf{X}_{ME} & \mathbf{Z} \end{array} \right]$$

where $\mathbf{Z}$ is a $n \times \nu$ matrix containing the ν columns corresponding to Hadamard product of the linear main effects for the relevant factors. Hence, $\mathbf{X}$ is of the order $n \times \upsilon$ where $\upsilon = \alpha + \nu$. □

To distinguish between the potential two-factor interactions that may be included in the model, we will introduce the notation $\mathbf{z}_{(ab)}$ to represent the linear-by-linear

component of the interaction between factors F_a and F_b, and the associated parameter in $\boldsymbol{\beta}$ will be labelled $\beta_{(ab)}$.

Since the concept of 'linear' is only applicable to quantitative factors, the following example extends the model matrix developed for the hole punching experiment by considering the linear-by-linear two-factor interactions between the three quantitative factors only.

Example 1.1.5. The main effects only model matrix in Example 1.1.4 requires $\alpha = 1 + 2 \times 3 + 1 \times 1 = 8$ degrees of freedom, hence we are left with $18 - 8 = 10$ degrees of freedom for additional terms. There are $\binom{3}{2} = 3$ potential two-factor interactions to be considered between the three quantitative factors, so we may be able to fit the linear-by-linear component of all of these interactions. Let $\nu = 3$, then the matrix of interactions to be appended to the model matrix defined in Example 1.1.4 is

$$\mathbf{Z} = \begin{array}{c} \begin{matrix} \mathbf{z}_{(12)} & \mathbf{z}_{(13)} & \mathbf{z}_{(23)} \end{matrix} \\ \begin{bmatrix} \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & -\frac{3}{2} & 0 \\ -\frac{3}{2} & 0 & 0 \\ -\frac{3}{2} & -\frac{3}{2} & \frac{3}{2} \\ 0 & 0 & \frac{3}{2} \\ 0 & 0 & -\frac{3}{2} \\ 0 & 0 & -\frac{3}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{3}{2} & 0 & 0 \\ -\frac{3}{2} & \frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 \\ \frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \end{bmatrix} \end{array}$$

and the extended vector of parameters is

$$\boldsymbol{\beta}' = \begin{bmatrix} \beta_0 & \beta_{11} & \beta_{12} & \beta_{21} & \beta_{22} & \beta_{31} & \beta_{32} & \beta_{41} & \beta_{(12)} & \beta_{(13)} & \beta_{(23)} \end{bmatrix}. \qquad \square$$

Notice that the construction of the model matrix requires us to assume, or 'know', which terms are to be included in the model. This raises questions about the

assumptions underlying the structure of the model and we will address these in this thesis. To facilitate such a discussion it will be helpful to define the following key terms that are used throughout the literature.

Aliasing. Two linearly dependent terms in the model are said to be *fully aliased.* Fully aliased terms are indistinguishable from one another and so we must rely on the assumption that one of them is negligible to be able to say anything about the other. Two terms in the model that are neither fully aliased nor independent from one another are said the be *partially aliased.*[3] Example 1.1.8 in the following section illustrates aliasing with a small binary design.

Effect Hierarchy. As defined by Wu and Hamada (2011), *effect hierarchy* is the assumption that (i) lower-order effects are more likely to be important than higher-order effects, (ii) effects of the same order are equally likely to be important. This implies that the main effects are more likely to be important than the interaction terms, and that two-factor interactions are more likely to be important than higher-order interactions.

Effect Sparsity. The assumption that "only a small number of factorial effects are important relative to the rest. This can be called the Pareto Principle in experimental designs or effect sparsity" (Hamada and Wu (1992)).

Effect Heredity. Hamada and Wu (1992) define *effect heredity* as "when a two factor interaction is significant, at least one of the corresponding factor main effects is also significant". This is also referred to as *weak heredity.*

Functional Marginality. Tsai et al. (2007) paraphrase the definition of *functional marginality* from McCullagh and Nelder (1989) to mean "that every term in the model must be accompanied by all terms marginal to it, whether these are large or small". This concept is also known as *strong heredity.* For example, suppose the linear-by-linear component of the interaction between factors F_a and F_b is included in the model, then the linear components of the main effects of both F_a and F_b must also be included in the model, but neither of the quadratic terms need to be included.

Note that we have so far only mentioned two-factor interactions without consideration of any higher-order interactions. In this thesis we will assume all higher order interactions are negligible, as is the case in the huge majority of the literature. This is justified by invoking a combination of effect hierarchy and effect sparsity.

[3] We note that aliasing is generally defined in the context of *defining relations* (see, for example, Wu and Hamada (2011)) which is not the focus of this thesis.

1.1.3 Measures of Goodness

In order to evaluate the various statistical properties of FFDs we require some measure of 'goodness'. The next example introduces three designs with competing virtues to illustrate the statistical properties we would like to assess.

Example 1.1.6. Suppose we are only interested in the three quantitative factors in the hole punching experiment, and we are constrained to $n = 14$ runs. The design matrices of three potential designs are listed in Table 1.1.4. □

Table 1.1.4: Three design matrices with 3 three-level factors in 14 runs

$\mathbf{R}_1$			$\mathbf{R}_2$			$\mathbf{R}_3$		
0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	1
0	0	1	0	1	0	0	1	0
0	0	1	0	1	2	0	1	2
0	1	0	0	2	1	0	2	1
0	1	0	1	0	0	0	2	2
0	1	2	1	0	2	1	0	0
0	1	2	1	0	2	1	0	2
0	2	1	1	1	1	1	1	1
0	2	1	1	1	1	1	1	2
1	0	0	1	2	0	1	2	0
1	0	0	1	2	0	1	2	1
2	0	1	2	0	1	2	0	1
2	0	1	2	0	1	2	0	2

Which design in Table 1.1.4 is 'best'? Notice that $\mathbf{R}_1$ contains 7 distinct runs, each appearing twice; $\mathbf{R}_2$ contains 10 distinct runs, a subset of which are repeated; and $\mathbf{R}_3$ is made up of 14 distinct runs. If an estimate of pure error is required then $\mathbf{R}_1$ will be preferred as it has the maximum number of repeated runs. If, however, the goal of the experiment is to maximise the coverage of the design space then $\mathbf{R}_3$ will be considered superior as it has the largest number of distinct runs. $\mathbf{R}_2$ can be viewed as a compromise between these two objectives as it contains more distinct runs than $\mathbf{R}_1$ (hence greater coverage of the designs space) but the presence of repeated runs also allows for the estimation of pure error. Hence, the decision as to which design is 'best' will depend on the perspective of the experimenter.

We summarise many of the numerous optimality criteria that have been developed to rank designs in Section 1.3.1, but to aid the current discussion it will be helpful to mention the key *alphabetic criteria* here.

Given an $n \times \upsilon$ model matrix $\mathbf{X}$, the $\upsilon \times \upsilon$ *information matrix* is given by $\mathbf{M} = \mathbf{X}'\mathbf{X}$, and, for a model of full rank, $\mathbf{C} = \mathbf{M}^{-1}$ is the covariance matrix of the parameter

estimates. In the one parameter case, we usually prefer estimates which minimise the variance of the parameter estimate. When there are several parameters there are many ways to 'minimise' the variance. Three of the most common are given below (Atkinson et al. (2007)).

Let $\lambda_1, \ldots, \lambda_\upsilon$ be the υ eigenvalues of the information matrix. Then the A-, D-, and E-criteria are defined as follows:

1. A**-optimality** minimises the sum of the variances of the parameter estimates $\min\left(\sum_{i=1}^{\upsilon} \frac{1}{\lambda_i}\right)$
2. D**-optimality** minimises the generalised variance of the parameter estimates $\min\left(\prod_{i=1}^{\upsilon} \frac{1}{\lambda_i}\right)$
3. E**-optimality** minimises the variance of the least well-estimated parameter $\min\left(\max\left(\frac{1}{\lambda_i}\right)\right)$

Note that A- and D- are equivalent to minimising the trace of $\mathbf{C}$ and minimising the determinant of $\mathbf{C}$ respectively, while E-optimality minimises the maximum eigenvalue of $\mathbf{C}$.

Example 1.1.7. The A-, D-, and E-criteria for each of the designs in Example 1.1.6 are presented in Table 1.1.4. These criteria have been calculated for both the main effects only model (ME) and the main effects and linear-by-linear interactions model involving all three interactions (ME+2fi).

Table 1.1.5: A-, D-, and E-criteria for each of the designs in Table 1.1.4

		$\mathbf{R}_1$	$\mathbf{R}_2$	$\mathbf{R}_3$
ME	A-	1.4444	0.8870	0.7406
	D-	1.6745×10^{-7}	1.5877×10^{-8}	9.4066×10^{-9}
	E-	0.6895	0.4367	0.3692
ME+2fi	A-	–	9.1019	3.1552
	D-	–	9.3027×10^{-9}	6.4434×10^{-10}
	E-	–	7.2912	1.7543

Given that we are aiming to minimise A-, D-, and E-criteria, it appears that $\mathbf{R}_3$ is the superior design on all counts − it has the smallest value for every criterion in both models. Recall, however, that $\mathbf{R}_3$ has no repeated runs, hence it is unable to estimate pure error. Choosing $\mathbf{R}_3$ would imply we are willing to sacrifice the estimation of pure error in order to maximise the precision of the parameter estimates. If pure error was of particular interest we may prefer to use $\mathbf{R}_1$ as it has the maximum number of repeated runs, but it covers such a small portion of the design

space that it isn't able to estimate a model including any two-factor interactions, so we would need to be confident that these terms are indeed negligible. □

This example highlights the compromises that needs to be made between competing objectives. In the next section we will give a comprehensive list to lay the ground work for the criteria to be discussed in Section 1.3. But first, the following example shows how the information matrix relates to the concept of aliasing defined above.

Example 1.1.8. Consider the following OA[4,2^3].

F_1	F_2	F_3
0	0	0
0	1	1
1	0	1
1	1	0

The model matrix and corresponding information matrix for a model containing the linear-by-linear interaction between F_2 and F_3 are

$$\mathbf{X} = \begin{array}{c} \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{f}_{(11)} & \mathbf{f}_{(21)} & \mathbf{f}_{(31)} & \mathbf{z}_{(23)} \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 \end{bmatrix} \end{array}, \quad \mathbf{M} = \begin{array}{c} \\ \mathbf{f}_{(11)} \\ \mathbf{f}_{(21)} \\ \mathbf{f}_{(31)} \\ \mathbf{z}_{(23)} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{f}_{(11)} & \mathbf{f}_{(21)} & \mathbf{f}_{(31)} & \mathbf{z}_{(23)} \end{array} \\ \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & -4 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & -4 & 0 & 0 & 4 \end{bmatrix} \end{array}$$

Notice from $\mathbf{X}$ that $\mathbf{f}_{11} = -\mathbf{z}_{(23)}$. Thus, $\mathbf{f}_{11}$ and $\mathbf{z}_{(23)}$ are linearly dependent on one another, hence the corresponding terms in the model are fully aliased. This can also be seen in the non-zero off-diagonal terms of the information matrix, each of which corresponds to a pair of terms in $\mathbf{X}$. The zero off-diagonal entries in $\mathbf{M}$ indicate that the relevant pair of terms are linearly independent. Notice that the off-diagonal term in $\mathbf{M}$ corresponding to the pair of terms ($\mathbf{f}_{11}$, $\mathbf{z}_{(23)}$) is equal in absolute value to the diagonal term corresponding to $\mathbf{f}_{11}$ and $\mathbf{z}_{(23)}$. This indicates $\mathbf{f}_{11}$ and $\mathbf{z}_{(23)}$ are fully aliased.

Now suppose we append the triple of levels 100 to the OA. The associated model matrix and information matrix for a model containing the linear-by-linear interaction between F_2 and F_3 are

$$\mathbf{X} = \begin{array}{c} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{f}_{(11)} & \mathbf{f}_{(21)} & \mathbf{f}_{(31)} & \mathbf{z}_{(23)} & \end{array} \\ \begin{bmatrix} -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \end{array}, \quad \mathbf{M} = \begin{array}{c} \\ \mathbf{f}_{(11)} \\ \mathbf{f}_{(21)} \\ \mathbf{f}_{(31)} \\ \mathbf{z}_{(23)} \end{array} \begin{array}{c} \begin{array}{ccccc} \mathbf{f}_{(11)} & \mathbf{f}_{(21)} & \mathbf{f}_{(31)} & \mathbf{z}_{(23)} \end{array} \\ \begin{bmatrix} 5 & 1 & -1 & -1 & 1 \\ 1 & 5 & -1 & -1 & -3 \\ -1 & -1 & 5 & 1 & -1 \\ -1 & -1 & 1 & 5 & -1 \\ 1 & -3 & -1 & -1 & 5 \end{bmatrix} \end{array}$$

Notice from $\mathbf{X}$ that the dependency $\mathbf{f}_{11} = -\mathbf{z}_{(23)}$ does not hold in the last row. Thus, $\mathbf{z}_{(23)}$ and $\mathbf{f}_{11}$ are no longer linearly dependent and the relevant off-diagonal entry in $\mathbf{M}$ is non-zero but less than the $\mathbf{f}_{11}$ and $\mathbf{z}_{(23)}$ diagonal terms in absolute value. Hence, the terms $\mathbf{f}_{11}$ and $\mathbf{z}_{(23)}$ are said to be partially aliased. In fact, all off-diagonal entries are non-zero and less than $n = 5$ in absolute value. Hence, there is some degree of partial aliasing within all pairs of terms in this new design. □

1.1.4 Experimenter's Objectives

Atkinson et al. (2007) give the following list of objectives, the first 14 of which were taken from Box and Draper (1975).

1. Generate a satisfactory distribution of information throughout the region of interest, which may not coincide with the design region.
2. Ensure that the fitted value, $\hat{y}(x)$ at x, be as close as possible to the true value $\eta(x)$ at x.
3. Make it possible to detect lack of fit.
4. Allow estimation of transformations of both the response and the quantitative experimental factors.
5. Allow experiments to be performed in blocks.
6. Allow designs of increasing order to be built up sequentially. Often a second-order designs will follow one of the first order.
7. Provide an internal estimate of error from replication.
8. Be insensitive to wild observations and the violation of the usual normal theory assumptions.
9. Require a minimum number of experimental runs.
10. Provide simple data patterns that allow ready visual appreciation.
11. Ensure simplicity of calculation.
12. Behave well when errors occur in the setting of the experimental variables.
13. Not require an impractically large number of levels of the experimental factors.
14. Provide a check on the 'constancy of variance' assumption.

15. Orthogonality: the designs have a diagonal information matrix, leading to uncorrelated estimates of the parameters.
16. Rotatability: the variance of $\hat{y}(x)$ depends only on the distance from the centre of the experimental region.

Similar lists have appeared throughout the literature, such as the simplified list of eight in Robinson and Anderson-Cook (2010). Different optimality criteria effectively focus on different points in the list above. For instance, point 3 regarding lack of fit relates to pure error. Recall that to obtain an estimate of pure error we require at least one repeated run in the design. Thus, an optimality criterion that focuses on point 3 will incorporate the number of repeated runs in the design.

Some designs may perform equally well based on a given criterion. We may be able to discriminate between these designs based on other criteria, but if they perform equally well on all criteria we are interested in then they are essentially the same for our purposes. Two designs are said to be 'essentially the same', or *isomorphic*, if they possess the same statistical properties. We will now discuss the different types of isomorphism that we will examine in this thesis.

1.1.5 Isomorphism

Note that neither the order of the rows nor the order of the columns in $\mathbf{X}$ has an effect on the eigenvalues of $\mathbf{M}$. Thus, in relation to the A-, D-, and E-criteria, row and column permutations have no effect on the statistical properties of a design. This is the case for all the criteria that we will consider. Hence, any two designs in which one can be constructed from one another via row and/or column permutations are considered to be essentially the same. But what happens when we permute the levels within the factors?

When the factors have more than two levels, isomorphism depends on whether the factors are qualitative or quantitative. Two designs consisting of qualitative factors are *combinatorially isomorphic* if one can be constructed from the other through any combination of row, column or level permutations. Quantitative designs, however, have an inherent ordering in the levels of the factors, hence any permutation that disrupts this order may not retain the same statistical properties. Thus, two designs consisting of quantitative factors are *geometrically isomorphic* if one can be constructed from the other through any combination of row or column permutations or reversing the levels in one or more factors. That is, level-permutations that disrupt the inherent ordering of the factors are not allowed. A set of designs that are all combinatorially or geometrically isomorphic to one another are said to

be in the same combinatorial or geometric *isomorphism class*.

Example 1.1.9. $\mathbf{R}_A$ and $\mathbf{R}_B$ in Table 1.1.6 have both been constructed through row, column or level permutations of $\mathbf{R}_1$ from Table 1.1.4.

Table 1.1.6: Two designs isomorphic to $\mathbf{R}_1$ from Table 1.1.4

$\mathbf{R}_1$			$\mathbf{R}_A^\dagger$			$\mathbf{R}_B^\dagger$		
0	0	0	0	0	0	2	0	0
0	0	1	0	0	0	2	0	0
0	1	0	0	0	1	2	1	0
0	1	2	0	0	1	2	1	0
0	2	1	0	1	0	2	0	1
0	2	2	0	1	0	2	0	1
1	0	0	0	1	2	2	2	1
1	0	2	0	1	2	2	2	1
1	1	1	0	2	1	2	1	2
1	1	2	0	2	1	2	1	2
1	2	0	2	0	0	1	0	0
1	2	1	2	0	0	1	0	0
2	0	1	1	0	1	0	1	0
2	0	2	1	0	1	0	1	0

†*Note: we avoided reordering the rows lexicographically for ease of comparison*

To obtain $\mathbf{R}_A$ we performed the level mapping $\{0, 1, 2\} \to \{0, 2, 1\}$ in the first factor. This is not a geometric level mapping as the order of the levels has neither been maintained nor reversed, so this design is combinatorially isomorphic but not geometrically isomorphic to $\mathbf{R}_1$. To obtain $\mathbf{R}_B$ we performed the level mapping $\{0, 1, 2\} \to \{2, 1, 0\}$ on the first factor, and switched the second and third columns. Both of these actions are consistent with geometric isomorphism, hence this design is both combinatorially and geometrically isomorphic to $\mathbf{R}_1$. □

As we shall see in Section 1.3.2, the distinction between combinatorial and geometric isomorphism becomes important when we use optimality criteria to compare designs.

1.2 Notation and Useful Results

Now that we have defined the key concepts at the core of this thesis, we will define some notation and useful results which we will require in later chapters.

1.2.1 Properties of OAs

As noted previously, the statistical properties of OAs make them a desirable basis for FFDs. The following Lemma outlines one of the most useful properties which we will exploit in Chapter 2.

Lemma 1.2.1. Let $\mathbf{X}$ be the $N \times \alpha$ main effects only model matrix associated with an OA$[N, s_1^{m_1} \times s_2^{m_2} \times \cdots \times s_k^{m_k}]$, constructed as per Definition 1.1.3. The information matrix, $\mathbf{M} = \mathbf{X}'\mathbf{X}$, is equal to $N\mathbf{I}_\alpha$.

Proof. Consider the first column of $\mathbf{X}$ which is $\mathbf{1}_N$. The entry in the first row and column of $\mathbf{X}'\mathbf{X}$ will be $\mathbf{1}_N'\mathbf{1}_N = N$, and all remaining entries in the first row and column of $\mathbf{X}'\mathbf{X}$ will be zero since $\mathbf{P}_s\mathbf{1}_s = \mathbf{0}_{s-1}$ for all factors. Now consider factor F_a at s_a levels. Let $\mathbf{X}_a$ be the $(s_a - 1)$ columns of $\mathbf{X}$ corresponding to the main effects of F_a. Since each level of F_a will appear exactly N/s_a times in the OA, $\mathbf{X}_a'\mathbf{X}_a = \frac{N}{s_a}\mathbf{P}_{s_a}\mathbf{P}_{s_a}' = \frac{N}{s_a}s_a\mathbf{I}_{s_a-1} = N\mathbf{I}_{s_a-1}$. Hence, after the first row/column, the diagonal of $\mathbf{X}'\mathbf{X}$ can be considered as blocks of $N\mathbf{I}_{s_j-1}$ sub-matrices, $1 \leq j \leq m$. Now consider the h_ath column corresponding to the main effects of factor F_a at s_a levels, and the h_bth column corresponding to the main effects of factor F_b at s_b levels. Let $\mathbf{p}_{sh}$ be the hth row of $\mathbf{P}_s$, and note that $\mathbf{P}_s\mathbf{1}_s = \mathbf{0}_{s-1}$ implies that $\mathbf{p}_{sh} \cdot \mathbf{1}_s = 0$, $1 \leq h \leq s-1$. Let p_{sh_i} be the ith entry in $\mathbf{p}_{sh}$. Since each pair of level combinations will appear exactly $N/(s_a s_b)$ times in the OA, we have

$$\begin{aligned}
\mathbf{f}_{ah_a}'\mathbf{f}_{bh_b} &= N/(s_a s_b)\mathbf{p}_{s_a h_a}\mathbf{p}_{s_b h_b}' \\
&= N/(s_a s_b)\sum_{i=0}^{s_a-1}\sum_{j=0}^{s_b-1} p_{s_a h_{a i}} p_{s_b h_{b j}} \\
&= N/(s_a s_b)\sum_{i=0}^{s_a-1} p_{s_a h_{a i}} \sum_{j=0}^{s_b-1} p_{s_b h_{b j}} \\
&= N/(s_a s_b)\sum_{i=0}^{s_a-1} p_{s_a h_{a i}} \left(\mathbf{p}_{s_b h_b} \cdot \mathbf{1}_{s_b}\right) \\
&= 0.
\end{aligned}$$

Hence every off-diagonal entry in $\mathbf{M}$ equals 0 and all diagonal entries equal N, that is, $\mathbf{M} = N\mathbf{I}_\alpha$. □

Example 1.2.1. The 18×8 matrix, $\mathbf{X}$, constructed in Example 1.1.4 is the main effects only model matrix for an OA$[18, 3^3 \times 2^1]$, hence the information matrix $\mathbf{M} = \mathbf{X}'\mathbf{X}$, is equal to $18\mathbf{I}_8$. □

Recall that D-optimality is equivalent to minimising the determinant of $\mathbf{C} = \mathbf{M}^{-1}$. That is, a design is said to be D-optimal if it has the smallest $|\mathbf{M}^{-1}|$ amongst

the class of competing designs. Of course, minimising $|\mathbf{M}^{-1}|$ is equivalent to maximising $|\mathbf{M}|$, so in order to avoid taking the inverse, we will redefine a D-optimal design to be the design with the largest $|\mathbf{M}|$ amongst the class of competing designs. Hence, following Lemma 1.2.1, any two OA$[N, s_1^{m_1} \times s_2^{m_2} \times \cdots \times s_k^{m_k}]$s are equally good under a main effects only model since $|\mathbf{M}| = |N\mathbf{I}_\alpha| = N^\alpha$ for both designs. We note that Lemma 1.2.1 does not extend to models that contain interaction effects, as we shall see in Chapter 4.

Another useful property of OAs which will form the basis of Chapter 2 is that, assuming a main effects only model, the off-diagonal entries in the matrix $\mathbf{XX}'$ depend only on the *Hamming distance* between the relevant pairs of runs in the OA, as we will see in Lemma 1.2.2 below.

Definition 1.2.1. The Hamming distance between two vectors of equal length is the number of positions in which the corresponding entries are different. □

We will use the notation $\mathbf{r}_i$ to represent the ith row of the design matrix $\mathbf{R}$, that is, the m-tuple of the factor-level combination of the ith run.

Lemma 1.2.2. Let $\mathbf{X}$ be the $\left(\prod_{i=1}^{k} s_i^{m_i}\right) \times \alpha$ the main effects only model matrix associated with the complete factorial of a design with m_i factors each at s_i levels, and let d_i be the Hamming distance between rows $\mathbf{r}_x$ and $\mathbf{r}_y$ of the complete factorial for the m_i factors at s_i levels, $1 \leq i \leq k$. Then the entry in the xth row and yth column of the matrix $\mathbf{XX}'$ is

$$\begin{aligned} 1 + \sum_{i=1}^{k}[(s_i - 1)(m_i - d_i) + (-1)d_i] &= 1 + \sum_{i=1}^{k}[(s_i - 1)m_i - s_i d_i] \\ &= \alpha - \sum_{i=1}^{k} s_i d_i. \end{aligned}$$

Proof. Let $\mathbf{P}_s$ be the contrast matrix of order $(s-1) \times s$ as per Definition 1.1.2. Consider $\mathbf{J}_s = \frac{1}{\sqrt{s}}\mathbf{P}_s$. Then $\mathbf{J}_s\mathbf{J}_s' = \frac{1}{\sqrt{s}}\mathbf{P}_s\frac{1}{\sqrt{s}}\mathbf{P}_s' = \frac{1}{s}\mathbf{P}_s\mathbf{P}_s' = \mathbf{I}_{s-1}$. So, $\mathbf{J}_s$ is an orthogonal matrix and $\mathbf{J}_s^{-1} = \mathbf{J}_s'$. Thus, $\mathbf{J}_s'\mathbf{J}_s = \mathbf{I}_{s-1}$, that is $\frac{1}{\sqrt{s}}\mathbf{P}_s'\frac{1}{\sqrt{s}}\mathbf{P}_s = \frac{1}{s}\mathbf{P}_s'\mathbf{P}_s = \mathbf{I}_{s-1}$. Hence,

$$P_s(a) \cdot P_s(b) = \begin{cases} s-1 & \text{if } a = b \\ -1 & \text{if } a \neq b \end{cases}$$

for $a, b \in \{0, 1, \ldots, s-1\}$ where $P_s(0), \ldots, P_s(s-1)$ denote the columns of $\mathbf{P}_s$. Thus, the xth row and yth column of the matrix $\mathbf{XX}'$ can be viewed as the sum of three parts: 1 for the intercept, $s_i - 1$ for each of the $m_i - d_i$ factors that are the same between $\mathbf{r}_x$ and $\mathbf{r}_y$ in the m_i factors at s_i levels, and -1 for each of the d_i

factors that are different between $\mathbf{r}_x$ and $\mathbf{r}_y$ in the m_i factors at s_i levels. Hence, $1 + \sum_{i=1}^{k}[(s_i - 1)(m_i - d_i) + (-1)d_i]$. □

We illustrate this result in the example below.

Example 1.2.2. Consider the first three runs of the lexicographically ordered OA$[18, 3^3 \times 2^1]$ given in Table 1.1.2, and reproduced below.

$$\begin{array}{c} \\ \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{array} \begin{array}{cccc} \overbrace{}^{s_1=3} & & & \overbrace{}^{s_2=2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{array}$$

The main effects only model matrix for these runs will be the first three rows of the model matrix constructed in Example 1.1.4, given below.

$$\mathbf{X} = \begin{bmatrix} 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -1 \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 1 \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 1 \end{bmatrix}$$

Hence

$$\mathbf{X}\mathbf{X}' = \begin{bmatrix} 8 & 3 & 3 \\ 3 & 8 & 2 \\ 3 & 2 & 8 \end{bmatrix}.$$

Rather than calculating $\mathbf{X}\mathbf{X}'$ directly, we can also calculate the entries in this matrix using Lemma 1.2.2. The entry in the ith row and jth column corresponds to the Hamming distance between the ith and jth pair of runs. First, consider the diagonal entries. The Hamming distance of any run with itself is always zero, hence the diagonal entries are always α. Now consider the off-diagonal entries. The set of Hamming distances, (d_1, d_2), between $\mathbf{r}_1$ and $\mathbf{r}_2$ is $(1, 1)$. Hence, the entry in the 1st row and 2nd column of $\mathbf{X}\mathbf{X}'$ is $\alpha - \sum_{i=1}^{k} s_i d_i = 8 - (3 \times 1 + 2 \times 1) = 3$. The Hamming distance between $\mathbf{r}_1$ and $\mathbf{r}_3$ is also $(1, 1)$. The Hamming distance between $\mathbf{r}_2$ and $\mathbf{r}_3$ is $(2, 0)$, hence the relevant entry in $\mathbf{X}\mathbf{X}'$ is $8-(3\times2+2\times0) = 2$. Of course, the Hamming distance between the ith and jth run is the same as the Hamming distance between the jth and ith run, hence $\mathbf{X}\mathbf{X}'$ is symmetric. □

1.2.2 Matrix Algebra

For many of the results in subsequent chapters, we will rely on standard results in matrix algebra concerning the determinant of particular types of matrices and the

inverse of partitioned matrices. We cite the key results from Harville (1997) here which we will use in later chapters.

Theorem 1.2.1. *[Theorem 18.1.1 from Harville (1997)]*
Let $\mathbf{G}$ represent an $n \times n$ matrix, $\mathbf{S}$ an $n \times m$ matrix, $\mathbf{H}$ an $m \times m$ matrix, and $\mathbf{L}$ an $m \times n$ matrix. If $\mathbf{G}$ and $\mathbf{H}$ are nonsingular, then

$$|\mathbf{G} + \mathbf{SHL}| = |\mathbf{G}||\mathbf{H}||\mathbf{T}^{-1} + \mathbf{LG}^{-1}\mathbf{S}|. \qquad \square$$

Theorem 1.2.2. *[Theorem 8.5.11 from Harville (1997)]*
Let $\mathbf{T}$ represent an $m \times m$ matrix, $\mathbf{U}$ an $m \times n$ matrix, $\mathbf{V}$ an $n \times m$ matrix, and $\mathbf{W}$ an $n \times n$ matrix. Suppose that $\mathbf{T}$ is nonsingular. Then, $\begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$ is nonsingular if and only if the $n \times n$ matrix

$$\mathbf{Q} = \mathbf{W} - \mathbf{VT}^{-1}\mathbf{U}$$

is nonsingular, in which case

$$\begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{T}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{T}^{-1}\mathbf{U} \\ \mathbf{I}_n \end{bmatrix} \mathbf{Q}^{-1} \begin{bmatrix} -\mathbf{VT}^{-1}, & \mathbf{I}_n \end{bmatrix}. \qquad \square$$

1.3 Literature Review

Much of the early research into the statistical properties of FFDs focused on binary factors. Since the distinction between geometric and combinatorial isomorphism is meaningless when the factors have only two levels, geometric isomorphism is a relatively recent concept. While investigations into designs with three-level factors have expanded over the past decade, many authors still choose to focus on the combinatorial properties of these designs, leaving research into the geometric properties as relatively fertile ground. For example, the hole punching experiment described above, which involves three quantitative factors, was taken from Schoen (2009). In that paper the orthogonal arrays in 18 runs are compared using a criterion developed by Xu and Wu (2001) that assumes qualitative factors, instead of Cheng and Ye (2004)'s geometric version. As noted in Schoen (2009), "the arrays developed in the present study permit a more easy calculation of optimum designs for quantitative factors under any criterion, because there are just 3^a-level permutations to consider once the combinatorially non-isomorphic designs are given"; so from the perspective of our research proposal, the literature review should not be narrowed to studies based on geometric isomorphism alone. Care needs to be

taken, however, to check the assumptions underlying the type of factors being studied.

In this section we will summarise some of the many optimality criteria that have been proposed for FFDs and the structural properties that have a bearing on the performance of FFDs for model estimation.

1.3.1 Optimality Criteria

Cheng (2013) stresses that "we cannot overemphasize that no single criterion works in all situations". With that in mind, we now investigate criteria that have been developed for a broad range of objectives.

Aberration-type Criteria

Minimum aberration is an optimality criterion based on the assumption of effect hierarchy. It was developed to minimise the aliasing amongst the lower-order factors, with the goal of finding good designs for a number of models, such designs being termed *model-robust* designs. The early development of this criterion is not directly relevant to our work in this thesis as the assumption that effects of the same order are equally important assumes that binary or qualitative factors are being used (when factors are quantitative we may expect the linear component to be more important than the quadratic). However, some authors have recently developed geometric versions of this criterion and this is important to our research.

Minimum aberration is based on the *wordlength pattern*, initially developed by Fries and Hunter (1980) for *regular* designs with binary factors. Tang and Deng (1999) extended this work to *non-regular* designs with the *minimum G_2-aberration* criterion. This work was extended to factors with more than two levels by Xu and Wu (2001) who proposed the *generalized wordlength pattern* (GWLP) and *generalised minimum aberration* criterion.

Cheng and Ye (2004) noted that "an immediate problem is that this aberration criterion completely fails to distinguish and rank combinatorially isomorphic but geometrically non-isomorphic designs". They developed what they called a β wordlength pattern in order to allow for quantitative factors. This paper was seminal in the research into the geometric properties of designs and is discussed further in Section 1.3.2.

These aberration-type criteria are very popular throughout the literature and are often used as benchmarks for development of other criteria. However, the original minimum aberration criterion was proposed for a "situation in which prior knowl-

edge is diffuse concerning the possible greater importance of certain effects" (Fries and Hunter (1980)), and so it is not necessarily the best option in all situations.

Alphabet Criteria

Most of the criteria that are known as the 'alphabet criteria' (as distinct from the aberration-type criteria) are based on the A-, D- and E- criteria defined in Section 1.1.3. These criteria have been developed to address the experimenter's desire to increase the precision of the parameter estimates and have been widely used and modified throughout the literature.

The calculation of the alphabet criteria requires that we 'know' (or have assumed) which terms are to be included in the model. In situations where the experimenter may not have well defined prior information and therefore would like to find model-robust designs, Cheng et al. (1999) establish the concept of *estimation capacity* (EC) in the context of two-level factors. EC is defined as the proportion of models within the model space which are estimable. Of course, the size of this 'model space' will depend on the assumptions made about effect hierarchy, effect sparsity, and effect heredity. The authors compare this measure of model-robustness with the minimum aberration criterion and conclude that minimum aberration is a good surrogate for EC. Li and Nachtsheim (2000) define a second measure of model-robustness, the *information capacity* (IC), as the average D-efficiency over all models. They also develop what they call *model-robust factorial designs* (MRFDs) by selecting designs with maximum IC from those with EC = 100%.

We will focus on D-optimality in Chapter 2 for main effects only models. In Chapters 4 and 5, when we examine models that contain a subset of the linear-by-linear interactions, we will also assess model-robustness in a very similar manner to MRFDs.

1.3.2 Structural Properties

While design optimality can capture certain design performance features, structural properties of the designs can also be important in determining the model-fitting capacity of the designs.

In this section we will focus on the structural papers that allow FFDs to be good screening designs, or good search designs. We also discuss different definitions of isomorphism and design augmentation.

Screening and Projection

When the experimenter has many factors to consider but is not sure which ones will be *active*, (that is, non-negligible in the model) it has been common practice to run a two-stage analysis. In the first stage, a *screening design* is used to screen out unimportant factors. Experimenters typically choose an OA for the initial-stage design to ensure that all main effects are orthogonal to one another. This guarantees that the main effects of all subsets of factors are estimable. Once the active factors have been determined, a second-stage experiment is run in order to allow for second-order terms in the model. This two-stage procedure is limited by the implicit assumption in the initial-stage design that all higher-order terms are negligible. Thus, active two-factor interactions may be aliased with the main effects and have a negative impact on which factors are deemed to be active.

Cheng and Wu (2001) proposed a procedure to combine these two stages into a single design, that is, the single-stage analysis would both screen out unimportant factors and estimate higher-order terms. Designs that are able to identify any set of active factors of the same size equally well are said to have good *projective properties*. Although Cheng and Wu (2001) were the first to look at projection in the context of a single-stage analysis, the concept of projection had been well researched before then (Lin and Draper (1992), Lin and Draper (1993), Wang and Wu (1995), Cheng (1998), Box and Tyssedal (1996)).

A *regular design* in one in which all pairs of terms are either orthogonal or fully aliased. *Non-regular designs* have various degrees of partial aliasing amongst the pairs of terms in the model and this leads to a much more complex aliasing structure. These complex structures can often lead to superior projection properties however (Cheng (2006)). Wang and Wu (1995) coined the term *hidden projection property* to describe a design that "allows some (or all) interactions to be estimated even when the projected design does not have the right resolution or other combinatorial design property for the same interactions to be estimated"; that is, a design with favourable complex aliasing.

Other authors have looked at issues similar to screening and projections. For example, Cheng et al. (1999) introduced the concept of *suspect* two-factor interactions, being those that are aliased with a large contrast and are therefore potentially active.

Search Designs

Srivastava (1975) was concerned with bias caused by fitting a model in which some

higher-order interactions were incorrectly assumed to be negligible. He developed the concepts of *search linear models* and *search designs* to investigate designs that allow for up to g unknown higher-order terms to be estimated. We will use the following expression for Srivastava's search linear model (Morgan et al. (2014)):

$$E(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,$$

where $\mathbf{Y}$ is a vector of n response variables, $\mathbf{X}_1$ $(n \times p_1)$ and $\mathbf{X}_2$ $(n \times p_2)$ are known model matrices, $\boldsymbol{\beta}_1$ is a vector of p_1 fixed unknown parameters (effects) of primary interest, and it is believed that at most g elements of $\boldsymbol{\beta}_2$ are non-zero but the identity of these is unknown.

A design that is able to estimate the non-zero elements of $\boldsymbol{\beta}_2$ alongside the elements of $\boldsymbol{\beta}_1$ is called a search design with *resolving power* $\{\boldsymbol{\beta}_1; \boldsymbol{\beta}_2; g\}$.

Sirvastava's early work focused on 2^k factorial experiments where $\boldsymbol{\beta}_1$ contains the overall mean, the k main effects, and the $\binom{k}{2}$ two-factor interactions, while $\boldsymbol{\beta}_2$ contains all higher-order interactions. Note that many of these models will violate effect hierarchy.

In later work he considered 2^k factorial experiments where $\boldsymbol{\beta}_1$ contains only the overall mean and the k main effects, while $\boldsymbol{\beta}_2$ contains *all* interactions. Srivastava (1975) called these the *main effect plus g* plans, also known as MPE.g plans.

Anderson and Thomas (1980) examined search designs for l^k factors where l is a prime or prime power. Chatterjee and Mukerjee (1986) extended this work by removing the restriction that l be a prime power. Ghosh and Burns (2002) looked at search designs for factor screening experiments with k factors each at three levels. Ghosh and Flores (2013) considered 2^k designs and introduced a new property that used the concept of common variance. By using OAs, Angelopoulos et al. (2013) constructed search designs containing, apart from the general mean and the main effects, up to five non-negligible two-factor interaction effects. Sarkar and Chatterjee (2010) discussed search designs for three-level factors, main effects with up to two 'hidden' two-factor interactions. They noted that although a lot of work has been done for two-level factors, not much has been done on three-level factors.

In addition to the research that considers screening designs specifically, many authors have published on very similar concepts. For example, Li and Nachtsheim (2000) "develop a class of balanced designs that can be used for estimation of main effects and any combination of up to g interactions, where g is specified by the user". Lekivetz and Tang (2014) consider a requirement set, a set which contains all main effects and a subset of two-factor interactions. They look for designs that

allow this set of effects to be estimated while also minimising contamination by effects outside of the set.

A direct investigation of the above-mentioned structural properties is outside the scope of this thesis. Since many of of the themes throughout the literature in this area are very similar to the issues we do discuss, we will revisit these properties in Chapter 5 to illustrate how the various concepts discussed above tie in with our research.

Isomorphism

As mentioned earlier, initial research of design optimality commonly assumed binary factors and so the distinction between geometric and combinatorial isomorphism was not needed. The 'discovery' of geometric isomorphism was made gradually as many authors began to notice that designs in the same 'class' (that is, the same combinatorial class) had different model-fitting properties. Tsai et al. (2000) use the term *design family* to describe a geometric isomorphism class, but they do not discuss isomorphism (or use the term geometric) per se. Clark and Dean (2001) discuss the early investigation into design equivalence such as the discovery of the necessary condition that both designs must have the same GWLP. They discuss the fact that the permissible re-ordering of factors levels are different for quantitative and qualitative factors. Cheng and Wu (2001) use the term *model isomorphism* to address the fact that two combinatorially equivalent designs may not be geometrically isomorphic when $l > 2$. Cheng and Ye (2004) are the first to define geometric isomorphism specifically.

Since then, the literature on methods for determining geometric non-equivalence and the enumeration of geometric isomorphism classes has grown. Clark and Dean (2001) presented a sufficient condition for determining geometric non-equivalence based on the Euclidean distance between pairs of design points. Katsaounis et al. (2007) modified this method by introducing "split weights" of the rows of a design matrix. In an unpublished manuscript, Tsai et al. (2006a) were interested in enumeration, and they claim to have enumerated all geometrically non-isomorphic 18-run three-level OAs using an algorithm for checking equivalence based on the indicator function of Cheng and Ye (2004). Pang and Liu (2011) also claim to have carried out a complete enumeration of all OA$[18, 3^m]$s using a similar indicator function, although their counts differ from those in Tsai et al. (2006a). We resolve this discrepancy in Chapter 3 by using a different method of enumeration and showing that it covers the entire design space.

With regards to combinatorial isomorphism, Schoen (2009) enumerated all combi-

natorially inequivalent orthogonal arrays with 18 runs and eight or fewer factors, and used a classification method based on GWLP of the original arrays and those of their projections into fewer factors. Schoen et al. (2010) and Schoen and Nguyen (2007) give an algorithm for complete combinatorial enumeration of pure-level and mixed level orthogonal arrays of given strength t, and run-size N. Angelopoulos et al. (2007) give an algorithm for the construction and identification of non-isomorphic two-level OAs.

1.3.3 Adjoined or Unavailable Runs

So far, our literature review has broadly summarised the general concepts surrounding the design and comparison of FFDs. In this section we will focus on a more specific area of the literature which is closer to the heart of this project, that is, the performance of OAs when runs are adjoined or unavailable.

Adjoined Runs

When adding one additional run, the most general result was obtained by Mukerjee (1999) who gave a condition that must be satisfied to obtain an OA augmented with one additional run which is optimal, in the set of all fractional factorial designs with that number of runs, with respect to every *generalised Type 1 criterion* − a general criterion of which D-optimality is a subset. Chai et al. (2002) have investigated the construction of orthogonal array plus one run designs when the conditions identified by Mukerjee (1999) are not satisfied. As a result, optimal resolution 3 orthogonal array plus one run designs with up to 100 runs in the original OA have been identified, except perhaps for some with 72 runs.

Other authors have investigated the form of optimal orthogonal array plus p run designs when $p > 1$. Hedayat and Zhu (2003) determined the best sets of runs to adjoin to saturated D-optimal binary orthogonal arrays of strength 2 for the estimation of the main effects only model. Tsai and Liao (2011) developed sufficient conditions for a partially replicated parallel-flats design, for a mix of binary and ternary factors, to be D-, A- or E-optimal. They considered the optimal regular designs for a specified set of model terms, which may be more extensive than a main effects only model. Chatzopoulos et al. (2011) identified some sufficient conditions to describe a set of p runs which, when adjoined to a symmetric orthogonal array, would result in a Type 1 optimal orthogonal array plus p run design for a main effects only model. Other results on binary factors are available in the literature. For example, Butler and Ramos (2007) (augmenting a resolution 5 design),

Tsai et al. (2012) (deleting factors from a Hadamard matrix and adjoining runs to the resulting binary OA), and Tsai and Liao (2014) (using extended minimum aberration as the design comparison criterion).

This thesis contributes to the literature in this area by giving general constructions for D-optimal designs, under a main effects only model, constructed from an OA with p adjoined runs. Most of this work, which appears in Chapter 2, has been published in Bird and Street (2016). We then extend this work to consider models that contain one or more linear-by-linear interactions in Chapters 4 and 5.

Unavailable Runs

There are two distinct situations in which a subset of runs in a design may be deemed 'unavailable'. Our primary interest is the case in which the resources available for an experiment are slightly less than what would be required to perform all the runs in an OA, in which case we consider 'unavailable runs' to refer to the runs that we intentionally remove from the OA in order to meet cost/practical constraints before the experiment is run. We are not aware of any papers that address the issue of intentionally removed runs.

The other situation in which runs may be deemed unavailable is when a subset of runs are missing or suspicious in some sense (outliers) for unforeseen circumstances. That is, although we initially have the resources for an experiment with N runs, we 'lose' t observations through the course of the experiment and are left with only $N - t$ units to analyse. As noted by Akhtar and Prescott (1986), "The risk of losing observations usually cannot be ignored in practice. When small response surface designs are used, the effects of missing points may be substantial. The ability to estimate all parameters could be completely lost, or the variances of predicted responses could be incredibly large in a certain part of an experimental region. With respect to design optimality, designs will usually no longer be optimal when missing values exist". We are not aware of any research into the issues of missing runs in the context of OAs specifically, and so we summarise the literature on missing runs in the context of FFDs more broadly below.

A design which performs well in the absence of one or more runs is said to be *robust* to missing runs. Herzberg (1982) gives Box (1953) the credit for the introduction of the term robust generally into statistics. In the context of designs she says that robustness can focus either on investigating how well a design optimal under one criterion performs under another, or on the construction of designs which guard against particular short-comings, say the consequences of missing runs when fitting a model. It is this second setting that we will concentrate on in this thesis.

Ghosh (1982) developed four equivalent conditions, all based on properties of the model matrix, which allow one to determine whether or not a design would be robust to the loss of any t runs. Let $t_{\max}$ denote the largest t such that any t runs can be missing but the model is still estimable whereas there is at least one set of $t+1$ runs for which the model is not estimable. The ideas in Ghosh (1982) are extended in MacEachern et al. (1995) to give an upper bound for $t_{\max}$ and this bound is easily calculated from the model matrix.

Tanco et al. (2013) considered the robustness performance for estimating the full second-order model when runs are dropped from the 9 designs that they have decided to focus on. These designs have been specifically developed for estimating the second-order model and vary greatly in terms of robustness (summarised in their Table 2) and also in terms of the number of runs required (for example, when $m = 3$, n varies from 14 to 18; when $m = 7$, n varies from 40 to 82).

Akhtar and Prescott (1986) developed a *minimax loss* criterion, based on D-efficiency, in the context of *central composite designs*. This work was adapted and extended by Ahmad and Gilmour (2010) in the context of *subset designs*. Arshad and Gilmour (2012) construct *augmented pairs minimax loss designs* from *Plackett-Burman plans* where the design construction is based on the minimax loss criterion given by Akhtar and Prescott (1986).

This thesis will contribute to the literature in this area by assessing designs constructed from OAs with t runs removed in the same way we assess designs constructed from OAs with p runs adjoined, as discussed above. That is, Chapter 2 considers D-optimal designs under a main effects only model, and Chapters 4 and 5 consider models that include a subset of the linear-by-linear interactions.

1.4 Research Questions

This project will investigate the properties of OAs with adjoined or unavailable runs. The primary motivation for this research arises from situations when an experimenter has the resources for slightly more (or slightly fewer) runs then an OA, in which case it is tempting to use an OA with 'a few' runs adjoined (or removed). For example, the resources available for the hole punching experiment allowed for up to 20 runs, and so a design based on an OA$[18, 3^3 \times 2^1]$ was used to exploit the desirable statistical properties of orthogonal arrays. Thus, the experimenter chose to forgo any benefit that might have been gained by fitting an additional 2 runs. This thesis considers the question: what are the best runs to adjoin or remove when the number of runs available for an experiment is only slightly larger

or smaller than the number in an OA?

Of course, as discussed above, an additional motivation for investigating the consequences of unavailable runs is the possibility that a random subset of runs may be missing, rather than being intentionally removed. Hence, we will use the terminology 'missing', 'removed' and 'unavailable' interchangeably throughout this thesis and we will also discuss the implications of the distinction between these terms along the way. We will often use the terminology 'altered deign' to refer to an OA that has been altered by either adjoining or removing runs.

The measure of goodness we have chosen to focus on in this thesis is D-optimality since it is so widely accepted in the literature. Additionally, following work such as Li and Nachtsheim (2000) and Cheng et al. (1999), we will also consider various measures of model-robustness − a design's capacity to estimate multiple models. This work contributes to the literature on orthogonal arrays by identifying the best altered designs for various circumstances.

In Chapter 2 we assess the D-optimality of OAs with some runs adjoined or unavailable, assuming a main effects only model. As mentioned above, Chatzopoulos et al. (2011) investigated augmenting p runs to symmetric OAs (that is, designs with $k = 1$), but they did not consider missing runs. Hence, we extend this work by considering all OAs with $N \leq 100$ runs for any k, for both adjoined and missing runs. We show that the best designs depend on the Hamming distance of the runs adjoined or removed and we give some construction methods for finding optimal sets of runs. The work in Chapter 2 on OAs with adjoined runs has been published in Bird and Street (2016).

We extend these results in Chapters 4 and 5 to consider a model containing one or more of the linear-by-linear interactions. However, in order to do so, it transpired that we require a complete set of designs to study so that we can make design recommendations based on knowledge of the entire design space. Recall that there is some discrepancy in the literature regarding the complete enumeration of all geometrically non-isomorphic OA$[18, 3^m]$s. Both Tsai et al. (2006a) and Pang and Liu (2011) claim to have complete enumerations despite citing different counts, and neither set of authors is able to make available the actual OAs they constructed. Hence, in Chapter 3 of this thesis, we will enumerate all geometric isomorphism classes for symmetric OAs with ternary factors, and prove that these classes cover the entire OA$[18, 3^m]$s space. These designs appear in the electronic appendix.

In Chapter 4 we considering OAs with adjoined or unavailable runs in a model containing the linear-by-linear component of a single two-factor interaction. We start with some general algebraic results that can be applied to any OA, then we

carry out a systematic assessment of each of the geometric classes enumerated in Chapter 3 using D-optimality and measures of model-robustness.

Chapter 5 extends the work in Chapter 4 by considering models that contain any subset of ν linear-by-linear interactions. Again, we focus on the geometric classes enumerated in Chapter 3 and illustrate the various (and often competing) properties of the most notable classes as assessed via D-optimality and measure of model-robustness. We also compare this work to that on search designs and discuss the distinction between model-robustness and model-discrimination.

We close the thesis in Chapter 6 with a summary of the work in this project. We also discuss some additional concepts in area of optimal FFDs that are not directly related to this project, hence they have been omitted from the above literature review. We have chosen to hold these concepts until the final chapter as they would only protract rather than enhance our initial discussion of the literature. Hence, Chapter 6 finishes with a broad discussion of where the work in this thesis sits in relation to these other concepts, and potentially interesting avenues of future research that stem from these observations.

Chapter 2

Main Effects Only Models

As mentioned in Chapter 1, OAs are a class of FFDs that are optimal according to a range of optimality criteria. This makes it tempting to construct fractional factorial designs by adjoining additional runs to an OA, or by removing runs from an OA, when the number of runs available for the experiment is only slightly larger or smaller than the number in the OA. This chapter investigates optimising designs under these conditions assuming a main effects only model. We give conditions for the theoretical optimal designs, and develop constructions for finding the best realisable designs.

We begin by defining the class of competing designs in Section 2.1. In Section 2.2 we examine the information matrix of these designs and establish that the optimisation methods for adjoining runs to an OA are essentially the same as the optimisation methods for removing runs. Given this similarity, the constructions in subsequent sections focus largely on augmenting runs to an OA with the understanding that these methods can be applied in the context of missing runs too. Of course, an obvious difference between the two processes is that we are free to adjoin any run from the complete factorial to an OA, whereas the set of potential runs to remove is limited to those within the OA. Section 2.7 concludes the chapter with an examination of the implications of this restriction in the context of symmetric ternary designs in 18 runs.

Much of the work in this chapter has appeared in Bird and Street (2016). The notation in this chapter differs slightly to the notation in that paper, most notably, this thesis uses $\mathbf{X}$ for the model matrix rather than Q.

2.1 The Class of Competing Designs

2.1.1 Augmenting p Runs

Suppose we require a $s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}$ design in n runs, and an OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$ exists such that n is only slightly larger than N. Hence, we will adjoin $p = n - N$ runs to this OA, which we refer to as an *OA plus p runs* design. The class of completing designs are all designs that can be obtained by adjoining any p runs to the OA. As we shall see in Section 2.2.1, it does not matter which OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$ is used, hence the class of competing designs is not limited to p runs adjoined to one particular OA, rather it encompasses all OA plus p runs designs for any OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$. We note that there are $\prod_{i=1}^{k} s_i^{m_i}$ level combinations in the complete factorial, hence there are $\left(\prod_{i=1}^{k} s_i^{m_i}\right)^p$ potential sets of p runs to augment.

2.1.2 Missing t Runs

Now suppose that the number of runs in the design, n, is slightly smaller than N. Hence, we will remove $t = N - n$ runs from the OA, which we refer to as an *OA minus t runs* design. The class of completing designs in this setting are all designs that can be obtained by removing any t runs from the OA. As we shall see in Section 2.2.2, it does not matter, theoretically, which OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$ is used when runs are removed, as is the case when runs are augmented. Hence, the class of competing designs encompasses all OA minus t runs designs for any OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$. Of course when we remove runs, our selection is limited to those within the OA, hence there are $\binom{N}{t}$ potential sets of t runs to remove. Thus, although we can theoretically use any OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$, we might prefer some OAs over others if they provide a more desirable choice of runs to remove. We will explore this concept further in Section 2.7.

This highlights the important distinction between adjoining and removing runs, that is, we have far more choice when adjoining runs then when removing runs. Furthermore, as discusses in Chapter 1, we might not have any choice in which runs are removed. Suppose we have resources for N runs, but we anticipate up to t of these may fail, and we have no knowledge which specific runs these might be. We can think of this as an OA minus t runs design, although rather than intentionally removing t runs in the construction of the design, we discover that t runs are missing at the conclusion of the experiment. Hence, recall from Chapter 1 that we use the terminology 'remove', 'missing' and 'unavailable' interchangeably.

2.1.3 Efficiency

Recall that a design is said to be D-optimal if its information matrix has the largest determinant amongst the class of competing designs. Let $|\mathbf{M}|$ be the determinant of the information matrix for design D, and $|\mathbf{M}|_{opt}$ be the optimal determinant, that is, the largest determinant of the information matrix from all designs in the same class as D. Then following Atkinson et al. (2007), we define the efficiency of design D as:

$$\left(\frac{|\mathbf{M}|}{|\mathbf{M}|_{opt}}\right)^{1/\alpha} \tag{2.1.1}$$

where $\alpha = 1 + \sum_1^k (s_i - 1)m_i$ is the number of parameters in the main effects only model. Atkinson et al. (2007) note that taking the ratio of determinants to the $(1/\alpha)$ power "results in an efficiency measure which is proportional to design size, irrespective of the dimension of the model. So, two replicates of a design measure for which [the efficiency]=0.5 would be as efficient as one replicate of the optimum measure". Hence, this chapter is focused on finding designs that maximise $|\mathbf{M}|$ in order to bring the efficiency as close to 1 as possible.

2.2 Information Matrix for the Altered Design

In this section, we define the information matrix for OA plus p designs and OA minus t designs in tern, then we establish a bound on the determinants of these matrices.

2.2.1 Augmenting p Runs

Suppose that we adjoin p runs to an OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$. Then we will write the model matrix of the augmented design as $\mathbf{X}_A = \begin{bmatrix} \mathbf{X} \\ \mathbf{A} \end{bmatrix}$ where $\mathbf{A}$ is a $p \times \alpha$ matrix that contains the p row vectors containing entries for the design matrix corresponding to the p augmented runs. The information matrix is $\mathbf{M}_A = \mathbf{X}'\mathbf{X} + \mathbf{A}'\mathbf{A}$. Hence, following Lemma 1.2.1, $\mathbf{M}_A = N\mathbf{I}_\alpha + \mathbf{A}'\mathbf{A}$.

To determine the D-optimal design, we require an expression for the determinant of the information matrix, $|\mathbf{M}_A|$. We see that $\mathbf{M}_A = N\mathbf{I}_\alpha + \mathbf{A}'\mathbf{A}$ has the form $\mathbf{G} + \mathbf{SHL}$ if we let $\mathbf{G} = N\mathbf{I}_\alpha$, $\mathbf{S} = \mathbf{A}'$, $\mathbf{H} = \mathbf{I}_p$ and $\mathbf{L} = \mathbf{A}$. $\mathbf{G}$ and $\mathbf{H}$ are nonsingular, so we apply Theorem 1.2.1 and see that

$$\begin{aligned} |\mathbf{M}_A| &= |\mathbf{G}||\mathbf{H}||\mathbf{H}^{-1} + \mathbf{L}\mathbf{G}^{-1}\mathbf{S}| \\ &= |N\mathbf{I}_\alpha||\mathbf{I}_p||\mathbf{I}_p^{-1} + \mathbf{A}\left(N\mathbf{I}_\alpha\right)^{-1}\mathbf{A}'| \\ &= N^\alpha|\mathbf{I}_p + \mathbf{A}\mathbf{A}'/N| \\ &= N^{\alpha-p}|N\mathbf{I}_p + \mathbf{A}\mathbf{A}'| \end{aligned} \tag{2.2.1}$$

$N^{\alpha-p}$ is constant for all competing designs in the class. We will use $\mathbf{\Omega}_A$ to denote the matrix $N\mathbf{I}_p + \mathbf{A}\mathbf{A}'$ to be optimised. Notice that $\mathbf{\Omega}_A$ is a $p \times p$ matrix, which is likely smaller in dimensions than $\mathbf{M}_A$ ($\alpha \times \alpha$) as we typically expect $p < \alpha$. Furthermore, we note that $\mathbf{\Omega}_A$ depends only on the p runs to be augmented, not on the runs of the OA, hence the use of any OA is equally good and does not effect the optimisation process.

2.2.2 Missing t Runs

Suppose that rather than augmenting p runs we are missing t runs from an OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$. Then we can partition the runs of $\mathbf{X}$ as $\begin{bmatrix} \mathbf{X}_B \\ \mathbf{B} \end{bmatrix}$ where $\mathbf{B}$ is a $t \times \alpha$ matrix that contains the t row vectors associated with the missing runs. The information matrix for the OA is $\mathbf{X}'\mathbf{X} = \mathbf{X}_B'\mathbf{X}_B + \mathbf{B}'\mathbf{B}$. Hence, since $\mathbf{X}'\mathbf{X} = N\mathbf{I}_\alpha$ (Lemma 1.2.1), the information matrix for the design with t missing runs is $\mathbf{M}_B = \mathbf{X}_B'\mathbf{X}_B = N\mathbf{I}_\alpha - \mathbf{B}'\mathbf{B}$.

The structure of the determinant of $\mathbf{M}_B$ is very similar to that of $\mathbf{M}_A$ in equation (2.2.1). We let $\mathbf{G} = N\mathbf{I}_\alpha$, $\mathbf{S} = \mathbf{B}'$, $\mathbf{H} = -\mathbf{I}_t$ and $\mathbf{L} = \mathbf{B}$, so that $\mathbf{M}_B = \mathbf{G} + \mathbf{SHL}$. $\mathbf{G}$ and $\mathbf{H}$ are nonsingular, so we apply Theorem 1.2.1 and see that

$$\begin{aligned} |\mathbf{M}_B| &= |\mathbf{G}||\mathbf{H}||\mathbf{H}^{-1} + \mathbf{L}\mathbf{G}^{-1}\mathbf{S}| \\ &= |N\mathbf{I}_\alpha||-\mathbf{I}_t||(-\mathbf{I}_t)^{-1} + \mathbf{B}\left(N\mathbf{I}_\alpha\right)^{-1}\mathbf{B}'| \\ &= N^\alpha(-1)^t|-\mathbf{I}_t + \mathbf{B}\mathbf{B}'/N| \\ &= (-1)^t N^{\alpha-t}|\mathbf{B}\mathbf{B}' - N\mathbf{I}_t| \end{aligned} \tag{2.2.2}$$

We will use $\mathbf{\Omega}_B$ to denote $\mathbf{B}\mathbf{B}' - N\mathbf{I}_t$, the matrix to be optimised. We note that $\mathbf{\Omega}_B$ depends only on the t missing runs, so the use of any OA is equally good, as is the case when runs are augmented, however when runs are missing we are restricted to considering only a subset of the runs in the OA rather than the complete factorial.

2.2.3 A Bound on the Determinant

Augmenting p Runs

We will start by considering the bound on $|\mathbf{\Omega}_A|$ when p runs are augmented to an OA. Let $\mathbf{a}_x$ be the xth row in $\mathbf{A}$. We note that $\mathbf{a}_x.\mathbf{a}_x' = \alpha$ for all runs in the complete factorial, hence all diagonal entries of $\mathbf{\Omega}_A$ are $N+\alpha$, and the trace of $\mathbf{\Omega}_A$ is $(N+\alpha)p$. Since the trace of a matrix is equal to the sum of its eigenvalues, we have $\sum_{i=1}^{p} \lambda_i = (N+\alpha)p$, where $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ are the eigenvalues of $\mathbf{\Omega}_A$. From the arithmetic-geometric mean inequality see that:

$$\begin{aligned} \left(\prod_{i=1}^{p} \lambda_i\right)^{1/p} &\leq \frac{\sum_{i=1}^{p} \lambda_i}{p} \\ \left(\prod_{i=1}^{p} \lambda_i\right)^{1/p} &\leq \frac{(N+\alpha)p}{p} \\ \prod_{i=1}^{p} \lambda_i &\leq (N+\alpha)^p \end{aligned} \tag{2.2.3}$$

Since the determinant of a matrix is equal to the product of its eigenvalues, Equation (2.2.3) implies $|\mathbf{\Omega}_A| \leq (N+\alpha)^p$. The arithmetic-geometric mean also states that $\left(\prod_{i=1}^{p} \lambda_i\right)^{1/p} = \frac{\sum_{i=1}^{p} \lambda_i}{p}$ if and only if all λ_i are equal, hence the upper bound of $|\mathbf{\Omega}_A|$ is realised when all eigenvalues are the same. Thus, if every value of $\mathbf{a}_x.\mathbf{a}_y'$, $x \neq y$, from amongst the p runs to be adjoined is equal to 0, then the augmented design is D-optimal.

From Equation (2.1.1), we see that the efficiency of an OA plus p runs design is

$$\begin{aligned} \left(\frac{|\mathbf{M}_A|}{|\mathbf{M}_A|_{opt}}\right)^{1/\alpha} &= \left(\frac{N^{\alpha-p}|\mathbf{\Omega}_A|}{N^{\alpha-p}(N+\alpha)^p}\right)^{1/\alpha} \\ &= \left(\frac{|\mathbf{\Omega}_A|}{(N+\alpha)^p}\right)^{1/\alpha}. \end{aligned} \tag{2.2.4}$$

Missing t Runs

Similar arguments can be applied to the upper bound on $|\mathbf{\Omega}_B|$ when t runs are missing from an OA. The trace of $\mathbf{\Omega}_B$ is $(\alpha - N)t$, and so, following the same logic described above, $|\mathbf{\Omega}_B| \leq (\alpha - N)^t$. Let $\mathbf{b}_x$ be the xth row in $\mathbf{B}$, then this bound will be realised when every value of $\mathbf{b}_x.\mathbf{b}_y'$, $x \neq y$, from amongst the t missing runs is equal to 0.

From Equation (2.1.1), we see that the efficiency of an OA minus t runs design is

$$\left(\frac{|\mathbf{M}_B|}{|\mathbf{M}_B|_{opt}}\right)^{1/\alpha} = \left(\frac{(-1)^t N^{\alpha-t}|\mathbf{\Omega}_B|}{(-1)^t N^{\alpha-t}(\alpha-N)^t}\right)^{1/\alpha} = \left(\frac{|\mathbf{\Omega}_B|}{(\alpha-N)^t}\right)^{1/\alpha}. \tag{2.2.5}$$

In general, we expect the number of parameters in a main effects only model, α, to be less than the number of runs in the OA, N. However, if the model is saturated $\alpha = N$ and Equation (2.2.5) is undefined.

Altering an OA by One Run

Consider the class of OA plus 1 run designs. We have $\mathbf{\Omega}_A = N\mathbf{I}_1 + \mathbf{a}_1.\mathbf{a}_1' = N + \alpha$. Hence, adjoining any single run is equally good according to the D-optimality criterion as $|\mathbf{M}_A| = N^{\alpha-1}(N+\alpha)$ regardless of which run is appended.

Similarly, when $t = 1$ run is missing from an OA, hence $\mathbf{B}$ contains a single row, $\mathbf{b}_1$, then $\mathbf{\Omega}_B = \mathbf{b}_1.\mathbf{b}_1' - N\mathbf{I}_1 = \alpha - N$, hence $|\mathbf{M}_B| = -N^{\alpha-1}(\alpha-N)$ regardless of which run is missing.

Altering an OA by Two Runs

To find the D-optimal OA plus 2 run design, we need to maximise the determinant of the 2×2 matrix $\mathbf{\Omega}_A$ which can be written as

$$\begin{bmatrix} N+\alpha & \mathbf{a}_1.\mathbf{a}_2' \\ \mathbf{a}_1.\mathbf{a}_2' & N+\alpha \end{bmatrix}.$$

The determinant of this matrix is $(N+\alpha)^2 - (\mathbf{a}_1.\mathbf{a}_2')^2$, thus the D-optimal design is obtained by adding a pair of rows with the smallest realisable absolute value of $\mathbf{a}_1.\mathbf{a}_2'$.

To find the D-optimal OA minus 2 run design, we need to maximise the determinant of the 2×2 matrix $\mathbf{\Omega}_B$ which can be written as

$$\begin{bmatrix} \alpha-N & \mathbf{b}_1.\mathbf{b}_2' \\ \mathbf{b}_1.\mathbf{b}_2' & \alpha-N \end{bmatrix}.$$

The determinant of this matrix is $(\alpha-N)^2 - (\mathbf{b}_1.\mathbf{b}_2')^2$, thus the D-optimal design is obtained when the pair of runs with the smallest realisable absolute value of

$\mathbf{b}_1.\mathbf{b}_2'$ is missing.

Altering an OA by More Than Two Runs

First we will consider altering designs by augmenting p runs where $p > 2$. We will label the xth run to be appended $\mathbf{r}_x$, with corresponding row $\mathbf{a}_x$ in $\mathbf{A}$. Recall from Lemma 1.2.2 in Chapter 1 that we let d_i be the Hamming distance between $\mathbf{r}_x$ and $\mathbf{r}_y$ for the m_i factors with s_i levels, $1 \leq i \leq k$, and that

$$\mathbf{a}_x.\mathbf{a}_y' = \alpha - \sum_{i=1}^{k} s_i d_i. \tag{2.2.6}$$

We seek a set of p runs in which $\mathbf{a}_x.\mathbf{a}_y' = 0$ between all pairs, hence Equation (2.2.6) implies we are essentially searching for a set of runs with associated set of pairwise Hamming distances such that $\sum_{i=1}^{k} s_i d_i = \alpha$. If no such set of runs exist we need to determine the structure of the optimal sets of runs to adjoin. The following example illustrates these ideas.

Example 2.2.1. Consider an OA$[32, 2^5 \times 4^7]$. That is, $N = 32$, $s_1 = 2$, $s_2 = 4$, $m_1 = 5$, $m_2 = 7$ and $\alpha = 1 + 5(2-1) + 7(4-1) = 27$. Since $\mathbf{a}_x.\mathbf{a}_y' = 27 - 2d_1 - 4d_2$ there are no integer values of d_1 and d_2 such that $\mathbf{a}_x.\mathbf{a}_y' = 0$. We get values of 1 when $(d_1, d_2) = (1, 6)$, $(3, 5)$ or $(5, 4)$, and of -1 when $(d_1, d_2) = (0, 7)$, $(2, 6)$ or $(4, 5)$. For $p = 3$, the 3×3 matrix $\mathbf{\Omega}_A$ is

$$\begin{bmatrix} 59 & \mathbf{a}_1.\mathbf{a}_2' & \mathbf{a}_1.\mathbf{a}_3' \\ \mathbf{a}_1.\mathbf{a}_2' & 59 & \mathbf{a}_2.\mathbf{a}_3' \\ \mathbf{a}_1.\mathbf{a}_3' & \mathbf{a}_2.\mathbf{a}_3' & 59 \end{bmatrix}.$$

Given that we can not find pairs of runs with $\mathbf{a}_x.\mathbf{a}_y' = 0$, the largest possible determinant for this matrix is 205,204, which occurs when the product of the off-diagonal entries is 1. If the product of the off-diagonal entries is -1 then the determinant is 205,200. Exhaustive searching of the complete factorial shows that it is not possible to find sets of 3 runs for which the product of the off-diagonal entries is 1 but we can find at least one set of 3 runs for which the product of the off-diagonal entries is -1, as given below.

$$\begin{array}{cc} & \overbrace{}^{s_1=2}\overbrace{}^{s_2=4} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1\,0 \\ \mathbf{r}_3 & 1\,1\,1\,1\,1\,0\,1\,0\,2\,2\,2\,0 \end{array}$$

The Hamming distance between $\mathbf{r}_1$ and $\mathbf{r}_2$ is $(2, 6)$, between $\mathbf{r}_1$ and $\mathbf{r}_3$ is $(5, 4)$, and between $\mathbf{r}_2$ and $\mathbf{r}_3$ is $(3, 5)$. Hence, the associated $\mathbf{\Omega}_A$ is equal to

$$\begin{bmatrix} 59 & -1 & 1 \\ -1 & 59 & 1 \\ 1 & 1 & 59 \end{bmatrix}$$

with determinant 205,200. As the upper bound of $|\mathbf{\Omega}_A|$ is $59^3 = 205,379$, the best realisable design is $\left(\frac{205,200}{205,379}\right)^{1/27} = 99.997\%$ efficient relative to the theoretical bound. □

Now we will consider altering designs by removing t runs where $t > 2$. Let $\mathbf{b}_x$ be the row in $\mathbf{B}$ associated with the xth missing run, $\mathbf{r}_x$, then following Equation (2.2.6), we have $\mathbf{b}_x.\mathbf{b}'_y = \alpha - \sum_{i=1}^{k} s_i d_i$.

As we saw in the beginning of this section, the optimisation of $\mathbf{\Omega}_B$ is essentially the same as the optimisation of $\mathbf{\Omega}_A$, that is, we hope to achieve every off-diagonal entry equal to zero. In the case of augmenting runs, we searched for a set of p runs with associated pairwise Hamming distances that realise $\sum_{i=1}^{k} s_i d_i = \alpha$. Similarly, in the case of missing runs, we seek a set of t runs that realise $\sum_{i=1}^{k} s_i d_i = \alpha$. Note that $\mathbf{a}_x$ and $\mathbf{b}_x$ are effectively different notations for the same concept, that is, the row vector in the model matrix associated with the run $\mathbf{r}_x$. The only distinction between the two processes is that $\mathbf{a}_x$ can come from any run in the complete factorial when we append runs, whereas $\mathbf{b}_x$ can only be selected from runs within the OA, as the following example illustrates.

Example 2.2.2. Suppose we intend to remove $t = 3$ runs from the OA$[32, 2^5 \times 4^7]$ given in Table 2.2.1 below. For $t = 3$, the 3×3 matrix $\mathbf{\Omega}_B$ is

$$\begin{bmatrix} -5 & \mathbf{b}_1.\mathbf{b}'_2 & \mathbf{b}_1.\mathbf{b}'_3 \\ \mathbf{b}_1.\mathbf{b}'_2 & -5 & \mathbf{b}_2.\mathbf{b}'_3 \\ \mathbf{b}_1.\mathbf{b}'_3 & \mathbf{b}_2.\mathbf{b}'_3 & -5 \end{bmatrix}.$$

Following Example 2.2.1, there are no integer values of d_1 and d_2 such that $\mathbf{b}_x.\mathbf{b}'_y = 0$, and so we seek a set of 3 runs in which each pairwise Hamming distance is either $(d_1, d_2) = (1, 6)$, $(3, 5)$ or $(5, 4)$ so that $\mathbf{b}_x.\mathbf{b}'_y = 1$, or $(d_1, d_2) = (0, 7)$, $(2, 6)$ or $(4, 5)$ so that $\mathbf{b}_x.\mathbf{b}'_y = -1$. Note that the runs $\mathbf{r}_2 = 000111111110$ and $\mathbf{r}_3 = 111110102220$ from Example 2.2.1 do not appear in this OA, hence we can not use them as part of our solution. It can be shown exhaustively that there are no two runs in this OA that realise $(d_1, d_2) = (1, 6)$ or $(5, 4)$, but there are many other pairs of runs that realise the other Hamming distances we are searching for. An example of a set of 3 runs that will optimise $\mathbf{\Omega}_B$ is the 1st, 2nd and 7th rows of the OA. The

Table 2.2.1: OA$[32, 2^5 \times 4^7]$

1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	2	2	2	2	2	2	2
3	0	0	0	0	1	1	1	1	1	1	1	1
4	0	0	0	0	1	3	3	3	3	3	3	3
5	0	0	1	1	0	0	1	1	3	3	2	2
6	0	0	1	1	0	2	3	3	1	1	0	0
7	0	0	1	1	1	1	0	0	2	2	3	3
8	0	0	1	1	1	3	2	2	0	0	1	1
9	0	1	0	1	0	1	2	3	0	1	2	3
10	0	1	0	1	0	3	0	1	2	3	0	1
11	0	1	0	1	1	0	3	2	1	0	3	2
12	0	1	0	1	1	2	1	0	3	2	1	0
13	0	1	1	0	0	1	3	2	3	2	0	1
14	0	1	1	0	0	3	1	0	1	0	2	3
15	0	1	1	0	1	0	2	3	2	3	1	0
16	0	1	1	0	1	2	0	1	0	1	3	2
17	1	0	0	1	0	0	2	0	3	1	3	1
18	1	0	0	1	0	2	0	2	1	3	1	3
19	1	0	0	1	1	1	3	1	2	0	2	0
20	1	0	0	1	1	3	1	3	0	2	0	2
21	1	0	1	0	0	0	3	1	0	2	1	3
22	1	0	1	0	0	2	1	3	2	0	3	1
23	1	0	1	0	1	1	2	0	1	3	0	2
24	1	0	1	0	1	3	0	2	3	1	2	0
25	1	1	0	0	0	1	0	3	3	0	1	2
26	1	1	0	0	0	3	2	1	1	2	3	0
27	1	1	0	0	1	0	1	2	2	1	0	3
28	1	1	0	0	1	2	3	0	0	3	2	1
29	1	1	1	1	0	1	1	2	0	3	3	0
30	1	1	1	1	0	3	3	0	2	1	1	2
31	1	1	1	1	1	0	0	3	1	2	2	1
32	1	1	1	1	1	2	2	1	3	0	0	3

Hamming distance between rows 1 and 2 is (0,7), between rows 1 and 7 is (3,5) and between rows 3 and 7 is (3,5). The associated $\mathbf{\Omega}_B$ is equal to

$$\begin{bmatrix} -5 & -1 & 1 \\ -1 & -5 & 1 \\ 1 & 1 & -5 \end{bmatrix}$$

with determinant -112. As the upper bound of $|\mathbf{\Omega}_B|$ is $(-5)^3 = -125$, the best realisable design is $\left(\frac{-112}{-125}\right)^{1/27} = 99.594\%$ efficient relative to the theoretical bound.

□

Note that the efficiency of an OA plus p runs design involves the ratio of two very large numbers compared to the efficiency of the analogous OA minus t runs design, hence the efficiency when augmenting runs will always be better then when the same solution is applied to missing runs. For example, both of the previous examples involved altering an OA$[32, 2^5 \times 4^7]$, and in both cases we found solutions such that the off-diagonal entries in the upper triangle of the relevant matrix was $(-1, 1, 1)$, but the efficiency when runs were augmented was $\left(\frac{205,200}{205,379}\right)^{1/27} = 99.997\%$, whereas when runs were missing the efficiency was $\left(\frac{-112}{-125}\right)^{1/27} = 99.594\%$. As we shall see in Section 2.7, this disparity is exacerbated when the solution moves further away from the theoretical best.

The previous two examples highlight the challenge of both determining which values are realisable in the off-diagonal entries of $\mathbf{\Omega}_A$ or $\mathbf{\Omega}_B$, and subsequently which values to choose if it is not possible to realise all zeros. We will now make some general observations about the how to maximise $|\mathbf{\Omega}_A|$ (or $|\mathbf{\Omega}_B|$) when $p = 3$ (or $t = 3$) which can also be applied to larger values of p (or t).

In general, if $p = 3$, the determinant of $\mathbf{\Omega}_A$ is

$$(N+\alpha)^3 + 2(\mathbf{a}_1.\mathbf{a}_2')(\mathbf{a}_1.\mathbf{a}_3')(\mathbf{a}_2.\mathbf{a}_3') - (N+\alpha)\left((\mathbf{a}_1.\mathbf{a}_2')^2 + (\mathbf{a}_1.\mathbf{a}_3')^2 + (\mathbf{a}_2.\mathbf{a}_3')^2\right). \tag{2.2.7}$$

Similarly, if $t = 3$, the determinant of $\mathbf{\Omega}_B$ is

$$(\alpha-N)^3 + 2(\mathbf{b}_1.\mathbf{b}_2')(\mathbf{b}_1.\mathbf{b}_3')(\mathbf{b}_2.\mathbf{b}_3') - (\alpha-N)\left((\mathbf{b}_1.\mathbf{b}_2')^2 + (\mathbf{b}_1.\mathbf{b}_3')^2 + (\mathbf{b}_2.\mathbf{b}_3')^2\right). \tag{2.2.8}$$

If the smallest realisable value of $\mathbf{a}_x.\mathbf{a}_y'$ (or $\mathbf{b}_x.\mathbf{b}_y'$) is 0 but if we can not get 3 runs which pairwise have $\mathbf{a}_x.\mathbf{a}_y' = 0$ (or $\mathbf{b}_x.\mathbf{b}_y' = 0$), our empirical calculations suggest that it is better to aim to minimise the absolute value of the maximum entry, and then to have as many small entries as possible. The following example illustrates this idea.

Example 2.2.3. Consider an OA$[48, 2^1 \times 3^1 \times 4^2]$, hence $\alpha = 10$, and suppose we would like to adjoin $p = 3$ runs. The value of $\mathbf{a}_x.\mathbf{a}_y'$ for each plausible set of (d_1, d_2, d_3) is given in Table 2.2.2. The following pair of runs $\mathbf{r}_1$ and $\mathbf{r}_2$ have

Table 2.2.2: Each plausible set of (d_1, d_2, d_3) for an OA$[48, 2^1 \times 3^1 \times 4^2]$

(d_1, d_2, d_3)	$\mathbf{a}_x.\mathbf{a}_y'$	(d_1, d_2, d_3)	$\mathbf{a}_x.\mathbf{a}_y'$	(d_1, d_2, d_3)	$\mathbf{a}_x.\mathbf{a}_y'$	(d_1, d_2, d_3)	$\mathbf{a}_x.\mathbf{a}_y'$
(0, 0, 0)	10	(0, 1, 0)	7	(1, 0, 0)	8	(1, 1, 0)	5
(0, 0, 1)	6	(0, 1, 1)	3	(1, 0, 1)	4	(1, 1, 1)	1
(0, 0, 2)	2	(0, 1, 2)	−1	(1, 0, 2)	0	(1, 1, 2)	−3

Hamming distance $(d_1, d_2, d_3) = (1, 0, 2)$, hence $\mathbf{a}_1.\mathbf{a}_2' = 0$.

$$\begin{array}{cc|cc|cc} & \overbrace{}^{s_1=2} & \overbrace{}^{s_2=3} & & \overbrace{}^{s_3=4} & \\ \mathbf{r}_1 & 0 & 0 & 0 & 0 \\ \mathbf{r}_2 & 1 & 0 & 1 & 1 \end{array}$$

However, we are not able to construct a third run, $\mathbf{r}_3$, such that $\mathbf{a}_x.\mathbf{a}_y' = 0$ is maintained for all pairs of runs, and so we seek to minimise the absolute value of the maximum $\mathbf{a}_x.\mathbf{a}_y'$. The run $\mathbf{r}_3 = 1102$ has Hamming distance (d_1, d_2, d_3) = (1, 1, 1) and (0, 1, 2) from $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively, hence $\mathbf{a}_1.\mathbf{a}_3' = 1$ and $\mathbf{a}_2.\mathbf{a}_3' = -1$. Thus, from Equation (2.2.7), $|\mathbf{\Omega}_A| = (48+10)^3 + 2(0)(1)(-1) - (48+10)\left((0)^2 + (1)^2 + (-1)^2\right) = 194,996$. As the upper bound of $|\mathbf{\Omega}_A|$ is $(48 + 10)^3 = 195,112$, this design is $\left(\frac{194,996}{195,112}\right)^{1/10} = 99.994\%$ efficient relative to the theoretical bound.

Suppose, we sought to maximise the number of $\mathbf{a}_x.\mathbf{a}_y' = 0$ rather than minimise the absolute value of the maximum $\mathbf{a}_x.\mathbf{a}_y'$. The run $\mathbf{r}_3 = 1011$ has Hamming distance (d_1, d_2, d_3) = (1, 0, 2) and (0, 0, 0) from $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively, hence $\mathbf{a}_1.\mathbf{a}_3' = 0$ and $\mathbf{a}_2.\mathbf{a}_3' = 10$. Thus, we have $\mathbf{a}_x.\mathbf{a}_y' = 0$ for two of the three pairs. From Equation (2.2.7), $|\mathbf{\Omega}_A| = (48+10)^3 + 2(0)(0)(10) - (48+10)\left((0)^2 + (0)^2 + (10)^2\right) = 189,312$, hence this design is $\left(\frac{189,312}{195,112}\right)^{1/10} = 99.699\%$ efficient relative to the theoretical bound. Hence, this design is less efficient than the design we obtained by minimising the the absolute value of the maximum $\mathbf{a}_x.\mathbf{a}_y'$. □

Example 2.2.4. Following from Example 2.2.3, consider an OA$[48, 2^1 \times 3^1 \times 4^2]$, hence $\alpha = 10$, but now suppose we would like to remove $t = 3$ runs. The optimal set of three runs to remove will be the same as the optimal set of three runs to augment, provided those runs are included in the particular OA we are interested in. Hence, if we can find an OA that contains the set of three runs identified in Example 2.2.3 ($\mathbf{r}_1 = 0000$, $\mathbf{r}_2 = 1011$ and $\mathbf{r}_3 = 1102$) then we know these are the optimal set of runs to remove. Table 2.2.3 below gives an example of an OA$[48, 2^1 \times 3^1 \times 4^2]$ which does contain these runs (run numbers 1, 28 and 34), hence this solution is optimal for both augmenting and missing runs. Thus, from Equation (2.2.8), the maximum realisable value of $|\mathbf{\Omega}_B|$ is $(10-48)^3 + 2(0)(1)(-1) - (10-48)\left((0)^2 + (1)^2 + (-1)^2\right) = -54,798$. As the upper bound of $|\mathbf{\Omega}_B|$ is $(10-48)^3 = -54,872$, this design is $\left(\frac{-54,798}{-54,872}\right)^{1/10} = 99.987\%$ efficient relative to the theoretical bound. □

Table 2.2.3: OA$[48, 2^1 \times 3^1 \times 4^2]$

1	0	0	0	0	13	0	1	2	0	25	1	0	0	0	37	1	1	2	0
2	0	0	0	0	14	0	1	2	3	26	1	0	0	1	38	1	1	2	3
3	0	0	1	1	15	0	1	3	1	27	1	0	1	0	39	1	1	3	1
4	0	0	1	1	16	0	1	3	2	28	1	0	1	1	40	1	1	3	2
5	0	0	2	2	17	0	2	0	2	29	1	0	2	2	41	1	2	0	3
6	0	0	2	2	18	0	2	0	3	30	1	0	2	3	42	1	2	0	3
7	0	0	3	3	19	0	2	1	2	31	1	0	3	2	43	1	2	1	2
8	0	0	3	3	20	0	2	1	3	32	1	0	3	3	44	1	2	1	2
9	0	1	0	1	21	0	2	2	0	33	1	1	0	1	45	1	2	2	1
10	0	1	0	2	22	0	2	2	1	34	1	1	0	2	46	1	2	2	1
11	0	1	1	0	23	0	2	3	0	35	1	1	1	0	47	1	2	3	0
12	0	1	1	3	24	0	2	3	1	36	1	1	1	3	48	1	2	3	0

As mentioned earlier, $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_B$ have such similar forms that the processes for optimising each are very similar. The main points of difference are:

1. The diagonal entries in $\mathbf{\Omega}_A$ are $N + \alpha$, whereas in $\mathbf{\Omega}_B$ the diagonal entries are equal to $\alpha - N$.

2. $|\mathbf{M}_A| = N^{\alpha-p}|\mathbf{\Omega}_A|$
   $|\mathbf{M}_B| = (-1)^t N^{\alpha-t}|\mathbf{\Omega}_B|$.

3. In the case of augmenting runs, the experimenter is free to choose any set of p runs from the complete factorial, whereas in the case of missing runs, the experimenter is constrained the runs within the OA (assuming the logistics of the experiment allows the designer to choose which runs are missing at all).

Hence, for simplicity, the remainder of this chapter will focus on the properties of $\mathbf{\Omega}_A$ in the context of augmenting p runs, however the same logic applies to $\mathbf{\Omega}_B$ in the context of missing t runs. We will close the chapter with a case study of missing runs for a particular set of designs in Section 2.7.

As we shall see in the next section, we can always determine the optimal value of $\mathbf{a}_1.\mathbf{a}_2'$ for any design when $p = 2$. However, for any $p > 2$, it is not as straight forward to determine the optimal structure for $\mathbf{\Omega}_A$ unless it is possible to obtain all $\mathbf{a}_x.\mathbf{a}_y' = 0$ for all pairs of p runs. We give some constructions which extend results for $p = 2$ to larger values of p in subsequent sections.

2.3 Augmenting OAs by Adjoining Two Runs

In this section we begin by expressing the Hamming distances associated with the last two sets of factors in terms of the Hamming distances for the first $k-2$ sets of factors. We then describe an algorithm for determining these values and finally discuss some construction considerations.

2.3.1 Optimal Hamming Distance

As before, we let d_i be the Hamming distance between $\mathbf{r}_1$ and $\mathbf{r}_2$ for the m_i factors with s_i levels, $1 \leq i \leq k$. Thus, we know that $0 \leq d_i \leq m_i$, $1 \leq i \leq k$.

Suppose we have a vector of $k-2$ integers $[d_1, d_2, \ldots, d_{k-2}]$ such that $0 \leq d_i \leq m_i$. We let $d_k = \delta$ for some integer δ such that $0 \leq \delta \leq m_k$. Suppose $\mathbf{a}_1.\mathbf{a}'_2 = q$. Then rearranging Equation (2.2.6) gives

$$d_{k-1} = m_{k-1} + \frac{1 - q - s_k\delta + (s_k - 1)m_k - m_{k-1} + \sum_{i=1}^{k-2}[(s_i - 1)m_i - s_i d_i]}{s_{k-1}}. \tag{2.3.1}$$

This solution will only be valid when d_{k-1} is an integer, that is,

$$s_{k-1} \;\Bigg|\; \left(1 - q - s_k\delta + (s_k - 1)m_k - m_{k-1} + \sum_{i=1}^{k-2}[(s_i - 1)m_i - s_i d_i]\right) \tag{2.3.2}$$

and when

$$0 \leq d_{k-1} \leq m_{k-1}. \tag{2.3.3}$$

We can ensure (2.3.3) is satisfied by strengthening the bounds on δ, which become

$$\begin{aligned} &\delta \leq \min\left\{\left\lfloor \frac{1 - q + \sum_{i=1}^{k}(s_i - 1)m_i - \sum_{i=1}^{k-2} s_i d_i}{s_k} \right\rfloor, m_k\right\}, \\ &\text{and} \\ &\delta \geq \max\left\{\left\lceil \frac{1 - q - s_{k-1}m_{k-1} + \sum_{i=1}^{k}(s_i - 1)m_i - \sum_{i=1}^{k-2} s_i d_i}{s_k} \right\rceil, 0\right\}, \end{aligned} \tag{2.3.4}$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x and $\lceil x \rceil$ is the smallest integer greater than or equal to x. To achieve the optimal value of $\mathbf{a}_1.\mathbf{a}'_2$, we start

by setting $q = 0$ and search for the vector of $k-2$ integers $[d_1, d_2, \ldots, d_{k-2}]$, and hence δ, such that Equations (2.3.2) and (2.3.4) are satisfied. If no such values exist, we set q to 1 and repeat the procedure. To keep $|q|$ as small as possible, the values of q to be considered are, in order, $0, 1, -1, 2, -2, 3, -3, \ldots$ and so on. The following example illustrates this.

Example 2.3.1. Consider an OA$[16, 2^3 \times 4^4]$. Here $[d_1, d_2, \ldots, d_{k-2}]$ is of length zero since $k = 2$. For $q = 0$, Equation (2.3.2) requires a value of δ such that $2|(1 - 4\delta - m_1 + 3m_2) = (10 - 4\delta)$, and so this constraint places no restrictions on δ. Thinking about the bounds on δ, Equation (2.3.4) requires that $\max\left\{\left\lceil\frac{1-m_1+3m_2}{4}\right\rceil, 0\right\} \leq \delta \leq \min\left\{\left\lfloor\frac{1+m_1+3m_2}{4}\right\rfloor, m_2\right\}$ which becomes $3 \leq \delta \leq 4$. When $\delta = 3$, hence $d_2 = \delta = 3$, then Equation (2.3.1) gives $d_1 = 3 + (1 - 12 + 12 - 3)/2 = 2$. An example of a pair of runs with Hamming distance equal to $(d_1, d_2) = (2, 3)$ is:

$$\begin{array}{cc} & \overbrace{}^{s_1=2}\overbrace{}^{s_2=4} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,1\,1\,0\,1\,1\,1 \end{array}$$

When $\delta = 4$ then $d_1 = 3 + (1 - 16 + 12 - 3)/2 = 0$ and $d_2 = \delta = 4$. An example of a pair of runs with Hamming distance equal to $(d_1, d_2) = (0, 4)$ is:

$$\begin{array}{cc} & \overbrace{}^{s_1=2}\overbrace{}^{s_2=4} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,0\,0\,1\,1\,1\,1 \end{array}$$

In both cases we have $\mathbf{a}_1.\mathbf{a}_2' = 0$, as expected, hence we have found two optimal solutions. □

In the example above, where Equation (2.3.2) required $2|(1 - q - 4\delta - m_1 + 3m_2)$, we were able to find a solution for $q = 0$ as m_1 was odd and m_2 was even. If, however, $m_1 \equiv m_2 \mod 2$, then the condition in Equation (2.3.2) will never be satisfied for $q = 0$ and $s_1 = 2$, regardless of δ. In this case, the best realisable value of q must be non-zero. The next example illustrates this.

Example 2.3.2. Consider an OA$[32, 2^4 \times 4^4]$. As $m_1 \equiv m_2 \mod 2$, we set $q = 1$ so that condition (2.3.2) becomes $2|(4\delta + m_1 - 3m_2) = (4\delta - 8)$. The bounds on δ are

$$\max\left\{\left\lceil\frac{-m_1+3m_2}{4}\right\rceil, 0\right\} \leq \delta \leq \min\left\{\left\lfloor\frac{m_1+3m_2}{4}\right\rfloor, m_2\right\}$$

which becomes $2 \leq \delta \leq 4$. The set of (d_1, d_2) for each allowable value of δ, where d_1 is calculated using Equation (2.3.1), and an example of a pair of runs that realise each of these Hamming distances is given in Table 2.3.1. In all three cases we have $\mathbf{a}_1.\mathbf{a}_2' = 1$, as expected, hence we have three solutions to choose from. □

Table 2.3.1: Optimal $p = 2$ runs to append to OA$[32, 2^4 \times 4^4]$

δ	(d_1, d_2)	Example Runs	
		$\overbrace{}^{s_1=2}\overbrace{}^{s_2=4}$	
2	(4, 2)	$\mathbf{r}_1$	0 0 0 0 0 0 0 0
		$\mathbf{r}_2$	1 1 1 1 0 0 1 1
3	(2, 3)	$\mathbf{r}_1$	0 0 0 0 0 0 0 0
		$\mathbf{r}_2$	0 0 1 1 0 1 1 1
4	(0, 4)	$\mathbf{r}_1$	0 0 0 0 0 0 0 0
		$\mathbf{r}_2$	0 0 0 0 1 1 1 1

We will now consider an example with larger k.

Example 2.3.3. Consider an OA$[72, 2^8 \times 3^2 \times 4^1 \times 6^1]$. Here $k = 4$ and so $[d_1, d_2, \ldots, d_{k-2}]$ will be of length 2. Since $0 \leq d_i \leq m_i$, we will potentially need to consider $(m_1 + 1)(m_2 + 1) = (8 + 1)(2 + 1) = 27$ possible values for $[d_1, d_2]$, that is $[d_1, d_2] = [0, 0], [0, 1], [0, 2], [1, 0], \ldots, [8, 2]$. We start by setting $q = 0$ and for each set of potential $[d_1, d_2]$ we iterate through all possible values of δ, given the bounds in Equation (2.3.4), and determine whether condition (2.3.2) is satisfied.

Table 2.3.2: Optimal $p = 2$ runs to append to OA$[32, 2^4 \times 4^4]$

(d_1, d_2, d_3, d_4)		Example Runs											
		$\overbrace{}^{s_1=2}$								$\overbrace{}^{s_2=3}$		$\overbrace{}^{s_3=4}$	$\overbrace{}^{s_4=6}$
(4, 1, 1, 1)	$\mathbf{r}_1$	0	0	0	0	0	0	0	0	0	0	0	0
	$\mathbf{r}_2$	0	0	0	0	1	1	1	1	0	1	1	1
(6, 1, 0, 1)	$\mathbf{r}_1$	0	0	0	0	0	0	0	0	0	0	0	0
	$\mathbf{r}_2$	0	0	1	1	1	1	1	1	0	1	0	1
(7, 1, 1, 0)	$\mathbf{r}_1$	0	0	0	0	0	0	0	0	0	0	0	0
	$\mathbf{r}_2$	0	1	1	1	1	1	1	1	0	1	1	0

We find multiple instances where these conditions are satisfied, as given above, and so the theoretical optimal value is realisable hence we do not require further searching with different values of q. □

The previous three examples illustrate the process described in pseudo-code in Algorithm 2.1. We note that when $k \leq 2$, the vector $[d_1, d_2, \ldots, d_{k-2}]$ will be of length zero, and so we are effectively ignoring lines 1 and 5.

Note that if $k = 1$ (i.e. the OA is symmetric) then Equations (2.3.1) and (2.3.2) are not applicable, hence all that remains is to find a value of δ such that

$$\max\left\{\left\lceil\frac{1-q+(s_1-1)m_1}{s_1}\right\rceil, 0\right\} \leq \delta \leq \min\left\{\left\lfloor\frac{1-q+(s_1-1)m_1}{s_1}\right\rfloor, m_1\right\}.$$

Thus we need to find the smallest value of q that satisfies $s_1 \mid (1 - q - m_1)$ and then let $\delta = \frac{1-q+(s_1-1)m_1}{s_1}$. Chatzopoulos et al. (2011) have shown for symmetric

Algorithm 2.1: Optimal Hamming distances when $p = 2$

Input: $(s_1, s_2, \ldots, s_k)$ and $(m_1, m_2, \ldots, m_k)$
Output: This best realisable q and a list of all associated $(d_1, d_2, \ldots, d_k)$

Generate all combinations of $[d_1, d_2, \ldots, d_{k-2}]$ such that $0 \leq d_i \leq m_i$
$solved \leftarrow$ FALSE
$q \leftarrow 0$
while *solved = FALSE* **do**
  for each *set of* $[d_1, d_2, \ldots, d_{k-2}]$ **do**
    Calculate the lower and upper bounds of δ
    for each *plausible value of* δ **do**
      if
      $s_{k-1} \mid \left(1 - q - s_k\delta + (s_k - 1)m_k - m_{k-1} + \sum_{i=1}^{k-2}[(s_i - 1)m_i - s_i d_i]\right)$
      then
        $solved \leftarrow$ TRUE
        $d_{k-1} \leftarrow m_{k-1} + \frac{1-q-s_k\delta+(s_k-1)m_k-m_{k-1}+\sum_{i=1}^{k-2}[(s_i-1)m_i-s_i d_i]}{s_{k-1}}$
        $d_k \leftarrow \delta$
        Output solution $(d_1, d_2, \ldots, d_k)$
  if *solved = TRUE* **then**
    Output q
  else
    if $q > 0$ **then**
      $q \leftarrow -q$
    else
      $q \leftarrow |q| + 1$

OAs that adjoining any two runs with Hamming distance $m_1 - \text{round}[(m_1 - 1)/s_1]$ gives designs which are equally good. Our representation of δ for $k = 1$ is equivalent to this result. Note that whether $1/2$ is rounded up or down makes no difference to the D-optimality value of the augmented design.

2.3.2 Designs with $N \leq 100$

Following the lead of Chai et al. (2002), suppose we restrict our attention to designs in which $N \leq 100$. These designs can be obtained by taking the 192 parent OAs from Kuhfeld (2006) with $N \leq 100$, systematically making all possible expansive replacements and level collapsings, and considering all the designs that can result by dropping one or more factors from each of these designs. The following example illustrates this process by describing how the OA$[32, 2^5 \times 4^7]$ seen previously in Table 2.2.1 was constructed.

Example 2.3.4. The OA$[32, 4^8 \times 8^1]$ from Kuhfeld (2006) is reproduced in Table 2.3.3 below.

Table 2.3.3: OA$[32, 4^8 \times 8^1]$ from Kuhfeld (2006)

1	0	0	0	0	0	0	0	0	0
2	0	0	1	1	3	3	2	2	1
3	0	1	2	3	0	1	2	3	2
4	0	1	3	2	3	2	0	1	3
5	0	2	0	2	1	3	1	3	4
6	0	2	1	3	2	0	3	1	5
7	0	3	2	1	1	2	3	0	6
8	0	3	3	0	2	1	1	2	7
9	1	0	2	3	2	3	1	0	3
10	1	0	3	2	1	0	3	2	2
11	1	1	0	0	2	2	3	3	1
12	1	1	1	1	1	1	1	1	0
13	1	2	2	1	3	0	0	3	7
14	1	2	3	0	0	3	2	1	6
15	1	3	0	2	3	1	2	0	5
16	1	3	1	3	0	2	0	2	4
17	2	0	2	0	3	1	3	1	4
18	2	0	3	1	0	2	1	3	5
19	2	1	0	3	3	0	1	2	6
20	2	1	1	2	0	3	3	0	7
21	2	2	2	2	2	2	2	2	0
22	2	2	3	3	1	1	0	0	1
23	2	3	0	1	2	3	0	1	2
24	2	3	1	0	1	0	2	3	3
25	3	0	0	3	1	2	2	1	7
26	3	0	1	2	2	1	0	3	6
27	3	1	2	0	1	3	0	2	5
28	3	1	3	1	2	0	2	0	4
29	3	2	0	1	0	1	3	2	3
30	3	2	1	0	3	2	1	0	2
31	3	3	2	2	0	0	1	1	1
32	3	3	3	3	3	3	3	3	0

We can use the OA$[8, 2^4 \times 4^1]$ from Kuhfeld (2006) to replace the 8-level factor with four binary factors and one 4-level factor as shown in Table 2.3.4. Once this expansive replacement has been made, the OA$[32, 4^8 \times 8^1]$ becomes a OA$[32, 2^4 \times 4^9]$. We can use the OA$[4, 2^3]$ from Kuhfeld (2006) to replace any of the 4-level factors with three binary factors as shown in Table 2.3.5. Replacing just one of the 4-level factors will convert the OA$[32, 2^4 \times 4^9]$ to a OA$[32, 2^7 \times 4^8]$, hence we can obtain a OA$[32, 2^5 \times 4^7]$ by dropping two of the binary factors and one of the 4-level factors. The OA$[32, 2^5 \times 4^7]$ in Table 2.2.1 was constructed by expansively

Table 2.3.4: Expansive replacement of an 8-level factor

8-lvl Factor		OA$[8, 2^4 \times 4^1]$				
0	$\rightarrow$	0	0	0	0	0
1	$\rightarrow$	0	0	1	1	2
2	$\rightarrow$	0	1	0	1	1
3	$\rightarrow$	0	1	1	0	3
4	$\rightarrow$	1	0	0	1	3
5	$\rightarrow$	1	0	1	0	1
6	$\rightarrow$	1	1	0	0	2
7	$\rightarrow$	1	1	1	1	0

Table 2.3.5: Expansive replacement of a 4-level factor

4-lvl Factor		OA$[4, 2^3]$		
0	$\rightarrow$	0	0	0
1	$\rightarrow$	0	1	1
2	$\rightarrow$	1	0	1
3	$\rightarrow$	1	1	0

replacing the first 4-level factor in the OA$[32, 2^4 \times 4^9]$ generated as described above, then dropping the 5th and 7th binary factors and the last 4-level factor from the resulting OA$[32, 2^7 \times 4^8]$. □

As we do not require N to find the best distance vectors, we consider only the distinct sets of $\{s_1^{m_1}, s_2^{m_2}, \ldots, s_k^{m_k}\}$ from these designs. There are 23,275 such sets, which are broken down by k and the best realisable value of $|q|$ in Table 2.3.6.

Table 2.3.6: Best $|q|$ when $p = 2$ for all OAs with $N \leq 100$

		$\lvert q\rvert$					
		0	**1**	**2**	**3**	**4**	**Total**
k	**1**	76	96	25	10	3	210
	2	3,300	1,863	87	7	–	5,257
	3	9,393	3,479	5	–	–	12,877
	4	4,231	700	–	–	–	4,931
Total		17,000	6,138	117	17	3	23,275

A Python implementation of Algorithm 2.1, run on a x64 PC with Intel Core i5, determined all optimal Hamming distances and associated q for each of the 23,275 designs in Table 2.3.6 in less than a minute. Of course when $p = 2$, the process of constructing a pair of runs for a given Hamming distance is trivial as we saw in Examples 2.3.1, 2.3.2 and 2.3.3 above.

In the case of missing runs, however, the parameters $\{s_1^{m_1}, s_2^{m_2}, \ldots, s_k^{m_k}\}$ and optimal $|q|$ are not enough to construct an optimal design as we also need to see if there are a pair of runs in the OA with the associated Hamming distance. Hence, rather

than considering only the distinct set of $\{s_1^{m_1}, s_2^{m_2}, \ldots, s_k^{m_k}\}$ described above, we also require an enumeration of all the non-isomorphic OAs. This additional requirement increases the search exponentially, as the following example illustrates.

Example 2.3.5. The construction of an OA$[32, 2^5 \times 4^7]$ described in Example 2.3.4 involved three choices:

1. Expand one of nine 4-level factors: $\binom{9}{1} = 9$ options
2. Drop one of eight 4-level factors: $\binom{8}{1} = 8$ options
3. Drop two of seven 2-level factors: $\binom{7}{2} = 21$ options

Hence there are at least $9 \times 8 \times 21 = 1,512$ initial designs to consider although the total may be reduced after the determination of isomorphic sets of designs (if any). There may also be other plausible combinations of starting OAs and expansive replacements to consider. In fact, using the same starting OA, we could have expanded two, rather than one, of the nine 4-level factors ($\binom{9}{2} = 36$ options), to give ten 2-level factors, of which we drop five ($\binom{10}{5} = 252$ options). In any case, all of these designs are associated with a single set, $\{2^5, 4^7\}$, out of the 23,275 distinct sets described in Table 2.3.6. □

The computational issues around the enumeration of OAs in the context of missing runs will be discussed in greater detail in the final section of this chapter. For now, we will return our focus to augmenting runs, hence the parameters $\{s_1^{m_1}, s_2^{m_2}, \ldots, s_k^{m_k}\}$ and optimal $|q|$ are enough to continue our investigation.

2.3.3 Further Construction Considerations

Although the algorithm is very fast it is nonetheless interesting from a theoretical perspective to make some comments about how it is possible to step from one optimal solution to another. We discuss three possible ways to step between solutions below.

Suppose that $(d_1,\ d_2)$ is an optimal Hamming distance for $\{s_1^{m_1}, s_2^{m_2}\}$ associated with $\mathbf{a}_1.\mathbf{a}_2' = q$. Let g be the greatest common denominator of s_1 and s_2, (which we will denote $\gcd(s_1, s_2)$ from now on), so that $s_i = gt_i$, $i = 1, 2$. Then $1 + (s_1 - 1)m_1 + (s_2 - 1)m_2 - s_1d_1 - s_2d_2 = q$. Rearranging we see that $d_1 = m_1 + (1 - q - gt_2d_2 + (gt_2 - 1)m_2 - m_1)/gt_1$. Suppose we increment d_2 by a. Then the associated d_1 is $d_1 - at_2/t_1$, and since $\gcd(t_1, t_2) = 1$, the possible values for a are multiples of t_1, such that d_1 and d_2 remain within the allowable range. We illustrated this idea in Example 2.3.1, where $s_1 = 2$, $s_2 = 4$, and so $g = 2$, $t_1 = 1$ and $t_2 = 2$. We showed that when $(m_1,\ m_2) = (3,\ 4)$, $(d_1,\ d_2) = (2,3)$ or $(0,4)$

gave $q = 0$. If we start with $(d_1,\ d_2)$ = (2,3) and increment d_2 by $a = 1$ to give $d_2 = 3 + 1 = 4$ then we must decrease d_1 by $at_2/t_1 = 2$ to give $d_1 = 0$. No other solutions for d_1 and d_2 satisfy the constraints, which we can see from Table 2.3.7.

Table 2.3.7: Optimal (d_1, d_2) for $(s_1, s_2) = (2, 4)$ and a select set of (m_1, m_2)

m_1 \ m_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	**(0, 1)** **(1, 1)**	(0, 2)	**(0, 3)** **(1, 2)**	(1, 3)	**(0, 4)** **(1, 4)**	(0, 5)	**(0, 6)** **(1, 5)**	(1, 6)
2	(1, 1)	**(0, 2)** **(1, 2)** **(2, 1)**	(0, 3) (2, 2)	**(0, 4)** **(1, 3)** **(2, 3)**	(1, 4)	**(0, 5)** **(1, 5)** **(2, 4)**	(0, 6) (2, 5)	**(0, 7)** **(1, 6)** **(2, 6)**
3	**(1, 1)** **(2, 1)** **(3, 0)**	(1, 2) (3, 1)	**(0, 3)** **(1, 3)** **(2, 2)** **(3, 2)**	(0, 4) (2, 3)	**(0, 5)** **(1, 4)** **(2, 4)** **(3, 3)**	(1, 5) (3, 4)	**(0, 6)** **(1, 6)** **(2, 5)** **(3, 5)**	(0, 7) (2, 6)
4	(2, 1) (4, 0)	**(1, 2)** **(2, 2)** **(3, 1)** **(4, 1)**	(1, 3) (3, 2)	**(0, 4)** **(1, 4)** **(2, 3)** **(3, 3)** **(4, 2)**	(0, 5) (2, 4) (4, 3)	**(0, 6)** **(1, 5)** **(2, 5)** **(3, 4)** **(4, 4)**	(1, 6) (3, 5)	**(0, 7)** **(1, 7)** **(2, 6)** **(3, 6)** **(4, 5)**
5	**(2, 1)** **(3, 1)** **(4, 0)** **(5, 0)**	(2, 2) (4, 1)	**(1, 3)** **(2, 3)** **(3, 2)** **(4, 2)** **(5, 1)**	(1, 4) (3, 3) (5, 2)	**(0, 5)** **(1, 5)** **(2, 4)** **(3, 4)** **(4, 3)** **(5, 3)**	(0, 6) (2, 5) (4, 4)	**(0, 7)** **(1, 6)** **(2, 6)** **(3, 5)** **(4, 5)** **(5, 4)**	(1, 7) (3, 6) (5, 5)
6	(3, 1) (5, 0)	**(2, 2)** **(3, 2)** **(4, 1)** **(5, 1)** **(6, 0)**	(2, 3) (4, 2) (6, 1)	**(1, 4)** **(2, 4)** **(3, 3)** **(4, 3)** **(5, 2)** **(6, 2)**	(1, 5) (3, 4) (5, 3)	**(0, 6)** **(1, 6)** **(2, 5)** **(3, 5)** **(4, 4)** **(5, 4)** **(6, 3)**	(0, 7) (2, 6) (4, 5) (6, 4)	**(0, 8)** **(1, 7)** **(2, 7)** **(3, 6)** **(4, 6)** **(5, 5)** **(6, 5)**
7	**(3, 1)** **(4, 1)** **(5, 0)** **(6, 0)**	(3, 2) (5, 1) (7, 0)	**(2, 3)** **(3, 3)** **(4, 2)** **(5, 2)** **(6, 1)** **(7, 1)**	(2, 4) (4, 3) (6, 2)	**(1, 5)** **(2, 5)** **(3, 4)** **(4, 4)** **(5, 3)** **(6, 3)** **(7, 2)**	(1, 6) (3, 5) (5, 4) (7, 3)	**(0, 7)** **(1, 7)** **(2, 6)** **(3, 6)** **(4, 5)** **(5, 5)** **(6, 4)** **(7, 4)**	(0, 8) (2, 7) (4, 6) (6, 5)
8	(4, 1) (6, 0)	**(3, 2)** **(4, 2)** **(5, 1)** **(6, 1)** **(7, 0)** **(8, 0)**	(3, 3) (5, 2) (7, 1)	**(2, 4)** **(3, 4)** **(4, 3)** **(5, 3)** **(6, 2)** **(7, 2)** **(8, 1)**	(2, 5) (4, 4) (6, 3) (8, 2)	**(1, 6)** **(2, 6)** **(3, 5)** **(4, 5)** **(5, 4)** **(6, 4)** **(7, 3)** **(8, 3)**	(1, 7) (3, 6) (5, 5) (7, 4)	**(0, 8)** **(1, 8)** **(2, 7)** **(3, 7)** **(4, 6)** **(5, 6)** **(6, 5)** **(7, 5)** **(8, 4)**

When $m_1 + m_2 = 1 \ (mod\, 2)$, optimal $q = 0$.
*When $m_1 + m_2 = 0 \ (mod\, 2)$, as indicated in **bold**, optimal $|q| = 1$.*

Similarly, we can step from an optimal Hamming distance for a given set of $\{s_1^{m_1}, s_2^{m_2}, \ldots, s_k^{m_k}\}$ to a design with $\{s_1^{m_1'}, s_2^{m_2'}, \ldots, s_k^{m_k'}\}$ for new values m_i'. Again, suppose $k = 2$, and for a given set of (m_1, m_2) and $(s_1, s_2) = (gt_1, gt_2)$ such that $\gcd(t_1, t_2) = 1$, and that the Hamming distance (d_1, d_2) corresponds to a specific q. We let $m_1' = m_1 + b$ and $m_2' = m_2 - b$, then using the original value for d_2 we have $d_1' = (m_1 + b) + (1 - q - gt_2 d_2 + (gt_2 - 1)(m_2 - b) - (m_1 + b))\,/gt_1 = d_1 - b(t_2/t_1 - 1)$. Starting from, say, (m_1, m_2) = (5, 6) and (d_1, d_2) = (4, 4) in Table 2.3.7, we let $b = 2$ so that $(m_1', m_2') = (5 + 2, 6 - 2) = (7,\ 4)$ and

$(d_1', d_2') = (4-2(2/1-1), 4) = (2, 4)$. We can see that this form of stepping is only possible when $m_2' \geq 0$, $d_1' \geq 0$, and $t_1 | bt_2$.

To step from k to $k+1$ factors, note that if $(d_1, d_2, \ldots, d_k)$ is a set of distances such that $\mathbf{a}_1.\mathbf{a}_2' = q$, then if $(s_{k+1}-1)m_{k+1} = s_{k+1}d_{k+1}$, a set of m_{k+1} factors with s_{k+1} levels can be adjoined to the original design and the value of q is unaltered. The constraint means that $s_{k+1} | m_{k+1}$ and $d_{k+1} = (s_{k+1}-1)m_{k+1}/s_{k+1}$. The next example illustrates this.

Example 2.3.6. Consider a $2^4 \times 4^1$ design. From Table 2.3.7 we see there are 2 sets of $(d_1, d_2) = (2,1)$ or $(4,0)$ with $q = 0$. Suppose we would like to append a set of 3-level factors so the design becomes $2^4 \times 4^1 \times 3^{m_3}$ for some m_3 which is a multiple of $s_3 = 3$. We require $d_3 = 2m_3/3$, and so when $m_3 = 3$, $(d_1, d_2, d_3) = (2, 1, 2)$ or $(4, 0, 2)$; when $m_3 = 6$, $(d_1, d_2, d_3) = (2, 1, 4)$ or $(4, 0, 4)$; when $m_3 = 9$, $(d_1, d_2, d_3) = (2, 1, 6)$ or $(4, 0, 6)$ and so on. □

Table 2.3.7 gives optimal (d_1, d_2) for $(s_1, s_2) = (2, 4)$ and $1 \leq m_1 \leq 8$, $1 \leq m_2 \leq 8$. Similar tables for $(s_1, s_2) = (2, 3)$, $(2, 5)$, $(3, 4)$ and $(3, 5)$ can be found in the appendix.

We can also drop the ith set of factors provided that $s_i d_i = (s_i - 1)m_i$ since the value of q will be unaltered. The following example illustrates the approach required.

Example 2.3.7. Consider a $2^{12} \times 3^3 \times 4^1 \times 6^6$ design. There are 12 sets of (d_1, d_2, d_3, d_4) with $q = 0$ and these are given below. In each set we have also indicated in bold which values of d_i are equal to $(s_i - 1)m_i/s_i$ so that $s_i d_i = (s_i - 1)m_i$ is satisfied. We see that d_3 never equals 3/4, but we have multiple occurrences of $d_1 = (2-1)12/2 = 6$, $d_2 = (3-1)3/3 = 2$ and $d_4 = (6-1)6/6 = 5$.

$$\begin{array}{lllllll} (3, \mathbf{2}, 1, 6) & (\mathbf{6}, 0, 1, 6) & (8, 0, 0, 6) & (9, 0, 1, \mathbf{5}) & (11, 0, 0, \mathbf{5}) & (12, 0, 1, 4) \\ (5, \mathbf{2}, 0, 6) & (\mathbf{6}, \mathbf{2}, 1, \mathbf{5}) & (8, \mathbf{2}, 0, \mathbf{5}) & (9, \mathbf{2}, 1, 4) & (11, \mathbf{2}, 0, 4) & (12, \mathbf{2}, 1, 3) \end{array}$$

□

In the following sections we will discuss how to augment OAs by adjoining more than two runs.

2.4 Augmenting OAs by Adjoining Three Runs

Ideally we would like to adjoin $p = 3$ runs in which the $\mathbf{a}_x.\mathbf{a}_y'$ for any pair is equal to the smallest possible $|q|$. As mentioned earlier, the construction of a pair of runs that realise a given Hamming distance (hence $|q|$) is trivial as we can assume

without loss of generality that the first run is a vector of zeros, and that for each set of m_i factors in the second run, d_i positions are 1 (or any non-zero value smaller than s_i) and the remaining $m_i - d_i$ positions are zero for each i, $1 \leq i \leq k$. Hence, when $p = 2$ the only concern is to find a Hamming distance that minimises $|q|$ as the pair of runs is implied. However, when $p > 2$, there is no guarantee that a set of p runs exists such that a given Hamming distance is realised for all pairs, as we saw in Example 2.2.3. Hence, it is not enough to find the optimal $|q|$, we also need to determine whether a set of p runs can be constructed such that the all pairs realise the associated Hamming distance. Hence, this section considers methods for finding a set of three runs that pairwise realise the best Hamming distance (or a mix, if multiple optimal Hamming distances exist).

It is best if all of the off-diagonal entries are 0. We saw in Table 2.3.6 that it is possible to find a Hamming distance such that $\mathbf{a}_1.\mathbf{a}_2' = 0$ for at least one pair of runs in the complete factorial for 17,000 of the 23,275 OAs with $N \leq 100$. As discussed above, this does not guarantee that we are able to construct three runs such that $\mathbf{a}_x.\mathbf{a}_y' = 0$ for all pairs. The number of designs where this is possible is given in Table 2.4.1 below.

Table 2.4.1: Number of designs where we can extend a pair of runs with $\mathbf{a}_1.\mathbf{a}_2' = 0$ to a triple of runs in which each pair has $\mathbf{a}_x.\mathbf{a}_y' = 0$

		All $\mathbf{a}_x.\mathbf{a}_y' = 0$		
		Yes	**No**	**Total**
k	**1**	52	24	76
	2	2,373	927	3,300
	3	6,804	2,589	9,393
	4	3,396	835	4,231
Total		12,625	4,375	17,000

If all $\mathbf{a}_x.\mathbf{a}_y'$ are non-zero and equal then it is better for the product of the values to be positive since in that case we would get the smallest realisable determinant. If neither of these is possible then, as we said in Section 2.2.3, we have found that it is better to aim to minimise the absolute value of the maximum entry, and then to have as many small entries as possible.

We first consider constructing three runs by taking columns of levels such that any pair of runs will realise a given q. We then discuss some further run construction considerations in which it is possible to adjoin either a single run to an optimal pair, or a set of columns to an optimal triple.

2.4.1 Using the Same Hamming distance For All Pairs

We will record the Hamming distances between pairs of runs in a matrix. For $p > 1$, we let $\mathbf{D}_{kp}$ be a $k \times \binom{p}{2}$ matrix of Hamming distances. The rows of $\mathbf{D}_{kp}$ are labelled by the k sets of factors with the same number of levels and the columns are labelled by pairs of adjoined runs. The runs are assumed to be in some fixed, but arbitrary, order, and the columns are then labelled in lexicographic order. Row vector i, denoted by $\mathbf{d}_i$, has entries $d_{i(xy)}$ which is the Hamming distance between $\mathbf{r}_x$ and $\mathbf{r}_y$ for the m_i factors with s_i levels. We call $\mathbf{d}_i$ the *distance vector* corresponding to the column of levels. The *(Hamming) weight* of a distance vector is the number of non-zero entries in the distance vector.

We let $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ be the three adjoined runs. Without loss of generality we assume that $\mathbf{r}_1 = \mathbf{0}$, and we write the runs as the rows of a matrix. Then there are, up to re-naming of levels, five possible columns of levels, and these are given in the final column of Table 2.4.2. Of course the column of levels $[012]'$ is only possible when $s_i \geq 3$.

Table 2.4.2: Hamming distance vectors when $p = 3$

Weight	$d_{i(12)}$	$d_{i(13)}$	$d_{i(23)}$	**Levels**
0	0	0	0	$[000]'$
2	0	1	1	$[001]'$
	1	0	1	$[010]'$
	1	1	0	$[011]'$
3	1	1	1	$[012]'$

Suppose that we would like to have all the elements in $\mathbf{d}_i$ equal to some specific d_i. We see that for $s_i \geq 3$ we can achieve this by using d_i columns with weight 3 (i.e. type $[012]'$) and $m_i - d_i$ of weight 0 (i.e. type $[000]'$). If $s_i = 2$ then we need to have $d_i/2$ of each column with weight 2 (i.e. types $[001]'$, $[010]'$ and $[011]'$) and $m_i - 3d_i/2$ of weight 0. This is only possible if d_i is even and $m_i \geq 3d_i/2$.

Example 2.4.1. Consider an OA$[16, 2^3 \times 4^4]$. As we saw in Example 2.3.1, $q = 0$ can be achieved for any pair of runs with $(d_1, d_2) = (2, 3)$. Hence, an optimal set of 3 runs would be achieved if the following matrix of distance vectors were possible:

$$\mathbf{D}_{23} = \begin{matrix} \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{matrix} \begin{matrix} \begin{matrix} d_{i(12)} & d_{i(13)} & d_{i(23)} \end{matrix} \\ \begin{bmatrix} 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \end{matrix}.$$

Given $s_1 = 2$, d_1 is even, and $m_1 \geq 3d_1/2$, we can get such a set of 3 runs for $\mathbf{d}_1$ by taking $d_1/2 = 1$ of each column with weight 2, and as $s_2 \geq 3$, we can use $d_2 = 3$

columns of weight 3 and $m_2 - d_2 = 1$ column of weight 0 to achieve the desired $\mathbf{d}_2$. Thus, an optimal set of 3 runs to adjoin to this OA appears below.

$$\begin{array}{cc} & \overbrace{}^{s_1=2}\overbrace{}^{s_2=4} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,1\,1\,0\,1\,1\,1 \\ \mathbf{r}_3 & 1\,0\,1\,0\,2\,2\,2 \end{array}$$

□

Notice that the example above used only one of the two solutions found in Example 2.3.1. The same method could be applied to the second solution, $(d_1, d_2) = (0, 4)$, to construct a different set of optimal runs such that

$$\mathbf{D}_{23} = \begin{array}{c} \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{array} \begin{array}{c} \begin{array}{ccc} d_{i(12)} & d_{i(13)} & d_{i(23)} \end{array} \\ \left[\begin{array}{ccc} 0 & 0 & 0 \\ 4 & 4 & 4 \end{array}\right] \end{array}.$$

We are not concerned with which pair of (d_1, d_2) is used, only that $\mathbf{a}_x.\mathbf{a}'_y = 0$ for all pairs of adjoined runs. Hence, if it is possible to do so, we could construct $\mathbf{D}_{kp}$ from a number of distinct solutions and we will illustrate this in the following example.

Example 2.4.2. We adjoin the following 3 runs to a $2^3 \times 4^4$ OA in $N = 16$ runs.

$$\begin{array}{cc} & \overbrace{}^{s_1=2}\overbrace{}^{s_2=4} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,1\,1\,0\,1\,1\,1 \\ \mathbf{r}_3 & 0\,1\,1\,1\,0\,2\,2 \end{array}$$

The corresponding matrix of distance vectors is

$$\mathbf{D}_{23} = \begin{array}{c} \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{array} \begin{array}{c} \begin{array}{ccc} d_{i(12)} & d_{i(13)} & d_{i(23)} \end{array} \\ \left[\begin{array}{ccc} 2 & 2 & 0 \\ 3 & 3 & 4 \end{array}\right] \end{array}.$$

Each column of $\mathbf{D}_{23}$ corresponds to a pair of $(d_{1(xy)}, d_{2(xy)})$ such that $\mathbf{a}_x.\mathbf{a}'_y = 0$, hence the augmented design is D-optimal. □

Our focus has been on the construction of sets of 3 runs in which all the elements in $\mathbf{d}_i$ are equal, although the previous example shows that this is not essential to get all $\mathbf{a}_x.\mathbf{a}'_y = 0$. But there are designs in which having more than one distinct value in $\mathbf{d}_i$ is essential to obtain the best design. The following example illustrates this.

Example 2.4.3. Consider a $2^6 \times 3^4$ design. The two sets (3,3) and (6,1) for (d_1, d_2) both have $q = 0$. It is not possible to construct 3 runs in which the Hamming distances between pairs are either all (3,3) or all (6,1), but the distances between pairs of the following runs is a mix of these two solutions.

$$\begin{array}{cc} & \overbrace{}^{s_1=2}\overbrace{}^{s_2=3} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,0\,0\,1\,1\,1\,0\,1\,1\,1 \\ \mathbf{r}_3 & 1\,1\,1\,1\,1\,1\,0\,0\,0\,2 \end{array}$$

The corresponding matrix of distance vectors is

$$\mathbf{D}_{23} = \begin{array}{c} \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{array} \begin{array}{c} \begin{array}{ccc} d_{i(12)} & d_{i(13)} & d_{i(23)} \end{array} \\ \left[\begin{array}{ccc} 3 & 6 & 3 \\ 3 & 1 & 3 \end{array}\right] \end{array}.$$

Each column of $\mathbf{D}_{23}$ corresponds to a pair of $(d_{1(xy)}, d_{2(xy)})$ such that $\mathbf{a}_x.\mathbf{a}'_y = 0$, hence the augmented design is D-optimal. □

2.4.2 Further run construction considerations

We have discussed methods for constructing $p = 3$ runs from columns of levels above, however when all $s_i \geq 3$, which is the case for 1,269 of the 23,275 OAs with $N \leq 100$, this is effectively the same as adjoining a single run to an optimal pair. The best pair of runs to adjoin can be extended to a best set of three runs by extending each column of the form $[00]'$ to one of the form $[000]'$ and each column of the form $[01]'$ to one of the form $[012]'$.

Example 2.4.4. Consider a $3^6 \times 4^1$ design. When $p = 2$, we get $q = 0$ for $(d_1, d_2) = (4, 1)$, which can be represented by the following pair of runs:

$$\begin{array}{cc} & \overbrace{}^{s_1=3}\overbrace{}^{s_2=4} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,0\,1\,1\,1\,1\,1 \end{array}$$

Since all $s_i \geq 3$, we can extend this to $p = 3$ by making the column extensions described above. The resulting triple of runs is:

$$\begin{array}{cc} & \overbrace{}^{s_1=3}\overbrace{}^{s_2=4} \\ \mathbf{r}_1 & 0\,0\,0\,0\,0\,0\,0 \\ \mathbf{r}_2 & 0\,0\,1\,1\,1\,1\,1 \\ \mathbf{r}_3 & 0\,0\,2\,2\,2\,2\,2 \end{array}$$

Since $(d_{1(xy)}, d_{2(xy)}) = (4, 1)$ for all pairs of runs, the design is optimal. □

Furthermore, recall from Section 2.3.3 that we can append an additional m_{k+1} factors with s_{k+1} levels without altering q as long as $d_{k+1} = (s_{k+1} - 1)m_{k+1}/s_{k+1}$. This means we can append a set of $4a$ binary factors to any set of 3 runs without changing the value of q for any pair of the runs as long as $d_{k+1} = 2a$. One way to do this is to adjoin 0000 a times to one of the runs, 0011 a times to another of the runs and 0101 a times to the final run.

Example 2.4.5. Consider a $2^4 \times 3^6 \times 4^1$ design. Note this is equivalent to appending $4a$ binary factors, with $a = 1$, to the $3^6 \times 4^1$ design in Example 2.4.4. Hence, we can adjoin the level combinations as described above to the optimal runs in Example 2.4.4. The resulting runs are:

	s_1=2				s_2=3						s_3=4
$\mathbf{r}_1$	0	0	0	0	0	0	0	0	0	0	0
$\mathbf{r}_2$	0	0	1	1	0	0	1	1	1	1	1
$\mathbf{r}_3$	0	1	0	1	0	0	2	2	2	2	2

The Hamming distance for all pairs of runs is $(d_{1(xy)}, d_{2(xy)}, d_{3(xy)}) = (2, 4, 1)$ which gives $q = 0$ and so the solution is optimal. □

2.5 Augmenting OAs by Adjoining Four Runs

As in the previous section, we ideally we would like to adjoin $p = 4$ runs in which the $\mathbf{a}_x.\mathbf{a}'_y$ for any pair is equal to the smallest possible $|q|$ for the given set of s_i and m_i.

As before, we first consider constructing four runs by taking columns of levels such that any pair of runs will realise a given q. We then discuss some further run construction considerations in which it is possible to adjoin either a set of two runs to an optimal pair, or a set of columns to an optimal quadruple.

2.5.1 Using the Same Hamming Distance For All Pairs

We let $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ and $\mathbf{r}_4$ be the four adjoined runs. As before, we assume that $\mathbf{r}_1 = \mathbf{0}$, and we write the runs as the rows of a matrix. Then there are, up to re-naming of levels, fifteen possible columns of levels, and these are given in the final column of Table 2.5.1. As in Section 2.4, 'Weight' is the number of non-zero entries in the associated distance vector. Of course the columns of levels with

weight 6 can only be utilised when $s_i \geq 4$, those with weight 5 can only be utilised when $s_i \geq 3$ but the remaining columns of levels can be used for any s_i.

Table 2.5.1: Hamming distance vectors when $p = 4$

Weight	$d_{i(12)}$	$d_{i(13)}$	$d_{i(14)}$	$d_{i(23)}$	$d_{i(24)}$	$d_{i(34)}$	**Levels**
0	0	0	0	0	0	0	$[0000]'$
3	0	0	1	0	1	1	$[0001]'$
	0	1	0	1	0	1	$[0010]'$
	1	0	0	1	1	0	$[0100]'$
	1	1	1	0	0	0	$[0111]'$
4	0	1	1	1	1	0	$[0011]'$
	1	0	1	1	0	1	$[0101]'$
	1	1	0	0	1	1	$[0110]'$
5	0	1	1	1	1	1	$[0012]'$
	1	0	1	1	1	1	$[0102]'$
	1	1	1	0	1	1	$[0112]'$
	1	1	0	1	1	1	$[0120]'$
	1	1	1	1	0	1	$[0121]'$
	1	1	1	1	1	0	$[0122]'$
6	1	1	1	1	1	1	$[0123]'$

Suppose that we would like to have all the elements in $\mathbf{d}_i$ equal to some specific d_i. We may be able to achieve this by taking all the columns of levels with a particular weight, provided certain conditions are met, as we saw in Section 2.4. Here we see that for $s_i \geq 4$ we can achieve this by using d_i columns of the type $[0123]'$ and $m_i - d_i$ of the type $[0000]'$. If $s_i = 3$ then any of the columns of levels except the final one are available to use. Thus, we need to choose subsets of columns for which the corresponding $d_{i(xy)}$ all sum to d_i. These include the four columns corresponding to weight 3 (so each of the columns of types $[0001]'$, $[0010]'$, $[0100]'$ and $[0111]'$ would appear $d_i/2$ times, and the remaining $m_i - 4d_i/2$ would be of the type $[0000]'$), the three columns corresponding to weight 4 (so each of the columns of types $[0011]'$, $[0101]'$ and $[0110]'$ would appear $d_i/2$ times, and the remaining $m_i - 3d_i/2$ would be of the type $[0000]'$), and the six columns corresponding to weight 5 (so each of the columns of types $[0012]'$, $[0122]'$, $[0112]'$, $[0120]'$, $[0121]'$ and $[0122]'$ would appear $d_i/5$ times, and the remaining $m_i - 6d_i/5$ would be of the type $[0000]'$). Of course for the columns of levels corresponding to a particular weight to work we need d_i to be appropriately divisible (here by 2 or 5) so that we have an integer number of columns. We also need m_i to be large enough to include the required number of columns.

Example 2.5.1. Consider an OA$[16, 2^3 \times 4^4]$. As we saw in Example 2.3.1, $q = 0$ can be acheived for any pair of runs with $(d_1, d_2) = (2, 3)$. Hence, an optimal set of

4 runs would be achieved if the following matrix of distance vectors were possible:

$$\mathbf{D}_{24} = \begin{array}{c} \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{array} \begin{array}{c} \begin{array}{cccccc} d_{i(12)} & d_{i(13)} & d_{i(14)} & d_{i(23)} & d_{i(24)} & d_{i(34)} \end{array} \\ \begin{bmatrix} 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 \end{bmatrix} \end{array}.$$

Given $s_1 = 2$, d_1 is even, and $m_1 \geq 3d_1/2$, this can be achieved for $\mathbf{d}_1$ by taking $d_1/2 = 1$ of each of the columns with weight 4, and as $s_2 \geq 4$, we can use $d_2 = 3$ columns of weight 6 and $m_2 - d_2 = 1$ column of weight 0 to achieve the desired $\mathbf{d}_2$. This is equivalent to adjoining the run $\mathbf{r}_4 = 1100333$ to the set of 3 runs obtained for $p = 3$ in Example 2.4.1. □

A similar approach applies when $s_i = 2$ although fewer level vectors are available, but the four columns corresponding to weight 3, and the three columns corresponding to weight 4 are available and so the construction above works when $s_i = 2$.

We note that taking all runs with a given weight is not the only way to construct $\mathbf{d}_i$ with all elements equal to some specific d_i. The next example illustrates this.

Example 2.5.2. Consider the set of runs below.

$$\begin{array}{ll} \mathbf{r}_1 & 0000 \\ \mathbf{r}_2 & 0111 \\ \mathbf{r}_3 & 1201 \\ \mathbf{r}_4 & 2210 \end{array}$$

These columns of levels are subsets of those with weights 4 and 5, and the Hamming distance between each pair of runs is 3. Hence, if we wanted to construct $\mathbf{d}_i$ with all elements equal to some specific d_i that is divisible by 3, we could use $d_i/3$ copies of these columns and $m_i - 4d_i/3$ columns of type $[0000]'$, provided $s_i \geq 3$ and $m_i \geq 4d_i/3$. □

2.5.2 Further Run Construction Considerations

We will now extend the run construction methods in Section 2.4.2 for $p = 3$ to $p = 4$. When all $s_i \geq 4$, which is the case for 383 of the 23,275 OAs with $N \leq 100$, we can adjoin a set of two runs to an optimal pair. The best pair of runs to adjoin can be extended to a best set of four runs by extending each column of the form $[00]'$ to one of the form $[0000]'$ and each column of the form $[01]'$ to one of the form $[0123]'$.

Example 2.5.3. Consider a $4^1 \times 6^4$ design. When $p = 2$, we get $q = 0$ for $(d_1, d_2) = (0, 4)$, which can be represented by the following pair of runs:

	s_1=4	s_2=6			
$\mathbf{r}_1$	0	0	0	0	0
$\mathbf{r}_2$	0	1	1	1	1

Since all $s_i \geq 4$, we can extend this to $p = 4$ by making the column extensions described above. The resulting quadruple of runs is:

	s_1=4	s_2=6			
$\mathbf{r}_1$	0	0	0	0	0
$\mathbf{r}_2$	0	1	1	1	1
$\mathbf{r}_3$	0	2	2	2	2
$\mathbf{r}_4$	0	3	3	3	3

And so $(d_{1(xy)}, d_{2(xy)}) = (0, 4)$ for all pairs of runs, hence the design is optimal. □

We will now consider appending an additional m_{k+1} factors. We can append a set of $4a$ binary factors to any set of 4 runs without changing the value of q for any pair of the runs as long as $d_{k+1} = 2a$. One way to do this is to adjoin a sets of the four columns of levels $[0000]'$, $[0011]'$, $[0101]'$ and $[0110]'$.

Example 2.5.4. Consider a $2^4 \times 4^1 \times 6^4$ design. Note this is equivalent to appending $4a$ binary factors, with $a = 1$, to the $4^1 \times 6^4$ design in Example 2.5.3. Hence, we can adjoin the columns of levels described above to the optimal runs in Example 2.5.3. The resulting runs are:

	s_1=2				s_2=4	s_3=6			
$\mathbf{r}_1$	0	0	0	0	0	0	0	0	0
$\mathbf{r}_2$	0	0	1	1	0	1	1	1	1
$\mathbf{r}_3$	0	1	0	1	0	2	2	2	2
$\mathbf{r}_4$	0	1	1	0	0	3	3	3	3

The Hamming distance for all pairs of runs is $(d_{1(xy)}, d_{2(xy)}, d_{3(xy)}) = (2, 0, 4)$ which gives $q = 0$ and so the solution is optimal. □

Similarly, we can append a set of $3a$ ternary factors to any set of 4 runs without changing the value of q for any pair of the runs as long as $d_{k+1} = 2a$. One way to do this is to adjoin a sets of the three columns of levels $[0011]'$, $[0101]'$ and $[0110]'$. We note that this is equivalent to the method described above for binary factors if we remove ignore the column of levels $[0000]'$. Hence, we can discard the first column of zeros from the optimal runs constructed for a $2^4 \times 4^1 \times 6^4$ design in

Example 2.5.4 to obtain an optimal set of four runs to adjoin to a $3^3 \times 4^1 \times 6^4$ OA.

2.6 Augmenting OAs by Adjoining p Runs

In the previous section we found constructions of $\mathbf{d}_i$ such that all the elements were equal to some specific d_i by taking all the columns of levels with a particular weight, provided certain conditions were met. In this section we extend this approach to more than 3 adjoined runs.

We will assume that the runs to be adjoined are $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_p$, that $\mathbf{r}_1 = \mathbf{0}$ and we will write the runs as the rows of a matrix. We now think about the possible columns of this matrix. Consider the set of distinct levels in a given column. We let h be the number of elements in this set, $1 \leq h \leq s_i$, and we construct a vector, $\mathbf{v}$, of length h where the jth entry is the number of times the jth element in the set appears in the column. The next example illustrates this.

Example 2.6.1. Consider the table of weights and associated columns of levels in Table 2.4.2. The column of levels associated with weight 0, $[000]'$, has $h = 1$ distinct level which appears three times, hence is represented by $\mathbf{v} = [3]$. The columns of levels associated with weight 2, $[001]'$, $[010]'$ and $[011]'$, each have $h = 2$ distinct levels which appear either once or twice, and so are represented by $\mathbf{v} = [1, 2]$. Similarly, the vector $\mathbf{v}$ associated with weight 3 is $[1, 1, 1]$. □

Each vector $\mathbf{v}$ corresponds to a *partition* of p. We let ℓ be the number of distinct elements in $\mathbf{v}$, and we write $\{p_1^{t_1}, p_2^{t_2}, \ldots, p_\ell^{t_\ell}\}$ for the generic partition, where t_j is the number of times p_j appears in $\mathbf{v}$, hence $p = \sum_{j=1}^{\ell} p_j t_j$.

Example 2.6.2. Consider the vectors constructed in Example 2.6.1. The vector associated with weight 0, $\mathbf{v} = [3]$, has one distinct element, 3, which appears once and so the partition is represented by $\{3^1\}$. The vector associated with weight 2, $\mathbf{v} = [1, 2]$, has two distinct elements, 1 and 2, each of which appear once, and so the partition is represented by $\{1^1, 2^1\}$. Similarly, the partitions associated with weight 3 is $\{1^3\}$. □

Each plausible weight is associated with a unique partition of p, so each column of levels corresponds to a partition of p (equivalently, weight) but each partition of p may correspond to more than one column of levels. We now determine the weight of any distance vector corresponding to a given partition, and count the number of columns of levels that correspond to a given partition.

Lemma 2.6.1. The weight of any distance vector associated with the partition $\mathbf{p} = \{p_1^{t_1}, p_2^{t_2}, \ldots, p_\ell^{t_\ell}\}$ is

$$w_{\mathbf{p}} = \binom{p}{2} - \sum_{j=1}^{\ell} \binom{p_j}{2} t_j.$$

Proof. The number of entries in the distance vector that are equal to 0 when there are p_j entries the same in the column is $\binom{p_j}{2}$. As there are $\binom{p}{2}$ entries altogether, the result follows. □

Lemma 2.6.2. The number of columns of levels corresponding to the partition $\mathbf{p} = \{p_1^{t_1}, p_2^{t_2}, \ldots, p_\ell^{t_\ell}\}$ is

$$n_{\mathbf{p}} = \frac{\binom{p}{p_\ell}\binom{p-p_\ell}{p_\ell}\cdots\binom{p-(t_\ell-1)p_\ell}{p_\ell}\binom{p-(t_\ell)p_\ell}{p_{\ell-1}}\cdots\binom{p-\sum_{j=2}^{\ell} t_j p_j-(t_1-1)p_1}{p_1}}{\prod_{j=1}^{\ell} t_j!}.$$

Proof. The denominator is the number of re-arrangements of the t_j levels that each appear p_j times. The numerator is the number of ways of partitioning the p positions to have t_j sets with p_j positions, $1 \leq j \leq \ell$. □

Lemma 2.6.3. If all of the columns corresponding to the partition $\mathbf{p} = \{p_1^{t_1}, p_2^{t_2}, \ldots, p_\ell^{t_\ell}\}$ are adjoined to the OA then the contribution to all the entries the vector $\mathbf{d}_i$ is equal to

$$\tau_{\mathbf{p}} = \frac{n_{\mathbf{p}} \times w_{\mathbf{p}}}{\binom{p}{2}}.$$

□

We are now able to extend the method established in the previous section for $p = 3$ to any p.

Theorem 2.6.1. Let $\mathbf{p}$ be a partition of p. Suppose, for some d_i, a set of runs is constructed by taking $d_i/\tau_{\mathbf{p}}$ copies of each of the columns associated with $\mathbf{p}$, and $m_i - n_{\mathbf{p}} d_i/\tau_{\mathbf{p}}$ copies of the column associated with the partition $\{p^1\}$, then all the elements in $\mathbf{d}_i$ will be equal to d_i. This will be possible when the following conditions are met:

1. $\tau_{\mathbf{p}} \mid d_i$;
2. $m_i \geq n_{\mathbf{p}} d_i/\tau_{\mathbf{p}}$;
3. $s_i \geq \sum_{j=1}^{\ell} t_j$. □

The first condition ensures the number of required columns is an integer, the second condition ensures there are enough factors with the same number of levels to include all of the required columns, and the final condition ensures the number of levels is large enough to accommodate the number of distinct levels in the required columns.

Example 2.6.3. Consider the method used to construct the runs in Example 2.4.1. When $i = 1$ we have $s_1 = 2, m_1 = 3$; the desired d_1 was 2 and so we used the columns of levels associated with weight 2. The relevant partition is $\mathbf{p} = \{1^1, 2^1\}$, as we saw in Example 2.6.2, hence $n_{\mathbf{p}} = 3$ and $\tau_{\mathbf{p}} = 2$. Thus, the conditions in Theorem 2.6.1 are met. □

Of course, there may be other sets of columns which have distance vectors that sum to a constant value, and Example 2.4.2 shows there may be D-optimal designs where not all the elements in $\mathbf{d}_i$ are the same.

2.7 Choosing Runs From Within the OA

As discussed earlier, the optimisation methods outlined in this chapter with regards to augmenting runs are also appropriate in the context of missing runs, however we must also confirm that the OA contains pairs of runs with the required Hamming distances rather than being able to rely on all runs in the complete factorial. Similar comments can be made with regards to augmenting runs when we would like to include repeated runs in the experiment in order to estimate pure error. In that case, we would rather augment runs that are already contained in the OA to ensure replication, and so an understanding of the pairwise distance structure of the OA is required.

As Example 2.3.5 illustrates, the task of enumerating and analysing all non-isomorphic OAs with $N \leq 100$ is not computationally feasible, so instead we will focus our attention on a specific subset of these designs. We have chosen to investigate symmetric ternary OAs in $N = 18$ runs, that is OAs with $k = 1$ and $s_1 = 3$ for some $m_1 = m$. Schoen (2009) states that the class of 18-run OAs "has considerable practical importance because its members are among the smallest arrays that permit factors to be investigated at more than two levels. Indeed, there are 18-run arrays with up to seven three-level factors". Hence, we will investigate all OA$[18, 3^m]$, $3 \leq m \leq 7$.

Recall that two designs are combinatorially isomorphic if one can be constructed from the other through any combination of row, column or level permutations. Row and column permutations are also allowed under geometric isomorphism, but the levels within a factor may only be reversed. In the context of augmenting or missing runs, we are interested in the distribution of Hamming distances between all pairs of runs.

Lemma 2.7.1. All designs in the same combinatorial isomorphism class have the same distribution of Hamming distances across all pairs of runs.

Proof. Consider two OAs, OA_1 and OA_2, with the same N and $\{s_1^{m_1}, s_2^{m_2}, \ldots, s_k^{m_k}\}$. If OA_1 and OA_2 are combinatorially isomorphic then there is a mapping which takes rows to rows, columns to columns and symbols to symbols (on a column by column basis). Imagine that runs $\mathbf{r}_{1_{(1)}}$ and $\mathbf{r}_{2_{(1)}}$ in OA_1 have Hamming distance vector $(d_1, d_2, \ldots, d_k)$ and that they map to the runs $\mathbf{r}_{1_{(2)}}$ and $\mathbf{r}_{2_{(2)}}$ in OA_2 respectively. Then the entries that are the same in $\mathbf{r}_{1_{(1)}}$ and $\mathbf{r}_{2_{(1)}}$ must still be the same in $\mathbf{r}_{1_{(2)}}$ and $\mathbf{r}_{2_{(2)}}$ (since each level in OA_1 maps to exactly one level in OA_2) and levels that are different in $\mathbf{r}_{1_{(1)}}$ and $\mathbf{r}_{2_{(1)}}$ will still be different in $\mathbf{r}_{1_{(2)}}$ and $\mathbf{r}_{2_{(2)}}$. So, $\mathbf{r}_{1_{(2)}}$ and $\mathbf{r}_{2_{(2)}}$ have Hamming distance $(d_1, d_2, \ldots, d_k)$ as well. □

Since the optimisation processes for augmenting or missing runs is concerned only with Hamming distances, Lemma 2.7.1 tells us that the issue of enumerating all designs that are 'essentially different' from one another is equivalent to enumerating all combinatorial isomorphism classes. That is, we do not require the more discriminatory enumeration of geometric isomorphism classes.

Schoen (2009) enumerated all combinatorially non-isomorphic $\text{OA}[18, 3^m]$, and found that there are 4, 12, 10, 8 and 3 isomorphism classes for $m = 3$, 4, 5, 6 and 7 respectively.

We have enumerated the Hamming distance between all $\binom{18}{2} = 153$ pairs of runs for each of these classes. The distribution of these Hamming distances, along with the associated q and the efficiency of an OA with two runs augmented (Eff_A) and an OA with two runs missing (Eff_B) are given at the end of this chapter in Tables 2.7.1, 2.7.2, 2.7.4, 2.7.5 and 2.7.6 below for $m = 3$, 4, 5, 6 and 7 respectively.

Table 2.7.1: Distribution of Hamming distances between all pairs of runs for each of the 4 $\text{OA}[18, 3^3]$ combinatorial isomorphism classes

	Hamming			**Count by Class**			
q	**Distance**	**Eff**$_A$[†]	**Eff**$_B$[‡]	**1**	**2**	**3**	**4**
1	2	99.98%	99.88%	90	84	81	108
−2	3	99.91%	99.52%	42	44	45	36
4	1	99.63%	97.99%	18	24	27	–
7	0	98.84%	92.85%	3	1	–	9

[†] *Efficiency of OA plus two run design compared to theoretical bound*

[‡] *Efficiency of OA minus two run design compared to theoretical bound*

We can see that the choice of OA will depend on the objectives of the experimenter. For example, when $m = 3$, class 4 has the most pairs of runs with the best realisable q, so if we intend to augment or remove a single pairs of runs we can maximise our options by using a design from this class. Furthermore, if we intend to augment or remove more than two runs, this class is more likely to contain a set of runs that pairwise realise the best q. However, this class also has the most pairs of runs with

the worst realisable q. Thus, if the experimenter doesn't have any control over which runs might be missing (for example, we anticipate up to t runs may fail, but we do not know in advance which t runs these will be), then the experimenter might choose to avoid class 4 to minimise the chance of realising the worst q. In that case, a trade-off needs to be made between minimising the probability of obtaining a less favourable Hamming distance and maximising the probability of obtaining the most favourable Hamming distance. So, class 3, for example, might be preferred as it does not contain any pairs of runs that realise the worst q, but this choice would be made at the expense of the probability of realising the best q, which appears the least number of times in this class.

As mentioned earlier, the efficiency of augmented designs is always better than the efficiency of designs with missing runs. We note that the efficiencies reduce much faster when runs are missing than when runs are augmented. For example, when $m = 3$ the efficiencies obtained when augmenting an OA with a pair of runs range between 99.98% and 98.84% whereas in the same setting when a pair of runs is missing the efficiency ranges between 99.88% and 92.85%. Hence, the loss of efficiency given the worst pair of runs might be tolerable in the context of augmenting runs, but not acceptable in the context of missing runs.

We have seen that all designs in a given combinatorial isomorphism class necessarily have the same distribution of Hamming distances, and we can now see from Table 2.7.2 that different classes do not necessarily have different distributions. For example, when $m = 4$, classes 3 and 10 are 'essentially the same' from our perspective as they have the same distribution of Hamming distances, as do classes 5 and 8.

The trade-off we described earlier between maximising the probability of obtaining the best q versus minimising the probability of obtaining the worst q is illustrated well in class 12 of the OA$[18, 3^4]$ designs. This class has considerably more incidences of 100% efficiency than any other class, yet it also has the highest incidence of the worst realisable Hamming distance, which in this case means that an OA with two runs missing will not be estimable. While a loss in efficiency may be considered undesirable yet tolerable in some circumstances, the inability to estimate the model will clearly never be acceptable. Thus, if the experimenter anticipates the loss of up to two runs but they have no knowledge of which two runs they might be, then it is advisable to avoid classes 1, 2, 6, 7 and 12 as they all risk obtaining a singular information matrix. If, however, the experimenter has control over which runs are to be missing, or they are interested in augmenting rather than removing runs, then class 12 gives the maximum number of choices of optimal pairs to drop/augment.

Table 2.7.2: Distribution of Hamming distances between all pairs of runs for each of the 12 OA$[18, 3^4]$ combinatorial isomorphism classes

	Ham.			**Count by Class**					
q	**Dist.**	**Eff**$_A$[†]	**Eff**$_B$[‡]	**1**	**2**	**3**	**4**	**5**	**6**
0	3	100%	100%	92	104	99	87	81	114
3	2	99.86%	98.70%	36	24	27	39	45	18
−3	4	99.86%	98.70%	20	16	18	22	24	12
6	1	99.44%	93.68%	4	8	9	5	3	6
9	0	98.67%	*NA*[§]	1	1	–	–	–	3
	Ham.			**Count by Class**					
q	**Dist.**	**Eff**$_A$[†]	**Eff**$_B$[‡]	**7**	**8**	**9**	**10**	**11**	**12**
0	3	100%	100%	80	81	78	99	72	144
3	2	99.86%	98.70%	48	45	48	27	54	–
−3	4	99.86%	98.70%	24	24	25	18	27	–
6	1	99.44%	93.68%	–	3	2	9	–	–
9	0	98.67%	*NA*[§]	1	–	–	–	–	9

[†] *Efficiency of OA plus two run design compared to theoretical bound*

[‡] *Efficiency of OA minus two run design compared to theoretical bound*

[§] $\mathbf{M}_B$ *is singular*

We note that some classes have pairs of runs with Hamming distance = 0 in both $m = 3$ and $m = 4$ designs. This means that the OA contains repeated runs. We mentioned earlier that an experimenter might be interested in augmenting runs from within the OA in order to ensure replication to obtain an estimate of pure error. Of course, we could achieve even greater replication if we start with an OA that already contains repeated runs. The following example illustrates this.

Example 2.7.1. Consider a OA$[18, 3^4]$ plus 2 run design and suppose we would like to estimate pure error. We choose an OA from class 12 in Table 2.7.2 because, as indicated by the distribution of Hamming distances, it contains 9 repeated runs. An example of one such OA is given in Table 2.7.3 below. If we choose to augment a pair of repeated runs from within the OA, say the 1st and 2nd runs, then we will obtain a design containing 8 repeated pairs plus the run 0000 appearing 4 times. This design will be 98.67% efficient. Alternatively, we could augment any pair of runs with Hamming distance not equal to zero, say the 1st and 3rd runs, then we will obtain a design containing 7 repeated runs plus the runs 0000 and 0122 each appearing 3 times. This design will be 100% efficient. □

Similarly, in the case of missing runs, if replication is a high priority we could chose an OA that contains repeated runs, but if we are not sure which particular pair of runs will be missing then we risk obtaining a singular information matrix, as demonstrated in the following example.

Table 2.7.3: OA$[18, 3^4]$ from combinatorial isomorphism class 12

1	0	0	0	0
2	0	0	0	0
3	0	1	2	2
4	0	1	2	2
5	0	2	1	1
6	0	2	1	1
7	1	0	1	2
8	1	0	1	2
9	1	1	0	1
10	1	1	0	1
11	1	2	2	0
12	1	2	2	0
13	2	0	2	1
14	2	0	2	1
15	2	1	1	0
16	2	1	1	0
17	2	2	0	2
18	2	2	0	2

Example 2.7.2. Following from Example 2.7.1, suppose we are interested in an OA$[18, 3^4]$ minus 2 run design, and we would like to estimate pure error. If we have control over which two runs are missing, we might still choose an OA from class 12 for the same reasons as the previous example. In the case of missing runs, however, we can not choose a pair of repeated runs to remove as this would result in a singular matrix, so we must choose two runs with non-zero Hamming distance, say the 1st and the 3rd runs. This will result in a design with 7 repeated pairs and two non-repeated runs. If we do not have control over which runs may be missing from the design then we would need to weigh the advantages of replication against the risk of obtaining a non-estimable design. □

When $m > 4$, there are no designs with repeated runs, as we can see from Tables 2.7.4, 2.7.5 and 2.7.6. We note that when $m = 5$, classes 3 and 5 are essentially the same in terms of the distribution of Hamming distances, as are classes 6 and 8. When $m = 6$, classes 1, 2, 4 and 5 all have the same distribution of Hamming distances and they are also the only classes that do not realise the worst q. When $m = 7$ the distribution of Hamming distances are the same for all three classes.

Table 2.7.4: Distribution of Hamming distances between all pairs of runs for each of the 10 OA$[18, 3^5]$ combinatorial isomorphism classes

	Ham.			Count by Class									
q	Dist.	$\mathbf{Eff}_A$[†]	$\mathbf{Eff}_B$[‡]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
-1	4	99.99%	99.82%	72	66	81	75	81	63	57	63	99	45
2	3	99.96%	99.23%	63	69	54	60	54	72	78	72	36	90
-4	5	99.82%	96.47%	9	11	6	8	6	12	14	12	–	18
5	2	99.73%	93.72%	9	7	12	10	12	6	4	6	18	–

[†] *Efficiency of OA plus two run design compared to theoretical bound*

[‡] *Efficiency of OA minus two run design compared to theoretical bound*

Table 2.7.5: Distribution of Hamming distances between all pairs of runs for each of the 8 OA$[18, 3^6]$ combinatorial isomorphism classes

	Ham.			Count by Class							
q	Dist.	$\mathbf{Eff}_A$[†]	$\mathbf{Eff}_B$[‡]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	4	99.99%	99.69%	81	81	93	81	81	99	111	135
-2	5	99.97%	98.67%	54	54	42	54	54	36	24	–
4	3	99.87%	92.44%	18	18	14	18	18	12	8	–
-5	6	99.80%	*NA*[§]	–	–	4	–	–	6	10	18

[†] *Efficiency of OA plus two run design compared to theoretical bound*

[‡] *Efficiency of OA minus two run design compared to theoretical bound*

[§] $\mathbf{M}_B$ *is singular*

Table 2.7.6: Distribution of Hamming distances between all pairs of runs for each of the 3 OA$[18, 3^7]$ combinatorial isomorphism classes

	Ham.			Count by Class		
q	Dist.	$\mathbf{Eff}_A$[†]	$\mathbf{Eff}_B$[‡]	**1**	**2**	**3**
0	5	100%	100%	108	108	108
3	4	99.94%	*NA*[§]	27	27	27
-3	6	99.94%	*NA*[§]	18	18	18

[†] *Efficiency of OA plus two run design compared to theoretical bound*

[‡] *Efficiency of OA minus two run design compared to theoretical bound*

[§] $\mathbf{M}_B$ *is singular*

In this chapter we have shown that the optimal pair of runs to adjoin to an OA (or remove from an OA) is any pair which minimises the inner product of the corresponding rows of the model matrix. When extending this idea to larger sets of runs, we have found that any set of p (or t) runs is optimal, provided that every pairwise inner product of the corresponding rows of the model matrix is 0. For certain parameter sets it is not possible to construct runs with inner product 0. In that case the best realisable designs result from finding sets of runs in which the pairwise inner products are as small in absolute value as possible. We have given some constructions for sets of runs which meet these objectives. We have shown that when the selection of runs is restricted to those within the OA, an enumeration of combinatorial isomorphism classes is sufficient for assessing the available distribution of Hamming distances to choose from.

An obvious extension of this work is to consider the implications of including an interaction in the model. Chapter 4 is focused on addressing this issue, but as we shall see in that chapter, a thorough examination requires an understanding of the geometric structure in the underlying designs. Hence, to enable the investigation in Chapter 4, we first enumerate all geometrically non-isomorphic OA$[18, 3^m]$s in Chapter 3.

Chapter 3

Geometric Isomorphism of Ternary OAs in 18 Runs

Following the final section of Chapter 2 in which we examined all combinatorially non-isomorphic OA$[18, 3^m]$s under the assumption of a main effects only model, we would now like to obtain a set of OAs that will allow us to investigate models that include the linear-by-linear component of at least one of the two-factor interactions. Cheng and Ye (2004) note that "for designs with quantitative factors, level permutation of one or more factors in a design matrix could result in different geometric structures, and, thus, different design properties". Hence, in order to facilitate the investigation in Chapters 4 and 5 regarding models including linear-by-linear interactions, we first require an enumeration of geometric isomorphism classes, rather than the combinatoric isomorphism classes studied in Chapter 2.

In this chapter we will enumerate all geometrically non-isomorphic OA$[18, 3^m]$s. We begin with a discussion of the distinction between geometric and combinatorial isomorphism in the context of OAs that have been altered by adjoining or removing some runs. We then describe an exhaustive enumeration process in Section 3.2.

3.1 Geometric Versus Combinatorial Isomorphism

In Chapter 2 we examined OA plus p run designs and OA minus t run designs under the assumption of a main effects only model. We showed that the performance of these designs depends on the pairwise Hamming distances of the runs that have been adjoined or removed. Hence, if we require these runs to come from within the OA (which is necessarily the case when runs are missing), then we require an understanding of the structure of the OAs in the design space. We showed that two designs in the same combinatorial isomorphism class have the same distribution

of pairwise Hamming distances, hence combinatorial isomorphism is a sufficient method for discriminating between OAs under a main effects only model.

In Chapter 4, we will relax the assumption of a main effects only model and investigate the implications of including a linear-by-linear interaction in the model. As noted in Cheng and Ye (2004), we will show that combinatorial isomorphism is not sufficient to classify OAs under this model, rather we require geometric isomorphism to determine whether a pair of OAs are essentially the same. The following example illustrates why this is the case.

Example 3.1.1. Suppose we construct a 16-run design with three ternary factors by removing a pair of runs with Hamming distance 2 from an OA$[18, 3^3]$. From Chapter 2, we know that the determinant of the information matrix of this design, assuming a main effects only model, is $|\mathbf{M}_B| = (-1)^2 18^{7-2} |\mathbf{\Omega}_B|$ where $\mathbf{\Omega}_B = \begin{bmatrix} 7-18 & \mathbf{b}_1.\mathbf{b}_2' \\ \mathbf{b}_1.\mathbf{b}_2' & 7-18 \end{bmatrix}$ and $\mathbf{b}_1.\mathbf{b}_2' = 7-3\times 2 = 1$, hence $|\mathbf{M}_B| = 18^5 \times ((-11)^2 - 1^2) = 226,748,160$, which is 99.8815% efficient. Table 3.1.1 below gives two OA$[18, 3^3]$s from the first combinatorial isomorphism class. Recall from Table 2.7.1 in the previous chapter that all designs in this class have a Hamming distance of 2 in 90 pairs of runs, hence each of these OAs contain 90 pairs of runs that could be removed to give a design with $|\mathbf{M}_B| = 226,748,160$. We will focus on the first 5 runs (notice that this subset of runs is the same in both designs), which we label $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$ and $\mathbf{r}_5$ respectively. The pairs ($\mathbf{r}_1$, $\mathbf{r}_3$), ($\mathbf{r}_1$, $\mathbf{r}_4$), ($\mathbf{r}_1$, $\mathbf{r}_5$), ($\mathbf{r}_2$, $\mathbf{r}_3$), ($\mathbf{r}_2$, $\mathbf{r}_4$), ($\mathbf{r}_2$, $\mathbf{r}_5$) and ($\mathbf{r}_4$, $\mathbf{r}_5$) are 7 pairs of runs with Hamming distance 2 (note that the other pairwise Hamming distances amongst these 5 runs are not equal to 2).

Now consider a model that contains the linear-by-linear component of one two-factor interaction. Table 3.1.2 gives the efficiency for each of the possible two-factor interactions to be included in the model when each of the pairs of runs mentioned above are removed. Notice that the range of efficiencies are very different between the two designs. Most notably, the efficiencies in OA_1 generally tend to be better than in OA_2. Under a main effects only model, we would consider these two designs to be essentially same since they belong to the same combinatorial isomorphism class, hence they have the same distribution of pairwise Hamming distances. Table 3.1.2 illustrates that the same Hamming distances do not necessarily lead to the same $|\mathbf{M}_B|$ when an interaction term is included in the model, hence the distribution of Hamming distances (that is, combinatorial isomorphism) is not sufficient to discriminate between designs for these larger models. □

The previous example examines the implications of including an interaction term in the model when runs are missing from an OA. We will now consider the situation when runs are adjoined to an OA.

Table 3.1.1: Two OA$[18, 3^3]$s from combinatorial isomorphism class 1

	OA_1				OA_2		
1	0	0	1	1	0	0	1
2	0	0	1	2	0	0	1
3	0	1	0	3	0	1	0
4	0	1	2	4	0	1	2
5	0	2	0	5	0	2	0
6	0	2	2	6	0	2	2
7	1	0	0	7	1	0	2
8	1	0	2	8	1	0	2
9	1	1	0	9	1	1	0
10	1	1	2	10	1	1	1
11	1	2	1	11	1	2	0
12	1	2	1	12	1	2	1
13	2	0	0	13	2	0	0
14	2	0	2	14	2	0	0
15	2	1	1	15	2	1	1
16	2	1	1	16	2	1	2
17	2	2	0	17	2	2	1
18	2	2	2	18	2	2	2

Table 3.1.2: Efficiency† of OA minus 2 run designs when one linear-by-linear interaction is included in the model for each of the OA$[18, 3^3]$s in Table 3.1.1

OA_1		**Pair of factors in the interaction**		
		(F_1, F_2)	(F_1, F_3)	(F_2, F_3)
Missing runs	$(\mathbf{r}_1, \mathbf{r}_3)$	96.8396%	97.1540%	100%
	$(\mathbf{r}_1, \mathbf{r}_4)$	96.8396%	97.1540%	100%
	$(\mathbf{r}_1, \mathbf{r}_5)$	94.9555%	97.1540%	97.1540%
	$(\mathbf{r}_2, \mathbf{r}_3)$	96.8396%	97.1540%	100%
	$(\mathbf{r}_2, \mathbf{r}_4)$	96.8396%	97.1540%	100%
	$(\mathbf{r}_2, \mathbf{r}_5)$	94.9555%	97.1540%	97.1540%
	$(\mathbf{r}_4, \mathbf{r}_5)$	96.1631%	94.2942%	97.1540%
OA_2		**Pair of factors in the interaction**		
		(F_1, F_2)	(F_1, F_3)	(F_2, F_3)
Missing runs	$(\mathbf{r}_1, \mathbf{r}_3)$	91.9839%	95.6128%	94.2942%
	$(\mathbf{r}_1, \mathbf{r}_4)$	93.2848%	93.5366%	94.2942%
	$(\mathbf{r}_1, \mathbf{r}_5)$	92.3980%	95.6128%	89.6901%
	$(\mathbf{r}_2, \mathbf{r}_3)$	91.9839%	95.6128%	94.2942%
	$(\mathbf{r}_2, \mathbf{r}_4)$	93.2848%	93.5366%	94.2942%
	$(\mathbf{r}_2, \mathbf{r}_5)$	92.3980%	95.6128%	89.6901%
	$(\mathbf{r}_4, \mathbf{r}_5)$	93.2848%	91.4107%	89.6901%

† *Relative to the best known OA$[18, 3^3]$ minus 2 run design*

Example 3.1.2. Suppose we intend to adjoin two runs with Hamming distance 2 to an OA$[18, 3^3]$. From Chapter 2, we know that the determinant of the information matrix of this design, assuming a main effects only model, is $|\mathbf{M}_A| = 18^{7-2}|\mathbf{\Omega}_A|$ where $\mathbf{\Omega}_A = \begin{bmatrix} 18+7 & \mathbf{a}_1.\mathbf{a}_2' \\ \mathbf{a}_1.\mathbf{a}_2' & 18+7 \end{bmatrix}$ and $\mathbf{a}_1.\mathbf{a}_2' = 7 - 3 \times 2 = 1$, hence $|\mathbf{M}_A| = 18^5 \times (25^2 - 1^2) = 1,179,090,432$, which is 99.9771% efficient. Following from Example 3.1.1, we will focus on the first 5 runs of the two OA$[18, 3^3]$s given in Table 3.1.1. As we noted in that example, the pairs ($\mathbf{r}_1$, $\mathbf{r}_3$), ($\mathbf{r}_1$, $\mathbf{r}_4$), ($\mathbf{r}_1$, $\mathbf{r}_5$), ($\mathbf{r}_2$, $\mathbf{r}_3$), ($\mathbf{r}_2$, $\mathbf{r}_4$), ($\mathbf{r}_2$, $\mathbf{r}_5$) and ($\mathbf{r}_4$, $\mathbf{r}_5$) are 7 pairs with Hamming distance 2, hence they are examples of OA plus 2 designs where $|\mathbf{M}_A| = 1,179,090,432$.

Now consider a model that contains the linear-by-linear component of a two-factor interaction. Table 3.1.3 gives the determinant of the information matrix for each of the possible two-factor interactions to be included in the model when each of the pairs of runs mentioned above are adjoined to each OA. Notice that the range of efficiencies are very different between the two designs, as was the case in the context of missing runs. The efficiencies in OA_1 generally tend to be better than OA_2, as was the case in the previous example. □

Table 3.1.3: Efficiency† of OA plus 2 run designs when one linear-by-linear interaction is included in the model for each of the OA$[18, 3^3]$s in Table 3.1.1

OA_1		**Pair of factors in the interaction**		
		(F_1, F_2)	(F_1, F_3)	(F_2, F_3)
Adjoined runs	($\mathbf{r}_1$, $\mathbf{r}_3$)	96.5552%	98.6685%	97.6097%
	($\mathbf{r}_1$, $\mathbf{r}_4$)	96.5552%	98.6685%	97.6097%
	($\mathbf{r}_1$, $\mathbf{r}_5$)	97.3843%	98.6685%	98.6685%
	($\mathbf{r}_2$, $\mathbf{r}_3$)	96.5552%	98.6685%	97.6097%
	($\mathbf{r}_2$, $\mathbf{r}_4$)	96.5552%	98.6685%	97.6097%
	($\mathbf{r}_2$, $\mathbf{r}_5$)	97.3843%	98.6685%	98.6685%
	($\mathbf{r}_4$, $\mathbf{r}_5$)	96.8719%	99.7292%	98.6685%
OA_2		**Pair of factors in the interaction**		
		(F_1, F_2)	(F_1, F_3)	(F_2, F_3)
Adjoined runs	($\mathbf{r}_1$, $\mathbf{r}_3$)	92.8137%	96.9298%	92.0403%
	($\mathbf{r}_1$, $\mathbf{r}_4$)	92.4488%	97.7375%	92.0403%
	($\mathbf{r}_1$, $\mathbf{r}_5$)	92.8762%	96.9298%	93.6036%
	($\mathbf{r}_2$, $\mathbf{r}_3$)	92.8137%	96.9298%	92.0403%
	($\mathbf{r}_2$, $\mathbf{r}_4$)	92.4488%	97.7375%	92.0403%
	($\mathbf{r}_2$, $\mathbf{r}_5$)	92.8762%	96.9298%	93.6036%
	($\mathbf{r}_4$, $\mathbf{r}_5$)	92.4488%	98.4078%	93.6036%

†*Relative to the best known OA*$[18, 3^3]$ *plus 2 internal run design*

Although the two OAs in Table 3.1.1 belong to the same combinatorial isomorphism class, they do not belong to the same geometric isomorphism class, which is

why their efficiencies differ so much when an interaction is included in the model. Thus, as illustrated by both of the previous two examples, we require geometric isomorphism to classify the performance of altered OAs for a model with main effects plus some linear-by-linear interaction terms. Hence, in order to facilitate the investigation in Chapters 4 and 5 regarding models including linear-by-linear interactions, we first require a complete enumeration which gives one representative from each of the geometric isomorphism classes in the design space.

3.2 Complete Enumeration

Most of the literature concerning the enumeration of OAs focuses on combinatorial isomorphism. Tsai et al. (2000) enumerate many geometrically non-isomorphic OA$[18, 3^m]$s with a columnwise procedure in which two designs are said to be in the same "design family" (that is, the same geometric isomorphism class) if a design criterion, $Q(\Gamma^{(m)})$, is the same for both designs. They state that "Although it is possible that some designs will be missed, there is no certain way to avoid this without looking at every possible design, when the problem soon becomes unmanageable", and give one representative from each of the 13, 129, 320, 440 and 253 geometric isomorphism classes they find for $m = 3, 4, 5, 6$ and 7 respectively.

In an unpublished manuscript, Tsai et al. (2006a) count 13, 133, 332, 478, and 284 geometrically non-isomorphic OA$[18, 3^m]$s for $m = 3, 4, 5, 6$ and 7 respectively, and they claim that this is a complete enumeration. Pang and Liu (2011) give an algorithm for checking geometric isomorphism, which they also claim to be a complete enumeration. They report the number of geometric isomorphism classes as 13, 137, 333, 485 and 291 for $m = 3, 4, 5, 6$ and 7 respectively. Neither of these sets of authors were able to publish the actual OAs.

Chapter 10 of Kaski and Östergård (2006) stresses the importance of using multiple methods to address enumeration problems. They suggest systems for checking the validity of computational results such as "consistency checking" in which "some relationships among the results − preferably involving the whole computation − are checked" (p299). Hence, since two independent enumerations have given inconsistent results, we thought it prudent to either confirm or otherwise the reported counts of geometric isomorphism classes by enumerating all of them ourselves. Additionally, validity checking aside, we were also faced with the practical problem of needing a representative of each of the classes in order to carry out an analysis of models including linear-by-linear interactions, thus the requirement to develop our own enumeration method was unavoidable.

We utilise the software package Nauty (McKay and Piperno (2014)) for pairwise comparisons. This method is distinct from those employed by previous authors. Each pairwise comparison is practically instant, which enables us to look at every possible design, hence we exhaustively enumerate all geometrically non-isomorphic OA$[18, 3^m]$s without missing any designs. We emphasise that our goal is simply to determine the number of geometric isomorphism classes and obtain one representative from each in order to analyse them in subsequent chapters, hence this chapter is only concerned with ensuring that the enumeration is exhaustive and can be carried out in a practical time frame, rather than in the fastest time possible.

Our method for enumerating all geometrically non-isomorphic m-factor ternary OAs in 18 runs involves systematically appending columns to each of the geometrically non-isomorphic designs in $m-1$ factors. As noted by Schoen (2009), the largest ternary 18-run OAs have $m = 7$ factors.

A ternary design in $N = 18$ runs will satisfy the properties of an OA of strength 2 if the $3^2 = 9$ pairs of level combinations appear exactly $18/9 = 2$ times in every pair of columns. When $3^m \leq 18$, there is only one design to consider, that is, all 3^m runs in the full factorial, each repeated $18/3^m$ times. Hence, we will begin by constructing all non-isomorphic designs with $m = 3$ factors since this is the smallest value of m where $3^m > 18$.

3.2.1 Three Factors

Recall that geometric isomorphism is invariant to row permutations, hence we always order the rows of an OA lexicographically. In the case of OA$[18, 3^m]$s, this means we can assume without loss of generality, that the first two columns of the OA are:

$$\mathbf{c}_1' = [\ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 2\ \ 2\ \ 2\ \ 2\ \ 2\ \ 2\ ]$$

$$\mathbf{c}_2' = [\ 0\ \ 0\ \ 1\ \ 1\ \ 2\ \ 2\ \ 0\ \ 0\ \ 1\ \ 1\ \ 2\ \ 2\ \ 0\ \ 0\ \ 1\ \ 1\ \ 2\ \ 2\ ]$$

With these two columns fixed, the task of enumerating all OAs with $m = 3$ factors has two stages:

1. Enumerate all possible third columns, which we denote by $\mathbf{c}_3$, such that all 9 pairs of treatment combinations appear exactly twice in each of the pairs of columns $[\mathbf{c}_1\ \ \mathbf{c}_3]$ and $[\mathbf{c}_2\ \ \mathbf{c}_3]$.

2. Compare all of the OAs defined by $[\mathbf{c}_1\ \ \mathbf{c}_2\ \ \mathbf{c}_3]$ for each $\mathbf{c}_3$ constructed in the previous step, and discard any geometrically isomorphic designs.

We will now outline a method for enumerating all possible third columns, $\mathbf{c}_3$, and we will then use the software package Nauty (McKay and Piperno (2014)) to discard isomorphic designs.

All Possible Third Columns

We can represent the entries in $\mathbf{c}_3$ as a contingency table of the number of levels within $\mathbf{c}_3$ versus the pairs of levels in $[\mathbf{c}_1 \ \mathbf{c}_2]$ across the rows. The following example illustrates this.

Example 3.2.1. Consider the following potential third column:

$$\mathbf{c}_3' = [\ 0 \ \ 1 \ \ 1 \ \ 2 \ \ 0 \ \ 2 \ \ 2 \ \ 2 \ \ 0 \ \ 0 \ \ 1 \ \ 1 \ \ 0 \ \ 1 \ \ 1 \ \ 2 \ \ 0 \ \ 2 \ ]$$

The number of levels in this column that appear in the same row as each of the pairs of levels in $[\mathbf{c}_1 \ \mathbf{c}_2]$ are given in the Table 3.2.1 below. □

Table 3.2.1: Contingency table of levels of $\mathbf{c}_3$ vs. pairs of levels in $[\mathbf{c}_1 \ \mathbf{c}_2]$

$\mathbf{c}_3$	Pairs in $[\mathbf{c}_1 \ \mathbf{c}_2]$								
	00	**01**	**02**	**10**	**11**	**12**	**20**	**21**	**22**
0	1	0	1	0	2	0	1	0	1
1	1	1	0	0	0	2	1	1	0
2	0	1	1	2	0	0	0	1	1

This contingency table can be thought of as three square tables, where each square is associated with a level of $\mathbf{c}_1$, as indicated by the dashed vertical lines in Table 3.2.1. Notice that the marginal totals for each square is equal to 2 for every row and column, as highlighted in Table 3.2.2 below. The row totals for each square ensure that each level of $\mathbf{c}_1$ appears with each level of $\mathbf{c}_3$ exactly twice, and the column totals ensure that each level of $\mathbf{c}_1$ appears with each level of $\mathbf{c}_2$ exactly twice (which is implicit given the specification of $\mathbf{c}_1$ and $\mathbf{c}_2$).

Table 3.2.2: Squares within Table 3.2.1

$\mathbf{c}_1 = 0$				
	$\mathbf{c}_2$			
$\mathbf{c}_3$	**0**	**1**	**2**	**Total**
0	1	0	1	2
1	1	1	0	2
2	0	1	1	2
Total	2	2	2	6

$\mathbf{c}_1 = 1$				
	$\mathbf{c}_2$			
$\mathbf{c}_3$	**0**	**1**	**2**	**Total**
0	0	2	0	2
1	0	0	2	2
2	2	0	0	2
Total	2	2	2	6

$\mathbf{c}_1 = 2$				
	$\mathbf{c}_2$			
$\mathbf{c}_3$	**0**	**1**	**2**	**Total**
0	1	0	1	2
1	1	1	0	2
2	0	1	1	2
Total	2	2	2	6

Finally, for the columns $[\mathbf{c}_1 \ \mathbf{c}_2 \ \mathbf{c}_3]$ to satisfy the properties of an OA, we must ensure that each level of $\mathbf{c}_2$ appears with each level of $\mathbf{c}_3$ exactly twice. This

can be done by superimposing all three squares and calculating the sum of the entries in each of the cells, as illustrated in Table 3.2.3 below. All counts in this superimposed square must equal 2 to ensure that each level of $\mathbf{c}_2$ appears with each level of $\mathbf{c}_3$ exactly twice, as is the case here.

Table 3.2.3: Squares within Table 3.2.2 superimposed

	$\mathbf{c}_2$		
$\mathbf{c}_3$	**0**	**1**	**2**
0	1+0+1=2	0+2+0=2	1+0+1=2
1	1+0+1=2	1+0+1=2	0+2+0=2
2	0+2+0=2	1+0+1=2	1+0+1=2

The first step in enumerating all possible vectors of $\mathbf{c}_3$ is to enumerate all possible squares. The required marginal totals of 2 give us an upper bound on the values in these squares, hence we can restrict our search for the values in each cell to 0, 1 and 2. So, there are three possible entries for each of the 9 cells, hence there are $3^9 = 19,683$ potential squares to be considered. After considering each of these squares and discarding any that do not meet the required marginal totals, we reduce the number of squares to be considered to 21, as given in Table 3.2.4.

Table 3.2.4: All valid squares

2	0	0	0	0	2	0	2	0	0	2	0	2	0	0	0	2	1	1	0
0	0	2	0	2	0	0	0	2	2	0	0	0	2	0	2	0	0	0	2
0	2	0	2	0	0	2	0	0	0	0	2	0	0	2	0	2	0	1	1	0
1	0	1	0	2	0	1	1	0	0	0	2	2	0	0	0	1	1	0	1	1
1	0	1	1	0	1	1	1	0	1	1	0	0	1	1	2	0	0	0	1	1
0	2	0	1	0	1	0	0	2	1	1	0	0	1	1	0	1	1	2	0	0
1	0	1	1	1	0	1	1	0	0	1	1	0	1	1	1	0	1	1	0	1
0	2	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	0	0	1	1
1	0	1	1	0	1	0	1	1	1	1	0	1	0	1	0	1	1	1	1	0

We now need to consider all $21^3 = 9,261$ sets of three of these squares. By exhaustively considering each of these 9,261 sets and discarding any in which the cell counts within the superimposed square are not all equal to 2, we reduce the number of sets to 132. Each of the 132 sets is associated with a column vector, $\mathbf{c}_3$. Of these 132 column vectors, 6 remain unchanged when their levels are reversed after the resulting rows of $[\mathbf{c}_1 \ \mathbf{c}_2 \ \mathbf{c}_3]$ are ordered lexicographically. The remaining 126 columns can be split into two groups with a one-to-one mapping when the levels are reversed. The next two examples illustrate these two types of columns.

Example 3.2.2. Consider the column $\mathbf{c}_3$ in OA_1 in Table 3.2.5. If we reverse the levels of this column we obtain OA_2. We then re-order the rows of OA_2

lexicographically to obtain OA_3, which is exactly the same as OA_1. Hence, this particular instance of $\mathbf{c}_3$ remains unchanged when its levels are reversed. □

Table 3.2.5: $\mathbf{c}_3$ remains unchanged after reversing its levels

OA_1			→	OA_2			→	OA_3		
$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$		$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$		$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$
0	0	1		0	0	1		0	0	1
0	0	1		0	0	1		0	0	1
0	1	0		0	1	2		0	1	0
0	1	2		0	1	0		0	1	2
0	2	0		0	2	2		0	2	0
0	2	2		0	2	0		0	2	2
1	0	0		1	0	2		1	0	0
1	0	2		1	0	0		1	0	2
1	1	0		1	1	2		1	1	0
1	1	2		1	1	0		1	1	2
1	2	1		1	2	1		1	2	1
1	2	1		1	2	1		1	2	1
2	0	0		2	0	2		2	0	0
2	0	2		2	0	0		2	0	2
2	1	1		2	1	1		2	1	1
2	1	1		2	1	1		2	1	1
2	2	0		2	2	2		2	2	0
2	2	2		2	2	0		2	2	2

Example 3.2.3. Consider the column $\mathbf{c}_3$ in OA_1 in Table 3.2.6. If we reverse the levels of this column we obtain OA_2. We then re-order the rows of OA_2 lexicographically to obtain OA_3, which is *not* exactly the same as OA_1. Since the operations performed to construct OA_3 from OA_1 are allowed under geometric isomorphism, these two designs are in the same geometric isomorphism class. Hence, we consider the two columns labelled $\mathbf{c}_3$ in each of OA_1 and OA_3 to be isomorphic. □

We require only one representative from each geometric isomorphism class, so we can discard the $126/2 = 63$ columns that can be generated by reversing the levels of another column, hence we are left with $63 + 6 = 69$ potential vectors for $\mathbf{c}_3$. These 69 columns are given in Table A5 of the Appendix.

Geometric Isomorphism

Although we were able to discard some vectors in the enumeration of all potential columns of $\mathbf{c}_3$ above, this process alone is not enough to guarantee the remaining 69 columns will each belong to their own isomorphic class when appended to $[\mathbf{c}_1 \ \mathbf{c}_2]$.

Table 3.2.6: $\mathbf{c}_3$ changes after reversing its levels

OA_1			$\rightarrow$	OA_2			$\rightarrow$	OA_3		
$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$		$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$		$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$
0	0	0		0	0	2		0	0	0
0	0	2		0	0	0		0	0	2
0	1	1		0	1	1		0	1	0
0	1	2		0	1	0		0	1	1
0	2	0		0	2	2		0	2	1
0	2	1		0	2	1		0	2	2
1	0	1		1	0	1		1	0	0
1	0	2		1	0	0		1	0	1
1	1	0		1	1	2		1	1	0
1	1	2		1	1	0		1	1	2
1	2	0		1	2	2		1	2	1
1	2	1		1	2	1		1	2	2
2	0	0		2	0	2		2	0	1
2	0	1		2	0	1		2	0	2
2	1	0		2	1	2		2	1	1
2	1	1		2	1	1		2	1	2
2	2	2		2	2	0		2	2	0
2	2	2		2	2	0		2	2	0

Suppose we intend to establish whether a pair of designs, OA_1 and OA_2, are non-isomorphic via exhaustive checking of all allowable operations under geometric isomorphism. We must first generate all designs in the geometrical isomorphism class of OA_1, say, and then compare each of these designs to the second design, OA_2. If none of the generated designs match OA_2 then we can conclude that OA_1 and OA_2 are geometrically non-isomorphic, otherwise they belong to the same isomorphism class. For a single design, the construction of all designs in the same isomorphism class would require reversing the levels of all $\sum_{i=0}^{3}\binom{3}{i} = 2^3 = 8$ possible subsets of columns for each of the $3! = 6$ permutations of columns. Hence, there are $8 \times 6 = 48$ designs in an isomorphism class for three-factor designs, assuming the rows of each design are ordered lexicographically. Since we have narrowed our search down to 69 OAs, there are potentially $\binom{69}{2} = 2,346$ pairs of designs to consider, each of which requires the construction of 48 designs. While this task may be manageable for three-factor designs, this naive approach quickly becomes computationally prohibitive when $m > 3$, as we illustrate below.

An m-factor design has $\sum_{i=0}^{m}\binom{m}{i} = 2^m$ possible subsets of columns to reverse for each of the $m!$ permutations of columns. Hence, a single pair of designs would require the construction of $2^m m!$ designs to determine isomorphism exhaustively, assuming the rows of each design are ordered lexicographically. In the case of, say, $m = 7$, there are up to $2^7 7! = 645,120$ designs in each isomorphism class,

hence if we have x designs to be classified into isomorphism classes, then there are potentially $\binom{x}{2}$ pairs to be considered, each of which requires the construction of $645,120$ designs.

Of course the calculations above describe a worst-case scenario. The implementation of an exhaustive method could be refined by, for example, abandoning the construction of isomorphic designs as soon as the isomorphism between the original pair has been established. Even after such adjustments have been made, however, the computational task remains impractically large for our purposes, hence we require a more efficient method for determining the isomorphism properties within a group of potential designs. We use Nauty (McKay and Piperno (2014)), a program for computing automorphism groups of graphs and digraphs, which allows us to determine the isomorphism between a pair of designs practically instantly. Algorithm 3.1 below gives details on how we use Nauty's pairwise comparison to filter out any isomorphic OAs within a set.

Algorithm 3.1: Keep one representative from each geometric isomorphism class from a set of OAs

Input: Set of i OAs, $A = (OA_1, OA_2, \ldots, OA_i)$
Output: Set of j non-isomorphic OAs, $B = (OA_1, OA_2, \ldots, OA_j)$, $j \leq i$

```
Initialise B ← empty set
for each OA_A in A do
    matched ← FALSE
    for each OA_B in B do
        Execute Nauty to determine isomorhpism between OA_A and OA_B
        if OA_A and OA_B are isomorphic then
            matched ← TRUE
            Exit inner loop
    if matched = FALSE then
        Append OA_A to B
Output B
```

A Python implementation of Algorithm 3.1 on the set of 69 three-factor OAs identified previously took approximately 35 seconds. We chose to use Python to be consistent with the coding used throughout this project, for which Python is well suited. Since the motivation of this chapter is simply to enumerate all geometrically non-isomorphic designs so that we are able analyse them in subsequent chapters, we have found this implementation to be sufficient for our purposes. We note, however, that Nauty is written in C, hence the efficiency of this algorithm could optimised with a C implementation by calling the functions in Nauty directly.

We determined that there are 13 geometric isomorphism classes, which is consistent with the literature. Both Tsai et al. (2000) and Pang and Liu (2011) give one

representative from each of the 13 geometric isomorphism classes they found, and we have confirmed that exactly one of our designs is geometrically isomorphic to exactly one of each of these sets of designs. For convenience, we have arbitrarily labelled these classes from 1 to 13, as given in Table A5 of the Appendix.

We note that despite there being 48 allowable permutations under geometric isomorphism for three-factor designs, the resulting designs will not necessarily be distinct, as illustrated in Example 3.2.2. This explains why the complete enumeration of all possible $\mathbf{c}_3$ in Section 3.2.1 only uncovered 132 columns rather than $13 \times 48 = 624$ which we would expect if all 48 designs within each isomorphism class were distinct. To confirm this finding, we have exhaustively generated all 48 designs for each of the 13 isomorphism classes and confirmed that when duplicates are removed only 132 remain across all classes.

3.2.2 Four Factors

We will enumerate all four-factor OAs by appending a column to one representative from each of the 13 three-factor classes. The following Lemma shows that this will be sufficient for finding all geometrically non-isomorphic four-factor designs in the design space provided that the set of potential columns we consider is complete.

Lemma 3.2.1. Let $\{\mathbf{c}_m\}$ be a complete set of all possible columns to append to an OA with $m-1$ factors. Let OA_1 and OA_2 be geometrically isomorphic OAs with $m-1$ factors, then every m-factor OA constructed by appending a column from $\{\mathbf{c}_m\}$ to OA_1 will be geometrically isomorphic to an m-factor OA that can be constructed by appending a column from $\{\mathbf{c}_m\}$ to OA_2.

Proof. Let $P(\cdot)$ denote the geometric operations that transforms OA_1 to OA_2. Suppose we append a particular column from $\{\mathbf{c}_m\}$, $\mathbf{c}_{m_1}$, to OA_1 and then perform $P(\cdot)$ on the resulting m-factor design so that the first $m-1$ columns now resemble OA_2. If there are any row permutations in $P(\cdot)$, then the corresponding entries in $\mathbf{c}_{m_1}$ will also be permuted. We will label this permuted column $\mathbf{c}_{m_2}$. Since $\{\mathbf{c}_m\}$ is a complete set of all possible columns, $\mathbf{c}_{m_2}$ must be in $\{\mathbf{c}_m\}$. Hence, the OA constructed by appending $\mathbf{c}_{m_1}$ to OA_1 is geometrically isomorphic to the OA constructed by appending $\mathbf{c}_{m_2}$ to OA_2. □

Following Lemma 3.2.1, the first step in enumerating all geometric isomorphism classes for four-factor OAs is the generation of a complete set of fourth columns, $\{\mathbf{c}_4\}$. We can use the results from Section 3.2.1 in which we enumerated all possible vectors of $\mathbf{c}_3$ as a starting point for finding all possible vectors of $\mathbf{c}_4$, however we will also need to consider some additional permutations to ensure that this set is complete. The following example illustrates why this is the case.

Example 3.2.4. The three-factor OA comprised of columns $[\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3]$ in Table 3.2.7 below is a representative of the first three-factor isomorphism class, where $\mathbf{c}_3$ is column #7 from Table A5 of the Appendix. The two potential columns to be appended, $\mathbf{c}_{4_1}$ and $\mathbf{c}_{4_2}$, are identical except for the pairs of rows (9, 10) and (17, 18) in which the order of the entries differs between the two columns. It can be shown that the pair of four-factor OAs $[\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3\ \mathbf{c}_{4_1}]$ and $[\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3\ \mathbf{c}_{4_2}]$ are geometrically non-isomorphic, yet $\mathbf{c}_{4_1}$ and $\mathbf{c}_{4_2}$ have the same contingency table of the levels of each of these columns versus the pairs of levels in $[\mathbf{c}_1\ \mathbf{c}_2]$, as given in Table 3.2.8. □

Table 3.2.7: Append $\mathbf{c}_4$ to a three-factor OA

	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_{4_1}$	$\mathbf{c}_{4_2}$
1	0	0	1	0	0
2	0	0	1	1	1
3	0	1	0	0	0
4	0	1	2	2	2
5	0	2	0	1	1
6	0	2	2	2	2
7	1	0	0	2	2
8	1	0	2	1	1
9	1	1	0	0	1
10	1	1	2	1	0
11	1	2	1	0	0
12	1	2	1	2	2
13	2	0	0	2	2
14	2	0	2	0	0
15	2	1	1	1	1
16	2	1	1	2	2
17	2	2	0	1	0
18	2	2	2	0	1

Table 3.2.8: Contingency table of levels of $\mathbf{c}_4$ (either $\mathbf{c}_{4_1}$ or $\mathbf{c}_{4_2}$) vs. pairs of levels in $[\mathbf{c}_1\ \mathbf{c}_2]$

$\mathbf{c}_4$	Pairs in $[\mathbf{c}_1\ \mathbf{c}_2]$								
	00	**01**	**02**	**10**	**11**	**12**	**20**	**21**	**22**
0	1	1	0	0	1	1	1	0	1
1	1	0	1	1	1	0	0	1	1
2	0	1	1	1	0	1	1	1	0

As the previous example illustrates, it is not sufficient to consider a single column associated with each of the 132 sets of squares identified in Section 3.2.1 since two non-isomorphic designs can be constructed from the same set of squares. This was not an issue previously when $m = 3$ since the order of the levels in $\mathbf{c}_3$ within each

pair of $[\mathbf{c}_1 \ \mathbf{c}_2]$ is irrelevant. For example, note that rows 9 and 10 of Table 3.2.7 are both associated with the pair of levels (1, 1) in $[\mathbf{c}_1 \ \mathbf{c}_2]$, and the levels of $\mathbf{c}_3$ in these rows are 0 and 2 respectively. Suppose we switch the order of these two entries in $\mathbf{c}_3$. When the rows of the entire design are ordered lexicographically, these two rows will be swapped, hence the initial order does not influence the final design. The same logic cannot be applied to pairs of entries in $\mathbf{c}_4$, however, as this column is not necessarily used in the lexicographical ordering of the rows[1]. Hence, for each of the 132 sets of squares (contingency tables) enumerated in Section 3.2.1, we need to consider all potential permutations within each pair of $[\mathbf{c}_1 \ \mathbf{c}_2]$. In a worse-case scenario, this means we will need to consider $2^9 = 512$ representations for a single contingency table. However in some cases some pairs of entries are equal, hence permuting within these pairs will result in a duplicated column.

We exhaustively constructed all permutations within the 9 pairs for each of the 132 sets of squares which resulted in 23,436 plausible columns of $\mathbf{c}_4$ after duplicates were removed. We note that this matches the number of columns quoted in Tsai et al. (2000) for their columnwise procedure, as well as the number of "balanced columns" used in Pang and Liu (2011)'s algorithm. We can discard exactly half of these columns as they have a one-to-one mapping with the retained columns when the levels over the entire column are reversed. Hence, we are left with 11,718 columns in the complete set of $\{\mathbf{c}_4\}$ to append to one representative from each of the 13 three-factor isomorphism classes. But before we use Algorithm 3.1 to determine the isomorphism of all of these designs, we need to determine whether they satisfy the properties of an OA.

Since all of the 11,718 columns in the complete set $\{\mathbf{c}_4\}$ are representatives of one of the sets of squares enumerated in the Section 3.2.1, we know that all pairs of levels will appear equally often in the pairs of columns $[\mathbf{c}_1 \ \mathbf{c}_4]$ and $[\mathbf{c}_2 \ \mathbf{c}_4]$. What remains to be seen, however, is whether all pairs of levels appear equally often in the pair of columns $[\mathbf{c}_3 \ \mathbf{c}_4]$. Hence, the list of designs to be inputed into Algorithm 3.1 will be smaller than $13 \times 11,718 = 152,334$ after invalid designs have been discarded. In fact, the number of potential OAs to be fed into Algorithm 3.1 was reduced to 1,944, which took less than an hour to run. This resulted in 137 geometric isomorphism classes, which matches the number quoted in Pang and Liu (2011). Although Pang and Liu (2011) do not provide representatives of the classes they found when $m \geq 4$, Tsai et al. (2000) give electronic copies

[1] $\mathbf{c}_4$ may be used in the lexicographical ordering of the rows if two triples in $[\mathbf{c}_1 \ \mathbf{c}_2 \ \mathbf{c}_3]$ happen to be the same. In that case, the method we describe in this section will produce two columns that are essentially the same, but since we are not overly concerned with calculation optimisation, there is no gain in identifying these columns in the initial set, rather we will let Algorithm 3.1 filter them out

of one representative from each of the OA$[18, 3^m]$ geometric isomorphism classes they found for all m. When $m = 4$ they have 129 classes and we have confirmed each of these is geometrically isomorphic to exactly one of our 137 designs.

3.2.3 m Factors

The process described in the previous section for generating all four-factor isomorphism classes is a specific example of the method we have applied for all m-factor isomorphism classes when $m > 3$. That is, we can view the 11,718 columns in the complete set of $\{\mathbf{c}_4\}$ identified in the previous section as 11,718 columns in the complete set of $\{\mathbf{c}_m\}$, which can be appended to one representative from each of the geometric isomorphism classes with $m - 1$ factors. When we generate an m-factor design by appending an additional column to an OA with $m - 1$ factors, we avoid the need to check that all 9 pairs of level combinations appear exactly twice in all pairs of the first $m - 1$ columns as we know this must already be the case. Additionally, the method described in the previous section to generate the complete set of $\{\mathbf{c}_m\}$ ensures that the pairs of columns $[\mathbf{c}_1 \ \mathbf{c}_m]$ and $[\mathbf{c}_2 \ \mathbf{c}_m]$ will also have exactly two incidence of each of the pairs of level combinations, so all that remains to be checked is the $m - 3$ pairs of columns $[\mathbf{c}_3 \ \mathbf{c}_m]$, $[\mathbf{c}_4 \ \mathbf{c}_m]$, ..., $[\mathbf{c}_{m-1} \ \mathbf{c}_m]$.

We summarise the process of generating a single representative from each of the geometric isomorphism classes of m-factor OAs, $m > 3$, below:

1. Find all geometric isomorphism classes for $m - 1$ factors.
2. Generate an initial set of OAs:
   (a) Append each of the 11,718 columns in the complete set $\{\mathbf{c}_m\}$ to one representative from each of the classes for $m - 1$ factors.
   (b) Discard any designs that do not have all pairs of level combinations appear exactly twice in each of the pairs of columns $[\mathbf{c}_3 \ \mathbf{c}_m]$, $[\mathbf{c}_4 \ \mathbf{c}_m]$, ..., $[\mathbf{c}_{m-1} \ \mathbf{c}_m]$.
3. Input the set of OAs into Algorithm 3.1 to filter out isomorphic designs.

We found the same number of geometrically non-isomorphic designs as reported in Pang and Liu (2011) for all m. Since we used a different method to those authors, we have successfully carried out "consistency checking" as defined by Kaski and Östergård (2006).

Table 3.2.9 summarises our findings for all OA$[18, 3^m]$ geometric isomorphism classes. The time taken to generate the initial set of OAs was negligible, hence we have presented the time taken to run Algorithm 3.1 once the initial set has been constructed.

Table 3.2.9: Number of geometric isomorphism classes for OA$[18, 3^m]$s

m	**No. Classes**	**No. Initial OAs to Filter**	**Time to Filter**
3	13	69	$<$ 1 min
4	137	1,944	$<$ 1 hr
5	333	2,481	$<$ 3 hr
6	485	3,384	$<$ 5 hr
7	291	1,893	$<$ 2 hr

One representative from each geometric isomorphism class is given in the electronic appendix. We have confirmed for all m that each of the designs given by Tsai et al. (2000) is geometrically isomorphic to exactly one of our designs.

As noted previously, our calculation times may not be optimal, but they are fast enough for our purposes, which is simply to obtain one representative from each geometric isomorphism class. Hence, we are now in a position to analyse all OA plus p run designs and all OA minus t run designs for ternary OAs in 18 runs when at least one linear-by-linear interaction is included in the model. In Chapter 4 we consider models that include one linear-by-linear interaction, and in Chapter 5 we consider models that contain any subset of ν linear-by-linear interactions.

Chapter 4

Models Containing One Linear-by-Linear Interaction

In Chapter 2 we derived $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_B$, matrices to be optimised when adding or removing runs from an OA under a main effects only model. These matrices have a well defined structure based on the pairwise Hamming distances of the runs in question, hence we were able to develop generic rules for the optimisation process.

In this chapter we will derive the analogous matrices under a model that includes the linear-by-linear component of a single two-factor interaction. We will show that the structure of these matrices are more complicated than those in Chapter 2 because $\mathbf{M}^{-1}$ is no longer a diagonal matrix, as was the case under a main effects model. The complicated nature of these matrices precludes the development of generic rules analogous to those developed in Chapter 2, hence we conduct an empirical investigation using the ternary 18-run OAs enumerated in Chapter 3 instead.

We begin in Section 4.1 by making some general observations about the structure of the information matrix for the OA when a single linear-by-linear interaction is included in the model. We then derive a general formulation for the matrices to be optimised when runs are augmented or removed in Section 4.2, including a general discussion of the properties of these matrices. In Section 4.3 we enumerate the entries in these matrices for the OA$[18, 3^m]$s enumerated in Chapter 3. We conclude with an empirical examination of OA$[18, 3^m]$ plus p designs and OA$[18, 3^m]$ minus t designs in Section 4.4.

4.1 Information Matrix for the OA

In order to consider the form of the information matrix for OA plus p runs designs or OA minus t runs designs, we first need to consider the form of the information matrix of the unaltered OA. Recall that to derive the matrices to be optimised in Chapter 2 we utilised Theorem 1.2.1 regarding the determinant of matrices. As we saw in Sections 2.2.1 and 2.2.2, this required that the information matrix of the altered design be in the form **G**+**SHL**. One of the substitutions required to achieve this form was to set **G** equal to **M**, the information matrix for the OA before any runs are adjoined or removed. Theorem 1.2.1 requires that we calculate the inverse of **G**. This was not an issue in Chapter 2, as we were assuming a main effects only model, hence we were able to exploit Lemma 1.2.1 and assume $\mathbf{M} = N\mathbf{I}_\alpha$, thus $\mathbf{G}^{-1} = \mathbf{M}^{-1} = \mathbf{I}_\alpha/N$. This chapter, however, considers models that contain a single linear-by-linear interaction, and so we can no longer assume that the information matrix of the OA is a multiple of the identity, hence $\mathbf{G}^{-1} = \mathbf{M}^{-1}$ may no longer have a well-defined structure.

In order to examine the structure of $\mathbf{M}^{-1}$, we begin by examining the structure of **M**. Recall that $\alpha = 1 + \sum_{i=1}^{k}(s_i - 1)m_i$ is the number of parameters in a main effects only model, hence a model that contains the linear-by-linear component of a single two-factor interaction has $\upsilon = \alpha + 1$ parameters.

Recall from Chapter 1 that the hth column within the set of $s_i - 1$ columns of the model matrix associated with the main effects factor F_j is labelled $\mathbf{f}_{jh}$. The main focus throughout the remainder of this thesis is on factors with three levels, hence for clarity, we will replace $h = 1$ with l and $h = 2$ with q to signify the linear and quadratic effects respectively. Hence, the two columns in **X** associated with the main effects of a three-level factor F_j are denoted by $\mathbf{f}_{jl}$ and $\mathbf{f}_{jq}$ respectively. We use f_{al_i} to denote the ith entry in $\mathbf{f}_{al}$, $1 \leq i \leq N$. Also recall that we use $\mathbf{z}_{(ab)}$ to denote the column in **X** corresponding to the linear-by-linear interaction between factors F_a and F_b, that is $\mathbf{z}_{(ab)} = \mathbf{f}_{al} \circ \mathbf{f}_{bl}$. In the remainder of this section we will present proofs that can be applied to any s-level factor, hence we continue to use some generic notation for the number of levels of each factor, but since we are interested in the linear-by-linear interactions, it helps to introduce notation such as $\mathbf{f}_{jl}$ now to aid clarity.

Let $\mathbf{X}_{ME}$ be the $N \times \alpha$ main effects only model matrix for an OA, then the partitioned $N \times \upsilon$ model matrix for the main effects and one linear-by-linear interaction model is

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{ME} & \mathbf{z}_{(ab)} \end{bmatrix} \tag{4.1.1}$$

where the interaction is assumed to be between factors F_a and F_b. The associated

information matrix is

$$\mathbf{M} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}'_{ME} \\ \mathbf{z}'_{(ab)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{ME} & \mathbf{z}_{(ab)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_{ME}\mathbf{X}_{ME} & \mathbf{X}'_{ME}\mathbf{z}_{(ab)} \\ \mathbf{z}'_{(ab)}\mathbf{X}_{ME} & \mathbf{z}'_{(ab)}\mathbf{z}_{(ab)} \end{bmatrix}. \quad (4.1.2)$$

We know $\mathbf{X}'_{ME}\mathbf{X}_{ME} = N\mathbf{I}_\alpha$ (Lemma 1.2.1), and in Lemma 4.1.1 below we show that $\mathbf{z}'_{(ab)}\mathbf{z}_{(ab)} = N$ by exploiting the properties of the polynomial contrasts used to construct $\mathbf{f}_{al}$ and $\mathbf{f}_{bl}$. The following lemma uses the notation $\mathbf{p}_{sl}$ to denote the first row of $\mathbf{P}_s$ (Definition 1.1.2), that is, $\mathbf{p}_{sl}$ is the row vector of length s representing the linear polynomial contrast for an s-level factor. We use p_{sl_j} to represent the jth entry in $\mathbf{p}_{sl}$, $1 \leq j \leq s$.

Lemma 4.1.1. Let $\mathbf{z}_{(ab)}$ be the Hadamard product between the linear main effects of factors F_a and F_b at levels s_a and s_b respectively for an OA in N runs, that is, $\mathbf{z}_{(ab)} = \mathbf{f}_{al} \circ \mathbf{f}_{bl}$. Then $\mathbf{z}'_{(ab)}\mathbf{z}_{(ab)} = N$.

Proof. $\mathbf{z}'_{(ab)}\mathbf{z}_{(ab)} = (\mathbf{f}_{al} \circ \mathbf{f}_{bl})' \cdot (\mathbf{f}_{al} \circ \mathbf{f}_{bl}) = \sum_{i=1}^{N} f_{al_i} f_{bl_i} f_{al_i} f_{bl_i}$. From Definition 1.1.2, $\mathbf{P}_s\mathbf{P}'_s = s\mathbf{I}_{s-1}$ which implies $\mathbf{p}_{sl}\mathbf{p}'_{sl} = s$. Since each pair of level combinations between factors F_a and F_b will appear exactly $N/(s_a s_b)$ times in the OA, we have

$$\begin{aligned}
\mathbf{z}'_{(ab)}\mathbf{z}_{(ab)} &= \frac{N}{s_a s_b} \sum_{i=1}^{s_a} \sum_{j=1}^{s_b} p^2_{s_a l_i} p^2_{s_b l_j} \\
&= \frac{N}{s_a s_b} \sum_{i=1}^{s_a} \left(p^2_{s_a l_i} \sum_{j=1}^{s_b} p^2_{s_b l_j} \right) \\
&= \frac{N}{s_a s_b} \sum_{i=1}^{s_a} \left(p^2_{s_a l_i} (\mathbf{p}_{s_b l} \mathbf{p}'_{s_b l}) \right) \\
&= \frac{N}{s_a} \sum_{i=1}^{s_a} p^2_{s_a l_i} \\
&= \frac{N}{s_a} \mathbf{p}_{s_a l} \mathbf{p}'_{s_a l} \\
&= N.
\end{aligned}$$

□

We will use $\mathbf{u}_{(ab)}$ to denote the $\alpha \times 1$ vector $\mathbf{X}'_{ME}\mathbf{z}_{(ab)}$. Hence, following Lemma 4.1.1, Equation (4.1.2) becomes

$$\mathbf{M} = \begin{bmatrix} N\mathbf{I}_\alpha & \mathbf{u}_{(ab)} \\ \mathbf{u}'_{(ab)} & N \end{bmatrix}. \quad (4.1.3)$$

Note that each entry in $\mathbf{u}_{(ab)}$ is associated with a particular term in the main effects only model. The following two Lemmas show that at least five of these entries are zero.

Lemma 4.1.2. Let $\mathbf{u}_{(ab)} = \mathbf{X}'_{ME}\mathbf{z}_{(ab)}$ for an OA in N runs. Then the first entry in $\mathbf{u}_{(ab)}$, which is associated with the intercept, is $\mathbf{1}'_N\mathbf{z}_{(ab)} = \sum_{i=1}^{N} f_{al_i}f_{bl_i} = 0$.

Proof. From Definition 1.1.2, $\mathbf{P}_s\mathbf{1}_s = \mathbf{0}_{s-1}$ which implies $\mathbf{p}_{sl} \cdot \mathbf{1}_s = 0$. Since each pair of level combinations between factors F_a and F_b will appear exactly $N/(s_a s_b)$ times in the OA, we have

$$\begin{aligned}
\mathbf{1}'_N\mathbf{z}_{(ab)} &= \frac{N}{s_a s_b}\sum_{i=1}^{s_a}\sum_{j=1}^{s_b} p_{s_a l_i}p_{s_b l_j} \\
&= \frac{N}{s_a s_b}\sum_{i=1}^{s_a}\left(p_{s_a l_i}\sum_{j=1}^{s_b} p_{s_b l_j}\right) \\
&= \frac{N}{s_a s_b}\sum_{i=1}^{s_a} p_{s_a l_i}\left(\mathbf{p}_{s_b l}\cdot\mathbf{1}_{s_b}\right) \\
&= 0.
\end{aligned}$$

□

Lemma 4.1.3. Let $\mathbf{u}_{(ab)} = \mathbf{X}'_{ME}\mathbf{z}_{(ab)}$ for an OA in N runs. Then the entries in $\mathbf{u}_{(ab)}$ associated with the main effects of factors F_a and F_b are zero. That is $\mathbf{f}'_{al}\mathbf{z}_{(ab)} = \mathbf{f}'_{aq}\mathbf{z}_{(ab)} = \mathbf{f}'_{bl}\mathbf{z}_{(ab)} = \mathbf{f}'_{bq}\mathbf{z}_{(ab)} = 0$.

Proof. Note that $\mathbf{f}'_{al}\mathbf{z}_{(ab)} = \frac{N}{s_a s_b}\sum_{i=1}^{s_a}\sum_{j=1}^{s_b} p^2_{s_a l_i}p_{s_b l_j}$, then $\mathbf{f}'_{al}\mathbf{z}_{(ab)} = 0$ follows from the proof in Lemma 4.1.2. The same can be said for $\mathbf{f}'_{aq}\mathbf{z}_{(ab)}$, $\mathbf{f}'_{bl}\mathbf{z}_{(ab)}$ and $\mathbf{f}'_{bq}\mathbf{z}_{(ab)}$. □

Lemmas 4.1.2 and 4.1.3 show that five of the entries in $\mathbf{u}_{(ab)}$ are zero. We will label the remaining (potentially) non-zero entries in $\mathbf{u}_{(ab)}$ according to the main effects they are associated with. We denote the entries corresponding to $\mathbf{f}'_{jl}\mathbf{z}_{(ab)}$ and $\mathbf{f}'_{jq}\mathbf{z}_{(ab)}$ as $u_{(ab)jl}$ and $u_{(ab)jq}$ respectively, $j \notin (a, b)$. Hence, these entries can be considered as pairs of $(u_{(ab)jl}, u_{(ab)jq})$ for each of the remaining $m-2$ factors. The following example uses this notation to illustrate the previous Lemmas.

Example 4.1.1. Consider a OA$[18, 3^4]$. We label the factors F_1, F_2, F_3 and F_4, and let $\mathbf{z}_{(12)}$ represent the linear-by-linear component of the interaction between factors F_1 and F_2. Then

$$
\begin{aligned}
\mathbf{u}_{(12)} &= \mathbf{X}'_{ME}\mathbf{z}_{(12)} \\
&= \left[\begin{array}{cccccccccc} \mathbf{1}_{18} & \mathbf{f}_{1l} & \mathbf{f}_{1q} & \mathbf{f}_{2l} & \mathbf{f}_{2q} & \mathbf{f}_{3l} & \mathbf{f}_{3q} & \mathbf{f}_{4l} & \mathbf{f}_{4q} \end{array} \right]' \cdot (\mathbf{f}_{1l} \circ \mathbf{f}_{2l}) \\
&= \begin{bmatrix} 1 & 1 & \dots & 1 \\ f_{1l_1} & f_{1l_2} & \dots & f_{1l_{18}} \\ f_{1q_1} & f_{1q_2} & \dots & f_{1q_{18}} \\ f_{2l_1} & f_{2l_2} & \dots & f_{2l_{18}} \\ f_{2q_1} & f_{2q_2} & \dots & f_{2q_{18}} \\ f_{3l_1} & f_{3l_2} & \dots & f_{3l_{18}} \\ f_{3q_1} & f_{3q_2} & \dots & f_{3q_{18}} \\ f_{4l_1} & f_{4l_2} & \dots & f_{4l_{18}} \\ f_{4q_1} & f_{4q_2} & \dots & f_{4q_{18}} \end{bmatrix} \begin{bmatrix} f_{1l_1} f_{2l_1} \\ f_{1l_2} f_{2l_2} \\ \vdots \\ f_{1l_{18}} f_{2l_{18}} \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ u_{(12)3l} \\ u_{(12)3q} \\ u_{(12)4l} \\ u_{(12)4q} \end{bmatrix}.
\end{aligned}
$$

The first five elements in $\mathbf{u}_{(12)}$ correspond to the intercept and the pairs of main effect components for factors F_1 and F_2, hence these values are equal to zero. The remaining (potentially) non-zero elements can be grouped into pairs corresponding to each of the remaining factors, $(u_{(12)3l}, u_{(12)3q})$ and $(u_{(12)4l}, u_{(12)4q})$.

Suppose we use a OA$[18, 3^4]$ from the first geometric isomorphism class enumerated in Chapter 3. Then

$$\mathbf{X}_{ME} = \begin{array}{c} \begin{array}{cccccccc} \mathbf{f}_{1l} & \mathbf{f}_{1q} & \mathbf{f}_{2l} & \mathbf{f}_{2q} & \mathbf{f}_{3l} & \mathbf{f}_{3q} & \mathbf{f}_{4l} & \mathbf{f}_{4q} \end{array} \\ \left[\begin{array}{cccccccc} 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 0 & -\sqrt{2} \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 0 & -\sqrt{2} \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & -\sqrt{2} & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} \\ 1 & 0 & -\sqrt{2} & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} \\ 1 & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 0 & -\sqrt{2} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} & 0 & -\sqrt{2} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} \\ 1 & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} & 0 & -\sqrt{2} \end{array}\right] \end{array}$$

and

$$\begin{aligned} \mathbf{z}_{(12)} &= \mathbf{f}_{1l} \circ \mathbf{f}_{2l} \\ &= \left[\begin{array}{cccccccccccccccc} \frac{3}{2} & \frac{3}{2} & 0 & 0 & -\frac{3}{2} & -\frac{3}{2} & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{3}{2} & -\frac{3}{2} & 0 & 0 & \frac{3}{2} & \frac{3}{2} \end{array}\right]'. \end{aligned}$$

Hence

$$\begin{aligned} \mathbf{u}_{(12)} &= \mathbf{X}'_{ME}\mathbf{z}_{(12)} \\ &= \left[\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & -\frac{9}{\sqrt{2}} & 0 & -9\sqrt{2} \end{array}\right]'. \end{aligned}$$

Thus, $(u_{(12)3l},u_{(12)3q}) = (0,-\frac{9}{\sqrt{2}})$ and $(u_{(12)4l},u_{(12)4q}) = (0,-9\sqrt{2})$. □

4.1.1 Inverse of the Information Matrix for the OA

As we have seen above, the inclusion of an interaction term in the model complicates the structure of the information matrix and so we can no longer assume $\mathbf{M}^{-1} = \mathbf{I}_\alpha/N$. This section examines the structure of $\mathbf{M}^{-1}$ when the linear-by-linear component of a single two-factor interaction is included in the model.

Given the partitioned representation of $\mathbf{M}$ in (4.1.3), let $\mathbf{T} = N\mathbf{I}_\alpha$, $\mathbf{U} = \mathbf{u}_{(ab)}$, $\mathbf{V} = \mathbf{u}'_{(ab)}$ and $\mathbf{W} = N$, then $\mathbf{M} = \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$. We can now apply Theorem 1.2.2 to find the inverse of $\mathbf{M}$. For this purpose, it will be useful to introduce the notation $\mathbf{0}_{i\times j}$ to represent a matrix of order $i \times j$ with all entries equal to zero.

$$\begin{aligned} \mathbf{M}^{-1} &= \begin{bmatrix} \mathbf{T}^{-1} & \mathbf{0}_{\alpha\times 1} \\ \mathbf{0}_{1\times\alpha} & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{T}^{-1}\mathbf{U} \\ \mathbf{I}_1 \end{bmatrix} \mathbf{Q}^{-1} \begin{bmatrix} -\mathbf{V}\mathbf{T}^{-1} & \mathbf{I}_1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_\alpha/N & \mathbf{0}_{\alpha\times 1} \\ \mathbf{0}_{1\times\alpha} & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{u}_{(ab)}/N \\ 1 \end{bmatrix} \mathbf{Q}^{-1} \begin{bmatrix} -\mathbf{u}'_{(ab)}/N & 1 \end{bmatrix} \end{aligned} \tag{4.1.4}$$

where

$$\begin{aligned} \mathbf{Q} &= \mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U} \\ &= N - \mathbf{u}'_{(ab)}\mathbf{u}_{(ab)}/N \\ &= N - \frac{1}{N}\sum_{\substack{j=1 \\ j\notin(a,b)}}^{m} \left(u^2_{(ab)j_l} + u^2_{(ab)j_q}\right). \end{aligned} \tag{4.1.5}$$

We note that $\mathbf{Q}$ must be non-singular to satisfy the condition of Theorem 1.2.2, hence $N - \frac{1}{N}\sum_{\substack{j=1 \\ j\notin(a,b)}}^{m} \left(u^2_{(ab)j_l} + u^2_{(ab)j_q}\right)$ must be non-zero, that is,

$$N \neq \sqrt{\sum_{\substack{j=1 \\ j\notin(a,b)}}^{m} \left(u^2_{(ab)j_l} + u^2_{(ab)j_q}\right)}.$$

The following example illustrates this result.

Example 4.1.2. Following Example 4.1.1, consider an OA in N runs with four three-level factors, hence $\alpha = 1+2\times 4 = 9$, and $\upsilon = \alpha+1 = 10$. Assume the interaction to be included in the model is between factors F_1 and F_2. For more compact notation within this example we will omit the subscript $_{(12)}$ from the entries in $\mathbf{u}_{(12)}$, that is, we will denote the entries as $\mathbf{u}_{(12)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & u_{3l} & u_{3q} & u_{4l} & u_{4q} \end{bmatrix}'$. From Equations (4.1.4) and (4.1.5) we have

$$\mathbf{M}^{-1} = \begin{bmatrix} \frac{1}{N}\mathbf{I}_9 & \mathbf{0}_{9\times 1} \\ \mathbf{0}_{1\times 9} & 0 \end{bmatrix} + N^2\zeta \begin{bmatrix} -\frac{1}{N}\mathbf{u}_{(12)} \\ 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{N}\mathbf{u}'_{(12)} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{N}\mathbf{I}_9 & \mathbf{0}_{9\times 1} \\ \mathbf{0}_{1\times 9} & 0 \end{bmatrix} + \zeta \begin{bmatrix} \mathbf{0}_{5\times 5} & & & \mathbf{0}_{5\times 5} & & \\ & u_{3l}^2 & u_{3l}u_{3q} & u_{3l}u_{4l} & u_{3l}u_{4q} & -Nu_{3l} \\ & u_{3l}u_{3q} & u_{3q}^2 & u_{3q}u_{4l} & u_{3q}u_{4q} & -Nu_{3q} \\ \mathbf{0}_{5\times 5} & u_{3l}u_{4l} & u_{3q}u_{4l} & u_{4l}^2 & u_{4l}u_{4q} & -Nu_{4l} \\ & u_{3l}u_{4q} & u_{3q}u_{4q} & u_{4l}u_{4q} & u_{4q}^2 & -Nu_{4q} \\ & -Nu_{3l} & -Nu_{3q} & -Nu_{4l} & -Nu_{4q} & N^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{N} & 0 & 0 & 0 & 0 & & & \mathbf{0}_{5\times 5} & & \\ 0 & \frac{1}{N} & 0 & 0 & 0 & & & & & \\ 0 & 0 & \frac{1}{N} & 0 & 0 & & & & & \\ 0 & 0 & 0 & \frac{1}{N} & 0 & & & & & \\ 0 & 0 & 0 & 0 & \frac{1}{N} & & & & & \\ & & & & & \frac{1}{N}+\zeta u_{3l}^2 & \zeta u_{3l}u_{3q} & \zeta u_{3l}u_{4l} & \zeta u_{3l}u_{4q} & -\zeta N u_{3l} \\ & & & & & \zeta u_{3l}u_{3q} & \frac{1}{N}+\zeta u_{3q}^2 & \zeta u_{3q}u_{4l} & \zeta u_{3q}u_{4q} & -\zeta N u_{3q} \\ & & \mathbf{0}_{5\times 5} & & & \zeta u_{3l}u_{4l} & \zeta u_{3q}u_{4l} & \frac{1}{N}+\zeta u_{4l}^2 & \zeta u_{4l}u_{4q} & -\zeta N u_{4l} \\ & & & & & \zeta u_{3l}u_{4q} & \zeta u_{3q}u_{4q} & \zeta u_{4l}u_{4q} & \frac{1}{N}+\zeta u_{4q}^2 & -\zeta N u_{4q} \\ & & & & & -\zeta N u_{3l} & -\zeta N u_{3q} & -\zeta N u_{4l} & -\zeta N u_{4q} & \zeta N^2 \end{bmatrix}$$

where $\zeta = 1/\left(N^3 - N\left(u_{3l}^2 + u_{3q}^2 + u_{4l}^2 + u_{4q}^2\right)\right)$. □

The previous example illustrates that if we assume all factors have three levels and rearrange the columns of $\mathbf{X}$ such that the first five columns are the intercept and the pair of main effects for both of the factors involved in the interaction, then the $\upsilon \times \upsilon$ inverse of the information matrix will be a block diagonal matrix with a 5×5 block in the top left corner equal to $\mathbf{I}_5/N$ and a $(\upsilon - 5) \times (\upsilon - 5)$ block in the bottom right corner with entries dependant on $\mathbf{u}_{(12)}$ and N.

4.2 Information Matrix of the Altered Design

We will now consider the structure of the altered design when runs are adjoined or removed. For the remainder of this chapter, we will assume that all factors appear at three levels. That is, $k = 1$, $s_1 = 3$ and $m_1 = m$.

4.2.1 Augmenting p Runs

Recall from Section 2.2.1 that we write the model matrix of the augmented design as $\mathbf{X}_A = \begin{bmatrix} \mathbf{X} \\ \mathbf{A} \end{bmatrix}$ with corresponding information matrix $\mathbf{M}_A = \mathbf{X}'\mathbf{X} + \mathbf{A}'\mathbf{A}$. That section focused on models containing only the intercept and main effects, so we

were able to utilise the result $\mathbf{M} = \mathbf{X}'\mathbf{X} = N\mathbf{I}_\alpha$ in conjunction with Theorem 1.2.1 to define the $p \times p$ matrix $\mathbf{\Omega}_A$. A similar approach can be taken here, however we can no longer use the result $\mathbf{M} = N\mathbf{I}_\alpha$ since the structure of $\mathbf{M}$ has changed. As in Section 2.2.1, we make the substitutions $\mathbf{S} = \mathbf{A}'$, $\mathbf{H} = \mathbf{I}_p$ and $\mathbf{L} = \mathbf{A}$, but now that we are dealing with an information matrix with a more complex structure, we will let $\mathbf{G} = \mathbf{M}$ in order to represent $\mathbf{M}_A = \mathbf{M} + \mathbf{A}'\mathbf{A}$ in the form $\mathbf{G} + \mathbf{SHL}$. Following Theorem 1.2.1:

$$\begin{aligned} |\mathbf{M}_A| &= |\mathbf{G}||\mathbf{H}||\mathbf{H}^{-1} + \mathbf{L}\mathbf{G}^{-1}\mathbf{S}| \\ &= |\mathbf{M}||\mathbf{I}_p||\mathbf{I}_p^{-1} + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'| \\ &= |\mathbf{M}||\mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'|. \end{aligned} \tag{4.2.1}$$

We now have two determinants to consider: $|\mathbf{M}|$ and $|\mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'|$. We note that $|\mathbf{M}|$, the determinant of the information matrix for the OA before any runs are adjoined, will be constant when we are considering which runs to adjoin to a specific OA, but this value may change from OA to OA, hence we can no longer treat it as a constant in the class of all OA plus p runs designs. We will define $\mathbf{\Psi}_A = \mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$. Note that in Chapter 2 the second term in $\mathbf{\Omega}_A$ was simply $\mathbf{A}\mathbf{A}'$, which allowed us to examine the entries in terms of Hamming distances. In this chapter, however, the second term in $\mathbf{\Psi}_A$ is $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ which complicates the optimisation process, as we shall see in Section 4.2.3.

Due to the complicated nature of $\mathbf{M}^{-1}$, hence $\mathbf{\Psi}_A$, the majority of this chapter will focus on the non-zero off-diagonal entries in $\mathbf{M}$ and the implication for the structure of $\mathbf{M}^{-1}$. But first we will look at the form of $\mathbf{\Psi}_B$ when t runs are missing from the OA.

4.2.2 Missing t Runs

Following Section 2.2.2, we partition the runs of $\mathbf{X}$, the model matrix for the full OA, as $\begin{bmatrix} \mathbf{X}_B \\ \mathbf{B} \end{bmatrix}$ where $\mathbf{B}$ is a $t \times \alpha$ matrix that contains the t row vectors associated with the missing runs. From here we see the information matrix for the design with t missing runs is $\mathbf{M}_B = \mathbf{X}_B'\mathbf{X}_B = \mathbf{X}'\mathbf{X} - \mathbf{B}'\mathbf{B}$. As discussed in the previous section, we can no longer assume $\mathbf{X}'\mathbf{X} = N\mathbf{I}_\alpha$.

As in Section 2.2.2, we let $\mathbf{S} = \mathbf{B}'$, $\mathbf{H} = -\mathbf{I}_t$, $\mathbf{L} = \mathbf{B}$, and as above we use $\mathbf{G} = \mathbf{M}$ in order to represent $\mathbf{M}_B = \mathbf{X}'\mathbf{X} - \mathbf{B}'\mathbf{B} = \mathbf{G} + \mathbf{SHL}$. Following Theorem 1.2.1,

we see that

$$\begin{aligned} |\mathbf{M}_B| &= |\mathbf{G}||\mathbf{H}||\mathbf{H}^{-1} + \mathbf{L}\mathbf{G}^{-1}\mathbf{S}| \\ &= |\mathbf{M}||-\mathbf{I}_t||(-\mathbf{I}_t)^{-1} + \mathbf{B}\mathbf{M}^{-1}\mathbf{B}'| \\ &= (-1)^t|\mathbf{M}||\mathbf{B}\mathbf{M}^{-1}\mathbf{B}' - \mathbf{I}_t|. \end{aligned} \tag{4.2.2}$$

As was the case when adjoining runs, we have two determinants to consider, one of which is $|\mathbf{M}|$, the determinant of the information matrix for the un-altered OA. We will use $\mathbf{\Psi}_B$ to denote $\mathbf{B}\mathbf{M}^{-1}\mathbf{B}' - \mathbf{I}_t$.

4.2.3 The Structure of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$

As we have seen above, the main term in $\mathbf{\Psi}_A$ is $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$, and the main term in $\mathbf{\Psi}_B$ is $\mathbf{B}\mathbf{M}^{-1}\mathbf{B}'$. The only difference between these two terms is that $\mathbf{A}$ contains p rows corresponding to the augmented runs, whereas $\mathbf{B}$ contains t rows corresponding to the missing runs, hence an examination of the structure of either of these matrices will be the same[1]. For the remainder of this section we will discuss the properties of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ with the understanding that the same logic can be applied to $\mathbf{B}\mathbf{M}^{-1}\mathbf{B}'$.

Let $\mathbf{a}_i$ be the ith row in $\mathbf{A}$, $1 \leq i \leq p$, that is, when the linear-by-linear interaction is between factors F_a and F_b, $\mathbf{a}_i = \begin{bmatrix} 1 & f_{1l_i} & f_{1q_i} & f_{2l_i} & f_{2q_i} & \dots & f_{ml_i} & f_{mq_i} & f_{al_i}f_{bl_i} \end{bmatrix}$.

In order to discuss the entries in $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ it will be helpful to use notation defined in Chapter 2. Recall that we use q as simplified notation for $\mathbf{a}_i\mathbf{a}_j'$ under a main effects only model. Let $\mathbf{a}_{ME_i}$ be $\mathbf{a}_i$ truncated to the first α entries (that is, the form $\mathbf{a}_i$ would take under a main effects only model), hence $q = \mathbf{a}_{ME_i}\mathbf{a}_{ME_j}'$. We can use this notation to simplify the expression for the entry in the ith row and jth column of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$, as the following example illustrates.

Example 4.2.1. Consider an OA in N runs with four three-level factors and assume the interaction to be included in the model is between factors F_1 and F_2, hence $\mathbf{u}_{(12)}$ and $\mathbf{M}^{-1}$ are as defined in Example 4.1.2, with the subscript (12) for the entries in $\mathbf{u}_{(12)}$ omitted. Then the entry in the ith row and jth column of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ is $\mathbf{a}_i\mathbf{M}^{-1}\mathbf{a}_j'$. When expanded out and simplified this gives:

$$\frac{1}{N}\left(1 + f_{1l_i}f_{1l_j} + f_{1q_i}f_{1q_j} + f_{2l_i}f_{2l_j} + f_{2q_i}f_{2q_j} + f_{3l_i}f_{3l_j} + f_{3q_i}f_{3q_j} + f_{4l_i}f_{4l_j} + f_{4q_i}f_{4q_j}\right) - \frac{\left(Nf_{1l_i}f_{2l_i} - f_{3l_i}u_{3l} - f_{3q_i}u_{3q} - f_{4l_i}u_{4l} - f_{4q_i}u_{4q}\right)\left(Nf_{1l_j}f_{2l_j} - f_{3l_j}u_{3l} - f_{3q_j}u_{3q} - f_{4l_j}u_{4l} - f_{4q_j}u_{4q}\right)}{N\left(-N^2 + u_{3l}^2 + u_{3q}^2 + u_{4l}^2 + u_{4q}^2\right)}$$

where $1 + f_{1l_i}f_{1l_j} + f_{1q_i}f_{1q_j} + f_{2l_i}f_{2l_j} + f_{2q_i}f_{2q_j} + f_{3l_i}f_{3l_j} + f_{3q_i}f_{3q_j} + f_{4l_i}f_{4l_j} + f_{4q_i}f_{4q_j}$ can be represented by q. □

[1]We just need to keep in mind that there are no restrictions on the runs in $\mathbf{A}$, whereas the runs of $\mathbf{B}$ are restricted to those in the OA.

The result in Example 4.2.1 can be generalised to m-factor ternary OAs as follows.

Theorem 4.2.1. Consider an OA$[N, 3^m]$ under a model containing the intercept, all main effects and the linear-by-linear component of the interaction between factors F_a and F_b. The entry in the ith row and jth column of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ is:

$$\frac{1}{N}\left(q - \frac{\left(Nf_{al_i}f_{bl_i} - \mathbf{a}_{ME_i}\mathbf{u}_{(ab)}\right)\left(Nf_{al_j}f_{bl_j} - \mathbf{a}_{ME_j}\mathbf{u}_{(ab)}\right)}{\mathbf{u}'_{(ab)}\mathbf{u}_{(ab)} - N^2}\right).$$

□

Theorem 4.2.1 shows that the off-diagonal entries in $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ are far more complicated than the analogous matrix under a main effects only model. Chapter 2 was devoted to finding optimal pairwise Hamming distances in the set of p runs to be adjoined in order to minimise the value of q, the off-diagonal entry in $\mathbf{\Omega}_A$. In this chapter, in order to minimise the off-diagonal entries in $\mathbf{\Psi}_A$, we need to consider not only q, but also $\mathbf{u}'_{(ab)}\mathbf{u}_{(ab)}$ for each OA, $\mathbf{a}_{ME_i}\mathbf{u}_{(ab)}$ for each adjoined run, and $f_{al_i}f_{bl_i}$, for each pair of adjoined runs. Hence, general constructions analogous to those in Chapter 2 are not possible here. Instead, we will carry out an empirical investigation on the OA$[18, 3^m]$s constructed in Chapter 3.

4.2.4 A Bound on the Determinants

In Chapter 2 we were able to derive a general expression for the theoretical upper bounds on $|\mathbf{M}_A|$ and $|\mathbf{M}_B|$ under a main effects only model due to the relatively simple structure of $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_B$. The complicated nature of $\mathbf{\Psi}_A$ and $\mathbf{\Psi}_B$ precludes us from deriving analogous results here, hence we will conclude this section by defining efficiency in terms of the empirical best, rather than the theoretical best. Nonetheless, it is still interesting to consider what the theoretical maximum of $|\mathbf{M}_A|$ and $|\mathbf{M}_B|$ might be, and so we begin this section by examining various features of the adjoined (or removed) runs that may influence these values.

Augmenting p runs

Recall that when an interaction term is included in the model we have two determinants to consider: $|\mathbf{M}|$ and $|\mathbf{\Psi}_A|$. We will begin by considering the simplest case, that is, when $\mathbf{u}_{(ab)}$ has all entries equal to zero.

We see from Equation (4.1.3) that the diagonal entries of $\mathbf{M}$ are all N. Thus, when $\mathbf{u}_{(ab)}$ is a vector of zeros, $\mathbf{M} = N\mathbf{I}_\upsilon$, hence $|\mathbf{M}| = |N\mathbf{I}_\upsilon| = N^\upsilon$. Therefore, under this simplistic setting, we can focus on optimising $|\mathbf{\Psi}_A|$ independently of $|\mathbf{M}|$ since the later is a constant.

Now consider $|\mathbf{\Psi}_A| = |\mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'|$. Following Theorem 4.2.1, when $\mathbf{u}_{(ab)}$ is the column vector of zeros, the entry in the ith row and jth column of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ is

$$\begin{aligned}\mathbf{a}_i\mathbf{M}^{-1}\mathbf{a}_j' &= \frac{1}{N}\left(q - \frac{\left(Nf_{al_i}f_{bl_i}\right)\left(Nf_{al_j}f_{bl_j}\right)}{-N^2}\right)\\ &= \frac{1}{N}\left(q + f_{al_i}f_{bl_i}f_{al_j}f_{bl_j}\right)\\ &= \frac{1}{N}\left(\mathbf{a}_i\mathbf{a}_j'\right).\end{aligned} \tag{4.2.3}$$

This follows from the fact that $\mathbf{M} = N\mathbf{I}_v$, hence $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}' = \mathbf{A}\mathbf{A}'/N$. Recall that the entries in the linear main effect of a three-level factor are $-\sqrt{3/2}$ for the low level, 0 for the middle level, and $\sqrt{3/2}$ for the high level. Hence, $f_{al_i}f_{bl_i}f_{al_j}f_{bl_j}$ will be equal to 0 when one or both of F_a or F_b are set to the middle level in either run, and $\frac{9}{4}$ otherwise. Also recall from Equation (2.2.6) in Chapter 2 that $q = \alpha$ when the Hamming distance between a given pair of runs is zero (which is obviously the case when we compare a single run to itself). Hence, when one of F_a or F_b is set to the middle level in the ith run then the ith diagonal entry of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ will be α/N, otherwise it will be $\left(\alpha + \frac{9}{4}\right)/N$. Let p_1 be the number of runs within the set of p runs to be adjoined in which at least one of F_a or F_b is set to the middle level. Then the trace of $\mathbf{\Psi}_A$ is

$$\begin{aligned}\operatorname{tr}\left(\mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'\right) &= p_1\left(1 + \frac{\alpha}{N}\right) + (p - p_1)\left(1 + \frac{\alpha + \frac{9}{4}}{N}\right)\\ &= \frac{p\left(9 + 4\left(\alpha + N\right)\right) - 9p_1}{4N}.\end{aligned} \tag{4.2.4}$$

Hence $\sum_{i=1}^{p}\lambda_i = \frac{p(9+4(\alpha+N))-9p_1}{4N}$, where $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ are the eigenvalues of $\mathbf{\Psi}_A$. From the arithmetic-geometric mean inequality see that

$$\begin{aligned}\left(\prod_{i=1}^{p}\lambda_i\right)^{1/p} &\leq \frac{\sum_{i=1}^{p}\lambda_i}{p},\\ \left(\prod_{i=1}^{p}\lambda_i\right)^{1/p} &\leq \frac{p\left(9 + 4\left(\alpha + N\right)\right) - 9p_1}{4Np},\\ \prod_{i=1}^{p}\lambda_i &\leq \left(\frac{p\left(9 + 4\left(\alpha + N\right)\right) - 9p_1}{4Np}\right)^p.\end{aligned} \tag{4.2.5}$$

Since the determinant of a matrix is equal to the product of its eigenvalues, Equation (4.2.5) implies $|\mathbf{\Psi}_A| \leq \left(\frac{p(9+4(\alpha+N))-9p_1}{4Np}\right)^p$. This bound decreases as p_1 increases, thus we can choose a set of p runs such that $p_1 = 0$ (that is, none of the

runs are set to the middle level in either F_a or F_b) so that $|\mathbf{\Psi}_A| \leq \left(\frac{9+4(\alpha+N)}{4N}\right)^p$. The arithmetic-geometric mean also states that the arithmetic mean is equal to the geometric mean if and only if all λ_i are equal, hence the upper bound of $|\mathbf{\Psi}_A|$ is realised when all eigenvalues are the same, which is achieved when all off-diagonal entries are equal to zero. However, since we chose the set of p runs such that none of them are set to the middle level in either F_a or F_b, Equation (4.2.3) becomes $\mathbf{a}_i\mathbf{M}^{-1}\mathbf{a}_j' = \frac{1}{N}\left(q + \frac{9}{4}\right)$. Recall that q is an integer, hence the off-diagonal entries of $\mathbf{\Psi}_A$ can never be zero and the theoretical upper bound $\left(\frac{9+4(\alpha+N)}{4N}\right)^p$ can never be realised in practice.

If, instead, we choose the set of p runs such that either F_a or F_b are set to the middle level in each run, then Equation (4.2.3) becomes $\mathbf{a}_i\mathbf{M}^{-1}\mathbf{a}_j' = \frac{q}{N}$, and it might be possible to find set of runs with pairwise Hamming distances such that all $q = 0$, hence all $\mathbf{a}_i\mathbf{M}^{-1}\mathbf{a}_j' = 0$. However, by choosing p runs such that either F_a or F_b are set to the middle level in every run, we are effectively setting $p_1 = p$ which decreases the bound in Equation (4.2.5). This bound becomes $|\mathbf{\Psi}_A| \leq \left(\frac{\alpha+N}{N}\right)^p$ which is smaller than $\left(\frac{9+4(\alpha+N)}{4N}\right)^p$. As we will show empirically throughout the remainder of this chapter, it is generally better to search for adjoined runs in which none are set to the middle level in either F_a or F_b, because although we can never realise the theoretical upper bound $\left(\frac{9+4(\alpha+N)}{4N}\right)^p$, we can still get quite close, that is, we can find determinants that are larger than the upper bound $\left(\frac{\alpha+N}{N}\right)^p$ which is imposed when either F_a or F_b are set to the middle level in the adjoined run. As we will discuss in in Section 4.2.4, this situation is reversed when runs are missing.

The following example illustrates these ideas regarding adjoined runs.

Example 4.2.2. Consider the OA$[18, 3^4]$ in Table 4.2.1, which is a representative of the third geometric isomorphism class for the four-factor designs enumerated in Chapter 3. Since we have four three-level factors, $\alpha = 1 + 2 \times 4 = 9$, and a model that includes a single linear-by-linear interaction has $\upsilon = \alpha + 1 = 10$ terms. When the interaction to be included in the model is between factors F_2 and F_3, the information matrix for the OA is $\mathbf{M} = 18\mathbf{I}_{10}$, that is, all entries in $\mathbf{u}_{(23)}$ are equal to zero.

Suppose $p = 2$ and we seek to maximise the bound on $|\mathbf{\Psi}_A|$. Thus, we require two runs in which neither run is set to the middle level in either F_2 or F_3. Then the diagonal entries of $\mathbf{\Psi}_A$ would be $1 + \left(\alpha + \frac{9}{4}\right)/N = \frac{13}{8}$, and hence the upper bound on $|\mathbf{\Psi}_A|$ would be $\frac{169}{64}$. We would also like the off-diagonal entries of $\mathbf{\Psi}_A$, which are equal to $\frac{1}{18}\left(q + \frac{9}{4}\right)$, to be as close to 0 as possible. Hence, we would like q to be as close to $-\frac{9}{4}$ as possible. Thus, we require that the pair of runs have Hamming distance 4 so that $q = \alpha - 3d = 9 - 3 \times 4 = -3$ (which is the closest obtainable

Table 4.2.1: OA$[18, 3^4]$s from geometric isomorphism class 3

	F_1	F_2	F_3	F_4
1	0	0	1	0
2	0	0	1	1
3	0	1	0	1
4	0	1	2	0
5	0	2	0	2
6	0	2	2	2
7	1	0	0	2
8	1	0	2	2
9	1	1	0	1
10	1	1	2	0
11	1	2	1	0
12	1	2	1	1
13	2	0	0	0
14	2	0	2	1
15	2	1	1	2
16	2	1	1	2
17	2	2	0	0
18	2	2	2	1

q to $-\frac{9}{4}$), so the diagonal entries of $\mathbf{\Psi}_A$ will be $\frac{1}{18}\left(-3+\frac{9}{4}\right) = -\frac{1}{24}$. The 5th and 14th runs meet both of these requirements, and give $\mathbf{\Psi}_A = \begin{bmatrix} \frac{13}{8} & -\frac{1}{24} \\ -\frac{1}{24} & \frac{13}{8} \end{bmatrix}$. Thus, $|\mathbf{\Psi}_A| = \frac{95}{36}$, which is 0.9993 of the upper bound in this setting. The determinant of the information matrix for the augmented design is $|\mathbf{M_A}| = |\mathbf{M}||\mathbf{\Psi}_A| = 18^{10} \times \frac{95}{36} = 9,422,066,292,480$. □

Example 4.2.3. Following Example 4.2.2, suppose rather than choosing two runs in which neither run is set to the middle level in either F_2 or F_3, we instead choose a pair of runs such that both are set to the middle level in either F_2 or F_3. The diagonal entries of $\mathbf{\Psi}_A$ in this setting would be $1 + \frac{\alpha}{N} = \frac{3}{2}$ hence the upper bound on $|\mathbf{\Psi}_A|$ would be $\frac{9}{4}$. Note this bound is smaller than the bound in Example 4.2.2 because the diagonal entries in $\mathbf{\Psi}_A$ are smaller. This bound can be realised when the off-diagonal elements are equal to 0, that is, $\frac{q}{18} = 0$. Hence, we require that the pair of adjoined runs has Hamming distance 3 which would give $q = 9 - 3 \times 3 = 0$. The 1st and 3rd runs are an example of such a pair. These two adjoined runs give $\mathbf{\Psi}_A = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}$. Thus, $|\mathbf{\Psi}_A| = \frac{9}{4}$, which is the upper bound in this setting. The determinant of the information matrix for the augmented design is $|\mathbf{M_A}| = |\mathbf{M}||\mathbf{\Psi}_A| = 18^{10} \times \frac{9}{4} = 8,033,551,259,904$. □

Notice that the determinant calculated in Example 4.2.2 is larger than the determinant calculated in Example 4.2.3. In the former example, neither adjoined run

was set to the middle level in either F_2 or F_3. This had the advantage that we were able to maximise the bound on $|\mathbf{\Psi}_A|$, but we were unable to achieve this bound. In the later example, either F_2 or F_3 were set to the middle level in both adjoined runs. This had the advantage that were were able to realise the upper bound on $|\mathbf{\Psi}_A|$, however this bound was smaller than that in the former example. These two examples suggest that when all entries in $\mathbf{u}_{(ab)}$ are zero, we should select a pair of runs in which neither are set to the middle level in either of the factors involved in the interaction, as this can result in a larger determinant than if both of the adjoined runs were set to the middle level in at least one of these factors.

Of course, the above discussion has only considered the simplest case, that is, when all entries in $\mathbf{u}_{(ab)}$ are zero. Furthermore, we have only discussed the most extreme cases within this scenario, that is when the diagonal entries of $\mathbf{A}_i\mathbf{M}^{-1}\mathbf{A}_j'$ are equal (either $\frac{\alpha}{N}$ or $\frac{1}{N}\left(\alpha+\frac{9}{4}\right)$). When $\mathbf{u}_{(ab)}$ contains non-zero values, the determinant of $\mathbf{M}$ is reduced (that is, $|\mathbf{M}| < N^v$), but the determinant of $\mathbf{\Psi}_A$ can increase such that their product is greater than when $\mathbf{u}_{(ab)}$ is a vector of zeros and $|\mathbf{M}| = N^v$. The following example illustrates this.

Example 4.2.4. An obvious question that arises from Examples 4.2.2 and 4.2.3 is: what is the largest value of $|\mathbf{M}_A|$ we can achieve in the class of all OA$[18, 3^4]$ plus 2 runs designs when a single linear-by-linear interaction is included in the model? We completed an exhaustive search of all 137 OA$[18, 3^4]$ in Chapter 3, for each of the $\binom{4}{2} = 6$ possible interactions to be included in the model, and each of the $\binom{18}{2} = 153$ pairs of runs from within the OA to be adjoined[2]. The maximum of these $137 \times 6 \times 153 = 125,766$ determinants is 9,744,400,139,328 which occurs only 4 times.

The representative from the 5th geometric isomorphism class enumerated in Chapter 3 achieves this determinant when the interaction is between factors F_3 and F_4 and the adjoined runs are 0222 and 1002. Notice that neither F_3 nor F_4 are set to the middle level in either of these runs. For this OA, the relevant entries in $\mathbf{u}_{(34)}$ are $(u_{(34)1l}, u_{(34)1q}) = \left(0, \frac{3}{2}\sqrt{\frac{3}{2}}\right)$ and $(u_{(34)2l}, u_{(34)2q}) = \left(-\frac{9}{4}\sqrt{2}, -3\sqrt{\frac{3}{2}}\right)$, hence $|\mathbf{M}| = 3,272,928,291,072$ which is 0.9167 of the value $|\mathbf{M}|$ takes when all entries in $\mathbf{u}_{(34)}$ are equal to zero, 18^{10}. Then $\mathbf{\Psi}_A = \begin{bmatrix} \frac{115}{66} & -\frac{8}{33} \\ -\frac{8}{33} & \frac{115}{66} \end{bmatrix}$, with $|\mathbf{\Psi}_A| = \frac{131}{44}$, which is 1.1282 of the theoretical upper bound when all entries in $\mathbf{u}_{(34)}$ are equal to zero.

The best determinant is also realised when we use the representative from the 93rd geometric isomorphism class enumerated in Chapter 3, the interaction is

[2]In this chapter we assume that the set of runs to choose from is restricted to the N runs within the OA. We discuss this further in section 4.2.4.

between factors F_3 and F_4 and the adjoined runs are 0222 and 2002. Again, note that neither F_3 nor F_4 are set to the middle level in either of these runs. For this OA, the relevant entries in $\mathbf{u}_{(34)}$ are $(u_{(34)1l}, u_{(34)1q}) = \left(0, 3\sqrt{\frac{3}{2}}\right)$ and $(u_{(34)2l}, u_{(34)2q}) = \left(0, -3\sqrt{\frac{3}{2}}\right)$, which has the same determinant as the determinant of the $\mathbf{M}$ given above, despite the matrix itself being different. The $\mathbf{\Psi}_A$ matrix is identical to that given above. □

As mentioned at the beginning of this section, given the complicated nature of the dependencies between $\mathbf{u}_{(ab)}$, $\mathbf{M}$, $\mathbf{\Psi}_A$, and the pairwise values of both $f_{al_i} f_{bl_i} f_{al_j} f_{bl_j}$ and q, we will redefine the efficiency of a design to be relative to the realisable best, rather than the theoretical best. But first we will consider the bound on $|\mathbf{\Psi}_B|$ when runs are missing.

Missing t Runs

The arguments for removing t runs are very similar to those above for augmenting p runs, and so we will only discuss the minor differences between the two methods here. The two determinants to consider when t runs are missing are $|\mathbf{M}|$ and $|\mathbf{\Psi}_B|$. Recall the only difference between $\mathbf{\Psi}_A$ and $\mathbf{\Psi}_B$ is that $\mathbf{\Psi}_A$ adds the identity to $\mathbf{AM}^{-1}\mathbf{A}'$ whereas $\mathbf{\Psi}_B$ subtracts the identity from $\mathbf{BM}^{-1}\mathbf{B}'$.

The information matrix for the OA before the design is altered, $\mathbf{M}$, is independent of whether we are considering adjoined or removing runs, hence the previous discussion regarding $\mathbf{M}$ is unchanged. The discussion above regarding the entries in $\mathbf{AM}^{-1}\mathbf{A}'$ are still relevant here in the context of $\mathbf{BM}^{-1}\mathbf{B}'$, however we need to make a minor adjustment to Equation (4.2.4) to reflect the negative identity in $\mathbf{\Psi}_B$. Let t_1 be the number of runs within the set of t runs to be removed in which at least one of F_a or F_b is set to the middle level. Then the trace of $\mathbf{\Psi}_B$ is

$$\begin{aligned} \operatorname{tr}\left(\mathbf{BM}^{-1}\mathbf{B}' - \mathbf{I}_t\right) &= t_1\left(\frac{\alpha}{N} - 1\right) + (t - t_1)\left(\frac{\alpha + \frac{9}{4}}{N} - 1\right) \\ &= \frac{t\left(9 + 4\left(\alpha - N\right)\right) - 9t_1}{4N}. \end{aligned} \tag{4.2.6}$$

Example 4.2.5. Following Example 4.2.2, consider the OA$[18, 3^4]$ in Table 4.2.1. Recall $\alpha = 1 + 2 \times 4 = 9$, $\upsilon = \alpha + 1 = 10$, and when the interaction to be included in the model is between factors F_2 and F_3, the information matrix for the OA is $\mathbf{M} = 18\mathbf{I}_{10}$, that is, all entries in $\mathbf{u}_{(23)}$ are equal to zero.

Suppose $t = 2$ and we seek to maximise the bound on $|\mathbf{\Psi}_B|$. Thus, we require two runs in which neither run is set to the middle level in either F_2 or F_3. Then the diagonal entries of $\mathbf{\Psi}_B$ would be $\left(\alpha + \frac{9}{4}\right)/N - 1 = -\frac{3}{8}$, hence the upper bound on

$|\mathbf{\Psi}_B|$ would be $\left(-\frac{3}{8}\right)^2 = \frac{9}{64}$. As in Example 4.2.2, we would also like the pair of runs to have Hamming distance 4 so that the off-diagonal entries of $\mathbf{\Psi}_B$ are as close to 0 as possible. Hence, the requirements here are the same as the analogous scenario when we are augmenting runs, so we can choose the same pair of runs as we did in Example 4.2.2, that is, the 5th and 14th. The only difference is that the diagonal entries of $\mathbf{\Psi}_B$ are 2 less than the diagonal entries of $\mathbf{\Psi}_A$. The off-diagonal entries remain the same. Hence, $\mathbf{\Psi}_B = \begin{bmatrix} -\frac{3}{8} & -\frac{1}{24} \\ -\frac{1}{24} & -\frac{3}{8} \end{bmatrix}$, and $|\mathbf{\Psi}_B| = \frac{5}{36}$, which is 0.9877 of the upper bound in this setting. The determinant of the information matrix for the augmented design is $|\mathbf{M_B}| = (-1)^2|\mathbf{M}||\mathbf{\Psi}_B| = 1 \times 18^{10} \times \frac{5}{36} = 495,898,225,920$. □

Example 4.2.6. Following Example 4.2.5, suppose rather than choosing two runs in which neither run is set to the middle level in either F_2 or F_3, we instead choose a pair of runs such that both are set to the middle level in either F_2 or F_3. The diagonal entries of $\mathbf{\Psi}_B$ are now $\frac{\alpha}{N} - 1 = -\frac{1}{2}$ hence the upper bound on $|\mathbf{\Psi}_B|$ is $\left(-\frac{1}{2}\right)^2 = \frac{1}{4}$ in this setting. Again, we will use the same runs we chose in the analogous situation in Example 4.2.3, that is, the 1st and 3rd runs. Hence, $\mathbf{\Psi}_B = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}$, and $|\mathbf{\Psi}_B| = \frac{1}{4}$, which is the upper bound in this setting. The determinant of the information matrix for the augmented design is $|\mathbf{M_B}| = (-1)^2|\mathbf{M}||\mathbf{\Psi}_B| = 1 \times 18^{10} \times \frac{1}{4} = 892,616,806,656$. □

Recall that when we were considering adjoining runs, we found it was generally better to search for runs in which neither of the factors involved in the interaction is set to the middle level so as to maximise the diagonal entries of $\mathbf{\Psi}_A$. That is, we chose runs such that the diagonal entries of $\mathbf{\Psi}_A$ were $1 + \left(\alpha + \frac{9}{4}\right)/N$, rather than $1 + \alpha/N$ which would be that case if at least one of the factors involved in the interaction was set to the middle level. In the case of missing runs, however, the situation is reversed, as the previous two examples illustrate. Recall that when runs are adjoined we add the identity to $\mathbf{AM}^{-1}\mathbf{A}'$, whereas when run are missing we subtract the identity from $\mathbf{BM}^{-1}\mathbf{B}'$. Hence, the diagonal entries of $\mathbf{\Psi}_B$ are 2 less than that of $\mathbf{\Psi}_A$. Thus, the diagonal entries of $\mathbf{\Psi}_B$ are $\left(\alpha + \frac{9}{4}\right)/N - 1$ when neither of the factors involved in the interaction is set to the middle level, or $\alpha/N - 1$ when at least one of the factors involved in the interaction is set to the middle level. Hence, when $\alpha = 9$ and $N = 18$, these values are negative. When $t = 2$ we seek to maximise the absolute value of the diagonal entries, hence $\alpha/N - 1$ is preferred over $\left(\alpha + \frac{9}{4}\right)/N - 1$. As we will see in Section 4.4.2, these 'rules' work most of the time, but given the complicated nature of the determinants, there are no general rules that work in all situations.

For completeness, we will now consider the best possible $|\mathbf{M}_B|$ across all OA$[18, 3^4]$s

for all possible interactions and all possible pairs of removed runs.

Example 4.2.7. We completed an exhaustive search of all 137 OA$[18, 3^4]$ from Chapter 3, for each of the $\binom{4}{2} = 6$ possible interactions to be included in the model, and each of the $\binom{18}{2} = 153$ pairs of runs to be removed. The maximum of these $137 \times 6 \times 153 = 125,766$ determinants is 892,616,806,656 which occurs 538 times. The design Example 4.2.6 is one such example. □

The Class of Competing Designs

In Section 2.1 we defined the class of competing designs as follows.

> Suppose we require a $s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}$ design in n runs, and an OA$[N, s_1^{m_1} \times s_2^{m_2} \times \ldots \times s_k^{m_k}]$ exists such that n is only slightly larger than N. Hence, we will adjoin $p = n - N$ runs to this OA, which we refer to as an *OA plus p runs* design. The class of completing designs are all designs that can be obtained by adjoining any p runs to the OA.

As we have seen above, we can no longer use theoretical results to derive general rules that suit all situations. Hence, in this chapter we will narrow our scope to symmetric three-level OAs in 18 runs. As noted in Chapter 2, these designs are of particular interest when we want to simultaneously allow for some curvature in the model (that is, we want to include factors with more than two levels) and minimise the number of required runs. This was our motivation in Chapter 2 for examining the distribution of pairwise Hamming distances of each of the combinatorially non-isomorphic OA$[18, 3^m]$s.

Now that we are interested in models that contain a single linear-by-linear interaction, we need to broaden our investigation from combinatorial isomorphism to geometric isomorphism. The enumeration of all geometric isomorphism classes for OA$[18, 3^m]$s in Chapter 3 allows us to carry out exhaustive calculations for the entire design space. Thus, although we may not be able to derive definitive rules, we hope to be able to develop some general guidelines and insights into sensible design construction procedures. Additionally, an exhaustive empirical examination allow us to make explicit recommendations about the construction of optimal OA$[18, 3^m]$ plus p runs designs ($p = 1$ or 2), and OA$[18, 3^m]$ minus t runs designs ($t = 1$ or 2).

We will limit the set of runs to choose from to the runs within the OA. That is, rather than choosing any p runs from the 3^m runs in the complete factorial, we will only consider the 18 runs in the OA. There are three reasons for this.

1. We are interested in the comparison between adjoined and removed runs, but it is only possible to remove runs that are within the OA.
2. Adjoining runs that are already in the OA guarantees repeated runs for an estimate of pure error.
3. As m increases, an exhaustive search of all sets of p runs in the complete factorial becomes computationally prohibitive. For example, when $m = 7$ there are $3^7 = 2,187$ runs in the complete factorial and $\binom{7}{2} = 21$ possible interactions to include in the model, hence for, say $p = 2$, there are $21\binom{2187}{2} = 50,198,211$ combinations to consider for each of the 291 geometric isomorphism classes.

The inclusion of an interaction term in the model also raises some distinctions that are worth clarifying here. Recall that the choice of p runs to adjoin to an OA under a main effects only model can be made independently of which OA we are using. As we have shown in this section, this is no longer the case when we include an interaction term in the model as the determinant of the information matrix for the OA can change from model to model and design to design. Hence, we need to simultaneously consider the choice of the OA, the adjoined runs, and the interaction. So, rather than posing our research question simply in terms of the optimal runs to adjoin, the class of competing designs is informed by the question: what is the best OA$[18, 3^m]$ plus p runs design, and which interaction should we include in the model[3], given we can choose any OA$[18, 3^m]$ in the design space and any p runs from within that OA?

Similarly, in the context of missing runs, the class of competing designs is all OA$[18, 3^m]$ minus t runs designs for all models that include a single linear-by-linear interaction.

Efficiency

The examples given above have illustrated why we are not able to derive a practical expression for the theoretical bound on $|\mathbf{M}_A|$ and $|\mathbf{M}_B|$. Hence, as we stated at the outset of this section, we will re-define efficiency in terms of the best realisable determinant rather than the theoretical best. Such an adjustment would not have been possible in Chapter 2 since we were making general constructions for a vast number of design spaces, so the exhaustive enumeration of all determinants would not have been feasible. For the remainder of this chapter, however, since we are

[3]In practice, the question "which interaction should we include in the model?" is equivalent to "given I suspect there might be a linear-by-linear interaction between factors F_a and F_b, which columns should I assign these factors to?"

focusing on the OA$[18, 3^m]$s enumerated in Chapter 3, an exhaustive computation is possible.

Let $|\mathbf{M}|$ be the determinant of the information matrix for design D, and $|\mathbf{M}|_{opt}$ be the optimal determinant, that is, the largest determinant of the information matrix from all designs in the same class as D. Then following Equation (2.1.1), we re-define the efficiency of design D as

$$\left(\frac{|\mathbf{M}|}{|\mathbf{M}|_{opt}}\right)^{1/\upsilon}. \tag{4.2.7}$$

The only noticeable difference between this definition and that in Equation (2.1.1) is that we have changed α to υ to reflect the additional term in the model.

Example 4.2.8. In Examples 4.2.2 and 4.2.3 we constructed two OA$[18, 3^4]$ plus 2 runs designs with $|\mathbf{M}_A| = 9,422,066,292,480$ and 8,033,551,259,904 respectively. We then showed in Example 4.2.4 that the best realisable $|\mathbf{M}_A|$ for an OA$[18, 3^4]$ plus 2 runs design is 9,744,400,139,328. Hence, the efficiency of the two designs in Examples 4.2.2 and 4.2.3 are $\left(\frac{9,422,066,292,480}{9,744,400,139,328}\right)^{1/10} = 99.6642\%$ and $\left(\frac{8,033,551,259,904}{9,744,400,139,328}\right)^{1/10} = 98.0879\%$ respectively. □

Example 4.2.9. In Examples 4.2.5 and 4.2.6 we constructed two OA$[18, 3^4]$ minus 2 runs designs with $|\mathbf{M}_B| = 495,898,225,920$ and 892,616,806,656 respectively. We then showed in Example 4.2.7 that the best realisable $|\mathbf{M}_B|$ for an OA$[18, 3^4]$ minus 2 runs design is 892,616,806,656. Hence, the efficiency of the two designs in Examples 4.2.5 and 4.2.6 are $\left(\frac{495,898,225,920}{892,616,806,656}\right)^{1/10} = 94.2915\%$ and $\left(\frac{892,616,806,656}{892,616,806,656}\right)^{1/10} = 100\%$ respectively. □

4.3 Entries in $\mathbf{u}_{(ab)}$

We showed in the previous section that the optimal values of $|\mathbf{M}_A|$ or $|\mathbf{M}_B|$ depend, amongst other things, on the off-diagonal entries in $\mathbf{M}$, that is, the entries in $\mathbf{u}_{(ab)}$. In this section we exhaustively enumerate these values for all OA$[18, 3^m]$ and make some observations about the effect on these values of moving between different OAs in the same geometric isomorphism class.

Recall from Section 4.1 that the inner product of $\mathbf{z}_{(ab)}$ with the intercept and each of the main effect components of the factors involved in the interaction will be zero, hence there are only $m-2$ pairs of potentially non-zero entries in $\mathbf{u}_{(ab)}$ to consider: $(u_{(ab)cl}, u_{(ab)cq}) = (\mathbf{z}_{(ab)}\mathbf{f}_{cl}, \mathbf{z}_{(ab)}\mathbf{f}_{cq})$, $c \in (1, 2 \ldots, m) \backslash (a, b)$. Let $z_{(ab)_i}$

be the ith entry in $\mathbf{z}_{(ab)}$. Then $u_{(ab)cl} = \sum_{i=1}^{18} z_{(ab)_i} f_{cl_i} = \sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cl_i}$ and $u_{(ab)cq} = \sum_{i=1}^{18} z_{(ab)_i} f_{cq_i} = \sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cq_i}$. Notice that $u_{(ab)cl}$, $u_{(ac)bl}$ and $u_{(bc)al}$ must be equal since all three are the sum of the product of the linear components of F_a, F_b and F_c. We cannot say the same for $u_{(ab)cq}$, $u_{(ac)bq}$ and $u_{(bc)aq}$ however, as they involve two linear components and one quadratic component, and the quadratic component changes across these three terms.

We consider the plausible values of the linear and quadratic terms, that is $u_{(ab)cl}$ and $u_{(ab)cq}$, in Sections 4.3.1 and 4.3.2 respectively. We then make some observations about the realisable pairs of $(u_{(ab)cl}, u_{(ab)cq})$ in Section 4.3.3.

4.3.1 Linear Terms

We will begin by considering what possible values the sum $\sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cl_i}$ can take. Recall that the entries in the linear main effect of a three-level factor are $-\sqrt{3/2}$ for the low level, 0 for the middle level, and $\sqrt{3/2}$ for the high level. Hence, the contribution to the sum when any of F_a, F_b or F_c are set to the middle level will be zero, and so we are only concerned with runs in which the levels of all three factors are either high or low.

Suppose we take any triple of columns from an OA$[18, 3^m]$ and order the rows lexicographically. Then we can immediately discard 10 of the rows from the summation based on the instances of the middle level in the first two columns. This is illustrated in Table 4.3.1 below.

The following examples investigate how many runs can be set to the middle level in the third factor.

Example 4.3.1. Suppose we would like to maximise the number of runs that contribute to the sum. We begin with two copies of all pairs of levels for the first two factors, in lexicographic order, as illustrated in the first array in Table 4.3.1. We would like the middle level of F_3 to appear in as few runs that have not already been ruled out as possible. We know that the middle level must appear in F_3 exactly twice in the 13th to 18th runs in order for the pair of levels 21 to appear exactly twice between F_1 and F_3, so we set F_3 to the middle level in the 15th and 16th runs, which have already been ruled out as contributing to the sum. This means we have exhausted our allocation for the pair of levels 11 between F_2 and F_3. We also know that the middle level must appear in F_3 exactly twice in the 1st to 6th runs in order for the pair of levels 01 to appear exactly twice between F_1 and F_3, but we cannot use the 3rd or 4th runs since this would produce another pair of levels 11 between F_2 and F_3. Hence, both instances of the middle level in the 1st to 6th runs must appear in two of the four rows that have not yet been

Table 4.3.1: Consider the number of triples without the middle level

	Initial†			**OA$_1$**			**OA$_2$**		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
1	0	0	?	~~0~~	~~0~~	~~1~~	0	0	0
2	0	0	?	~~0~~	~~0~~	~~1~~	0	0	0
3	~~0~~	~~1~~	~~?~~	~~0~~	~~1~~	~~0~~	~~0~~	~~1~~	~~2~~
4	~~0~~	~~1~~	~~?~~	~~0~~	~~1~~	~~2~~	~~0~~	~~1~~	~~2~~
5	0	2	?	0	2	0	~~0~~	~~2~~	~~1~~
6	0	2	?	0	2	2	~~0~~	~~2~~	~~1~~
7	~~1~~	~~0~~	~~?~~	~~1~~	~~0~~	~~0~~	~~1~~	~~0~~	~~2~~
8	~~1~~	~~0~~	~~?~~	~~1~~	~~0~~	~~2~~	~~1~~	~~0~~	~~2~~
9	~~1~~	~~1~~	~~?~~	~~1~~	~~1~~	~~0~~	~~1~~	~~1~~	~~1~~
10	~~1~~	~~1~~	~~?~~	~~1~~	~~1~~	~~2~~	~~1~~	~~1~~	~~1~~
11	~~1~~	~~2~~	~~?~~	~~1~~	~~2~~	~~1~~	~~1~~	~~2~~	~~0~~
12	~~1~~	~~2~~	~~?~~	~~1~~	~~2~~	~~1~~	~~1~~	~~2~~	~~0~~
13	2	0	?	2	0	0	~~2~~	~~0~~	~~1~~
14	2	0	?	2	0	2	~~2~~	~~0~~	~~1~~
15	~~2~~	~~1~~	~~?~~	~~2~~	~~1~~	~~1~~	2	1	0
16	~~2~~	~~1~~	~~?~~	~~2~~	~~1~~	~~1~~	~~2~~	~~1~~	~~0~~
17	2	2	?	2	2	0	2	2	2
18	2	2	?	2	2	2	2	2	2

†*Start by discarding rows that contain the middle level in F_1 or F_2*

OA$_1$ *maximises the number of rows with no middle level*

OA$_2$ *minimises the number of rows with no middle level*

ruled out as contributing to the sum, thereby ruling out two more runs. OA$_1$ in Table 4.3.1, which is a representative of the first geometric isomorphism class for OA$[18, 3^3]$ designs, is an example of three columns with this structure. □

The previous example has shown that we can have at most 6 runs in three columns of an OA$[18, 3^m]$ that do not contain the middle level in any factor. Hence, we can have at most 6 non-zero terms in the sum $\sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cl_i}$.

Example 4.3.2. Suppose we would like to minimise the number of runs that contribute to the sum $\sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cl_i}$, that is, we would like to set F_3 to the middle level in as many of the rows that have not already been ruled out as possible. Since the middle level of F_3 must appear exactly twice in each of the first six runs and each of the last six runs, we can rule out a maximum of 4 additional runs. Consider OA$_2$ in Table 4.3.1, which is a representative of the 13th geometric isomorphism class for OA$[18, 3^3]$ designs. Four of the 8 rows that were not initially ruled out are set to the middle level in the third column, hence these rows will be ruled out at well. Hence, only 4 rows will contribute to the sum. □

Based on the previous two examples we know that there will be somewhere between 4 and 6 non-zero terms in the sum $\sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cl_i}$. Since $f_{al_i} = -\sqrt{3/2}$ or $\sqrt{3/2}$

when F_a is set to the low or high level respectively, $f_{al_i} f_{bl_i} f_{cl_i} = \pm\frac{3}{2}\sqrt{\frac{3}{2}}$ where the sign depends on the number of factors that are set to the low level in the ith run. When there are 0 or 2 occurrences of the low level in a given triple (that is, the triples 000, 022, 202 and 220) then $f_{al_i} f_{bl_i} f_{cl_i}$ will be positive. Otherwise $f_{al_i} f_{bl_i} f_{cl_i}$ will be negative. These concepts are formalised in the following Lemma.

Lemma 4.3.1. Consider any three columns from an OA$[18, 3^m]$, which we will label F_a, F_b, and F_c respectively. Let γ_{abc+} be the number of occurrences of any of the triples (002, 020, 200, 222), and γ_{abc-} be the number of occurrences of any of the triples (000, 022, 202, 220). Then $u_{(ab)cl} = u_{(ac)bl} = u_{(bc)al} = \sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cl_i} = \gamma_{abc}\frac{3}{2}\sqrt{\frac{3}{2}}$ where $\gamma_{abc} = \gamma_{abc+} - \gamma_{abc-}$. □

We have already established that there will be somewhere between 4 and 6 non-zero terms in the sum $\sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cl_i}$, hence $4 \leq \gamma_{abc+} + \gamma_{abc-} \leq 6$. Note that if we were to reverse the levels of a single factor or all three factors then the values of both γ_{abc+} and γ_{abc-} would be reversed, hence γ_{abc} would stay the same in absolute value, but the sign would change. If, however, we were to reverse the levels of any two factors then both γ_{abc+} and γ_{abc-} would remain unchanged. The following example illustrates this.

Example 4.3.3. Suppose a set of three columns from an OA$[18, 3^m]$ contained the following 5 triples in which none of the factors are set to the middle level: 020, 022, 200, 202, and 220. By counting the number of occurrences of any of the triples (002, 020, 200, 222) or (000, 022, 202, 220) respectively, we see that $\gamma_{abc+} = 2$ and $\gamma_{abc-} = 3$, hence $\gamma_{abc} = 2 - 3 = -1$. Table 4.3.2 below gives the values of γ_{abc+}, γ_{abc-} and γ_{abc} when the levels within each possible subset of columns is reversed. □

Table 4.3.2: Effect of reversing levels on the value of γ_{abc}

Reversed Factors	Runs (sets of levels)	γ_{abc+}	γ_{abc-}	γ_{abc}
None (original design)	020, 022, 200, 202, 220	2	3	−1
F_1	220, 222, 000, 002, 020	3	2	1
F_2	000, 002, 220, 222, 200	3	2	1
F_3	022, 020, 202, 200, 222	3	2	1
F_1, F_2	200, 202, 020, 022, 000	2	3	−1
F_1, F_3	222, 220, 002, 000, 022	2	3	−1
F_2, F_3	002, 000, 222, 220, 202	2	3	−1
F_1, F_2, F_3	202, 200, 022, 020, 002	3	2	1

Recall that reversing the levels of a factor is allowed under geometric isomorphism, hence any two geometrically isomorphic designs will have the same $|\gamma_{abc}|$ but the

sign may be different. As we shall see in Section 4.3.3, the sign of γ_{abc} does not affect $|\mathbf{M}_A|$ or $|\mathbf{M}_B|$, hence it is irrelevant for our purposes. We have exhaustively enumerated γ_{abc+}, γ_{abc-} and γ_{abc} for all triples of columns in a representative of each OA$[18, 3^m]$ geometric isomorphism classes enumerated in Chapter 3. We present the values for $m = 3$ in Table 4.3.3 below. Since any triple of runs from an OA$[18, 3^m]$ with $m > 3$ will be geometrically isomorphic to one of the OA$[18, 3^3]$ classes, all $\binom{m}{3}$ sets of columns are associated with one of the $|\gamma_{abc}|$ in Table 4.3.3. As discussed above, the values for γ_{abc+} and γ_{abc-} cited here are specific to the particular representative we used from each OA$[18, 3^3]$ geometric isomorphism classes, hence they can be interchanged by reversing 1 or 3 of the factors.

Table 4.3.3: Values of $|\gamma_{abc}|$ for each OA$[18, 3^3]$ geometric isomorphism class

Class	γ_{abc+}	γ_{abc-}	$\lvert\gamma_{abc}\rvert$	**Class**	γ_{abc+}	γ_{abc-}	$\lvert\gamma_{abc}\rvert$
1	3	3	0	**8**	1	4	3
2	2	3	1	**9**	2	4	2
3	4	1	3	**10**	0	4	4
4	2	2	0	**11**	3	3	0
5	1	5	4	**12**	0	6	6
6	0	6	6	**13**	2	2	0
7	0	5	5				

When $m = 3$ there is only one value to consider, $|\gamma_{123}|$. However when $m > 3$ we have $\binom{m}{3}$ values of $|\gamma_{abc}|$ to calculate. The following example illustrates this.

Example 4.3.4. When $m = 4$ there are $\binom{4}{3} = 4$ triples of columns to consider. We found through exhaustive checking of one representative from the first OA$[18, 3^4]$ geometric isomorphism class that $\gamma_{123} = 0$, $\gamma_{124} = 0$, $\gamma_{134} = 2$ and $\gamma_{234} = -2$. Hence, for the particular representative we used[4], $u_{(12)3l} = u_{(13)2l} = u_{(23)1l} = u_{(12)4l} = u_{(14)2l} = u_{(24)1l} = 0$, $u_{(13)4l} = u_{(14)3l} = u_{(34)1l} = 2 \times \frac{3}{2}\sqrt{\frac{3}{2}} = 3\sqrt{\frac{3}{2}}$ and $u_{(23)4l} = u_{(24)3l} = u_{(34)2l} = -2 \times \frac{3}{2}\sqrt{\frac{3}{2}} = -3\sqrt{\frac{3}{2}}$. If we were to choose any other representative from the same class, then we know $\gamma_{123} = \gamma_{124} = 0$, and $|\gamma_{134}| = |\gamma_{234}| = 2$. □

4.3.2 Quadratic Terms

As noted at the beginning of this section, the values $u_{(ab)cq}$, $u_{(ac)bq}$ and $u_{(bc)aq}$ are not necessarily equal since the quadratic component varies across these three terms. The quadratic main effects are $1/\sqrt{2}$ when the factor is set to either the

[4]Note that we are referring to specific columns here, so when columns are permuted (which is allowed under geometric isomorphism) the labels in the subscript of u need to be updated to reflect the new column positions

high or low levels, and $-\sqrt{2}$ when the factor is set to the middle level. Since $u_{(ab)cq} = \mathbf{z}_{(ab)}\mathbf{f}_{cq} = \sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cq_i}$, we can disregard the runs that are set to the middle level in either of factors F_a or F_b, as we did above when considering the linear terms, but we can no longer disregard the runs that are set to the middle level of factor F_c. Hence, as we can see from the initial set of three columns in Table 4.3.1, assuming the interaction is between factors F_1 and F_2, there will be exactly 8 non-zero terms to consider in the sum.

If the pair of factors F_a and F_b are set to 00 or 22 in the ith run, then $f_{al_i} f_{bl_i} = \frac{3}{2}$, otherwise $f_{al_i} f_{bl_i} = -\frac{3}{2}$, hence:

$$f_{al_i} f_{bl_i} f_{cq_i} = \begin{cases} \frac{3}{2} \times -\sqrt{2} = -\frac{3\sqrt{2}}{2} & \text{if } (F_a, F_b) \in (00, 22) \text{ and } F_c = 1 \\ \frac{3}{2} \times \frac{1}{\sqrt{2}} = \frac{3\sqrt{2}}{4} & \text{if } (F_a, F_b) \in (00, 22) \text{ and } F_c \neq 1 \\ -\frac{3}{2} \times -\sqrt{2} = \frac{3\sqrt{2}}{2} & \text{if } (F_a, F_b) \in (02, 20) \text{ and } F_c = 1 \\ -\frac{3}{2} \times \frac{1}{\sqrt{2}} = -\frac{3\sqrt{2}}{4} & \text{if } (F_a, F_b) \in (02, 20) \text{ and } F_c \neq 1 \\ 0 \quad \text{otherwise} \end{cases} \tag{4.3.1}$$

In Lemma 4.3.1 we calculated the value of $u_{(ab)cl}$ by counting the number of runs that belonged to one of two sets. As we can see from Equation (4.3.1), the calculation of $u_{(ab)cq}$ will require four sets of runs to be considered.

Lemma 4.3.2. Consider any three columns from an OA$[18, 3^m]$, which we will label F_a, F_b and F_c, in that order, and assume the interaction is between, F_a and F_b. Let $\omega_{(ab)c-2}$ be the number of occurrences of any of the triples (001, 221), $\omega_{(ab)c+1}$ be the number of occurrences of any of the triples (000, 002, 220, 222), $\omega_{(ab)c+2}$ be the number of occurrences of any of the triples (021, 201), and $\omega_{(ab)c-1}$ be the number of occurrences of any of the triples (020, 022, 200, 202). Then $u_{(ab)cq} = \sum_{i=1}^{18} f_{al_i} f_{bl_i} f_{cq_i} = \omega_{(ab)c} \frac{3\sqrt{2}}{4}$ where $\omega_{(ab)c} = \omega_{(ab)c+1} - \omega_{(ab)c-1} + 2\left(\omega_{(ab)c+2} - \omega_{(ab)c-2}\right)$. □

Aside from the additional sets of runs to consider, a notable difference between Lemmas 4.3.1 and 4.3.2 is that Lemma 4.3.1 calculates a single value, γ_{abc}, which can be applied to the triple of columns regardless of which pair is involved in the interaction, where as the value calculated in Lemma 4.3.2, $\omega_{(ab)c}$, corresponds specifically to the interaction between F_a and F_b, and so the values $\omega_{(ac)b}$ and $\omega_{(bc)a}$ require additional calculations. The following example steps through the entire process of using the results in this section to calculate the values of $\mathbf{u}_{(ab)}$.

Example 4.3.5. Consider the OA$[18, 3^3]$ in Table 4.3.4 below (which is the same as OA_1 in Table 4.3.1). This table shows how each run contributes to each of

γ_{123}, $\omega_{(12)3}$, $\omega_{(13)2}$ and $\omega_{(23)1}$. Since $\gamma_{123} = 0$, we know the entry in $u_{(ab)cl}$ will be 0 regardless of which pair of factors are involved in the interaction. $u_{(ab)cq}$ will also be 0 when the interaction is between either the first and third factors or the second and third factors, but when the interaction is between the first two factors $u_{(12)3q} = 6 \times \frac{3\sqrt{2}}{4} = -\frac{9\sqrt{2}}{2}$. Thus, $\mathbf{u}_{(13)}$ and $\mathbf{u}_{(23)}$ are both columns vectors of length $\alpha = 1 + 2 \times 3 = 7$ with all entries equal to zero, and

$$\begin{aligned}\mathbf{u}'_{(12)} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & u_{(12)3l} & u_{(12)3q} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -\frac{9\sqrt{2}}{2} \end{bmatrix}.\end{aligned}$$

□

Table 4.3.4: Calculation of γ_{abc} and $\omega_{(ab)c}$ for an OA$[18, 3^3]$ from the first geometric isomorphism class

	$\mathbf{OA}_1$			γ_{abc}	$\omega_{(ab)c}$		
	F_1	F_2	F_3	γ_{123}	$\omega_{(12)3}$	$\omega_{(13)2}$	$\omega_{(23)1}$
1	0	0	1	0	−2	0	0
2	0	0	1	0	−2	0	0
3	0	1	0	0	0	−2	0
4	0	1	2	0	0	2	0
5	0	2	0	1	−1	1	−1
6	0	2	2	−1	−1	−1	1
7	1	0	0	0	0	0	−2
8	1	0	2	0	0	0	2
9	1	1	0	0	0	0	0
10	1	1	2	0	0	0	0
11	1	2	1	0	0	0	0
12	1	2	1	0	0	0	0
13	2	0	0	1	−1	−1	1
14	2	0	2	−1	−1	1	−1
15	2	1	1	0	0	0	0
16	2	1	1	0	0	0	0
17	2	2	0	−1	1	−1	−1
18	2	2	2	1	1	1	1
			Sum:	**0**	**−6**	**0**	**0**

We will now consider the effect on $\omega_{(ab)c}$ when the levels of one or more of the columns in the triple are reversed. Suppose for a given triple of ordered columns, F_a, F_b and F_c, the interaction is between factors F_a and F_b. We first consider what happens when we reverse the levels of F_c. We can see from Equation (4.3.1) regarding F_c that we are only interested in whether or not this factor is set to 1. Reversing the levels of a factor will not effect any runs set to the middle level (reversing only swaps high and low), hence reversing the levels of F_c has no effect on $\omega_{(ab)c}$.

Now suppose we reverse the levels of just one factor involved in the interaction, say F_a, and consider what happens to the counts in Lemma 4.3.2. Recall that $\omega_{(ab)c-2}$ is the number of occurrences of any of the triples (001, 221). When the levels of F_a are reversed, these runs become (201, 021), which is the set of runs being counted by $\omega_{(ab)c+2}$, hence the values of $\omega_{(ab)c-2}$ and $\omega_{(ab)c+2}$ are swapped. Similarly, $\omega_{(ab)c+1}$ is the number of occurrences of any of the triples (000, 002, 220, 222), which becomes (200, 202, 020, 022) when the levels of F_a are reversed. This is equivalent to the set being counted by $\omega_{(ab)c-1}$, hence $\omega_{(ab)c+1}$ and $\omega_{(ab)c-1}$ swapped when the levels of F_a are reversed. Thus, the sign of all coefficients in the expression for $\omega_{(ab)c}$ are reversed, hence the absolute value of $\omega_{(ab)c}$ remains the same, but the sign changes. The same arguments can be applied for reversing the levels of F_b.

Finally, suppose we reverse the levels of both F_a and F_b. This can be considered in sequential steps. First reverse the levels of F_a, hence the sign of $\omega_{(ab)c}$ changes. Then reverse the levels of F_b, hence the sign of $\omega_{(ab)c}$ changes back again. Thus, there is no net change in $\omega_{(ab)c}$ when the levels of both F_a and F_b are reversed.

The following example illustrates these results. To simplify the number of runs to be examined in this example, recall that the 8 runs that contribute to $\omega_{(ab)c}$ are those that are not set to the middle level in either F_a or F_b, hence reversing the levels of either of these factors will neither remove nor introduce any new runs to this set of runs to be considered. Thus, in order to consider the effect of reversing the levels of F_a or F_b, we need only consider the set of 8 runs that contribute to $\omega_{(ab)c}$.

Example 4.3.6. Consider OA_1 in Table 4.3.4 for which $\omega_{(12)3} = -6$. The 8 runs that contribute to $\omega_{(12)3}$ are 001, 001, 020, 022, 200, 202, 220 and 222. The table below gives the values of $\omega_{(12)3}$ when the levels of the two factors involved in the interaction are reversed. We can see that $\omega_{(12)3}$ is not effected by both factors in the interaction being reversed, but it changes sign when only one of the factors is reversed. □

Table 4.3.5: Effect of reversing levels on the value of $\omega_{(12)3}$

Reversed Factors	Runs (sets of levels)	$\omega_{(12)3}$
None (original design)	001, 001, 020, 022, 200, 202, 220, 222	-6
F_1	201, 201, 220, 222, 000, 002, 020, 022	6
F_2	021, 021, 000, 002, 220, 222, 200, 202	6
F_1, F_2	221, 221, 200, 202, 020, 022, 000, 002	-6

The previous example focused specifically on the non-zero value of $\omega_{(ab)c}$, that is $\omega_{(12)3} = -6$, because changing the sign of $\omega_{(ab)c}$ is meaningless when $\omega_{(ab)c} = 0$.

Hence, the fact that $\omega_{(13)2} = \omega_{(23)1} = 0$ effectively means that reversing the levels of F_3 has no effect on the design. In order to discuss how this situation arises, we will first present the values of all $|\omega_{(ab)c}|$, $|\omega_{(ac)b}|$ and $|\omega_{(bc)a}|$ for each class of OA$[18, 3^3]$. We present these values in Table 4.3.6 along with the values of $|\gamma_{abc}|$ from Table 4.3.3 to aid further discussion.

Table 4.3.6: $|\gamma_{abc}|$ and $|\omega_{(ab)c}|$ for each OA$[18, 3^3]$ geometric isomorphism class

Class	$\lvert\gamma_{abc}\rvert$	$\lvert\omega_{(ab)c}\rvert$	$c^\dagger$	**Class**	$\lvert\gamma_{abc}\rvert$	$\lvert\omega_{(ab)c}\rvert$	$c^\dagger$
1	0	6	3	**8**	3	3	3
		0	2			3	2
		0	1			3	1
2	1	9	3	**9**	2	0	3
		3	2			0	2
		3	1			6	1
3	3	9	3	**10**	4	0	3
		3	2			0	2
		9	1			0	1
4	0	12	3	**11**	0	6	3
		0	2			6	2
		0	1			6	1
5	4	6	3	**12**	6	6	3
		6	2			6	2
		6	1			6	1
6	6	0	3	**13**	0	12	3
		6	2			12	2
		0	1			12	1
7	5	3	3				
		3	2				
		3	1				

$^\dagger c$ *indicates* F_c*, that is, the column number of the quadratic factor not included in the interaction*

Recall from Chapter 3 that some of the columns we used to construct the OA$[18, 3^3]$s remain unchanged after the levels are reversed. These columns, which are marked with (*) in Table A5 of the appendix, appear in the 1st and 4th geometric isomorphism classes for three-factor designs. Notice that for both of these classes, in addition to having two $\omega_{(ab)c}$ values equal to zero, γ_{abc} is also equal to zero. This is because if we reverse the levels of the factor that is unchanged by this action (F_3 in both classes), then the design remains exactly the same, hence the values of γ_{123}, $\omega_{(13)2}$ and $\omega_{(23)1}$ must also stay the same, but the sign of these variables must also change. Of course, the only way a variable can change sign yet remain the same is if it is equal to zero.

We note that these are not the only zero values in Table 4.3.6. In order to discuss how the other zero values have arisen, it will be helpful to summarise the findings in this section for both linear and quadratic terms in a single table.

Table 4.3.7: Effect on γ_{abc} and $\omega_{(ab)c}$ of reversing levels of factors

Reversed Factors	γ_{abc}	$\omega_{(ab)c}$	Comments
None	γ_{abc}	$\omega_{(ab)c}$	Original design
$F_a^\dagger$	$-\gamma_{abc}$	$-\omega_{(ab)c}$	One of the factors involved in interaction
F_c	$-\gamma_{abc}$	$\omega_{(ab)c}$	Factor not involved in interaction
F_a and F_b	γ_{abc}	$\omega_{(ab)c}$	Both factors involved in interaction
$F_a^\dagger$ and F_c	γ_{abc}	$-\omega_{(ab)c}$	One factor involved in interaction, and one factor not involved in interaction
F_a and F_b and F_c	$-\gamma_{abc}$	$\omega_{(ab)c}$	All factors

$F_a^\dagger$ can be interchanged with F_b as they both represent a factor involved in the interaction

In addition to the 1st and 4th classes discussed above, notice that γ_{abc} also equals zero in the 11th and 13th classes. Although neither of these classes contain columns that remain the same when reversed, they both remain the same if all three columns are reversed simultaneously. That is, if we take one representative from either class, reverse the levels of all three factors, then reorder the rows lexicographically, we will end up with exactly the same design that we started with. We see from Table 4.3.7 that $\omega_{(ab)c}$ is not effected if all three factors are reversed, but the sign of γ_{abc} will change. Hence, since the design remains unchanged when the levels of all factors are reversed, γ_{abc} must equal zero.

Next, consider the 6th and 9th classes in which the values for $|\omega_{(ab)c}|$ are 0 or 6. Designs from each of these classes will remain unchanged if two of the columns are reversed. In the representative we used from the 6th class, the pair of columns that can be reversed together without changing the design are the first and third. From Table 4.3.7 we see that γ_{abc} is not effected if any two factors are reversed, but $\omega_{(ab)c}$ will change sign if one of the reversed factors is not involved in the interaction. Hence, in the 6th class, $\omega_{(12)3} = \omega_{(23)1} = 0$. If we reverse the levels of, say, the first and second factors, then the resulting design will be different (albeit isomorphic) to the original representative, hence $\omega_{(13)2}$, which was equal to -6 in the original design, will change sign. In the 9th class, $\omega_{(12)3} = \omega_{(13)2} = 0$, which implies the pair of columns that can be reversed together without changing the design is the second and third.

Finally, consider the 10th class. Designs in this class remain the same if any pair of factors are reversed simultaneously, hence all $\omega_{(ab)c}$ are equal to 0.

4.3.3 Properties of $|\gamma_{abc}|$ and $|\omega_{(ab)c}|$

We will now make some general observations about the enumerated values of $|\gamma_{abc}|$ and $|\omega_{(ab)c}|$ in Table 4.3.6 and then conclude this section with a proof that geometric

permutations as described in Table 4.3.7 do not effect the value of $|\mathbf{M}_A|$ or $|\mathbf{M}_B|$.

Notice from Table 4.3.6 that all triple of columns have either one or two distinct values of $|\omega_{(ab)c}|$, never three.

Another observation we can make about Table 4.3.6 is that the pairs of $(|\gamma_{abc}|, |\omega_{(ab)c}|)$ have the same parity. This is due to the number of occurrences of the triple of levels 111 (all factors set to the middle level) in the set of columns. Since the pair of levels 11 must appear exactly twice in each pair of columns, and there are $\binom{3}{2} = 3$ pairs of columns, then the pair of middle levels must appear $2 \times 3 = 6$ times throughout the triple of columns. The triple of levels 111 takes up 3 of these occurrences since it represents one instance for each of the 3 pairs of columns. Hence, if the triple of middle levels appears in the triple of columns twice then there will be no other occurrences of the middle level appearing more than once in any other run. In general, the middle level will appear in the triple of columns $3 \times 6 = 18$ times in total. Hence, when the triple 111 appears in the set of columns twice the remaining $18 - 2 \times 3 = 12$ middle levels will appear in 12 separate runs. If 111 appears once in the triple of columns then there will be 3 other runs in which two of the factors are set to the middle level, hence the remaining $18 - 1 \times 3 - 3 \times 2 = 9$ middle levels will appear in 9 separate runs. If 111 does not appear in the triple of columns at all then there will be 6 runs that contain the middle level twice and the remaining $18 - 6 \times 2 = 6$ middle levels will appear in 6 separate runs. Of course, the calculations of both $|\gamma_{abc}|$ and $|\omega_{(ab)c}|$ depend on the occurrences of the middle level. Runs that contain the middle level in any of the three factors do not contribute to the calculation of $|\gamma_{abc}|$ at all. Runs that contain the middle level in either with F_a or F_b do not contribute to the calculation of $|\omega_{(ab)c}|$, but those runs that are set to the middle level in the F_c do.

Before we enumerate $|\mathbf{M}_A|$ and $|\mathbf{M}_B|$ for one representative from each OA$[18, 3^m]$ geometric isomorphism class, we must first show that when we permute each OA to obtain another design in the same class, the change of sign in $|\gamma_{abc}|$ and/or $|\omega_{(ab)c}|$ as outlined in Table 4.3.7 does not effect the final value of $|\mathbf{M}_A|$ and $|\mathbf{M}_B|$. That is, we need to show that all designs in the same geometric isomorphism class have the same statistical properties.

Recall from Theorem 4.2.1 that for an OA$[N, 3^m]$ and a model containing the linear-by-linear component of the interaction between F_a and F_b, the entry in the ith row and jth column of $\mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$ is

$$\frac{1}{N}\left(q - \frac{\left(Nf_{al_i}f_{bl_i} - \mathbf{a}_{ME_i}\mathbf{u}_{(ab)}\right)\left(Nf_{al_j}f_{bl_j} - \mathbf{a}_{ME_j}\mathbf{u}_{(ab)}\right)}{\mathbf{u}'_{(ab)}\mathbf{u}_{(ab)} - N^2}\right). \tag{4.3.2}$$

For the purpose of this discussion, let us assume that there are three factors in the OA. $1/N$ and q are not effected by reversing the levels of any of the factors, hence we are only concerned with the fraction in (4.3.2) which can be written as

$$\frac{\left(Nf_{al_i}f_{bl_i} - \left(f_{cl_i}u_{(ab)cl} + f_{cq_i}u_{(ab)cq}\right)\right)\left(Nf_{al_j}f_{bl_j} - \left(f_{cl_j}u_{(ab)cl} + f_{cq_j}u_{(ab)cq}\right)\right)}{u^2_{(ab)cl} + u^2_{(ab)cq} - N^2}$$
$$= \frac{\left(Nf_{al_i}f_{bl_i} - \left(\frac{3}{2}\sqrt{\frac{3}{2}}f_{cl_i}\gamma_{abc} + \frac{3\sqrt{2}}{4}f_{cq_i}\omega_{(ab)c}\right)\right)\left(Nf_{al_j}f_{bl_j} - \left(\frac{3}{2}\sqrt{\frac{3}{2}}f_{cl_j}\gamma_{abc} + \frac{3\sqrt{2}}{4}f_{cq_j}\omega_{(ab)c}\right)\right)}{\frac{27}{8}\gamma^2_{abc} + \frac{9}{8}\omega^2_{(ab)c} - N^2}. \tag{4.3.3}$$

Assume we start with one representative from a particular geometric isomorphism class with a particular pair of values $(\gamma_{abc}, \omega_{(ab)c})$. Let us consider what will happen to each of the terms in the equation above when the levels of one or more of the columns in the OA are reversed. We note that a change of sign in either γ_{abc} and/or $\omega_{(ab)c}$ does not effect the value of the denominator, $\frac{27}{8}\gamma^2_{abc} + \frac{9}{8}\omega^2_{(ab)c} - N^2$. To simplify the notation throughout this discussion, let

$$G(i) = Nf_{al_i}f_{bl_i} - \left(\frac{3}{2}\sqrt{\frac{3}{2}}f_{cl_i}\gamma_{abc} + \frac{3\sqrt{2}}{4}f_{cq_i}\omega_{(ab)c}\right). \tag{4.3.4}$$

Thus the numerator in (4.3.3) is $G(i)G(j)$.

Suppose we reverse the levels of F_a. Then the sign of both γ_{abc} and $\omega_{(ab)c}$ will change (see Table 4.3.7). If F_a was originally set to the middle level in the ith run, then $f_{al_i}f_{bl_i} = 0$ will remain unchanged when the levels of F_a are reversed. Let $G(i)'$ be the value of $G(i)$ after the level reversing occurs. Then $G(i)' = -G(i)$ due to the change of sign in γ_{abc} and $\omega_{(ab)c}$. If F_a was not originally set to the middle level in the ith run then $f_{al_i}f_{bl_i} = \pm\sqrt{3/2}f_{bl_i}$, which will change sign when the levels of F_a are reversed, so $G(i)' = -G(i)$. Hence, when the levels of F_a are reversed, $G(i)' = -G(i)$ regardless of what level F_a was originally set to in the ith run. Thus, the numerator becomes $G(i)'G(j)' = (-G(i))(-G(j)) = G(i)G(j)$, and remains unchanged.

Now suppose we reverse the levels of F_c. Then the sign of γ_{abc} will change but $\omega_{(ab)c}$ will remain the same. When level F_c are reversed f_{cq_i} will not change, regardless of which level F_c was originally set to in the ith run, but f_{cl_i} might. If F_c was originally set to the middle level in the ith run then $f_{cl_i} = 0$ will remain unchanged, hence the change in sign of γ_{abc} in $G(i)$ is irrelevant since it is multiplied by zero. Thus, $G(i)' = G(i)$. If F_c was not originally set to the middle level in the ith run then f_{cl_i} will change signs when the levels of F_c are reversed, and hence the new value of $f_{cl_i}\gamma_{abc}$ will become $(-f_{cl_i})(-\gamma_{abc}) = f_{cl_i}\gamma_{abc}$, and so there is no net change in $G(i)$.

Now suppose we reverse the level of F_a and F_c. Then the sign of $\omega_{(ab)c}$ will change but γ_{abc} will remain the same. As mentioned above, f_{cq_i} will not change, regardless what level F_c was originally set to in the ith run. If both F_a and F_c were originally set to the middle level in the ith run then $f_{al_i} = f_{cl_i} = 0$, hence $G(i) = -\left(\frac{3\sqrt{2}}{4} f_{cq_i} \omega_{(ab)c}\right)$. When the levels of each of F_a and F_c are reversed, the change of sign in $\omega_{(ab)c}$ means $G(i)' = -G(i)$.

If F_a was originally set to the middle level in the ith run, but the level of F_c in the ith run was not, then $f_{al_i} = 0$. Thus, $G(i) = -\left(\frac{3}{2}\sqrt{\frac{3}{2}} f_{cl_i} \gamma_{abc} + \frac{3\sqrt{2}}{4} f_{cq_i} \omega_{(ab)c}\right)$. Hence, when the levels of each of F_a and F_c are reversed, the sign of both f_{cl_i} and $\omega_{(ab)c}$ will change while the other variables in $G(i)$ remain constant. Thus, $G(i)' = -G(i)$.

If neither F_a nor F_c were originally set to the middle level in the ith run, then each of the three terms in $G(i)$ will change sign when the levels of each of these two factors are reversed. Hence, $G(i)' = -G(i)$. Therefore, no matter what levels each of F_a and F_c were originally set to in either the ith or jth runs, the numerator will become $G(i)'G(j)' = (-G(i))(-G(j)) = G(i)G(j)$ when the levels of each of these two factors are reversed.

4.4 Empirical Results

As mentioned above, the absence of generic rules has compelled us to carry out an empirical enumeration instead. The goal of performing the empirical enumeration is to make some general comments that assist in the selecting a good design. While the algebraic results derived in Section 4.2 do not lead to general rules, these results are still valuable in greatly reducing the calculation time of the empirical enumeration. Rather than calculating the determinant of the $\upsilon \times \upsilon$ matrix $\mathbf{M}_A$ (or $\mathbf{M}_B$) directly for every combination of OA, set of p (or t) runs, and interaction, the results in Section 4.2 allow us to simply calculate $|\mathbf{M}|$ once for each OA, then we only need to calculate the determinant of the $p \times p$ matrix $\mathbf{\Psi}_A$ (or the $t \times t$ matrix $\mathbf{\Psi}_B$) for each set of p (or t runs) with each interaction. This sizeable increase in the calculation efficiency allows to enumerate many more results then would otherwise be possible.

Using the results in Section 4.2, we have exhaustively calculated $|\mathbf{M}_A|$ and $|\mathbf{M}_B|$ for each of the OA$[18, 3^m]$s enumerated in Chapter 3. Since OA plus p runs designs (or OA minus t runs designs) are appropriate when p (or t) is reasonably small (otherwise it might be more appropriate to construct the design from first principles), we have considered all designs with either 1 or 2 adjoined (or removed)

runs. As discussed in Section 4.2.4, we have restricted the class of OA plus p runs designs to designs in which the adjoined p runs already appear in the OA. Hence, when $p = 1$ (or $t = 1$), there are 18 runs to choose from. When $p = 2$, in addition to the $\binom{18}{2} = 153$ pairs of runs within the OA, we have also considered all 18 repeated pairs, thus we have considered $153 + 18 = 171$ pairs of runs to adjoin to each of the OA$[18, 3^m]$s. Of course, in the context of missing runs, it is not possible to remove the same run twice, hence when $t = 2$ we have considered 153 pairs of runs to remove from each of the OA$[18, 3^m]$s.

In the following sections we will make reference to, and give recommendations for, OAs in specific geometric isomorphism classes. We have shown that all OAs in a given geometric isomorphism class have the same statistical properties, hence when we state the optimal value can be achieved by a particular class we are referring to all OAs within that class. However, we also give specific examples of which interaction is included in the model and which set of runs should be adjoined or removed. These examples assume that the OA in question is the representative enumerated in Chapter 3. If a different OA is used, then the specifics of the interaction and/or adjoined or missing runs may change depending on the permutations required to step from the OA being used to the OA assumed in the example.

There are hundreds of isomorphism classes across all m, hence a thorough discussion of the properties of each is not feasible. However, there are only 13 geometric isomorphism classes when $m = 3$ and only $\binom{3}{2} = 3$ interactions to be considered in each, thus a discussion of the entire design space is feasible. In Section 4.4.1 we examine the properties of all OA$[18, 3^3]$ geometric isomorphism classes and make some general observations about the effect of the entries in $\mathbf{u}_{(ab)}$ on the performance of these designs. In Section 4.4.2 we summarise our findings for the OA$[18, 3^m]$ geometric isomorphism classes when $m > 3$, and make recommendations about the best designs for each m. We finish with a discussion of how this investigation relates to specific optimality criteria from the literature in Section 4.4.3.

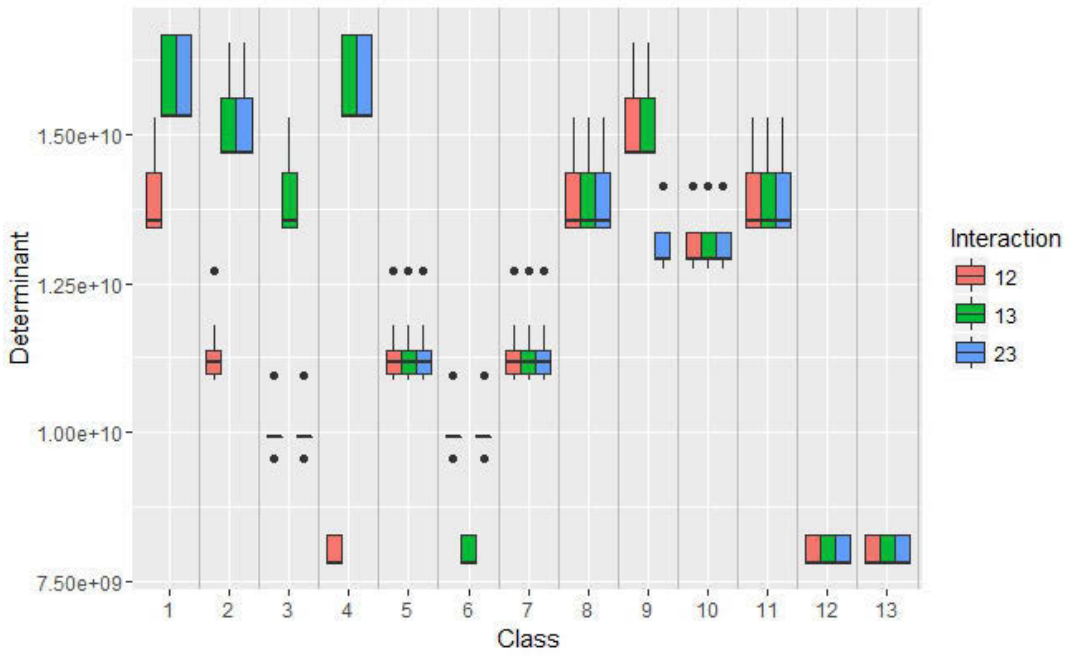
4.4.1 Three Factors

In this and the following sections we will present graphs containing multiple, coloured boxplots for various geometric isomorphism classes. Each boxplot represents a linear-by-linear interaction that is included in the model. In order to distinguish between these interactions, a legend is provided on the right of each graph indicating which colour is associated with which interaction, where the interaction between factors F_x and F_y is denoted by xy.

Adjoining $p = 1$ run

Figure 4.4.1 shows the distribution of $|\mathbf{M}_A|$ across all 18 single runs to be adjoined for each of the 13 OA$[18, 3^3]$ geometric isomorphism classes and for each of the possible linear-by-linear interactions to include in the model.

Figure 4.4.1: $|\mathbf{M}_A|$ for all OA$[18, 3^3]$ plus 1 run designs



The best obtainable $|\mathbf{M}_A|$ is $16,682,995,872$ which is realised in classes 1 and 4 when particular single runs are adjoined and when the interaction is between factors F_1 and F_3 (green), or F_2 and F_3 (blue). We would expect these classes to behave similarly for these two interactions since their values of $|\gamma_{123}|$, $|\omega_{(13)2}|$, and $|\omega_{(23)1}|$ are the same (see Table 4.3.6). All three of these values are equal to zero in both classes, hence the information matrix for each of these un-altered OAs is $\mathbf{M} = 18\mathbf{I}_8$, which gives the maximum possible $|\mathbf{M}|$, that is, 18^8. Since $|\mathbf{M}_A| = |\mathbf{M}||\mathbf{\Psi}_A|$, it is not surprising that the best $|\mathbf{M}_A|$ corresponds to the best $|\mathbf{M}|$. However, recall from Section 4.2.4 that we are often better to have non-zero values in $\mathbf{u}_{(ab)}$ (hence non-zero values of $|\gamma_{abc}|$ and/or $|\omega_{(ab)c}|$, and non-zero off-diagonal entries in $\mathbf{M}$) because the resulting decrease in $|\mathbf{M}|$ is off-set by a greater increase in $|\mathbf{\Psi}_A|$. This is not the case when $p = 1$ because $\mathbf{\Psi}_A$ is only a 1×1 matrix hence its determinant has less influence on $|\mathbf{M}_A|$ then for larger p.

The values of $|\omega_{(12)3}|$ for the first and fourth classes are 6 and 12 respectively. This is reflected in the boxplot for the interaction between F_1 and F_2 (red), which is considerably lower for the fourth class than the first. The maximum determinant in the first class when the interaction is between the F_1 and F_2 is $15,267,237,048$

which is $\left(\frac{15,267,237,048}{16,682,995,872}\right)^{1/8} = 98.8976\%$ efficient, and the maximum determinant in the fourth class when the interaction is between the F_1 and F_2 is $8,264,970,432$ which is 91.5948% efficient.

Notice there are pairs of classes that appear to have the same properties, such as the pairs (5,7), (8,11) and (12,13) which have identical boxplots within each pair. It is not immediately obviously why this is the case as the values of $|\gamma_{abc}|$ and $|\omega_{(ab)c}|$, which are reproduced in Table 4.4.1 below, are not the same within each of these pairs.

Table 4.4.1: Subset of results from Table 4.3.6

Class	$\lvert\gamma_{abc}\rvert$	$\lvert\omega_{(ab)c}\rvert^{\dagger}$	**Class**	$\lvert\gamma_{abc}\rvert$	$\lvert\omega_{(ab)c}\rvert^{\dagger}$	**Class**	$\lvert\gamma_{abc}\rvert$	$\lvert\omega_{(ab)c}\rvert^{\dagger}$
5	4	6	**8**	3	3	**12**	6	6
7	5	3	**11**	0	6	**13**	0	12

†*Note* $|\omega_{(12)3}| = |\omega_{(13)2}| = |\omega_{(23)1}|$ *for each of these classes*

From Theorem 4.2.1 we see that the values of $|\gamma_{abc}|$ and $|\omega_{(ab)c}|$ will effect $\mathbf{a}_{ME_i}\mathbf{u}_{(ab)}$ in the numerator and $\mathbf{u}'_{(ab)}\mathbf{u}_{(ab)}$ in the denominator of the entry(s) in $\mathbf{\Psi}$. In the context of three-factor designs $\mathbf{u}'_{(ab)}\mathbf{u}_{(ab)} = u^2_{(ab)cl} + u^2_{(ab)cq} = \left(\frac{3}{2}\sqrt{\frac{3}{2}}|\gamma_{abc}|\right)^2 + \left(\frac{3\sqrt{2}}{4}|\omega_{(ab)c}|\right)^2 = \frac{9}{8}\left(3\gamma^2_{abc} + \omega^2_{(ab)c}\right)$. Hence, $\mathbf{u}'_{(ab)}\mathbf{u}_{(ab)}$ is the same within each pair of classes ($\mathbf{u}'_{(ab)}\mathbf{u}_{(ab)} = 94.5$, 40.5 and 162 for the pairs of classes (5,7), (8,11) and (12,13) respectively). Now consider $\mathbf{a}_{ME_i}\mathbf{u}_{(ab)}$ in the numerator. For three-factor designs, $\mathbf{a}_{ME_i}\mathbf{u}_{(ab)} = f_{cl_i}u_{(ab)cl} + f_{cq_i}u_{(ab)cq} = \frac{3}{2}\sqrt{\frac{3}{2}}\gamma_{abc}f_{cl_i} + \frac{3\sqrt{2}}{4}\omega_{(ab)c}f_{cq_i}$. This becomes $\frac{3}{4}\left(-3\gamma_{abc} + \omega_{(ab)c}\right)$ when F_c is set to the low level, $-\frac{3\omega_{(ab)c}}{2}$ when F_c is set to the middle level, or $\frac{3}{4}\left(3\gamma_{abc} + \omega_{(ab)c}\right)$ when F_c is set to the high level. The table below gives the value for each of these expressions in Table 4.4.1.

Table 4.4.2: $|\mathbf{a}_{ME_i}\mathbf{u}_{(ab)}|$ for each of the classes in Table 4.4.1

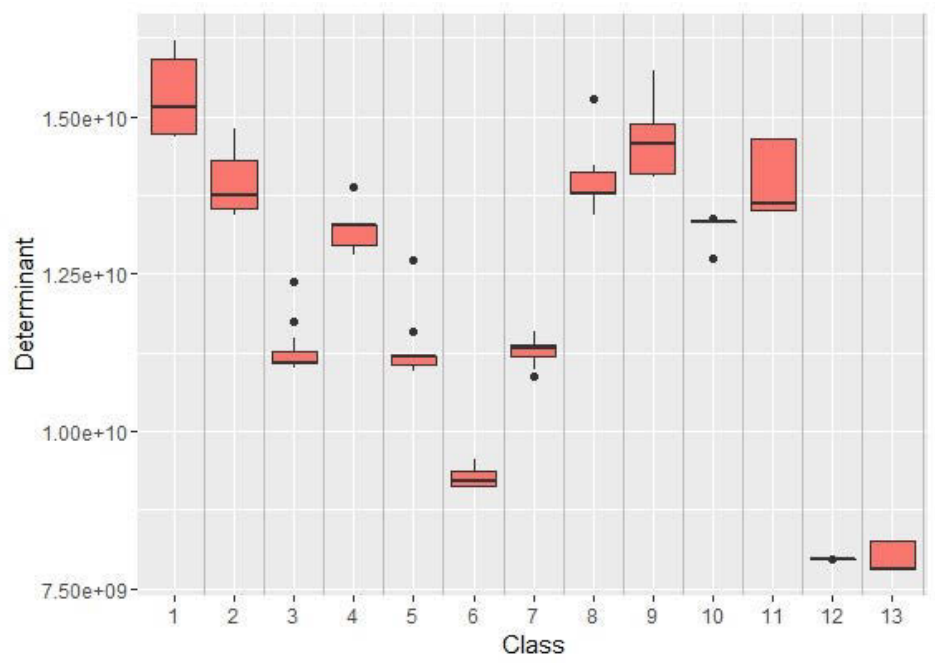
	F_c				F_c				F_c		
Class	**0**	**1**	**2**	**Class**	**0**	**1**	**2**	**Class**	**0**	**1**	**2**
5	$\frac{9}{2}$	9	$\frac{27}{2}$	**8**	$\frac{9}{2}$	$\frac{9}{2}$	9	**12**	9	9	18
7	9	$\frac{9}{2}$	$\frac{27}{2}$	**11**	$\frac{9}{2}$	9	$\frac{9}{2}$	**13**	9	18	9

The three possible values of $|\mathbf{a}_{ME_i}\mathbf{u}_{(ab)}|$ are the same within each pair, albeit in different orders. Hence, the classes within each pair may have difference values of $|\mathbf{a}_{ME_i}\mathbf{u}_{(ab)}|$ for a given i (that is, for a specific run), but the tally of all values will be the same across all adjoined runs since F_c will be set to each of the three levels exactly 6 times in the 18 runs. Hence, the boxplot across all 18 runs will be the same.

Of course, if we know which interaction is to be included in the model, then the

distribution of $|\mathbf{M}_A|$ across all of the class may not be of interest, rather we would simply like to know: what is the best OA$[18, 3^3]$ plus 1 run design to use? In this case, there are multiple optimal settings to choose from, one example being: start with an OA$[18, 3^3]$ from the first class, adjoin the 3rd run (010), and assign the two factors that we know will be involved in the interaction to the first and third columns. However, as discussed in Chapter 1, if we do not know which interaction will be active in the model we would like to find a model-robust deign that performs well regardless of which interaction is active. Following Jones et al. (2007), we will evaluate model-robustness by considering the average of the determinant across all models. Figure 4.4.2 illustrates this.

Figure 4.4.2: Average $|\mathbf{M}_A|$ for all OA$[18, 3^3]$ plus 1 run designs

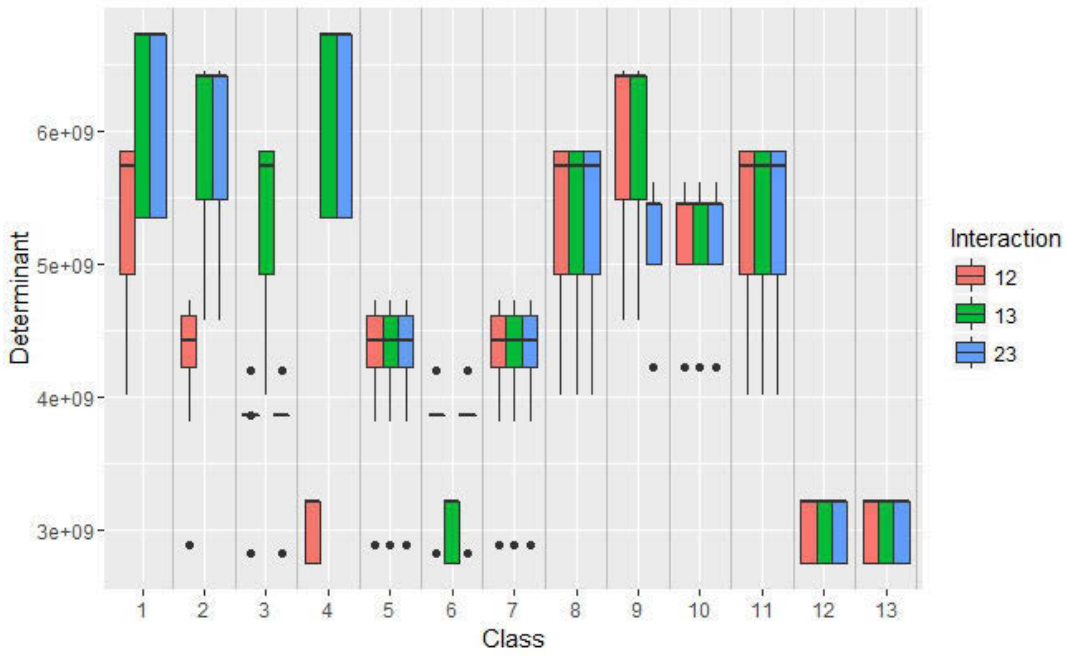


Notice the difference between classes 1 and 4 now. These classes perform equally well when the pair of factors involved in the interaction is (F_1, F_3) or (F_2, F_3), and both can achieve the empirically optimal $|\mathbf{M}_A|$ for each of these interactions. However, when the determinant is averaged over all three interactions, the poor performance of class 4 in the interaction between factors F_1 and F_2 makes this class much less desirable from a model-robust perspective. But again, we have only provided this chart to illustrate the model-robustness properties of each class − from a practical perspective, the distributions themselves may not be of interest, rather we just want to know which design/run combination will give the largest possible average $|\mathbf{M}_A|$. In that case, the answer is the first class with either the 17th or 18th run adjoined (220 or 222), which will give an average $|\mathbf{M}_A|$ of 16,211,076,264.

Missing $t = 1$ run

Figure 4.4.3 shows the distribution of $|\mathbf{M}_B|$ across all 18 single run deletions for each of the 13 OA$[18, 3^3]$ geometric isomorphism classes and for each of the possible linear-by-linear interactions to include in the model.

Figure 4.4.3: $|\mathbf{M}_B|$ for all OA$[18, 3^3]$ minus 1 run designs

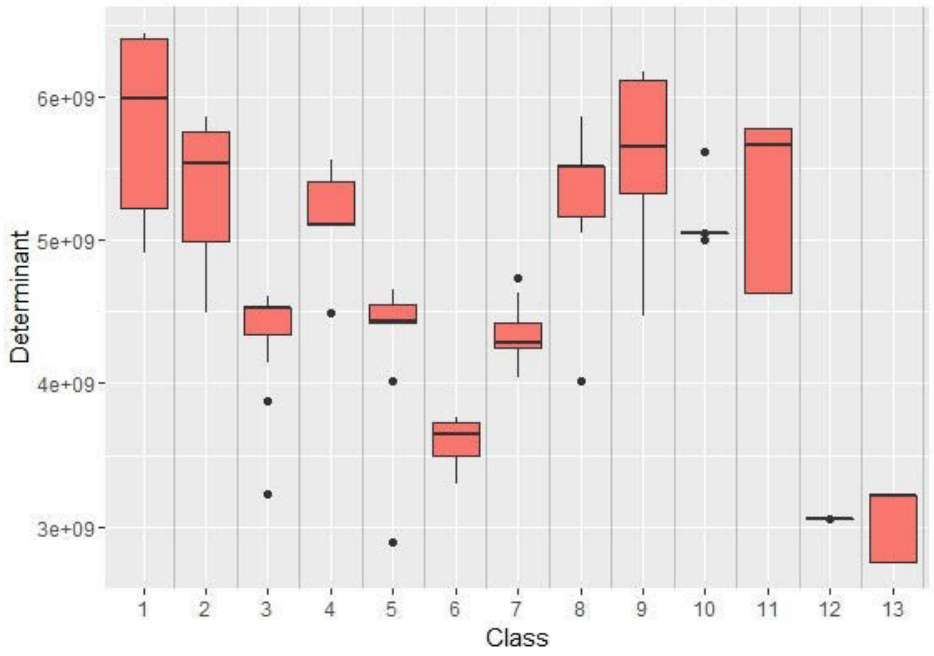


The similarities we observed when adjoining 1 run between particular pairs of class can be seen here in the context of removing 1 run too. For example, the green and blue boxplots are the same between classes 1 and 4, and all three boxplots are the same between the pairs of classes (5,7), (8,11) and (12,13). Again, the best determinant, which in this case is $6,734,420,352$, can be realised in a range of runs in both class 1 and 4 when the interaction is between either F_1 and F_3, or F_2 and F_3. If we know which interaction is to be included in the model, an example of an optimal experimental setting is to use an OA$[18, 3^3]$ from the first class, remove the first run (001), and assign the two factors that are involved in the interaction to either the first and third, or second and third columns.

If we do not know which interaction will be included in the model then we will choose the design based on the average across all three models, as illustrated in Figure 4.4.4. As was the case when adjoining a single run, the first class appears to perform the best overall, with the best average $|\mathbf{M}_B|$ equal to 6,441,064,920 when the 9th or 10th runs are removed (110 or 112).

There is another perspective to consider when removing runs that was not an issue in the context of adjoining runs: what if the run is missing at random? That is,

Figure 4.4.4: Average $|\mathbf{M}_B|$ for all OA$[18, 3^3]$ minus 1 run designs



what if we do not have control over which run will be removed, rather we suspect a random run might fail or be rendered invalid? In this case, we need to consider the whole distribution of each interaction (or the average of the interactions if we do not know which one will be active) rather than focusing on the maximum achievable. Now the preceding charts are valuable tools for assessing the range of D-optimality values or model-robustness within each class.

Finally, to wrap up the investigation into adjoining or removing a single run when $m = 3$, it is worth commenting on the 12th and 13th classes, which stand out in each of the preceding figures as consistently under-performing. The drawback of this method of data-visualisation is that it does not capture the degrees of freedom for pure error, that is, the number of repeated runs in the altered design. The OA$[18, 3^3]$s in classes 12 and 13 consist of 9 pairs of repeated runs, hence while they appear to under-perform when our goal is solely to maximise $|\mathbf{M}_A|$ or $|\mathbf{M}_B|$, their properties would be more desirable if we were were to balance this goal against a desire to maximise the degrees of freedom for pure error.

Adjoining $p = 2$ runs

Figure 4.4.5 shows the distribution of $|\mathbf{M}_A|$ across all 171 pairs of runs ($\binom{18}{2} = 153$ pairs of runs from within the OA plus 18 pairs of repeated runs) which might be adjoined to each of the 13 OA$[18, 3^3]$ geometric isomorphism classes, and for each of the possible linear-by-linear interactions to include in the model.

While similar patterns can be seen between classes that were seen in the context of adjoining a single run, it is no longer the case that the best determinant is realised in classes 1 and 4. The largest determinant, which is $25,755,756,624$, only appears four times across all $13 \times 171 \times 3 = 6,669$ calculated values (13 designs each with 171 pairs of adjoined runs for each of 3 interactions). In class 2, the optimal determinant is achieved when the interaction is between F_1 and F_3 and the pair of runs (012, 020) are adjoined, or when the interaction is between F_2 and F_3 and the pair of runs (102, 200) are adjoined. In class 9 the optimal determinant is achieved when the pair of runs (200, 222) are adjoined and the interaction is between either F_1 and F_2, or F_1 and F_3. The best realisable determinant in both class 1 and class 4 is $25,254,076,320$ which is $\left(\frac{25,254,076,320}{25,755,756,624}\right)^{1/8} = 99.7544\%$ efficient.

Regarding model-robustness, Figure 4.4.6 illustrates the average $|\mathbf{M}_A|$ for all OA$[18, 3^3]$ plus 2 runs designs. The best average determinant is 24,290,396,640 which is achieved in class 9 when the 13th and 18th runs (200 and 222) are removed.

Missing $t = 2$ runs

Figure 4.4.7 shows the distribution of $|\mathbf{M}_B|$ across all 153 pairs of runs to be removed from each of the 13 OA$[18, 3^3]$ geometric isomorphism classes and for each of the possible linear-by-linear interactions to include in the model.

The optimal determinant, which is $4,081,466,880$ is realised in Classes 1 and 4. One example of an optimal design setting is to use an OA from the first class, remove the runs 001 and 121, and to assign the factors involved in the interaction to the first and third, or second and third columns.

An important feature of Figure 4.4.7 that we have not yet seen is the occurrence of zero determinants. So, far, if the active interaction or the pair of missing runs is not known then we have advocated considering the average determinant across all interactions. However, as we can see in Figure 4.4.8, this does not provide us with information about whether all models are estimable for all pairs of missing runs. For example, based on Figure 4.4.8 alone, we might consider class 1 to be favourable. However, from Figure 4.4.7, we can see that when the interaction is between F_1 and F_2, there is at least one pair of runs in class 1 that results in a

Figure 4.4.5: $|\mathbf{M}_A|$ for all OA$[18, 3^3]$ plus 2 runs designs

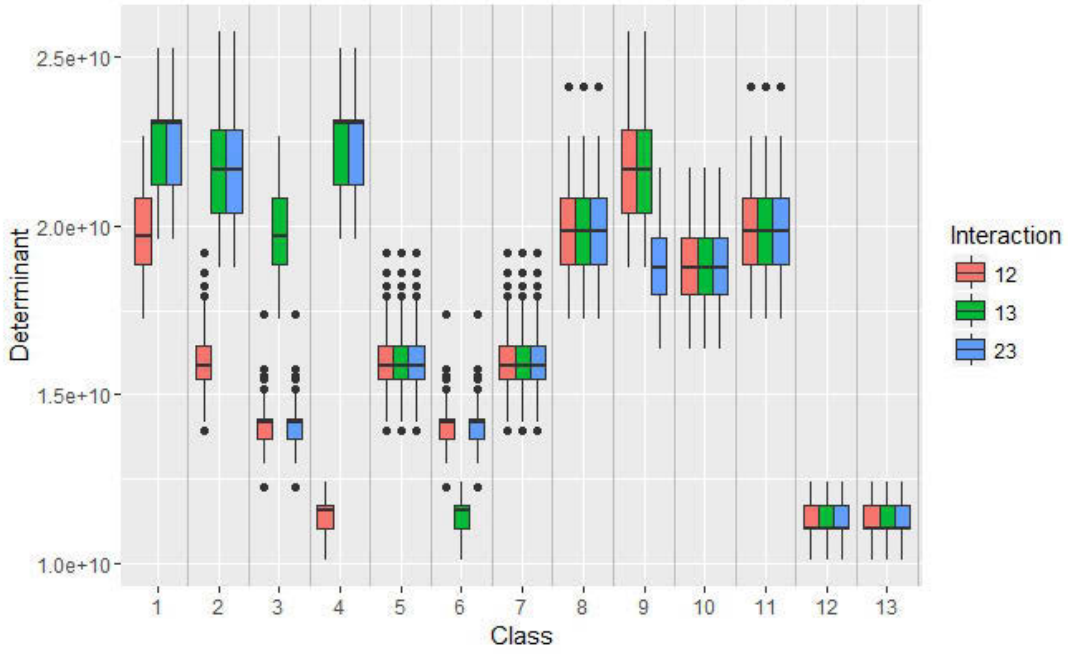


Figure 4.4.6: Average $|\mathbf{M}_A|$ for all OA$[18, 3^3]$ plus 2 runs designs

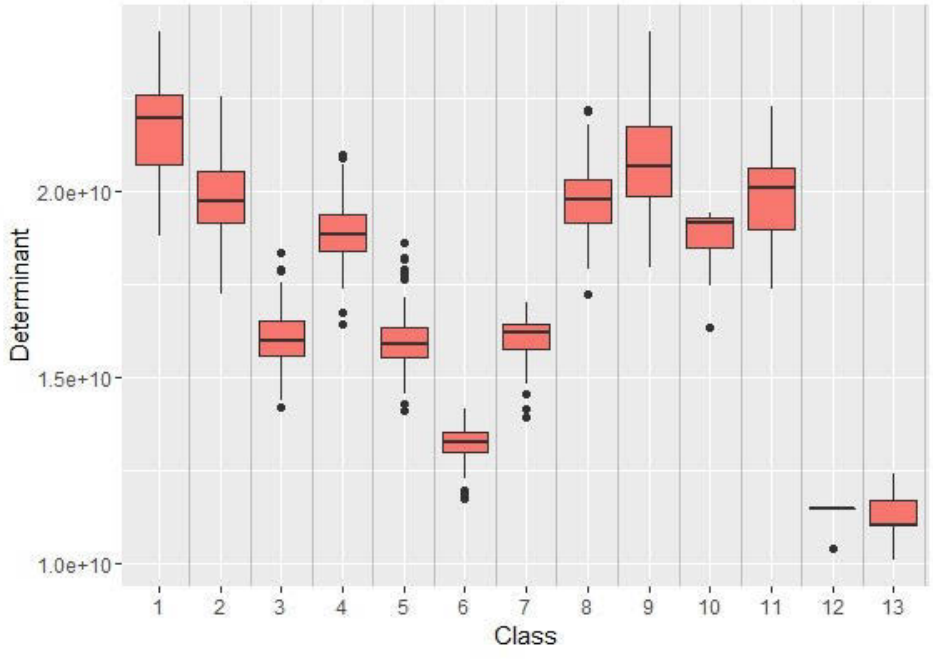


Figure 4.4.7: $|\mathbf{M}_B|$ for all OA$[18, 3^3]$ minus 2 runs designs

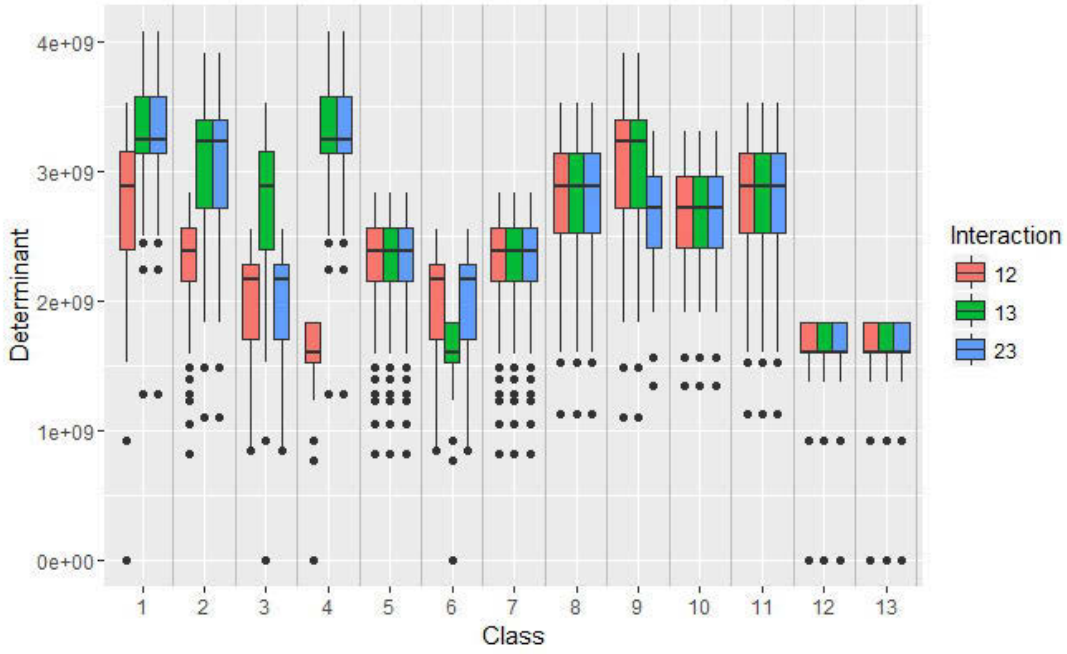
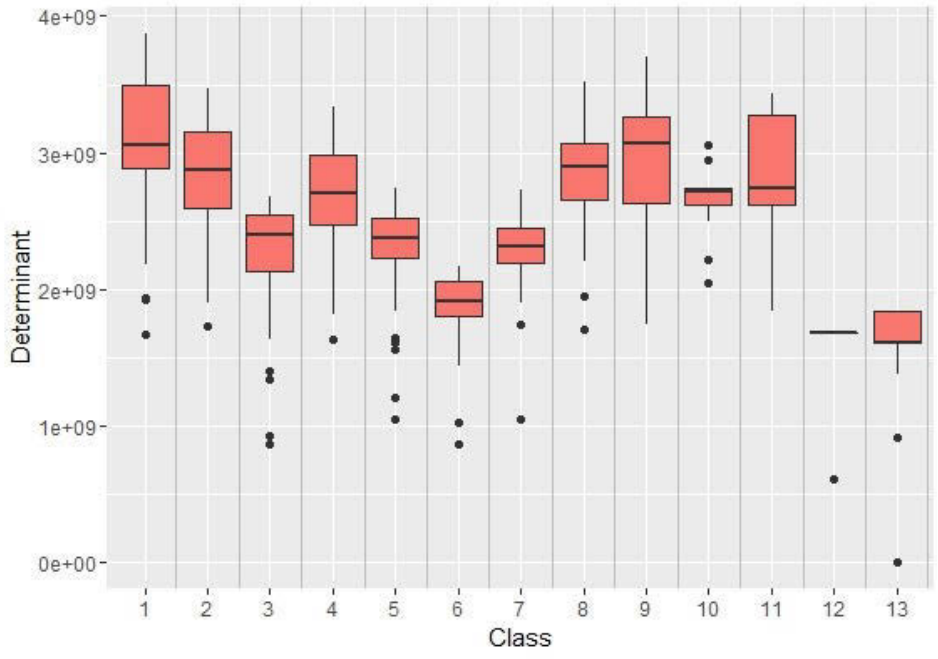


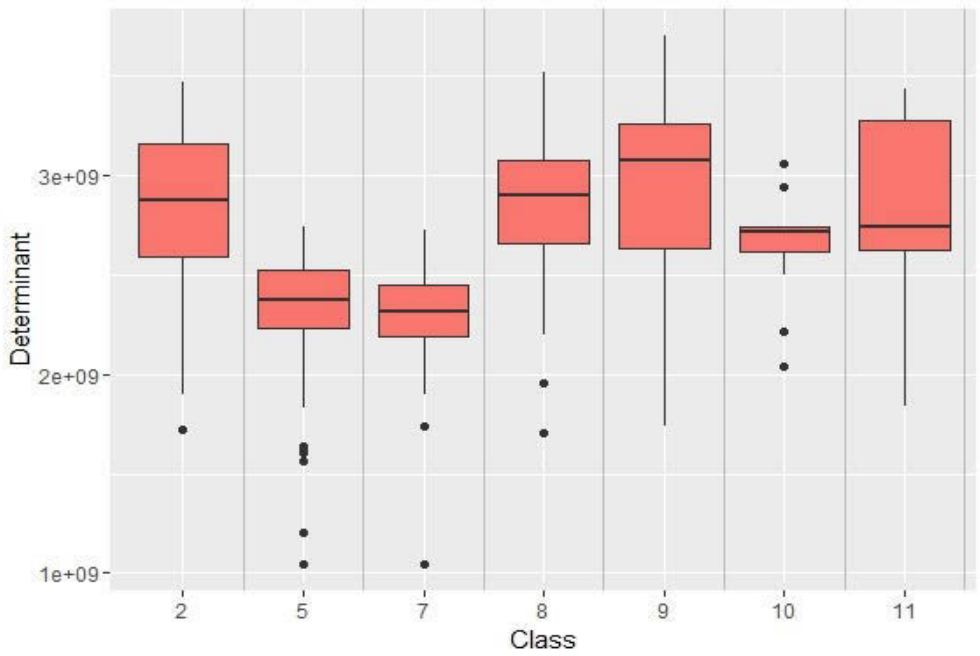
Figure 4.4.8: Average $|\mathbf{M}_B|$ for all OA$[18, 3^3]$ minus 2 runs designs



non-estimable model when it is removed. If we have control over which runs to remove from the OA then we can ensure the pair we chose will not result in a zero determinant. However, if we are uncertain which pair of runs will be missing, then we should avoid class 1 altogether to ensure that the final design and model is estimable.

Figure 4.4.9 is a subset of the boxplots in Figure 4.4.8, where all those designs that contain some run/interaction combinations that result in a zero determinant have been omitted. Based on this figure, we would choose class 9 under run and model uncertainty, because although we can not realise the optimal determinant, we can at least be confident that the final design and model will be estimable.

Figure 4.4.9: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^3]$ minus 2 runs designs



†*Any classes that result in a zero determinant for any combination of pair of missing runs/interaction has been omitted.*

4.4.2 More Than Three Factors

In this section we will summarise the results for all OA$[18, 3^m]$ designs when $m > 3$. Tables of optimal $|\mathbf{M}_A|$ and $|\mathbf{M}_B|$ are provided that include examples of optimal design settings for situations when the experimenter has prior information about which interaction will be active in the model. For situations when the experimenter is uncertain which interaction will be active in the model, tables of the best determinants averaged across all interactions are provided. In the case of missing runs, when the experimenter does not have control over which runs are to be removed, plots of the distribution of both the optimal value and the best

average value are presented for a subset of designs.

Adjoining $p = 1$ Run

Table 4.4.3 below gives the optimal $|\mathbf{M}_A|$ for all OA$[18, 3^m]$ plus 1 run designs when $m > 3$.

Table 4.4.3: Optimal $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 1 run designs

m	$\lvert\mathbf{M}_A\rvert$	No.†	Example		
			Class	Interaction	Run
4	5.8020×10^{12}	128	3	(F_2, F_3)	0202
5	1.9053×10^{15}	72	38	(F_1, F_3)	01022
6	5.5788×10^{17}	114	11	(F_3, F_6)	012202
7	1.5517×10^{20}	1260	14	(F_3, F_7)	0100121

†*Number of occurrences of optimal* $|\mathbf{M}_A|$ *across all classes and all* $18 \times \binom{m}{2}$ *run/interaction combinations*

This information will only be useful if the experimenter has prior information about which interaction will be active in the model. For example, suppose we seek to construct an OA$[18, 3^7]$ plus 1 run design and we are confident that the interaction will be between factors F_a and F_b. Then, following Table 4.4.3, we will use an OA$[18, 3^7]$ from the 14th geometric isomorphism class, adjoin the run 0100121, and ensure factors F_a and F_b are assigned to the third and seventh columns. However, as discussed above, if we are unsure which interaction to include in the model, we would prefer a model-robust design. As we saw in the context of three-factor OAs, care must be taken to filter out classes that can result in zero determinants before we average $|\mathbf{M}_A|$ across all interactions in order to assess model-robustness. Hence, we will introduce a two-stage method for determining the best model-robust design based on the number of estimable models. Our first priority is to minimise the number of zero determinants. Thus, if there are any combinations of class and runs in which $|\mathbf{M}_A|$ is non-zero for all interactions, then we will consider only these class/run combinations. If no class/run combination exists such that all $|\mathbf{M}_A|$ are non-zero for all interactions then we will determine the smallest number of zero determinants per class/run combination and consider the average $|\mathbf{M}_A|$ of only the subset of class/run combinations that realise this minimum. For example, when $m = 7$, none of the 291 classes achieve a non-zero $|\mathbf{M}_A|$ for all $\binom{7}{2} = 21$ interactions, regardless of which run is adjoined. Hence, we search for the minimum number of zero determinants across all class/runs combinations, which turns out to be 4. Thus, we calculate the average $|\mathbf{M}_A|$ of only the class/run combinations that have exactly 4 zero determinants.

Table 4.4.4 below gives the best $|\mathbf{M}_A|$ averaged across all interaction for all OA$[18, 3^m]$

plus 1 run designs that realise the specified number of zero determinants. Hence, when $m = 7$ and we are uncertain which interaction to include in the model, following Table 4.4.4 we would use an OA$[18, 3^7]$ from the 62nd geometric isomorphism class and adjoin the run 0100101. This would not guarantee an estimable model as there is a 4 in 21 chance that the active interaction will result in a zero determinant, which is the smallest possible chance of a non-estimable model that we can obtain. We note that this two-stage process may result in a smaller average $|\mathbf{M}_A|$ then would otherwise be possible if we were to ignore the number of non-estimable models. For example, if we use an OA$[18, 3^7]$ from the 230th geometric isomorphism class and adjoin the run 0220200 then the average $|\mathbf{M}_A|$ is 7.1322×10^{19} which is larger than that reported in Table 4.4.4. However, this class/run combination includes 5 non-estimable models, and we are electing to prioritise model-estimablity over average D-optimality.

Table 4.4.4: Best average $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 1 run designs

m	No. Zeros†	Avg.$\lvert\mathbf{M}_A\rvert$	No.‡	Example	
				Class	Run
4	0	4.8391×10^{12}	2	12	2200
5	0	1.3091×10^{15}	2	38	01022
6	0	3.5781×10^{17}	6	469	001220
7	4	6.7587×10^{19}	6	62	0100101

†*Number of zero determinants amongst all* $\binom{m}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_A|$ *across all classes with given number of zeros and all 18 runs*

We assume the experimenter has control over which OA to select and which runs to adjoin, hence the information contained in these two tables are all that is required to construct either a D-optimal design (when the experimenter has prior information about which interaction will be active in the model) or a model-robust design with the best average determinant (under model uncertainty). However, as noted in Section 4.4.1, we may require further information when runs are missing rather than adjoined. This is illustrated in the next section.

Missing $t = 1$ Run

Table 4.4.5 below gives the optimal $|\mathbf{M}_B|$ for all OA$[18, 3^m]$ minus 1 run designs when $m > 3$.

As noted above, if we have control over which OA to use and which runs to remove, then the information contained in this table is sufficient to construct a D-optimal design. However, these tables do not provide information about the range of values across all missing runs within each class. If we do not have

Table 4.4.5: Optimal $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 1 run designs

m	$\lvert\mathbf{M}_B\rvert$	No.†	Example		
			Class	**Interaction**	**Run**
4	1.7852×10^{12}	160	3	(F_2, F_3)	0010
5	4.1240×10^{14}	108	38	(F_1, F_3)	00101
6	8.1123×10^{16}	228	11	(F_3, F_6)	001112
7	1.3493×10^{19}	630	14	(F_3, F_7)	0011112

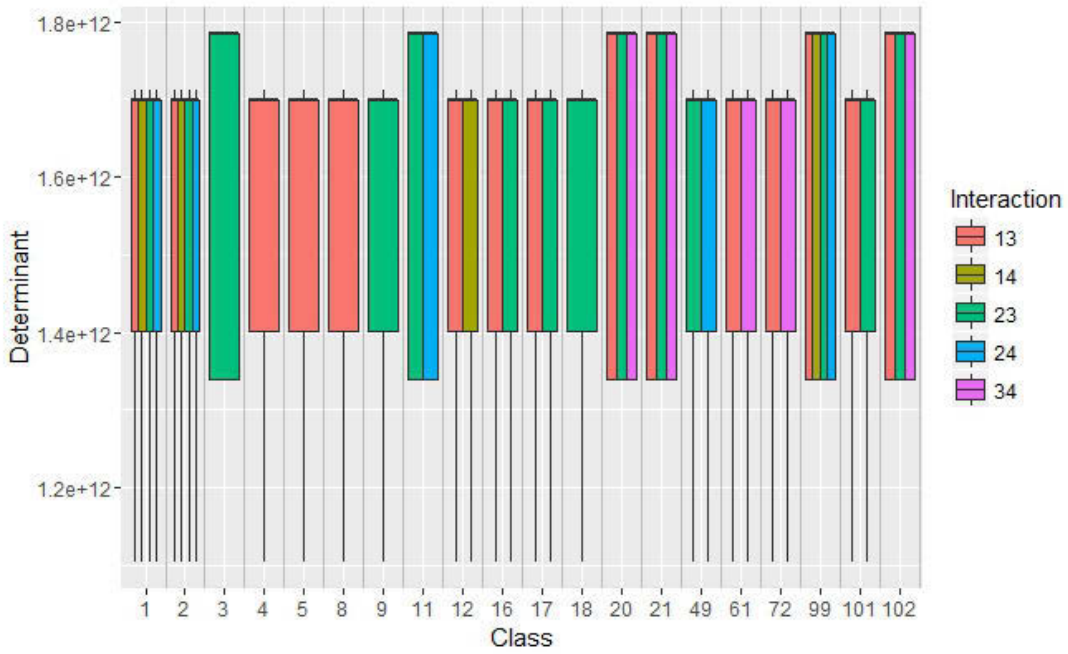
†*Number of occurrences of optimal* $|\mathbf{M}_B|$ *across all classes and all* $18 \times \binom{m}{2}$ *run/interaction combinations*

control over which runs will be missing from the OA, then we would like some further information on the distribution of values across each class, such as the data-visualisation methods used in Section 4.4.1. As the number of classes increases it becomes cumbersome to use data-visualisation methods for the entire design space, hence it is tempting to use some kind of filter to plot only those classes that seem 'promising' in some sense. In the context of D-optimality (that is, we know which interaction will be active in the model, and so we are not concerned about model-robustness) we are interested in class/interaction combinations that produce reasonably large $|\mathbf{M}_B|$ for all adjoined runs. To this end, we will plot only those class/interaction combinations for which the average $|\mathbf{M}_B|$ is in the top 5 percent. Note the distinction between the average in this context (that is, the average $|\mathbf{M}_B|$ across all runs within a given class/interaction) and the average we consider in the context of model-robustness (the is, the average $|\mathbf{M}_B|$ across all interactions within a given class/adjoined run).

When $m = 4$ there are 137 OA$[18, 3^4]$s to choose from and $\binom{4}{2} = 6$ potential interactions that could be included in the model, hence there are $137 \times 6 = 822$ distributions of $|\mathbf{M}_B|$ to consider. We calculated the average $|\mathbf{M}_B|$ for each of these class/interaction combinations and selected those in the top 5 percent. The distributions of these class/interaction combinations are presented in Figure 4.4.10. Note that although there are 6 interactions that could be included in the model, only 5 are presented in this chart as the average $|\mathbf{M}_B|$ for the interaction between F_1 and F_2 does not make it into the top 5 percent for any of the OAs.

As it happens, the optimal design given in Table 4.4.5 remains one of the best designs to use even when we don't know which run will be missing from the OA. That is, if we use an OA from the 3rd geometric class and assign the factors involved in the interaction to the second and third columns, then there is a chance that resulting OA$[18, 3^4]$ minus 1 run design we will obtain the optimal $|\mathbf{M}_B|$ for a run that is missing at random, but if not, we will at least achieve 97.1642% efficiency. As we can see from Figure 4.4.10, there are 16 other class/interaction combinations that can also achieve this outcome. Note that if we were to rely on Table 4.4.5

Figure 4.4.10: $|\mathbf{M}_B|$ for subset† of OA$[18, 3^4]$ minus 1 run designs

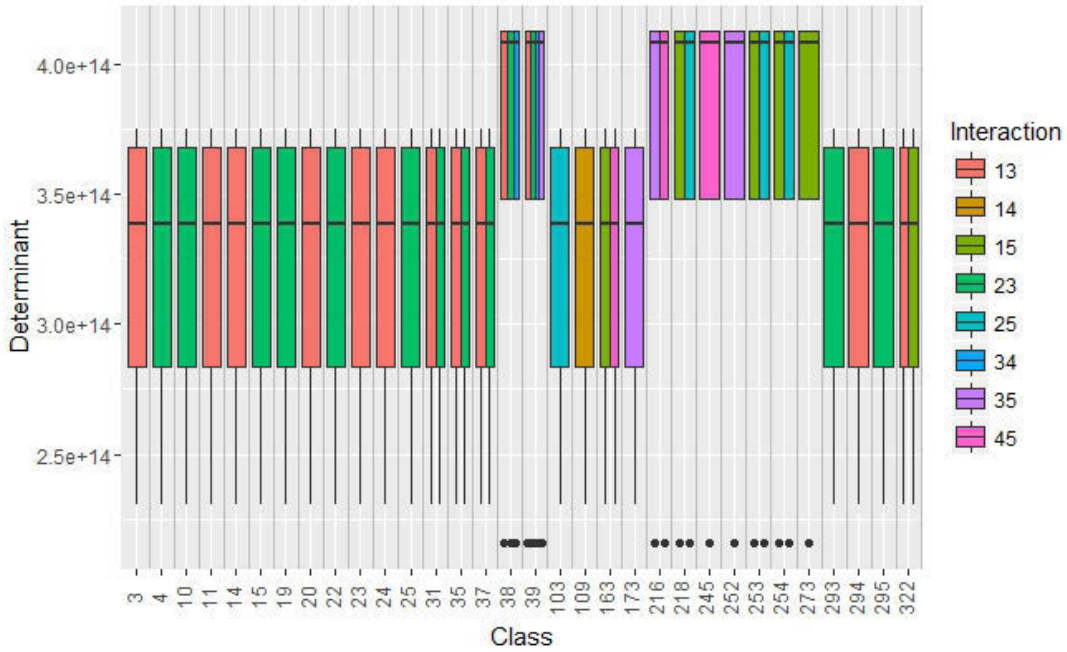


†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_A|$ *across all adjoined runs in the top 5 percent of all class/interaction combinations.*

alone, without plotting the distribution across all potential missing runs, then we would essentially be making the assumption that any class/interaction combination that achieves the optimal determinant for a particular missing run will also perform well when the missing run is different. Although this assumption turns out to be warranted in this case, a more prudent approach is to check the distribution for the given class/interaction to ensure that the design performs well regardless of which run is missing.

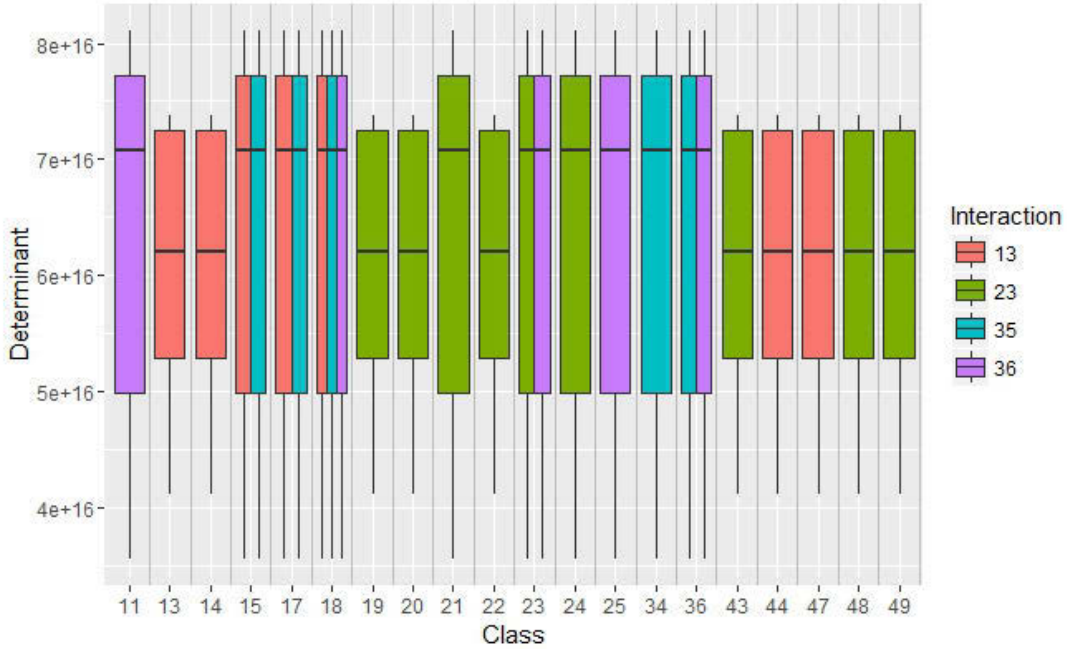
The analogous plots for $m = 5, 6$ and 7 are given in Figures 4.4.11, 4.4.12 and 4.4.13 respectively. For each of these design spaces we have restricted the filter to the top 1 percent rather than the top 5 percent because these spaces are so large that there are too many class/interaction combinations in the top 5 percent to consider. Furthermore, when $m = 6$ and 7, even when we restrict our attention to only the class/interactions in the top 1 percent, the number still is so large that we have reduced the chart to only the first 20 OAs. To ensure that we have not inadvertently filtered out note-worthy designs, we have checked that the distributions of each of the omitted class/interaction combinations are the same as one of those that has been plotted.

Figure 4.4.11: $|\mathbf{M}_B|$ for subset† of $\mathrm{OA}[18, 3^5]$ minus 1 run designs



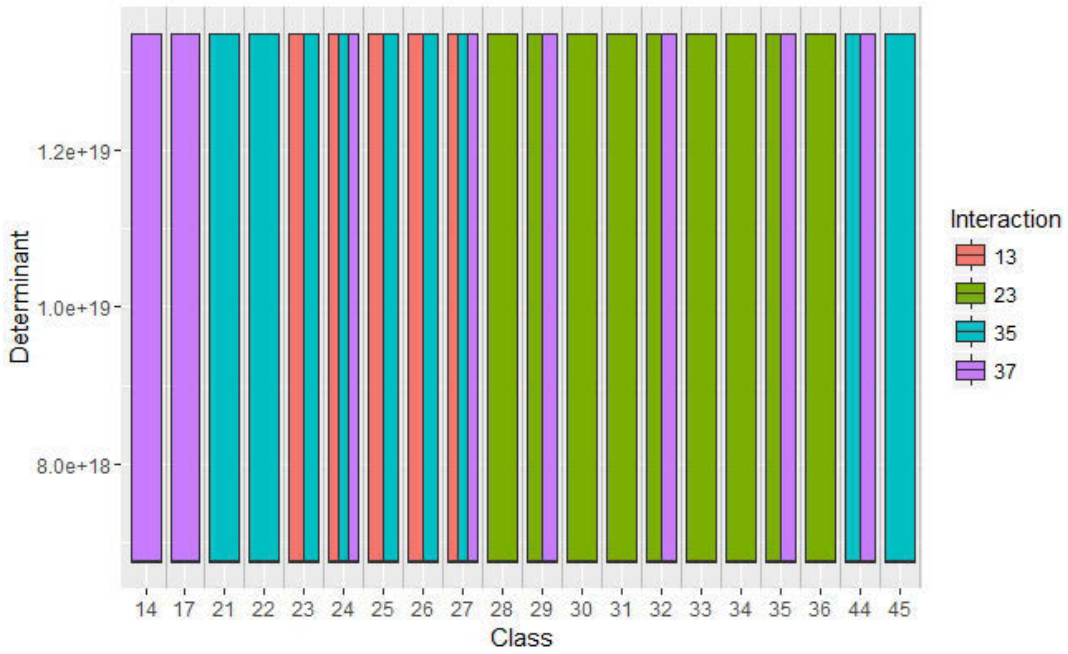
†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 1 percent of all class/interaction combinations.*

Figure 4.4.12: $|\mathbf{M}_B|$ for subset† of $\mathrm{OA}[18, 3^6]$ minus 1 run designs



†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 1 percent of all class/interaction combinations. There are 101 class/interaction combinations in the top 1 percent, hence we have only presented the first 20 classes.*

Figure 4.4.13: $|\mathbf{M}_B|$ for subset† of OA$[18, 3^7]$ minus 1 run designs



†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 1 percent of all class/interaction combinations. There are 78 class/interaction combinations in the top 1 percent, hence we have only presented the first 20 classes.*

We note that all graphs show only one or two distinct distributions to choose from. When $m = 4$ the maximum and minimum for one of the distributions is larger than the respective maximum and minimum of the other distribution, hence the choice is reasonably easy. When $m = 5$, however, the choice may not be as clear. One of the distributions has a majority of its values above the median of the other distribution, however it also has an outlier well below the minimum value of the other distribution. Is the good performance in a majority of the designs worth the risk of the outlier? This question can only be answered in the context of the practical logistics of the experiment. For example, if we suspect for some reason that some runs towards the end of the experiment might fail, but we are reasonably confident that the first 'few' runs will be valid, then we could use a class/interaction combination that has mostly large values of $|\mathbf{M}_B|$ with some small outliers, we just need to ensure that the run which produces a small determinant is run early in experiment so we can be confident it will not fail. If however, the missing run is genuinely missing at random, then perhaps we would prefer a 'safer' class/interaction so that we can control the minimum achievable $|\mathbf{M}_B|$.

Now suppose that we are uncertain which interaction should be included in the model. Then we are interested in a model-robust design. Table 4.4.6 below gives the best $|\mathbf{M}_B|$, averaged across all interactions, for OA$[18, 3^m]$ minus 1 run designs when $m > 3$. Note that there are no class/run combinations that realise non-zero

$|\mathbf{M}_B|$ across all interactions, yet previously, in Figure 4.4.13, we saw that there are multiple class/interaction combinations that realise non-zero $|\mathbf{M}_B|$ across all runs. This highlights the importance of being clear about whether our uncertainty surrounds which interaction should be included in the model, which run should be missing from the OA, or both. Figure 4.4.13 was addressing run uncertainty, hence we looked at specific class/interaction combinations that were able to estimate the model regardless of which run was missing. Table 4.4.6 is addressing model uncertainty, hence we are interested in those class/run combinations that can estimate as many models as possible regardless of which interaction is included.

Table 4.4.6: Best average $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 1 run designs

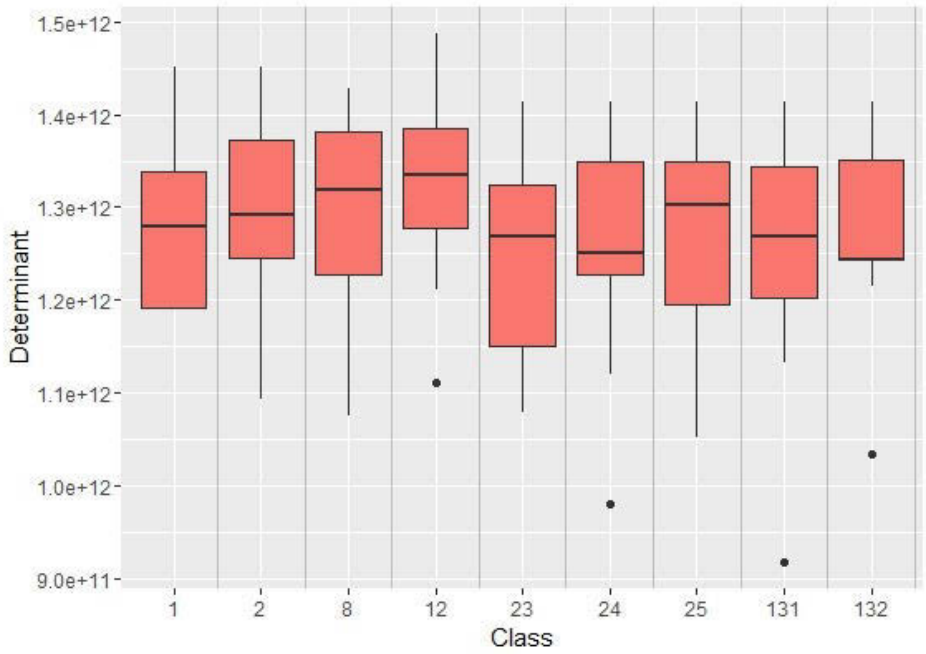
m	No. Zeros†	Avg.$\lvert\mathbf{M}_B\rvert$	No.‡	Example	
				Class	Run
4	0	1.4877×10^{12}	2	12	2111
5	0	2.9081×10^{14}	1	38	11120
6	0	4.7199×10^{16}	6	469	000112
7	4	4.2769×10^{18}	36	14	0011112

†*Number of zero determinants amongst all* $\binom{m}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_B|$ *across all classes with given number of zeros and all 18 runs*

Again, if we have control over which run to remove from the OA then this table provides sufficient information for constructing a model-robust design. However if we are not sure which run will be missing then we need to consider the distribution of the average $|\mathbf{M}_B|$ across all missing runs for each OA. Figures 4.4.10, 4.4.11 and 4.4.12 present the distribution of the average $|\mathbf{M}_B|$ for each of $m = 4, 5$ and 6 respectively. When $m = 7$, we have illustrated the effect of the two-stage process described earlier for minimising the number of zero determinants before calculating the average $|\mathbf{M}_B|$ in Figures 4.4.17 and 4.4.18. Figure 4.4.17 gives the distributions in the top 5 percent of average $|\mathbf{M}_B|$ with no concern for the number zero determinants in each class/run combination. Figure 4.4.18 represents the two-stage process where we first filter out class/run combinations that have more than the minimum number of zeros. Note that all classes have the same number of zero determinants across all interactions for each missing run. So, although we introduced the concept of the two-stage process in the context of eliminating class/run combinations that contain more the the minimum number of zeros, this is effectively the same as eliminating all classes in which all runs have more than the minimum number of zeros. For example, class 35, which appears in Figure 4.4.17, has 5 zero determinants (the minimum is 4) in every class/run combination. Thus, we eliminate the entire class (not specific class/run combinations) when compiling Figure 4.4.18. We note that class 14, which is featured in in Table 4.4.6 as an example of a class/run combination that can achieve the best average $|\mathbf{M}_B|$,

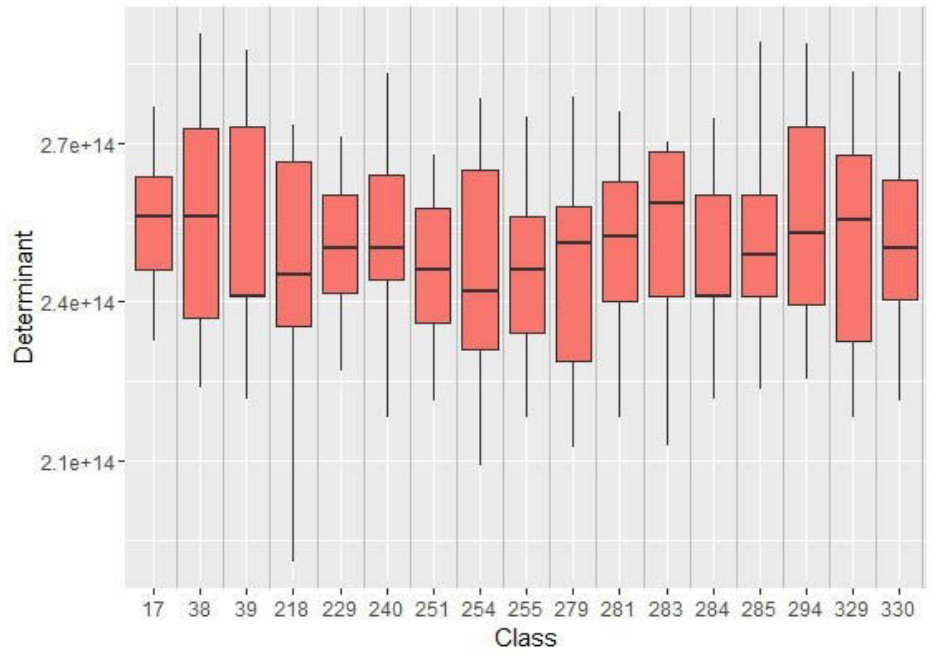
appears in the Figure 4.4.17 but not Figure 4.4.18. This class has not been filtered out due its number of zero determinants, rather we have limited the top 5 percent calculation in the second chart to only those that made it through the first stage of filtering (that is, only those classes with 4 zeros in each missing run), hence, although class 14 can achieve the best average $|\mathbf{M}_B|$ for at least one of the missing runs, is does not fare as well when we consider its distributions over all missing runs. Notice that classes 46, 53, 62 and 68 can achieve the maximum average $|\mathbf{M}_B|$ in Figure 4.4.18 too. Figure 4.4.17 suggests class 230 can do even better, but it has 5 zero determinants across all interactions for each missing run.

Figure 4.4.14: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^4]$ minus 1 run designs



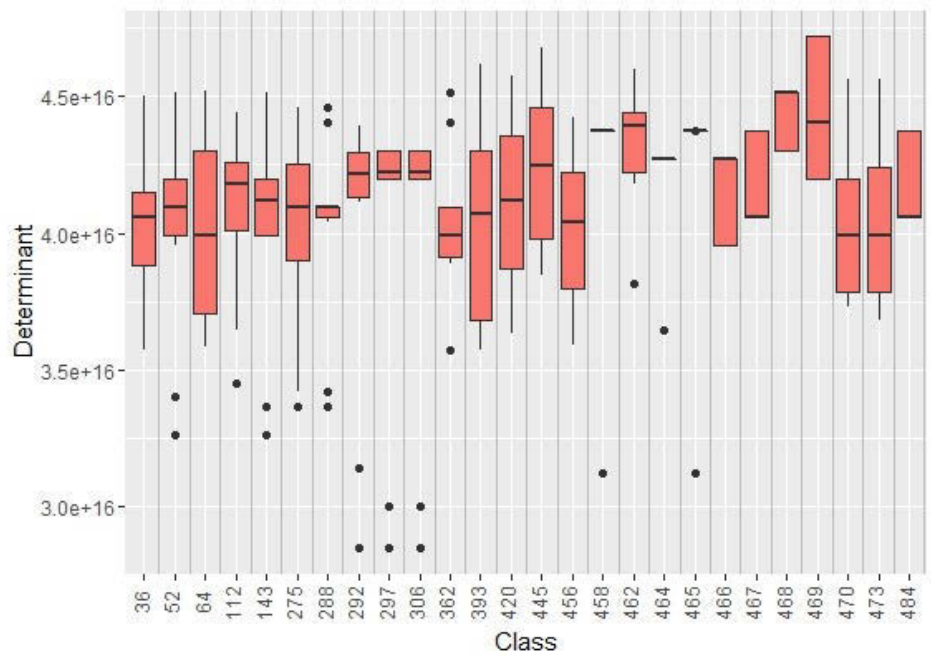
†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs.*

Figure 4.4.15: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^5]$ minus 1 run designs



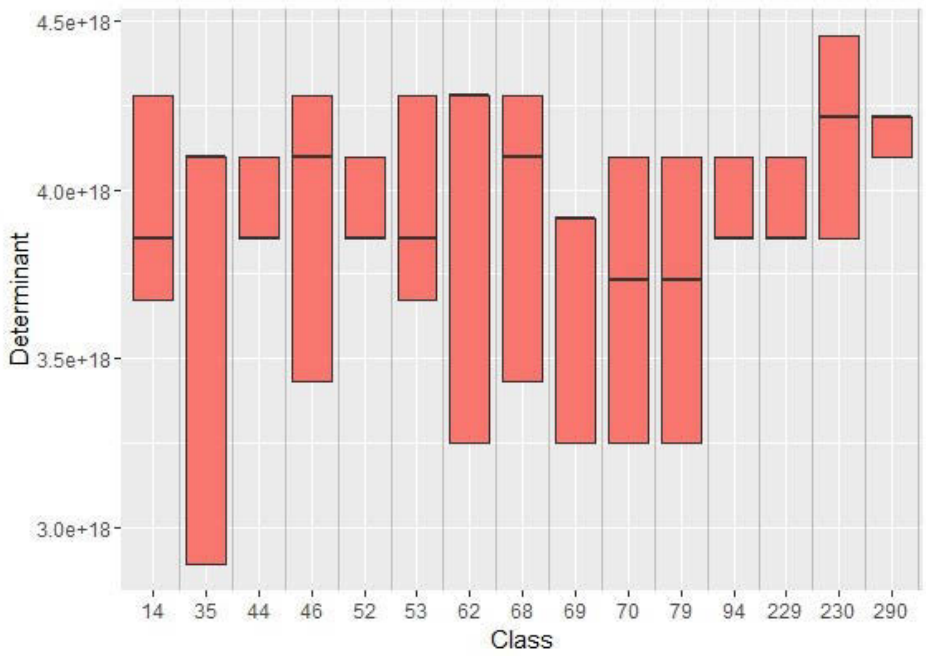
†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs.*

Figure 4.4.16: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^6]$ minus 1 run designs



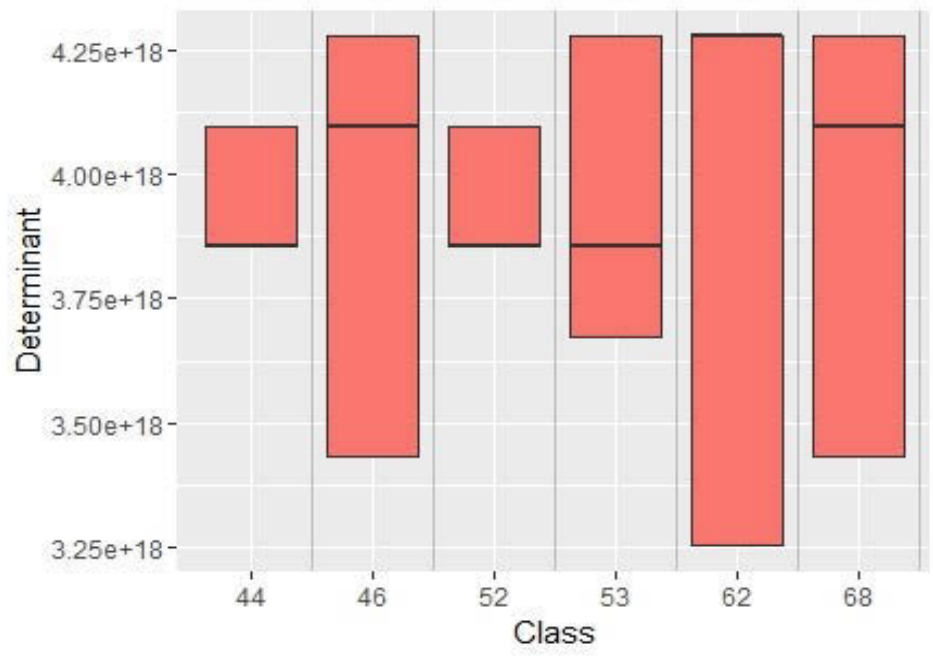
†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs.*

Figure 4.4.17: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^7]$ minus 1 run designs



†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs.*

Figure 4.4.18: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^7]$ minus 1 run designs



†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs. Any classes that results in a zero determinant for any combination of missing run/interaction has been omitted.*

Adjoining $p = 2$ Runs

For situations when we know which interaction will be included in the model, Table 4.4.7 gives the optimal $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 2 runs designs for $m > 3$.

Table 4.4.7: Optimal $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 2 runs designs

m	$\lvert\mathbf{M}_A\rvert$	No.†	Example		
			Class	Interaction	Runs
4	9.7444×10^{12}	4	5	(F_3, F_4)	0222 1002
5	3.4174×10^{15}	36	38	(F_1, F_3)	01022 01210
6	1.0482×10^{18}	57	11	(F_3, F_6)	012202 100100
7	2.9685×10^{20}	6,930	14	(F_3, F_7)	0100121 0122002

†*Number of occurrences of optimal* $|\mathbf{M}_A|$ *across all classes and all* $171 \times \binom{m}{2}$ *combinations of pairs of runs and interactions*

Recall from Section 4.2.4 we found it was generally better to search for runs in which neither of the factors involved in the interaction is set to the middle level. The example optimal designs in Table 4.4.7 confirm this observation when $m = 4$, 5 or 6 but this is not the case when $m = 7$. As noted in Section 4.2.4, the complicated nature of $|\mathbf{M}_A|$ when an interaction term is included in the model means we can not find simple general rules that can be applied to all situations.

For situations when we do not know which interaction will be included in the model, the best average $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 2 runs designs is given in Table 4.4.8 for $m > 3$.

Table 4.4.8: Best average $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 2 runs designs

m	No. Zeros†	Avg.$\lvert\mathbf{M}_A\rvert$	No.‡	Example	
				Class	Runs
4	0	7.7939×10^{12}	1	12	2200 2222
5	0	2.2266×10^{15}	1	38	01022 10222
6	0	6.3930×10^{17}	5	469	001220 012202
7	4	1.2867×10^{20}	15	62	0100101 0220122

†*Number of zero determinants amongst all* $\binom{m}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_A|$ *across all classes with given number of zeros and all 171 pairs of runs*

Missing $t = 2$ Runs

For situations when we know which interaction will be included in the model, Table 4.4.9 gives the optimal $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 2 runs designs when $m > 3$.

Table 4.4.9: Optimal $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 2 runs designs

m	$\lvert\mathbf{M}_B\rvert$	No.†	Example		
			Class	Interaction	Runs
4	8.9262×10^{11}	538	3	(F_2, F_3)	0010 0101
5	1.5710×10^{14}	162	38	(F_1, F_3)	00101 11120
6	4.0601×10^{17}	1	323	(F_1, F_2)	022101 220222
7	1.1244×10^{18}	28,260	10	(F_4, F_5)	0011111 0100102

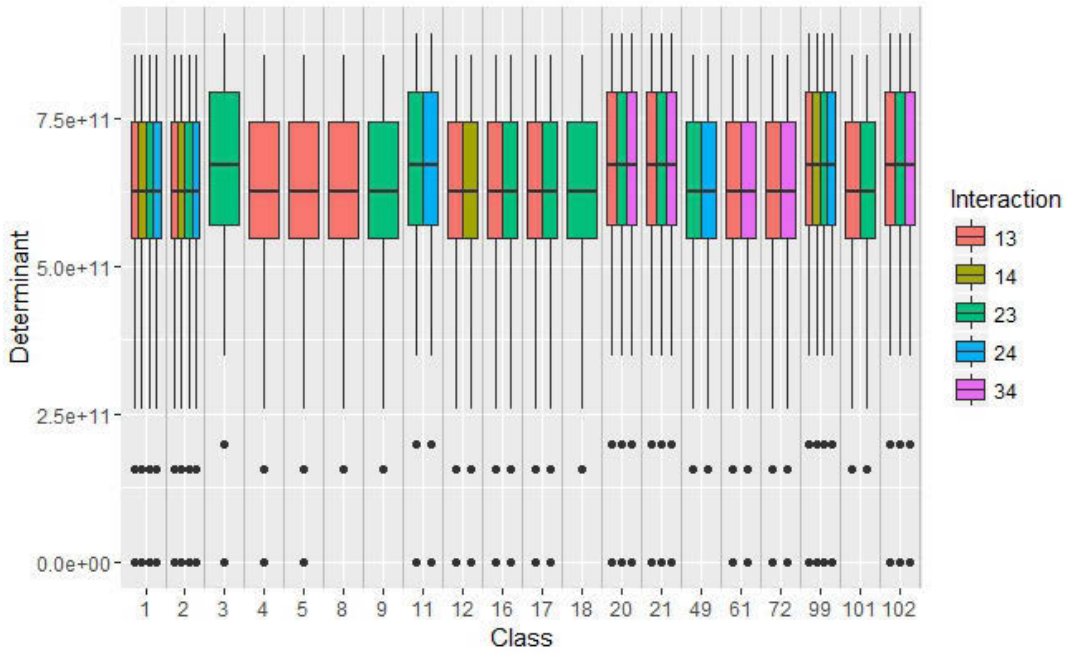
†*Number of occurrences of optimal* $|\mathbf{M}_B|$ *across all classes and all* $153 \times \binom{m}{2}$ *combinations of pairs of runs and interactions*

As discussed in Section 4.4.2, if we do not have control over which pair of runs will be missing from the OA, we require further information about the distribution of each class/interaction across all pairs of missing runs. Now that we are missing a pair of runs, rather than a single run, there are more class/interaction combinations that result in non-estimable models, hence there are a larger number of class/interactions to be excluded from the charts. For example, consider the distribution of OA$[18, 3^4]$ minus 2 runs designs in Figure 4.4.19. A majority of these class/interaction combinations contain at least one pair of runs that result in a zero determinant when removed from the OA. This includes the OA from the 3rd class with interaction between factors F_2 and F_3, which features in Table 4.4.9 for being optimal with a particular set of runs.

Since we want to protect against the possibility of obtaining a pair of missing runs that results in a non-estimable model for our choice of class/interactions combination, we will filter these distributions out of the chart. As we can see from Figure 4.4.20, this effectively leaves us with only one option in terms of the shape of the distribution across all pairs of missing runs.

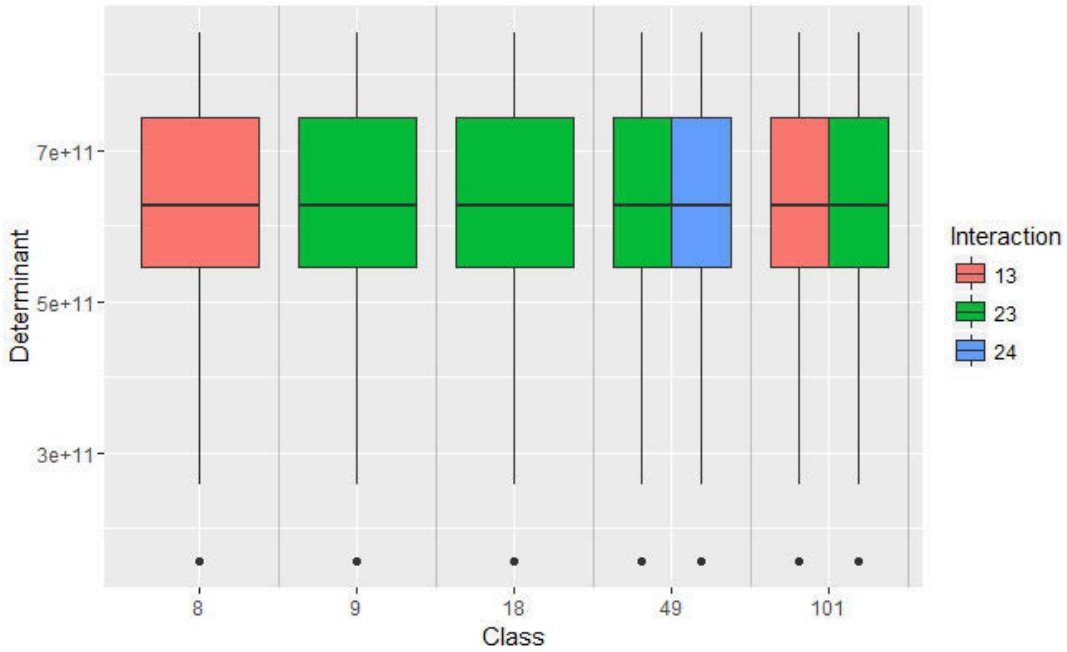
None of the class/interaction combinations in the top 5 percent of the OA$[18, 3^5]$ minus 2 runs designs have any pairs of runs that result in a non-estimable model, but the number of combinations to consider is very large, hence we have presented only those in the top 1 percent in Figure 4.4.21. Note that the optimal design given in Table 4.4.9, that is, class 38 with an interaction between F_1 and F_3, appears in this chart.

Figure 4.4.19: $|\mathbf{M}_B|$ for subset† of OA$[18, 3^4]$ minus 2 runs designs



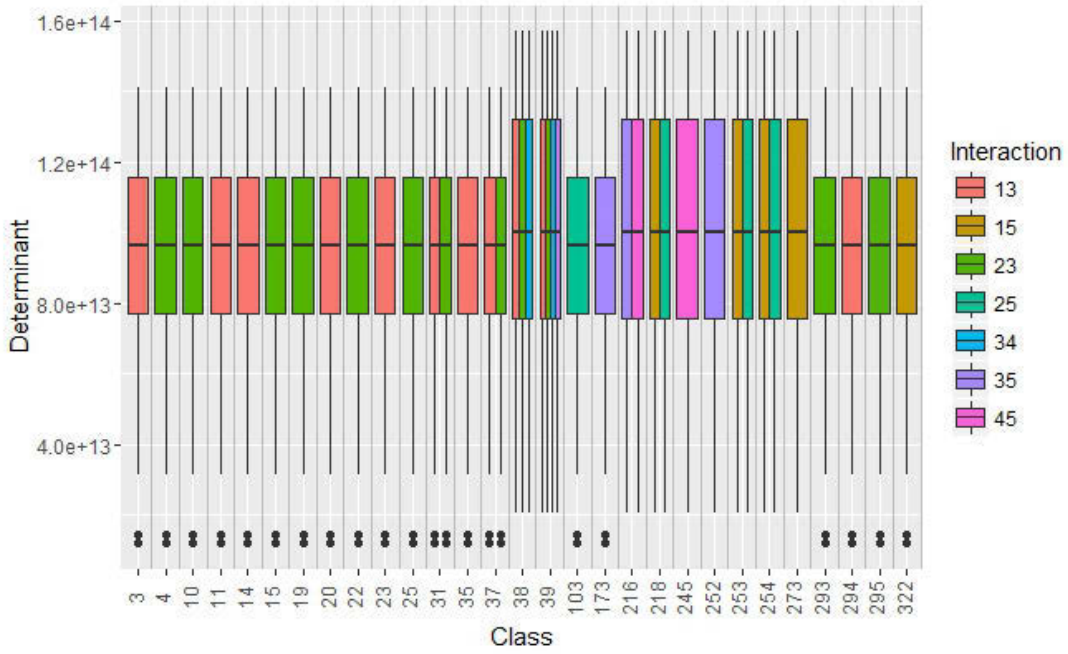
†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all class/interaction combinations.*

Figure 4.4.20: $|\mathbf{M}_B|$ for subset† of OA$[18, 3^4]$ minus 2 runs designs



†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all class/interaction combinations. Any class/interaction combination that results in a zero determinant for any pair of missing runs has been omitted.*

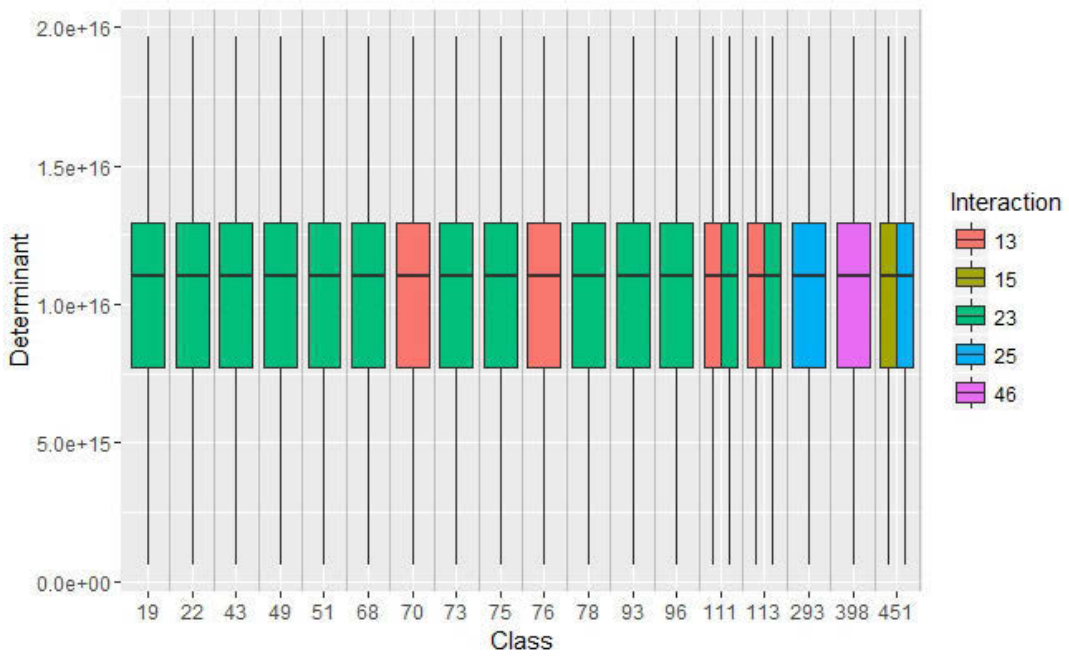
Figure 4.4.21: $|\mathbf{M}_B|$ for subset† of OA$[18, 3^5]$ minus 2 runs designs



†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 1 percent of all class/interaction combinations.*

Similarly for the OA$[18, 3^6]$ minus 2 runs designs, there are too many class/interactions in the top 5 percent to consider in a single chart, even after filtering out numerous combinations that contain zero values. Hence, Figure 4.4.22 presents the top 1 percent of class/interactions excluding those that contain zero determinants. Note that the optimal design given in Table 4.4.9, that is, class 323 with an interaction between F_1 and F_2, does not appear in this chart as there are 55 pairs of runs that result in a non-estimable model for this class/interaction combination.

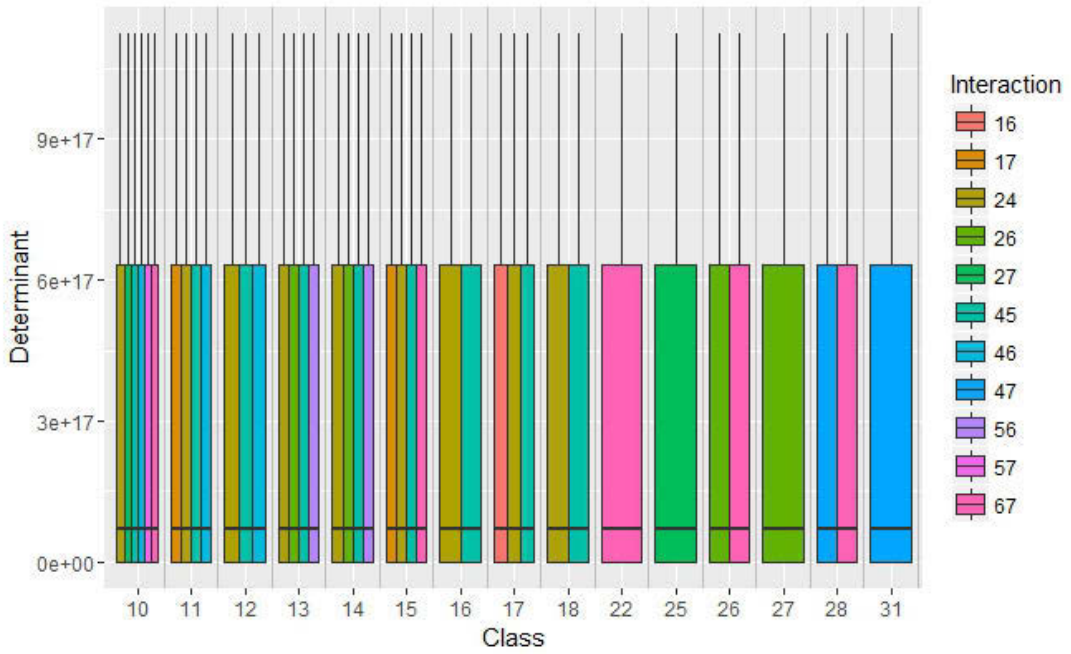
Figure 4.4.22: $|\mathbf{M}_B|$ for subset† of OA$[18, 3^6]$ minus 2 runs designs



†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 1 percent of all class/interaction combinations. Any class/interaction combination that results in a zero determinant for any pair of missing runs has been omitted.*

This systematic method of removing all class/interaction combinations that can result in non-estimable models is not appropriate when we consider OA$[18, 3^7]$ minus 2 runs designs because none of these combinations have any distributions without at least one zero. In fact, when $m = 7$ and $t = 2$, all class/interaction combinations result in a zero determinant in either 45, 81 or all 153 pairs of missing runs. Hence, Figure 4.4.23 gives the class/interactions that contain only 45 pairs of runs that result in a zero determinant. There are 371 such combinations in the top 1 percent, hence we only present the first 40 here. To ensure that we have not inadvertently filtered out note-worthy designs, we have checked that the distributions of each of the omitted class/interaction combinations are the same as those that have been plotted. Note that the optimal design given in Table 4.4.9, that is, class 10 with interaction between F_4 and F_5, appears in this chart.

Figure 4.4.23: $|\mathbf{M}_B|$ for subset† of OA$[18, 3^7]$ minus 2 runs designs



†*Each of the class/interaction combinations in this chart have average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 1 percent of all class/interaction combinations. All class/interaction combinations have result in a zero determinant for exactly 45 pairs of runs. There are 371 class/interaction combinations in the top 1 percent, hence we have only presented the first 40.*

Finally, in the case of model uncertainty, Table 4.4.10 gives the best average $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 2 runs designs when $m > 3$.

Table 4.4.10: Best average $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 2 runs designs

m	No. Zeros†	Avg.$\lvert\mathbf{M}_B\rvert$	No.‡	Example	
				Class	Runs
4	0	7.1699×10^{11}	4	12	1102 2111
5	0	1.0890×10^{14}	1	285	00111 11001
6	0	3.0321×10^{16}	1	323	022101 220222
7	5	4.4174×10^{17}	216	34	0011222 0122211

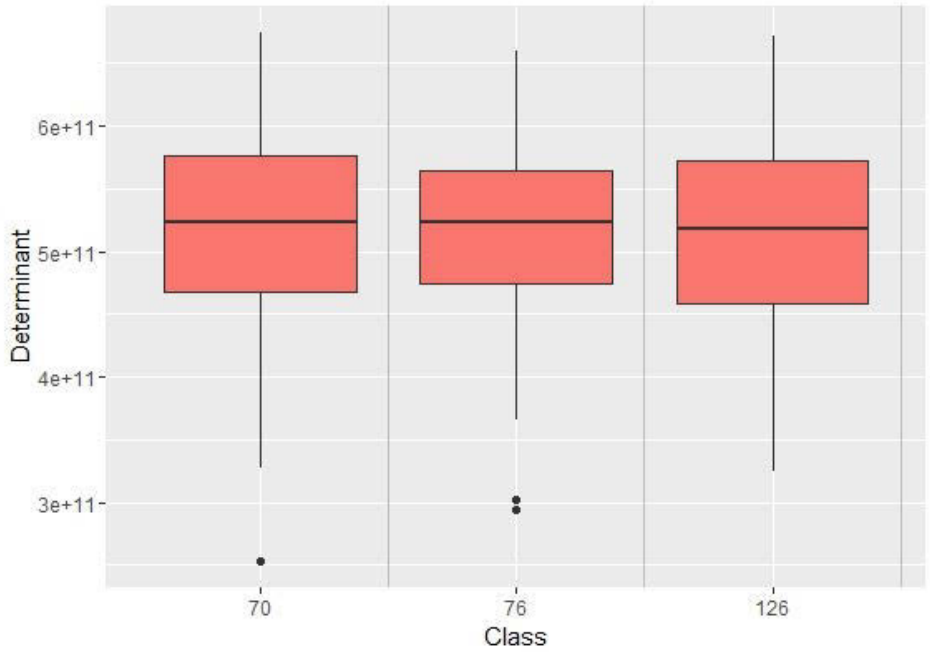
†*Number of zero determinants amongst all* $\binom{m}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_B|$ *across all classes with given number of zeros and all 153 pairs of runs*

For the situation when the experimenter does not have control over which pair of runs to remove, Figures 4.4.24, 4.4.25, 4.4.26 and 4.4.27 present the distribution of the average $|\mathbf{M}_B|$ for $m = 4$, 5, 6 and 7 respectively.

When $t = 2$, it is no longer the case that all classes have the same number of zero determinants across all interactions for each set of missing runs. For example,

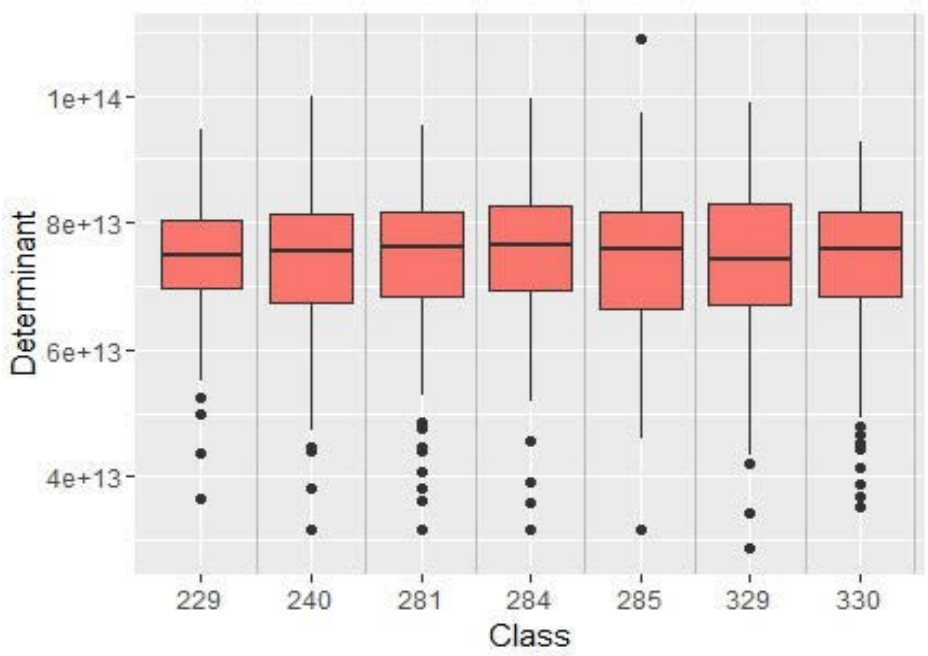
when $m = 4$, the 1st class achieves non-zero determinants in all 6 interactions for a majority of the 153 pairs of runs. However, when the first two runs are removed all 6 interactions produce a zero determinant. There are also 4 pairs of runs in which $|\mathbf{M}_B| = 0$ for 1 of the 6 interactions. Hence, despite this design achieving all non-zero determinants in 148 of the 153 pairs of missing runs, we will omit the entire class from the chart. Hence, the classes presented in the following figures achieve the minimum number of non-zero determinants in all pairs of runs.

Figure 4.4.24: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^4]$ minus 2 runs designs



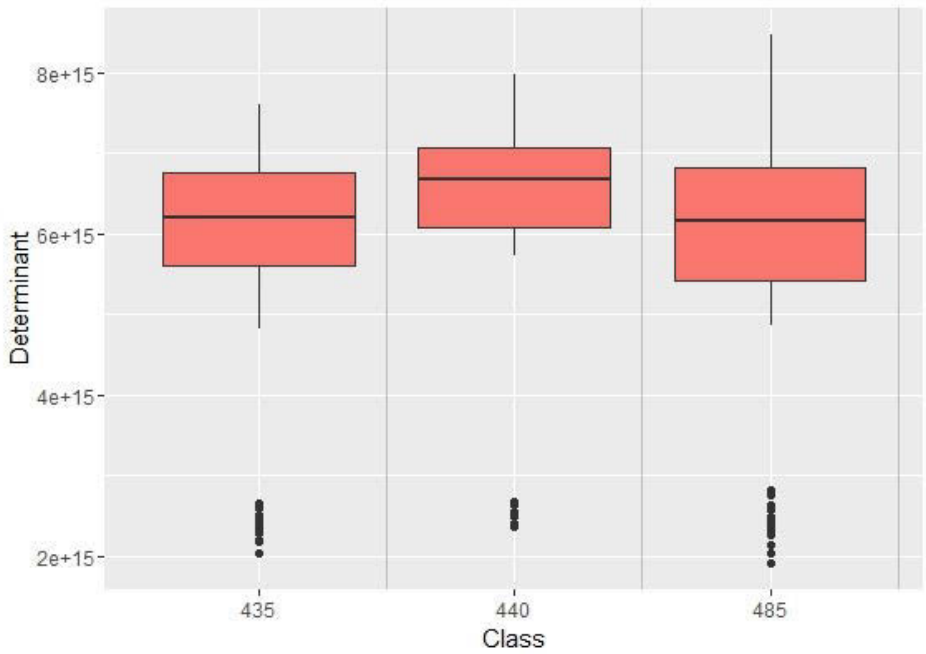
†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs. Any classes that results in a zero determinant for any combination of missing run/interaction has been omitted.*

Figure 4.4.25: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^5]$ minus 2 runs designs



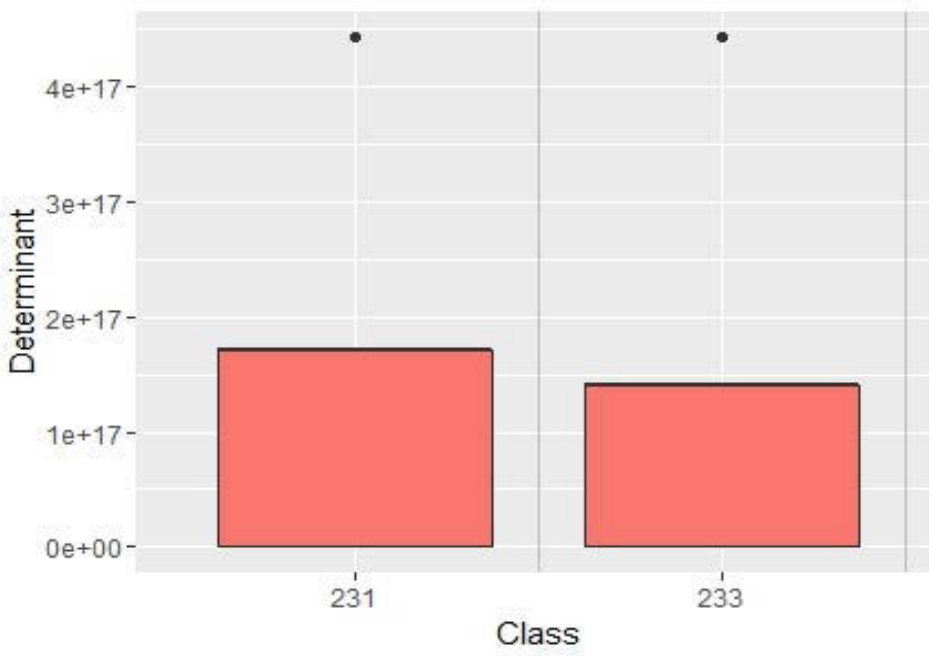
†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs. Any classes that results in a zero determinant for any combination of missing run/interaction has been omitted.*

Figure 4.4.26: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^6]$ minus 2 runs designs



†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs. Any classes that results in a zero determinant for any combination of missing run/interaction has been omitted.*

Figure 4.4.27: Average $|\mathbf{M}_B|$ for subset† of OA$[18, 3^7]$ minus 2 runs designs



†*Each of the OAs in this chart have average average* $|\mathbf{M}_B|$ *across all adjoined runs in the top 5 percent of all OAs. Any classes that results in more than 5 zero determinants for any combination of missing run/interaction has been omitted.*

4.4.3 Model-Robustness Criteria

Throughout this empirical investigation we have used a two-stage procedure to assess the model-robustness of each design across all models containing a single linear-by-linear interaction. In the first stage, we have filtered out any designs that are not able to estimate the maximum number of models. Ideally, we would like all models to be estimable, hence if at least one design exists such that all models are estimable, then we will filter out any designs that do not meet this requirement. If, however, no design exists such that all models are estimable, then we determine the maximum number of estimable models across all designs and filter out any designs that do not realise this maximum. In the second stage, we take the average $|\mathbf{M}_A|$ (or $|\mathbf{M}_B|$) across all models. Hence, our dual-purpose criterion for model-robustness prioritises model-estimablity over D-optimality.

Recall from Chapter 1 that Cheng et al. (1999) established the concept of estimation capacity (EC) to assess model-robustness. EC is defined as the proportion of estimable models within the model space. Furthermore, Li and Nachtsheim (2000) defined information capacity (IC) as the average D-efficiency over all models. They also developed model-robust factorial designs (MRFDs) by selecting designs with maximum IC from those with EC = 100% (if possible).

Our work is essentially equivalent to the MRFDs in Li and Nachtsheim (2000). That is, in our first stage we seek designs in which all models are estimable (this essentially means we seek EC = 100%), and then we assess the average $|\mathbf{M}_A|$ (or $|\mathbf{M}_B|$) within these designs (this is similar to the IC criterion). A superficial point of difference between the two procedures is that our version of IC takes the average of the 'raw' determinants, whereas Li and Nachtsheim (2000) use the efficiency. Since the number of parameters and the optimal determinant are constant across all models within a given model space, these two averages are assessing the same outcome. Hence, the tables of the best average $|\mathbf{M}_A|$ (or $|\mathbf{M}_B|$) presented in this chapter are essentially tables of the best MRFDs.

Of course, our motivation differs from that of most authors in the area of optimality criteria. We are not interested in the efficiency of the criteria per se, rather we are simply using these criteria as tools for assessing the performance of OA plus p runs designs and OA minus t runs designs.

We note that the measure of a design's model-robustness does not necessarily speak to its model-discrimination properties. This distinction is addressed in the next chapter along with discussions of similar areas of research, such as search designs.

Chapter 5

Models Containing More Than One Linear-by-Linear Interaction

In Chapter 4 we considered the performance of OA plus p runs designs and OA minus t runs designs when a single linear-by-linear interaction is included in the model. We demonstrated how a thorough investigation can be carried out via empirical calculations and gave the results of such a study for all OA$[18, 3^m]$ plus p runs designs and all OA$[18, 3^m]$ minus t runs designs, $p, t \in (1, 2)$. We showed that our derivation of the $p \times p$ matrix $\mathbf{\Psi}_A$ and the $t \times t$ matrix $\mathbf{\Psi}_B$ greatly increased calculation efficiency which made the exhaustive empirical study possible.

In this chapter we extend this work by considering models that include ν linear-by-linear interactions, where the bounds on ν are discussed in Section 5.1.3. We show how the inclusion of additional interaction terms in the model affects $\mathbf{M}^{-1}$, and that the dimensions of $\mathbf{\Psi}_A$ and $\mathbf{\Psi}_B$ do not change. Hence, the process of an empirical investigation in the context of $\nu > 1$ will be very similar to the process demonstrated in Chapter 4 when $\nu = 1$. In Section 5.2, we give the results of such a study for $\nu = 2$ to illustrate these similarities.

The primary focus of this chapter, however, is not the results of the empirical studies per se, but rather on how these investigations tie into other areas of the literature, such as search designs. Specifically, we are interested in the distinction between model-robustness and model-discrimination. Hence, we focus on these concepts in Section 5.3.

5.1 Information Matrix for the OA

In Section 4.1 we examined the structure of the information matrix for an OA when a single linear-by-linear interaction is included in the model. We will extend that work in this section by examining the structure of the information matrix for an OA when ν linear-by-linear interactions are included in the model.

Following Definition 1.1.4 in Chapter 1, we use $\mathbf{Z}$ to denote the $N \times \nu$ matrix containing the ν columns corresponding to the linear-by-linear interaction terms to be included in the model. Thus, the partitioned $N \times \upsilon$ model matrix for a model containing ν linear-by-linear interactions is

$$\mathbf{X} = \left[\begin{array}{cc} \mathbf{X}_{ME} & \mathbf{Z} \end{array} \right] \tag{5.1.1}$$

where $\mathbf{X}_{ME}$ is the $N \times \alpha$ model matrix for the main effects only model, and $\upsilon = \alpha + \nu$.

Example 5.1.1. Following Example 4.1.1, suppose we use a OA$[18, 3^4]$ from the first geometric isomorphism class. Further suppose $\nu = 6$, that is, we would like to include all $\binom{4}{2} = 6$ linear-by-linear interactions in the model. Then the 18×6 matrix of interaction terms is

$$\mathbf{Z} = \begin{array}{c} \begin{array}{cccccc} \mathbf{z}_{(12)} & \mathbf{z}_{(13)} & \mathbf{z}_{(14)} & \mathbf{z}_{(23)} & \mathbf{z}_{(24)} & \mathbf{z}_{(34)} \end{array} \\ \left[\begin{array}{cccccc} \frac{3}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{3}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{3}{2} & 0 & 0 & \frac{3}{2} \\ 0 & -\frac{3}{2} & -\frac{3}{2} & 0 & 0 & \frac{3}{2} \\ -\frac{3}{2} & \frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} & \frac{3}{2} & -\frac{3}{2} \\ -\frac{3}{2} & -\frac{3}{2} & \frac{3}{2} & \frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & 0 & \frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & 0 & -\frac{3}{2} & \frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{2} & 0 \\ -\frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \\ -\frac{3}{2} & \frac{3}{2} & \frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} & \frac{3}{2} \\ 0 & 0 & -\frac{3}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{2} & 0 & 0 & 0 \\ \frac{3}{2} & -\frac{3}{2} & 0 & -\frac{3}{2} & 0 & 0 \\ \frac{3}{2} & \frac{3}{2} & 0 & \frac{3}{2} & 0 & 0 \end{array} \right] \end{array}.$$

□

The associated information matrix is

$$\mathbf{M} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}'_{ME} \\ \mathbf{Z}' \end{bmatrix} \begin{bmatrix} \mathbf{X}_{ME} & \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_{ME}\mathbf{X}_{ME} & \mathbf{X}'_{ME}\mathbf{Z} \\ \mathbf{Z}'\mathbf{X}_{ME} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}. \tag{5.1.2}$$

We note that the columns of the $\alpha \times \nu$ matrix $\mathbf{X}'_{ME}\mathbf{Z}$ are equivalent to the $\mathbf{u}_{(ab)}$ column vectors we studied in Chapter 4 for each of the interactions. Hence, we will denote $\mathbf{X}'_{ME}\mathbf{Z}$ by $\mathbf{U}$.

Example 5.1.2. Consider the matrix of interactions, $\mathbf{Z}$, constructed in Example 5.1.1. Then

$$\mathbf{U} = \mathbf{X}'_{ME}\mathbf{Z} = \begin{array}{c} \begin{matrix} \mathbf{u}_{(12)} & \mathbf{u}_{(13)} & \mathbf{u}_{(14)} & \mathbf{u}_{(23)} & \mathbf{u}_{(24)} & \mathbf{u}_{(34)} \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3\sqrt{\frac{3}{2}} \\ 0 & 0 & 0 & 0 & 0 & \frac{9}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & 0 & -3\sqrt{\frac{3}{2}} \\ 0 & 0 & 0 & 0 & 0 & -\frac{9}{\sqrt{2}} \\ 0 & 0 & 3\sqrt{\frac{3}{2}} & 0 & -3\sqrt{\frac{3}{2}} & 0 \\ -\frac{9}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 3\sqrt{\frac{3}{2}} & 0 & -3\sqrt{\frac{3}{2}} & 0 & 0 \\ -9\sqrt{2} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}. \qquad \square$$

We note that although the number of columns in $\mathbf{Z}$ and $\mathbf{U}$ increases as ν increases, so far all of the concepts we have considered in this chapter are the same as those seen in Chapter 4. For example, the columns of $\mathbf{U}$ given in the previous example are the same columns we enumerated in Chapter 4, that is, they can all be constructed from the values of γ_{abc} and $\omega_{(ab)c}$ that we have already enumerated. We now define a $\nu \times \nu$ matrix, $\mathbf{W}$, which will be the only new notation in this chapter.

Recall from Chapter 4 that when we consider only a single linear-by-liner interaction (that is, $\nu = 1$), the bottom right corner of the partitioned information matrix for the OA is $\mathbf{z}'_{(ab)}\mathbf{z}_{(ab)}$, which we showed to be equal to N. In this chapter, the bottom right corder of the partitioned information matrix is equal to the matrix $\mathbf{Z}'\mathbf{Z}$.

Let $\mathbf{W}$ be the $\nu \times \nu$ matrix:

$$\begin{aligned}\mathbf{W} &= \mathbf{Z}'\mathbf{Z} \\ &= \begin{bmatrix} N & \mathbf{z}'_{(12)}\mathbf{z}_{(13)} & \mathbf{z}'_{(12)}\mathbf{z}_{(14)} & \cdots & \mathbf{z}'_{(12)}\mathbf{z}_{(m-1;m)} \\ \mathbf{z}'_{(13)}\mathbf{z}_{(12)} & N & \mathbf{z}'_{(13)}\mathbf{z}_{(14)} & \cdots & \mathbf{z}'_{(13)}\mathbf{z}_{(m-1;m)} \\ \vdots & & \ddots & & \vdots \\ \mathbf{z}'_{(m-1;m)}\mathbf{z}_{(12)} & \mathbf{z}'_{(m-1;m)}\mathbf{z}_{(13)} & \mathbf{z}'_{(m-1;m)}\mathbf{z}_{(14)} & \cdots & N \end{bmatrix} \\ &= \begin{bmatrix} N & w_{1123} & w_{1124} & \cdots & w_{12;m;m-1} \\ w_{1123} & N & w_{1134} & \cdots & w_{13;m;m-1} \\ \vdots & & \ddots & & \vdots \\ w_{12;m;m-1} & w_{13;m;m-1} & w_{14;m;m-1} & \cdots & N \end{bmatrix}\end{aligned} \tag{5.1.3}$$

where $w_{abcd} = \sum_{i=1}^{N} f_{al} f_{bl} f_{cl} f_{dl}$.

Then Equation (5.1.2) becomes

$$\mathbf{M} = \begin{bmatrix} N\mathbf{I}_\alpha & \mathbf{U} \\ \mathbf{U}' & \mathbf{W} \end{bmatrix}. \tag{5.1.4}$$

Example 5.1.3. Following Examples 5.1.1 and 5.1.2, the $\mathbf{W}$ matrix for the OA$[18, 3^4]$ from the first geometric isomorphism class under a model containing all $\nu = 6$ linear-by-linear interactions is

$$\mathbf{W} = \mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 18 & 0 & 0 & 0 & 0 & 0 \\ 0 & 18 & \frac{9}{2} & -\frac{9}{2} & 0 & 0 \\ 0 & \frac{9}{2} & 18 & 0 & -9 & 0 \\ 0 & -\frac{9}{2} & 0 & 18 & -\frac{9}{2} & 0 \\ 0 & 0 & -9 & -\frac{9}{2} & 18 & 0 \\ 0 & 0 & 0 & 0 & 0 & 18 \end{bmatrix}.$$

□

5.1.1 Inverse of the Information Matrix for the OA

Given the partitioned representation of $\mathbf{M}$ in (5.1.4), let $\mathbf{T} = N\mathbf{I}_\alpha$ and $\mathbf{V} = \mathbf{U}'$. Then $\mathbf{M} = \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$. We can now apply Theorem 1.2.2 to find the inverse of

M.

$$\begin{aligned}\mathbf{M}^{-1} &= \begin{bmatrix} \mathbf{T}^{-1} & \mathbf{0}_{\alpha\times\nu} \\ \mathbf{0}_{\nu\times\alpha} & \mathbf{0}_{\nu\times\nu} \end{bmatrix} + \begin{bmatrix} -\mathbf{T}^{-1}\mathbf{U} \\ \mathbf{I}_\nu \end{bmatrix} \mathbf{Q}^{-1} \begin{bmatrix} -\mathbf{V}\mathbf{T}^{-1} & \mathbf{I}_\nu \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_\alpha/N & \mathbf{0}_{\alpha\times\nu} \\ \mathbf{0}_{\nu\times\alpha} & \mathbf{0}_{\nu\times\nu} \end{bmatrix} + \begin{bmatrix} -\mathbf{U}/N \\ \mathbf{I}_\nu \end{bmatrix} \mathbf{Q}^{-1} \begin{bmatrix} -\mathbf{U}'/N & \mathbf{I}_\nu \end{bmatrix}\end{aligned} \tag{5.1.5}$$

where

$$\begin{aligned}\mathbf{Q} &= \mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U} \\ &= N - \mathbf{U}'\mathbf{U}/N.\end{aligned} \tag{5.1.6}$$

Let the ith interaction be between factors F_a and F_b. Then the diagonal entry in the ith row/column of $\mathbf{U}'\mathbf{U}$ is:

$$\sum_{\substack{j=1 \\ j\notin(a,b)}}^{m} \left(u^2_{(ab)jl} + u^2_{(ab)jq}\right).$$

Let the jth interaction be between factors F_c and F_d. Then the entry in the ith row and j column of $\mathbf{U}'\mathbf{U}$ is:

$$\sum_{\substack{j=1 \\ j\notin(a,b,c,d)}}^{m} \left(u^2_{(ab)jl}u^2_{(cd)jl} + u^2_{(ab)jq}u^2_{(cd)jq}\right).$$

5.1.2 Information Matrix for the Altered Design

We examined the determinant of the information matrix for OA plus p runs designs and OA minus t runs designs when a single interaction is included in the model in Sections 4.2.1 and 4.2.2 respectively. The arguments in each of those sections remain the same in the context of model that contain ν interactions, so we will simply report the results here.

Adjoining p runs

Recall from Equation (4.2.1) that the determinant of the information matrix for an OA plus p runs design is

$$|\mathbf{M}_A| = |\mathbf{M}||\mathbf{\Psi}_A| \tag{5.1.7}$$

where $\mathbf{A}$ contains the p rows from the model matrix corresponding to the adjoined runs, and $\mathbf{\Psi}_A = \mathbf{I}_p + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}'$. In Chapter 4 we assumed there was only one interaction in the model, hence $\mathbf{M}$ was a $(\alpha+1)\times(\alpha+1)$ matrix and the dimensions of $\mathbf{A}$ were $p \times (\alpha+1)$. Thus, $\mathbf{\Psi}_A$ was a $p \times p$ matrix. In this chapter, $\mathbf{M}$ is a

$\upsilon \times \upsilon$ matrix, so the calculation of $|\mathbf{M}|$ may be slightly slower now that we have increased the number of terms in the model, but this only needs to be calculated once for each OA. The dimensions of $\mathbf{A}$ in this chapter are $p \times \upsilon$, hence $\mathbf{\Psi}_A$ is still a $p \times p$ matrix, and so the calculation of $|\mathbf{\Psi}_A|$ for each set of p runs should be roughly the same, regardless of how many terms are included in the model. Thus, despite the increase in size and complexity of $\mathbf{M}$, we can still reap the benefits of efficient calculation times for empirical studies as ν increases.

Missing t runs

Similarly, from Equation (4.2.2), the determinant of the information matrix for an OA minus t runs design is

$$|\mathbf{M}_B| = (-1)^t |\mathbf{M}||\mathbf{\Psi}_B| \tag{5.1.8}$$

where $\mathbf{B}$ contains the t rows from the model matrix corresponding to the adjoined runs, and $\mathbf{\Psi}_B = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}' - \mathbf{I}_t$. Hence, the discussion about the dimensions of $\mathbf{M}$ and $\mathbf{\Psi}_A$ are applicable here as well but for $\mathbf{\Psi}_B$ rather than $\mathbf{\Psi}_A$.

5.1.3 Bound on ν

The number of terms we are able to fit will depend on the number of unique runs in the experiment. A design in n runs, none of which are repeated, will allow for the estimation of n terms in the model including the intercept. Suppose r of these runs are repeated, then we have $n - r$ degrees of freedom for model fitting, thus we can fit a maximum of $n - r$ terms. These r terms could be in the form of a single run being represented $r + 1$ times in the design, or r pairs of repeated runs, or something between these two extremes. The following example illustrates this.

Example 5.1.4. Consider the three 6-run 2^2 designs below.

$\mathbf{R}_1$	$\mathbf{R}_2$	$\mathbf{R}_3$
00	00	00
00	00	00
00	01	00
00	01	01
01	11	11
11	11	11

Each of these designs has $r = 3$ repeated runs. $\mathbf{R}_1$ has 4 representatives of the run 00, hence 3 of these are repeated. The runs 00, 01 and 11 are each repeated pairs

in $\mathbf{R}_2$. $\mathbf{R}_3$ has 3 representatives of the run 00 (contributes 2 to r), and the run 11 appears as a repeated pair (contributes 1 to r). Hence, we can fit models with a maximum of $6 - 3 = 3$ terms to each of these designs. □

An underlying assumption we have made throughout this thesis is that all models contain the intercept and all main effect terms for all factors. This requires α degrees of freedom. Hence, we have $n - r - \alpha$ degrees of freedom left over for linear-by-linear interaction terms.

Notice that the desire to maximise the number of estimable terms in the model is in opposition to the desire to maximise the number of degrees of freedom available for estimating pure error (that is, the number of repeated runs). We will discuss this issue further in section 6.2.2 of the next chapter.

In addition to the number of unique runs in the design, the number of interactions we can include in the model also depends on the number factors in the design. A design with m factors has $\binom{m}{2}$ potential two-factor interactions to include in the model. This discussion is formalised in the following Lemma.

Lemma 5.1.1. Let $\mathbf{R}$ be an m-factor design in n runs, r of which are repeated, and let α be the number of terms in the main effects only model. Then the bounds on ν, the number of linear-by-linear interactions that can be included the model for design $\mathbf{R}$, are

$$0 \leq \nu \leq \min\left(n - r - \alpha, \binom{m}{2}\right).$$

□

Now consider the designs that we have been focusing on in Chapter 4, that is, OA$[18, 3^m]$ plus p runs designs and OA$[18, 3^m]$ minus t runs designs. The number of terms in the main effects only model for these designs is $\alpha = 1 + 2m$. Hence, the bound on ν for OA$[18, 3^m]$ plus p runs designs is $0 \leq \nu \leq \min\left(18 + p - r - (1 + 2m), \binom{m}{2}\right)$ where r is the number of repeated runs in the set of all $18 + p$ runs. The bound on ν for OA$[18, 3^m]$ minus t runs designs is $0 \leq \nu \leq \min\left(18 - t - r - (1 + 2m), \binom{m}{2}\right)$ where r is the number of repeated runs in the subset of $18 - t$ runs from the OA$[18, 3^m]$.

5.2 Empirical Results For $\nu = 2$

Recall from Chapter 4 that there are two measures of goodness we consider in our empirical studies: D-optimality and model-robustness. D-optimality is appropriate when we have certainty about which interaction(s) will be included in

the model, while model-robustness is appropriate under model uncertainty. Furthermore, there are two methods for presenting the results of empirical studies for each of these criterion: tables and graphs. Tables are appropriate in the context of adjoined runs since we have control over which runs to adjoin, hence we simply require information about which set of runs is best and, assuming we know which factors will be involved in the interaction(s), which columns to assign these factors to. Similar comments can be made in the context of missing runs when we have control over which runs to remove. However, when we are not certain which runs will be missing from the OA, graphs are more appropriate than tables as they give us a picture of the entire distribution for a given optimality criterion over all possible sets of t missing runs. Hence, our empirical investigation in Chapter 4 involved both tables and graphs.

As noted in Chapter 1, our primary interest in the context of missing runs is when the resources available for an experiment are slightly less than an OA. Hence, we usually consider 'unavailable runs' to refer to the runs that we intentionally remove. Since we have already demonstrated how to assess empirical investigations when runs are missing at random in Chapter 4, we will omit the graphs in this section as they are not our primary concern. Instead, we will present the tables of the D-optimal and model-robust designs for OA plus p runs designs and OA minus t runs designs, $p, t \in (1, 2)$.

As noted above, our derived matrices $\mathbf{\Psi}_A$ and $\mathbf{\Psi}_B$ facilitate efficient calculations for empirical studies, even as ν increases. Hence, it is not difficult to obtain tables of optimal deigns for any value of ν. However, as stated at the beginning of this chapter, now that we have demonstrated the use of these matrices for empirical studies, our focus for this chapter is to relate these methods to other areas of the literature. Hence, we will limit the presentation of empirical results to $\nu = 2$. These results will be useful to refer back to in Section 5.3 where we discuss the similarities between this research and search designs.

5.2.1 Adjoining $p = 1$ Run, $\nu = 2$

Table 5.2.1 gives the best values of $|\mathbf{M}_A|$ (D-optimality) for each of the OA$[18, 3^m]$ plus 1 run designs, along with examples of designs that achieve these values.

Table 5.2.2 gives the best values of the average $|\mathbf{M}_A|$ (model-robustness) for each of the OA$[18, 3^m]$ plus 1 run designs, along with examples of designs that achieve these values. Recall from Chapter 4 that we have a two-stage method for determining model-robustness. First, we determine the smallest possible number of non-zero determinants across all models within each design, then we calculate the

Table 5.2.1: Optimal $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 1 run designs

m	$\lvert\mathbf{M}_A\rvert$	No.†	Example		
			Class	Interactions	Run
3	3.2027×10^{11}	2	1	(F_1, F_3),(F_2, F_3)	220
4	1.1046×10^{14}	10	20	(F_1, F_3),(F_3, F_4)	2002
5	3.0817×10^{16}	16	38	(F_1, F_3),(F_3, F_4)	01022
6	8.2251×10^{18}	23	15	(F_1, F_3),(F_3, F_5)	022000
7	1.9430×10^{21}	516	19	(F_1, F_3),(F_3, F_5)	0100022

†*Number of occurrences of optimal* $|\mathbf{M}_A|$ *across all classes and all* $18 \times \binom{\binom{m}{2}}{2}$ *run/interaction combinations*

average determinant for those designs that satisfy this minimum. When $m \leq 6$ the smallest number of zeros across all models is zero, hence it is possible to construct OA$[18, 3^m]$ plus 1 run designs such that any model that contains $\nu = 2$ linear-by-linear interactions is estimable. When $m = 7$, there are $\binom{7}{2} = 21$ potential interactions to be included in the model, hence there are $\binom{21}{2} = 210$ models to consider when $\nu = 2$. As we can see from Table 5.2.2, the best number of non-zero determinants we can achieve from all 210 models is 72.

Table 5.2.2: Best average $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 1 run designs

m	No. Zeros†	Avg.$\lvert\mathbf{M}_A\rvert$	No.‡	Example	
				Class	Run
3	0	3.0443×10^{11}	2	1	220
4	0	7.0110×10^{13}	2	12	2200
5	0	1.4617×10^{17}	1	17	22020
6	0	2.6793×10^{18}	1	445	022012
7	72	1.5791×10^{20}	6	201	0010100

†*Number of zero determinants amongst all* $\binom{\binom{m}{2}}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_A|$ *across all classes with given number of zeros and all 18 runs*

We note that our two-stage method can sometimes result in average $|\mathbf{M}_A|$ values that are smaller then would be the case if we simply took the average without regard for the number of estimable models. For example, when $m = 7$, the 290th geometric isomorphism class with the 5th run adjoined has average $|\mathbf{M}_A| = 2.852 \times 10^{20}$ across all 210 models, which is better than the value given in Table 5.2.2. However, 92 of these determinants are zero. Thus, although the non-zero determinants for this design are generally better than those for the design given in Table 5.2.2, this design will have $92/210 = 43.81\%$ chance of not being able to estimate the model. Our-two step procedure prioritises the minimisation of non-estimable models, hence we prefer the design in Table 5.2.2 since it only has a $72/210 = 34.29\%$ chance of not being able to estimate the model.

5.2.2 Missing $t = 1$ Run, $\nu = 2$

Table 5.2.3 gives the best values of $|\mathbf{M}_B|$ (D-optimality) for each of the OA$[18, 3^m]$ minus 1 run designs, along with examples of designs that achieve these values.

Table 5.2.3: Optimal $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 1 run designs

m	$\lvert\mathbf{M}_B\rvert$	No.†	Example		
			Class	Interactions	Run
3	1.1364×10^{11}	8	1	(F_1, F_3),(F_2, F_3)	001
4	3.2134×10^{13}	4	99	(F_1, F_3),(F_2, F_4)	1111
5	6.2983×10^{15}	48	38	(F_1, F_3),(F_3, F_4)	00101
6	1.1140×10^{18}	93	15	(F_1, F_3),(F_3, F_5)	001210
7	1.6192×10^{20}	258	19	(F_1, F_3),(F_3, F_5)	0011212

†*Number of occurrences of optimal* $|\mathbf{M}_B|$ *across all classes and all* $18 \times \binom{\binom{m}{2}}{2}$ *run/interaction combinations*

Table 5.2.4 gives the best values of the average $|\mathbf{M}_B|$ (model-robustness) for each of the OA$[18, 3^m]$ minus 1 run designs, along with examples of designs that achieve these values.

Table 5.2.4: Best average $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 1 run designs

m	No. Zeros†	Avg.$\lvert\mathbf{M}_B\rvert$	No.‡	Example	
				Class	Run
3	0	1.0813×10^{11}	2	1	110
4	0	2.0207×10^{13}	2	12	2111
5	0	2.8967×10^{15}	1	38	11120
6	0	2.9486×10^{17}	1	445	110212
7	84	4.1474×10^{18}	12	95	0001121

†*Number of zero determinants amongst all* $\binom{\binom{m}{2}}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_B|$ *across all classes with given number of zeros and all 18 runs*

5.2.3 Adjoining $p = 2$ Runs, $\nu = 2$

Table 5.2.5 gives the best values of $|\mathbf{M}_A|$ (D-optimality) for each of the OA$[18, 3^m]$ plus 2 runs designs, along with examples of designs that achieve these values.

Table 5.2.5: Optimal $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 2 runs designs

m	$\lvert\mathbf{M}_A\rvert$	No.†	Example		
			Class	Interactions	Runs
3	5.4928×10^{11}	1	1	(F_1, F_3),(F_2, F_3)	220 222
4	2.0251×10^{14}	5	20	(F_1, F_3),(F_3, F_4)	2002 2022
5	5.8634×10^{16}	8	38	(F_1, F_3),(F_3, F_4)	01022 21002
6	1.5784×10^{19}	46	15	(F_1, F_3),(F_3, F_5)	022000 110220
7	1.3111×10^{24}	1	8	(F_1, F_3),(F_5, F_7)	0220020 2022101

†*Number of occurrences of optimal* $|\mathbf{M}_A|$ *across all classes and all* $171 \times \binom{m}{2}$ *combinations of pair of runs and interaction*

Table 5.2.6 gives the best values of the average $|\mathbf{M}_A|$ (model-robustness) for each of the OA$[18, 3^m]$ plus 2 runs designs, along with examples of designs that achieve these values.

Table 5.2.6: Best average $|\mathbf{M}_A|$ for OA$[18, 3^m]$ plus 2 runs designs

m	No. Zeros†	Avg.$\lvert\mathbf{M}_A\rvert$	No.‡	Example	
				Class	Runs
3	0	5.1094×10^{11}	1	1	220 222
4	0	1.2087×10^{14}	1	12	2200 2222
5	0	2.5516×10^{16}	1	17	00112 22020
6	0	4.9703×10^{18}	4	445	000122 022012
7	66	4.7657×10^{20}	1	200	0222212 1011010

†*Number of zero determinants amongst all* $\binom{m}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_A|$ *across all classes with given number of zeros and all 171 pairs of runs*

5.2.4 Missing $t = 2$ Runs, $\nu = 2$

Table 5.2.7 gives the best values of $|\mathbf{M}_B|$ (D-optimality) for each of the OA$[18, 3^m]$ minus 2 runs designs, along with examples of designs that achieve these values.

Note that $m = 7$ is omitted from this table. This follows from our discussion of degrees of freedom for model fitting in Section 5.1.3. Since we are considering 18-run OAs with 2 runs removed, we have 16 degrees of freedom for model fitting. When $m = 7$, the number of degrees of freedom required for the intercept and main effects is $\alpha = 1 + 2 = 15$, hence we are left with only one degree of freedom for interaction terms. Thus, OA$[18, 3^7]$ minus 2 runs designs are not able to estimate any models containing more than $\nu = 1$ interaction.

Table 5.2.7: Optimal $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 2 runs designs

m	$\lvert\mathbf{M}_B\rvert$	No.†	Example		
			Class	Interactions	Runs
3	6.8875×10^{10}	20	1	(F_1, F_3),(F_2, F_3)	001 121
4	1.5062×10^{13}	120	20	(F_1, F_3),(F_3, F_4)	0010 1001
5	2.3994×10^{15}	72	38	(F_1, F_3),(F_3, F_4)	00101 11120
6	1.4577×10^{19}	1	323	(F_1, F_2),(F_4, F_5)	202110 220011

†*Number of occurrences of optimal* $|\mathbf{M}_B|$ *across all classes and all* $153 \times \binom{m}{2}$ *combinations of pair of runs and interaction*

Table 5.2.8 gives the best values of the average $|\mathbf{M}_B|$ (model-robustness) for each of the OA$[18, 3^m]$ minus 2 runs designs, along with examples of designs that achieve these values.

Table 5.2.8: Best average $|\mathbf{M}_B|$ for OA$[18, 3^m]$ minus 2 runs designs

m	No. Zeros†	Avg.$\lvert\mathbf{M}_B\rvert$	No.‡	Example	
				Class	Runs
3	0	6.4513×10^{10}	8	1	110 121
4	0	9.6892×10^{12}	4	1	0011 1101
5	0	9.6715×10^{14}	1	38	00101 11120
6	0	4.8316×10^{16}	1	64	112110 211121

†*Number of zero determinants amongst all* $\binom{m}{2}$ *interactions within each average*
‡*Number of occurrences of best Avg.*$|\mathbf{M}_B|$ *across all classes with given number of zeros and all 153 pairs of runs*

5.3 Model-Robustness Vs. Model-Discrimination

In addition to D-optimality, we have chosen to focus on model-robustness to assess the performance of designs when linear-by-linear interactions are included in the model. However, as noted in Chapter 4, a model-robust design does not necessarily have good model-discrimination properties. Jones et al. (2007) note that "for a given design, two competing models may be both estimable and yet difficult or impossible to discriminate in the model selection procedure". They give a range of optimality criteria for the specific purpose of assessing model-discrimination properties. These types of criteria are further investigated in Agboto et al. (2010).

Although a thorough examination of model-discrimination is beyond the scope of this thesis, it is nevertheless interesting to note that the two-stage model-robust procedure we have demonstrated is closely linked to well-established model-discrimination procedures. In this section we will show how our empirical methods for assessing model-robustness can be used to assess model-discrimination with the use of search designs.

The following example will aid further discussion about the distinction between model-robustness and model-discrimination.

Example 5.3.1. Consider the OA$[18, 3^6]$ in Table 5.3.1, and suppose we remove the third and twelfth runs, as indicated.

Table 5.3.1: OA$[18, 3^6]$ from the 85th geometric isomorphism class

	F_1	F_2	F_3	F_4	F_5	F_6
1	0	0	1	0	1	1
2	0	0	1	1	2	2
3	~~0~~	~~1~~	~~0~~	~~2~~	~~1~~	~~0~~
4	0	1	2	1	0	1
5	0	2	0	0	2	0
6	0	2	2	2	0	2
7	1	0	0	2	0	1
8	1	0	2	1	1	0
9	1	1	1	0	0	0
10	1	1	1	2	2	2
11	1	2	0	1	2	1
12	~~1~~	~~2~~	~~2~~	~~0~~	~~1~~	~~2~~
13	2	0	0	0	0	2
14	2	0	2	2	2	0
15	2	1	0	1	1	2
16	2	1	2	0	2	1
17	2	2	1	1	0	0
18	2	2	1	2	1	1

When $\nu = 1$, there are $\binom{6}{2} = 15$ single linear-by-linear interactions we could include in the model. All 15 of these models are estimable, hence this design is considered to be model-robust for $\nu = 1$. When $\nu = 2$, there are $\binom{15}{2} = 105$ pairs of linear-by-linear interactions we could include in the model. 57 of these models are not estimable, hence this design is not model-robust for $\nu = 2$.

We note that the design has no repeated runs, and that $\alpha = 1 + 2 \times 6 = 13$. Thus, we have $N - t - r - \alpha = 18 - 2 - 0 - 13 = 3$ degrees of freedom for interaction terms. Thus, in theory, we should be able to estimate models containing up to three linear-by-linear interactions. Indeed, this design is able to estimate 48 of the 105 models containing two linear-by-linear interactions, which confirms that the reason this design cannot estimate the remaining 57 models is not due to insufficient degrees of freedom. Hence, the non-estimability of some models must be due to the complicated aliasing structure of the design. □

The previous example illustrated the model-robustness of the design in Table 5.3.1 when $\nu = 1$, but gave no reference to the design's model-discrimination properties. We will show that this design's performance when $\nu = 2$ has implications for its model-discrimination properties when $\nu = 1$. In order to illustrate why this is the case, it is useful to revisit the definition of search designs.

Recall from Chapter 1 that a search design is defined as

$$E(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \tag{5.3.1}$$

where $\mathbf{Y}$ is a vector of n response variables, $\mathbf{X}_1$ $(n \times p_1)$ and $\mathbf{X}_2$ $(n \times p_2)$ are known model matrices, $\boldsymbol{\beta}_1$ is a vector of p_1 fixed unknown parameters (effects) of primary interest, and it is believed that at most g elements of $\boldsymbol{\beta}_2$ are non-zero but the identity of these is unknown.

Let $\boldsymbol{\beta}_1$ contain the intercept and all main effects, hence $\mathbf{X}_1 = \mathbf{X}_{ME}$ and $p_1 = \alpha$. Let $\boldsymbol{\beta}_2$ contain all possible linear-by-linear interactions. Then $\mathbf{X}_2 = \mathbf{Z}_{all}$, where $\mathbf{Z}_{all}$ indicates an interaction matrix containing all $\binom{m}{2}$ combinations of $\mathbf{z}_{(ab)}$, and $p_2 = \binom{m}{2}$. Then if we suspect ν linear-by-linear interactions are active, but we are not sure which interactions these might be, we are essentially dealing with a search design for $g = \nu$.

A design that is able to estimate the non-zero elements of $\boldsymbol{\beta}_2$ alongside the elements of $\boldsymbol{\beta}_1$ is called a search design with resolving power $\{\boldsymbol{\beta}_1; \boldsymbol{\beta}_2; g\}$. Morgan et al. (2014) give a necessary condition for discrimination of the correct model among the $\binom{p_2}{g}$ models as $\text{rank}\Big(\begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_{2i} & \mathbf{X}_{2j} \end{bmatrix}\Big) = p_1 + 2g$ for every $i, j = 1, \ldots, \binom{p_2}{g}$ such that $\boldsymbol{\beta}_{2i}$ and $\boldsymbol{\beta}_{2j}$ are disjoint $(g \times 1)$ subvectors of

$\boldsymbol{\beta}_2$. Hence, when we suspect ν linear-by-linear interactions are active, we require $\text{rank}\Big(\begin{bmatrix} \mathbf{X}_{ME} & \mathbf{Z}_{2i} & \mathbf{Z}_{2j} \end{bmatrix}\Big) = \alpha + 2\nu$ for every $i, j = 1, \ldots, \binom{\binom{m}{2}}{\nu}$. We note that this requirement will be satisfied when the information matrix associated with the partitioned model matrix $\begin{bmatrix} \mathbf{X}_{ME} & \mathbf{Z}_{2i} & \mathbf{Z}_{2j} \end{bmatrix}$ has a non-zero determinant for every $i, j = 1, \ldots, \binom{\binom{m}{2}}{\nu}$. The following example illustrates this.

Example 5.3.2. Consider the OA$[18, 3^6]$ minus 2 runs design given in Example 5.3.1. Note that $\alpha = 13$, hence $p_1 = 13$. Furthermore, $n = N - t = 18 - 2 = 16$, hence $\mathbf{X}_1 = \mathbf{X}_{ME}$ is a 16×13 matrix. There are $\binom{6}{2} = 15$ possible linear-by-linear interactions, hence $p_2 = 15$ and $\mathbf{X}_2 = \mathbf{Z}_{all}$ is the 16×15 matrix containing all $\mathbf{z}_{(ab)}$, $a, b \in (1, 2, \ldots, 6), a \neq b$.

Suppose we suspect $\nu = 1$ interaction will be active in the model but we are not sure which one. Then to meet the requirements of search designs and ensure that this design can discriminate between each of these 15 models, we must check that the rank of each of the $\binom{15}{2} = 105$ partitioned matrices $\begin{bmatrix} \mathbf{X}_{ME} & \mathbf{z}_{(12)} & \mathbf{z}_{(13)} \end{bmatrix}$, $\begin{bmatrix} \mathbf{X}_{ME} & \mathbf{z}_{(12)} & \mathbf{z}_{(23)} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{X}_{ME} & \mathbf{z}_{(46)} & \mathbf{z}_{(56)} \end{bmatrix}$ are all equal to $\alpha + 2\nu = 15$. Note that each of these partitioned matrices is equivalent to a model matrix for a model containing two interactions, that is, a model matrix when $\nu = 2$. We saw in Example 5.3.1 that 57 of the models with $\nu = 2$ are not estimable, which implies that the 57 associated model matrices are not of full rank. Thus, the OA$[18, 3^6]$ minus 2 runs design given in Example 5.3.1 does not meet the requirements for the discrimination of the correct model among the 15 models that contain a single interaction. □

The previous example illustrates how we can use the results of our empirical studies for models containing 2ν interactions to assess the model-discrimination properties for models containing ν interactions. The tables of results presented in Section 5.2 concern $\nu = 2$, hence we can use these to find designs with good model discrimination proprieties when $\nu = 1$. For example, we see from Table 5.2.8 that at least one OA$[18, 3^6]$ minus 2 runs design exists that achieves no zero determinants across all 105 models that include $\nu = 2$ linear-by-linear interactions. Hence, this design is able to discriminate between all 15 models that include $\nu = 1$ linear-by-linear interaction. We note, however, that Table 5.2.8 does not provide information on the number of designs that achieve these model-discriminating properties. Recall that the number of designs given in these tables (in the column labelled "No.") represent the number of designs that achieve the minimum number of zero determinants across all models *and* the best average $|\mathbf{M}_A|$ (or $|\mathbf{M}_B|$). Hence, although Table 5.2.8 tell us there is only one OA$[18, 3^6]$ minus 2 runs design that achieves no zero determinants *and* the best average $|\mathbf{M}_B|$, there may be more designs that

achieve no zero determinants with smaller average $|\mathbf{M}_B|$.

It is worth noting that our discussion in Section 5.1.3 regarding the bounds on ν has implication for how applicable search designs are for assessing the model-discrimination properties of designs as small as the ones we have been focusing on in this thesis. For example, suppose we would like to the assess the model-discrimination properties of the OA$[18, 3^6]$ minus 2 runs design in Example 5.3.1 for a model that contains $\nu = 2$ interactions. In order to do so, we require the results of an empirical study for models containing $2\nu = 4$ interactions. Recall, however, that OA$[18, 3^6]$ minus 2 runs designs have at most 3 degrees of freedom for fitting interaction terms. Hence, it will not be possible to fit any models containing 4 interactions terms, and so an assessment of the model-discrimination properties of an OA$[18, 3^6]$ minus 2 runs design is not possible when two linear-by-linear interactions are included in the model.

Search designs are also limited in their application when the number of possible interactions is small. For example, when $m = 3$ there are only $\binom{3}{2} = 3$ interactions that we could include in any given model. Hence, we cannot use search designs to assess the model-discrimination properties of models containing $\nu = 2$ interactions as any two pairs of linear-by-linear interactions must intersect and so are not disjoint.

Chapter 6

Concluding Remarks

In this chapter we will review the work in this thesis and suggest further avenues of research. In Section 6.1 we summarise the research and results within this project. Then in Section 6.2 we discuss other areas of research that are similar, and suggest how these branches might be tied together in future studies.

6.1 Review

The primary motivation of this thesis has been to examine the performance of OA plus p runs designs and OA minus t runs designs. We began in Chapter 1 by presenting some fundamental results pertaining to the statistical properties of orthogonal arrays. We also reviewed the literature regarding the construction of experimental designs, adjoined or missing runs, and the concept of design optimality in general.

In Chapter 2 we showed that under a main effects only model, optimal OA plus p runs designs (or OA minus t runs designs) depend only on the pairwise Hamming distance of the adjoined (or removed) runs. We presented an algorithm for finding optimal Hamming distances, and enumerated these values for all possible parameters of OAs with $N \leq 100$. We then provided general methods for constructing optimal sets of runs once the optimal pairwise Hamming distances have been identified. Finally, in Section 2.7 we addressed the limitation of OA minus t runs designs, that is, that the choice of runs to drop is restricted to sets of t runs within the OA. We showed that this restriction requires an enumeration of the combinatorial isomorphism classes for the OA parameters of interest, which are readily available in the literature, hence we examined the pairwise Hamming distances of all combinatorially non-isomorphic OA$[18, 3^m]$ designs. Most of the work in this chapter was published in Bird and Street (2016).

We then considered the implication of including linear-by-linear interactions in the model. An initial investigation into this idea quickly established that such an investigation would require a complete set of geometric isomorphism classes for our design space of interest, OA$[18, 3^m]$s. Geometric isomorphism is not as well studied as combinatorial isomorphism, and discrepancies exist in the literature as to the number of geometrically non-isomorphic OA$[18, 3^m]$s that are possible. Hence, we enumerated all geometrically non-isomorphic OA$[18, 3^m]$ in Chapter 3 and provide copies in the electronic appendix.

In Chapter 4, we considered the inclusion of a single linear-by-linear interaction in the model. We derived matrices to be optimised for OA plus p runs designs and OA minus t runs designs. We showed that the complicated nature of these matrices did not lend itself to general constructions analogous to those in Chapter 2, hence we conducted an empirical study on the OA$[18, 3^m]$ enumerated in Chapter 3. We gave examples of D-optimal and model-robust designs for each of the design spaces, that is, each of $m = 3$, 4, 5, 6 and 7.

We extended this work in Chapter 5 by considering models that contain ν linear-by-linear interactions terms. The results of an empirical study of the OA$[18, 3^m]$ plus p runs designs and OA$[18, 3^m]$ minus t runs designs were presented for ν. We then showed how our study of model-robustness relates to the concept of model-discrimination, particularly in the context of search designs. If we wish to find a design that can identify up to ν linear-by-linear interaction effects then we can look at the properties of the designs in estimating 2ν linear-by-linear interactions.

We will now consider other related areas of research. These ideas are not directly relevant to the work in this project, but we they provide interesting avenues of future research.

6.2 Possible Future Research

While the results in Chapter 2 are applicable to any OA of strength 2, the empirical studies in Chapters 4 and 5 focused on OA$[18, 3^m]$s. Thus, an obvious extension of this work could be to consider other families of OAs, for example OA$[27, 3^m]$s or asymmetric OAs. Of course any empirical study of this kind would require the complete enumeration of all geometrically non-isomorphic designs in the given design space, which could be formidably large.

A core underlying assumption of this research has been that all OAs are of strength 2. That is, all possible level-combinations between two factors in the complete factorial appear equally often in all pairs of columns in the OA. Hence, another

obvious extension to this work could be to consider the performance of OA plus p runs designs and OA minus t runs designs where the OAs have strength greater than 2. That is, all possible level-combinations between x factors in the complete factorial appear equally often in all x-tuples of columns of the OA, where $x > 2$. Interaction terms are independent of main effect terms when an OA has strength greater than 2, hence the issues we encountered in Chapters 4 and 5 regarding non-zero off-diagonal entries in $\mathbf{M}$ do not occur. This advantage comes at a cost in run-size however, as the number of runs required to satisfy the constraints of being of greater strength increases substantially.

Another extension that could be made is the inclusion of more interaction terms in the model. For example, we could consider the linear-by-quadratic, quadratic-by-linear and quadratic-by-quadratic terms.

Finally, we could view the concept of optimality through a different lens by considering different optimality criteria. We will now briefly summarise some of these possibilities.

6.2.1 Q_B Criterion

Tsai et al. (2000) introduce the Q-criterion which is the average of the A-criterion over all models. They use an approximation to the trace of $\mathbf{C}$ to save on calculation time and avoid the problem of singular matrices (which arise for non-full rank models). They continue this work in Tsai et al. (2004) and Tsai et al. (2006b), and then develop an extension, Q_B, in Tsai et al. (2007) which incorporates the experimenter's prior knowledge about the parameters in the model. Q_B is relatively 'advanced' compared to other alphabet criteria in that it encapsulates multiple objectives in a single value. Relationships can be drawn between Q_B and concepts such as factor screening and projectivity (discussed in Section 1.3.2) as well as model-robustness.

> The definition of the Q_B criterion originated from finding good three-level main-effects designs for quantitative factors. Depending on the prior knowledge of effects being non-negligible, different designs will be chosen among the class of main effects designs to ensure robustness to the two-factor interactions. However, Q_B can be applied to measure the performance of a design under any given maximal model and so can be used in different design problems under model uncertainty. For example, if the prior knowledge is such that large interactions are expected, then a broader class of designs than main-effects plans should

be considered (Tsai and Gilmour (2010))

Our study of model-robustness, in which we average the D-criterion over all models, is analogous to the work of these authors but in the context of D- rather than A-optimality. We note that our derivations of $\mathbf{\Omega}_A$, $\mathbf{\Omega}_B$, $\mathbf{\Psi}_A$ and $\mathbf{\Psi}_B$ are based on the determinant of the information matrix, whereas the A-criterion is concerned with the trace of the inverse of the information matrix. Hence, the methods here are not readily transferable from D- to A-optimality. However, a comparison of the two methods could be interesting.

Furthermore, the authors weight their criterion by incorporating prior knowledge of which models are most likely to be correct. This is similar to the concept of compound criteria, which we discuss in Section 6.2.3.

6.2.2 Pure Error

Gilmour and Trinca (2012) are highly critical of the traditional alphabet criteria, saying "the usual definitions of optimum design criteria no longer have all of the statistical interpretations claimed for them, since the sizes of confidence intervals and regions, or equivalently the power of hypothesis tests, depend not only on the variance matrix of the parameter estimates, but also on the degrees of freedom for pure error". They propose a modification of the main alphabet criteria to incorporate the degrees of freedom for pure error.

In place of D-optimality, which can be viewed as minimising $1/|\mathbf{X}'\mathbf{X}|$, they define $DP = (F_{\upsilon,d;1-\alpha})^{\upsilon}/|\mathbf{X}'\mathbf{X}|$, where υ is the number of parameters in the model and d is the number of pure error degrees of freedom (of course if $d = 0$ then DP is undefined).

D_s-optimality is a modification of D-optimality that is used throughout the literature in various forms. Gilmour and Trinca (2012) summarise these modifications as $|(\mathbf{M}^{-1})_{22}|$ for a given subset of υ_2 parameters (e.g. the higher order terms), where $\mathbf{M} = \mathbf{X}'\mathbf{X}$ and $(\mathbf{M}^{-1})_{22}$ is the portion of the inverse corresponding to the subset of the parameters of interest. They propose replacing this with $(DP)_S = (F_{\upsilon_2,d;1-\alpha})^{\upsilon_2}/|(\mathbf{M}^{-1})_{22}|$, and note that unlike the standard D_S which reduces to D in some situations, $(DP)_S$ does not reduce to DP due to the reduction in the numerator degrees of freedom.

Although we have not assessed the performance of designs in relation to pure error directly in this thesis, the OA$[18, 3^m]$ plus p runs designs we assessed in our empirical studies in Chapters 4 and 5 were implicitly selected in order to maximise their ability to estimate pure error. Recall from Section 4.2.4 that the size of the

designs space for all OA$[18, 3^m]$ plus p runs designs where the adjoined runs can be any set of p from the complete factorial is prohibitively large. This compelled us to limit the scope of this space, which we did by limiting the p runs to be adjoined to those within the OA. One of our arguments for doing so is that this guarantees at least p degrees of freedom for pure error.

Of course, as discussed in Section 5.1.3, there is a trade-off between maximising the degrees of freedom for pure error and maximising the degrees of freedom for model fitting. Hence, an interesting future study could be assess the model-robustness of designs that perform well under the criteria proposed by Gilmour and Trinca (2012).

6.2.3 Compound Criteria

A common theme throughout the design literature is that multiple objectives often compete with one another and a balance needs to be struck between them. Robinson and Anderson-Cook (2010) discuss this issue, paying special attention the D-criterion, and suggest a number of other criteria that should be considered simultaneously. Lu et al. (2011) and Lu and Anderson-Cook (2012) address the problem of competing objectives with Pareto frontier methods that produce a suite of 'good' designs, rather than attempting to single out the 'best' one. As discussed above, Gilmour and Trinca (2012), who are concerned with the estimation of pure error as well as the precision of parameter estimates, use a compound version of their modified alphabet criteria to allow the experimenters to give appropriate weight to the different components.

Recall that our assessment of model-robustness is a two-stage process in which we first identify the best realisable rate of model-estimability, then we assess the average D-criterion for the subset of designs that achieve this rate. Hence, this criteria can, in a sense, be considered as a version of compound criteria. Another interesting avenue of research could be to assess the performance of OA plus p runs designs and OA minus t runs designs with more comprehensive compound criteria that incorporate concepts such a pure error, A-optimality, and the experimenter's prior knowledge of which interactions might be active in the design.

Appendix

Chapter 2

Tables A1, A2, A3 and A4 give the optimal (d_1, d_2) for $(s_1, s_2) = (2, 3)$, $(2, 5)$, $(3, 4)$ and $(3, 5)$ respectively. As Chapter 2 considers designs with at most N=100 runs, only one of m_1 and m_2 can be greater than 1 for $(s_1, s_2) = (3, 4)$ and $(3, 5)$ (since, for example, 3^2, 3×4 and 4^2 would all need to divide N otherwise, and hence the smallest N would be 144).

Table A1: Optimal (d_1, d_2) for $(s_1, s_2) = (2, 3)$

m_1 \ m_2	1	2	3	4	5	6	7	8
1	$(\mathbf{0}, \mathbf{1})$ $(\mathbf{1}, \mathbf{1})$	$(0, 2)$	$(1, 2)$	$(\mathbf{0}, \mathbf{3})$ $(\mathbf{1}, \mathbf{3})$	$(0, 4)$	$(1, 4)$	$(\mathbf{0}, \mathbf{5})$ $(\mathbf{1}, \mathbf{5})$	$(0, 6)$
2	$(1, 1)$	$(2, 1)$	$(0, 3)$	$(1, 3)$	$(2, 3)$	$(0, 5)$	$(1, 5)$	$(2, 5)$
3	$(3, 0)$	$(1, 2)$	$(2, 2)$	$(0, 4)$ $(3, 2)$	$(1, 4)$	$(2, 4)$	$(0, 6)$ $(3, 4)$	$(1, 6)$
4	$(2, 1)$	$(3, 1)$	$(1, 3)$ $(4, 1)$	$(2, 3)$	$(0, 5)$ $(3, 3)$	$(1, 5)$ $(4, 3)$	$(2, 5)$	$(0, 7)$ $(3, 5)$
5	$(4, 0)$	$(2, 2)$ $(5, 0)$	$(3, 2)$	$(1, 4)$ $(4, 2)$	$(2, 4)$ $(5, 2)$	$(0, 6)$ $(3, 4)$	$(1, 6)$ $(4, 4)$	$(2, 6)$ $(5, 4)$
6	$(3, 1)$	$(4, 1)$	$(2, 3)$ $(5, 1)$	$(3, 3)$ $(6, 1)$	$(1, 5)$ $(4, 3)$	$(2, 5)$ $(5, 3)$	$(0, 7)$ $(3, 5)$ $(6, 3)$	$(1, 7)$ $(4, 5)$
7	$(5, 0)$	$(3, 2)$ $(6, 0)$	$(4, 2)$ $(7, 0)$	$(2, 4)$ $(5, 2)$	$(3, 4)$ $(6, 2)$	$(1, 6)$ $(4, 4)$ $(7, 2)$	$(2, 6)$ $(5, 4)$	$(0, 8)$ $(3, 6)$ $(6, 4)$
8	$(4, 1)$	$(5, 1)$	$(3, 3)$ $(6, 1)$	$(4, 3)$ $(7, 1)$	$(2, 5)$ $(5, 3)$ $(8, 1)$	$(3, 5)$ $(6, 3)$	$(1, 7)$ $(4, 5)$ $(7, 3)$	$(2, 7)$ $(5, 5)$ $(8, 3)$

*When $(m_1, m_2) = (1, 1)$, $(1, 4)$ or $(1, 7)$, as indicated in **bold**, optimal $|q| = 1$. Otherwise, optimal $q = 0$.*

Table A2: Optimal (d_1, d_2) for $(s_1, s_2) = (2, 5)$

m_1 \ m_2	1	2	3	4	5	6	7	8
1	$\mathbf{(0,1)}$ $\mathbf{(1,1)}$	$(0,2)$	$\mathbf{(0,3)}$	$\mathbf{(1,3)}$	$(1,4)$	$\mathbf{(0,5)}$ $\mathbf{(1,5)}$	$(0,6)$	$\mathbf{(0,7)}$
2	$(1,1)$	$\mathbf{(0,2)}$ $\mathbf{(1,2)}$	$(0,3)$	$(2,3)$	$\mathbf{(1,4)}$ $\mathbf{(2,4)}$	$(1,5)$	$\mathbf{(0,6)}$ $\mathbf{(1,6)}$	$(0,7)$
3	$\mathbf{(1,1)}$ $\mathbf{(2,1)}$	$(1,2)$	$(3,2)$	$(0,4)$	$(2,4)$	$\mathbf{(1,5)}$ $\mathbf{(2,5)}$	$(1,6)$	$(3,6)$
4	$(2,1)$	$(4,1)$	$(1,3)$	$(3,3)$	$(0,5)$	$(2,5)$	$(4,5)$	$(1,7)$
5	$(5,0)$	$(2,2)$	$(4,2)$	$(1,4)$	$(3,4)$	$(0,6)$ $(5,4)$	$(2,6)$	$(4,6)$
6	$(3,1)$	$(5,1)$	$(2,3)$	$(4,3)$	$(1,5)$ $(6,3)$	$(3,5)$	$(0,7)$ $(5,5)$	$(2,7)$
7	$(6,0)$	$(3,2)$	$(5,2)$	$(2,4)$ $(7,2)$	$(4,4)$	$(1,6)$ $(6,4)$	$(3,6)$	$(0,8)$ $(5,6)$
8	$(4,1)$	$(6,1)$	$(3,3)$ $(8,1)$	$(5,3)$	$(2,5)$ $(7,3)$	$(4,5)$	$(1,7)$ $(6,5)$	$(3,7)$ $(8,5)$

When $(m_1, m_2) = (1,1)$, $(1,3)$, $(1,4)$, $(1,6)$, $(1,8)$, $(2,2)$, $(2,5)$, $(2,7)$, $(3,1)$ *or* $(3,6)$, *as indicated in* ***bold*** *optimal* $|q| = 1$. *Otherwise, optimal* $q = 0$.

Table A3: Optimal (d_1, d_2) for $(s_1, s_2) = (3, 4)$

(a) $m_1 = 1$

m_2	1	2	3	4	5	6	7	8
	$\mathbf{(1,1)}$	$\mathbf{(0,2)}$	$(0,3)$	$(1,3)$	$\mathbf{(1,4)}$	$\mathbf{(0,5)}$	$(0,6)$	$(1,6)$

When $(m_1, m_2) = (1,1)$, $(1,2)$, $(1,5)$ *or* $(1,6)$, *as indicated in* ***bold***, *optimal* $|q| = 1$. *Otherwise, optimal* $q = 0$.

(b) $m_2 = 1$

m_1	1	2	3	4	5	6	7	8
	$\mathbf{(1,1)}$	$\mathbf{(1,1)}$	$(2,1)$	$(4,0)$	$\mathbf{(3,1)}$ $\mathbf{(5,0)}$	$(4,1)$	$(6,0)$	$\mathbf{(5,1)}$ $\mathbf{(7,0)}$

When $(m_1, m_2) = (1,1)$, $(2,1)$, $(5,1)$ *or* $(8,1)$, *as indicated in* ***bold***, *optimal* $|q| = 1$. *Otherwise, optimal* $q = 0$.

Table A4: Optimal (d_1, d_2) for $(s_1, s_2) = (3, 5)$

(a) $m_1 = 1$

m_2	1	2	3	4	5	6	7	8
	$\mathbf{(1,1)}$	$\mathbf{(0,2)}$	$(0,3)$	$\mathbf{(0,4)}$ $\mathbf{(1,3)}$	$(1,4)$	$\mathbf{(1,5)}$	$\mathbf{(0,6)}$	$(0,7)$

When $(m_1, m_2) = (1,1)$, $(1,2)$, $(1,4)$, $(1,6)$ *or* $(1,7)$, *as indicated in* ***bold***, *optimal* $|q| = 1$. *Otherwise, optimal* $q = 0$.

(b) $m_2 = 1$

m_1	1	2	3	4	5	6	7	8
	$\mathbf{(1,1)}$	$\mathbf{(1,1)}$	$(2,1)$	$\mathbf{(3,1)}$ $\mathbf{(4,0)}$	$(5,0)$	$(4,1)$	$\mathbf{(5,1)}$ $\mathbf{(6,0)}$	$(7,0)$

When $(m_1, m_2) = (1,1)$, $(2,1)$, $(4,1)$ *or* $(7,1)$, *as indicated in* ***bold***, *optimal* $|q| = 1$. *Otherwise, optimal* $q = 0$.

Chapter 3

Table A5: All possible $\mathbf{c}_3$

Column	Class	$\mathbf{c}_3'$																	
1	1	0	1	1	2	0	2	2	2	0	0	1	1	0	1	1	2	0	2	
2	1	0	2	0	2	1	1	1	1	0	2	0	2	0	2	1	1	0	2	*
3	1	0	2	1	1	0	2	0	1	2	2	0	1	1	2	0	0	1	2	
4	1	0	2	1	1	0	2	0	2	0	2	1	1	1	1	0	2	0	2	*
5	1	0	2	1	1	0	2	1	1	0	2	0	2	0	2	0	2	1	1	*
6	1	0	2	1	2	0	1	1	1	0	0	2	2	0	2	1	2	0	1	
7	1	1	1	0	2	0	2	0	2	0	2	1	1	0	2	1	1	0	2	*
8	1	1	2	0	0	1	2	0	1	2	2	0	1	0	2	1	1	0	2	
9	2	0	1	0	2	1	2	1	2	1	2	0	0	0	2	0	1	1	2	
10	2	0	1	1	2	0	2	0	2	1	2	0	1	1	2	0	0	1	2	
11	2	0	1	1	2	0	2	1	2	0	1	0	2	0	2	0	2	1	1	
12	2	0	1	1	2	0	2	2	2	0	1	0	1	0	1	0	2	1	2	
13	2	0	1	2	2	0	1	1	2	0	1	0	2	0	2	0	1	1	2	
14	2	0	2	0	1	1	2	0	2	1	2	0	1	1	1	0	2	0	2	
15	2	0	2	0	1	1	2	1	2	0	1	0	2	0	1	2	2	0	1	
16	2	0	2	0	1	1	2	1	2	1	2	0	0	0	1	0	2	1	2	
17	2	0	2	0	2	1	1	0	1	1	2	0	2	1	2	0	1	0	2	
18	2	1	1	0	2	0	2	0	2	1	2	0	1	0	2	0	1	1	2	
19	2	1	2	0	0	1	2	0	2	1	2	0	1	0	1	1	2	0	2	
20	2	1	2	0	2	0	1	0	0	1	2	1	2	1	2	0	1	0	2	
21	3	0	0	2	2	1	1	1	2	0	1	0	2	1	2	0	1	0	2	
22	3	0	1	0	1	2	2	1	2	1	2	0	0	0	2	0	2	1	1	
23	3	0	1	0	1	2	2	2	2	0	1	0	1	0	1	2	2	0	1	
24	3	0	1	2	2	0	1	2	2	0	1	0	1	0	1	0	1	2	2	
25	3	0	2	0	2	1	1	1	2	1	2	0	0	0	1	0	1	2	2	
26	3	0	2	1	2	0	1	0	2	1	2	0	1	1	1	0	0	2	2	
27	3	1	1	0	2	0	2	2	2	0	1	0	1	0	0	1	2	1	2	
28	3	1	1	2	2	0	0	0	2	0	1	1	2	0	2	0	1	1	2	
29	3	1	2	0	0	1	2	1	2	1	2	0	0	0	0	1	2	1	2	
30	3	1	2	0	1	0	2	1	2	0	1	0	2	0	0	2	2	1	1	
31	3	2	2	0	1	0	1	0	0	1	2	1	2	1	1	0	2	0	2	
32	3	2	2	0	1	0	1	0	1	0	1	2	2	0	1	2	2	0	1	

continued...

... Table A5 continued

Column	Class	$\mathbf{c}'_3$																	
33	4	0	1	2	2	0	1	0	2	1	1	0	2	1	2	0	0	1	2	
34	4	0	2	0	2	1	1	0	2	1	1	0	2	1	1	0	2	0	2	*
35	4	1	1	0	2	0	2	0	2	1	1	0	2	0	2	0	2	1	1	*
36	4	1	2	0	2	0	1	0	0	1	1	2	2	1	2	0	2	0	1	
37	5	0	0	1	2	1	2	1	2	0	2	0	1	1	2	0	1	0	2	
38	5	0	1	1	2	0	2	0	1	0	2	1	2	2	2	0	1	0	1	
39	5	0	2	1	2	0	1	1	2	0	2	0	1	0	1	0	1	2	2	
40	5	1	2	1	2	0	0	0	1	0	2	1	2	0	2	0	1	1	2	
41	6	0	0	1	2	1	2	1	2	0	0	1	2	1	2	1	2	0	0	
42	6	0	1	0	1	2	2	0	1	2	2	0	1	2	2	0	1	0	1	
43	6	0	1	0	1	2	2	0	2	0	2	1	1	1	2	1	2	0	0	
44	6	0	1	0	2	1	2	0	1	0	2	1	2	2	2	1	1	0	0	
45	6	2	2	0	1	0	1	1	1	0	2	0	2	0	0	1	2	1	2	
46	6	2	2	1	1	0	0	0	1	0	2	1	2	0	1	0	2	1	2	
47	7	0	1	0	2	1	2	0	2	0	1	1	2	1	2	1	2	0	0	
48	7	1	2	0	2	0	1	1	2	0	1	0	2	0	0	1	2	1	2	
49	7	1	2	1	2	0	0	0	2	0	1	1	2	0	1	0	2	1	2	
50	7	2	2	0	1	0	1	0	1	1	2	0	2	0	1	0	2	1	2	
51	8	0	1	0	2	1	2	0	2	1	2	0	1	1	2	0	1	0	2	
52	8	0	1	1	2	0	2	0	2	0	1	1	2	1	2	0	2	0	1	
53	8	0	2	0	1	1	2	0	1	1	2	0	2	1	2	0	2	0	1	
54	8	1	2	0	2	0	1	0	1	1	2	0	2	0	2	0	1	1	2	
55	9	0	1	1	2	0	2	0	2	0	2	1	1	1	2	0	1	0	2	
56	9	0	1	1	2	0	2	1	2	0	0	1	2	0	2	1	2	0	1	
57	9	0	2	0	1	1	2	0	1	2	2	0	1	1	2	0	1	0	2	
58	9	0	2	1	1	0	2	1	2	0	2	0	1	0	1	0	2	1	2	
59	9	0	2	1	2	0	1	1	1	0	2	0	2	0	2	0	1	1	2	
60	9	1	2	0	2	0	1	0	1	0	2	1	2	0	2	1	1	0	2	
61	10	1	2	0	2	0	1	0	2	1	1	0	2	0	1	0	2	1	2	
62	11	0	2	1	2	0	1	0	1	0	2	1	2	1	2	0	1	0	2	
63	11	1	2	0	1	0	2	0	1	0	2	1	2	0	2	1	2	0	1	
64	12	0	0	1	1	2	2	1	1	2	2	0	0	2	2	0	0	1	1	
65	12	0	0	2	2	1	1	1	1	0	0	2	2	2	2	1	1	0	0	
66	12	1	1	2	2	0	0	2	2	0	0	1	1	0	0	1	1	2	2	
67	12	2	2	1	1	0	0	0	0	2	2	1	1	1	1	0	0	2	2	
68	13	0	0	2	2	1	1	2	2	1	1	0	0	1	1	0	0	2	2	
69	13	1	1	0	0	2	2	2	2	1	1	0	0	0	0	2	2	1	1	

*Columns marked with * remain unchanged after the levels have been reversed*

Electronic Appendix

This thesis is accompanied by five electronic files, as detailed below.

Table E1: Geometric isomorphism classes for OA$[18, 3^3]$s

One representative from each of the 13 geometric isomorphism class for OA$[18, 3^3]$s
File name: *"E1. OA geometric isomorphism classes. N=18, s=3, m=3.csv"*

Table E2: Geometric isomorphism classes for OA$[18, 3^4]$s

One representative from each of the 137 geometric isomorphism class for OA$[18, 3^4]$s
File name: *"E2. OA geometric isomorphism classes. N=18, s=3, m=4.csv"*

Table E3: Geometric isomorphism classes for OA$[18, 3^5]$s

One representative from each of the 333 geometric isomorphism class for OA$[18, 3^5]$s
File name: *"E3. OA geometric isomorphism classes. N=18, s=3, m=5.csv"*

Table E4: Geometric isomorphism classes for OA$[18, 3^6]$s

One representative from each of the 485 geometric isomorphism class for OA$[18, 3^6]$s
File name: *"E4. OA geometric isomorphism classes. N=18, s=3, m=6.csv"*

Table E5: Geometric isomorphism classes for OA$[18, 3^7]$s

One representative from each of the 291 geometric isomorphism class for OA$[18, 3^7]$s
File name: *"E5. OA geometric isomorphism classes. N=18, s=3, m=7.csv"*

Bibliography

Agboto, V., Li, W., and Nachtsheim, C. (2010). Screening designs for model discrimination. *Journal of Statistical Planning and Inference*, 140(3):766–780.

Ahmad, T. and Gilmour, S. (2010). Robustness of subset response surface designs to missing observations. *Journal of Statistical Planning and Inference*, 140(1):92–103.

Akhtar, M. and Prescott, P. (1986). Response surface designs robust to missing observations. *Communications in Statistics-Simulation and Computation*, 15(2):345–363.

Anderson, D. A. and Thomas, A. M. (1980). Weakly resolvable IV. 3 search designs for the p^n factorial experiment. *Journal of Statistical Planning and Inference*, 4(3):299–312.

Angelopoulos, P., Chatterjee, K., and Koukouvinos, C. (2013). Construction of search designs from orthogonal arrays. *Journal of Statistical Theory and Practice*, 7(4):774–782.

Angelopoulos, P., Evangelaras, H., Koukouvinos, C., and Lappas, E. (2007). An effective step-down algorithm for the construction and the identification of non-isomorphic orthogonal arrays. *Metrika*, 66(2):139–149.

Arshad, H. M. and Gilmour, S. G. (2012). The performance of subset response surface designs for estimating third order terms. *Journal of Statistical Planning and Inference*, 142(1):232–247.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Bird, E. M. and Street, D. J. (2016). D-optimal asymmetric orthogonal array plus p run designs. *Journal of Statistical Planning and Inference*, 170:64–76.

Box, G. and Tyssedal, J. (1996). Projective properties of certain orthogonal arrays. *Biometrika*, 83(4):950–955.

Box, G. E. and Draper, N. R. (1975). Robust designs. *Biometrika*, 62(2):347–352.

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, 40(3/4):318–335.

Butler, N. A. and Ramos, V. M. (2007). Optimal additions to and deletions from two-level binary arrays. *Journal of the Royal Statistical Society, Series B*, 69(1):51–61.

Chai, F. S., Mukerjee, R., and Suen, C.-Y. (2002). Further results on orthogonal array plus one run plans. *Journal of Statistical Planning and Inference*, 106(1):287–301.

Chatterjee, K. and Mukerjee, R. (1986). Some search designs for symmetric and asymmetric factorials. *Journal of Statistical Planning and Inference*, 13:357–363.

Chatzopoulos, S. A., Kolyva-Machera, F., and Chatterjee, K. (2011). Optimality results on orthogonal arrays plus p runs for s^m factorial experiments. *Metrika*, 73(3):385–394.

Cheng, C.-S. (1998). Some hidden projection properties of orthogonal arrays with strength three. *Biometrika*, 85(2):491–495.

Cheng, C.-S. (2006). Projection properties of factorial designs for factor screening. In Dean, A. and Lewis, S., editors, *Screening. Methods for Experimentation in Industry, Drug Discovery, and Genetics*, pages 156–168. Springer, New York.

Cheng, C.-S. (2013). *Theory of Factorial Design: Single-and Multi-Stratum Experiments*. CRC Press, New York.

Cheng, C.-S., Steinberg, D. M., and Sun, D. X. (1999). Minimum aberration and model robustness for two-level fractional factorial designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):85–93.

Cheng, S.-W. and Wu, C. (2001). Factor screening and response surface exploration. *Statistica Sinica*, 11(3):553–579.

Cheng, S.-W. and Ye, K. Q. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *Annals of Statistics*, 32(5):2168–2185.

Clark, J. B. and Dean, A. (2001). Equivalence of fractional factorial designs. *Statistica Sinica*, 11(2):537–548.

Fries, A. and Hunter, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics*, 22(4):601–608.

Ghosh, S. (1982). Robustness of designs against the unavailability of data. *Sankhya Series B*, 44(1):50–62.

Ghosh, S. and Burns, C. (2002). Theory & methods: Comparison of four new general classes of search designs. *Australian & New Zealand Journal of Statistics*, 44(3):357–366.

Ghosh, S. and Flores, A. (2013). Common variance fractional factorial designs and their optimality to identify a class of models. *Journal of Statistical Planning and Inference*, 143(10):1807–1815.

Gilmour, S. G. and Trinca, L. A. (2012). Optimum design of experiments for statistical inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):345–401.

Hamada, M. and Wu, C. J. (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, 24(3):130–137.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer-Verlag, New York.

Hedayat, A. and Zhu, H. (2003). Adding more observations to saturated D-optimal resolution III two-level factorial designs. *Journal of Combinatorial Designs*, 11(1):51–77.

Herzberg, A. (1982). The robust design of experiments: A review. *Serdica*, 8:223–228.

Jones, B. A., Li, W., Nachtsheim, C. J., and Ye, K. Q. (2007). Model discrimination - another perspective on model-robust designs. *Journal of Statistical Planning and Inference*, 137(5):1576–1583.

Kaski, P. and Östergård, P. R. (2006). *Classification Algorithms for Codes and Designs*. Springer-Verlag, Berlin Heidelberg.

Katsaounis, T., Dingus, C., and Dean, A. (2007). On the geometric equivalence and non-equivalence of symmetric factorial designs. *Journal of Statistical Theory and Practice*, 1(1):101–115.

Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis*. Duxbury/Thomson Learning, US, 2nd edition.

Kuhfeld, W. F. (2006). Orthogonal arrays. Technical report, SAS Institute. Available at http://support.sas.com/techsup/technote/ts723.html.

Lekivetz, R. and Tang, B. (2014). Multi-level orthogonal arrays for estimating main effects and specified interactions. *Journal of Statistical Planning and Inference*, 144:123–132.

Li, W. and Nachtsheim, C. J. (2000). Model-robust factorial designs. *Technometrics*, 42(4):345–352.

Lin, D. K. and Draper, N. R. (1992). Projection properties of Plackett and Burman designs. *Technometrics*, 34(4):423–428.

Lin, D. K. and Draper, N. R. (1993). Generating alias relationships for two-level Plackett and Burman designs. *Computational statistics & data analysis*, 15(2):147–157.

Lu, L. and Anderson-Cook, C. M. (2012). Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature. *Quality Engineering*, 24(3):404–422.

Lu, L., Anderson-Cook, C. M., and Robinson, T. J. (2011). Optimization of designed experiments based on multiple criteria utilizing a pareto frontier. *Technometrics*, 53(4):353–365.

MacEachern, S., Notz, W., Whittinghill, D., and Zhu, Y. (1995). Robustness to the unavailability of data in the linear model, with applications. *Journal of Statistical Planning and Inference*, 48(2):207–213.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman & Hall/CRC, Boca Raton, Florida, second edition.

McKay, B. D. and Piperno, A. (2014). Practical graph isomorphism, II. *Journal of Symbolic Computation*, 60(0):94 – 112.

Morgan, J. P., Ghosh, S., and Dean, A. (2014). J.N. Srivastava and experimental design. *Journal of Statistical Planning and Inference*, 144:3–18.

Mukerjee, R. (1999). On the optimality of orthogonal array plus one run plans. *Annals of Statistics*, 27(1):82–93.

Pang, F. and Liu, M.-Q. (2011). Geometric isomorphism check for symmetric factorial designs. *Journal of Complexity*, 27(5):441–448.

Robinson, T. J. and Anderson-Cook, C. M. (2010). A closer look at D-optimality for screening designs. *Quality Engineering*, 23(1):1–14.

Sarkar, A. and Chatterjee, K. (2010). An MEP.2 plan in 3^n factorial experiment and its probability of correct identification. *Journal of Statistical Planning and Inference*, 140(11):3531–3539.

Schoen, E. D. (2009). All orthogonal arrays with 18 runs. *Quality and Reliability Engineering International*, 25(4):467–480.

Schoen, E. D., Eendebak, P. T., and Nguyen, M. V. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *Journal of Combinatorial Designs*, 18(2):123–140.

Schoen, E. D. and Nguyen, M. V. (2007). *Enumeration and Classification of Orthogonal Arrays*. University of Antwerp, Faculty of Applied Economics.

Srivastava, J. (1975). Designs for searching non-negligible effects. In Srivastava, J., editor, *A Survey of Statistical Design and Linear Models*, pages 507–519. Amsterdam: North-Holland.

Tanco, M., del Castillo, E., and Viles, E. (2013). Robustness of three-level response surface designs against missing data. *IIE Transactions*, 45(5):544–553.

Tang, B. and Deng, L.-Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Annals of Statistics*, 27(6):1914–1926.

Tsai, K., Ye, K., and Li, W. (2006a). A complete catalog of geometrically non-isomorphic 18-run orthogonal arrays. *Preprint. http://202.113.29.3/ stat/doe.files/CP/Ye,Kenny.pdf.*

Tsai, P. W. and Gilmour, S. G. (2010). A general criterion for factorial designs under model uncertainty. *Technometrics*, 52(2):231–242.

Tsai, P. W., Gilmour, S. G., and Mead, R. (2000). Projective three-level main effects designs robust to model uncertainty. *Biometrika*, 87(2):467–475.

Tsai, P. W., Gilmour, S. G., and Mead, R. (2004). Some new three-level orthogonal main effects plans robust to model uncertainty. *Statistica Sinica*, 14(4):1075–1084.

Tsai, P. W., Gilmour, S. G., and Mead, R. (2006b). Statistical isomorphism of three-level fractional factorial designs. *Utilitas Mathematica*, 70:3–9.

Tsai, P. W., Gilmour, S. G., and Mead, R. (2007). Three-level main-effects designs exploiting prior information about model uncertainty. *Journal of Statistical Planning and Inference*, 137(2):619–627.

Tsai, S.-F. and Liao, C.-T. (2011). Optimal partially replicated parallel-flats designs for $2^{n_1} \times 3^{n_2}$ mixed factorial experiments. *Journal of Statistical Planning and Inference*, 141(8):2803–2810.

Tsai, S.-F. and Liao, C.-T. (2014). Selection of partial replication on two-level orthogonal arrays. *Canadian Journal of Statistics*, 42(1):168–183.

Tsai, S.-F., Liao, C.-T., and Chai, F.-S. (2012). *D*-optimal partially replicated two-level factorial designs. *Statistica Sinica*, 22(1):419–432.

Wang, J. and Wu, C. J. (1995). A hidden projection property of Plackett-Burman and related designs. *Statistica Sinica*, 5(1):235–250.

Wu, C. J. and Hamada, M. S. (2011). *Experiments: Planning, Analysis, and Optimization*. John Wiley & Sons, US, 2nd edition.

Xu, H., Cheng, S.-W., and Wu, C. (2004). Optimal projective three-level designs for factor screening and interaction detection. *Technometrics*, 46(3).

Xu, H. and Wu, C. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Annals of Statistics*, 29(4):549–560.