OPTIMAL STATISTICAL DESIGN
FOR VARIANCE COMPONENTS IN
MULTISTAGE VARIABILITY
MODELS

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2014

By
Sergio Ivan Loeza-Serrano
School of Mathematics
## Contents

Abstract .................................................. 9

Declaration ............................................... 10

Copyright Statement ................................... 11

Acknowledgements ........................................ 12

Dedication ................................................ 13

Publications .............................................. 14

1 Introduction ............................................. 15
   1.1 Background ........................................ 17
   1.2 Experimental designs for variance components 20
   1.3 Structure of the thesis ............................. 26

2 Preliminaries .......................................... 28
   2.1 Design of experiments ............................. 28
      2.1.1 Basic ideas and terminology .................. 28
      2.1.2 Fundamental considerations ................... 35
2.2 Statistical models ........................................ 39
  2.2.1 Linear fixed models ............................... 40
  2.2.2 Linear mixed models ............................. 46
  2.2.3 Nonlinear models ................................. 49
2.3 Design optimality theory ............................. 51
2.4 Optimality criteria and the general equivalence theorem ........ 57
  2.4.1 $D$-optimality .................................. 57
  2.4.2 $G$-optimality .................................. 59
  2.4.3 The general equivalence theorem .................. 59
  2.4.4 $D_A$-optimality (Generalised $D$-optimality) ...... 62
  2.4.5 $D_S$-optimality ................................ 63
  2.4.6 $c$-optimality .................................. 64
  2.4.7 $V$-optimality .................................. 64
2.5 Summary ............................................. 65

3 Parameter estimation .................................. 67
  3.1 Model fitting ....................................... 67
  3.2 Fixed effects models ............................... 68
    3.2.1 Analysis of variance (ANOVA) ................. 69
    3.2.2 Least squares estimation ...................... 73
  3.3 Mixed effects models .............................. 75
    3.3.1 Maximum likelihood estimation ............... 76
    3.3.2 MLE information matrix ....................... 79
    3.3.3 Restricted maximum likelihood estimation .... 82
    3.3.4 REML information matrix ..................... 84
    3.3.5 Bayesian framework ............................ 85
  3.4 Further results for variance components ............ 87
    3.4.1 Information matrix for ratios of the variance components 87
3.4.2 Changes in total variance ........................................... 90
3.5 Dispersion-Mean model for the
estimation of variance components ............................... 94
3.6 Summary ............................................................... 100

4 Design structures .................................................. 102
  4.1 Factors and experimental units .............................. 103
     4.1.1 Factors ..................................................... 103
     4.1.2 Experimental units ...................................... 104
  4.2 Design, treatment and variability structures ............ 105
  4.3 Primary structures ............................................. 107
     4.3.1 $m$-factor layout ........................................ 107
     4.3.2 Crossed layout .......................................... 108
     4.3.3 Nested layout .......................................... 109
     4.3.4 Model formulation ...................................... 111
  4.4 Split units and restricted randomisation ................. 112
  4.5 Useful structures ............................................. 115
     4.5.1 Split-Plot structures ................................. 115
     4.5.2 Split-Block design ..................................... 117
  4.6 Summary of structures ........................................ 119
  4.7 Comparison with alternative descriptions of design structures ........................................ 120
     4.7.1 Split-Plot design ....................................... 122
     4.7.2 Split-Split-Plot design ............................... 124
     4.7.3 Split-Block design ..................................... 126
  4.8 Summary ............................................................... 128

5 Some information matrices for variance components .......... 130
5.1 Design optimality for $\sigma^2$ ............................... 131
  5.1.1 Review of $A$- and $D$-optimality ......................... 131
  5.1.2 $A_v$- and $D_v$-optimality ............................ 132
  5.1.3 Pseudo-Bayesian optimality .......................... 133
5.2 The $\{1\}$\{3-stage(2$\cdot$)$\}$ structure using $\sigma^2$ ............................... 134
  5.2.1 The $\{1\}$\{3-stage(2C)$\}$ structure ...................... 135
  5.2.2 The $\{1\}$\{3-stage(2Ci)$\}$ structure ...................... 136
  5.2.3 The $\{1\}$\{3-stage(2N)$\}$ structure ...................... 136
5.3 The $\{1\}$\{3-stage(2$\cdot$)$\}$ structure using $\eta$ ............................... 141
  5.3.1 The $\{1\}$\{3-stage(2C)$\}$ structure ...................... 141
  5.3.2 The $\{1\}$\{3-stage(2N)$\}$ structure ...................... 146
5.4 Summary ...................................................... 148

6 Design construction algorithms ........................ 149
  6.1 Introduction ............................................. 150
    6.1.1 The purpose of an algorithm ......................... 150
    6.1.2 Continuous designs ................................ 150
    6.1.3 Exact designs ......................................... 152
    6.1.4 Update formulae .................................... 153
  6.2 Algorithm for variance components models .......... 155
    6.2.1 Preliminaries ...................................... 155
    6.2.2 Optimum designs requiring $\sigma_0^2$ and $N$ .......... 156
    6.2.3 Optimum designs when $\sigma_0^2$ is not specified .... 159
    6.2.4 Computed designs for variance components .......... 164
  6.3 Designs for the $\{1\}$\{3-stage(2C)$\}$
    structure to estimate $\sigma^2$ .............................. 165
    6.3.1 $N=16$ and MLE .................................... 166
    6.3.2 $N=16$, MLE and pseudo-Bayesian optimality ......... 171
6.3.3 N=16 and REML ............................................ 174
6.3.4 N=16, REML and pseudo-Bayesian optimality .......... 177
6.3.5 N=24 and MLE vs REML ................................. 180
6.3.6 N=24, MLE vs REML and pseudo-Bayesian
   optimality ...................................................... 189
6.3.7 N=32 and MLE vs REML ................................. 194
6.3.8 N=32, MLE vs REML and pseudo-Bayesian
   optimality ...................................................... 199

6.4 Designs to estimate \( \sigma^2 \) for the \{1\}{3-stage(2C)} structure
   (interaction model) ............................................. 205
6.4.1 N=16 and MLE vs REML ................................. 205
6.4.2 N=16, MLE vs REML and pseudo-Bayesian
   optimality ...................................................... 210
6.4.3 N=24 and MLE vs REML ................................. 214
6.4.4 N=24, MLE vs REML and pseudo-Bayesian
   optimality ...................................................... 221
6.4.5 N=32 and MLE vs REML ................................. 225
6.4.6 N=32, MLE vs REML and pseudo-Bayesian
   optimality ...................................................... 231

6.5 Designs to estimate \( \eta \) in the
   \{1\}{3-stage(2C)} structure .................................... 235
6.5.1 N=16 and MLE vs REML ................................. 236
6.5.2 N=16, MLE vs REML and pseudo-Bayesian
   optimality ...................................................... 241
6.5.3 N=24 and MLE vs REML ................................. 244
6.5.4 N=24, MLE vs REML and pseudo-Bayesian
   optimality ...................................................... 251
6.5.5 \(N=32\) and MLE vs REML .......................... 254
6.5.6 \(N=32\), MLE vs REML and pseudo-Bayesian
optimality ............................................. 260

6.6 Designs to estimate \(\eta\) in the
\{1\}\{3-stage(2Ci)\} structure (with interaction) ............... 264
6.6.1 \(N=16\) and MLE vs REML .......................... 264
6.6.2 \(N=16\), MLE vs REML and pseudo-Bayesian
optimality ............................................. 269
6.6.3 \(N=24\) and MLE vs REML .......................... 272
6.6.4 \(N=24\), MLE vs REML and pseudo-Bayesian
optimality ............................................. 279
6.6.5 \(N=32\) and MLE vs REML .......................... 283
6.6.6 \(N=32\), MLE vs REML and pseudo-Bayesian
optimality ............................................. 286

6.7 The \{1\}\{3-stage(2N)\} structure to estimate \(\sigma^2\) ............... 291
6.7.1 \(N=16\) and MLE vs REML .......................... 292
6.7.2 \(N=16\), MLE vs REML and pseudo-Bayesian
optimality ............................................. 297
6.7.3 \(N=24\) and MLE vs REML .......................... 299
6.7.4 \(N=24\), MLE vs REML and pseudo-Bayesian
optimality ............................................. 304
6.7.5 \(N=32\) and MLE vs REML .......................... 306
6.7.6 \(N=32\), MLE vs REML and pseudo-Bayesian
optimality ............................................. 310

6.8 Optimum designs using the Dispersion-Mean model ............. 317
6.8.1 Ordinary least squares ................................ 318
6.8.2 Generalised least squares ........................... 329
6.9 Summary .................................................. 348

7 Discussion and conclusions .................................. 351
  7.1 Discussion .............................................. 351
    7.1.1 Unbalanced cases ................................ 351
  7.2 Conclusions ............................................ 357

A Mathematics results ......................................... 361
  A.1 Matrix Algebra ........................................ 361
  A.2 Statistics .............................................. 362
  A.3 vec operator and permutation matrix .................. 364

B Industrial internship ......................................... 366
  B.1 Introduction .......................................... 366
  B.2 Description of the project ............................ 367
    B.2.1 Justification ...................................... 367
    B.2.2 Objectives ......................................... 368
    B.2.3 Outcomes ........................................... 369
  B.3 Final outcome .......................................... 371
    B.3.1 Design with trend consideration for process optimisation 371
    B.3.2 Variance study ..................................... 372
  B.4 Conclusion ............................................. 379
  B.5 Summary ............................................... 381
This thesis focuses on the construction of optimum designs for the estimation of the variance components in multistage variability models. Variance components are the model parameters that represent the different sources of variability that affect the response of a system.

A general and highly detailed way to define the linear mixed effects model is proposed. The extension considers the explicit definition of all the elements needed to construct a model. One key aspect of this formulation is that the random part is stated as a functional that individually determines the form of the design matrices for each random regressor, which gives significant flexibility. Further, the model is strictly divided into the treatment structure and the variability structure. This allows separate definitions of each structure by using the single rationale of combining, with little restrictions, simple design arrangements called factor layouts.

To provide flexibility for considering different models, methodology to find and select optimum designs for variance components is presented using MLE and REML estimators and an alternative method known as the dispersion-mean model. Different forms of information matrices for variance components were obtained. This was mainly done for the cases when the information matrix is a function of the ratios of variances. Closed form expressions for balanced designs for random models with 3-stage variability structure, in crossed and nested layouts were found. The nested case was obtained when the information matrix is a function of the variance components. A general expression for the information matrix for the ratios using REML is presented. An approach to using unbalanced models, which requires the use of general formulae, is discussed. Additionally, $A$- and $D$- criteria of design optimality are restated for the case of variance components, and a specific version of pseudo-Bayesian criteria is introduced.

Algorithms to construct optimum designs for the variance components based on the aforementioned methodologies were defined. These algorithms have been implemented in the R language. The results are communicated using a simple, but highly informative, graphical approach not seen before in this context. The proposed plots convey enough details for the experimenter to make an informed decision about the design to use in practice.

An industrial internship allowed some of the results herein to be put into practice, although no new research outcomes originated. Nonetheless, this is evidence of the potential for applications. Equally valuable is the experience of providing statistical advice and reporting conclusions to a non statistical audience.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s Policy on Presentation of Theses.
Acknowledgements

My deepest and most sincere gratitude goes to all the hard working people of Mexico, whose efforts have allowed me to pursue this special goal. These efforts provide means to Mexico’s Consejo Nacional de Ciencia y Tecnología, CONACYT, to offer grants like the one I obtained.

I am also grateful to my supervisor Dr. Alex Donev for his advice, support and patience. Mr. Alastair Bissett deserves my appreciation for his support before, during and after the shorterKTP internship. Thanks to them, I could take advantage of unique opportunities.

Special thanks to the lovely ladies that have been next to me in this journey. Irene and Samara, I thank you both specially and infinitely for this, and in advance for what is yet to come.

I am deeply grateful to my family in Mexico, that I have left behind one way or another. I have always been unreservedly supported and I am and will always thank you all for that. I will do my best to give back as much as I can give.

To my most special friend, Hector, and his father whose help came at the most important time of my education, I can only sincerely say Thank you! I will always be in debt to you.
Dedication

a mi Madre

junto con todos mis logros de ayer, hoy y mañana.
Publications

Chapter 1

Introduction

This thesis is about statistical design of experiments. It focuses on the construction of optimum designs for the estimation of the variance components in mixed models arising from multi-stage variability structures. Three stage variability is mostly used throughout to exemplify the methodologies and tools introduced in the thesis. This is done using crossed, with and without interaction, and nested layouts for the random effects. The theoretical base allows straightforward extensions to more variability stages. However, specific development of computer code and use of more powerful numerical methods may be required to deal with more complex models. The methodology has been implemented using the R language.

The results presented in the thesis are mainly obtained under the following conditions:

1. knowledge about the real values of the variance components is non existent or very limited,

2. the total number of observations for the experiment is known.

The first condition requires the use of a fine grid to find an optimum design. The second condition allows for the generation of the largest possible list of suitable candidate designs. Relaxing the first condition implies that preliminary values for
the variance components are given and therefore, the search for an optimum design is not exhaustive. Similarly, if candidate designs are provided, the comparison is limited to those given. These last two situations are totally plausible, but are considered less general and thus results for them are not reported herein. However, the developed R code can work with options from the two scenarios.

Most of the developments focus on balanced designs. Advantages of such designs are explained. However, practical conditions often impede the execution of a fully balanced experiment. Therefore, a class of least-unbalanced designs is introduced. This is a class that, as much as possible, tries to keep the balance of the candidate designs for a given variability structure and number of observations.

The optimum designs for variance components are computed based on the information matrices obtained from MLE and REML estimators. Additionally, a method known as the dispersion-mean (DM) model is presented to compare designs found using different methods of estimation. Moreover, the DM model is used to find optimum designs assuming the data are non normal by including a kurtosis parameter. The selection of an optimum design is aided by the use of graphical output from a purpose built algorithm. These consist of triangular, rank ordered and robustness plots. The latter were developed specifically for the thesis and the use of triangular plots in this context is being pioneered here.

This is done within a framework that allows the definition of the design structure in conjunction with the formulation of the model for analysis. The framework presented here tries to facilitate the understanding of the experimental workflow to the practitioner rather than the statistician. Therefore, not everyone will find it relevant, but based on my industrial experience, it would be welcomed by a large audience. The proposed approach is compared to other alternatives available in the literature.

For completeness of the thesis, relevant preliminary information is provided as
1 Introduction

1.1 Background

The study of variation is a very important topic. In a scientific investigation the need to explain variability may arise from the following situations. In the first scenario the overall variation needs to be reduced. This constitutes the problem of identifying the sources of variability affecting the outcome, and the estimation of their influence. The second scenario is when the sources of variability are key characteristics of the subject matter and, therefore, are known. The objective then is the precise estimation of their values. In either case, understanding or explaining the variability will contribute to making better decisions. In a statistical model, the parameters representing the sources of variability are called variance components.

Different improvement philosophies can be found in all types of industries. Consider the following examples. Robust engineering is most often associated with designing products taking into consideration the tolerances (random variation) in the specifications and the components used. Such tolerances can be modelled as variance components. Higher quality is obtained by reducing the number of defective items in assembly lines, i.e. by increasing yield. In the health/safety and drug industry, quality and robustness are measured by the effectiveness of a health policy or a medicine across the general population in a given time period. What these examples have in common is that a high level of performance can only be attained by reducing variation.

For a discipline where the main interest of the studies is the precise estimation of different variances, consider quantitative genetics. Variance components are
used to estimate the genetic merit, *e.g.* of plants or cattle, which are then ranked and classified by their potential for improving next progenies. Other areas where estimating variances is the purpose of the statistical studies include modern theory of communications systems, image analysis and computer vision. In these cases, the reference signal, *e.g.* channel noise or the initial location, from which the subsequent measurements are taken, is best characterised as a variance component.

In summary, important phenomena to be understood are increasingly complex and sources of variability, other than the random error, have to be explained by the associated models.

Traditional statistical modelling and experimental design deal with models that focus on fixed parameters. Using design optimality theory, experiments that minimise the variance of these parameter estimates can be obtained. For estimating variance components, extensions to the methods and techniques are needed. The linear mixed effects model has been used to accommodate the added difficulties. Customarily, efficient designs and good estimates of the fixed parameters in the presence of random effects are obtained. However, the focus remains on the fixed elements, while the random parameters are modelled as nuisances.

Less attention has been given to the cases where the accuracy of the estimation of the variances in the models is the main objective. Consequently, fewer methods are available for the optimum estimation of the variance components. Detailed surveys of the work done in this area are Khuri and Sahai (1985) and Khuri (2000).

**Estimation of variance components**

The fundamental methodology for the estimation of variance components is due to R. A. Fisher. It is the widely known ANOVA method of estimation (Fisher, 1941). Most authors have used this method and some authors found some refinements for
1 Introduction

1.1 Background

certain models. To embrace the maximum likelihood method (MLE) for variance components, the most influential work is Hartley and Rao (1967). MLE techniques for mixed models have become standard in many computer packages such as SAS and R, but ANOVA is still used by some authors. In a paper as recent as 2002, Ankenman et al. use ANOVA estimators.

The intricacy of estimating variance components is heavily driven by the balancedness of the data. For balanced data, the use of the ANOVA method is common. Results for models with two to four variability stages, using crossed and nested random layouts, are readily available, e.g. see Burdick and Graybill (1992) and Rao (1997). Extensions of ANOVA estimators for more elaborated balanced models should be straightforward, although algebraically cumbersome. Results for MLE and REML are also available for a few specific models, however, expressions in closed form cannot be obtained even in some simple balanced cases. However, general expressions that can be applied directly to a variety of models are available. This is an advantage of likelihood over ANOVA estimation despite the computational demands.

It is known that under normality assumptions, ANOVA estimators are the same as restricted maximum likelihood (REML) estimators (Rao, 1997). Although normality is not required for the equivalency in the case of balanced data, on a model to model basis, closed form expressions for the standard error of each variance component estimator have to be provided when ANOVA is used. This is limiting for constructing and comparing optimum designs. MLE is more suitable computationally given the expressions available for the information matrices of the general mixed model. These apply to balanced as well as unbalanced data.

When the data are unbalanced, the problem becomes more difficult. The equivalency of methods is not as strong as for balanced cases, therefore results from different methods are not guaranteed to have the same properties. Depending
1.2 Experimental designs for variance components

on the degree of unbalancedness, one method could be more favourable. The general ANOVA method can be used, and in particular the framework given by Henderson (1953). Maximum likelihood on the contrary, provides a unified approach for estimation of the general mixed model as pioneered by Hartley and Rao (1967), although computations can still be demanding.

In this thesis ANOVA estimators are not used for design construction. MLE and REML are preferred. Solving the ML equations is no longer a major impairment; the asymptotic behaviour of the estimators is desirable and the variance-covariance matrix is computable in most cases through the information matrix. Moreover, the same general formulation can be used for different models with balanced or unbalanced data. Additionally, the dispersion-mean model for finding optimum designs is presented here. There is no available literature where this method is used for designing experiments.

1.2 Experimental designs for variance components

The theory of design optimality for fixed effects models is now well established, see for example Atkinson et al. (2007). It can roughly be described as the problem of defining which and how many treatments (observations) in the design region are required in order for the estimates to exhibit minimum variance in some respect.

Common criteria of optimality are $\mathcal{A}$- and $\mathcal{D}$-optimality, see Section 2.4 for more criteria. An $\mathcal{A}$-optimum design minimises the average variance of the estimators by minimising the trace of the inverse of the information matrix. A $\mathcal{D}$-optimum design minimises the confidence ellipsoid of the estimates by maximising the determinant of the information matrix.

To construct an optimum design for the fixed effects, the standard approach is to
employ an exchange algorithm. Interchanging support points varies the optimality criterion because the information matrix, for a given variance-covariance structure, depends on the design matrix, which itself is the collection of the support points of the design. It must be noted that the same philosophy cannot be directly applied to the case of variance components, because the design criterion depends on the variance-covariance matrix and not the design matrix. One important reason for this is that the variability structure of the design, reflected in the form of the variance-covariance matrix, is what determines the information matrix.

In contrast to the case of fixed effects, there is no unified approach for design optimality for variance components. The first full development of criteria based estimation is that of MINQUE by Rao who presented the method in a series of papers following a single exposition in Rao (1973). Further, using MINQUE in an iterative manner gives I-MINQUE estimates. It is known that I-MINQUE and REML are equivalent under normality, as stated in Rao (1997) and Searle et al. (1992). However, the design problem is not addressed by either authors. Khuri (2000) mentions that all early literature reports optimum designs only for some specific models, more notably for one-way and two-way models, crossed and nested. An attempt to provide a general method for an $m$-way crossed model using the method of unweighted squares of means was given by Mukerjee and Huda (1988). They concluded that when all factors have an equal number of levels, a balanced design is $A$-optimum. Similarly, Ankenman et al. (2003) found that highly balanced designs are $D$-optimum (they don’t explore $A$-optimality). These conclusions are important given that balanced designs may be easier to calculate and conduct. It yet appears that there is not a complete general theory for optimum designs for variance components.

It is not always necessary to consider the treatment design to find an optimum design for the variance components. For example, the information matrix based
on MLE is block diagonal. The block corresponding to the random effects is independent of the fixed effects and vice versa. This indicates that the optimality of fixed effects is only related to the treatment design (choice of support points), and the optimality of the variance components depends mainly on the variability structure.

Comparatively, the information matrix based on REML involves the structure of the fixed and the random effects. This allows to design optimum experiments for the variance components that would potentially yield unbiased estimates for a given treatment design. Moreover, designs for simultaneous estimation of the fixed effects and variance components are theoretically possible.

In general, proposed designs based on MLE or REML are not common in the literature. Li and Klotz (1978) describe the estimation but not the design construction of the variance components for the Split-Plot experiment using MLE and REML. Ankenman et al. (2002) introduced split factorial experiments, that using the ANOVA method would allow the estimation of both type of effects in a mixed model, but with an even number of variance components only and the fixed part must have a crossed layout.

For nested models, Bainbridge (1965) introduced the staggered nested designs with the intention of spreading the available degrees of freedom as evenly as possible over all the levels of nesting. This would allow almost equally “good” estimation of all the variance components. The work of Goldsmith and Gaylor (1970) compares 61 designs for a random two-fold nested model. These authors compare the trace and the determinant criteria based on the variance-covariance matrix of ANOVA estimators; they preferred the trace criterion (A-optimality). Along similar arguments and also using ANOVA estimators, Delgado and Iyer (1999) find staggered designs for the 3-stage nested model using A-optimality. Specifically, by performing numerical optimisation and using search algorithms,
they were able to compare more designs than those shown by Goldsmith and Gaylor (1970).

Compound design criteria based on MLE information matrices for simultaneous estimation of the fixed effects and the variance components have been used by Atkinson and Cook (1995) and Giovagnoli and Sebastiani (1989). The former also include a Bayesian criterion and the latter obtained the optimum number of observations only for the balanced one-way nested random model. Ankenman et al. (2003) introduced assembled designs and used the MLE method to derive $\mathcal{D}$-optimum designs. The work of Avilés (2001) in her PhD thesis is the basis for the latter paper. The results are calculated using combinatorial arguments and use only MLE information matrices. She presented the REML formulation but did not use it. The fixed effects include the overall mean alone. The results are specific for models with two variance components only.

It is important to note that there are computational complexities when dealing with the information matrices and the corresponding evaluation of a design criterion. Finding an optimum design requires a large number of iterations, therefore a relatively small number of variance components can become very computationally demanding such that it is not very practical. Most of the burden lies on the dependence of the information matrix on the parameters. There is need to circumvent this dependence. Bayesian methods can give a solution for well defined problems. Nonetheless, the literature on approaches of this kind is not vast. A significant contribution in this arena is presented by Tack et al. (2002). Similarly, Mylona et al. (2014) provide an innovative approach that includes a Bayesian composite optimality criterion. They use state of the art quadrature methods to tackle the optimisation of the design criterion which involves the challenging integration task associated with Bayesian formulations.

The information matrix for the variance components, include the parameters
1.2 Experimental designs for variance components

to be estimated. Consequently, the associated designs depend on the true values of the variance components. Hence, preliminary values for the parameters have to be used. This implies that the optimum designs are locally optimum. For example, preliminary values are needed when the objective is to select the optimum number of levels in a balanced design. This also means that designing experiments for estimating variance components is a problem similar to that of finding designs for nonlinear fixed models. This remark is important and local optimality is exploited in this thesis for the construction of designs for variance components.

Most results have used the general axiom that all variance components are to be estimated using the same number of degrees of freedom, thus giving equal precision for the estimates. A key example of this is the staggered nested design. This suggests that the more balanced the design, the more optimum or efficient it is. Evidence that balanced designs are optimal for variance components can be found e.g. in Mukerjee and Huda (1988), and considering different levels of balancedness, Ankenman et al. (2003). The latter found that the most balanced design is optimum. However, the consideration that all variance components can be estimated with equal precision is not relevant in all cases, such as in Split-Plot designs. Consequently, there is value in considering unbalanced designs that do not explicitly try to use the same number of degrees of freedom along the variability structure but still attempt to keep a high level of balancedness. One class of such type of designs is introduced in Section 7.1.1.

A graphical approach to discriminate among designs is the one developed by Khuri (1997), which is based on the quantiles of the variances of the variance components but, although clever and informative, is not very easy to interpret.

From the pioneering work of R. L. Anderson in the 1960’s and 1970’s to the computer generated designs of Delgado and Iyer (1999) and the assembled designs of Ankenman et al. (2003), the search for optimum designs for variance components...
1.2 Experimental designs for variance components

shares some key characteristics as follows:

1. Variance components estimators are based on information matrices that are nonlinear functions of their unknown true values. Thus, pre-specified values are required to construct $A$- and $D$-optimality designs. Consequently, the designs are only locally optimum.

2. There is no single design that is optimum at once for all the variance components of a model. The relative sizes of the individual variance components are important and they relate to each other by their ratio. For a large random error, the ratio is small and a balanced design will most likely be optimum. As the ratio increases, the optimum design is expected to concentrate the experimental effort where the largest variance component is.

3. For balanced designs, the optimisation problem seems to focus on finding the number of observations that meets the criteria. For unbalanced designs, the equivalent problem is to find the optimum number of levels of the random factors in the model.

4. The reported optimum designs are highly tied to a specific model and estimation method, and most reported designs are based on ANOVA estimators. Generic methods and software that allow the construction of optimum designs for different models and design structures is not available in general.

The work presented in this thesis provides tools to deal with some of these complexities. Methodology to find and select optimum designs for variance components is developed using MLE and REML estimators and, additionally, an alternative method known as the dispersion-mean model. The methodology has been implemented using the R language and the results are reported using a simple, but highly informative graphical approach, not seen before in this context.
Furthermore, it is important to make the methodology accessible to experimenters in any field. To contribute to this challenging task, a framework is developed with the objective of strengthening the relationship between the definition of an experimental design and the formulation of the model for data analysis.

1.3 Structure of the thesis

This document is organised as follows. At the end of each chapter, a summary of the most relevant content is given.

Chapter 2 gives fundamental concepts about the design of experiments and introduces terminology needed afterward. It continues with the presentation of statistical models including notation used throughout the thesis. Finally, it gives the essentials of the theory of design optimality.

Chapter 3 presents methods used to estimate the parameters in a model. In each case, the formulation of the information matrix is described. This is fundamental since optimum experimental designs are based on the information matrix. Also included is a discussion of the dispersion-mean model which allows the variance components to be explained using the concepts of the standard linear model instead of nonlinear models. This formulation is later used to compare optimum designs for the variance components and to find designs when the data are non normal.

Chapter 4 gives the full description of the more flexible formulation of the linear mixed model that is proposed here. Required additional ideas and the definitions of the key concepts of crossed and nested structures within the context of the thesis are introduced. The implementation of the extended formulation of the mixed model is exemplified using common design structures. A comparison with other available formulations is made.
Chapter 5 presents results of information matrices for the variance components for specific cases. Subsequently, these are used to find optimum designs.

Chapter 6 presents the fundamental theory needed to justify the design construction algorithms. It contains the description of the algorithms developed for this thesis and the corresponding results.

In Chapter 7, the class of least-unbalanced designs is introduced as an approach to the design of unbalanced experiments for the estimation of variance components. This class allows seamless use of the same tool set as for the balanced designs presented earlier.

It was made possible to take part in an industrial internship program that added extra dimensions to the learning outcomes of the PhD research. This created the invaluable opportunity to put into practice the developments made, although new research results were not directly obtained. The outcomes of this internship are detailed in Appendix B.

As part of the discussion on Chapter 7, further comments about extensions of the results, topics for future work and overall conclusions finalise the chapter. To complete the presentation, Appendix A contains known results in statistics and matrix algebra that are used in the thesis.
Chapter 2

Preliminaries

In this chapter, the important role of the experimentation process in a scientific study is stressed. First, it presents the fundamental ideas related to the design of experiments and introduces new terminology. The second part is an account of the most encountered statistical models. Finally, the approach of design optimality is described.

2.1 Design of experiments

2.1.1 Basic ideas and terminology

A scientific study or investigation is a project that aims to learn about a specific phenomenon in a rigorous way. The phenomenon under study can be conceptualised as a system whose output is determined by the settings of several inputs, as shown in Figure 2.1. The learning process relies heavily on experiments. These are carried out through several experimental phases.

An experimental phase consist of different steps and each of them should have specific objectives; an experimental plan should inform on all the details, constraints, and instructions to run the experiment. Generally, the process is
iterative. Consequently, the experimental plan has to be updated as the study develops. It is important to realise that different information is needed for every part of the study, and each one may need an experimentation phase. This requires an established methodology to design the experiments. Effective experimentation is key because the actual knowledge and learning in a study are primarily obtained by applying statistical methods to the outcomes of experiments. Design of Experiments is the area of statistics that provides methodology and strategies for planning experiments. Other common names are Experimental Design and Statistical Design.

Statistical science is involved in the experimentation process because all systems have a very important element: variation. The total variation is caused by different sources each contributing a different amount, and their effects cannot be completely removed. When the system is controlled and stable, the variability is expected to be random. Completely random variation is referred to as random error and it is associated to variables that are not under the control of the experimenter. In Figure 2.1, the arrows entering from the bottom labeled $\epsilon_i$, $i = 1, 2, \ldots, N$ represent the random errors.
The technical concept is that the impact on the output due to these sources of variability is unknown, but it is known that the experimental results are influenced by them. Therefore, the output of the system is considered a random variable because it captures the random error. Given that the behaviour of this error is not systematic its overall effect can be modelled as a random variable.

An experiment consists of applying specific inputs to the system under investigation, and obtaining data from the output by measuring a chosen characteristic (presumably affected by the inputs’ settings). Each measurement forms a data point or sample, called an **observation**. The overall outcome is known as the response variable or simply **response**. The response exemplified in Figure 2.1 is denoted by \( y_i, i = 1, 2, \ldots, N \). Note that each \( y_i \) has an embedded random error \( \epsilon_i \) and it can have additional random effects \( \tau_i, i = 1, \ldots, q_i \). Also, more than one response can be measured. Using appropriate statistical methods, the information contained in the measured output can be extracted, and knowledge about the system is gained. Informed decisions can then be made based on the analysis.

The set of variables acting on the system are known as **factors**. For each factor, **factor levels** are the predetermined values (set points) that can be used in the experiment. Commonly, multiple factors are involved and a properly designed experiment must allow to study their **effects** on the response simultaneously. The factors and their effects can be distinguished as follows:

**Fixed** When it is clear which levels are to be used due to specific interests of the study or the experimental conditions, and the values assigned to those levels will not change during the experiment, then that factor is defined as fixed and it contributes **fixed effects** to the response.

**Random** If it can be considered that the factor levels included in the experiment are only a random sample of a population of levels, then that factor and its effects are defined as random.
A further distinction is between *input* and *nuisance* factors. Input factors are explicitly studied and their levels have to be adjusted (reset) during the experiment; in Figure 2.1 they are labeled $x_i$, $i = 1, 2, \ldots, m$. Nuisance factors also enter the system, but cannot be changed purposefully; they can be subdivided in two types:

**Controllable** When present, the experiment has to be designed using a technique called *blocking* (see Section 2.1.2), such that the nuisance factor do not contribute to the random error.

**Non controllable** These may or may not be known, but no control or explicit condition can be set on them; their effects add up to the random error.

Figure 2.1 shows controllable nuisance factors as the arrows on the top labeled $\zeta_i$, $i = 1, 2, \ldots, s$. In some applications, studying the controllable nuisance factors can be the objective of an experimental phase since they are sources of variability. In those cases they are referred to as *variance components*. Chapter 5 contains developments on variance components that are relevant to designing experiments for models with more than one source of variability.

Factors can be further classified as *quantitative* or *qualitative*. Quantitative factors can be related to continuous variables, since their levels can be set to any value within its range. In contrast, the levels of qualitative factors can only be set to a few pre-specified values as in a discrete variable. Additionally, factors can be defined as *hard to change* (HC). A HC factor occurs when adjusting its levels would disrupt the execution of the experiment, or when such change is expensive. Common examples are environmental factors *e.g.* temperature and pressure. Location and machinery setups are often hard to change as well. If a factor does not fall into the HC category, then it is *easy to change* (EC).

The way factors are arranged with respect to other factors is very important.
Arrangements of \( m \) factors are known as \textbf{m-factor} layouts. Chapter 4 explores the concept in detail. There are two fundamental ideas. First, factors are \textit{crossed} (C) if all the level combinations can be included in the experiment. This is also called a \textit{factorial} array or an \( m \)-factor crossed layout, depending on the context (see Section 3.2.1). The second arrangement, known as \textit{nesting}, happens when the levels of a factor are uniquely associated to one specific level of a different factor; it can then be said that the second factor is \textit{nested} (N) within the first one. Complex experiments can be defined by combining simple factor layouts yielding a full design structure.

The presence of nesting, in most cases, gives the experiment a \textit{hierarchical structure}. Each level in the hierarchy is called a \textit{stage}, and any number of stages can be used. Usually, a variance component is associated to each stage, and according to the hierarchy an experiment (and the corresponding model) acquires a specific \textit{variability structure}. The basic case is a crossed structure which has only one variability stage. The crossed structure itself can be part of a nested arrangement; this would involve more than one variability stage. The presence of HC factors usually establishes a hierarchy and the need to use special blocking techniques, \textit{e.g.} a \textit{Split Plot}. See Chapter 4 for a full description of these ideas.

A central concept in experimental design is that of a \textit{treatment}. It is defined as a combination of factor levels. Roughly speaking, treatments are experimental conditions of particular interest in the study, hence an experiment consists of sets of treatments that are “given” or applied to the experimental units being used. The effect of a treatment on the response variable(s) is known as the \textit{treatment effect} (can be fixed or random). Each time a treatment is applied to an experimental unit, an \textit{experimental run} or \textit{trial} has taken place. Each run gives rise to an \textit{observation}.

It is also important to have a clear definition of what the \textit{experimental units}
are. This depends on the specific system under investigation. For example, when
the dose of a drug is administered to a patient, each dose will be a treatment
and each patient an experimental unit. Another example is a production process
where the manufactured parts from an ultrasonic welding machine are the result
of a treatment involving settings for ultrasonic frequency, pressure, contact area
and time, whilst the set of components being welded form the experimental unit.
Identifying different types of experimental units is key. If more than one type is
involved, the design structure is hierarchical and a variability structure has to be
defined accordingly (see Chapter 4).

Using the terminology above, an experiment can be defined as the operation
of applying treatments to experimental units to record an observation. The
design of an experiment is the process of assigning selected treatments to
experimental units and the sequence in which the treatments have to be applied.
This can be understood as the data generation process.

The set of all possible treatments available to use in the experiment define the
design region. Consider the case of fixed factors where this region is denoted by
\( \mathcal{X} \). In an unrestricted design region, the operational limits of the experiment are
determined by the minimum and maximum levels of every factor. For example,
if two quantitative factors are of interest and both are used at two levels only, a
rectangular region is defined by the four values as depicted in Figure 2.2a. If a
third quantitative experimental factor is included, the region will be a cube (see
Figure 2.2b). For random factors, the analogous concept is called parameter
space and is denoted by \( \mathcal{Z} \). It refers to the values of the variance components; it
is defined by the set of values that are going to be considered acceptable for the
experiment according to the amount of knowledge of the system.

Often, there are restrictions on the factors or on certain treatments. A restricted
treatment that is often found is when the extreme levels of the factors cannot
2 Preliminaries

2.1 Design of experiments

all be used at the same time, e.g. all factors at the maximum level, which is usually related to a physical limitation. This causes the design region to be of irregular shape. For experiments with restrictions in the factors, consider the following example. Figure 2.2c shows, for two quantitative factors, two different (independent) possibilities for the case when one factor has to be larger than the other one. The resulting design regions are of triangular shape. For the condition $x_1 > x_2$, the restricted design region is shown in light grey. The second possibility is when $x_2 > x_1$ and the resulting design region appears in dark grey.

The approach of design optimality is a powerful tool to design experiments with restricted design regions; see Section 2.3 for a full description of the approach, which plays a key role in this thesis.

In summary, most scientific and engineering studies have an experimentation
phase. An experimental plan is required if this phase is to be effective. More
than one iteration of an experiment may be needed. In addition, there may
be an experimentation phase at each step of the study. For these reasons, an
established methodology to create the experimental plan is required. The design
of the experiment specifies the treatments that will be used on each trial, thus
defining what data will be collected.

2.1.2 Fundamental considerations

The statistical design of experiments has been studied extensively. It is recognised
that the first authoritative account of the subject was given by Sir. R. A Fisher in
his book “Statistical Methods for Research Workers” (see Fisher (1941)), followed
by “The Design of Experiments” (see Fisher (1949)). Excellent work that followed
include Cochran and Cox (1957), Federer (1955), Cox (1958), Box et al. (1978)
and Box and Draper (1987). More recent books are Wu and Hamada (2009) and
Montgomery (2001) that focus on engineering experiments, whilst Kutner et al.
(2005) present examples of social and biological sciences.

Nowadays, the use of statistical software is unavoidable. Some packages allow
for an efficient workflow from the design to the analysis phases and even the
creation of a report. Useful references describing software that can be used for
statistical design are e.g., Dean and Voss (1999) which is accompanied by examples
and code for SAS®, and optimum experimental design is treated by Atkinson
et al. (2007), making extensive use of SAS. Milliken and Johnson (2009) also use
JMP®. There are few sources with code written for the R language, for example
Casella (2008) and Rash et al. (2011).

The use of a designed experiment should make the generated data as reliable as
possible so that these data provide enough information for analyses. All literature
point out three fundamental principles that statistically design experiments have
to comply with. **Randomisation** assures that potential systematic variability, such as machinery wear out or warm up, is only due to the random error. This principle states that experimental runs have to be executed in random order. **Replication** provides the data needed for estimation of the variance components. To replicate is to measure the response in more than one trial, by recreating the experimental conditions and the treatment. The third principle is named **blocking**, where controllable nuisance factors are accounted for explicitly in the experimental design. This helps in keeping all unexplained variation random, thus potentially increasing the precision of the random error estimate. It is achieved by grouping experimental units under the influence of a level of the nuisance factor such that the variation among the units within a group, or **block**, is at its minimum and among blocks is maximum. The number of observations per block is known as **block size**.

A good experimental design has to consider statistical aspects of the study as well as non-statistical ones (Federer, 1955). For this reason, the design of an experiment requires an overall understanding of the system under investigation, information about the experimental conditions, and statistical knowledge. The experimenter’s plan has to include an evaluation of the circumstances under which the experiment is to be ran in order to be aware of possible matters that could affect the response, but do not help answering experimental questions. Consequently, effective communication between the statistician and the experimenter is critical. All possible effort must be made to design an experiment that is insensible to, or **robust** against, unfavourable nuisance effects.

Robustness can be defined in many different ways depending on the context. Another common interpretation is that of being robust against miss specification of statistical assumptions of the design and model under consideration. Details on statistical modelling are given in Section 2.2.
The first step in the design phase is to define the specific questions that need to be answered. A second step is characterising the system and the experimental conditions. The system characterisation involves determining the response variable and the shape and size of the design region, \textit{i.e.} the factors to be studied, the range of the factors levels and possible restrictions. The experimental units also have to be defined considering aspects of size, type, and crucially, the number of units available, \textit{i.e.} the total number of observations, \( N \). For the experimental conditions, any constraint for the realisation of the experiment has to be taken into account. This should lead to the determination of the blocking scheme, the arrangement of the factors (crossed or nested) and thus the variability structure, and the randomisation format. Additionally, the block sizes and the number of replications per block could be defined here, or they can be determined later based on the design of the experiment.

<table>
<thead>
<tr>
<th>Type of Experiment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparative</td>
<td>Used to identify if the different treatments yield a statistically significant response, \textit{i.e.} if one treatment is better (or worse) than the others.</td>
</tr>
<tr>
<td>Screening</td>
<td>In early phases, they help identify which factors are relevant to the system and conditions of interest.</td>
</tr>
<tr>
<td>Response surface (RSM)</td>
<td>Used to establish or confirm the type of functional relationship between the response and the factors.</td>
</tr>
<tr>
<td>Optimisation</td>
<td>Find the treatments that yield the optimum (maximum or minimum) response.</td>
</tr>
<tr>
<td>Variability assessment</td>
<td>These are of particular importance if heterogeneity is expected or is part of the model. When the analysis is oriented to power calculations, precise estimation of the variability is important, and when the variability structure is complex.</td>
</tr>
</tbody>
</table>

These two steps combined should amalgamate into a few clear goals, and
the experiment has to be designed with those intentions in mind. According to these goals, designed experiments can be broadly classified as shown in Table 2.1 on the preceding page. The last category in the table, Variability assessment, is an important topic in this thesis. Some developments on experiments for estimating variance components are presented in later chapters.

The conceptualisation of the statistical model to be used is very important. The model has to be defined according to the objectives, but it also has to be able to account for the experimental conditions. The type of design, the characteristics of the factors, their arrangement and their effects are influential in the choice of a suitable model. An important distinction to be made is between random and fixed effects. The decision of which effects will be considered fixed and which ones random, involves statistical considerations, but mainly it is driven by the application area and/or the experimental conditions.

The properties of the response can also define the model needed. Two common examples are a dichotomous response which requires a model for non-normal data, and when the response comes from a growth (decay) study, which requires a nonlinear model. The way in which the statistical analysis of the data is to be performed and how the conclusions (inference) are reached, is strongly related to the model selection. Since the experiment has to be designed in connection with the assumed model; Section 2.2 introduces the most useful statistical models.

As was mentioned in Section 2.1.1, the statistical design will dictate which treatments will be applied to the experimental units. It also specifies the randomisation scheme to follow during the execution. The experimental plan is completed by specifying all other elements, e.g. personnel, time, machinery, measuring equipment, and how the data are to be collected and handled. Several conditions have been established for an experiment to be reliable. Box and Draper (1987) provided 14 points for a good experiment to comply with. Some are not as relevant now
with the availability of computer software. Others may not be applicable in some experimental conditions. Either way, it is advisable to keep them in mind and use them whenever possible. Also, the basic ideas of randomisation and replication must be applied during the planning and execution phases.

In summary, an experiment has to be designed purposefully to meet the particular needs of the study at the corresponding phase. Clearly not all statistically designed experiments are appropriate for all circumstances, since each one may seek different objectives. For this reason, knowledge and understanding of strategies that allow the experimenter and statistician to define a purposeful experimental plan are of surmounting importance.

2.2 Statistical models

The action performed by the system, i.e. the connection between the response and the factors, is described by a presumed functional relationship called model. In general, there is one function applied to the fixed factors and a different one for the random factors. The fixed part of the model is also known as the deterministic part. The random effects belong to the random or stochastic part of the model.

The determination of an appropriate model, alongside a suitable analysis technique, is an integral part of the statistical design of experiments. It is a team effort between the experimenter and the statistician to choose the best model possible. Once it is defined, the task is to find values for the parameters so that the model is indeed a good representation of the data. Alternative expressions are that the model accurately explains the variation in the data or that the model fits the data with high accuracy.

This section presents the most used statistical models. Methods to estimate the parameters involved are described in Chapter 3.
2.2.1 Linear fixed models

The linear statistical model with only fixed factors and fixed effects has the most widespread use. It is extensively documented in the literature and heavily used in comparative and response surface experiments (see Table 2.1). A good reference for the former is Hinkelman and Kempthorne (2008) and for the latter is Box and Draper (2007).

Consider an experiment with $N$ observations and $m$ fixed factors. The different factors are represented by the vector $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_m)^T$. Each factor has an associated vector $\mathbf{\ell}_{x_i}$ describing its levels such that $\mathbf{\ell}_{x_i} = (\ell^{(1)}_{x_i} \ \ell^{(2)}_{x_i} \ \cdots \ \ell^{(k_i)}_{x_i})^T$ where $i = 1, 2, \ldots, m$, and $k_i$ is the number of levels for the $i$th factor. For every $x_i$, only one element of $\mathbf{\ell}_{x_i}$ is used in a given experimental run. Hence, a treatment for the $j$th observation can be denoted by the vector $\mathbf{t}_j = (\ell_1 \ \ell_2 \ \cdots \ \ell_m)^T$ where $j = 1, 2, \ldots, N$ and $\ell_i \in \mathbf{\ell}_{x_i}$.

For the statistician or statistics savvy experimenter, this extended notation using $\mathbf{\ell}_{x_i}$ and $\mathbf{t}_j$ will appear excessive. However, for the experimenter or engineer not well acquainted with statistics, this simple extension may help introduce the concepts and methodology using a similar notation throughout the experimentation process. This also remarks the need for thinking holistically in terms of the statistical model and design from the definition of a specific experimental phase.

Using the extended notation, a linear model for the $i$th observation can be written as

$$y_i = f^T(\mathbf{x}|_{t_i})\boldsymbol{\theta} + \epsilon_i \quad i = 1, 2, \ldots, N,$$

where $y_i$ is the response. $f(\mathbf{x})$ is a vector of known linear functions of the $m$ factors and each term is called a regressor whose values are defined by $\mathbf{t}_i$, condition that is denoted $\mathbf{x}|_{t_i}$. Finally, $\boldsymbol{\theta}$ is the $p \times 1$ vector of the fixed parameters associated with those regressors. The random error is represented by the terms $\epsilon_i$. 

The model (2.1) can be written in standard notation and matrix form as

\[ y = X\theta + \epsilon, \quad (2.2) \]

where \( y \) is the \( N \times 1 \) vector of responses. The matrix \( X \) is of size \( N \times p \) where the \( i \)th row is defined by the regressors of \( f(x|t_i) \) for each experimental run, that is

\[
X = \begin{bmatrix}
f^T(x|t_1) \\
f^T(x|t_2) \\
\vdots \\
f^T(x|t_N)
\end{bmatrix}.
\]

\( X \) is called design matrix. Its rows are commonly sorted to facilitate the data analysis. The list of treatments \( t_i \) must be randomised for the execution of the experiment. The vector \( \theta \) of length \( p \) was defined in (2.1). The vector \( \epsilon \) contains the random error for each of the \( N \) observations.

**Example 1**

An engineer suspects that the performance of a car is affected by tyre temperature. An experiment is going to be conducted at the closest race track to verify the suspicion. The car’s performance is measured by the speed (km/h) and the temperature in degrees Celsius (°C). The response variable is the speed and the factor is tyre temperature. The maximum speed of the car over a lap will be recorded once the tyres have reach a predetermined temperature. Three temperatures will be tested: 100 °C, 130 °C and 160 °C. Two observations at each temperature will be recorded. The same car and driver will be used. A one to one relationship is going to be considered appropriate.

The engineer proceeds to document the conditions of the experiment. A formal, structured and informative scheme is required. The following format is suggested:
Experiment: Performance vs Tyre temperature
Date: dd/mm/yyyy  Location: The race track
Response: Speed (km/h)
Factors: $x = x_1$, $m = 1$

$x_1$=Tyre temperature ($^\circ$C); levels=(100 130 160)

$\ell_{x_1} = (-1 \ 0 \ 1)^T$ (coded levels)

Model: $y_i = \theta_0 + \theta_1 x_1 + \epsilon_i, \ p = 2$

$f(x) = (1 \ x_1) \ \theta = (\theta_0 \ \theta_1)^T$

Total number of observations N=6

This format provides all the details of the experiment in a succinct manner with
information for both, the execution of the experiment and, along with the data
sheet, the statistical analysis.

Example 2
Consider a general experimental situation. It is desired to estimate a full quadratic
model in two variables. All combination of the factor levels will be included. The
experimental sheet appears below; only one replication is shown.
Experiment: A very general experiment

Date: dd/mm/yyyy       Location: The laboratory

Response: output (units)

Factors: \( \mathbf{x} = (x_1 \ x_2)^T, \ m = 2 \)

\( x_1 \) = name 1 (units); levels = (l1a l1b l1c)

\( \ell_{x_1} = (-1 \ 0 \ 1)^T \) (coded levels)

\( x_2 \) = name 2 (units); levels = (l2a l2b l2c)

\( \ell_{x_2} = (-1 \ 0 \ 1)^T \) (coded levels)

Model: \( y_i = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2 + \theta_{12} x_1 x_2 + \epsilon_i, \ p = 6 \)

\( f(x) = (1 \ x_1 \ x_2 \ x_1^2 \ x_2^2 \ x_1 x_2) \quad \mathbf{\theta} = (\theta_0 \ \theta_1 \ \theta_2 \ \theta_{11} \ \theta_{22} \ \theta_{12})^T \)

Total number of observations \( N=9 \)

\[
\begin{array}{c}
  f(x|t_1) = (1 \ -1 \ 1) \\
  f(x|t_2) = (1 \ 0 \ -1) \\
  f(x|t_3) = (1 \ 1 \ -1) \\
  f(x|t_4) = (1 \ -1 \ 0) \\
  f(x|t_5) = (1 \ 0 \ 1) \\
  f(x|t_6) = (1 \ 1 \ 1) \\
  f(x|t_7) = (1 \ 0 \ 0) \\
  f(x|t_8) = (1 \ -1 \ -1) \\
  f(x|t_9) = (1 \ 1 \ 0) \\
\end{array}
\]

\[
\mathbf{X} = \begin{bmatrix}
1 & -1 & -1 & 1 & 1 & 1 \\
1 & -1 & 0 & 1 & 0 & 0 \\
1 & -1 & 1 & 1 & 1 & -1 \\
1 & 0 & -1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\]

\[
\mathbf{y} = \begin{bmatrix}
y_8 \\
y_4 \\
y_1 \\
y_2 \\
y_7 \\
y_5 \\
y_3 \\
y_9 \\
y_6
\end{bmatrix}
\]

The list of treatments \( t_i \) is known in the analysis phase. In the design phase, its definition is the objective. The two models presented above include an intercept parameter \( \theta_0 \) whose regressor value is 1. In Example 2, the term \( x_1 x_2 \) is called an \emph{interaction} between factor \( x_1 \) and \( x_2 \); its corresponding model parameter is \( \theta_{12} \). An interaction means that the effect of one factor is linked with the levels of
another factor. Here the factors form a 2-factor crossed layout. Different factor layouts can be used, see Chapter 4.

These models are called linear because the parameters $\theta$ appear linearly in the equations. Also, there is only one stochastic part which is the random error $\epsilon$. Still, because there are none random input factors, the models are considered as fixed effects only. The random errors are usually assumed to be independent and identically distributed with zero mean and variance $\sigma_\epsilon^2$, i.e.

$$
E(\epsilon_i) = 0, \quad \text{cov}(\epsilon_i, \epsilon_j) = 0, \quad \text{var}(\epsilon_i) = \sigma_\epsilon^2,
$$

(2.3)

where $i = j = 1, \ldots, N$. It is also common to assume they are normally distributed, which is written as

$$
\epsilon \sim N(0_N, \sigma_\epsilon^2 I_N),
$$

(2.4)

and implies that the variance-covariance matrix of the response vector is

$$
\text{var}(y) = V = \sigma_\epsilon^2 I_N.
$$

(2.5)

Equations (2.4) and (2.3) state that the model assumptions for the random errors are normality, homogeneous variance (homoscedasticity) and independence. This is the simplest set of assumptions. It has been found that from those, the least restrictive is normality; the variance changes from observation to observation are of more importance (McCullagh and Nelder, 1989). Therefore, there is need to extend the linear model to allow for the variance to be explained. This is done by defining a less restrictive variance-covariance matrix $V$.

For example, when not all observations have the same variance, a heteroscedastic model is needed. For this case, equation (2.2) is still a valid expression, but
the following assumptions are in place

\[ E(\epsilon_i) = 0, \quad (2.6a) \]
\[ \text{cov}(\epsilon_i, \epsilon_j) = 0, \quad (2.6b) \]
\[ \text{var}(\epsilon_i) = \sigma_i^2, \quad (2.6c) \]
\[ \epsilon \sim N(0_N, V), \quad (2.6d) \]

where the parameters \( \sigma_i^2 \) in (2.6c) are called the \textit{variance components}; they are commonly unknown and have to be estimated (see Sections 3.3 and 3.4). In this case, the variance-covariance matrix in (2.6d) is of the form

\[
V = \begin{bmatrix}
\sigma_1^2 & 0 & \cdots & 0 \\
0 & \sigma_2^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_N^2
\end{bmatrix}.
\]

(2.7)

The most general case is \( \sigma_1^2 \neq \sigma_2^2 \neq \cdots \neq \sigma_N^2 \). However, this results in an over parameterised model that is intractable and therefore simpler models with significantly fewer variance components are mostly used.

If the condition in (2.6b) is additionally relaxed so that \( \text{cov}(\epsilon_i, \epsilon_j) \neq 0 \), then the observations are correlated and the variance-covariance matrix can be written as

\[
V = \begin{bmatrix}
V_1 & 0 & \cdots & 0 \\
0 & V_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & V_r
\end{bmatrix} \sigma^2 + \sigma^2 \mathbf{I}_N,
\]

(2.8)

where the matrices \( V_1, V_2, \ldots, V_r \) give the correlation structure of the observations associated with \( \sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2 \), respectively. Therefore \( \sigma^2 = (\sigma_1^2 \ \sigma_2^2 \ \cdots \ \sigma_r^2) \); note
2 Preliminaries

2.2 Statistical models

how the random error is added in (2.8). This form of variance-covariance matrix is a suitable option for blocked and highly structured experiments.

Remember that (2.1), or equivalently (2.2), is a fixed effects model since all factors were assumed fixed. If only random factors were to be included, the model would be a random effects model. In the latter, distributional assumptions have to be made for the $i$th factor, fundamentally that it has variance $\sigma_i^2$ and zero expectation. Equation (2.5) would no longer apply, instead (2.8) can be used. This model will not be explicitly described further. Rather, more general models that include fixed effects and random effects, known as mixed effects models, are presented next.

2.2.2 Linear mixed models

The linear models suitable when fixed and random factor effects are both of interest are called linear mixed models. These models find application in a wide range of areas. Longitudinal and repeated measurements studies require the use of the mixed model (Davidian and Giltinan, 1995). Spatial models and clinical trials (Brown and Prescott, 2006) also use it. Industrial experiments (Milliken and Johnson, 2009) and pharmacokinetic studies also combine fixed and random effects. Models for variance components are mixed models as well (Searle et al., 1992; Malley, 1986).

A general form of the mixed effects model is

$$y_{ij} = f^T(x_{ij})\theta + g(\zeta_{ij})\tau + \epsilon_{ij}, \quad (2.9a)$$

which can be written in matrix form as

$$y = X\theta + Z\tau + \epsilon, \quad (2.9b)$$
where \( i = 1, \ldots, N \) is the index for the observations and \( j = 1, \ldots, v \) reflects the grouping (e.g. blocks or stages) of the observations. Further, \( \mathbf{y} \) is the \( N \times 1 \) response vector, \( f(\mathbf{x}) \) is the functional relationship for the fixed factors (as in (2.1)) and \( \mathbf{\theta} \) is the vector of fixed parameters. For the stochastic part, the vector of random factors is \( \mathbf{\zeta} = (\zeta_1, \ldots, \zeta_s)^T \) and \( \ell_{\zeta_i} \), \( i = 1, \ldots, s \) is the integer number of levels of \( \zeta_i \). 

\( g(\mathbf{\zeta}|t_j) \) are matrices that associate the random effects to an observation, commonly known as indicator matrices, where 1 is used to indicate presence and 0 otherwise; the function \( g(\mathbf{\zeta}) \) is used to contemplate extensions that specify a variance function (see e.g. Atkinson and Cook (1995)). The vector \( \mathbf{\tau} = (\tau_1, \ldots, \tau_r)^T \) contains the parameters for the random effects which include main effects plus interactions between them. These effects are symbolically represented in Figure 2.1 on page 29 as the arrows labeled \( \tau_1, \ldots, \tau_q \) entering the system from the bottom.

The matrix \( \mathbf{Z} \) is the design matrix of the stochastic part which is analogous with \( \mathbf{X} \) for the deterministic part, as defined in (2.2). It can be partitioned into \( \mathbf{Z}_j \) matrices each corresponding to a factor layout in a stage of the design structure. Note however, that the parameter space for the deterministic part corresponds to the column space of \( \mathbf{X} \), but the parameter space for the stochastic part is defined in a more complex way. The effect of each regressor has its own parameter in the fixed part, but in the random part, for each random regressor, there is one effect for every level in addition to one variance component.

Let \( s \) be the number of main random factors and \( r \) be the total number of random regressors in the model, \( r \geq s \). The latter have an associated variance component \( \sigma_j^2 \) with \( j = 0, 1, \ldots, s, s+1, \ldots, r \); these variances can be put in vector notation as \( \mathbf{\sigma}^2 = (\sigma_0^2 \ \sigma_1^2 \ \cdots \ \sigma_r^2) \). Let \( \mathbf{\tau} \) be partitioned into \( \mathbf{\tau}_j \) vectors. Each effect in \( \mathbf{\tau}_j \) corresponds to one of the \( \ell_{\zeta_j} \) levels of the \( j \)th random regressor. Therefore, the length of \( \mathbf{\tau} \) is \( q \) where \( q = \sum_{j=0}^{r} \ell_{\zeta_j} \). This formulation includes the random
error $\epsilon$ by specifying

$$
\tau_0 = \epsilon, \quad \ell_{\zeta_0} = N, \quad \text{and} \quad Z_0 = I_N,
$$

such that the random part in (2.9b) can be written more succinctly as

$$
Z\tau = \sum_{j=0}^{r} Z_j \tau_j. \quad (2.10)
$$

Consequently, the parameter space of the stochastic part consists of the vectors $\sigma^2$ and $\tau$ with a total of $r + 1$ variance components and $q$ random effects. If customary assumptions of normality and independence of the variance components are considered, then

$$
E(\tau_i) = 0, \quad \text{var}(\tau_i) = \sigma^2_i I_{\ell_{\zeta_i}}, \quad \text{cov}(\tau_i, \tau_j^T) = 0 \quad \text{for } i \neq j,
$$

$$
\text{var}(\tau) = \text{diag}(\sigma^2_0 I_{\ell_{\zeta_0}}, \sigma^2_1 I_{\ell_{\zeta_1}}, \sigma^2_2 I_{\ell_{\zeta_2}}, \ldots, \sigma^2_r I_{\ell_{\zeta_r}}),
$$

where $i = j = 0, \ldots, r$. The expectation and variance of the response can respectively be written as

$$
E(y) = X\theta, \quad \text{(2.11a)}
$$

$$
\text{var}(y) = V = \sum_{i=0}^{r} \sigma^2_i Z_i Z_i^T. \quad \text{(2.11b)}
$$

Both, the $m$ main factors in the fixed part and the $s$ main factors in the random part can be arranged in different ways. The two basic examples are the crossed and the nested layouts. The combination of such arrangements defines the structure of the corresponding part of the model, i.e. of $f(x)$ and $g(\zeta)$. Intuitively, the two structures are constructed using the same rationale. To distinguish the “structured” parts of the model, the deterministic structure will
be called the \textit{treatment structure}, and the stochastic one will be called the \textit{variability structure}. The treatment structure is usually organised in \textit{blocks} and the variability structure consists of \textit{stages}. The naming is primarily aimed at avoiding confusion as to which part of the model is being referred to, but it also helps in pointing out the different implications and role played by each part.

The formulation of the linear mixed effects model as in (2.9) implies that there is more than one variability stage. For example, the simplest case with one random factor is a two stage variability linear mixed effects model. The structure of every stage is defined in every term in (2.8). The complete variability structure of the model is therefore represented by equation (2.11b). Chapter 4 is dedicated to the topic of design structures.

\subsection*{2.2.3 Nonlinear models}

Many areas of applications often deal with systems whose variables are related in a highly complicated way that cannot be expressed as a linear combination of the regressors. Thus, linear models from previous sections are not suitable for these kind of relationships. They would require a high order model with many parameters to be estimated, and it will be very difficult to interpret them. Additionally, the computational load for the estimation may become impractical.

This sort of complex relationships may be better represented by means of \textit{nonlinear} models. These models are more complicated than linear ones, with estimation procedures only described numerically and estimators having non standard statistical properties. Ratkowsky (1983) gives a concise introduction to nonlinear regression and a detailed presentation of several models. A more theoretical presentation, but still with extensive details on applications, is given by Seber and Wild (1989). Further, a good number of application areas can be found in Davidian and Giltinan (1995) and Bates and Watts (1988).
2 Preliminaries

The idea of linearity thus far is taken with respect to the parameters of the model. Nonlinear models have the parameters appearing in a nonlinear fashion (e.g., as exponents or factors of other parameters). A nonlinear model for individual observations $i = 1, \ldots, N$, is of the form

$$y_i = f(x_i, \theta) + \epsilon_i,$$  \hfill (2.12a)

or in matrix notation

$$y = f(X, \theta) + \epsilon,$$  \hfill (2.12b)

where $f(X, \theta)$ represents the nonlinear function of the regressors that depends on the matrix $X$ and the parameter vector $\theta$. The random error $\epsilon$ term is assumed identically and independently distributed following the normal distribution, i.e., as in equation (2.3).

For example, with

$$x = (x_1 \ x_2 \ x_3),$$

$$\theta = (\theta_1 \ \theta_2 \ \theta_3 \ \theta_4),$$

$$f(x, \theta) = \theta_1 x_1 + \theta_2 x_2 + \theta_4 e^{\theta_3 x_3},$$

equation (2.12) gives,

$$y = \theta_1 x_1 + \theta_2 x_2 + \theta_4 e^{\theta_3 x_3} + \epsilon.$$

This clearly shows that the parameters appear nonlinearly.

As with the linear models, the purpose is to find the values of the parameters so that the model is a good representation of the data. For ease of interpretation
of the parameters, understanding of the statistical properties of their estimators and use of known analysis methods, it is useful to linearise a nonlinear model. The extent to which a linearisation is possible for a given nonlinear relationship is assessed by a measure of nonlinearity. As presented in (Bates and Watts, 1988) such assessment falls into two categories: a) close to linear or b) intrinsic nonlinear. For models in the first category, the estimators of the parameters, asymptotically, have good statistical properties. This implies that standard inferential procedures can be used.

The specification of a nonlinear model can be extended to include mixed effects. Longitudinal and repeated measurements studies are common areas of application. Excellent references are Davidian and Giltinan (1995) and Pinheiro and Bates (2000).

## 2.3 Design optimality theory

The discussion in Section 2.1 stresses the importance of making all possible efforts to assure that the chosen experimental design is an adequate one for the experimental goals. A good experimental plan would yield maximum information from the experimental runs with the allotted resources.

Designing experiments such that the outcome provides the highest amount of information given the conditions and resources is not an easy task. This section presents the theory of design optimality which gives tools to design experiments that fulfil the experimental purpose under the specific restrictions, as opposed to using a predefined design that does not fully fit into the experimental conditions and thus will not provide the maximum possible amount of information.

The theory of design optimality is based on the Gauss-Markov theorem, convex optimisation theory and results of matrix algebra. The material presented in
2 Preliminaries

2.3 Design optimality theory

This section is heavily based on Silvey (1980) and Atkinson et al. (2007). One fundamental component of this approach is the information matrix that summarises the information in the experiment and it is denoted by $\mathbf{M}(\mathbf{X}, \mathbf{V})$. The inverse of the information matrix \( i.e. \mathbf{M}^{-1}(\mathbf{X}, \mathbf{V}) \), is the variance-covariance matrix of the estimates of the parameters. This information matrix is a function of the structure of the design given by $\mathbf{X}$ (as defined in (2.2)) and of the variance-covariance matrix $\mathbf{V}$.

In the optimum design theory, the task is to optimise an objective function of the information matrix; it is denoted by $\Psi(\mathbf{M}(\mathbf{X}, \mathbf{V}))$. The choice of the objective function determines the design optimality criterion. The fundamental criterion is that of $\mathcal{D}$-optimality followed by $\mathcal{G}$-optimality (refer to Section 2.4).

The history of optimum experimental design can be traced back to a paper from Smith (1918). However, the topic started being effectively treated during the 1940's. $\mathcal{D}$-optimality is discussed in Wald (1943); the developments of Plackett and Burman (1946) are related to general weighting designs. $\mathcal{A}$- and $c$-optimality appear in Elfving (1952). Local optimality of nonlinear models is treated by Chernoff (1953). It was Elfving who consistently worked with the optimality approach, although the fundamental result of the General Equivalence Theorem was introduced by Kiefer and Wolfowitz (1960). A first complete account is given by Fedorov (1972), and afterwards the most serious presentation to cover general optimality criteria comes from Kiefer (1974). Relevant surveys on the topic at the time and other key results come from Pukelsheim (1980) and Silvey (1980). A good theoretical exposition is found in Pukelsheim (1993) but only focuses on fixed effects and linear models. The most extensive presentation, covering a wide range of models, is found in Atkinson et al. (2007).

Regarding the construction of optimal designs, current work is based on the foundations set by the exchange algorithm of Fedorov (1972); more refined
algorithms are found in Atkinson and Donev (1992). Recent work is concerned with
the optimality for designs with complex structures such as random blocks including
Split-Plot designs, designs involving several numbers of factors of different nature
or incorporating some linear restrictions in the model. All standard algorithms
are described in Atkinson et al. (2007). More examples of specific applications
can be found in Goos (2002), Goos and Donev (2006), Goos and Donev (2007),
and Jones and Goos (2009).

Direct application of the theory will result in designs that require a non-integer
number of experimental runs, known as continuous designs. In practice, the
actual number of observations is an integer. Use of the approximate theory as
named by Kiefer, will yield designs with finite integer number of runs, known as
discrete or exact designs. When the sample size $N$ is relatively large, exact
designs can be obtained from the continuous one by multiplying the corresponding
weight of each set of treatments by $N$ and rounding off to the nearest integer. In
most cases, to find an optimum experimental design under a particular optimality
criterion, a design construction algorithm has to be employed. Construction
algorithms will be described in Chapter 6.

Consider model (2.2) with the assumptions given in (2.3) to (2.5). A common
use of the estimated model is to make predictions of the response at arbitrary
settings of $x$, i.e. any treatment $t$. The predicted response is given by

$$
\hat{y}(x) = f^T(x|t)\hat{\theta}
$$

(2.13)

where $\hat{\theta}$ is the vector of the estimates of the fixed effects. The information matrix
of these estimates is

$$
M(X, V) = M = X^TV^{-1}X,
$$

(2.14)

obtained by using the generalised least squares method of estimation, but other
forms are possible when using different methods (see Chapter 3).

The variance of the predicted responses is called prediction variance, which serves as a measure of accuracy or usefulness of the model, and can be written as

\[
\text{var} \left( \hat{y}(x) \right) = f^T(x|t)M^{-1}f(x|t) .
\]

When doing calculations and for purposes of design comparisons, the matrix \( V \) is considered fixed. Similarly, \( \sigma^2 \) is assumed to be equal for the same experimental conditions, hence it does not need to be included in the computations; the same applies for the number of experimental runs. Further, the prediction variance can be scaled by \( \sigma^2 \) and \( N \) so that it is not affected by the relative sizes of these quantities; this is particularly useful when comparing designs. The result is called the standardised prediction variance and is written as

\[
d(x) = \frac{N}{\sigma^2} \text{var} \left( \hat{y}(x) \right) .
\]

Note that the prediction variance does not depend on \( \theta \).

According to the above framework, the observation of a random variable of interest (response), has a distribution that depends on the input factors \( x \), a function of the model parameters, \( h(\theta) \) say, as well as in the assumptions made for the random error \( \epsilon \). Then, the general experimental design problem can be stated as follows: given that an experimenter is allowed to take \( N \) independent observations of \( y \) and is interested in the precise estimation of the values of \( h(\theta) \), which set of \( N \) treatments \( t \) from the design region \( \mathcal{X} \) should be chosen?

Assume that the information matrix \( M(X, V) \) is available. As mentioned before, \( V \) does not change in the process of designing the experiment, so when it is not necessary to be explicit, it suffices to write \( M(X) \), or even simply \( M \) if the context allows it. The original problem is now seen as that of choosing
2 Preliminaries  

2.3 Design optimality theory

\( f(x|_{t_i}), \ i = 1, \ldots, N \), so that the objective function \( \Psi(M(X)) \) is large. This function is regularly chosen such that it can be associated with a statistical property. The many different options are known as *design optimality criteria* and include \( D_-, D_{S-}, A-, E-, G-, V-, \) and \( c \)-optimality, among others. Some of these will be reviewed in Section 2.4.

The flexibility of having different criteria of optimality to work with, brings the limitation that a chosen design that is optimum for one optimality criterion is, in general, not optimum for another criterion. However, the central result known as the *General Equivalence Theorem*, described in Section 2.4.3, states conditions under which a design can indeed be optimum under more than one criterion of optimality.

To tackle the optimisation problem, it is mathematically convenient to consider it in the framework of continuous designs and then adjust for the exact case. That is, search for the optimum continuous design, and then find the exact design that is still optimum in practice.

Continuous designs have an associated measure \( \xi \) on the design region \( \mathcal{X} \); informally \( \xi \) can be thought of as the equivalent of \( X \) in the exact case. Consider a design where observations are taken from \( n \) different points \( t_i \), with \( n \leq N \) and \( x \mapsto t_i \in \mathcal{X} \), then

\[
\xi = \begin{cases} 
t_1 & t_2 & \cdots & t_n \\
w_1 & w_2 & \cdots & w_n
\end{cases},
\]

where has the following conditions

\[ \int_{\mathcal{X}} \xi(dx) = 1; \quad 0 \leq w_i \leq 1, \quad \sum w_i = 1 \quad \forall i. \]

The top row of (2.17) specifies the treatments in the design and the weight of each treatment, denoted by \( w_i \), given in the bottom row. In this context each \( t_i \),
is called a *support point* and \( n \) is the number of support points in the design.

The information matrix is

\[
M(\xi) = \sum_{i=1}^{n} w_i f(x|t_i) f^T(x|t_i)
\]  

(2.18)

with standardised prediction variance

\[
d(x, \xi) = f^T(x|t) M^{-1}(\xi) f(x|t).
\]  

(2.19)

For exact designs, when the number of observations is \( N \) and \( n \) is an integer submultiple of \( N \), the design measure is

\[
\xi_N = \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ n_1/N & n_2/N & \cdots & n_n/N \end{pmatrix},
\]  

(2.20)

where the number of observations taken at \( t_i \) is \( n_i \), an integer, with the condition \( \sum_{i=1}^{n} n_i = N \). Using this terminology, exact designs have information matrix

\[
M(\xi_N) = \frac{1}{N\sigma^2} \sum_{i=1}^{n} n_i f(x|t_i) f^T(x|t_i)
\]  

(2.21)

with standardised prediction variance

\[
d(x, \xi_N) = f^T(x|t) M^{-1}(\xi_N) f(x|t) = \frac{N}{\sigma^2} \text{var}(\hat{y}(x)).
\]  

(2.22)

An optimum design that is constructed for a given model and optimality criterion is denoted by the measure \( \xi^* \), and is that design that optimises the
criterion value over the set $\Xi$ of all feasible designs, formally

$$\Psi(M(\xi^*)) = \text{opt}_{\xi \in \Xi} \Psi(M(\xi)). \tag{2.23}$$

What is needed now are ways to search for and identify which of all possible designs is optimum, a process called design construction. This process relies on computer searches, therefore design construction algorithms have to be used; some of these are reviewed in Chapter 6.

## 2.4 Optimality criteria and the general equivalence theorem

It has been established that the problem of design optimality is that of optimising a target function related to the information of the experiment. Now it is convenient to define the set target functions $\Psi(M(\xi))$, known as design optimality criteria. Two of the most commonly used criteria, $D$- and $G$-optimality are presented in this section leading to the discussion of the General Equivalence Theorem.

### 2.4.1 $D$-optimality

The best known design criterion is $D$-optimality. It minimises the generalised variance of the parameter estimators $\hat{\theta}$. The usefulness of this criterion is that it minimises the volume of the ellipsoidal confidence region of $\hat{\theta}$. In addition, optimum designs under this criterion usually perform well with respect to other optimality criteria, and they are also invariant to linear transformations of the design matrix. Therefore, $D$-optimum designs are insensitive to the scale of measurement of the variables, which adds to their convenience.
The objective function is

\[ \Psi(M(\xi)) = \det M(\xi) = |M(\xi)| \quad (2.24) \]

and the \( \mathcal{D} \)-optimality criterion is denoted by

\[ \mathcal{D} = \min |M^{-1}(\xi)|, \quad (2.25a) \]

which is the minimisation of the determinant of the variance-covariance matrix of \( \hat{\theta} \), or equivalently,

\[ \mathcal{D} = \max |M(\xi)|, \quad (2.25b) \]

which is the maximisation of the determinant of the information matrix. The resulting figure is called the criterion value. A more general description is as follows. Let the eigenvalues of the information matrix \( M(\xi) \) be \( \lambda_1, \ldots, \lambda_p \) then, the corresponding eigenvalues of \( M^{-1}(\xi) \) are \( 1/\lambda_1, \ldots, 1/\lambda_p \). The \( \mathcal{D} \)-optimality criterion can now be written as

\[ \mathcal{D} = \min \prod_{i=1}^{p} \frac{1}{\lambda_i}. \quad (2.25c) \]

Using these eigenvalues, other relevant criteria can be stated as follows:

\textbf{\( \mathcal{A} \)-optimality} The objective function is \( \Psi(M(\xi)) = \sum_{i=1}^{p} 1/\lambda_i \) and the criterion minimises the sum, or average, of the variances in \( M^{-1}(\xi) \):

\[ \mathcal{A} = \min \sum_{i=1}^{p} \frac{1}{\lambda_i}; \quad (2.26) \]

\textbf{\( \mathcal{E} \)-optimality} The objective function is \( \Psi(M(\xi)) = \max_i \frac{1}{\lambda_i} \) and the criterion minimises the variance of \( a^T\hat{\theta} \) with \( a^T a = 1 \), i.e. the least precise estimate
2 Preliminaries

2.4 Optimality criteria and the general equivalence theorem

of a linear combination of the model parameters:

$$\mathcal{E} = \min \max_i \frac{1}{\lambda_i}.$$  \hspace{1cm} (2.27)

$\mathcal{A}$- and $\mathcal{E}$-optimum designs are well suited for experiments with all qualitative factors. In such case, the criteria can be redefined, for example to express the interest in particular contrasts. $\mathcal{D}$-optimum designs are a better fit for experiments with quantitative factors, as reasoned in Atkinson et al. (2007).

2.4.2 $\mathcal{G}$-optimality

A $\mathcal{G}$-optimum design minimises the maximum of the prediction variance over the design region $\mathcal{X}$. The objective function can be written as

$$\Psi(M(\xi)) = \max_{\mathbf{x} \in \mathcal{X}} d(\mathbf{x}, \xi) = \bar{d}(\xi)$$

where $\mathbf{x} \mapsto \mathbf{t}$ and $d(\mathbf{x}, \xi)$ is the standardised prediction variance as introduced in (2.19). The design criterion is

$$\mathcal{G} = \min \bar{d}(\xi).$$  \hspace{1cm} (2.28)

The optimum design measure is $\xi_\mathcal{G}^*$, and when the design is continuous $\bar{d}(\xi_\mathcal{G}^*) = p$. This makes the $\mathcal{G}$-optimum design also $\mathcal{D}$-optimum, a condition known as design equivalency.

2.4.3 The general equivalence theorem

Experiments are planned for specific purposes and consequently an experimental design is constructed such that it yields more efficient results for the chosen property than any other design. In the optimality framework, one design criterion
will result in one experimental design and an alternative criterion may need a different design. However, there is a theoretical result which states some conditions for a continuous optimum design that can make it equally useful under more than one criterion of optimality, i.e. equivalent.

The discussion in Section 2.3 leading to equation (2.23) points out that the optimisation of a design is done in reference to a function of the information matrix \( \Psi(M(\xi)) \), which is to be minimised since it represents a general measure of imprecision. The task can also be formulated as a maximisation problem, depending on the context.

The General Equivalence Theorem states the following. Assume that the design region \( \mathcal{X} \) is compact and that \( \Psi(M(\xi)) \) is convex and differentiable with first derivative \( \phi(x, \xi) \). There are three conditions on \( \xi^* \) that are equivalent, with a fourth condition being a consequence of the third one; these conditions are:

1. The design \( \xi^* \) minimises \( \Psi(M(\xi)) \).
2. The design \( \xi^* \) maximises the minimum over \( \mathcal{X} \) of \( \phi(x, \xi) \), \( x \mapsto t \).
3. The minimum over \( \mathcal{X} \) of \( \phi(x, \xi) = 0 \) is achieved at the points of support of the design.
4. For any non-optimum design \( \xi \) the minimum over \( \mathcal{X} \) of \( \phi(x, \xi) < 0 \).

To understand the idea, consider \( D \)-optimality that minimises the generalised variance of the estimator of the model parameters. The objective function (2.24) can be expressed alternatively as

\[
\Psi(M(\xi)) = \log |M^{-1}(\xi)| = - \log |M(\xi)|
\]

such that the \( D \)-optimum design \( \xi^* \) maximises the determinant of \( M(\xi) \). It follows
that
\[ \phi(x, \xi) = p - d(x, \xi), \]
and given that \( \phi(x, \xi^*) \geq 0 \), the inequality
\[ d(x, \xi^*) \leq p \]
holds. It is clear that by the General Equivalence Theorem, a \( D \)-optimum design is also \( G \)-optimum (see Section 2.4.2) since it maximises the determinant of \( M(\xi) \) but it also minimises the maximum of the prediction variance \( d(x, \xi), x \mapsto t \) over the design region.

Proofs of the theorem can be found in Fedorov (1972), Silvey (1980) and Pázman (1986). The theorem may be considered a consequence of the assumed convexity of \( \Psi(M(\xi)) \) that makes its directional derivatives in the direction of \( \bar{\xi} \), the design with unit mass at the support point \( t \), to be zero at the minimum, which guarantees that if the minimum is found it will be global. The theorem can be used to construct and check designs.

When two designs have very similar or even equal value of the optimality criterion, the designs can be discriminated based on the maximum number \( n \) of support points required by each design. An upper limit for \( n \) can be obtained with the following rationale. From (2.18) the information matrix can be formed as a weighted sum of information matrices of individual design points, \( i.e. \) of rank-one matrices, which are evaluated at distinct \( \bar{\xi}_i \). By taking advantage of this additivity of \( M(\xi) \), which is a \( p \times p \) matrix, it can be shown using Carathéodory’s theorem (Silvey, 1980), that an optimum design requires no more than \( p(p + 1)/2 \) support points. Different designs can now be compared in terms of \( n \), where smaller criterion value is more desirable.

The General Equivalence Theorem in general does not apply for exact designs.
\( \xi_N \). More specifically, a given design with \( N \) observations that is optimum for one criterion, will not be optimum under a second criterion. The optimum design for the second criterion will require a different value of \( N \). This makes the equivalency hard to establish. In most cases, computer searches are required to find exact optimum designs for a specific \( N \). The search has to ensure that the optimisation of the target function does not correspond to local minima of other designs having similar values of \( \Psi(M(\xi_N)) \). Algorithms for construction of designs are presented in Chapter 6.

### 2.4.4 \( D_A \)-optimality (Generalised \( D \)-optimality)

When the experimenter is only interested in a subset \( s \), which is a linear combination of the model parameters \( \theta \), instead of all \( p \), a more general concept of \( D \)-optimality is used. These \( s \) linear combinations are the elements of \( A^T \theta \), where matrix \( A \) is \( p \times s \) with rank \( s < p \). The covariance matrix of the subset of interest is

\[
M_A(\xi) = A^T M^{-1}(\xi) A.
\]

A Generalised \( D \)-optimum design has \( \Psi(M(\xi)) = \log |M_A(\xi)| \) and the criterion is specified as

\[
D_A = \min \log |A^T M^{-1}(\xi) A|, \tag{2.29}
\]

and the predictive variance function is

\[
d_A(x, \xi) = f^T(x|\xi) M^{-1} A (A^T M^{-1} A)^{-1} A^T M^{-1} f(x|\xi). \tag{2.30}
\]

Define now \( \xi_{D_A}^* \) as the continuous \( D_A \)-optimum design. The value that maximises (2.30) using \( \xi_{D_A}^* \) is \( s \) and the corresponding maxima occur at the support points, therefore the General Equivalence Theorem is satisfied.
2.4.5 $D_S$-optimality

The particular case of $D_A$-optimality when $A = (I_s \ 0)^T$ is called $D_S$-optimality, and it is a recommended criterion when the interest is in the precise estimation of a subset $s$ of $\hat{\theta}$; the remaining parameters are treated as nuisances. By rewriting the model as

$$y = f^T(x|t)\theta = f_1^T(x|t)\theta_1 + f_2^T(x|t)\theta_2,$$

then $\theta_1$ is the vector of the parameters of interest and $\theta_2$ contains the $p-s$ nuisance parameters. Allow the information matrix to be partitioned as

$$M(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{12}^T(\xi) & M_{22}(\xi) \end{bmatrix}.$$

The variance-covariance matrix of the estimates $\hat{\theta}_1$ is $M_{11}^{-1}(\xi)$, the $s \times s$ submatrix of $M^{-1}(\xi)$ in the (1, 1) position. Using results of matrix algebra for the inverse of a partitioned matrix

$$M_{11}^{-1}(\xi) = \{M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}^T(\xi)\}^{-1},$$

and so the $D_S$-criterion of optimality for $\theta_1$ is

$$D_S = \max \left| M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}^T(\xi) \right| = \max \left| \frac{M(\xi)}{M_{22}(\xi)} \right|,$$

and the prediction variance is

$$d_s(x, \xi) = f^T(x|t)M^{-1}(\xi)f(x|t) - f_2^T(x|t)M_{22}^{-1}(\xi)f_2(x|t).$$
Given that $D_S$-optimality is a special case of $D_A$-optimality with

$$d_s(x, \xi^\ast_{D_S}) \leq s, \quad (2.35)$$

where $\xi^\ast_{D_S}$ is the $D_S$-optimum design and with equality only at the support points. This fact indicates that the General Equivalence Theorem still holds.

### 2.4.6 $c$-optimality

When there is interest in minimising the variance of the linear combination $c^T \theta$, where $c$ is a known vector of constants, the criterion of choice is $c$-optimality.

The formal statement for this criterion is

$$c = \min \operatorname{var}(c^T \hat{\theta}) \propto \min c^T M^{-1}(\xi)c, \quad (2.36)$$

where $c$ is $p \times 1$. For the General Equivalence Theorem to hold in this case

$$\{f^T(x|t)M^{-1}(\xi^\ast)c\}^2 \leq c^T M^{-1}(\xi^\ast)c, \quad (2.37)$$

when $x \in \mathcal{X}$.

### 2.4.7 $\mathcal{V}$-optimality

If the interest is in the average variance over a region $\mathcal{R}$, and this averaging is done with respect of a probability distribution $\mu$ on $\mathcal{R}$, the adequate optimum design should follow the criterion

$$\mathcal{V} = \min \int_\mathcal{R} f^T(x|t)M^{-1}(\xi)f(x|t)\mu(dx) = \min \int_\mathcal{R} d(x, \xi)\mu(dx). \quad (2.38)$$
This criterion is used in practice when comparing designs found by other criteria. It is also called $\mathcal{I}$-optimality.

In general, numerical approximations over a grid in $\mathcal{R}$ are commonly needed to compute the $\mathcal{V}$-criterion value. However, design construction algorithms have been developed that do not require this numerical approximation. By writing the $\mathcal{V}$-criterion as a particular case of linear optimality as in Meyer and Nachtsheim (1995), Jones and Goos (2012) developed an analytic expression for the average variance when the design region is $[-1, 1]^m$ and for a full quadratic model. They proceed to use it in an coordinate exchange algorithm to find $\mathcal{I}$-optimum designs.

### 2.5 Summary

The majority of the terms and concepts used throughout the thesis were introduced in this chapter. This was done by assuming a generic system that reflects the characteristics and relationships of the potentially different sources of variability involved in the generation of experimental data.

In addition to the fundamental concepts in experimental design, the definitions most directly relevant are the distinction between fixed and random factors and their corresponding effects. The discussion regarding random factors that are controllable or non-controllable is key since it leads to the important concept of variance components. The main body of the presented work is about extending the standard theory of statistical design optimality to models with random effects. Such models are a natural choice for studying variance components, which is the focus here.

A second set of important concepts includes the ideas of easy to change (EC) and hard to change (HC) factors. These in turn introduced the concepts of crossed (C) and nested layouts (N). Such layouts allow to proceed to the concept
of hierarchical design structure and in particular the key distinction between the
treatment structure and the variability structure of a design.

Common statistical models were introduced with emphasis on the linear mixed
effects model. At a glance, the linear models are presented as is customary in the
statistical literature, however there has been a very conscious effort to introduce a
more explicit notation. In particular, the vector $x$ is used exclusively for the main
input factors and to each element of $x$ there is an associated vector $\ell_x$, specifying
the levels. Subsequently, a vector $t$ of treatments is defined using the different
elements of $\ell_x$. Therefore, $x|_t$ denotes a specific treatment and $f(x|_t)$ is its
corresponding polynomial expansion based on the function $f(\cdot)$. Thus the design
matrix $X$ is constructed in a row wise fashion using the elements of $f(x|_t)$ for
all the treatments involved. For the random part of the mixed model, equivalent
definitions as given for $\zeta$, $\ell_\zeta$, $\zeta|_t$, $g(\zeta|_t)$ and $Z$. This notation is more flexible
and more detailed, with the added benefit of facilitating the transition from the
abstract model to its implementation in a computer algorithm. Additionally, the
deterministic part of the mixed model is explicitly associated with the treatment
structure of a design and the stochastic part with the variability structure.

Lastly, the fundamentals of the theory of design optimality are presented.
Instead of the common $x$ notation for support points, the $t$ notation is also used
here. This uniforms the presentation and also gives a subtle generalisation of the
idea of a support “point”. As reproduced here, as in most of the literature, the
theory models the random effects as nuisance parameters, and the focus remains
the estimation of the fixed effects.
Parameter estimation

The previous chapter described the need for experimental design and pointed out that a reasonable statistical model has to be defined before the experiment can be designed. With the additional knowledge of the method of estimation to be used, an optimum experiment can indeed be designed. The link between an estimation procedure and the optimum experimental design is the information matrix.

Estimators obtained under different methods posses different statistical properties. The variances of those estimates is of interest here. They are obtained via the information matrix, which is written differently depending on the chosen method of estimation. Relevant methods are described next. Each one correspond to a principle used to explain the variability of data. The most useful here are the least squares and the maximum likelihood. Different variations of estimation procedures can be derived from these two principles.

3.1 Model fitting

Parameter estimation will be considered here as part of the larger process of model fitting. A simplified model fitting workflow includes model definition, parameter estimation and residual analysis. Statistical models were reviewed in Chapter 2.
Parameter estimation is covered in this chapter. Statistical analysis is presented in standard design of experiments texts such as Box and Draper (2007), Hinkelmann and Kempthorne (2008) and Milliken and Johnson (2009). Residual analysis is not reviewed explicitly here. However, the concept of a residual is important. Residuals are used to assess the goodness of fit, but they are also used in the development of statistical theory and methodology.

The standard notation $\hat{\theta}$ has already been used in Chapter 2 to denote the vector of estimated values of the fixed model parameters. Similarly, $\hat{\sigma}^2$ refers to the estimates of the variance components. When these estimates are used in the model equation, with the same treatments used in the experiment i.e. $t$, the resulting values are called fitted values and are denoted by $\hat{y}$. How accurately $\hat{y}$ describes $y$ is known as the goodness of fit. This in turn can be assessed by analysing the residuals of the fitted model which are defined as

$$ e = y - \hat{y}. $$

(3.1)

Residuals are the basis for obtaining estimates of the random error. Also, the common assumptions of normality and independence are judged using residual analysis.

### 3.2 Fixed effects models

The fixed effects model is presented in Section 2.2 as equation (2.2). It is repeated now for convenience. The model equation is

$$ y = X\theta + \epsilon, $$

(3.2)
with the underlying assumptions

$$E(\epsilon_i) = 0, \quad \text{var}(\epsilon_i) = \sigma^2, \quad (3.3a)$$

$$\text{cov}(\epsilon_i, \epsilon_j) = 0 \quad i \neq j, \quad \epsilon \sim N(0_N, V). \quad (3.3b)$$

where $i = j = 0, \ldots, N$, and all the terms are as defined in Section 2.2.1.

### 3.2.1 Analysis of variance (ANOVA)

This method is key for understanding the fundamentals of extracting information from the data. That is one reason to be presented here. The second reason is to highlight the fact that different methods of estimation can give a different information matrix, mainly for the variance components. The ANOVA method is widely used since it is a natural way of analysing comparative experiments (see Table 2.1 on page 37); it is attributed to Sir. R. A. Fisher.

The primary idea behind ANOVA is the partitioning of the total variability of the model into its different sources. This is done via calculation of different sum of squares, each one associated with a source of variability. The decomposition is tabulated in what is known as the ANOVA table. The desired effect is that by isolating the elements of the variability it is possible to assess their influence. This would enable an experimenter to identify where to direct efforts to reduce variation.

To show a straightforward use of the ANOVA method, consider an example of model (3.2) when the response depends on one factor with $a$ levels, known as the
1-way classification or 1-factor layout\(^1\), which is written as

\[ y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad (3.4) \]

where \( \mu \) is the overall mean and \( \alpha_i \) is the \( i \)th treatment effect, \( i = 1, \ldots, a \) with \( j = n_1, \ldots, n_a \), thus \( n_i \) is the number of observations per treatment; \( \sum_{i=1}^{a} n_i = N \).

Linear models written in this fashion are referred to as *means models*, and are extensively used for comparison of treatment effects. This is in contrast with *regression* or *response surface* models, which are used for describing the behaviour or shape of the response variable as a function of the regressors. For simplicity consider the balanced case of model (3.4), so that \( n_i = n \ \forall i \) and \( N = an \). In addition,

\[
\begin{align*}
y_h &= \sum_{j=1}^{n} y_{ij}, \\
y. &= \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \\
y_{i.} &= n_i \\
y.. &= \frac{N}{N} \\
y_{i.} &= \frac{y_{i.}}{n_i} \\
y.. &= \frac{y.}{N}.
\end{align*}
\]

The decomposition of sources of variability and corresponding sum of squares is shown in Table 3.1. Note that each sum of squares represents a sample variance. These estimated variances come from 1) the difference between the treatment means and the overall mean (SSA), 2) the difference between the observations of a treatment and the treatment mean (SSe), and 3) the difference between an observation and the overall mean (SST). Therefore, SST represents the total variability, SSA represents the variability due to the treatments and SSe is the variability within each treatment which is attributed to the random error. It also

\(^1\)The term “\( m \)-way classification” is commonly used in this context of comparison of treatment means, and implies fully crossed factors. The “\( m \)-factor layout” terminology is more general and will be used onwards. Refer to Section 4.3 for more details.
Table 3.1 1-factor layout Sum of Squares

<table>
<thead>
<tr>
<th>Source of Variability</th>
<th>Sum of squares (SS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor (A)</td>
<td>$SS_A = n \sum_{i=1}^{a} (\bar{y}_i - \bar{y})^2$</td>
</tr>
<tr>
<td>Error (e)</td>
<td>$SS_e = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2$</td>
</tr>
<tr>
<td>Total (T)</td>
<td>$SST = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y})^2$</td>
</tr>
</tbody>
</table>

follows that

$$SST = SSA + SSE.$$  

The number of independent elements involved in the calculation of the sum of squares is called the number of degrees of freedom (d.f). The sum of squares divided by its corresponding d.f yields the mean squares (MS). The expectation of the MS, the expected mean squares (EMS), also need to be evaluated. When Table 3.1 is supplemented with these additional calculations, an standard ANOVA table is obtained as shown in Table 3.2.

Table 3.2 ANOVA table for the 1-factor layout model

<table>
<thead>
<tr>
<th>$SS$</th>
<th>d.f</th>
<th>Mean Squares (MS)</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS_A$</td>
<td>$a-1$</td>
<td>$SS_A/(a-1)$</td>
<td>$\sigma^2 + \frac{n}{a-1} \sum_{i=1}^{a} (\bar{y}_i - \bar{y})^2$</td>
</tr>
<tr>
<td>$SS_e$</td>
<td>$a(n-1)$</td>
<td>$SS_e/a(n-1)$</td>
<td>$\frac{SS_e}{a(n-1)}$</td>
</tr>
<tr>
<td>$SST$</td>
<td>$N-1$</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

For variance components, the ANOVA method of estimation consists of equating a given mean squares to its expected value and solving for the corresponding variance. In model (3.4) there is only one variance component, which corresponds to the random error $\sigma^2_\epsilon$. Using the ANOVA method, its estimator is

$$\hat{\sigma}^2_\epsilon = MSe = \frac{SS_e}{a(n-1)},$$  (3.5)

Given the independence and normality assumptions for the random error in (3.3), the ANOVA estimator has a $\chi^2$ distribution with $a(n-1)$ degrees of freedom.
This follows from general results for the sampling distribution of quadratic forms applied to the mean squares (see, e.g., Rao (1973)). For ANOVA models, the specific results are presented in Appendix A.2. Using the latter, it is possible to find an estimator for the variance of the estimated variances. For ANOVA in general, unbiased estimators can be found. Considering equation (3.5),

\[
\text{var} \left( \frac{SSe}{EMSe} \right) = 2 \, a(n - 1),
\]

\[
\text{var}(MSe) = \frac{\text{var}(SSe)}{(a(n - 1))^2} = \frac{2 \, EMSe^2}{a(n - 1)},
\]

and additionally

\[
\text{var}(MSe) = \text{E}(MSe^2) - EMSe^2.
\]

By equating both expressions for \( \text{var}(MSe) \),

\[
\frac{\text{E}(MSe^2)}{a(n - 1) + 2} = \frac{EMSe^2}{a(n - 1)}.
\]

Now write

\[
\text{var}(\hat{\sigma}_e^2) = \text{var}(MSe) = \frac{2 \, \sigma_e^4}{a(n - 1)},
\]

whose unbiased estimator is

\[
\hat{\text{var}}(\hat{\sigma}_e^2) = \frac{2 \, \hat{\sigma}_e^4}{a(n - 1) + 2}.
\]

The procedure described above is used in the same manner when the model includes more than one variance component, e.g. a 2-factor crossed layout random
model. It is imperative to be able to find estimates of the variances of the estimators of the variance components in order to find optimum designs. These estimated variances and covariances of the parameters estimates can be arranged in matrix form to get the corresponding variance-covariance matrix of the model and consequently the information matrix.

It must be noted that the ANOVA method and the 1-factor model are presented in this thesis mainly to introduce fundamental ideas on how to work with different variability sources and also for completeness. In fact, to take advantage of the theory of optimum design of experiments, information matrices are obtained using other estimation methods that can also allow the estimation of response surfaces while keeping desirable statistical properties for the estimators, e.g., least squares. In particular, maximum likelihood methods are primarily used. The former is described next and the latter in Section 3.3.1. Both approaches require the use of numerical methods in most practical cases.

### 3.2.2 Least squares estimation

Least squares estimation is the most commonly used method for estimating the fixed effects in a model. It is normally not used in the means model e.g. (3.4), but mainly in general regression or response surface settings. This method consists of minimising the discrepancy between the measured response and the assumed model. For the general fixed effects model (3.2)-(3.3) this discrepancy is described by the weighted error sum of squares defined as

\[
SSe = e^T V^{-1} e = (y - X\theta)^T V^{-1} (y - X\theta),
\tag{3.6}
\]
(different from SS\textsubscript{e} in the ANOVA method) which is a quadratic form of the theoretical residuals \( e = (y - X\theta) \) and has the property

\[
E(SS\textsubscript{e}) = V.
\]

The process of minimising \( SS\textsubscript{e} \) to obtain estimates of the vector of parameters \( \theta \) is known as the \textit{generalised least squares} (GLS) method. The minimisation is done by differentiating (3.6) with respect to \( \theta \), setting the derivative equal to zero and trying to solve for \( \theta \). This sequence, using the linear algebra results in Appendix A.1, yields

\[
(X\theta)^T \frac{\partial}{\partial \theta} SS\textsubscript{e} = \frac{\partial}{\partial \theta} \left[ -2(X\theta)^TV^{-1}y + (X\theta)^TV^{-1}(X\theta) \right]
\]

\[
= -2X^TV^{-1}y + 2X^TV^{-1}X\theta,
\]

\[
X^TV^{-1}X\theta = X^TV^{-1}y, \tag{3.7}
\]

which are known as the \textit{normal equations} and are then solved to give the parameter estimates

\[
\hat{\theta} = (X^TV^{-1}X)^{-1}X^TV^{-1}y. \tag{3.8}
\]

The GLS method does not require any distributional assumption for the model. In (3.8) different forms of \( V \) can be used, like for heteroscedastic or correlated observations as in equations (2.7) and (2.8). Further, when assumptions (3.3b) are fulfilled, the GLS estimates have two interesting properties:

i) By the Gauss-Markov theorem, the estimators are normally distributed, unbiased and have minimum variance with variance-covariance matrix

\[
\text{var}(\hat{\theta}) = (X^TV^{-1}X)^{-1}. \tag{3.9}
\]
ii) \( \hat{\theta} \) is also the \textit{maximum likelihood estimate} (MLE) of \( \theta \). Maximum likelihood estimation is described in Section 3.3.1.

When the simplest form of the variance-covariance matrix is an appropriate choice, \( i.e. \ V = \sigma^2 I \), the GLS estimator \( \hat{\theta} \) reduces to

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

which is known as the the ordinary least squares (OLS) estimator, and in light of the Gauss-Markov theorem, it has the same properties as the GLS estimates.

For designing an optimum experiment in some sense, the information matrix of the parameters estimates is needed. It is obtained as the inverse of the variance-covariance matrix given by (3.9), that is

\[
M(X, V) = X^T V^{-1} X,
\]

(3.10)

for the GLS estimator, and it obviously simplifies to

\[
M(X) = X^T X
\]

(3.11)

for the OLS estimator.

### 3.3 Mixed effects models

The mixed model is an extension of model (3.2) where there random part is more complex than having a single random term. It is formally introduced in Section 2.2.2, where all its elements are defined, and it is repeated here for
convenience. The model equation, using (2.10), is

\[ y = X\theta + \sum_{i=0}^{r} Z_i\tau_i. \]  (3.12)

For notational consistency, recall that \( \ell_{\xi_0} = N, \tau_0 = \epsilon \) and therefore \( \sigma_0^2 = \sigma_\epsilon^2 \).

Using assumptions (2.6)

\[ E(\tau_i) = 0 \quad \text{var}(\tau_i) = \sigma_i^2 I_{\ell_{\xi_i}}, \]  (3.13a)

\[ \text{cov}(\tau_i, \tau_j) = 0 \quad \text{for } i \neq j. \]  (3.13b)

The model mean and variance-covariance matrix can be written, respectively, as

\[ E(y) = X\theta \]  (3.14)

\[ \text{var}(y) = V = \sum_{i=0}^{r} \sigma_i^2 Z_i Z_i^T. \]  (3.15)

In a mixed model the estimation procedure must be capable of estimating the \( p \) fixed parameters in \( \theta \) and the \( r + 1 \) variance components \( \sigma^2 \) in matrix \( V \). The preferred estimation methods are likelihood based followed by Bayesian methods. However, least squares is still useful for the fixed effects.

### 3.3.1 Maximum likelihood estimation

The maximum likelihood method of estimation (MLE) consists of maximising the likelihood function of the model given the observed data \( y \) with respect to the parameters \((\theta, \sigma^2)\). The maximisation is done by differentiating the likelihood function, or more conveniently, the log-likelihood with respect to the parameters of interest. The resulting derivatives are set to zero and then solved for either \( \theta \)
or $\sigma_i^2, \ i = 0, \ldots, r$.

This method is chosen based on its good statistical (asymptotic) properties, from which a useful consequence is that, asymptotically, the information matrix of the estimators is always available. Additionally, the formulation is very general, so different types of models can be evaluated without major adjustments in the methodology.

Consider the model specified by (3.12) and (3.14), but allow a general variance-covariance matrix $V$. Assuming the response variable to be normally distributed, the model can be stated as $y \sim N_N(X\theta, V)$ which gives the likelihood function

$$L = L(\theta, V|y) = \frac{\exp\left(-\frac{1}{2}(y - X\theta)^T V^{-1}(y - X\theta)\right)}{(2\pi)^{\frac{1}{2}N} |V|^{\frac{1}{2}}}, \quad (3.16)$$

and the log-likelihood is

$$l = \log(L) = -\frac{1}{2}N \log(2\pi) - \frac{1}{2} \log(|V|) - \frac{1}{2}(y - X\theta)^T V^{-1}(y - X\theta). \quad (3.17)$$

Derivatives of (3.17) can be found using the matrix algebra results in Appendix A.1. The first derivative with respect to $\theta$ is

$$\frac{\partial l}{\partial \theta} = \frac{\partial}{\partial \theta} \left[ -\frac{1}{2}(-y^TV^{-1}X\theta - \theta^TX^TV^{-1}y + \theta^TX^TV^{-1}X\theta) \right]$$

$$= X^TV^{-1}y - X^TV^{-1}X\theta, \quad (3.18)$$

and with respect to $\sigma^2$ is

$$\frac{\partial l}{\partial \sigma_i^2} = \frac{\partial}{\partial \sigma_i^2} \left[ -\frac{1}{2} \log(|V|) - \frac{1}{2}(y - X\theta)^T V^{-1}(y - X\theta) \right]$$
The maximum likelihood equations are obtained when (3.18) and (3.19) are equated to zero. When solving them for the parameters, the values that are obtained maximise the log-likelihood, but strictly speaking, they are not always the maximum likelihood estimates. For this to happen, the second derivatives have to be checked and the solutions must not lie on the boundary of the parameters space. However, with the model assumptions that have been made, these solutions will indeed yield the MLEs.

The MLEs of the fixed parameters of model (3.12) can be written explicitly as

\[
\hat{\theta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y, \tag{3.20}
\]

which reveals, by comparing equations (3.8) and (3.20), that under normality and i.i.d assumptions the MLE is equivalent to the GLS estimator for \( \theta \). For the vector of variance components \( \sigma^2 \), the equation

\[
\text{tr}(V^{-1} \frac{\partial V}{\partial \sigma_i^2}) = (y - X\theta)^T V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} (y - X\theta) \tag{3.21}
\]

is nonlinear with respect to the elements \( \sigma_i^2 \), since they appear on both sides of the equation through \( V^{-1} \) and possibly in \( \partial V / \partial \sigma_i^2 \), and consequently it has to be solved numerically. Therefore, in general, there are no explicit expressions for the estimates \( \hat{\sigma}^2 \). Naturally, when \( \hat{\sigma}^2 \) is used in \( V \), the estimated variance-covariance matrix \( \hat{V} \) is obtained.

Commonly used iterative procedures for nonlinear estimation can be applied in this context using equations (3.21) and (3.20).
3.3.2 MLE information matrix

The information matrix for a mixed model also consists of two main parts, one for $\theta$ and one for $\sigma^2$. For the MLE method, what it is obtained is the Fisher information matrix defined as

$$
M \begin{bmatrix} \theta \\ \sigma^2 \end{bmatrix} = -E \begin{bmatrix} \frac{\partial^2 l}{\partial \theta \partial \theta^T} & \frac{\partial^2 l}{\partial \theta \partial \sigma^2} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \theta^T} & \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix},
$$

that is, the negative of the expectation of the second derivatives of the log-likelihood function with respect to both vectors of parameters. The first derivatives of $l$ with respect to $\theta$ and $\sigma^2$ have already been calculated in equations (3.18) and (3.19), respectively. To obtain the second derivatives, the formulae from Appendix A.1 will again be used.

Starting with equation (3.18), the second derivative of the log-likelihood with respect to the fixed parameters is easily found to be

$$
\frac{\partial^2 l}{\partial \theta \partial \theta^T} = \frac{\partial l}{\partial \theta} \left[ X^T V^{-1} y - X^T V^{-1} X \theta \right] = -X^T V^{-1} X.
$$

The two cross partial derivatives (off diagonal elements) in the right hand side of (3.22) are the same by symmetry, in other words

$$
\frac{\partial^2 l}{\partial \theta \partial \sigma^2} = \frac{\partial^2 l}{\partial \sigma^2 \partial \theta^T}.
$$

From the left hand side of the expression above and using (3.18), the corresponding
second derivative is

\[
\frac{\partial^2 l}{\partial \theta \partial \sigma^2_i} = \frac{\partial l}{\partial \sigma^2_i} \left[ x^T v^{-1} y - x^T v^{-1} x \theta \right] = \frac{\partial l}{\partial \sigma^2_i} \left[ x^T v^{-1} (y - x \theta) \right] \\
= - \left[ x^T v^{-1} \frac{\partial v}{\partial \sigma^2_i} v^{-1} (y - x \theta) \right].
\] (3.24)

Making use of (3.19), the second derivative of the log-likelihood with respect to the variance components is

\[
\frac{\partial^2 l}{\partial \sigma^2_i \partial \sigma^2_j} = \frac{\partial l}{\partial \sigma^2_j} \left[ -\frac{1}{2} \left( \text{tr} \left( v^{-1} \frac{\partial v}{\partial \sigma^2_i} \right) - (y - x \theta)^T v^{-1} \frac{\partial v}{\partial \sigma^2_i} v^{-1} (y - x \theta) \right) \right] \\
= -\frac{1}{2} \text{tr} \left( - v^{-1} \frac{\partial v}{\partial \sigma^2_j} v^{-1} \frac{\partial v}{\partial \sigma^2_i} \right) \\
+ \frac{1}{2} \left[ -2(y - x \theta)^T v^{-1} \frac{\partial v}{\partial \sigma^2_i} v^{-1} \frac{\partial v}{\partial \sigma^2_j} v^{-1} (y - x \theta) \right] \\
= -\frac{1}{2} \text{tr} \left( - v^{-1} \frac{\partial v}{\partial \sigma^2_i} v^{-1} \frac{\partial v}{\partial \sigma^2_j} \right) \\
- \left[ (y - x \theta)^T v^{-1} \frac{\partial v}{\partial \sigma^2_i} v^{-1} \frac{\partial v}{\partial \sigma^2_j} v^{-1} (y - x \theta) \right].
\] (3.25)

Evaluating the expectation of (3.23) and multiplying by -1, gives

\[-E(-x^T v^{-1} x) = x^T v^{-1} x.\]

For the cross derivative of (3.24), the result is

\[-E \left[ \frac{\partial^2 l}{\partial \theta \partial \sigma^2_i} \right] = -E \left( - \left[ x^T v^{-1} \frac{\partial v}{\partial \sigma^2_i} v^{-1} (y - x \theta) \right] \right) \\
= x^T v^{-1} \frac{\partial v}{\partial \sigma^2_i} v^{-1} E(y - x \theta) = 0, \] (3.26)

which follows from (3.14) since it gives \(E(y - x \theta) = 0\). An obvious consequence
of (3.26) is
\[-E \left[ \frac{\partial^2 l}{\partial \sigma_i^2 \partial \theta^T} \right] = 0.\]

The corresponding result for the second derivatives with respect to the variance components only, using (3.25), is
\[-E \left[ \frac{\partial^2 l}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = -E \left( -\frac{1}{2} \text{tr} \left( -V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} \frac{\partial V}{\partial \sigma_j^2} \right) \right. \]
\[\left. \quad - \left[ (y - X\theta)^T V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} \frac{\partial V}{\partial \sigma_j^2} V^{-1} (y - X\theta) \right] \right) \]
\[= -\frac{1}{2} \text{tr} \left( -V^{-1} \frac{\partial V}{\partial \sigma_j^2} V^{-1} \frac{\partial V}{\partial \sigma_i^2} \right) + \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} \frac{\partial V}{\partial \sigma_j^2} \right) \]
\[= \frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} \frac{\partial V}{\partial \sigma_j^2} \right). \quad (3.27)\]

This completes the elements on the right hand side of (3.22). First, note that the information matrix is block diagonal due to (3.26). This implies that the entries in the main diagonal are independent. It is then possible to write information matrices for the fixed part and for the random part separately. The (1, 1) element corresponds to the information matrix of the fixed parameters \( \theta \), and it is expressed as
\[M(\theta) = X^T V^{-1} X. \quad (3.28)\]

Similarly, the (2, 2) element refers to the information matrix of the variance parameters \( \sigma^2 \) whose corresponding expression is
\[M(\sigma^2) = \frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} \frac{\partial V}{\partial \sigma_j^2} \right). \quad (3.29)\]
The complete MLE information matrix of the general mixed model (3.12) is

\[
M \begin{bmatrix} \theta \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} X^T V^{-1} X & 0 \\ 0 & \frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2} V^{-1} \frac{\partial V}{\partial \sigma^2} \right) \end{bmatrix}.
\]

(3.30)

It also holds that

\[
\text{var} \begin{bmatrix} \hat{\theta} \\ \hat{\sigma}^2 \end{bmatrix} = M^{-1} \begin{bmatrix} \theta \\ \sigma^2 \end{bmatrix}.
\]

(3.31)

Lastly, note that the expression in (3.28) is the same as the information matrix obtained by the method of generalised least squares in Section 3.2.2 and given by (3.10). This is expected from the equivalency of MLE and GLS that had been mentioned in Section 3.3.1. An alternative form of \( M(\sigma^2) \) could be obtained via the ANOVA method, as outlined in Section 3.2.1. However it will not be equivalent to (3.29), there is no general expression for it and it would be obtained as the inverse of the variance-covariance matrix of the estimators. This approach is not explored in the literature of optimum designs and neither is here.

### 3.3.3 Restricted maximum likelihood estimation

The MLE’s in (3.21) are biased; in fact, the amount by which they underestimate the variances is \( \text{rank}(X) \sigma^2 / N \) (see e.g. Rao, 1997). This situation arises from the fact that the MLE procedure for variance components does not take into consideration the degrees of freedom needed to estimate the fixed parameters. An estimation method that treats the fixed parameters as nuisance elements can alleviate the problem.

The preferred such method is known as restricted maximum likelihood estimation (REML). The most intuitive interpretation of this method is that the variance components are estimated using the residuals obtained after fitting the fixed part of the model by the least squares method. This implies that the fixed
effects are not used (or eliminated) for the estimation of the variance components.

Removing the fixed effects is formally achieved by doing a transformation of the data. With \( \text{rank}(X) = p \), define \( A \) as a \( N \times (N - p) \) matrix with \( \text{rank}(A) = N - p \). The transformation consists of forming the vector \( A^T y \) of linear combinations of \( y \) under the restriction

\[
A^T X = 0 \quad \text{and therefore} \quad A^T y \sim N(0, A^T VA).
\]

Using \( A^T y \) instead of \( y \) in the likelihood function (3.16) gives

\[
L = L(V|A^T y) = \exp \left( -\frac{1}{2} y^T A (A^T VA)^{-1} A^T y \right) \frac{1}{(2\pi)^{\frac{1}{2}(N-p)} |(A^T VA)|^{\frac{1}{2}}},
\]

which is known as marginal or restricted likelihood. The restricted log-likelihood is

\[
l = -\frac{1}{2}(N - p) \log(2\pi) - \frac{1}{2} \log(|A^T VA|) - \frac{1}{2} y^T A (A^T VA)^{-1} A^T y.
\]  

(3.33)

The REML estimates are those that maximise (3.33). The optimisation is done as before by equating the first derivative to zero. In what follows, the results in Appendix A.1 are used repeatedly. The first derivative of (3.33) is

\[
\frac{\partial l}{\partial \sigma_i^2} = \frac{1}{2} \left\{ y^T P \frac{\partial V}{\partial \sigma_i^2} P y - \text{tr} \left( P \frac{\partial V}{\partial \sigma_i^2} \right) \right\},
\]  

(3.34)

where \( P \) is the orthogonal projector of \( X \) into \( V \) and it is defined as

\[
P = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} = A (A^T V^{-1} A)^{-1} A^T.
\]

This projector matrix, also known as the \( P \) matrix, has the characteristics

\[
P VP = P \quad \text{and} \quad PX = 0.
\]
The REML estimation equations can now be written in the following form

\[ y^T P \frac{\partial V}{\partial \sigma_i^2} P y = \text{tr} \left( P \frac{\partial V}{\partial \sigma_i^2} \right) \] (3.35)

which, as the MLE equations, require an iterative procedure for their solution.

### 3.3.4 REML information matrix

The steps to obtain the information matrix based on REML estimation are the same as for the MLE method. However, for REML there are no fixed parameters. Consequently, equation (3.22) reduces to only its \((2, 2)\) element so that

\[ M(\sigma^2) = -E \left[ \frac{\partial^2 l}{\partial \sigma_i^2 \partial \sigma_j^2} \right], \] (3.36)

but in this case \(l\) refers to the restricted log-likelihood (3.33). This is the REML information matrix which only exists for the variance components.

Given that

\[ \frac{\partial P}{\partial \sigma_i^2} = -A(A^T V^{-1} A)^{-1} \frac{\partial A^T V^{-1} A}{\partial \sigma_i^2} (A^T V^{-1} A)^{-1} A^T \]

\[ = -P \frac{\partial V}{\partial \sigma_i^2} P, \]

and using (3.34), the second derivative of the restricted log-likelihood is

\[ \frac{\partial^2 l}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{\partial l}{\partial \sigma_j^2} \left[ \frac{1}{2} \left\{ y^T P \frac{\partial V}{\partial \sigma_i^2} P y - \text{tr} \left( P \frac{\partial V}{\partial \sigma_i^2} \right) \right\} \right] \]

\[ = \frac{1}{2} \text{tr} \left( P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} \right) - \frac{1}{2} y^T P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} P y - \frac{1}{2} y^T P \frac{\partial V}{\partial \sigma_i^2} P \frac{\partial V}{\partial \sigma_j^2} P y \]

\[ = \frac{1}{2} \text{tr} \left( P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} \right) - y^T P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} P y. \] (3.37)
To evaluate the expectation of (3.37), note that the first term of the right hand side being the trace of a matrix it is a constant and its expectation is itself. For the second term,

\[-E \left( y^T P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} P y \right) = \text{tr} \left( P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} P V \right) - \theta^T X^T P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} P X \theta \]

\[= 0.\]

It then follows that

\[-E \left[ \frac{\partial^2 l}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = \frac{1}{2} \text{tr} \left( P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} P \right),\]

and by (3.36)

\[M(\sigma^2) = \frac{1}{2} \text{tr} \left( P \frac{\partial V}{\partial \sigma_j^2} P \frac{\partial V}{\partial \sigma_i^2} P \right) \]

(3.38)

is the REML information matrix. Equation (3.38) is a general expression; the derivatives involved take specific forms depending on the choice of the \( V \) matrix. See Chapter 5 for specific examples.

### 3.3.5 Bayesian framework

The mixed effects model leads naturally into a Bayesian framework. The theory of Bayesian statistics is not reviewed here though. References treating the topic are, for example, Pilz (1991) which covers design optimality, Broemeling (1985) which is about general modelling, and applications can be found in Davidian and Giltinan (1995) and Searle et al. (1992).

In a Bayesian context, model (3.12) is defined as a hierarchy. First the random error is modelled as the sampling distribution, and then the other parameters, fixed and random, are specified as a prior distribution. That is, the first level of
the hierarchy deals with a random variable $y$ whose distribution is conditional on a parameter $\Theta$. In turn, the second level models the distribution of $\Theta$.

Under the assumption of normality, a hierarchy for model (3.12) is described in stages as

$$\theta \sim N(\theta_0, T), \quad \tau \sim N(0, D), \quad \text{and} \quad \epsilon \sim N(0, R)$$

where, as mentioned above, $\epsilon$ is at the top of the hierarchy. With this structure, a full likelihood function can be written as

$$L(\theta_0, T, D, R | y) = \int \int f(y|\theta, \tau, R) f_\theta(\theta|\theta_0, T) f_\tau(\tau|D) d\theta \ d\tau$$

where $f(\cdot)$ are the corresponding known densities of the parameters.

The specific choices of $f(\cdot)$ will define the likelihood function from which to estimate the unobservable parameters. When the estimation is conditional on $\theta$, and $\tau$ is integrated out, the standard MLE results. If instead $\theta$ is integrated out, usually by specifying a flat or non-informative prior, REML is obtained.

The Bayesian approach is not going to be fully explored here. The variance of the estimates of the variance components is not straightforward to obtain. Considerable underestimation occurs arising from the assumptions made in the specification of the hierarchy. This would need to be corrected to find reliable variance estimates. Techniques to do this correction are available, but by construction they are intricate and difficult to implement. Also, none of these is a recognized method that is best for most situations and such techniques are applied on a case by case basis. See (Searle et al., 1992) for more details. This difficulty does not make this method attractive for designing optimum experiments for the variance components. A formal account of Bayesian optimum designs for the fixed effects is presented in Pilz (1991) and Atkinson et al. (2007). However, the idea
of incorporating available information on the values of the parameters can be
exploited in the experimental design phases. This scenario will rather be called
pseudo-Bayesian, but it still proves to be useful for variance components. This
possibility will be studied later.

3.4 Further results for variance components

3.4.1 Information matrix for ratios of the variance components

Writing the information matrix so that the variance components are expressed
as ratios with respect to $\sigma^2_i$ is very relevant. These ratios appear naturally in
the estimation process as elements of $V$. In the same way they appear in the
information matrix. Consequently, the optimality of experimental designs for
variance components depends on these ratios.

On the other hand, variance ratios are important on their own right. For
example, in blocked designs with random blocks, these ratios represent the degree
of correlation between observations in the same block (e.g. Goos (2002)). More
generally, the intra-class correlation in random models can be defined as

$$\rho = \frac{\sigma^2_i}{\sigma^2_i + \sigma^2}_i$$

where $i = 1, \ldots, r$ represents one class. These measures are common in psycho-
logical studies (Searle et al., 1992). Heritability, a measure of the relationship
between members of a species, is very important in genetic studies and is defined
as

$$h = \frac{4\sigma^2_i}{\sigma^2_i + \sigma^2}_i,$$

which is commonly associated with a 1-factor random model. See, e.g. Ghosh
et al. (2005) and Rao (1997). Other examples of the use of variance ratios in experiments with animals are described in Johnson and Thompson (1995).

**MLE information matrix**

The following notation was introduced by Hartley and Rao (1967), and is obtained by dividing the variance-covariance matrix $V$ by $\sigma^2_\epsilon$, more formally

$$H = V/\sigma^2_\epsilon, \quad \text{and} \quad H^{-1} = V^{-1} \sigma^2_\epsilon.$$  \hfill (3.39)

Now define the variance ratios as

$$\eta_i = \frac{\sigma^2_i}{\sigma^2_\epsilon} \quad i = 1 \ldots r \quad \text{and} \quad \eta_0 = \sigma^2_\epsilon,$$

so that

$$\begin{bmatrix} \eta_0 \\ \eta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I_r/\sigma^2_\epsilon \end{bmatrix} \sigma^2 = \Delta,$$

then by using the invariance principles involved in (A.19) from Appendix A.2, the information matrix can be written as

$$M[\Delta] = \left[ \frac{\partial \sigma^2_i}{\partial \Delta_j} \right] M(\sigma^2) \left[ \frac{\partial \sigma^2_i}{\partial \Delta_j} \right]^T.$$

The MLE information matrix can be obtained using (3.29). Next, finding the derivatives above and expanding the vector of variance components as $\sigma^2 = [\sigma^2_\epsilon \eta]$ result in

$$M \begin{bmatrix} \sigma^2_\epsilon \\ \eta \end{bmatrix} = \begin{bmatrix} 1 & \eta^T \\ 0 & \eta_0 I \end{bmatrix} \frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_\epsilon} V^{-1} \frac{\partial V}{\partial \sigma^2_j} \right) \begin{bmatrix} 1 & 0 \\ \eta & \eta_0 I \end{bmatrix}.$$
3 Parameter estimation

3.4 Further results for variance components

\[
\begin{bmatrix}
1 & \eta^T \\
0 & \eta_0 I
\end{bmatrix}
\begin{bmatrix}
\frac{1}{2} \text{tr}(V^{-2}) & \text{tr}\left(V^{-2} \frac{\partial V}{\partial \sigma_i^2}\right)^T \\
\frac{1}{2} \text{tr}\left(V^{-2} \frac{\partial V}{\partial \sigma_j^2}\right) & \text{tr}\left(V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} \frac{\partial V}{\partial \sigma_j^2}\right)
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
\eta & \eta_0 I
\end{bmatrix},
\]

which is a block matrix formed according to the subscripts \(i = j = 1, \ldots, r\), that after simplification reduces to

\[
M \begin{bmatrix}
\sigma_i^2 \\
\eta
\end{bmatrix} = \frac{1}{2}
\begin{bmatrix}
\frac{N}{\sigma_i^2} & \frac{\text{tr}\left(H^{-1} \frac{\partial V}{\partial \sigma_i^2}\right)^T}{\sigma_i^2} \\
\text{tr}\left(H^{-1} \frac{\partial V}{\partial \sigma_i^2}\right) & \text{tr}\left(H^{-1} \frac{\partial V}{\partial \sigma_i^2} H^{-1} \frac{\partial V}{\partial \sigma_j^2}\right)
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
\eta & \eta_0 I
\end{bmatrix},
\]

(3.40)

REML information matrix

The REML information matrix for the ratios can be obtained via the same method used in the MLE case. Initially, consider equation (3.38) for the REML information matrix for \(\sigma^2\). Then, as done above and using the parameter transformation properties, i.e. (A.19) once, write

\[
M \begin{bmatrix}
\sigma_i^2 \\
\eta
\end{bmatrix} = \frac{1}{2}
\begin{bmatrix}
\frac{N}{\sigma_i^2} & \frac{\text{tr}\left(P^2 \frac{\partial V}{\partial \sigma_i^2}\right)^T}{\sigma_i^2} \\
\text{tr}\left(P^2 \frac{\partial V}{\partial \sigma_i^2}\right) & \text{tr}\left(P^{-1} \frac{\partial V}{\partial \sigma_i^2} P^{-1} \frac{\partial V}{\partial \sigma_j^2}\right)
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
\eta & \eta_0 I
\end{bmatrix}.
\]

After a second use of (A.19), with

\[Q = P \sigma_i^2,\]

and further simplification the result is

\[
M \begin{bmatrix}
\sigma_i^2 \\
\eta
\end{bmatrix} = \frac{1}{2}
\begin{bmatrix}
\frac{N - p}{\sigma_i^2} & \frac{\text{tr}\left(Q \frac{\partial V}{\partial \sigma_i^2}\right)^T}{\sigma_i^2} \\
\text{tr}\left(Q \frac{\partial V}{\partial \sigma_i^2}\right) & \text{tr}\left(Q \frac{\partial V}{\partial \sigma_i^2} Q \frac{\partial V}{\partial \sigma_j^2}\right)
\end{bmatrix},
\]

(3.41)
where $p = \text{rank}(X)$. This procedure is outlined in Corbeil and Searle (1976) but is developed in terms of a modification of the W-transform first introduced in Hemmerle and Hartley (1973).

An alternative procedure is outlined by Johnson and Thompson (1995) starting directly from the restricted maximum likelihood function. However, they are concerned with the average information which includes the observed values, rather that the expected information only. Therefore, the resulting expressions are not equivalent to (3.41). Focusing on the expected information matrix, which is the interest here, the procedure is as follows. The restricted maximum likelihood function given in terms of $Q$ and $H$ is

$$
l = -\frac{1}{2} (N - p) \log(2\pi) - \frac{1}{2} \log(|H|) - \frac{1}{2} \log(|X^T H^{-1} X|) - \frac{1}{2\sigma^2_{\epsilon}} y^T Q y
$$

whose second derivatives are

$$
\frac{\partial^2 l}{\partial \eta_i \partial \eta_j} = \frac{1}{2} \text{tr} \left( Q \frac{\partial H}{\partial \eta_i} Q \frac{\partial H}{\partial \eta_j} \right) - \frac{1}{\sigma^2_{\epsilon}} y^T Q \frac{\partial H}{\partial \eta_i} Q \frac{\partial H}{\partial \eta_j} y.
$$

(3.42)

After taking the negative of the expectation of (3.42) and arranging in matrix form, the REML information matrix for the variance ratios can also be written as

$$
M = \frac{1}{2} \begin{bmatrix}
\frac{N-p}{\sigma^2_{\epsilon}} & \text{tr} \left( Q \frac{\partial H}{\partial \eta_i} Q \frac{\partial H}{\partial \eta_j} \right) \\
\text{tr} \left( Q \frac{\partial H}{\partial \eta_i} Q \frac{\partial H}{\partial \eta_j} \right) & \text{tr} \left( Q \frac{\partial H}{\partial \eta_i} Q \frac{\partial H}{\partial \eta_j} \right)
\end{bmatrix},
$$

(3.43)

which is equivalent to (3.41).

### 3.4.2 Changes in total variance

It is useful to consider the situation in which the total variance is changed by a fixed amount $k$. Given that designs for variance components depend on their ratios,
it is important to verify that changes in total variance do not affect the form of the
information matrix and its values only get linearly scaled by the constant amount
$k$. This is relevant as the parameter space of the optimum designs evaluated
in Chapter 6 is rescaled. Additionally, if the form of the information matrix for
the variance components does not change, the asymptotic characteristics of the
estimates remain.

Let $k$ be a known fixed constant, $k > 0$. Refer to model (3.12) and define

$$\text{var}(y) = V_k = kV, \tag{3.44}$$

with

$$V_k^{-1} = \frac{1}{k} V^{-1}. \tag{3.45}$$

By using (3.44) in (3.16), the likelihood function is now a function of $\theta$ and
$V_k$, which can be written as

$$L = L(\theta, V_k | y) = \frac{\exp \left( -\frac{1}{2} (y - X\theta)^T V_k^{-1} (y - X\theta) \right)}{(2\pi)^{\frac{1}{2}}N |V_k|^\frac{1}{2}},$$

and the log-likelihood becomes

$$\ell = -\frac{1}{2} N \log(2\pi) - \frac{1}{2} \log(|V_k|) - \frac{1}{2} (y - X\theta)^T V_k^{-1} (y - X\theta).$$

The maximisation is done with respect to $\theta$ and $\sigma^2_{k_i}$. The corresponding
derivatives involving $\theta$ are

$$\frac{\partial \ell}{\partial \theta} = \frac{1}{k} X^T V^{-1} y - \frac{1}{k} X^T V^{-1} X \theta,$$

$$\frac{\partial^2 \ell}{\partial \theta \partial \theta^T} = -\frac{1}{k} X^T V^{-1} X,$$
\[
\frac{\partial^2 l}{\partial \theta \partial \sigma^2_{k_i}} = \frac{\partial}{\partial \sigma^2_{k_i}} \mathbf{X}^T \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta) = -\frac{1}{k^2} \left[ \mathbf{X}^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta) \right]
\]

where \( \mathbf{V}^{-1}_k \) is defined by (3.45).

For the derivatives with respect to \( \sigma^2_{k_i} \), the starting point is

\[
\frac{\partial}{\partial \sigma^2_{k_i}} \left[ -\frac{1}{2} \log(||\mathbf{V}_k||) - \frac{1}{2} (\mathbf{y} - \mathbf{X} \theta)^T \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta) \right]
\]

and by using the results in Appendix A.1, the first term becomes

\[
\frac{\partial}{\partial \sigma^2_{k_i}} \left( -\frac{1}{2} \log(||\mathbf{V}_k||) \right) = -\frac{1}{2} \text{tr} \left( \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \right),
\]

from which the second derivative is obtained as

\[
\frac{\partial}{\partial \sigma^2_{k_j}} \left( -\frac{1}{2} \text{tr} \left( \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \right) \right) = -\frac{1}{2} k^{-2} \text{tr} \left( \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_j} \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \right).
\]

(3.46)

For the second term,

\[
\frac{\partial}{\partial \sigma^2_{k_i}} \left( -\frac{1}{2} (\mathbf{y} - \mathbf{X} \theta)^T \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta) \right)
\]

\[
= -\frac{1}{2} (\mathbf{y} - \mathbf{X} \theta)^T \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta)
\]

which leads to the second derivative

\[
\frac{\partial}{\partial \sigma^2_{k_j}} \left( -\frac{1}{2} (\mathbf{y} - \mathbf{X} \theta)^T \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta) \right)
\]

\[
= \frac{1}{2} \left[ -2 (\mathbf{y} - \mathbf{X} \theta)^T \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_j} \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta) \right]
\]

\[
= -k^{-3} \left[ (\mathbf{y} - \mathbf{X} \theta)^T \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_i} \mathbf{V}^{-1}_k \frac{\partial \mathbf{V}}{\partial \sigma^2_j} \mathbf{V}^{-1}_k (\mathbf{y} - \mathbf{X} \theta) \right].
\]

(3.47)
Combining (3.46) and (3.47) gives

\[
\frac{\partial^2 l}{\partial \sigma^2_{k_i} \partial \sigma^2_{k_j}} = -\frac{1}{2} k^{-2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_j} \right) \\
- k^{-3} \left[ (y - X\theta)^T V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_j} V^{-1} (y - X\theta) \right].
\]

To form the information matrix, the negative of the expectations of the second derivatives have to be evaluated as shown below:

\[
-E \left( \frac{\partial^2 l}{\partial \theta \partial \theta} \right) = k^{-1} X^T V^{-1} X \tag{3.48}
\]
\[
-E \left( \frac{\partial^2 l}{\partial \theta \partial \sigma^2_k} \right) = 0 \tag{3.49}
\]
\[
-E \left( \frac{\partial^2 l}{\partial \sigma^2_k \partial \sigma^2_k} \right) = -\frac{1}{2} k^{-2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_i} \right) \\
+ k^{-3} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_j} V^{-1} \frac{\partial V}{\partial \sigma^2_k} \right),
\]

and using (3.44) the second term becomes

\[
k^{-3} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_j} V^{-1} k V \right) = k^{-2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_j} V^{-1} V \right) \\
= k^{-2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_j} \right).
\]

Adding now the first and second terms gives

\[
-E \left( \frac{\partial^2 l}{\partial \sigma^2_k \partial \sigma^2_k} \right) = \frac{1}{2} k^{-2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma^2_i} V^{-1} \frac{\partial V}{\partial \sigma^2_j} \right). \tag{3.50}
\]

The information matrix is now formed by arranging (3.48), (3.49) and (3.50) which
results in

\[
M_k \begin{bmatrix} \theta \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} k^{-1}X^T\mathbf{V}^{-1}X & 0 \\ 0 & \frac{1}{2}k^{-2}\text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \right) \end{bmatrix} \quad (3.51)
\]

It can be seen that the estimates of the variance components do not depend on the total variance, although they are scaled by a factor \(k^2\) since

\[
\text{var}(\hat{\mathbf{V}}_k) = \text{var} \begin{bmatrix} \hat{\theta} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} k(X^T\mathbf{V}^{-1}X)^{-1} & 0 \\ 0 & 2k^2 \left( \text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \right) \right)_{i=j=0,...,r}^{-1} \end{bmatrix}.
\]

### 3.5 Dispersion-Mean model for the estimation of variance components

This approach involves a transformation of the mixed model so that the variance components \(\sigma^2\) appear linearly, as in a simple linear model. Estimation of variance components is in general a nonlinear problem. Consequently, linearising the model simplifies the problem and permits standard methods to be used. This is regularly done in nonlinear modelling. A similar idea is exploited here. After the linearisation, the method of least squares can be used for the variance components. Depending on which type of least squares method is used, the results will be equivalent to those obtained using other known estimation procedures.

The dispersion-mean model method was first introduced by Pukelsheim (1976). More details and examples of applications are presented in Searle et al. (1992). Extensions to the Gauss-Markov theorem can be applied to this transformation. The Gauss-Markov theorem can be extended to variance components as shown by Harville (1976). Malley (1986) presents such extension for the dispersion-mean linearisation. The presentation here is mainly based on Searle et al. (1992).
3 Parameter estimation

3.5 Dispersion-Mean model for the estimation of variance components

The technique uses the vec operator and the vec-permutation matrix as described in Appendix A.1 and A.3. In addition, define a projection matrix as

\[ M = I - X (X^T X)^{-1} X^T = I - XX^+, \]  

(3.52)

where \((X^T X)^{-1}\) corresponds to any generalised inverse of \((X^T X)\) and \(X^+\) is the Moore-Penrose inverse of \(X\). Also \(M\) is an idempotent matrix, that is

\[ M = M^T = M^2. \]

Consider the linear mixed effects model

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

(3.53)

with elements as defined in Section 2.2.2. To linearise the model for the variance components, since they are second moments or quadratic forms (from an algebraic point of view), the response \(y\) has to be expressed also in a quadratic form. Let \(A\) be a symmetric matrix so that \(y^T Ay\) is quadratic and unbiased (for linear combinations of \(\sigma^2\)) under the following conditions

\[ X^T AX = 0, \quad AX = 0, \quad \text{and} \quad MX = 0. \]

Also, consider the property

\[ MAM = AM = MA = A. \]

Now, it is possible to write

\[ y^T Ay = (My)^T A (My) \]
3 Parameter estimation

3.5 Dispersion-Mean model for the estimation of variance components

\[
\begin{align*}
= y^T M A M y \\
= \text{vec}(y^T M A M y) = \left( (M y)^T \otimes y^T M \right) \text{vec} A \\
= (\text{vec} A)^T (M y \otimes M y)
\end{align*}
\] (3.54)

where \( \otimes \) denotes the Kronecker product. Next, define

\[
y = M y \otimes M y,
\]

and note that by (3.54) the expectation of \( y^T A y \) depends on the matrix \( A \) and on \( E(y) \). Then

\[
E(M y \otimes M y) = E \left[ M(y - X \theta) \otimes M(y - X \theta) \right] \\
= (M \otimes M) E \left[ (y - X \theta) \otimes (y - X \theta) \right] \\
= (M \otimes M) E \left[ \text{vec}((y - X \theta)^T(y - X \theta)) \right] \\
= (M \otimes M) \text{vec} V,
\]

and using the basic form of \( V \) as in (3.15) it simplifies to

\[
E(M y \otimes M y) = (M \otimes M) \sum_{i=0}^{r} \text{vec}(Z_i Z_i^T) \sigma_i^2.
\] (3.55)

Define

\[
M_i = M Z_i Z_i^T M \quad i = 0, \ldots, r,
\]

then (3.55) can be written as the following matrix

\[
E(M y \otimes M y) = [M_0 | M_1 | \cdots | M_r] \sigma^2,
\]
which finally leads to
\[ E(\mathbf{y}) = \mathbf{X} \sigma^2, \tag{3.56} \]
where \( \mathbf{X} = [\mathbf{M}_0 \mid \mathbf{M}_1 \mid \cdots \mid \mathbf{M}_r] \). This is the required expression for the linear model for \( \sigma^2 \), which is known as the \textit{dispersion-mean} model. This formulation has the characteristic that it makes no assumption about the distribution of the response vector neither on the transformed variable \( \mathbf{y} \). Also the residuals obtained after fitting model (3.53) by OLS combine into squares and products to form \( \mathbf{y} \).

To complete the formulation of the dispersion-mean model, \( \text{var}(\mathbf{y}) \) is required. Clearly, this is the variance of a variable of quadratics, so it involves fourth moments. The procedure is as follows. From the model equation (3.53), define a general expression for the variance-covariance matrix of the random effects as
\[ \text{var}(\tau) = E(\tau \tau^T) = \mathbf{D}, \]
and consequently,
\[ \text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZDZ}^T + \sigma_t^2 \mathbf{I}_N, \]
where \( \mathbf{D} \) can take different forms depending on the context, such that it can contain, in addition to the variances, covariances between effects of the same random factor or between different random factors. This is customary notation used in the mixed models literature, for example Brown and Prescott (2006) and Jiang (2007). Next, define
\[ \mathbf{F} = \text{var}[(\mathbf{y} - \mathbf{X}\theta) \otimes (\mathbf{y} - \mathbf{X}\theta)] \]
\[ = \text{var}(\mathbf{Z}\tau \otimes \mathbf{Z}\tau) \]
\[ = (\mathbf{Z} \otimes \mathbf{Z}) \text{var}(\tau \otimes \tau)(\mathbf{Z}^T \otimes \mathbf{Z}^T), \tag{3.57} \]
then

\[
\text{var}(\tau \otimes \tau) = E[(\tau \otimes \tau)(\tau \otimes \tau)^T] - E(\tau \otimes \tau)(E(\tau \otimes \tau))^T
\]

\[
= E(\tau\tau^T \otimes \tau\tau^T) - E \text{vec}(\tau\tau^T)(E \text{vec}(\tau\tau^T))^T
\]

\[
= E(\tau\tau^T \otimes \tau\tau^T) - \text{vec}D(\text{vec}D)^T, \tag{3.58}
\]

and

\[
E(\tau\tau^T \otimes \tau\tau^T) = I + S_q + \text{vec}I_q(\text{vec}I_q)^T + \Gamma, \tag{3.59}
\]

where

\[
\Gamma = \text{diag} \left[ \text{vec}(\text{diag}(\gamma_i I_q)) \right],
\]

and

\[
\gamma_i = \frac{\mu_i}{(\sigma^2)^2} - 3
\]

is the common kurtosis parameter of the elements of \( \tau \) and \( \mu_i \) is the corresponding fourth central moment. For details of how to construct (3.59) see Malley (1986). Now, making use of (3.58) and (3.59)

\[
\text{var}(\tau \otimes \tau) + \text{vec}D(\text{vec}D)^T = (D^{1/2} \otimes D^{1/2}) \left[ I + S_q + \text{vec}I_q(\text{vec}I_q)^T + \Gamma \right] (D^{1/2} \otimes D^{1/2}),
\]

which after expanding and simplifying becomes

\[
\text{var}(\tau \otimes \tau) = (D \otimes D)(I + S_q) + (D^{1/2} \otimes D^{1/2})\Gamma(D^{1/2} \otimes D^{1/2}).
\]

Going back now to (3.57), while using the above equation, gives

\[
F = (Z \otimes Z) \left[ (D \otimes D)(I + S_q) + (D^{1/2} \otimes D^{1/2})\Gamma(D^{1/2} \otimes D^{1/2}) \right] (Z^T \otimes Z^T)
\]

\[
= (V \otimes V)(I + S_N) + (ZD^{1/2} \otimes ZD^{1/2})\Gamma(D^{1/2}Z^T \otimes D^{1/2}Z^T).
\]
Finally, the variance of $Y$ can be expressed as

$$
\text{var}(Y) = \text{var}(M_y \otimes M_y)
= \text{var}[M(y - X\theta) \otimes M(y - X\theta)]
= (M \otimes M)F(M \otimes M)
= (MV_M \otimes MV_M)(I + S_N)
+ (MZD^{1/2} \otimes MZD^{1/2})\Gamma(D^{1/2}Z^T M \otimes D^{1/2}Z^T M),
$$

and this completes the specification of the dispersion-mean model. It is clear that equation (3.60) consist of two parts. The second summand on the right hand side becomes null when $\gamma_i = 0 \forall i$ and therefore $\Gamma = 0$. This is the case under normality. This development could be used to design optimum experiments for variance components for non-normal responses.

These results can be extended by using the equations for $F$ and $\Gamma$ above to obtain an expression for $\text{cov}(y^T MA_M, y^T MB_M)$ as detailed in Searle et al. (1992) and Malley (1986).

The dispersion-mean model (3.56) can now be used to estimate the variance components from a linear mixed effects model using the OLS and GLS methods described in previous sections. If OLS is chosen, the resulting estimates are equivalent to MINQUE0 as specified by Rao (1971). The GLS method, with (3.60), will yield estimates equivalent to REML. When using a preliminary known value of $X\theta$ and GLS, MLE equivalent estimates are obtained.

**Information matrix**

The approach to find the information matrix of the variance components based on the dispersion-mean transformation will recourse on the fact that the estimation is done via the least squares method. This combines with the ability to use extended
versions of the Gauss-Markov theorem (Harville, 1976; Malley, 1986). Therefore, the estimates obtained from the dispersion-mean model using least squares are best linear unbiased estimates (BLUE).

The information matrix for the dispersion-mean model is then written as

$$ M_y = (X^T V^{-} X)^{-}. \quad (3.61) $$

It is known that the results presented in this section are invariant to the choice of generalised inverse. Thus, $V^-$ is changed to $V^+$ and $(X^T V^- X)^- \to (X^T V^+ X)^+.$

Equation (3.61) correspond to using GLS. The resulting information matrix will be equivalent to that found for REML in (3.38). Similarly, when $V = \sigma^2 \epsilon I,$ the procedure reduces to OLS and the information matrix will be equivalent to the MLE version in (3.30).

The extensions to the Gauss-Markov theorem that have been mentioned have not been proven for the case when $V$ involves kurtosis parameters. However, it is acknowledged (Searle et al., 1992) that including the kurtosis is a viable way to estimate variance components of non-normal response variables, and therefore (3.61) has been used here to find optimum designs. However, as concluded by Malley (1986) more work on the extension of the Gauss-Markov theorem to non zero kurtosis cases is needed.

### 3.6 Summary

This chapter presents an account of relevant methods of estimation of typical statistical models. Specifically, methods that are suitable to use when different sources of variability are included.

The main focus is on the mixed effects models and estimation based on maximum likelihood. Attention is given to the information matrices obtained
using the standard MLE and the REML methods. The principal objective is to present the information matrix for the variance components. The Bayesian approach is noted, but it is not investigated further given that information matrices are not easy to obtain. However, the concept of pseudo-Bayesian optimality is introduced. This idea is used later in the thesis to select optimum designs for the variance components.

In addition to the REML information matrix, two specific results for the variance components are presented. The information matrix for ratios of the type $\sigma_i^2/\sigma_c^2$ is reported. Given the importance of such ratios for the construction of optimum designs, the effect of changing the total variance was considered. This is shown in equation (3.51) where it can be seen that the form of the information matrices are indeed not affected by the change. An additional consequence worth noting is that when finding optimum designs, the parameter space of the variance components can be rescaled without affecting the results of the optimisation.

Estimation of nonlinear models is also mentioned but for the fixed case using the least squares method. References for using a Bayesian framework are given.

A method, known as the dispersion-mean model, that transforms the mixed model so that the variance components appear linearly is presented. The ordinary least squares method can be directly used with the dispersion-mean model. Therefore the information matrix of the variance components is written in the same form as for the case of fixed parameters, but using the transformed matrices $\mathbf{X}$ and $\mathbf{V}$, and generalised inverses.

Optimum designs for variance components can be found using the dispersion-mean model. A consequence of the transformation is that the variance-covariance matrix includes a kurtosis parameter. This allows to find optimum experimental designs for non-normal responses. It seems reasonable to keep such extension within the framework of the exponential family of distributions.
Experimental complexity is the result of having restrictions to incorporate. Budgetary constraints often determine the quantity of runs that can be done, and in some cases across how many sites and in which time frame. Physical and contextual limitations are usually reflected in the shape and size of the design region and the feasibility to conduct the experimental runs.

The experiment has to be constructed in ways that account for the restrictions imposed, the latter themselves manifest in how the key design elements are defined in relation with the system under investigation. Additionally, the subject matter and the objectives have an influence in the design.

A fundamental idea here is that by defining the characteristics of the factors and the experimental units, the design acquires a structure that can then be reflected in the statistical model. There is no definite rule for doing this. This chapter describes some of the ways in which a design can be arranged in relatively complex situations. These different arrangements are called *design structures*.

The key experimental elements of response, factors, levels, treatments and model, along with the principles of randomisation, replication and blocking were introduced in Chapter 2. In Section 2.1.1 the ideas of crossed and nested factors and hierarchical structure were briefly explained. In Section 2.2.2 the concepts of
treatment and variability structure were introduced. All these ideas intertwine in the design process to accommodate the complexity of an experimental phase.

4.1 Factors and experimental units

4.1.1 Factors

Clear definitions on the type of factors to be used are very important. In first instance, these definitions are completely related to the subject matter. Thereafter, the classification of the factors is made with statistical considerations as well. The number of stages are related to the conditions of the experiment but also to the model under consideration. Similar care is needed to decide the role and relative importance of each factor.

When the factors are defined, the following is assumed known:

- the total number of factors involved,
- the set of fixed and of random factors,
- the quantitative and qualitative fixed factors,
- the range of the levels of fixed factors and the number of levels of random factors,
- all factor restrictions, including hard to change (HC) factors.

Before transporting this information into a statistical model, other considerations are important. The definition of crossed and nested factors is of particular relevance. However, these characteristics may not be known at the time of factor selection. Their recognition may be coupled with the definition of other key elements of the complete design structure, e.g. the choice of experimental units. The latter, in particular type and size, is explained next. Crossed and nested factors are reviewed later.
4.1.2 Experimental units

It is widely recognised that the adequate identification of the units in an experiment is critical, see e.g. Hinkelmann and Kempthorne (2005) and Casella (2008).

Failure to do so correctly can lead to a design that does not satisfy the objectives of the experimental phase yielding invalid conclusions. Actually, this endeavour may be one of the most obscure of the design phase. The simplest example is that experimental units may be different if the same system is studied in a separate experiment under different objectives. Several definitions of what units are can be found across the design of experiments literature. Useful presentations and discussions are given by Bailey (2008) and Milliken and Johnson (2009).

The determination of unit size is context dependent and thus is subjective. This may be smallest, small, large, larger, etc. For multistage experiments, it is key to determine how many different sizes of units are involved and which units belong to which stage of the design structure. There is an interplay between these concepts. Sometimes it may be possible to determine the hierarchy first and then the size of the units is defined accordingly. Other times, the size of the unit may establish its location in the hierarchy.

Two different types of units will be distinguished:

**Observational units** These are the smallest size units and are assigned to the lowest stage of the design hierarchy. Observations are acquired at this stage, thus the name *observational* units.

**Structural units** Larger, intermediate or even conceptual (non physical) units will be associated with all other stages of the hierarchy. These will irrespectively be named structural units.

Concurrent with the definitions definitions above:

**Observational treatments** These are the set of treatments $t_i$, $i = 1, \ldots, N$, applied to the observational units.
4 Design structures

4.2 Design, treatment and variability structures

Structural treatments These are the set of treatments \( t_j, \ j = 1, \ldots, s \), applied to the structural units; they are obtained from the factor arrangement used in the corresponding stage (see Section 4.3).

When it is necessary to distinguish between the two types of treatments, \( t_i \) will denote the observational and \( t_{(j)} \) the structural treatments. It is generally assumed that the largest unit belongs to the top hierarchy stage, the second largest to the next stage down, and so on, until the observational units are found at the lowest stage.

The use of different size of units implies a nested arrangement (see Section 4.3.3). This in turn means a hierarchy ought to be specified, but not always. That is, nesting in itself does not require a model for multistage variability but a hierarchy does. By merging these ideas, it is clear that the precise definition of the experimental units is important since this may dictate the overall structure of the design. Conversely, if a design structure is specified first, the experimental units have to be allocated to the correct hierarchy stage.

4.2 Design, treatment and variability structures

A design structure can be created once the factors and the experimental units have been identified and classified. The structure is specified through the definition of the two parts of the mixed model; note that the two parts can be defined separately, but the same principles are used for both.

From Section 2.2.2, remember the general linear mixed effects model:

\[
y_{ij} = f^T(x|t_i)\theta + g(\zeta|t_{(j)})\tau + \epsilon_{ij} \tag{4.1a}
\]

\[
y = X\theta + Z\tau + \epsilon. \tag{4.1b}
\]
It is presented this way with the intention to make it general and flexible. This
generality includes hierarchical (multistage) models. It covers the combinations of
\( f(x) = 1 \) and \( g(\zeta) = 0 \). This allows the description of fixed, random and mixed
models reflecting complex designs by clearly defining the blocking schemes and
the treatment allocation “separately”.

Considering (4.1), the fixed part is defined by \( f(x|t_i), i = 1, \ldots, N \), where
\( x \mapsto t \) with \( t_i \) indicating a fixed treatment. The structure associated with the
fixed part is called the \textit{treatment structure}. This is the structure of the
observational treatments. These treatments will mainly be arranged in blocks
when necessary.

Likewise, \( g(\zeta|t_j) \) describes how the variability is to be explained. Therefore,
the structure associated with the random part is called \textit{variability structure}.
If this is hierarchical, then it is composed of different \textit{stages}. A stage can be
defined by the characteristics of one factor. The respective structural units will
receive a treatment that coincides with the levels of the factor. Further, a stage
can be related to more than one factor, and the structural treatments are given
by the factor layout in that stage.

A consistent way of defining each structure is proposed in this chapter. This
unified way of describing a design is not found in the literature. The structure of
the design is conveyed in the form of \( f(x) \) and \( g(\zeta) \). A full design structure can
be understood as the merging of a treatment and a variability structure.

\textbf{Definition} A design is described by specifying its treatment and variability struc-
tures separately. The full structure is given as
\[
\{m\text{-factor} \ (1^{st}\text{layout}, \ 2^{nd}\text{layout}, \ldots)\}\{v\text{-stage} \ (1^{st}\text{layout}, \ 2^{nd}\text{layout}, \ldots)\}.
\]

The first curly bracket specifies the treatment structure and the second one
specifies the variability structure. In the treatment structure, the total number
of factors in all the layouts equals $m$. In the variability structure, the number of layouts is $v - 1$ irrespectively of the number of factors per layout. The layouts in the treatment structure are independent from the ones in the variability structure, i.e. the $1^{\text{st}}$ layout in the left bracket is different from the $1^{\text{st}}$ layout in the right bracket, etc. For each structure, the $1^{\text{st}}$ layout corresponds to the treatments with the largest units, the $2^{\text{nd}}$ to the second largest units and so forth.

The full design structure is formed by combining different factor layouts in both sub-structures. Each layout can be crossed (C) or nested (N). If all the layouts in the structure are crossed, it means that there is no hierarchy. The details of these definitions are presented next.

### 4.3 Primary structures

#### 4.3.1 m-factor layout

The $m$-factor layout is not a design structure on its own. It means that $m$ factors have been combined or arranged in a well specified manner. The result of the combination is called a layout and each element is a treatment. A layout is defined in the same way for fixed and random factors.

The $m$-factor terminology is used here to say that the factors belonging to a stage of a hierarchy are arranged. A complete layout defines the treatments in the corresponding stage. Therefore, one set of factor layouts will define the observational treatments and a different set will define the structural treatments.

In relatively simple cases, the $m$-factor layout describes the full design structure, e.g. a \{m-factor\}{1-stage} design. The 1-stage variability corresponds to the random error. This represents a fixed model and there is no need to mention the variability structure.

Further, a set of $w$-factor layouts can be accommodated in one single hierarchy
stage, thus determining the structure of that stage. Then it can be said, e.g. that
the structural treatments of the top level of a 2-stage variability structure are given
by a $w$-factor crossed layout (defined below). Commonly, for the observational
units the layout defines a blocking scheme unless a completely randomised array
(crossed layout) is suitable.

The two main layouts are the crossed (C) and the nested (N). Using these two
repeatedly in combination allows the description of complex design structures.

### 4.3.2 Crossed layout

The basic design layout involves crossed factors. Fully crossed, or factorial, means
that all levels of all factors are combined and all resulting treatments can be
included in the design (see Section 2.1.1). These arrangements are called **crossed
layouts** (C). This layout is appropriate when all the involved factors have the
same relative importance.

The fully crossed layout is the simplest case, thus forming the basic building
block of a design. This layout can be used anywhere in complex structures.
Namely, a number of crossed layouts can be combined to form complicated designs.
One crossed layout can be placed at each treatment of another crossed structure,
and so on.

The simplest possibility is the 1-factor layout (1C) which means that each fac-
tor level is an available treatment. In practice, these levels have to be randomised.
The 2-factor(2C) layout is the elementary array. It is done by using each level of
the first factor with all the levels of the second factor. Consider the following case

\[ x = (x_1, x_2) \quad \ell_{x_1} = (-1, 0, 1) \quad \ell_{x_2} = (-1, 1) \]

that after being arranged as a 2-factor crossed layout, denoted by 2-factor(2C),
gives the treatment set $t_i$ shown in Table 4.1. This configuration can be extended easily by using more factors.

| Table 4.1 2-factor crossed layout i.e. 2-factor(2C) |
|-----------------|-----------------|
| Treatment $t_i$ | $\ell_{x_{1}}$ | $\ell_{x_{2}}$ |
| $t_1$            | -1              | -1              |
| $t_2$            | -1              | 1               |
| $t_3$            | 0               | -1              |
| $t_4$            | 0               | 1               |
| $t_5$            | 1               | -1              |
| $t_6$            | 1               | 1               |

Not many factors are needed to create complex arrangements. Consider three factors. Using crossed layouts, a completely randomised block design (CRB) is the simplest combination. The standard case has one blocking factor and two experimental factors. Each level of the blocking factor constitute a treatment and forms a block; the order of the blocks is randomised, i.e. a randomised 1-factor(1C) crossed layout. The other two factors form a 2-factor(2C) layout which is repeated inside each block using independent randomisation sequences. In each block, the observational units receive the treatments given by the 2-factor crossed layout. It is clear how the simple CRB design is a combination of two crossed layouts. A CRB design with $u$ blocking factors and $w$ experimental factors can be denoted as a $\{u+w\text{-factor}(uC, wC)\}$ structure. The case $\{3\text{-factor}(1C, 2C)\}$ is shown in Figure 4.1.

### 4.3.3 Nested layout

**Nested layouts** (N) are those that involve nested factors. Hence, units in one stage are unique to the units of the previous stage, which are themselves unique to the stage above, and so on. One level of nesting is the least complicated structure,
4 Design structures

4.3 Primary structures

Figure 4.1 The \{3-factor(1C, 2C)\} design structure

which involves two stages. The simplest case has the factors of the two stages defined as fixed. The next simple case has random factors in the top stage, and fixed factors in the bottom stage. The converse may also be applicable.

The basic nested structure requires two factors, one per stage. The second factor corresponds to the bottom stage, and is placed at each level of the first factor. The latter is the top stage. The units used at another factor level in the top stage are different from those used at any other factor level. The terminology is that the second factor is nested within the first one. Note that the randomisation is restricted (see Section 4.4).

The nested layout is the second building block for design structures. As in the crossed case, nested structures combine to form more complex ones. For the simplest combination think of one factor for the top stage and two factors in the bottom stage. Consider the following

\[
\mathbf{x} = (x_1 \ x_2 \ x_3) \quad \ell_{x_1} = (-1 \ 0 \ 1) \quad \ell_{x_2} = (-1 \ 1) \quad \ell_{x_3} = (-1 \ 1).
\]

The two bottom stage factors can be arranged in a 2-factor crossed layout which is applied to each level of the top stage factor. The top stage factor is a 1-factor layout (not necessarily randomised). The resulting treatments of the array are shown in Table 4.2.

Therefore, unlike the CRB design, the placement of the observational units
4.3 Primary structures

### Table 4.2 3-factor nested layout i.e. 3-factor(1N, 2C)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Factor levels</th>
<th>Treatments $t_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\ell_{x_1}$</td>
<td>$t_{11}$ $t_{21}$ $t_{31}$ $t_{41}$ $t_{12}$ $t_{22}$ $t_{32}$ $t_{42}$ $t_{13}$ $t_{23}$ $t_{33}$ $t_{43}$</td>
</tr>
<tr>
<td>2</td>
<td>$\ell_{x_2}$ $\ell_{x_3}$</td>
<td>-1 -1 1 1 -1 -1 1 1 -1 -1 1 1</td>
</tr>
</tbody>
</table>

from the 2-factor crossed layout is not random because the levels of the first factor are unique; however the run sequence of the 2-factor layout observational treatments is randomised. This arrangement is referred to as the 3-factor(1N, 2C) layout and is shown in Figure 4.2.

![Figure 4.2](image-url)

**Figure 4.2** The \{3-factor(1N, 2C)\} design structure

### 4.3.4 Model formulation

Consider the crossed layout. The form of the statistical model depends on the type of factors used. Assume $m$ factors; if all are assumed fixed, then $x$ will contain all of them. The experimenter decides which response surface model to use and $f(x)$ is defined. In the example of Section 4.3.2 there is a 2-factor crossed layout inside a 1-factor layout that is known as the CRB design. Then it can be described as a \{3-factor(1C,2C)\}\{1-stage\} design structure for a fixed effects model.

If the blocking factor is random, i.e. $s = 1$, then $m = 2$. The fixed part $f(x)$ is defined in the same way as above. Now in addition there is a term $g(\zeta_j)$, such
that
\[ g(\zeta_j) = I_{t_{is}} |_{s=1} \]

and additionally \( g(\zeta_j) \) holds information about how many observations are taken at each level of the blocking random factor (\( i.e. \) in each block). In general, \( Z \) is formed by vertically stacking the \( r \) matrices defined by \( g(\zeta_j) \). This mixed model can be described as a \{2-factor(2C)\}\{2-stage(1C)\} structure design. As given by the notation, the variability structure has a 1-factor crossed layout in the first stage; the bottom variability stage is always the random error to which a specific structure is not attached. The extensions to more factors are straightforward.

For the nested layout, the construction of the model follows the same ideas as above, with only a few changes. For the fixed part one column is added in \( X \) for each level of the nesting factor (top stage). For the random part, the matrices given by \( g(\zeta_j) \) are stacked in a block diagonal manner instead of vertically. Again, it is easy to extend the methodology to more factors.

The crossed and nested layouts are the building blocks to form complex design structures. Any type of structure can be constructed using these building blocks. Useful “intricate” structures are presented shortly.

4.4 Split units and restricted randomisation

It is not rare to find conditions that require the use of different sizes of units throughout the experiment. Examples include the case when some factors are more difficult to control or dealt with than others, and when the experimental runs have to be spread over different sites and/or different occasions. These conditions can be described using hard to change (HC) factors.

The practical impact of HC factors is that they cause the experimental units to be split. The \textit{split-unit} principle invokes a hierarchical structure. The idea is
Design structures 4.4 Split units and restricted randomisation

simple: considering the characteristics of the factors affecting the units

s) Apply a corresponding structural treatment at random to the largest structural unit. Assign the random error term $\sigma_s^2$.

s-1) Split the largest unit into smaller structural units; apply structural treatments at random to the split units. Assign the random error term $\sigma_{s-1}^2$.

;:

0) Split units into observational units and apply observational treatments at random. Assign the random error term $\sigma_0^2$.

It is required to set a structural treatment $t(j)$ defined by the HC factors, run all the required trials under that treatment at once, then split the unit to apply a different set of treatments on the smaller units, and repeat the process until the observational treatments $t_i$ are applied; EC factors are used to define $t_i$. The HC treatment is changed afterwards as needed, and the procedure is done as many times as HC treatments and stages have to be studied.

Splitting the units means changing their size. Then, each set of unit size can only be assigned to one HC structural treatment and this in turn is done at one specific hierarchical stage. Another split will correspond to another stage of the hierarchy and so on. This requires each stage to have an independent randomisation process. Therefore, units cannot be freely assigned to any part of the structure. This is known as restricted randomisation and it implies that different error terms (variance components) have to be assigned to each stage.

Restricted randomisation justifies the need for a hierarchical structure. It is rare though to use more than two splits. The case of two splits defines a 3-stage hierarchical structure as depicted in Figure 4.3. One replication can be seen and the splitting of the units is shown.

Blocking is done to deal with the heterogeneity of experimental units, with the
differentiating characteristics being identifiable. Blocking restricts the randomisation; it can only be done within blocks. In a simple case, sets of treatments are applied at random to units having similar characteristics separated by the blocks. Each block will follow a different randomisation scheme. Consequently, the random error is formed by the interaction between observational treatments and blocks.

Figure 4.3 is a \{2-factor(2C)\}\{3-stage(1N, 1C)\} structure, also known as Split-Split-Plot design. Four factors are involved. For the variability structure, Stage 1 has two structural units and each one receives one factor level (A, B) as its treatment. Stage 2 is nested within Stage 1, and it has two structural units that receive one factor level as treatment (a, b) for each unit in Stage 1. Two crossed factors define the observational treatments $t_1$ to $t_4$ in the treatment structure which is nested within Stage 2; one 2-factor(2C) layout is used for each structural unit from Stage 2. The randomisation sequences per stage can also be identified.
4.5 Useful structures

4.5.1 Split-Plot structures

The term Split-Plot comes from the agricultural setting, and refers to the split of big pieces of land; the terminology is still in use. Remember that units are split when HC factors are present. Here, the HC are called whole-plot factors and the structural units to which their levels are applied to are named whole-plots. The observational units treated with EC factors, nested within the whole-plots, are called split-plots and the associated factors are known as split-plot factors. The variance components in the 2-stage nested variability structure are called the whole-plot error and the split-plot error, respectively.

A Split-Plot experimental design uses the mixed model. The first procedure of randomisation and its corresponding error variance, is due to the random allocation of the whole-plot structural treatments to the available whole-plots. The second error variance comes from randomly assigning the observational treatments to the split-plot units.

Since the whole-plot effects are random, the experimental units within a whole-plot are correlated. The term $\eta$ is used to denote the degree of correlation and is defined as the ratio of the whole-plot error $\sigma_1^2$ to the split-plot error $\sigma_\epsilon^2$ (see Section 3.4.1).

When there are more hierarchical stages, the small units are split further, for example into split-split-plot observational units. In such case the experimental design contains the whole-plot error, split-plot error and the split-split-plot error (3-stage). In practice, the theoretical limitation difficult to be guaranteed as the number of stages increases, is that there should be at least one degree of freedom for every error term. This imposes the condition that the minimum block size must be two. Certainly, the more treatments available at each stage, the better
the estimates that can be obtained for the respective variance component. This is simply because more observations (information) can be obtained. All this in addition to the available number of replicates.

Designs beyond the Split-Split-Plot structure (three hierarchical stages) do not seem to be common. Federer and King (2007) present the ANOVA model for the Split-Split-Split-Plot design (4 stages), and Casella (2008) mentions the possibility of further splitting, but neither introduce examples of this nor cite other authors’ work with more than two subdivisions.

The difference between standard blocking and Split-Plots, is that the former does not require the units to be divided and there is thus no distinction between structural and observational units. In the second case, this division is key because it gives the experiment a hierarchical structure. Consequently, the total variance is formed by the variability contribution from the upper stage plus the contribution from the lower stage.

Two important aspects must be noted in Split-Plot experiments:

1. There will always be more experimental units at the split-plot stage to apply treatments to. Thus, this stage has the largest number of degrees of freedom in the structure. Compared to the upper stages, estimation procedures will have more power in this bottom stage.

2. In most situations, the whole-plot effects are considered random. It is a key consideration, because it implies that the observations in the same whole-plot have a degree of correlation, this in turn is related to the way the whole-plot error is determined. As a result, and in conjunction with point 1, if the experimenter wants to gain crucial knowledge about the effects of a HC factor, a Split-Plot design is not the correct option. At the split-plot stage, the choice of the type of factor is not mandated by the structure and it can be made according to the conditions and the objectives of the experiment.
There is extensive literature about Split-Plot designs. Most books on design of experiments include the topic, e.g. Dean and Voss (1999), Hinkelman and Kempthorne (2008), Milliken and Johnson (2009), among others. Vining et al. (2005) look at standard response surface designs to be used in a Split-Plot configuration. In contrast, Goos and Donev (2006) use design construction algorithms. Optimality of Split-Plot designs is presented in depth by Goos (2002) including design construction algorithms. Goos and Donev (2007) show a way to combine mixture and process factors in a Split-Plot configuration.

### 4.5.2 Split-Block design

A variation of the standard Split-Plot experimental design arises when the levels of the split-plot factor, instead of being randomised independently within each level of the whole-plot factor, are applied individually across all levels of the whole-plot.

This variation is known as Split-Block experimental design, but not exclusively. For example, Cochran and Cox (1957) refer to it as “sub-unit Treatment in Strips”, Casella (2008) uses the term “Strip plot” and Goos and Gilmour (2012) call it Row-Column design.

The simplest Split-Block design involves two factors. The standard version has the factor A and factor B treatments in a CRBD arrangement. There is no restriction, however, on the type of layout that can be used.

The steps for constructing a Split-Block design are given in Figure 4.4 based on a generic experiment having three levels of factor A, and four levels of factor B. The figure shows only one replicate of the experiment. The description of the design structure is

\[
\{m\text{-factor}(1\text{st layout}, \ldots, v\text{th layout})\}\{v+1\text{-stage}(vC)\}.
\]

For the case in the figure, the structure is \{2-factor(1C, 1C)\}{3-stage(2C)\}. 

117
Note that both factor A and factor B are whole-plot (structural) treatments. The whole-plot A experimental units are different (in size, say) from the whole-plot B experimental units. The levels of factor A are randomly assigned to the whole-plot A experimental units. Then, the levels of factor B are randomly allocated to the whole-plot B experimental units. Next, the two whole-plot treatments are arranged perpendicularly to each other such that the levels of factor A go across all levels of factor B and the levels of factor B go across all levels of factor A.

In the Split-Block design, the treatments applied to the observational units, still the smallest in the experiment, are in fact the result of the interaction between the factor A and the factor B treatments. There are independent randomisation procedures for the factors. Both are performed for each replicate. The most common scenario is when the two factors are considered fixed. But even in this case, there will be different variance components, one associated to each randomisation. These unique features of the design make it a multistage non-hierarchical structure.

A suitable model for the Split-Block experimental design is the mixed model.
The same guidelines to determine the structure by identifying the hierarchies and layouts also apply. Extensive descriptions on this structure and other variations can be found in Federer and King (2007). Design and analysis of this type of structure are presented by Arnouts et al. (2010). Optimality and construction algorithms are given by Arnouts and Goos (2010).

4.6 Summary of structures

This section presents a summary of how the proposed notation is used with the different design structures mentioned. Remember that these can be combined to represent different levels of complexity. The summary is presented in Table 4.3.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossed layout</td>
<td>( (mC) )</td>
<td>(2C)</td>
</tr>
<tr>
<td>Crossed layout with interaction</td>
<td>( (mCi) )</td>
<td>(2Ci)</td>
</tr>
<tr>
<td>Random error</td>
<td>{1-stage}</td>
<td>Not required</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>{m-factor}</td>
<td>{2-factor}</td>
</tr>
<tr>
<td>Random effects</td>
<td>{1}</td>
<td>{3-stage(2C)}</td>
</tr>
<tr>
<td>Mixed effects</td>
<td>{v-stage(1^{st}layout,...,v-1^{th}layout)}</td>
<td>{2-factor(2C)}{2-stage(1C)}</td>
</tr>
<tr>
<td>CRD</td>
<td>{m-factor(mC)}</td>
<td>{3-factor(3C)}</td>
</tr>
<tr>
<td>CRBD</td>
<td>{u + w-factor(uC,wC)}</td>
<td>{3-factor(1C,2C)}</td>
</tr>
<tr>
<td>Nested</td>
<td>{m-factor(1N,m-1C)}</td>
<td>{3-factor(1N,2C)}</td>
</tr>
<tr>
<td>Split-Plot</td>
<td>{m-factor(1^{st}layout,2^{nd}layout,...)}</td>
<td>{2-factor(2C)}{2-stage(1N)}</td>
</tr>
<tr>
<td>Split-Split-Plot</td>
<td>{m-factor(1^{st}layout,2^{nd}layout,...)}</td>
<td>{2-factor(2C)}{3-stage(1N,2C)}</td>
</tr>
<tr>
<td>Split-Block</td>
<td>{m-factor(1^{st}layout,...,v^{th}layout)}</td>
<td>{2-factor(1C,1C)}{3-stage(2C)}</td>
</tr>
</tbody>
</table>

119
4.7 Comparison with alternative descriptions of design structures

The framework proposed in this chapter was developed to provide a way of describing the structure of a design (even in complex cases) in a “single sentence” that can be matched with ease with the statistical model to be analysed. For this reason, the treatment and the variability structures are specified separately so that in the models they correspond with the fixed and the random terms, respectively.

Frameworks with the described objective are not easily found in the literature. Bailey (2008) is explicit in describing the fixed part and the random parts independently. Her framework, given in book length, proposes a sequential scheme for creating a complete design structure. When following the sequence, the variability structure is defined first and the treatment structure last. The same general idea is used in the framework proposed in this thesis. Bailey’s work focuses on comparative experiments i.e. first order fixed effects models. This means that the aim is to analyse the results using ANOVA methods. Bailey uses Hasse diagrams to aid the understanding of the structures.

Goos and Gilmour (2012) present a strategy which aim is more closely related to the one herein. Their Figure 1 is a nice and concise summary of the procedure that defines, separately, the variability structure and the treatment structure. They also covered the analysis portion, which it is not explicitly treated in this thesis. To a very large extend, the same foundations that motivated the framework presented here, were found relevant and challenging by these authors; they state this at the end of the third paragraph of their introduction. It has been unfortunate for the work in this thesis that Goos and Gilmour (2012) paper was published while this document was being written. As the authors recognise, at that time, there was no other proposed framework to achieve our common aim.
Similar to Bailey (2008), Goos and Gilmour (2012) use Hasse diagrams as the pictorial tool for representing design structures. Here, layered figures are used. It is simpler to draw complete Hasse diagrams than layered figures. However in a layered figure, only one complete block or branch of the hierarchy has to be drawn completely and this is repeated for the other branches. If one branch is different, then it is also drawn. Some structures can be fully represented. Information about the factors, the layouts and the randomisation accompanies the layered figure. The Hasse diagram includes the degrees of freedom and the layout can be read from the diagram. The layered figure is mainly intended to help in the definition of the model whilst the Hasse diagram is highly oriented to help in the analysis. The layered figure is also expected to guide the execution of the experiment. This is more difficult to achieve with the Hasse diagram. In summary, both types of graphics have advantages however they are meant to help in different areas. Concurrent use of both tools can be very effective during an experimental phase. Table 4.4 shows a summary of the main differences.

**Table 4.4  Features of Layered figures and Hasse diagrams**

<table>
<thead>
<tr>
<th>Layered figure</th>
<th>Hasse diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>“One sentence” expression to describe the structure</td>
<td>———</td>
</tr>
<tr>
<td>Aid the definition of the model</td>
<td>Aid the analysis</td>
</tr>
<tr>
<td>Try to help include the required terms in the model</td>
<td>Try to help construct an ANOVA table</td>
</tr>
<tr>
<td>Drawing may be elaborated for some structures requiring longer time</td>
<td>Easy and quick to draw</td>
</tr>
<tr>
<td>All structures look different so they can be identified visually</td>
<td>Some structures look the same e.g. CRBD and Split-Plot; they can’t always be identified visually</td>
</tr>
<tr>
<td>Additional information about the factors and their layout is given explicitly</td>
<td>Information about degrees of freedom given explicitly; layout is only read in the diagram</td>
</tr>
</tbody>
</table>

121
Three examples are presented below that show how a given structure would be represented in the framework proposed in this thesis and compared with a framework based on Hasse diagrams.

4.7.1 Split-Plot design

Example 3 Drug dose bioassay

Consider a bioassay in which certain compounds are tested to find out what is the efficacy of each one. The response is the relative growth of culture. Seven compounds are involved in the experiment, each one at seven different doses plus a control dose with no compound. The preparations under test are to be placed in the wells of laboratory plates and each compound will be tested on three plates. Experiments will be repeated on three different occasions, using different stock of each compound every time.

A Split-Plot experimental design is well suited for this situation. The response is the measured growth at each well. There are four possible experimental factors to consider. The first one may be the occasions with three levels, then the plates with three levels, next is the compounds used with seven levels, and finally the dose with a total of eight levels. The occasions, however, are not very suitable to be included as a treatment factor. Since every occasion uses all the material in a compound stock and compounds from different stock show different characteristics due to conditions in storage, date of production and preparation time, it is reasonable to consider the occasions as forming three blocks, or using the Split-Plot terminology, occasions are whole-plots that are then replicated three times. The experimental units \textit{i.e.} the wells on the plates, are nested within the whole-plots.

It is known that the variability of the response in the wells of the same plate may be very small, but this does not reflect the practical use of the compounds. So, there is interest in having different plates, and three plates will make use of
all the compound from one stock of material. Consequently, the factor “plate” with three levels, is nested within occasions. Since occasions are the whole-plots, then each plate is a level of the whole-plot factor “plate”. The first randomisation sequence takes place here.

The compounds and the doses are at the split-plot stage i.e. there are two split-plot factors. Since the objective is to assess the overall effect of the compounds, the compound/dose combination in every well i.e. the observational units, are obtained from a 2-factor crossed layout. There is now a second randomisation sequence here.

This experiment can be properly conducted as a Split-Plot design with occasions as three whole-plots, plate is the whole-plot factor with three levels; the physical plates are the structural units. Compound with seven levels and dose with eight levels are the split-plot factors, respectively, giving 56 treatments from a 2-factor(2C) layout. The observational units are the wells of each plate.

Using the “one sentence” notation, this design structure is described as \{2-factor(2C)\}{2-stage(1N)}. The layered figure is shown in Figure 4.5. The Hasse diagram is presented in Figure 4.6. The simplicity of the Hasse diagram

---

**Figure 4.5** The \{2-factor(2C)\}{2-stage(1N)} Split-Plot
is remarkable. However, more information can be obtained at a glance on the layered figure. Other differences are as summarised in Table 4.4.

![Hasse diagram of a Split-Plot structure](image)

**Figure 4.6** Hasse diagram of a Split-Plot structure

### 4.7.2 Split-Split-Plot design

**Example 4**

The data come from an experiment to test the effect of ozone gas on plants. This example is given in Casella (2008). There is a measured response variable $y$, and the different experimental factors are Chamber, Ozone, Variety, and Location.

The experiment was conducted by assigning two environmental chambers to each of four concentrations of ozone ($O_3$). Eight chambers were used. In each chamber, six varieties of plants were placed. After running the process, the plants were removed and examined. Data were recorded for two locations on each plant: one at the root (R) and one at the top (T). A name for the response was not indicated.

The experiment is laid out as a Split-Split-Plot design. It is important to notice how the blocks are constructed. The whole-plot treatments are the ozone concentrations, but they are arranged as a CRD with two observations per treatment, these observations being the environmental chambers. This implies that the whole-plot error is represented by the replication of the whole-plot treatments, *i.e.* ozone concentration.
The experiment has four whole-plots (or replicates), but due to the CRD configuration of ozone concentrations and chambers, only two “full” whole-plots are formed because there are two chambers in each block. The concentrations of ozone alone constitute the whole-plot factor A with \( a = 4 \) levels, the plant varieties are randomised in every chamber making variety the split-plot factor B with \( b = 6 \) levels, and the last factor, location, is a physical division of the experimental units (plants) into two parts forming the split-split-plot factor C at \( c = 2 \) levels. It seems appropriate to consider the Chamber effects as random, since the plants can get affected in any location. For the Ozone factor, it makes sense to focus on critical concentrations that could represent critical cases, e.g. minimum, medium, normal and large, and therefore Ozone is considered to have fixed effects. The same rationale follows for factors B and C.

The structure of this design is described as \{4-factor(2Ci, 1C, 1C)\}\{3-stage(1N, 1N)\} using the proposed concise notation. The layered figure for this structure appears in Figure 4.7. Note that the crossed layout at the top stage is well described. The corresponding Hasse diagram is shown in Figure 4.8. The top stage layout is also evident. More specific differences follow those listed in Table 4.4.
4.7.3 Split-Block design

Example 5

This is a hypothetical example based on industry experience. Injection moulding is a manufacturing process in which plastic materials are formed into specific shapes. The process requires a press. The press holds a mould whose different cavities give the plastic the desired shape. Plastic granules are fed into the press through a heater barrel that melts them and injects the softened plastic into the mould. After a cooling stage, the plastic hardens and the moulded parts can be collected. The quality of the final parts depends on the type of plastic used, the settings of the press and the specific characteristics of the mould.

Figure 4.8 The Hasse diagram for a Split-Split-Plot

Figure 4.9 The \{2-factor(1C,1C)\}\{3-stage(2C)\} Split-Block
In one city, a company is updating its facilities for security and productivity reasons. The supplier offers presses featured with different technologies for controlling the process parameters. Since the offer is interesting, the company’s managers decide to run an experiment to see which new technology enhances the cooling schemes that are being tested in the moulds.

For each of the three working shifts, the manufacturing process will be run with the four moulds tested in each of three different presses being offered. That is, the same press will produce parts using the four different types of moulds.

To fit the description into a Split-Block structure, the personnel shifts are identified as three replicates. The four different types of moulds form the four levels of factor A. The injection technologies are factor B and the three machines with different technologies give three levels. Note that the interest is not alone on the presses nor the moulds, but rather the combined effect, i.e. the interaction. So, the basic Split-Block design is indeed suitable for this experiment.

This Split-Block structure is described in the shot hand notation being introduced as \{2-factor(1C, 1C)\}\{3-stage(2C)\}. The respective layered figures is shown in Figure 4.9 whilst the Hasse diagram is presented in Figure 4.10. As in the previous examples, the differences can be related to those in Table 4.4

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{hasse_diagram.png}
\caption{The Hasse diagram for a Split-Block}
\end{figure}
4.8 Summary

The linear mixed model is presented in a different notation to what is customary in the literature. The model is made more general by explicitly stating that the stochastic part can have any desired structure. Such structure is defined according to the design matrices associated to each random regressor which themselves can be constructed in any way specified by the functional \( g(\zeta | t_j) \). For example, \( g_{\zeta_1}(\cdot) \) can correspond to an identity matrix, \( g_{\zeta_2}(\cdot) \) can be the results of the Kronecker product of vectors indicating the number of plots and observations, and \( g_{\zeta_3}(\cdot) \) could have no systematic pattern. Furthermore, fixed factors can also be included in the definition of \( g(\cdot) \). This notation also makes exclusive use of the word treatment to specify the settings of the factors since it applies equally for the fixed and random factors. This also reinforces the notion that the design matrices \( X \) and \( Z \) are the sets of treatments needed or used in the experiment.

Additionally, some definitions are proposed to help in constructing a direct link between the structure of an experimental design and its associated model. The developed concepts include the clear distinction between the treatment and the variability structures (fixed and random, respectively). Any design can be constructed by separate determination of each structure and combining them afterwards. It is important that the same approach is used to define both. A consequence is that the experimental units corresponding to each structure also have to be distinguished explicitly. Hence, the concepts of observational and structural units are introduced. The approach mentioned recourses to the repeated use and combination of simple design structures: crossed (C) and nested (N). Such combinations form layouts. Each of these can have any number of factors and it is important to identify that, so the \( m \)-factor layout is introduced. All the above materialises into a new notation that conveys at once all the information needed to understand the structure of the experimental design and the set up of
the analysis for the collected data. It provides experimenters with an approach to think about the data generation process and the model definition concurrently. The notation is

\[
\{m\text{-factor}(1^{\text{st}}\text{layout}, 2^{\text{nd}}\text{layout}, \ldots)}\{v\text{-stage}(1^{\text{st}}\text{layout}, 2^{\text{nd}}\text{layout}, \ldots)}\,
\]

where the information in the first curly brackets defines the treatment structure, while that in the second defines the variability structure. This “one sentence” notation should generally be paired with a layered figure that shows the stages of the design structure.

Since the model is presented in more detail with most elements defined explicitly, it sets a strong framework for describing complex statistical design problems. The methods for designing experiments can also be implemented more naturally following the notation introduced.

Examples of the use of the new notation are given by applying it to well known design structures such as the randomised block design and the Split-Plot family of designs. The layered figures are also included. The proposed framework is compared with Hasse diagrams. The comparison reveals that both approaches can be used concurrently to create a stronger framework for describing complex experimental phases.
Some information matrices for variance components

Good parameter estimates are always desirable. The variance of the estimators is of importance as it allows to assess the quality of the estimates. Good quality mainly refers to minimum variance and unbiasedness. In this chapter, the objective is to establish what is needed to find designs that minimise the variance of the estimates of the variance components or their ratios. The amount of bias is mainly determined by the estimation method.

For illustration, the 2-factor random model is explored using two different structures. The purpose is to understand the behaviour of the variance components and how this impacts the construction of an optimum design. For this reason, the focus is on the expressions for the information matrices. Additionally, the most common optimality criteria are restated for the variance components’ case. Estimation via maximum likelihood always permits to find the asymptotic information matrix of the estimates. This allows to construct an optimum design for a specified model and design structure. However, even for some simple models, closed form expressions for the estimates and their information matrix can not be obtained. Although this is not a severe limitation given the current computing
resources, it is always informative to have closed form equations. Often, the process of determining such expressions helps to better understand the methodology, partly because some limitations can then be realised. Using these expressions instead of general forms can facilitate and speed up the computational workload.

In the case of the MLE and REML methods, the variance of the parameters is found as the inverse of the information matrix given in equations (3.30) and (3.38), respectively. However, in what follows MLE is used since closed form expressions for REML do not exist in general. In this chapter consider $M[\theta \sigma^2]^T = M$.

## 5.1 Design optimality for $\sigma^2$

The theory of design optimality was presented in Section 2.4. Well known criteria are $A$, $V$- and $D$-optimality. Greater details on these criteria and others are found in the aforementioned section and in Atkinson et al. (2007). The discussion herein focuses on expressions for $A$- and $D$-optimality for the variance components.

### 5.1.1 Review of $A$- and $D$-optimality

$A$-optimality was introduced in Section 2.4.1 in terms of the eigenvalues of the information matrix. An alternative expression for this criterion function is

$$\Psi(M) = \text{tr} \ M^{-1},$$

i.e. the trace of the inverse of the information matrix. The corresponding design optimality criterion requires the minimisation of this function.

$D$-optimality specifies the function

$$\Psi(M) = |M|,$$
namely the determinant of the information matrix. For design construction, the optimisation should maximise such determinant.

5.1.2 $A_V$- and $D_V$-optimality

The Generalised $D$- criterion was introduced in Section 2.4.4 for the fixed model case. Accordingly, since $D$-optimality can be easily extended by using linear functions of the determinant of the information matrix, optimality criteria expressions will be written separately for the fixed parameters and the variance components.

The MLE information matrix of the mixed model was given in equation (3.30). It is easy to appreciate that the estimation of the fixed parameters and of the variance components can be done independently since (3.30) is block diagonal. Define $A_\theta = (I_p \ 0)$ and $A_{\sigma^2} = (0 \ I_{r+1})$. Then under the mixed model framework

$$A_\theta^T M A_\theta = M(\theta),$$

which is the $(1,1)$ element of (3.30). Similarly,

$$A_{\sigma^2}^T M A_{\sigma^2} = M(\sigma^2) \quad (5.1)$$

corresponds to the $(2,2)$ element. With this separation, it is straightforward to write, for example, the $D_\theta$-optimality criterion which maximises $|M(\theta)|$, and the $D_V$- criterion which maximises $|M(\sigma^2)|$.

Direct evaluation of $D_\theta$- is possible provided the matrix $V$ is available or has been estimated. In contrast, the evaluation of optimality criteria based on (5.1) is only possible if preliminary values for the vector of variance components are given, $\sigma^2_0$ say. This implies that the constructed designs will only be locally optimum.
Accordingly, a better description of the $D$- criterion is

$$D_V = \max \left| \left. M(\sigma^2) \right|_{\sigma_0^2} \right|, \quad (5.2)$$

and for $A$-optimality the corresponding expression is

$$A_V = \min \left| \left. \text{tr} \ M^{-1}(\sigma^2) \right|_{\sigma_0^2} \right|, \quad (5.3)$$

### 5.1.3 Pseudo-Bayesian optimality

The idea behind this alternative is to make use of all the information about the variance components that may be available. Different versions of Bayesian optimality criteria are available in the literature, e.g. Kessels et al. (2008) and Atkinson et al. (2007). A Bayesian $D$-optimality criterion can be stated as

$$D_B = \int_{\sigma^2} \log |M(\sigma^2)| p(\sigma^2) \ d\sigma^2, \quad (5.4)$$

and an $A$-optimality criterion as

$$D_A = \int_{\sigma^2} \text{tr} \left( M^{-1}(\sigma^2) \right) p(\sigma^2) \ d\sigma^2. \quad (5.5)$$

Since the optimality of designs for variance components is similar to the case of optimality for nonlinear models, the dependance of the information matrix on $\sigma^2$ can be dealt with by using a point prior instead of the prior distribution. However, for the present case, the strategy to be followed is that presented in Section 18.5 of Atkinson et al. (2007) where sampled values from the prior distribution are used to replace the integral in (5.4) and (5.5) by summing over the sampled values. Assuming $p(\sigma^2)$ to be a uniform distribution from which the sample preliminary values of the variance components are taken, the optimality criteria become for
5 Some information matrices for variance components

5.2 The $\{1\}{3\text{-stage}(2\text{.})}$ structure using $\sigma^2$

$D$-optimality

$$D_V = \min_{\sigma^2 \in \Sigma} \left| M^{-1}(\sigma^2) \right|_{\sigma^2 \in \Sigma},$$

and for $A$-optimality

$$A_V = \min \text{tr} M^{-1}(\sigma^2) \bigg|_{\sigma^2 \in \Sigma}.$$

The certainty on preliminary values is generally low and there is not a established approach to define them. Given the simplification of $p(\sigma^2)$ to a uniform distribution, the assumed sample of preliminary values and its uncertainty, the above criteria is considered to be pseudo-Bayesian. Other definitions of pseudo-Bayesian optimality are available in the literature. A good starting point to survey the options are the references given in this section.

For the proposed pseudo-Bayesian criteria, $\Sigma$ represents the subset of $\mathcal{Z}$, i.e. the parameter space of the variance components, that is determined by the a priori values to be investigated. The generality of the approach allows these a priori values to consist of the whole $\mathcal{Z}$ (paralleling the idea of using a non-informative prior in formal Bayesian settings).

5.2 The $\{1\}{3\text{-stage}(2\text{.})}$ structure using $\sigma^2$

The most employed model for variance components uses the balanced 1-factor layout with the factor being random. It is found across the literature to explain the statistical theory and applications involving variability structures. However, despite the random 2-factor layout possibly being equally needed in practice, it is not reported as frequently. Explicit expressions for the MLE information matrix using 2-factor structures will be established for some cases. Then, the optimality criteria described in Section 5.1 can be used and locally optimum designs be constructed. Note that all the random models that will be presented consist of a 3-stage variability structure.
5 Some information matrices for variance components

5.2 The \{1\} (3-stage(2C)) structure using \( \sigma^2 \)

Customary assumptions as given in equations (3.13)–(3.15) will be used. Thus, the information matrix in (3.30) reduces to

\[
M = \begin{bmatrix} X^T V^{-1} X & 0 \\ 0 & \frac{1}{2} \text{tr} \left( V^{-1} Z_i Z_i^T V^{-1} Z_j Z_j^T \right)_{i=j=0, \ldots, r} \end{bmatrix}
\]

(5.6)

and the one in equation (3.40) takes the form

\[
M \begin{bmatrix} \sigma^2_c \\ \eta \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \frac{N}{\sigma^2} & \frac{\text{tr} \left( H^{-1} Z_i Z_i^T \right)^T}{\sigma^2} \\ \frac{\text{tr} \left( H^{-1} Z_i Z_i^T \right)}{\sigma^2} & \text{tr} \left( H^{-1} Z_i Z_i^T H^{-1} Z_j Z_j^T \right) \end{bmatrix}
\]

(5.7)

since using Appendix A.1, equation (3.15) yields

\[
\frac{\partial V}{\partial \sigma^2_i} = Z_i Z_i^T.
\]

5.2.1 The \{1\} (3-stage(2C)) structure

Starting with the general model equation (2.9), a random model is specified by setting \( X = 1_N \) and \( \theta = \mu \). The 2-factor crossed model is fully described in Searle et al. (1992) and it is reproduced in this section.

Let the random factors be \( \zeta = (\alpha, \beta) \) with \( \ell_\alpha = a \) and \( \ell_\beta = b \), respectively. The model in question can be expanded as

\[
y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} = \mu 1 + Z_1 \alpha + Z_2 \beta + Z_0 \epsilon
\]

(5.8)

where a treatment is given by \( t_{ij} = (\ell_\alpha^{(i)} \ell_\beta^{(j)}) \) with \( i = 1 \ldots a, j = 1 \ldots b \). The number of observations per treatment is \( k = 1 \ldots n_{ij} \) so that \( \sum_i^{a} \sum_j^{b} n_{ij} = N \). When the data are balanced, \( n_{ij} = n \forall i \text{ and } j \) i.e. all block sizes are equal. Also,
5.2.2 The \{1\}\{3-stage(2Ci)\} structure

As in the case with out interaction, the 2-factor crossed with interaction model is fully described in Searle et al. (1992). Refer to Section 4.7 of the book for the expression of the variance-covariance matrix. Since it is given as an inverse, the information matrix is obtained directly.

5.2.3 The \{1\}\{3-stage(2N)\} structure

For the 2-factor nested layout, Searle et al. (1992) present the variance-covariance matrix. Here, closed form expressions for the information matrix are derived following the methodology presented in Searle and Henderson (1979).
The model under consideration is:

\[ y_{ij} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk} \]
\[ = \mu 1 + Z_1 \alpha + Z_2 \beta + Z_0 \epsilon \] (5.10)

where a treatment is given by \( t_{ij} = (\ell_i(1) \ell_j(2)) \) with \( i = 1 \ldots a, j = 1 \ldots b \) and \( k = 1 \ldots n_{ij} \) so that \( \sum_i \sum_j n_{ij} = N \), and \( r + 1 = 3 \). When the data are balanced, \( b_i = b \) and \( n_{ij} = n \forall \ i, j \). Consequently, the model equation can be expanded using Kronecker products. For example, \( (1_a \otimes 1_b \otimes J_n) \) is the triple Kronecker product of a vector of ones of size \( a \), the identity matrix of size \( b \), and an \( n \times n \) matrix of ones. This notation will be used throughout.

In matrix form the model can then be represented as

\[ y = (1_a \otimes 1_b \otimes 1_n)\mu + (1_a \otimes 1_b \otimes 1_n)\alpha + (1_a \otimes 1_b \otimes 1_n)\beta + (1_a \otimes 1_b \otimes 1_n)\epsilon \] (5.11)

which implies

\[ Z_1 = (1_a \otimes 1_b \otimes 1_n) \quad Z_2 = (1_a \otimes 1_b \otimes 1_n) \quad Z_0 = (1_a \otimes 1_b \otimes 1_n) \]

and therefore

\[ Z_1^T Z_1 = (1_a \otimes 1_b \otimes 1_n) \]
\[ Z_2^T Z_2 = (1_a \otimes 1_b \otimes 1_n) \] (5.12)
\[ Z_0^T Z_0 = (1_a \otimes 1_b \otimes 1_n) , \]

that leads to

\[ V = (1_a \otimes 1_b \otimes 1_n)\sigma^2_\alpha + (1_a \otimes 1_b \otimes 1_n)\sigma^2_\beta + (1_a \otimes 1_b \otimes 1_n)\sigma^2_\epsilon . \] (5.13)
5 Some information matrices for variance components

5.2 The $\{1\}\{3\text{-stage(2)}\}$ structure using $\sigma^2$

Remember from equation (5.6) that the information matrix for the variance components is given by $\frac{1}{2} \text{tr} \left( V^{-1} Z_i Z_i^T V^{-1} Z_j Z_j^T \right)$ but for the calculation task here, the following identity will be used

$$\text{tr} \left( V^{-1} Z_i Z_i^T V^{-1} Z_j Z_j^T \right) = \text{sesq} \left( Z_i^T V^{-1} Z_j^T \right)$$

where $\text{sesq}(\cdot)$ is the sum of the squared elements of the matrix inside the brackets.

The procedure begins with calculating $V^{-1}$.

Define,

$$\Sigma^T = \begin{bmatrix} \sigma^2_e & \sigma^2_{\beta} & 0 & \sigma^2_a & 0 & 0 & 0 \\ \end{bmatrix}, \quad (5.15)$$

$$\delta = T \Sigma = \begin{bmatrix} 1 & 0 \\ 1 & a \\ 1 & b \\ 1 & n \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 1 & b \\ 1 & n \end{bmatrix} \Sigma \quad (5.16)$$

where, as before, $\otimes$ denotes a Kronecker product. It is to note that $\Sigma_0$, $\Sigma_1$ and $\Sigma_{11}$ correspond to the expected mean squares for each variance component as in the ANOVA method of estimation. Now define

$$v = \frac{1}{\rho} \quad \text{and} \quad \delta = T^{-1} v$$
5 Some information matrices for variance components

5.2 The \( \{1\}\{3\text{-stage}(2)\} \) structure using \( \sigma^2 \)

such that (refer to Searle and Henderson (1979) for the interpretation of this expression)

\[
V^{-1} = \sum_{i=0}^{p} \delta_i (J_{np}^{i_p} \otimes J_{np-1}^{i_{p-1}} \otimes \ldots \otimes J_{n1}^{i_1})
\]

where \( p - 1 \) is the number of main effect factors in the model. Then

\[
\delta = \frac{1}{abn} \begin{bmatrix} a & 0 \\ -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} b & 0 \\ -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} n & 0 \\ -1 & 1 \end{bmatrix} v
\]

where

\[
\begin{bmatrix}
abn & \cdots & \cdots & \cdots & \cdots & \cdots \\
-\ab & \ab & \cdots & \cdots & \cdots & \cdots \\
-\an & \an & \cdots & \cdots & \cdots & \cdots \\
a & -a & -a & a & \cdots & \cdots \\
bn & \cdots & bn & \cdots & \cdots & \cdots \\
b & -b & \cdots & -b & b & \cdots \\
n & -n & -n & n & \cdots & \cdots \\
-1 & 1 & 1 & -1 & 1 & -1 & 1
\end{bmatrix}
\left(\begin{array}{c}
\frac{1}{\Sigma_0} \\
\frac{1}{\Sigma_1} \\
\frac{1}{\Sigma_0} \\
\frac{1}{\Sigma_1} \\
\frac{1}{\Sigma_0} \\
\frac{1}{\Sigma_1} \\
\frac{1}{\Sigma_0} \\
\frac{1}{\Sigma_1}
\end{array}\right) = \delta_{000}
\]

For the case being reviewed, the inverse is expressed as

\[
V^{-1} = \delta_{000}(I_a \otimes I_b \otimes I_n) + \delta_{001}(I_a \otimes I_b \otimes J_n) + \delta_{011}(I_a \otimes J_b \otimes J_n).
\]

which after simplification by joining similar terms and carrying out the Kronecker products results in,

\[
= \frac{1}{\Sigma_0} \left[(I_a \otimes I_b) \otimes (I_n - \frac{1}{n} J_n)\right] + \frac{1}{n \Sigma_1} \left[I_a \otimes (I_b - \frac{1}{b} J_b) \otimes J_n\right] + \frac{1}{bn \Sigma_{11}} (I_a \otimes J_b \otimes J_n)
\]
and substituting $C = I - \bar{J}$ where appropriate, the inverse being sought is

$$V^{-1} = \frac{1}{\sum_0} (I_a \otimes I_b \otimes C_n) + \frac{1}{\sum_1} (I_a \otimes C_b \otimes \bar{J}_n) + \frac{1}{\sum_{11}} (I_a \otimes \bar{J}_b \otimes \bar{J}_n). \quad (5.17)$$

The latter notation is very convenient since the Kronecker products of $C$ matrices with either $J$ or $\bar{J}$ are zero. Using (5.14) and (5.17),

$$\text{sesq}(Z_0^T V^{-1} Z_0) = \text{sesq}(V^{-1}) = \frac{ab(n - 1)}{\Sigma_0^2} + \frac{a(b - 1)}{\Sigma_1^2} + \frac{a}{\Sigma_{11}^2} = d_{\epsilon\epsilon}$$

$$\text{sesq}(Z_0^T V^{-1} Z_1) = \text{sesq}(V^{-1} Z_1) = \text{sesq}(V^{-1} (I_a \otimes I_b \otimes I_n))$$
$$= \text{sesq}\left(\frac{1}{\sum_{11}} (I_a \otimes I_b \otimes I_n)\right) = \frac{abn}{\Sigma_{11}^2} = d_{\epsilon\alpha}$$

$$\text{sesq}(Z_0^T V^{-1} Z_2) = \text{sesq}(V^{-1} Z_2) = \text{sesq}(V^{-1} (I_a \otimes I_b \otimes I_n))$$
$$= \text{sesq}\left(\frac{1}{\sum_1} (I_a \otimes C_b \otimes I_n) + \frac{1}{\sum_{11}} (I_a \otimes \bar{J}_b \otimes I_n)\right)$$
$$= \frac{a(b - 1)n}{\Sigma_1^2} + \frac{an}{\Sigma_{11}^2} = d_{\epsilon\beta}$$

$$\text{sesq}(Z_1^T V^{-1} Z_1) = \text{sesq}\left(\frac{1}{\sum_0} 0 + \frac{1}{\sum_1} 0 + \frac{1}{\sum_{11}} (I_a \otimes b \otimes n)\right)$$
$$= \text{sesq}\left(\frac{bn}{\sum_{11}} I_a\right) = \frac{bn^2}{\Sigma_{11}^2} a = d_{\alpha\alpha}$$

$$\text{sesq}(Z_1^T V^{-1} Z_2) = \text{sesq}\left(\frac{1}{\sum_0} 0 + \frac{1}{\sum_1} 0 + \frac{1}{\sum_{11}} (I_a \otimes I_b^T \otimes n)\right)$$
$$= \frac{abn^2}{\Sigma_{11}^2} = d_{\alpha\beta}$$

140
5 Some information matrices for variance components

5.3 The \{1\}{-stage(2\cdot)} structure using \(\eta\)

\[
\text{sesq}(Z_2^T V^{-1} Z_2) = \text{sesq}\left(\frac{1}{\Sigma_0} 0 + \frac{1}{\Sigma_1} (I_a \otimes C_b \otimes n) + \frac{1}{\Sigma_{11}} (I_a \otimes \bar{J}_b \otimes n)\right)
\]
\[
= n^2 \left( \frac{a(b-1)}{\Sigma_2} + \frac{a}{\Sigma_{11}} \right) = d_{\beta\beta}.
\]

Using the following relationships

\[d_{\epsilon \epsilon} = \frac{d_{\alpha\alpha}}{bn} \quad d_{\epsilon \beta} = \frac{d_{\beta\beta}}{n} \quad d_{\alpha \beta} = \frac{d_{\alpha\alpha}}{b}\]

the information matrix can finally be written as

\[
M = \frac{1}{2} \begin{bmatrix}
\hat{\sigma}_r^2 \\
\hat{\sigma}_a^2 \\
\hat{\sigma}_\beta^2
\end{bmatrix} = \begin{bmatrix}
d_{\epsilon \epsilon} & d_{\alpha\alpha}/bn & d_{\beta\beta}/n \\
\frac{d_{\alpha\alpha}}{bn} & d_{\alpha\alpha}/b & d_{\beta\beta}
\end{bmatrix}.
\] (5.18)

5.3 The \{1\}{-stage(2\cdot)} structure using \(\eta\)

This way of presenting the information matrix was introduced in Section 3.4. The process of obtaining the information matrices for the 2-factor crossed and nested layouts is outlined below. The reported expressions are not available in the literature.

5.3.1 The \{1\}{-stage(2C)} structure

When the ratios of the variance components are of interest in the 2-factor layout, the model can still be written as in (5.8). For the crossed case

\[
Z_1 = (I_a \otimes I_b \otimes I_n) \quad Z_1^T Z_1 = (I_a \otimes J_b \otimes J_n)
\]
\[
Z_2 = (I_a \otimes I_b \otimes I_n) \quad Z_2^T Z_2 = (J_a \otimes I_b \otimes J_n)
\]
\[
Z_0 = (I_a \otimes I_b \otimes I_n) \quad Z_0^T Z_0 = (I_a \otimes I_b \otimes I_n).
\]
Using the same procedure as in Section 5.2.3, in the present case the inverse of the variance-covariance matrix is given by

\[ V^{-1} = \Sigma_0^{-1} \left[ (I_a \otimes I_b \otimes C_n) + (C_a \otimes C_b \otimes \bar{J}_n) \right] + \Sigma_{11}^{-1} (C_a \otimes \bar{J}_b \otimes \bar{J}_n) \]

\[ + \Sigma_{12}^{-1} (\bar{J}_a \otimes C_b \otimes \bar{J}_n) + \Sigma_{4}^{-1} (\bar{J}_a \otimes \bar{J}_b \otimes \bar{J}_n). \]

Using \( \Sigma_0 \) as in (5.9) in addition to equation (3.39)

\[ H^{-1} = V^{-1} \Sigma_0. \] (5.20)

and the entries of the information matrix in (5.7) for model (5.8) can be calculated now as shown below.

The (1, 1) block entry of the information matrix is a trivial calculation and needs no special attention. The (2, 1) block entry is the transpose of the (1, 2) entry, so there is no point in computing both independently. Thus, consider the expression

\[ \frac{\text{tr}(H^{-1}Z_iZ_i^T)}{\Sigma_0}, \]

and since \( i = 1, 2 \), this block is a \( 1 \times 2 \) row vector given by

\[ \frac{1}{\Sigma_0} \left[ \text{tr}(H^{-1}Z_1Z_1^T), \text{tr}(H^{-1}Z_2Z_2^T) \right]. \]

Making appropriate use of (5.14) and (5.20)

\[ \frac{1}{\Sigma_0} \text{tr}(H^{-1}Z_1Z_1^T) = \frac{1}{\Sigma_0} \text{tr} \left( H^{-1}(I_a \otimes J_b \otimes J_n) \right) = \text{sesq} \left( \frac{1}{\Sigma_0} H^{-1}(I_a \otimes J_b \otimes J_n) \right) \]

\[ = \text{sesq} \left( \frac{1}{\Sigma_0} \left[ (I_a \otimes I_b \otimes C_n) + (C_a \otimes C_b \otimes \bar{J}_n) \right] (I_a \otimes J_b \otimes J_n) \right) \]
Some information matrices for variance components

5.3 The \{11\}-stage(2.) structure using \( \eta \)

\[
\begin{align*}
&= \text{sesq} \left( \frac{1}{\Sigma_0} \left[ (I_a \otimes I_b \otimes C_n \otimes J_n) + (C_a I_a \otimes C_b J_b \otimes \bar{J}_n J_n) \right] \\
&\quad + \frac{1}{\Sigma_{11}} (C_a I_a \otimes \bar{J}_b J_b \otimes \bar{J}_n J_n) + \frac{1}{\Sigma_{12}} (\bar{J}_a I_a \otimes C_b J_b \otimes \bar{J}_n J_n) \\
&\quad + \frac{1}{\Sigma_4} (\bar{J}_a I_a \otimes \bar{J}_b J_b \otimes \bar{J}_n J_n) \right) \\
&= \text{sesq} \left( \frac{1}{\Sigma_0} \left( C_a \otimes J_b \otimes J_n \right) \right) + \frac{1}{\Sigma_{11}} (C_a I_a \otimes \bar{J}_b J_b \otimes \bar{J}_n J_n) + \frac{1}{\Sigma_{12}} (\bar{J}_a I_a \otimes C_b J_b \otimes \bar{J}_n J_n) \\
&= \text{sesq} \left( \frac{1}{\Sigma_0} \left( C_a \otimes J_b \otimes J_n \right) \right) \\
\end{align*}
\]

and when computing the sums of squares (sesq(·)) of the Kronecker products lead to

\[
\frac{1}{\Sigma_0} \text{tr} \left( \left( H^{-1} Z_i Z_i^T \right) \right) = \frac{(a - 1)bn}{\Sigma_{11}^2} + \frac{bn}{\Sigma_4^2} = bn \left( \frac{a - 1}{\Sigma_{11}^2} + \frac{1}{\Sigma_4^2} \right); \quad (5.22)
\]

similarly,

\[
\frac{1}{\Sigma_0} \text{tr} \left( \left( H^{-1} Z_i^T Z_i \right) \right) = \text{sesq} \left( \frac{1}{\Sigma_0} H^{-1} (J_a \otimes I_b \otimes J_n) \right)
\]

\[
= \text{sesq} \left( \frac{1}{\Sigma_0} \left( J_a \otimes C_b \otimes J_n \right) \right) + \frac{1}{\Sigma_{12}} (J_a \otimes \bar{J}_b \otimes \bar{J}_n J_n) \]

\[
\frac{1}{\Sigma_0} \text{tr} \left( \left( H^{-1} Z_i Z_i^T \right) \right) = \frac{a(b - 1)n}{\Sigma_{12}^2} + \frac{an}{\Sigma_4^2} = an \left( \frac{b - 1}{\Sigma_{12}^2} + \frac{1}{\Sigma_4^2} \right). \quad (5.23)
\]

For the (2, 2) entry, now consider \( i = j = 1 \), so that

\[
\text{tr} \left( \left( H^{-1} Z_i Z_i^T H^{-1} Z_j Z_j^T \right) \right) = \text{sesq} (Z_i^T H^{-1} Z_j)
\]

is a \( 2 \times 2 \) matrix. Proceeding as above for \( i = j = 1 \) leads to

\[
\text{sesq} (Z_i^T H^{-1} Z_i) = \text{sesq} \left( (I_a \otimes 1_b \otimes 1_n)^T H^{-1} (I_a \otimes 1_b \otimes 1_n) \right)
\]
\[
\begin{align*}
&= \text{sesq} \left( (I_a \otimes I_b \otimes 1_n)^T (I_a \otimes I_b \otimes C_n) + (C_a \otimes C_b \otimes \bar{J}_n) (I_a \otimes I_b \otimes 1_n) \right) \\
&\quad + \sum_0 \frac{1}{\Sigma_{11}} (I_a \otimes I_b \otimes 1_n)^T (C_a \otimes \bar{J}_b \otimes \bar{J}_n) (I_a \otimes I_b \otimes 1_n) \\
&\quad + \frac{1}{\Sigma_{12}} (I_a \otimes I_b \otimes 1_n)^T (\bar{J}_a \otimes C_b \otimes \bar{J}_n) (I_a \otimes I_b \otimes 1_n) \\
&\quad + \frac{1}{\Sigma_4} (I_a \otimes I_b \otimes 1_n)^T (\bar{J}_a \otimes \bar{J}_b \otimes \bar{J}_n) (I_a \otimes I_b \otimes 1_n) \right) \\
&= \text{sesq} \left( [(I_a^T I_a \otimes I_b^T I_b \otimes 1_n^T C_n) + (I_a^T C_a \otimes I_b^T C_b \otimes 1_n^T \bar{J}_n)] (I_a \otimes I_b \otimes 1_n) \right) \\
&\quad + \sum_0 \frac{1}{\Sigma_{11}} (I_a^T C_a \otimes 1_b^T \bar{J}_b \otimes 1_n^T \bar{J}_n) (I_a \otimes I_b \otimes 1_n) \\
&\quad + \frac{1}{\Sigma_{12}} (I_a^T \bar{J}_a \otimes 1_b^T C_b \otimes 1_n^T \bar{J}_n) (I_a \otimes I_b \otimes 1_n) \\
&\quad + \frac{1}{\Sigma_4} (I_a^T \bar{J}_a \otimes 1_b^T \bar{J}_b \otimes 1_n^T \bar{J}_n) (I_a \otimes I_b \otimes 1_n) \right) \\
&= \text{sesq} \left( (I_a \otimes I_b^T \otimes 0) + (C_a \otimes 0 \otimes b I_b^T) \right) (I_a \otimes I_b \otimes 1_n) \\
&\quad + \sum_0 \frac{1}{\Sigma_{11}} (C_a \otimes 1_b^T \otimes 1_n^T) (I_a \otimes I_b \otimes 1_n) \\
&\quad + \frac{1}{\Sigma_{12}} (\bar{J}_a \otimes 0 \otimes 1_n^T) (I_a \otimes I_b \otimes 1_n) \\
&\quad + \frac{1}{\Sigma_4} (\bar{J}_a \otimes 1_b^T \otimes 1_n^T) (I_a \otimes I_b \otimes 1_n) \right) \\
&= \text{sesq} \left( 0 + \sum_0 \left[ \frac{1}{\Sigma_{11}} (C_a \otimes 1_b^T \otimes 1_n^T) - \frac{1}{\Sigma_4} (\bar{J}_a \otimes 0 \otimes 1_n^T) \right] \right) \\
&= \text{sesq} \left( \frac{1}{\Sigma_{11}} (C_a \otimes b \otimes n) + \frac{1}{\Sigma_4} (\bar{J}_a \otimes b \otimes n) \right) \\
&= \text{sesq} \left( \frac{1}{\Sigma_{11}} b m C_a + \frac{1}{\Sigma_4} b m \bar{J}_a \right) = \sum_0^2 \left( \frac{(a - 1)b^2 n^2}{\Sigma_{11}} \right) + \frac{b^2 n^2}{\Sigma_4}.
\end{align*}
\]

For the case \(i = j = 2\) the calculations are

\[
\text{sesq} \left( Z_2^T H^{-1} Z_2 \right) = \text{sesq} \left( (I_a \otimes I_b \otimes 1_n)^T H^{-1} (I_a \otimes I_b \otimes 1_n) \right)
\]

\[
= \text{sesq} \left( 0 (I_a \otimes I_b \otimes 1_n) + \sum_0 (I_a \otimes I_b \otimes 1_n) \right)
\]

144
5 Some information matrices for variance components

5.3 The \{1\}\{3-stage(2)\} structure using \(\eta\)

\[
+ \frac{1}{\Sigma_{12}} (1_a^T \otimes C_b \otimes 1_n^T)(1_a \otimes I_b \otimes 1_n)
\]

\[
+ \frac{1}{\Sigma_4} (1_a^T \otimes J_b \otimes 1_n^T)(1_a \otimes I_b \otimes 1_n)
\]

\[
= \text{sesq}\left[ \Sigma_0 \left[ \frac{1}{\Sigma_{12}} (1_a^T 1_a \otimes C_b 1_b \otimes 1_n^T 1_n) + \frac{1}{\Sigma_4} (1_a^T 1_a \otimes J_b 1_b \otimes 1_n^T 1_n) \right] \right]
\]

\[
= \text{sesq}\left[ \Sigma_0 \left[ \frac{1}{\Sigma_{12}} (a \otimes C_b \otimes n) + \frac{1}{\Sigma_4} (a \otimes J_b \otimes n) \right] \right] = \left( b - 1 \right) \frac{\Sigma_2}{\Sigma_4} + \frac{1}{\Sigma_4} \frac{\Sigma_2^2}{\Sigma_4^2}.
\]

Given that \(\text{tr}(H^{-1}Z_iZ_i^TH^{-1}Z_jZ_j^T)\) is a variance-covariance matrix, it is thus symmetric and consequently the expressions for the cases \((i = 1, j = 2)\) and \((i = 2, j = 1)\) yield the same answer. Evaluate

\[
\text{sesq}( Z_1^T H^{-1} Z_2 ) = \text{sesq}( (1_a \otimes 1_b \otimes 1_n)^T H^{-1} (1_a \otimes 1_b \otimes 1_n) )
\]

\[
= \text{sesq}\left[ \left( 0 (1_a \otimes 1_b \otimes 1_n) + \Sigma_0 \left[ 0 (1_a \otimes 1_b \otimes 1_n) + 0 (1_a \otimes 1_b \otimes 1_n) \right] + \frac{1}{\Sigma_4} (J_a \otimes 1_b^T \otimes 1_n^T)(1_a \otimes I_b \otimes 1_n) \right) \right]
\]

\[
= \text{sesq}\left[ \Sigma_0 \left[ \frac{1}{\Sigma_4} (J_a 1_a \otimes 1_b^T 1_b \otimes 1_n^T 1_n) \right] \right] = \text{sesq}\left[ \Sigma_0 \left[ \frac{1}{\Sigma_4} (1_a \otimes 1_b^T \otimes n) \right] \right]
\]

\[
= abn^2 \Sigma_0^2 \frac{\Sigma_2^2}{\Sigma_4^2}.
\]

Before putting together the information matrix, the following relationships involving the above results are defined for notational convenience

\[
h_{\alpha\alpha} = bn \left( a - 1 \frac{\Sigma_2}{\Sigma_{11}} + 1 \frac{\Sigma_2}{\Sigma_4} \right) \quad h_{\beta\beta} = an \left( b - 1 \frac{\Sigma_2}{\Sigma_{12}} + 1 \frac{\Sigma_2}{\Sigma_4} \right) \quad h_{\alpha\beta} = abn^2 \Sigma_0^2 \frac{\Sigma_2^2}{\Sigma_4^2}
\]

and the information matrix (5.7) for model (5.8) can be written as

\[
M_\eta = \begin{bmatrix}
\frac{N}{\Sigma_0} & h_{\alpha\alpha} & h_{\beta\beta} \\
\frac{N}{\Sigma_0} bn \Sigma_0^2 h_{\alpha\alpha} & h_{\alpha\beta} \\
h_{\beta\beta} & h_{\alpha\beta} & an \Sigma_0^2 h_{\beta\beta}
\end{bmatrix}
\]

(5.24)
5.3.2 The \{1\}{3-stage(2N)} structure

For the case of the two-way nested model, equation (5.8) can still be used with the following changes

\[
\begin{align*}
Z_1 &= (I_a \otimes I_b \otimes 1_n) \\
Z_2 &= (I_a \otimes I_b \otimes 1_n) \\
Z_0 &= (I_a \otimes I_b \otimes I_n)
\end{align*}
\]

\[
\begin{align*}
Z_1^T &= (I_a \otimes J_b \otimes J_n) \\
Z_2^T &= (I_a \otimes I_b \otimes J_n) \\
Z_0^T &= (I_a \otimes I_b \otimes I_n).
\end{align*}
\]

The inverse of the variance-covariance matrix, \(V^{-1}\), for this model was found in Section 5.2.3 and reported as equation (5.17), from which \(H^{-1}\) can be written as

\[
H^{-1} = (I_a \otimes I_b \otimes C_n) + \frac{\sum_0}{\sum_1} (I_a \otimes C_b \otimes \tilde{J}_n) + \frac{\sum_0}{\sum_11} (I_a \otimes \tilde{J}_b \otimes \tilde{J}_n) \tag{5.26}
\]

and the process of finding the corresponding information matrix mimics that of the two-way crossed model of Section 5.3.1. Begin by writing

\[
\frac{1}{\sum_0} tr(H^{-1}Z_1Z_1^T) = \text{sesq} \left( \frac{1}{\sum_0} H^{-1}(I_a \otimes J_b \otimes J_n) \right)
\]

\[
= \text{sesq} \left( \frac{1}{\sum_0} \left[ (I_a I_a \otimes I_b J_b \otimes C_n J_n) + \frac{1}{\sum_1} (I_a I_a \otimes C_b J_b \otimes \tilde{J}_n J_n) \right. \right.
\]

\[
\left. \left. + \frac{1}{\sum_11} (I_a I_a \otimes \tilde{J}_b J_b \otimes \tilde{J}_n J_n) \right) = \text{sesq} \left( 0 + 0 + \frac{1}{\sum_11} (I_a \otimes J_b \otimes J_n) \right) \right)
\]

\[
= \frac{abn}{\sum_11}.
\]

Proceed with

\[
\frac{1}{\sum_0} tr(H^{-1}Z_2Z_2^T) = \text{sesq} \left( \frac{1}{\sum_0} H^{-1}(I_a \otimes I_b \otimes J_n) \right)
\]
5 Some information matrices for variance components

5.3 The \{1\}{3-stage(2:)} structure using \( \eta \)

\[
\begin{align*}
&= \text{sesq} \left( \frac{1}{\Sigma_0} \left( (I_a \otimes I_b \otimes C_n J_n) \right) + \frac{1}{\Sigma_1} (I_b I_a \otimes C_b I_b \otimes \bar{J}_n J_n) \right. \\
&\quad + \left. \frac{1}{\Sigma_{11}} (I_a I_a \otimes \bar{J}_b I_b \otimes \bar{J}_n J_n) \right) \\
&= \text{sesq} \left( 0 + \frac{1}{\Sigma_1} (I_a \otimes C_b \otimes J_n) + \frac{1}{\Sigma_{11}} (I_a \otimes \bar{J}_b \otimes J_n) \right) \\
&= an \left( b - \frac{1}{\Sigma_2^2} + \frac{1}{\Sigma_{11}^2} \right).
\end{align*}
\]

Continuing with reference to equation (5.7), the \((2, 2)\) entries when \( i = j = 1 \) are

\[
\begin{align*}
\text{sesq} \left( Z_1^T H^{-1} Z_1 \right) &= \text{sesq} \left( (I_a \otimes I_b \otimes I_n)^T H^{-1} (I_a \otimes I_b \otimes I_n) \right) \\
&= \text{sesq} \left( \left\{ (I_a^T I_a \otimes I_b^T I_b \otimes I_n^T C_n) + \Sigma_0 \left[ \frac{1}{\Sigma_1} (I_b^T I_a \otimes I_b^T C_b \otimes I_n^T J_n) \right. \\
&\quad \left. + \frac{1}{\Sigma_{11}} (I_a^T I_a \otimes I_b^T \bar{J}_b \otimes I_n^T \bar{J}_n) \right] \right\} (I_a \otimes I_b \otimes I_n) \right) \\
&= \text{sesq} \left( \left\{ 0 + 0 + \frac{\Sigma_0}{\Sigma_{11}} (I_a \otimes I_b^T \otimes I_n^T) \right\} (I_a \otimes I_b \otimes I_n) \right) \\
&= \text{sesq} \left( \frac{\Sigma_0}{\Sigma_{11}} (I_a \otimes b \otimes n) \right) = ab^2 n^2 \Sigma_0 \Sigma_{11}.
\end{align*}
\]

Now, evaluate the case \( i = 1, j = 2 \) to obtain

\[
\begin{align*}
\text{sesq} \left( Z_1^T H^{-1} Z_2 \right) &= \text{sesq} \left( (I_a \otimes I_b \otimes I_n)^T H^{-1} (I_a \otimes I_b \otimes I_n) \right) \\
&= \text{sesq} \left( \sum_0 \frac{(I_a \otimes I_b^T \otimes I_n^T) (I_a \otimes I_b \otimes I_n)}{\Sigma_{11}} \right) \\
&= \text{sesq} \left( \sum_0 \frac{(I_a \otimes I_b^T \otimes n) \otimes (I_a \otimes I_b \otimes n)}{\Sigma_{11}} \right) = abn^2 \Sigma_0 \Sigma_{11}.
\end{align*}
\]

Lastly, with \( i = j = 2 \) the expressions are

\[
\begin{align*}
\text{sesq} \left( Z_2^T H^{-1} Z_2 \right) &= \text{sesq} \left( (I_a \otimes I_b \otimes I_n)^T H^{-1} (I_a \otimes I_b \otimes I_n) \right)
\end{align*}
\]
\[ \text{sesq}\left\{ \left( I_a^T I_a \otimes I_b^T I_b \otimes 1_n^T C_n \right) + \Sigma_0 \left[ \frac{1}{\Sigma_1} \left( I_a^T I_a \otimes I_b^T C_b \otimes 1_n^T \bar{J}_n \right) \right] \right\} (I_a \otimes I_b \otimes 1_n) \]
\[ = \text{sesq}\left\{ 0 + \Sigma_0 \left( I_a \otimes C_b \otimes 1_n^T \right) + \Sigma_0 \left( I_a \otimes \bar{J}_b \otimes 1_n^T \right) \right\} (I_a \otimes I_b \otimes 1_n) \]
\[ = \text{sesq}\left\{ \Sigma_0 \left( I_a \otimes C_b \otimes n \right) + \Sigma_0 \left( I_a \otimes \bar{J}_b \otimes n \right) \right\} = an^2 \Sigma_0^2 \left( b - 1 \Sigma_1^2 + \frac{1}{\Sigma_{11}^2} \right). \]

Using the following convenient relationships in the same way as in the last section,
\[ h_{\alpha\alpha} = abn \frac{\Sigma_0^2}{\Sigma_{11}^2} \quad h_{\beta\beta} = an \left( b - 1 \Sigma_1^2 + \frac{1}{\Sigma_{11}^2} \right). \]

Finally the information matrix takes the form
\[ M_\eta = \begin{bmatrix} N & h_{\alpha\alpha} & h_{\beta\beta} \\ h_{\alpha\alpha} & bnh_{\alpha\alpha} & nh_{\alpha\alpha} \\ h_{\beta\beta} & nh_{\alpha\alpha} & n\Sigma_0^2 h_{\beta\beta} \end{bmatrix}. \quad (5.27) \]

5.4 Summary

This chapter gives results specific to variance components and particularly those needed to construct optimum designs, namely, the information matrices and design optimality criteria.

To distinguish between design optimality for the fixed parameters and for the variance components, the criteria of $D_V$- and $A_V$-optimality are formally stated. This is done for the classical and for the pseudo-Bayesian approaches.

For the case of balanced data, MLE information matrices can be derived explicitly. This is done for the $\{1\}\{3\text{-stage}(2C)\}$ and the $\{1\}\{3\text{-stage}(2N)\}$ structures (2-factor random layouts in the variability structure, crossed and nested, respectively), and for the scenarios where $\sigma^2$ is of interest and when $\eta$ is of interest.
When the ideas of design optimality criteria and construction of optimum designs were introduced, it was made clear that in most circumstances construction algorithms are needed. These algorithms aim to find the set of support points within the design region that would yield an optimum design under a selected optimality criterion.

This chapter describes the most well known approaches for such algorithmic tasks. Their main features are to be outlined, and the conditions for their usage are also reviewed. As it happens with the different optimality criteria, a chosen algorithm may be best suited for certain experimental conditions and/or objectives, or even to deal with some constraints in the setting of the experiment. However, it seems that the most versatile options are “exchange” algorithms.

As is the case with most computationally intensive methods, the selection of a design construction algorithm should be done carefully considering the specific characteristics of the experiment. In general, a compromise between accuracy and computational efficiency has to be found.

The chapter focuses on a design construction algorithms for the variance components. The implementation is done in R. Descriptions for exchange algorithms for the fixed effects are given as an introduction to the topic. Examples for fixed
effects algorithms are not given.

6.1 Introduction

6.1.1 The purpose of an algorithm

The general objective of a design construction algorithm is to find the location of all the support points of the design within the design region such that the chosen optimality criterion is satisfied. Although a support point is actually a treatment, the customary terminology is used here.

Most construction algorithms for design optimality are based on the general class of hill-climbing mathematical optimisation routines. The use of these ideas in the context of experimental design were formalised by Fedorov (1972). The chosen criterion of optimality defines the way the optimisation is done, i.e. either up-hill or down-hill (also commonly known as mini-max and maxi-min, respectively). Silvey (1980) presents an example of Fedorov’s approach for $D$-optimality along with an important variation due to Wynn (1970). Many modifications to Fedorov’s algorithm have been made. Most established alternatives focus on $D$-optimality and fall in the category of exchange algorithms, which are well described in Atkinson et al. (2007, Chapter 12). A more specific account of alternatives for Blocked and Split-Plot experiments is given by Goos (2002).

6.1.2 Continuous designs

Recall from Section 2.3 that in general an optimum design will be so only for a specified criterion $\Psi(M(\xi))$, and it will comprise the set of support points and the design measure $\xi$, that achieves the objective.

For continuous designs, optimality theory uses the fundamental idea that functions of design matrices $\Psi(M(\cdot))$ are all convex functions, therefore, the
extremes of the convex set defining the function are contained within $X^T X$ or equivalently in $F^T F$. This in turn means that the maxima has to lie somewhere between these extremes.

Then, given a design measure $\xi$ with information matrix $M(\xi)$ and $\Psi(\cdot)$ design criterion function that is differentiable at $M(\xi)$, the objective is to find a support set $\xi'$ that maximises $\Psi(M(\xi))$ over the design region. This is achieved by the evaluation of the first derivative of $\Psi$ evaluated at $\xi$ in the direction of $\xi'$. If $\xi$ is not optimal, then the value of this derivative is greater than zero (see e.g. Theorem 3.7 of Silvey, 1980). So a new design measure has to be found that is closest to the maxima by moving in the direction of $F' F'$; this task is achieved by putting more weight on $f(x')$ than it is currently being put into it by $\xi$ by an amount $\alpha$. Let $\bar{\xi}$ be a measure that puts unit mass at the support point $f(x)$, then $\xi'$ is given by

$$\xi' = (1 - \alpha)\xi + \alpha \bar{\xi} \quad (6.1)$$

and the relevant derivative to evaluate is

$$\phi(f(x), \xi) = \lim_{\alpha \to 0^+} \frac{1}{\alpha} \left[ \Psi\left\{ (1 - \alpha)M(\xi) + \alpha M(\bar{\xi}) \right\} - \Psi\{M(\xi)\} \right]. \quad (6.2)$$

The key condition here is to make $\alpha$ sufficiently small so that the step length taken is not so big that there is risk of going over the maximum and not achieve the optimisation. For example, Fedorov’s original algorithm for $D$-optimality chooses $\alpha$ to give the maximal increase in $\Psi(\xi)$ using $\phi(f(x), \xi) = 0$ at every iteration, while Wynn’s approach uses a predetermined sequence of values for $\alpha$ for each iteration with the intuitive condition of making the step length to decrease as the size of the experiment increases, i.e. $\alpha \to 0$ as $n \to \infty$.  

151
6.1.3 Exact designs

In practice all designs are exact and the optimisation of the criterion of optimality as described above is only feasible to compute in a limited number of cases. These rarely reflect the experimental conditions faced in practice by scientists and engineers.

The theoretical constraint involved when moving from a continuous design to an exact one is that the convexity of the information matrix is not guaranteed. The tentative option of rounding off the continuous optimum design will yield nearly optimum exact designs, good enough for practical use, only in certain cases when the size of the experiment is large. In general, therefore, exact designs are “only” approximately optimum.

The idea of exchange algorithms was introduced to deal with the optimality problem for exact designs. In a standard scenario, an exchange algorithm starts with the construction of an initial $n$-point design that is non-singular. This starting design is based on a set of candidate support points spread across the entire design region. Subsequently, the initial design is improved by adding and/or removing observations so that the value of the optimality criterion $\Psi$ is improved. Evaluation of the candidate point that is best to be replaced is done by calculating the predictive variance of all the candidates and those with high values are added and the ones with low values are removed. The process continues until the specified number of observations are included or a predefined improvement tolerance is achieved. There is a risk of finding a local optimum, and this is overcome by repeating the procedure a predetermined number of times, called tries, using different starting designs. It thus seems a good approach to construct the different starting designs at random.

As described in the paragraph above, the algorithms are sequential. For $D$-optimality, the argument used to remove or add support points has an impact in
computation time and also in the relative efficiency, $D_{\text{eff}}$, of the selected design. However, there seems to be no general option that works best for most cases. Certainly, the rationale used to choose how and which points from the candidate list are to be evaluated leads to different algorithms.

Fedorov’s exchange algorithm (Fedorov, 1972) is the best known. Several techniques have been proposed in order to improve the outcome of this algorithm or to make it more efficient. Many of these are described in Atkinson et al. (2007). The KL and BLKL algorithms made available in Atkinson and Donev (1992) are a significant contribution. More extensions for blocked and Split-Plot designs are provided by Goos (2002). One particular successful approach is the coordinate-exchange algorithm (Meyer and Nachtsheim, 1995). The key difference being that it doesn’t require a set of candidate points and the exchange is not done point by point, but by exchanging different factor levels.

### 6.1.4 Update formulae

As the initial design is being constructed, the very first step involves the regularisation of the information matrix so that it does not become singular before its size is at least equal to the total number of parameters $p$. This is done by adding a small multiple of the information matrix, i.e.

$$M_\epsilon = M(\xi) + \epsilon I.$$  \hspace{1cm} (6.3)

Subsequently, the evaluation of the candidate points with highest or lowest prediction variance is done. After the corresponding exchange(s) takes place, the information of the design under construction has to be updated.

Updating and inverting $M$ for every iteration can prove to be computationally inefficient. Knowing the structure of the information matrix and by using important
results from matrix algebra, formulae to update the information matrix at every step without carrying out an actual matrix inversion has been proposed. Atkinson et al. (2007) present concise formulas that allow making the updates in the cases of adding, removing and replacement of points. It is attractive that the form of the expressions are the same for all cases and only appropriate settings of the elements according to a particular case are needed. The formulae cover the actual update of the information matrix, the update of the determinant, and also one for its inverse.

Of particular interest here is the more special cases of blocked, and mainly multi-stage experimental designs. The construction algorithms for more sophisticated structures like these, require the evaluation of exchanges at the different levels of nesting and among different blocks or both. This is to say that after certain exchange, the information matrix will be affected in a slightly different way depending on where within the design structure the exchange takes place.

Taking these facts into consideration allows for the definition of update formulae for the specific cases, and thus “complete” recalculation of $\mathbf{M}$, $|\mathbf{M}|$, and $\mathbf{M}^{-1}$ is not always necessary. The algorithms can be made much more efficient in computing time by using update formulae. To this endeavour, Arnouts and Goos (2010) present the corresponding formulae for blocked and Split-Plot designs. They cover the cases of exchanges at the split-plot level, whole-plot level and, swap of points between two whole-plots; additionally they cover the case when the size of a whole-plot is changed. For more than two stages, Jones and Goos (2009) use a Split-Split-Plot experimental design and give the appropriate update formulae.
6.2 Algorithm for variance components models

Design construction algorithms, which can also be seen as computerised search, are not widely documented in the literature for the case of variance components estimation. An extended literature review on this topic is presented in Section 1.2.

The algorithmic approach developed for this thesis using MLE and REML information matrices is described in this section. Additionally, designs are evaluated using the Mean-Dispersion model. The use of this approach for the design of experiments is not documented elsewhere.

6.2.1 Preliminaries

To construct an optimum design for the fixed effects, the standard approach is to employ an exchange algorithm. The same approach cannot be applied directly when the interest is in the variance components.

In this project, computer code to find an optimum design for variance components has been generated in the R language. Recall that the information matrix for the variance components depends on the parameters, and the problem can be seen as that of designing for nonlinear models in the fixed case. Therefore, global optimality is difficult to address in a general way, so locally optimum designs are evaluated here. However, using the pseudo-Bayesian criteria of optimality of Section 5.1.3, the procedure becomes less restricted. By using a likelihood based method, extensions to different models and design structures are relatively straightforward, so formulations that give different importance to each variance component do not need special attention.

Two different approaches have been investigated. First, the relatively conventional approach finds a design using a preliminary vector of values of the variance
components ($\sigma^2_0$) and the total number of observations ($N$); this algorithm is described in Section 6.2.2. The second approach proposed here, somewhat opposite from the first one, had previously been unseen since it does not require $\sigma^2_0$ and it can be used in two ways. If one specific design structure is provided, the algorithm finds the $\sigma^2$ vector for which the given provided design is optimum. On the other hand, if $N$ is provided, the parameter space $\sigma^2_0$ is mapped over a constructed list of plausible designs and the subset of the space, say $\sigma^2_{D}$, for which one particular design is optimum is then determined using computer searching. In other words, given a model and design structure the question to be answered is, what set of values of the variance components will be optimum under a selected criterion?

The second approach is also the option to use when the chosen design criterion of optimality is pseudo-Bayesian. This approach is described in Section 6.2.3.

Examples of the implementation of both of these algorithms will be presented later. The techniques are exemplified using the 2-factor crossed and the 2-factor nested models that were developed in Section 5.2. Initially, closed form expressions for MLE information matrices were used for the calculations. Subsequently, as presented in Section 3.3 and Section 3.5, general expressions for the methods of maximum likelihood (MLE), restricted maximum likelihood (REML), and the less explored dispersion-mean model were used.

### 6.2.2 Optimum designs requiring $\sigma^2_0$ and $N$

The implementation of an algorithm for finding an optimum design that is based on preliminary values $\sigma^2_0$ and the total number of observations $N$ can be considered the intuitive option. The comparison of competing designs is based on the values of a design optimality criterion, here $D$-optimality and $A$-optimality. In other words, the design problem is that of finding the appropriate number of levels for each main random factor in the model structure that maximises the criterion of
choice.

Before doing any computations, the statistical linear model of interest has to be defined and an expression for the information matrix $M(\sigma^2)$, either general or in closed form, has to be available. The specific definition of $M(\sigma^2)$ is directly related to the parameter estimation method that is to be used. Ideally, this would have already been chosen. In practice, statisticians report that this is not always feasible. Therefore, it seems relevant to allow flexibility for determining which design is optimal under different methods of estimation. Here MLE and REML are compared. Further, the mean-dispersion model implementation allows to find equivalent designs for different methods using the same general formulation. For example, one can find MINQUE, MLE or REML based designs.

The inputs to the algorithm are the following. The specification of the model provides the matrices $X$ and the form of $Z$, and the $r$ number of variance components to be considered. Additionally, the matrix $M(\sigma^2)$ is also available, which in turn specifies the estimation method. For design selection, these inputs do not have to be changed or updated. So, they will be referred to as implicit inputs in contrast to the explicit inputs that directly affect the output.

The explicit inputs required are the desired vector $\sigma^2_0$ of preliminary values of the variance components and the required total number of observations $N$. It is possible to specify more than one $\sigma^2_0$ vector and an optimum design for each vector will be provided. In addition, although the appropriate expression of the matrix $M(\sigma^2)$ to be used is an implicit input, the users can choose their preferred option: MLE, REML or dispersion-mean. In summary, the implicit inputs are $X$, the form of $Z$ and $M(\sigma^2)$ (via specification of an estimation method), and the explicit inputs are the preliminary values $\sigma^2_0$ and the number of observations $N$. 

157
Creation of the list of candidate designs

The first step that is executed creates the list, \( D \), of candidate designs that are going to be compared. Given \( N \), a decomposition into its divisors is performed. These divisors represent the number of levels that each main random factor in the model could have. The minimum number that is allowed is two, namely, at least two blocks of size two. Using \( s + 1 \), the number of main factors in the model plus the random error, permutations of the levels taken \( s + 1 \) at a time are found. If the product of the elements of a permutation does not equal \( N \), that permutation is eliminated. The remaining \( h \) permutations that satisfy the latter condition represent the design structures \( d_w, w = 1, \ldots, h \), that are to be compared against each other. The list of candidate designs is then \( D = \{d_1, d_2, \ldots, d_h\} \).

For example, in a 1-factor layout, the vector \((a, n)\) represents a balanced design with \( \ell_{ci} = a \), \( n \) observations per factor level and \( s = 1 \) main random regressors plus \( \epsilon \); in this case \( s + 1 = r = 2 \). For \( N = 6 \) total observations, the divisors of 6 to be permuted are 3 and 2. Then the list of candidate designs \( D \) is formed by the two possible permutations of 3 and 2, \( d_1 = (3, 2) \) and \( d_2 = (2, 3) \) since \( an = 3 \cdot 2 = N = 6 \), i.e. \( D = \{(3, 2) \ (2, 3)\} \). In fact, the final output of the algorithm is indeed the one permutation that gives the optimised criterion value.

The algorithm

Once the list of candidate designs is available, the calculations and comparisons are executed according to the following sequence:

1. For each \( d_w \) in \( D \)
   
   (a) Form matrix \( Z \) for the model using the number of levels of each factor indicated by \( d_w \).

   (b) Form the list of products \( Z_iZ_i^T \) \( i = 0, \ldots, r \).
6 Design construction algorithms

6.2 Algorithm for variance components models

(c) Using the provided $\sigma^2_0$ vector, form the variance-covariance matrix $V$.

(d) Compute $V^{-1}$.

(e) Calculate the information matrix.

(f) Evaluate the design criterion.

(g) Store the criterion value.

2. Compare all stored values from step 1g and select the design $d^*$ that gives the optimum result.

As mentioned before, the output $d^*$ is a vector of length $s + 1$ containing the permutation of the divisors of $N$ that optimised the criterion value for the given vector $\sigma^2_0$ of preliminary values of the $r + 1$ variance components, i.e. the locally optimum design.

The actions in 1d to 1g are the core steps of the algorithm. These will appear in most of the options under which the algorithm is used, as described next.

6.2.3 Optimum designs when $\sigma^2_0$ is not specified

The first step in finding an optimum design under this approach is to simulate $\sigma^2$ vectors to act as preliminary values of the variance components, then arrange them in a fine grid. Generally speaking, the chosen range for the set of values for a particular variance component reflect the degree of belief or current knowledge that the experimenter may have. This notion is very important given the local optimality characteristic of the designs.

One difference between this approach and those reported in the literature is that instead of looking for a design structure to be optimum for a specified design size, the experimenter chooses, from a list of candidates, an optimum design that is feasible to run according to the experimental conditions and budget.
The theory behind this idea imposes no limitation on the number of variance components that can be used. It is relevant though that when the number of variance components does not exceed three, simple but informative plots can be produced to summarise the results. The optimum criterion values, e.g. for $D$-optimality the maximum determinant values, are plotted against each variance component or each ratio. With 3-stage structures this happens inside a triangular simplex from which it is immediately evident which design from $D$ is optimum for any given $\sigma^2$ vector. The idea extends naturally since, in fact, areas inside the triangle can be identified where each one represents the section of the random parameter space $Z$ for which one design is optimum.

The triangle plot is a good tool to identify a suitable design for a particular $\sigma^2$ vector of interest, in addition to aid in choosing a more suitable design (e.g. cheaper) if $\sigma^2$ is on the edge of two or more regions. In addition, the relative efficiency of competing design has to be assessed. This is done by introducing robustness plots. They show a panel with two plots stacked one above the other; the plot on the top is the self efficiency of a design and the bottom plot corresponds to the relative efficiency of the other candidate designs. Examples and more details about these triangle and robustness plots are given in Section 6.2.4.

One advantage of this way of comparing designs is that it is equally applicable to balanced or unbalanced designs. The general forms of the information matrices, as presented in Chapter 3, have to be used in unbalanced cases since closed forms are not easily obtained. There is no constraint on the structures that can be used. A limitation of the triangle plots is that only three variances can be included at the same time. However, in Section 6.4 triangle plots are used when the fourth variance component is taken as the interaction of $\sigma_1^2$ and $\sigma_2^2$. 

160
Option 1: one specific design is provided

This option allows the experimenter to evaluate if a specific design of interest would be optimum in a target section of the parameter space \(Z\), i.e. in a range of values of the variance components that are expected to be applicable.

As before, \(X\), the form of \(Z\) and \(M\) are intrinsic inputs provided by the definition of the model. The explicit input is the design to be evaluated \(d_0\). This input will consist of a vector of length \(r + 1\) containing the levels for each random factor and interaction term in the model along with the number of observations per cell, as explained in Section 6.2.2.

The algorithm proceeds as follows.

1. Create a grid of \(\sigma^2\) vectors.

2. Form matrix \(Z\) for \(d_0\).

3. Form the list of products \(Z_iZ_i^T\) \(i = 0, \ldots, r\).

4. For each simulated \(\sigma^2\) vector:
   (a) Form the variance-covariance matrix \(V\).
   (b) Compute \(V^{-1}\).
   (c) Calculate the information matrix.
   (d) Evaluate the design criterion.
   (e) Store the criterion value.

5. Match the criterion values stored in 4e with the respective \(\sigma^2\) vector.

The output can be reported as a table of the \(\sigma^2\) vectors with the corresponding criterion value, or as a plot showing the portion of the parameter space \(Z\) at which the design gives locally optimum criterion values.
Option 2: N is provided

When the value of \( N \) is given, the situation is similar to the one described in Section 6.2.2. The difference is that now the list of candidate designs is evaluated at each of the simulated \( \sigma^2 \). The key implication is that most of the parameter space \( Z \) is considered, resulting in an exhaustive search for the optimum design.

The simulated values in the grid have to be chosen to give full coverage of the parameter space \( Z \). As mentioned before, when \( r = 3 \) the results can be plotted in a triangular simplex. For the simulated values to cover all the simplex, it has been suggested to use random draws from the exponential distribution rather than the uniform distribution. The main concern when using the uniform random draws is that the centre of the simplex is oversampled. Using exponential random draws, the quantity of samples in the centre of the simplex is more likely to be similar to the number of samples extending from the centre to the corners.

The simulated grid was created using the exponential distribution with \( \lambda = 3 \) as this was found to give enough extra sampling in the corners. Independent draws are made for each variance component involved, and afterwards are randomly combined to create a random \( \sigma^2_0 \) vector. The sampling algorithm was designed to create the \( \sigma^2_0 \) vectors under the restriction \( \sum_{i=0}^{r} \sigma^2_i = 1 \). In the case of variance ratios, one of the random draws is used to divide the other two draws to set the ratios randomly whilst keeping the random error value as a reference for further evaluation if required.

Once the grid is calculated, then the list of candidate designs is generated using the provided \( N \) input, and the algorithm is executed according to the following steps:

1. For each \( d_w \) in \( D \):
   
   (a) Form matrix \( Z \) for the model using \( d_w \).
(b) Form the list of products $Z_iZ_i^T$ for $i = 0, \ldots, r$.

(c) For each simulated $\sigma_0^2$ vector:
   
   i. Form the variance-covariance matrix $V$.
   
   ii. Compute $V^{-1}$.
   
   iii. Calculate the information matrix.
   
   iv. Evaluate the design criterion.
   
   v. Store the criterion value for each $d_w$ with the corresponding $\sigma_0^2$ vector.

2. for each $\sigma_0^2$ vector and using the $t$ different values from step 1(c)v, find the optimum design $d^*$.  

3. store the pair $(\sigma_0^2, d^*)$.

The resulting optimal pairs $(\sigma_0^2, d^*)$ can be reported as a table containing the values or as a plot that highlights the portion of the parameter space $Z$ at which each design $d_w$ is optimum. Examples of such plots are first presented in Section 6.3.

**Option 3: pseudo-Bayesian criteria**

The algorithmic approach reviewed so far implicitly assumes that there is no specific knowledge about the true or possible values of the variance components. This is reflected by creating the fine grid of $\sigma_0^2$ vectors. If the experimenter has some knowledge, it can be incorporated in the algorithm by using one of the pseudo-Bayesian design criteria presented in Section 5.1.3.

For example, the experimenter may know that one variance component is always larger than the rest, or it may be believed that all components are larger than a certain value, etc. Any of the versions of the algorithm explained above
can be used with a pseudo-Bayesian criterion of optimality. The difference will be captured in step “Evaluate design criterion”, where only the $\sigma^2_0$ vectors satisfying the condition of knowledge will be evaluated. Additionally, it is still possible to consider the case of total lack of knowledge where all the simulated $\sigma^2_0$ vectors are used. Finally, the criterion values for each design are compared to find the optimum design. The first example of this approach is presented in Section 6.3.2.

### 6.2.4 Computed designs for variance components

All designs presented in this section are balanced designs. The 3-stage layouts described in Section 5.2 and Section 5.3 are used. Three different total numbers of observations were used for computing and comparing optimum designs, namely $N = 16, 24$ and $32$.

For all models, there are $s = 2$ main random regressors plus the random error. For the crossed models, the cases with and without interaction are considered. Consequently, $r = s + 0 = 2$ and there are $r + 1 = 3$ total number of variance components when the interaction is not included, and $r = s + 1 = 3$ and $r + 1 = 4$ variance components for the cases with interaction. For the locally optimum designs for the 3-stage variability, the spaces

$$\sigma^2 = (\sigma^2_\epsilon, \sigma^2_1, \sigma^2_2) \quad \text{and} \quad \sigma^2 = (\sigma^2_\epsilon, \sigma^2_1, \sigma^2_2, \sigma^2_3)$$

are mapped, respectively, in a fine grid of 50,000 values for each element with the following conditions

$$\sigma^2_i = [0, 1] \quad r = \epsilon, 1, \ldots, r \quad \text{and} \quad \sum_{i=1}^{r} \sigma^2_i = 1.$$

Using the notation introduced in this chapter, an arbitrary design has the form $d_w = (a_w, b_w, n_w)$. By choosing appropriate permutations of the triplet
(a, b, n) of a levels of the first random factor, b levels of the second random factor and n observations per cell, different candidate designs can be generated. The permutations that are suitable depend on the number of variability stages and the total number of observations N. The candidate design lists generated for 3-stage variability with N = 16, 24 and 32 are shown in Table 6.1, Table 6.2 and Table 6.3, respectively.

Table 6.1 Candidate designs list generated by the algorithm for N = 16

<table>
<thead>
<tr>
<th></th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 6.2 Candidate designs list generated by the algorithm for N = 24

<table>
<thead>
<tr>
<th></th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>d7</th>
<th>d8</th>
<th>d9</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3 Candidate designs list generated by the algorithm for N = 32

<table>
<thead>
<tr>
<th></th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

6.3 Designs for the \( \{1\}\{3\text{-stage}(2C)\} \) structure to estimate \( \sigma^2 \)

The results for the \( \{1\}\{3\text{-stage}(2C)\} \) structure based on equation (5.9) are summarised below. Recall that this structure considers the overall mean in the fixed
part \(i.e. X = 1_N\). The three stages of variability are formed by a 2-factor crossed layout plus the random error.

For 3-stage variability, with \(N = 16\) there are three different candidate designs to compare. These are given in Table 6.1, where it can be seen that each design puts the highest possible replication in one stage, with the other two stages having equal replication. This is interesting because it is possible to see what effect replication has on the optimality of the designs. Unless otherwise noted, the following results were obtained by the algorithm as described in Section 6.2.3 under options two and three.

### 6.3.1 \(N=16\) and MLE

**Example 6**

The triangular plot for \(D^*_V\)-optimality is shown in Figure 6.1. The first thing to note in the triangle is the symmetry line that appears cutting the triangle in half. There are two main reasons why this happens. First, the parameter space \(Z\) is defined to be \([0, 1]\) in all dimensions. Second, the crossed nature of the factors implies that the proportion of the variability is maintained within \(Z\) for any pair.
of variance components. The main symmetry line is established for the condition \( \sigma_1^2 = \sigma_2^2 > 0.1 \). The line stops when \( \sigma_1^2 > 0.8 \). These inequalities are approximate and can be read from the plot. Alternatively, this condition is \( \sigma_1^2 / \sigma_2^2 = 1 \) and it indicates that both variances contribute equally to the total variability. Further, because in a triangular plot as one variable increases another one decreases at the same ratio, the line creates a mirror image to either side.

The top of the triangle shows that \( d_3 \) is the optimum design whenever \( \sigma_2^2 > 0.8 \). This design puts most of the experimental effort at the observational stage whose corresponding variance component is \( \sigma_1^2 \). This is as expected. To the left of the symmetry line, \( d_2 \) is optimum for \( \sigma_1^2 / \sigma_2^2 > 1 \), i.e. where \( \sigma_1^2 \) dominates and the design puts more replication in the first stage as expected. When the condition is \( \sigma_2^2 / \sigma_1^2 > 1 \) design \( d_1 \) is optimum and the same reasoning applies. To the right of the symmetry line, the contribution of each variance component remains, but the conditions are reversed, i.e. \( d_1 \) is optimum when \( \sigma_1^2 / \sigma_2^2 < 1 \) and \( d_2 \) when \( \sigma_2^2 / \sigma_1^2 < 1 \). This is the aforementioned mirror effect.

The most important question to address is how to select a design for practical use. Recall that a sigma vector is of the form \((\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2)\). For example, if the vector is \((0.3 \quad 0.4 \quad 0.3)\) design \( d_1 \) to the left of the symmetry line will be chosen. To select \( d_1 \) but on the right of the line, the conditions could be \((0.6 \quad 0.1 \quad 0.3)\). In this fashion, any \( \sigma^2 \) vector of interest to the experimenter can be located in the simplex and the corresponding optimum design is easily identified.

It is also important to assess the performance of the candidate designs across the parameter space \( Z \). To do this, robustness plots are introduced. Figure 6.2 shows three panels, each corresponds to a region of optimality as identified in Figure 6.1. For each panel, the plot on the top shows the efficiency of a particular design along its own region of optimality. It is called the self efficiency plot. The efficiency is plotted against the distance from where the design with the
most extreme criterion value (maximum for \(D\)- or minimum for \(A\)-optimality) is located within the region. The distance is calculated as the \(L^r\)-norm with \(r\) being the number of variance components. This norm has the characteristic that it represents the absolute distance, irrespective of direction or size of the region. All lines are smoothed and each one represents the lost of efficiency of the optimum design across its own region.

Although the efficiency drops in general, the given design is still optimum with respect to other designs in most of the cases. To verify this, relative efficiency plots are shown below the self efficiency plot. Figures 6.1 and 6.2 have the same colour code. This will be the case for all the plots shown unless otherwise noted.

![Figure 6.2](image)

**Figure 6.2** MLE \(D_V\)-robustness plot for \(\sigma^2\) in the \{1\}\{3-stage(2C)\} structure.

Regions 1 and 2 are associated with \(d_1\) and \(d_2\), respectively. The performance of the designs is very similar. In both regions, the self efficiency drops as low as 0.2. Where the self efficiency is the lowest, the mirror design is, at best, having the same performance as the optimum design, i.e. the relative efficiency is never greater or equal to one. That point effectively signals the location of the line of symmetry as described for Figure 6.1. The fact that the self efficiency increases on the other side of the symmetry lines indicates the mirror effect. Design \(d_3\) does not
have good performance at any point in neither of Regions 1 or 2. The robustness plot shows that along the symmetry line $d_1$ and $d_2$ have the same performance. Being able to identify this is relevant in practice, e.g. if the conditions of the experiment suggest that $\sigma_1^2 = \sigma_2^2$, then the most economical of the two designs may be preferred. In its own Region 3, $d_3$ shows good performance not dropping in self efficiency below 0.9. Neither of the other two designs is as good even when $d_3$ is at its lowest point.

Figure 6.3 MLE $A_V$- triangle plot for $\sigma^2$ in the $\{1\}{3\text{-stage}(2C)}$ structure.

Figure 6.4 MLE $A_V$- robustness plot for $\sigma^2$ in the $\{1\}{3\text{-stage}(2C)}$ structure.
The triangular plot for the $A_V$-optimality criterion is shown in Figure 6.3. As for the case of $D_V$-optimality, a symmetry line is present. Here, the condition is $\sigma_1^2 = \sigma_2^2 > 0.13$ but there is no mirror effect as in Figure 6.1 on page 166. Design $d_1$ is optimum for $\sigma_1^2 > \sigma_2^2$ whereas $d_2$ is optimum for $\sigma_2^2 > \sigma_1^2$. For $\sigma_2^2 > 0.77$, $d_3$ is optimum almost regardless of the values of $\sigma_1^2$ and $\sigma_2^2$ since the intersection with the other two regions is almost horizontal.

The $A_V$-optimality robustness plot appears in Figure 6.4. The three designs perform well across their own regions. Designs $d_1$ and $d_2$ show a minimum efficiency of about 93%, while $d_3$ does not go lower than 0.97 in efficiency. Only at the intersection of all regions, the designs seem to give the same performance. This is seen as $d_3$ has almost 100% efficiency at zero distance in all regions, as well as $d_1$ and $d_2$ approaching unity when $d_3$ is at its lowest. In more practical scenarios, no design is reasonably good outside their own region.

Looking at Table 6.1 on page 165, it is easily seen that design $d_1$ puts more weight in the first stage, where $\sigma_1^2$ is expected to dominate, $d_2$ does the same for the second stage, and $d_3$ puts the most weight at the observational stage. Therefore, the algorithm effectively chooses the design with most replication where the most variability is, e.g. $d_2$ is optimal when $\sigma_2^2$ is proportionally the largest.

Finally, comparing the results for $D_V$- and $A_V$-optimality an obvious observation is that there is no mirror effect when using $A_V$-optimality. This suggests that $A_V$-optimality is less sensitive to changes in $\sigma^2$, or alternatively, more robust against misspecification of $\sigma^2$. The performance of $D_V$-optimality across regions 1 and 2 is not good, while for $A_V$-optimality the performance is high and very stable.
6.3 Designs for the \{1\}\{3-stage(2C)\} structure to estimate \(\sigma^2\)

6.3.2 N=16, MLE and pseudo-Bayesian optimality

When the experimenter needs more flexibility in the assessment of the preliminary values \(\sigma_0^2\), it is possible to treat such information only as approximate. In these situations, the pseudo-Bayesian optimality criteria introduced in Section 5.1.3 can be used, and the provided information constitutes \(\Sigma\). The information matrix used in the algorithm in this case is given in equation (5.9).

The output of the algorithm is reported as a bar chart of the ranked order of the competing designs under the different criteria of optimality. Two important conditions are examined next.

Example 7 Total lack of knowledge for \(\sigma^2\)

This is the case where there is actually no prior knowledge about the values of the variance components. The adjustments to the algorithm needed to evaluate the pseudo-Bayesian criteria of optimality were described in Section 6.2.3.

![Figure 6.5](image)

**Figure 6.5** MLE pseudo-Bayesian \(D_V\)- and \(A_V\)-optimality plots when \(\sigma^2\) is unknown in the \{1\}\{3-stage(2C)\} structure.

For the given conditions, the algorithm creates the fine grid of simulated \(\sigma_0^2\) values as was described in Section 6.2.3. The values included in \(\Sigma\) actually span all the grid. The criterion values obtained when one specific design is
optimum, *e.g. d*₁, are identified, which gives a subset of all the criterion values over $\sigma^2_0$, and then averaged. These subset averages are rank ordered and the bar chart is created. The results for $D_V$- and $A_V$-optimality are presented in Figure 6.5 on the previous page.

The vertical axis corresponds to the criterion values. The optimum design is located in the far left of the horizontal axis. Some features of the plots can be easily observed. For example, design $d_1$ is the $D_V$-optimum since it has the maximum criterion value, however $d_2$ should be equally good. Design $d_3$ is not as good as the others. Similar remarks can be made for the $A_V$- criterion shown on the right of the figure. For this, the optimum design is the one with the lowest criterion value.

**Figure 6.6 MLE pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the {1}{3-stage(2C)} structure.**

By comparing bars of similar height, a quick evaluation of the efficiency of a design can be made. However, exact calculations of relative efficiencies are shown in Figure 6.6. It can be seen that $d_2$ and $d_1$ are equally good and yet they are markedly different. As mentioned in the last section, they give opposite weight to $\alpha$ and $\beta$. Therefore, if observations at the levels of $\beta$ are more expensive, the experimenter can choose $d_1$ on cost grounds without losing much statistical
efficiency. This applies for $D_V$- and $A_V$-optimality. Lastly, under both criteria, $d_3$ is never optimum.

**Example 8** $\sigma_1^2 > \sigma_2^2 > \sigma_\epsilon^2$

The situation where all values of the variance components are different and in addition there is an ordering is commonly encountered. An implicit assumption of the case under consideration is that the magnitude of the difference between the $\sigma^2$ values is not known. From the algorithmic point of view, in this case $\Sigma$ is formed by subsetting the grid of simulated values according to the condition, i.e. $\sigma_\alpha^2 > \sigma_\beta^2 > \sigma_\epsilon^2$. If more information is supplied, it will be treated as an additional restriction and the resulting subset will be smaller. After this critical step, the results are reported as before.

![Figure 6.7](image)

**Figure 6.7** MLE pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_\epsilon^2$ in the $\{1\}$\{3-stage(2C)\} structure.

The corresponding plots are shown in Figure 6.7. Design $d_1$ is $A_V$-optimum, followed by $d_2$ and $d_3$ respectively. As before, it is clear that the optimum design is the one putting more resources at the variability stage with larger variance. This fully reflects the conditions being evaluated under the pseudo-Bayesian criterion.

Surprisingly, under $D_V$-optimality design $d_2$ was given preference. Although $d_1$
6.3 Designs for the \{1\}\{3-stage(2C)\} structure to estimate $\sigma^2$

Figure 6.8  MLE pseudo-Bayesian $\mathcal{D}_V$- and $\mathcal{A}_V$-optimality efficiency plots when $\sigma^2$ is unknown in the \{1\}\{3-stage(2C)\} structure.

seems almost equally good. More properly, this can be assessed with the relative efficiency plots shown in Figure 6.8. After inspection of the plot, for $\mathcal{D}_V$-optimality, it is seen that $d_1$ is no more than 1% less efficient than $d_2$, whereas $d_3$ is always at least 2% lower. Again, the experimenter can choose a design based on cost considerations and still have good statistical efficiency. Under $\mathcal{A}_V$-optimality $d_1$ is about 15% more efficient than $d_2$ and $d_3$ gets to be as low as 20% less efficient than the optimum design.

6.3.3 N=16 and REML

The results of the algorithm when the REML information matrix is used are shown next. Recall that for all REML cases in this section $X = 1_N$.

Example 9

For $\mathcal{D}_V$-optimality the triangle and robustness plots appear in Figure 6.9 and Figure 6.10, respectively. As seen in previous examples, there is a symmetry line in the triangle plot. In this case, it reflects the condition $\sigma_1^2 = \sigma_2^2 > 0.07$. The region established by $d_3$ starts from $\sigma_e^2 \geq 0.85$ which means it is smaller than
6.3 Designs for the $\{1\}\{3\text{-stage}(2C)\}$ structure to estimate $\sigma^2$

Figure 6.9 REML $D_V$-triangle plot for $\sigma^2$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure. Before. However, it also shows more sensitivity to the values of $\sigma_1^2$ and $\sigma_2^2$ given its diamond shape.

Figure 6.10 REML $D_V$-robustness plot for $\sigma^2$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure.

The robustness plot shows that $d_1$ and $d_2$ have almost the same performance in their own regions. Both drop to as low as 0.2 in self efficiency, but each one is still the best design across their respective regions since the relative efficiency of the other designs is never close to one. They do not even look equally good along the symmetry line. In region 3 $d_3$ is never worse than 89% self efficient, outperforming $d_1$ and $d_2$, as expected.

175
6 Design construction algorithms

6.3 Designs for the \{1\}\{3-stage(2C)\} structure to estimate \(\sigma^2\)

For \(A_V\)-optimality the triangle and robustness plots appear in Figure 6.11 and Figure 6.12, respectively. The triangle plot is very similar to that for \(D_V\)-optimality. The difference is that region 3 is slightly larger. It starts at \(\sigma^2 = 0.79\) instead of 0.85. All other remarks apply equally to this plot. The robustness plot shows that the three designs perform with high efficiency in their own regions. For regions 1 and 2, the corresponding designs have a lowest self efficiency of about 0.93, and the competing designs are never equally good according to their relative efficiency. Design \(d_3\) is 0.98% self efficient and the other designs are only equally
good at the intersection point of the regions.

As seen in previous examples, the self efficiency under $D_V$-optimality across regions 1 and 2 is not good, whilst for $A_V$-optimality is high and very stable. The similarity of the triangle plots for the two criteria is remarkable, but the robustness plots show that one criterion has better performance in the parameter space $Z$.

**MLE vs REML**

Comparing the triangle plot in Figure 6.9 for $D_V$-optimality using REML and in Figure 6.1 using MLE, it is evident that the mirror effect does not appear when using REML estimation. This suggests that $D_V$-optimality under MLE is the most sensitive to changes in $\sigma^2$. It must be noted that the self efficiency of $D_V$-optimality for both methods is highly comparable. It has been more instructive to note the difference between $D_V$- and $A_V$-optimality, and being able to discriminate in favour of the latter based on the robustness plots.

**6.3.4 $N=16$, REML and pseudo-Bayesian optimality**

**Example 10** Total lack of knowledge for $\sigma^2$

The results for $D_V$- and $A_V$-optimality are presented in Figure 6.13 on the next page, and the corresponding relative efficiency plots are shown in Figure 6.14.

For $D_V$-optimality $d_1$ is optimum followed by $d_2$. The relative efficiency plot show that these two designs are almost identical. The situation is no different under $A_V$-optimality in spite of $d_2$ appearing as optimum. Design $d_3$ is always far from optimum. With these conditions, $d_1$ or $d_2$ can be chosen based on cost or feasibility with no efficiency loss.
6 Design construction algorithms

6.3 Designs for the \{1\}\{3-stage(2C)\} structure to estimate \(\sigma^2\)

Figure 6.13 REML pseudo-Bayesian \(D_V\)- and \(A_V\)-optimality plots when \(\sigma^2\) is unknown in the \{1\}\{3-stage(2C)\} structure.

Figure 6.14 REML pseudo-Bayesian \(D_V\)- and \(A_V\)-optimality efficiency plots when \(\sigma^2\) is unknown in the \{1\}\{3-stage(2C)\} structure.

Example 11 \(\sigma_1^2 > \sigma_2^2 > \sigma_\epsilon^2\)

The results for \(D_V\)- and \(A_V\)-optimality are presented in Figure 6.15, and the corresponding relative efficiency plots are shown in Figure 6.16.

The bar plot for \(D_V\)-optimality is very similar to the total lack of knowledge case, where it is not readily obvious how much better \(d_1\) is with respect to \(d_2\). In the present case, the relative efficiency plot clearly shows that \(d_2\) is very close to one losing only 0.01% of efficiency. Thus, similar conclusions as in the previous case can be drawn here. For the \(A_V\)-optimality criterion, \(d_1\) is optimum and the
6 Design construction algorithms

6.3 Designs for the \{1\}{3-stage(2C)} structure to estimate $\sigma^2$

Figure 6.15 REML pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the \{1\}{3-stage(2C)} structure.

Figure 6.16 REML pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the \{1\}{3-stage(2C)} structure.

relative efficiency of $d_2$ and $d_3$ is only 0.79 and 0.77, respectively.

MLE vs REML

For the $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ case, MLE $D_V$-optimality indicates $d_2$ is optimum. Although $d_1$ is the design that puts more weight to the first stage where $\sigma_1^2$ is largest. Under REML this does not happen. However, the relative efficiency plots show that for both pseudo-Bayesian cases, $d_1$ and $d_2$ are almost equally $D_V$-efficient. Conclusions drawn for all cases are the same. As before, the main differences come from the
optimality criteria rather than the estimation method.

6.3.5 N=24 and MLE vs REML

When the total number of observations is $N = 24$, there can be nine candidate designs under the variability structure being considered. The list of candidates is shown in Table 6.2 on page 165. The experimenter can choose an optimum design to use in practice from the triangular and robustness plots in the same way as was described in Section 6.3.1.

Example 12 $D_V$-optimality

The triangular plot for MLE $D_V$-optimality is shown in Figure 6.17 followed by the robustness plot in Figure 6.18 on page 182.

The first thing to note in the triangle is the symmetry line that appears cutting it in half. This was first seen in Section 6.3.1 and explained therein. Here, it can be considered that the line is clearly defined below the region were $d_9$ is optimum. This will happen for the condition $\sigma_1^2 = \sigma_2^2 > 0.04$, which implies $\sigma_1^2 < 0.92$. This is the design that puts most of the experimental effort at the observational stage whose corresponding variance component is $\sigma_2^2$. The alternative condition for where the line appears is $\sigma_1^2/\sigma_2^2 = 1$ as before. Further, the mirror effect described for the MLE $D_V$-triangle in Section 6.3.1, is also present here. This is no surprise. The mirror effect is vertical, but interestingly it only applies for $d_1$ to $d_4$. For the other designs there is a complementary rather than mirror effect.

The top area of the triangle, where $\sigma_1^2 \geq 0.8$, shows the complementary pattern. To the right of the symmetry line, designs $d_5$, $d_7$ and $d_9$ appear as optimum, and to the left of the line designs $d_6$, $d_8$ and $d_9$ are optimum. By pairing these designs with respect to the line of symmetry as $(d_5, d_6)$, $(d_7, d_9)$ and $(d_9, d_9)$, it is easy to see the complementary effect. For design $d_9$, $a = b = 2$ is obviously symmetrical.
on both random factors. For the other two pairs, the values for $a$ and $b$ are reversed. The other designs across the triangle can also be paired with respect to the symmetry line. The pairs in this case are $(d_1, d_4)$ and $(d_2, d_3)$. As expected these pairs have the values of $a$ and $b$ reversed. This ordering can be seen in the area defined by $0 < \sigma^2_e \leq 0.47$. Between $0.47 < \sigma^2_e \leq 0.85$ the ordering of each pair is reversed giving the mirror effect. For every pair, the value of $n$ is the same.

A more interesting pattern is found by realising that as $\sigma^2_e$ increases while $\sigma^2_{\alpha}$ and $\sigma^2_{\beta}$ decreases, the optimum design is more closely related to a fixed blocks design. This is reflected in that the optimum design progressively puts more experimental effort in pure replication by increasing the value of $n$, thus giving less importance to the variability structure. This is the reason why design $d_9$ is optimum at the very corner, since it is the design with the largest possible $n$ in the candidate list.

The robustness plot is shown in Figure 6.18. The self efficiency plots for Region 1 to Region 4 are U shaped. This is due to the mirror effect. However it is not exactly symmetric and the side with the lowest self efficiency corresponds to the smallest area shown in the corresponding triangle plot.

Regions 1 and 4 are paired as described above. Their respective self efficiency
Design construction algorithms

6.3 Designs for the \{1\}\{3-stage(2C)\}

structure to estimate $\sigma^2$

curves follow the same pattern. However, $d_4$ falls down to 0.2 self efficiency whilst $d_1$ gets no lower than 0.3. The same can be observed for the ($d_2$, $d_3$) pair. In Region 2 and Region 3 the complementary design is almost as good as the optimum design across the region based on the relative efficiency. In Region 1 $d_2$ is the closest design and not $d_4$ which is the complementary. The same happens in Region 4 with $d_3$. Also note that in Region 1 and Region 4, the complementary design is rather poor. Neither design $d_5$ to $d_9$ are competitive in these four regions.

For Region 5 and Region 6, designs $d_5$ and $d_6$ are highly self efficient as their

Figure 6.18 MLE $D^*_V$-robustness plot for $\sigma^2$ in the \{1\}\{3-stage(2C)\} structure.
self efficiency drops no lower than 0.91 and 0.95, respectively. Note that the designs in the contiguous regions in the increasing direction of $\sigma^2_\epsilon$ (with respect to their location in the triangular plot) are almost as good as the optimum design across the corresponding region in the relative efficiency sense, and not the complementary designs which are actually very poor. The same is observed for Region 7 and Region 8 where the contiguous design is $d_9$, but additionally the contiguous designs in the decreasing direction of $\sigma^2_\epsilon$ are second best. Also, the self efficiency of these two regions gets as low as 0.85. In Region 9, no design is close in efficiency to the optimum.

![Figure 6.19](image)

**Figure 6.19** REML $D_V$-triangle plot for $\sigma^2$ in the \{1\}{3-stage(2C)} structure.

The plots for REML $D_V$-optimality appear in Figure 6.19 and in Figure 6.20. An obvious difference in the MLE and REML triangle plots is the absent mirror effect in the latter. Also, Region 2 and Region 3 extend to higher values of $\sigma^2_\epsilon$, thus reducing the area of the other regions. Region 9 shows a more diamond like shape than triangular as in the MLE case, which means that $d_9$ is more sensitive to changes in the values of $\sigma^2_1$ and $\sigma^2_2$. This was also seen in Section 6.3.1. The rest of the plot can be interpreted as before.

More insight is obtained from the robustness plot. As expected due to the lack of mirror effect, the self efficiency plots are not U shaped. It was noted in the MLE case, that designs $d_2$ and $d_3$ show virtually the same self and relative
efficiency performance. These designs also take over $d_4$ and $d_1$ respectively for $\sigma^2 < 0.15$. It seems that in the case of REML estimation, the design optimality is not as sensitive to changes in certain values of the variance components.

The self efficiency of the optimum designs in Region 2 and Region 4 is 10% higher than under MLE. In Region 3 the opposite is true. Designs $d_5$ to $d_8$ are 1% to 7% more self efficient under REML. Region 1 and Region 9 show no difference. As it was identified in the MLE case, the relative efficiency is high for the designs in any given neighbouring region. In neither case, the optimum design
is outperformed in its own region.

The main difference between the optimum designs for MLE and REML estimation is the mirror effect. The self efficiency in Region 1 to Region 4 is kept or slightly improved when using REML. There is improvement in the other regions, but marginal. The self efficiency is also more uniform in all regions under REML. However, the REML version decreases the area of optimality of Region 5 to Region 9. This suggests that MLE may be preferred if the range of $\sigma^2$ is within these regions, and REML for Region 1 to Region 4.

**Example 13 $A_V$-optimality**

For the MLE $A_V$-optimality criterion, the triangular plot is shown in Figure 6.21

![MLE $A_V$- triangle plot for $\sigma^2$ in the $\{1\}$-{3-stage($2C$)} structure.](image)

The symmetry line in the triangular plot also appears here as is expected due to the crossed variability structure. The line is clearly defined below the region where $d_9$ is optimum, Region 9. This happens for the condition $\sigma_1^2 = \sigma_2^2 > 0.04$, which implies $\sigma_2^2 \leq 0.87$. Recall that the symmetry line also reflect the condition $\sigma_1^2/\sigma_2^2 = 1$. The complementary effect can also be seen.

For MLE $D_V$-optimality, $d_2$ and $d_3$ cover most of the area of the triangle. In the present case, $d_1$ and $d_4$ cover most of the area. This reveals the nature of
the design criterion. $A_V$-optimality minimises the the average variance and the complementary designs $d_1$ and $d_4$ put most of the experimental effort in the first and second stages, respectively. This implies that these designs have the largest variance possible for the corresponding stage. Therefore, $A_V$-optimality gives more coverage for these regions.

The robustness plot shows that in all the regions, the optimum design is highly self efficient. Designs $d_1$ and $d_4$ are no lower than 0.95 self efficient, whilst $d_2$ and $d_3$ are no lower than 0.92 self efficient. Designs $d_5$ to $d_9$ almost show no drop in
self efficiency across their regions where the worst scenario is a maximum 1% loss.

For Region 2 and Region 3, \((d_1, d_3)\) and \((d_2, d_4)\) respectively, are almost equally good as the optimal design in relative efficiency in all the region. Note that only the neighbour designs have such good performance.

Consider the case of REML estimation next. For the REML \(A_V\)-optimality criterion, the triangular plot is shown in Figure 6.23 and the robustness plot in Figure 6.24.

The REML triangle plot is very similar to the MLE version. In the REML case, the cross section along the symmetry line is larger extending until \(\sigma^2_1 = 0.9\). Also, as seen before, Region 9 becomes diamond shaped which makes it slightly more sensitive to changes in \(\sigma^2_1\) and \(\sigma^2_2\). Further, the area coverage of \(d_1\) and \(d_4\), or equivalently \((d_2, d_3)\), is larger. For example, under MLE Region 1 spans the area outlined by \(\sigma^2_1 > 0.37\) and \(\sigma^2_2 < 0.28\), but for REML this is \(\sigma^2_1 > 0.2\) and \(\sigma^2_2 < 0.2\).

In terms of robustness, Figure 6.24 shows that self efficiencies are virtually the same as in the MLE case. The ordering of the designs, according to the relative efficiency in each region, is also the same as for MLE, although the relative efficiency is, in average, 0.5 lower. Further, although the relative efficiency curves do not have the same shape as in the MLE case, the conclusions are the same here.

\[\text{Figure 6.23 REML } A_V\text{- triangle plot for } \sigma^2 \text{ in the } \{1\}\{3\text{-stage}(2C)\} \text{ structure.}\]
When comparing the MLE and REML versions of $\mathcal{A}_V$-optimality, the differences are not striking. Therefore, it is reasonable to consider that MLE and REML are almost everywhere $\mathcal{A}_V$-equivalent.
6.3.6 N=24, MLE vs REML and pseudo-Bayesian optimality

The results when considering the pseudo-Bayesian optimality criteria are presented below.

Example 14 Total lack of knowledge for $\sigma^2$

The bar plots for MLE $D_V$- and $A_V$-optimality are presented in Figure 6.25 and the corresponding relative efficiency plots in Figure 6.26. For $D_V$-optimality $d_3$ appears as the optimum design. However, the criterion value for $d_2$ is almost identical. The $D_V$-efficiency plot confirms that they are equally good. Although no longer optimum, the same phenomenon can be seen in the pairs $(d_4, d_1)$, $(d_5, d_6)$ and $(d_7, d_8)$. Design $d_9$ is always far from optimum, although the lowest relative efficiency is 0.96.

In the case of $A_V$-optimality the same general behaviour can be observed. Design $d_3$ is optimum but $d_2$ is equally good in relative efficiency. The conclusions for the pair $(d_4, d_1)$ also remain, as well as those for the other pairs, although on these the order is reversed. The lowest relative efficiency for $d_9$ in this case is
6 Design construction algorithms

6.3 Designs for the \{1\}(3-stage(2C)) structure to estimate \(\sigma^2\)

Figure 6.26 MLE pseudo-Bayesian \(D\) and \(A\)-optimality efficiency plots when \(\sigma^2\) is unknown in the \{1\}(3-stage(2C)) structure.

0.85, which is lower. Similarly, the maximum relative efficiency in the plot is also higher, but only marginally.

The bar plots for REML \(D\) and \(A\)-optimality are presented in Figure 6.27 and the corresponding relative efficiency plots in Figure 6.28. As it can be seen from the two figures, apart from the slightly different scale on the optimality criteria, there is no difference between the MLE and REML methods.

Figure 6.27 REML pseudo-Bayesian \(D\)- and \(A\)-optimality plots when \(\sigma^2\) is unknown in the \{1\}(3-stage(2C)) structure.
6.3 Designs for the \( \{1\}\{3\text{-stage}(2C)\} \) structure to estimate \( \sigma^2 \)

Example 15 \( \sigma_1^2 > \sigma_2^2 > \sigma^2 \)

The situation where all values of the variance components are different and in addition there is an ordering is reviewed next. The corresponding bar plots for MLE \( D_V \)- and \( A_V \)-optimality can be seen in Figure 6.29 and the efficiency plots in Figure 6.30.

Design \( d_4 \) is \( D_V \)-optimum with \( d_3 \) showing equally good performance according to the relative efficiency. The next two designs with high relative efficiency are \( d_2 \) and \( d_1 \). The ordering of these four designs is interesting. The more intuitive choice
would be for the design with the highest weight in $\sigma^2_1$ to be optimum. Nonetheless, $D_V$-optimality chooses $d_4$ which puts more emphasis on $\sigma^2_2$. In different random runs of the algorithm (not shown for brevity) $d_3$ appeared as optimum most of the time followed by $d_2$. $D_V$-optimality therefore seems to be highly sensitive to small changes in $\sigma^2$. However, the maximum difference in relative efficiency among $d_1$ to $d_4$ is less than 0.01, which suggests that actually the MLE $D_V$- criterion is not sensitive to small to medium differences in $\sigma^2$.

Contrastingly, with MLE $A_V$-optimality the ordering of the designs is completely intuitive. The design with the largest experimental effort in the stage associated with $\sigma^2_1$ is the optimum, this one being $d_1$, and this happens progressively over the candidate designs. Moreover, the maximum values in relative efficiency are more pronounced than in the $D_V$- case, with a maximum difference of 0.4.

The REML $D_V$- and $A_V$-optimality bar plots are presented in Figure 6.31 and the efficiency plots in Figure 6.32.

When using REML, the $D_V$- criterion gives more intuitive results than in the MLE case. Design $d_2$ is optimum and $d_3$ appears almost equally good in terms of

**Figure 6.30** MLE pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2_1 > \sigma^2_2 > \sigma^2_3$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure.
6 Design construction algorithms

6.3 Designs for the \( \{1\}\{3\text{-stage}(2C)\}\) structure to estimate \( \sigma^2 \)

Figure 6.31 REML pseudo-Bayesian \( D_V \)- and \( A_V \)-optimality plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}\{3\text{-stage}(2C)\}\) structure.

Figure 6.32 REML pseudo-Bayesian \( D_V \)- and \( A_V \)-optimality efficiency plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}\{3\text{-stage}(2C)\}\) structure.

Relative efficiency and \( d_1 \) follows. The ordering of the candidate designs reveals that in these case \( D_V \)-optimality tends to balance the variability stages of the experiment.

As in the previous example, the differences among MLE and REML \( A_V \)-optimality reside in the scale of the criterion value. Despite that, the conclusions drawn for the MLE case do not change for REML.

\( D_V \)-optimality when using MLE gave counter intuitive results. On the other hand, \( A_V \)-optimality is consistent under the two methods of estimation. The
results are also intuitive in both scenarios of amount of knowledge about \( \sigma^2 \).

### 6.3.7 \( N=32 \) and MLE vs REML

When the total number of observations is \( N = 32 \), the list of candidate designs has six members as shown in Table 6.3 on page 165. The results for this case are presented next.

**Example 16 \( D_V \)-optimality**

The triangular plot for MLE \( D_V \)-optimality is shown in Figure 6.33 followed by the robustness plot in Figure 6.34.

![Figure 6.33](image)

**Figure 6.33** MLE \( D_V \)-triangle plot for \( \sigma^2 \) in the \( \{1\} \{3\text{-stage(2C)}\} \) structure.

The triangle plot exhibits the symmetry line as in previous examples, and also the mirror effect seen in Example 12. Moreover, the shape of the regions are in general the same, with the obvious difference that only six designs are present here. For example, Region 1 and Region 3 have almost exactly the same shape and extension as Region 1 and Region 4, respectively, in Example 12. Comparing Table 6.2 and Table 6.3 on page 165, it is clear that the aforementioned regions correspond to the designs that put more weight on the first variability stage and the second stage, respectively. For \( N = 32 \), \((d_4, d_5)\) alternate the weight between the stages and \( d_6 \) puts most of the weight at the observational stage.
With this in mind, it is reasonable to argue that the MLE $D_V$-optimality results for $N = 24$ and $N = 32$ are similar. That is, the regions with largest area correspond to the designs that spread the weight of the design evenly across the variability stages. Designs $d_1$ and $d_3$ are complementary putting most of the weight in one variability stage. Same reasoning applies to the other designs.

The symmetry line extends up to the condition $\sigma^2 \leq 0.87$. The region where $d_6$ is optimum represents $\sigma^2 \geq 0.94$. Similarly as in Example 12, the mirror effect applies only to $d_1$ to $d_3$. For the other designs the effect is complementary.

In terms of self and relative efficiency, in general, the same observations can be made. The self efficiency plots for $(d_1, d_2, d_3)$ are U shaped and they go as low as 0.3. Naturally, the optimum designs is never outperformed in its region, and the neighbour and complementary designs show the highest relative efficiency.
6 Design construction algorithms

6.3 Designs for the \{1\}\{3-stage(2C)\} structure to estimate $$\sigma^2$$

in all regions.

The plots for REML $$D_V$$-optimality appear next where Figure 6.35 is the triangle plot and Figure 6.36 is the robustness plot.

Once more the observations made in Example 12 stand here. Namely, the area of the middle regions of optimality are larger. Consequently, the other regions are narrower and $$d_6$$, which puts most of the weight at the observational stage, becomes diamond shaped. Also, the mirror effect is not present, only the complementary effect.

In the robustness plot, only Region 2 shows a U shaped self efficiency curve. This is expected. Here, it depicts the symmetry along the middle of the triangle as well as the lack of the mirror effect. The self efficiency of the smaller regions, i.e. Region 5 and Region 6, are in average 0.03 lower than for MLE. As before, the neighbour and complementary designs show the highest relative efficiency, and the performance is very similar to the MLE version.

![Image](image.png)

**Figure 6.35** REML $$D_V$$-triangle plot for $$\sigma^2$$ in the \{1\}\{3-stage(2C)\} structure.

The relative efficiency performance of REML is in average lower than for MLE. The implication is that REML is more sensitive to changes in the values of $$\sigma^2$$.  

196
6 Design construction algorithms

6.3 Designs for the \( \{1\}\{3\text{-stage}(2C)\} \) structure to estimate \( \sigma^2 \)

Example 17 \( \mathcal{A}_V \)-optimality

For the MLE \( \mathcal{A}_V \)-optimality criterion, the triangular plot is shown in Figure 6.37 and the robustness plot in Figure 6.38 on the following page.

The symmetry line in the triangular plot also appears here as is expected due to the crossed variability structure. The line is clearly defined below the region where \( d_6 \) is optimum, Region 6. This happens for the condition \( \sigma_1^2 = \sigma_2^2 > 0.05 \), which implies \( \sigma_1^2 \leq 0.9 \). Recall that the symmetry line also reflect the condition \( \sigma_1^2 / \sigma_2^2 = 1 \). The complementary effect can also be seen. As observed in Example 13, compared to \( \mathcal{D}_V \)-optimality, the preference is not given to the most uniform design but to the designs that put most of the experimental effort in the first and second stages, respectively.

The robustness plot shows that in all the regions, the optimum design is highly

Figure 6.36 REML \( \mathcal{D}_V \)-robustness plot for \( \sigma^2 \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.
6 Design construction algorithms

6.3 Designs for the \{1\}{3-stage(2C)} structure to estimate \(\sigma^2\)

self efficient. Designs \(d_1\) to \(d_3\) are no lower than 0.91 self efficient. Designs \(d_4\) to \(d_6\) show almost no drop in self efficiency across their regions where the worst scenario is a maximum 3% loss. In terms of relative efficiency, no design appears

--

Figure 6.37  MLE \(A_V\)- triangle plot for \(\sigma^2\) in the \{1\}{3-stage(2C)} structure.

Figure 6.38  MLE \(A_V\)- robustness plot for \(\sigma^2\) in the \{1\}{3-stage(2C)} structure.
to be equally good as the optimum design in neither region.

Consider the REML case next. For the REML $A_V$-optimality criterion, the triangular plot is shown in Figure 6.39 and the robustness plot in Figure 6.40. The REML triangle plot is very similar to the MLE version. All remarks made in Example 13 apply to the present case. In terms of robustness, Figure 6.24 shows that self efficiencies are slightly lower than in the MLE case. The ordering of the designs, according to the relative efficiency in each region, is virtually the same as for MLE. In Region 1 and Region 3, the relative efficiency is as low as 0.5, but the conclusions are still the same.

When comparing the MLE and REML versions of $A_V$-optimality, the differences are not striking. Therefore, it is reasonable to consider that the designs for MLE and REML estimation are almost everywhere $A_V$-equivalent. The self efficiency is almost the same for both methods of estimation.

### 6.3.8 N=32, MLE vs REML and pseudo-Bayesian optimality

The results when considering the pseudo-Bayesian optimality criteria are presented below.
6.3 Designs for the \{1\}3-stage(2C) structure to estimate $\sigma^2$

Figure 6.40 REML $A_V$-robustness plot for $\sigma^2$ in the \{1\}3-stage(2C) structure.

Example 18 Total lack of knowledge for $\sigma^2$

Figure 6.41 MLE pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma^2$ is unknown in the \{1\}3-stage(2C) structure.
6 Design construction algorithms

6.3 Designs for the \( \{1\text{-stage}(2C)\} \) structure to estimate \( \sigma^2 \)

Figure 6.42 MLE pseudo-Bayesian \( D_V \) - and \( A_V \)-optimality efficiency plots when \( \sigma^2 \) is unknown in the \( \{1\text{-stage}(2C)\} \) structure.

The bar plots for MLE \( D_V \)- and \( A_V \)-optimality are presented in Figure 6.41 and the corresponding relative efficiency plots in Figure 6.42.

For \( D_V \)-optimality, \( d_2 \) is the optimum design. The criterion value for \( d_1 \) and \( d_3 \) is almost identical, but undoubtedly lower than that of \( d_2 \). The same can be seen in the pairs \((d_4, d_5)\). The \( D_V \)-efficiency plot confirms this. Design \( d_6 \) is always far from optimum, although the lowest relative efficiency is 0.96.

In the case of \( A_V \)-optimality, design \( d_2 \) is optimum and the relative efficiency of the competing designs is noticeable lower. The order of the pairs \((d_3, d_1)\) and \((d_5, d_4)\) is reversed with respect to \( D_V \)-optimality. The lowest relative efficiency for \( d_9 \) in this case is 0.79, which is significantly lower than in the \( D_V \)-case. However, the conclusions do not change.

The bar plots for REML \( D_V \)- and \( A_V \)-optimality are presented in Figure 6.43 and the corresponding relative efficiency plots in Figure 6.44. As it can be seen from the two figures, apart from the slightly different scale on the optimality criteria, there is no difference between the optimum designs for MLE and REML estimation.

\[ \]
6 Design construction algorithms

6.3 Designs for the \{1\}{3-stage(2C)} structure to estimate $\sigma^2$

**Figure 6.43** REML pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma^2$ is unknown in the \{1\}{3-stage(2C)} structure.

**Figure 6.44** REML pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the \{1\}{3-stage(2C)} structure.

**Example 19** $\sigma_1^2 > \sigma_2^2 > \sigma_\varepsilon^2$

The corresponding bar plots for MLE $D_V$- and $A_V$-optimality can be seen in Figure 6.45 and the efficiency plots in Figure 6.46.

Similarly as in Example 25, pseudo-Bayesian MLE $D_V$-optimality for the present case, does not give the expected results. Design $d_2$ is reported as optimum. This design gives uniform weight to the first two variability stages. The second best design, $d_3$, puts the maximum effort in the second stage, which is not the intuitive result. In general, this criterion does not choose the design that puts
more resources in the variability stage with more uncertainty. The criterion does not seem to capture the restriction imposed.

On the other hand, $A_V$-optimality results in a design ordering that is much more intuitive. Here $d_1$ is chosen as optimum as expected. Moreover, the relative efficiency of the competing designs is progressively smaller ranging from 0.92 to 0.7.

---

**Figure 6.45** MLE pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_\epsilon^2$ in the $\{1\}{3\text{-stage}(2C)}$ structure.

**Figure 6.46** MLE pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma_1^2 > \sigma_2^2 > \sigma_\epsilon^2$ in the $\{1\}{3\text{-stage}(2C)}$ structure.

The REML $D_V$- and $A_V$-optimality bar plots are presented in Figure 6.47 and the efficiency plots in Figure 6.48 on the next page. In this case, using REML
Design construction algorithms

6.3 Designs for the \{1\{3-stage(2C)\}\} structure to estimate $\sigma^2$

Figure 6.47 REML pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the \{1\{3-stage(2C)\}\} structure.

Figure 6.48 REML pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the \{1\{3-stage(2C)\}\} structure.

does not significantly improve the performance of the $D_V$- criterion. It still chooses $d_2$ as optimum and gives $d_3$ and $d_1$ a very similar relative efficiency. The same conclusion as in Example 25 is reached here.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}{3\text{-stage}(2Ci)}$ structure (interaction model)

The results for the $\{1\}{3\text{-stage}(2Ci)}$ structure which is an interaction model based on equation (5.9) are summarised below. Recall that this structure considers the overall mean in the fixed part i.e. $X = 1_N$. The three stages of variability are formed by a 2-factor crossed layout and the model contains the first order interaction of the two random factors in addition to the random error.

The total number of variance components in this model is $r = 4$. This means that the output of the algorithm cannot be plotted directly in a triangular simplex. However, a representation using triangle plots can still be constructed. In order to allow for comparison with the cases in Section 6.3, triangular plots with $\sigma_1^2$, $\sigma_2^2$ and $\sigma_3^2$ were drawn for a subset of values of the interaction variance component $\sigma_3^2$. The range of values of $\sigma_3^2$ is indicated by the horizontal line underneath each triangle. In this way, the effect that the interaction has on the regions of optimality can be assessed.

6.4.1 N=16 and MLE vs REML

Example 20

The triangular plot for MLE $D_V$-optimality is shown in Figure 6.49 and the robustness plot is shown in Figure 6.50 on the following page.

As it was observed in the case without interaction, the triangle plot shows the symmetry line along the middle of the simplex. When the interaction effect is small, it is expected that the triangle plot resembles the case without interaction. This can indeed be seen in the left most triangle where $0 < \sigma_3^2 \leq 0.25$. However, although the main optimality regions can be identified in shape and size, the optimum design for a noticeable number of $\sigma_0^2$ is spread in different regions.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}{\text{3-stage(2Ci)}}$ structure

As the value of $\sigma^2_3$ increases, there are fewer combinations of $\sigma^2_0$ remaining. This is reflected in the triangle plot in that $d_3$ does not appear in the next subset of values of $\sigma^2_3$. A hint of the mirror effect can still be seen when $0.25 < \sigma^2_3 \leq 0.5$. In this condition, only the complementary designs $d_1$ and $d_2$ have a region of optimality. The mirror effect completely disappears in the range $0.5 < \sigma^2_3 < 1$ which is shown in the two right most triangles.

![Figure 6.49](image1.png)  
**Figure 6.49** MLE $\mathcal{D}_V$- triangle plots for $\sigma^2$ in the $\{1\}{\text{3-stage(2Ci)}}$ structure with interaction.

![Figure 6.50](image2.png)  
**Figure 6.50** MLE $\mathcal{D}_V$- robustness plot for $\sigma^2$ in the $\{1\}{\text{3-stage(2Ci)}}$ structure with interaction.

The robustness plots show that the self efficiency of $d_1$ and $d_2$ is very low. Design $d_3$ does reasonably good in comparison with a lowest self efficiency of 0.65.
6 Design construction algorithms

6.4 Designs to estimate $\sigma^2$ for the $\{1\}$-stage(2Ci) structure

The relative efficiency of $d_2$ in Region 3 is very high. This is positive given the unreliable behaviour of the designs, since for large values of $\sigma^2$, $d_2$ and $d_3$ could be used interchangeably.

The triangular plot for the MLE $A_V$-optimality criterion is shown in Figure 6.51 and the robustness plot appears in Figure 6.52. The triangle for $0 < \sigma^2 \leq 0.25$

![Figure 6.51](image1)

Figure 6.51 MLE $A_V$-triangle plots for $\sigma^2$ in the $\{1\}$-stage(2Ci) structure with interaction.

![Figure 6.52](image2)

Figure 6.52 MLE $A_V$-robustness plot for $\sigma^2$ in the $\{1\}$-stage(2Ci) structure with interaction.

has the optimality regions much better defined than in the respective $D_V$-triangle. However, Region 3 also fade noticeably. There is no mirror effect for $A_V$-optimality.
and designs $d_2$ and $d_1$ are optimum for $\sigma_1^2 < 0.5$ and $\sigma_1^2 > 0.5$, respectively, when $\sigma_3^2 > 0.25$.

The self efficiency of the three designs is high being no lower than 0.94 in Region 1. The relative efficiency plot shows that the designs are equally good at the intersection. This can be seen in Region 3 as well, and the self efficiency plot reflects the fading of the region.

Comparing $D_V$- and $A_V$-optimality, it is obvious that under MLE and the model with interaction, $D_V$-optimality is not strong.

The REML $D_V$-optimality triangular plots are shown in Figure 6.53 and the robustness plots in Figure 6.56 on page 210.

The triangle plot for $0 < \sigma_3^2 < 0.25$ shows a very well contained Region 3. For $\sigma_3^2 > 0.25$ the situation is the same as in Figure 6.51 on the preceding page, where only $d_2$ and $d_3$ can be optimum. The main difference between the $D_V$-optimality plots when using MLE and REML is in the self efficiency. In the present case, the behaviour is more uniform and the minimum value is 0.15. Region 3 is comparatively highly self efficient despite it fading noticeably.

For the case of REML $A_V$-optimality, the triangle plot is presented in Figure 6.55 and the robustness plot in Figure 6.56 on page 210.
6.4 Designs to estimate \( \sigma^2 \) for the \{1\}{3-stage(2Ci)} structure (interaction model)

There is no clear difference in the triangular plots of when using MLE or REML. The observations and interpretations are the same. Differences in the curves can be seen in the robustness plots. REML shows a smoother plot, but the remarks are still the same as in the MLE case.

The previous example showed that the main differences are found in the optimality criteria rather than the method of estimation, as it was also found in the case without interaction.
6 Design construction algorithms

6.4 Designs to estimate $\sigma^2$ for the \{1\}3-stage(2Ci) structure

(enumeration model)

Figure 6.56  REML $A_V$-robustness plot for $\sigma^2$ in the \{1\}3-stage(2Ci) structure with interaction.

6.4.2 N=16, MLE vs REML and pseudo-Bayesian optimality

Example 21 Total lack of knowledge for $\sigma^2$

The results for $D_V$- and $A_V$-optimality are presented in Figure 6.57 as bar plots.

Figure 6.57  MLE pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma^2$ is unknown in the \{1\}3-stage(2Ci) structure with interaction.

The corresponding relative efficiency plots appear in Figure 6.58.

Both criteria give the same result showing $d_2$ as the optimum design followed
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$-3-stage(2Ci) structure

The relative efficiency of these two designs with respect to each other is very close to one. For both criteria, $d_3$ is comparatively poor in relative efficiency.

The interpretation of these results coincides with that of the case without interaction in Example 7, although the pair $(d_2, d_1)$ is reversed there.

The bar plots with the results for REML $D_Y$- and $A_Y$-optimality are presented in Figure 6.59, and the corresponding relative efficiency plots are shown in Figure 6.60.

The REML results here for both criteria are comparable to those in Example 11.
6.4 Designs to estimate $\sigma^2$ for the \{1\}{3-stage(2Ci)} structure

6. Design construction algorithms

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure660.png}
\caption{REML pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.}
\end{figure}

for the case without interaction. As it happened with the MLE plots above, apart from the difference in the scale of the optimality criterion values, the conclusions are the same as in Example 11.

\textbf{Example 22} \(\sigma_1^2 > \sigma_2^2 > \sigma_t^2\)

The bar plots for the MLE $D_V$- and $A_V$-optimality are shown in Figure 6.61. The relative efficiency plots for the two criteria are given as Figure 6.62. The results and interpretation in this case are the same as the one given in Example 8 for the model without interaction. The only difference, as in the MLE case above, is the scale of the values of the optimality criteria which is irrelevant for this comparison. Now consider, the REML method of estimation. The REML $D_V$- and $A_V$-optimality bar plots are shown in Figure 6.63. The relative efficiency plots are shown in Figure 6.64 on page 214.

Compare the $A_V$-optimality plots on the right of Figure 6.63 and the corresponding plot in Figure 6.61. It is clear that the results for REML in this example and the MLE results above convey the same information. The relative efficiency plots are also the same. Additionally, the optimum $D_V$-designs for MLE and REML swap positions. As was explained in Example 11, that MLE $D_V$-optimality
6.4 Designs to estimate $\sigma^2$ for the \{1\}\{3-stage(2Ci)\} structure

Design construction algorithms (interaction model)

Figure 6.61  MLE pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the \{1\}\{3-stage(2Ci)\} structure with interaction.

Figure 6.62  MLE pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the \{1\}\{3-stage(2Ci)\} structure with interaction.

chooses $d_2$ is counter intuitive.

The differences between optimality criteria and between method of estimation are difficult to find in the current scenario of pseudo-Bayesian optimality. Similarly, the differences between the model with interaction and without interaction for this case of $N = 16$ are almost imperceptible.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}\{3\text{-stage}(2Ci)\}$ structure

6.4.3 $N=24$ and MLE vs REML

Example 23

The triangular plot for MLE $D_V$-optimality is shown in Figure 6.65 followed by the robustness plot in Figure 6.66 on page 216.

As in Section 6.4.1, for $\sigma^2_3 < 0.25$ the triangle plot is very similar to the case without interaction. The left most triangle in Figure 6.65 shows the main or general characteristics of the triangle plot that were described before, such as...
symmetry and mirror and complementary effects. However, not all the regions have a well defined edge. This may indicate that the optimality criteria becomes very sensitive for certain $\sigma^2$ ranges.

![D-optimality plots](image)

**Figure 6.65** MLE $D_V$-triangle plots for $\sigma^2$ in the $\{1\}$-3-stage(2Ci) structure with interaction.

For $0.25 < \sigma^2_3 \leq 0.75$, designs $d_5$ to $d_9$ are no longer optimum. Only $d_1$ to $d_4$ have a region of optimality. When $\sigma^2_3 > 0.75$ virtually all the remaining region is covered by $d_2$ and $d_3$. These two designs spread the design weight more uniformly over the first two variability stages. This is not surprising since if the interaction is large, then more information about the two random factors is desirable to be able to better estimate the interaction parameter as well.

Surprisingly, $d_2$ and $d_3$ have a very poor self efficiency. Since the mirror and complementary effect on the optimality regions can still be observed, as seen in previous examples, the neighbour and complementary designs for any given region exhibit high relative efficiency.

The triangular plot for MLE $A_V$-optimality is shown in Figure 6.67 and the associated robustness plot is shown in Figure 6.68. In this case, the optimality regions are much better defined than with $D_V$-optimality where $\sigma^2_3 < 0.6$. For $\sigma^2_2 > 0.6$, the regions are outlined, but not very well defined. When $\sigma^2_3 > 0.25$ only designs $d_1$ to $d_4$ have optimality regions. For example, $d_4$ is optimum when $0.25 < \sigma^2_3 \leq 0.5, \sigma^2_2 > 0.6$ and $\sigma^2_1 < 0.32$. For even greater values of $\sigma^2_3$ only $d_2$
6.4 Designs to estimate $\sigma^2$ for the $\{1\}{3\text{-stage}(2^3)}$ structure (interaction model)

and $d_3$ can be optimum. The rationale behind this is the same as described in Section 6.4.1.

The self efficiency is generally good not being lower than 0.95 in any region. As seen before when the model had no interaction term, the complementary designs are the most relative efficient in any given region. Although the optimum designs are not outperformed in their own regions.

Next, results using REML are presented. The triangular plot for REML $D_V$-optimality appears in Figure 6.69 and the associated robustness plot is shown in

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6_66.png}
\caption{MLE $D_V$-robustness plot for $\sigma^2$ in the $\{1\}{3\text{-stage}(2^3)}$ structure with interaction.}
\end{figure}
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$ 3-stage($2^2$) structure

6 Design construction algorithms

(interaction model)

Figure 6.67 MLE $A_Y$-triangle plots for $\sigma^2$ in the $\{1\}$ 3-stage($2^2$) structure with interaction.

Figure 6.68 MLE $A_Y$-robustness plot for $\sigma^2$ in the $\{1\}$ 3-stage($2^2$) structure with interaction.
6.4 Designs to estimate $\sigma^2$ for the \{1\}\{3-stage(2Ci)\} structure

Figure 6.70 on the next page.

REML $D_V$-optimality does not present the mirror effect. This is still the case as can be seen in the leftmost triangle of Figure 6.69. Also, it is difficult to clearly define a region of optimality for designs $d_5$ to $d_9$. The robustness plot of Figure 6.70 on the following page indicates that Region 7 to Region 9 do not extend far into the simplex since their absolute distance goes from 0.3 on Region 7 down to 0.1 for Region 9.

For the range $0.25 < \sigma_3^2 \leq 0.5$ and $0.5 < \sigma_3^2 \leq 0.75$ designs $d_1$ to $d_4$ have a small region of optimality. For the range $0.75 < \sigma_3^2 < 1$ only $d_2$ and $d_3$ are optimum in their respective regions. However, self efficiency is considerably low. This has been the case with $D_V$-optimality in most examples. Region 2 and Region 3 self efficiency is as low as 0.2. The effect of the complementary designs can be seen in the relative efficiency plots.

![D-optimality](image)

**Figure 6.69** REML $D_V$- triangle plots for $\sigma^2$ in the \{1\}\{3-stage(2Ci)\} structure with interaction.

The triangular plot for REML $A_V$-optimality can be consulted in Figure 6.71 and the associated robustness plot is presented as Figure 6.72 on page 220. Is not easy to distinguish between the leftmost triangles of Figure 6.67 and Figure 6.71. But in the latter is possible to note that $d_9$ is hardly present. In fact, Region 9 of Figure 6.72 only extends to 0.1 in absolute distance. Therefore, in practice there is no plausible region where $d_9$ is optimum. As seen in previous examples
6.4 Designs to estimate $\sigma^2$ for the $\{1\}{3\text{-stage}(2Ci)}$ structure (interaction model)

Figure 6.70 REML $D_V$-robustness plot for $\sigma^2$ in the $\{1\}{3\text{-stage}(2Ci)}$ structure with interaction.

Figure 6.71 REML $A_V$-triangle plots for $\sigma^2$ in the $\{1\}{3\text{-stage}(2Ci)}$ structure with interaction.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}\{3\text{-stage(2Ci)}\}$ structure (interaction model)

Figure 6.72  REML $A_V$- robustness plot for $\sigma^2$ in the $\{1\}\{3\text{-stage(2Ci)}\}$ structure with interaction.

of $A_V$-optimality when using REML, the optimality regions are more uniform than with MLE. This can also be seen here. Apart from these considerations, the interpretation is the same as the one given for MLE before.

Most of the differences seen between the cases presented in this section are found for the condition $0 < \sigma^2 \leq 0.25$. This is where similarities to the results for the model without interaction are expected. For the condition $\sigma^2 > 0.25$ differences are not obvious. REML seems to have better defined regions of optimality and
more uniform.

### 6.4.4 N=24, MLE vs REML and pseudo-Bayesian optimality

The results when considering the pseudo-Bayesian optimality criteria are presented below.

**Example 24** Total lack of knowledge for \( \sigma^2 \)

The bar plots for MLE \( D_V \)- and \( A_V \)-optimality are presented in Figure 6.73 and the corresponding relative efficiency plots in Figure 6.74.

**Figure 6.73** MLE pseudo-Bayesian \( D_V \)- and \( A_V \)-optimality plots when \( \sigma^2 \) is unknown in the \( \{1\}\{3\text{-stage}(2Ci)\} \) structure with interaction.

Both criteria show the same result in terms of optimum designs and relative efficiencies. For the condition being considered, the ordering \((d_3, \ d_2, \ d_1, \ d_4)\) is the most intuitive. Also, the complementary designs are highly relative efficient among each other.

The bar plots for REML \( D_V \)- and \( A_V \)-optimality are presented in Figure 6.75 and the corresponding relative efficiency plots in Figure 6.76. As with the MLE information matrix, both criteria in this case, show the same result in terms of
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$-3-stage(2Ci) structure

6 Design construction algorithms

(Interaction model)

optimum designs and relative efficiencies. For the condition being considered, the ordering ($d_2, d_3, d_1, d_4$) is also intuitive. In fact, using the relative efficiency plots, the pair ($d_2, d_3$) seem to be equivalent, similarly ($d_1, d_4$). Remember that the complementary designs are highly relative efficient.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$\{3-stage(2Ci)\} structure

(Interaction model)

Figure 6.76  REML pseudo-Bayesian $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the $\{1\}$\{3-stage(2Ci)\} structure with interaction.

Example 25 $\sigma_1^2 > \sigma_2^2 > \sigma_\epsilon^2$

The bar plots for MLE $D_V$- and $A_V$-optimality are presented in Figure 6.77 and the corresponding relative efficiency plots in Figure 6.78 on the following page.

Figure 6.77  MLE pseudo-Bayesian $D_V$- and $A_V$-optimality plots when $\sigma^2$ is unknown in the $\{1\}$\{3-stage(2Ci)\} structure with interaction.

$D_V$-optimality is once more unable to provide results that are intuitive. On the other hand, $A_V$-optimality gives expected results and it offers good discrimination between designs as can be seen in the relative efficiency plot.

For the case of REML $D_V$- and $A_V$-optimality, bar plots are shown in Figure 6.79 and the corresponding relative efficiency plots in Figure 6.80 on page 225.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$-{3-stage(2Ci)} structure

The results for REML $D_V$-optimality select $d_2$ as the optimum design. This option is acceptable. Not having $d_1$ as the second best, though, is not necessarily expected. As in the MLE case, $A_V$-optimality gives more intuitive and consistent results.

As it was done in the case without interaction, $D_V$-optimality does not seem to properly respond the restriction given. Once more, $A_V$-optimality behaves consistently in both methods of estimation.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$3-stage(2Ci) structure

6.4.5 $N=32$ and MLE vs REML

Example 26 $D_v$-optimality

The triangular plot for MLE $D_v$-optimality is shown in Figure 6.81 followed by the robustness plot in Figure 6.82 on the following page. The leftmost triangle plot exhibits the complementary symmetry similar to the case with no interaction presented in Section 6.3.7. Here, the mirror effect is not present though. The regions for $d_4$ and $d_5$ are not very well defined. In this case, for the condition $\sigma_3^2 \leq 0.25$, the triangle plot does not resemble the one for the model without interaction. For larger values of $\sigma_3^2$ designs $d_4$, $d_5$ and $d_6$ do not have a region of optimality. Conversely, $d_1$ and $d_3$ always have an optimality region. For example, when $0.25 < \sigma_3^2 \leq 0.5$ Region 1 is defined for $\sigma_1^2 < 0.09$ and $\sigma_2^2 > 0.1$; when $0.5 < \sigma_3^2 \leq 0.75$ the region in delimited by $\sigma_1^2 < 0.11$ and $\sigma_2^2 > 0.25$, and when $\sigma_3^2 > 0.75$ the conditions are $\sigma_1^2 < 0.1$ and $\sigma_2^2 > 0.5$. Region 2 spans the majority of the simplex for all conditions of $\sigma_3^2$.

The robustness plot show the same as in previous cases, Region 1 to Region 3 have very low self efficiency, but the corresponding design is never outperformed. The complementary and even neighbourhood effect in relative efficiency takes
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$-$\{3$-stage$(2Ci)\}$ structure

(Interaction model)

Figure 6.81 MLE $D_\gamma$-triangle plots for $\sigma^2$ in the $\{1\}$-$\{3$-stage$(2Ci)\}$ structure with interaction.

Figure 6.82 MLE $D_\gamma$-robustness plot for $\sigma^2$ in the $\{1\}$-$\{3$-stage$(2Ci)\}$ structure with interaction.
place here as well.

The triangular plot for MLE $\mathcal{A}_V$-optimality is shown in Figure 6.83 and the associated robustness plot is shown in Figure 6.84 on the next page.

![Figure 6.83](image-url)  
**Figure 6.83** MLE $\mathcal{A}_V$-triangle plots for $\sigma^2$ in the \{1\}{3-stage(2Ci)} structure with interaction.

The triangle plot corresponding to $\sigma^2_3 \leq 0.25$ in Figure 6.83 shows Region 1 to Region 3 very well defined, and comparatively with $\mathcal{D}_V$-optimality, Region 4 to Region 6 are discernible. As it has happened in all previous cases, for the condition $\sigma^2_3 > 0.25$ designs other than $(d_1, d_2, d_3, d_4)$ do not possess a region of optimality. For the present case of MLE $\mathcal{A}_V$-optimality, clearly defined Region 2 and Region 3 can be seen for the condition $0.25 < \sigma^2_3 \leq 0.5$. For $\sigma^2_3 > 0.5$, only Region 2 remains optimal.

Self efficiencies are high for all regions, as has been the case in most scenarios under $\mathcal{A}_V$-optimality. The relative efficiencies also are high and uniform across the regions and they reflect the neighbour and complement effect.

The plots for REML $\mathcal{D}_V$-optimality appear below where Figure 6.85 is the triangle plot and Figure 6.86 is the robustness plot.

The triangular plots for MLE and REML are not significantly different. In the REML leftmost triangle, the regions are arguably better defined, but they look smaller for $d_4$ to $d_6$. The main difference is that for $\sigma^2_3 > 0.75$ virtually only Region 2 appears. However, the self efficiency for Region 2 is improved. All other
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$\{3-stage(2Ci)\} structure

interaction model)

Figure 6.84 MLE $A_V$-robustness plot for $\sigma^2$ in the $\{1\}$\{3-stage(2Ci)\} structure with interaction.

remarks made for MLE are also applicable in this case.

The set of plots for REML $A_V$-optimality are shown in Figure 6.87 on page 230 and Figure 6.88 for the triangle plot and the robustness plot, respectively.

In this instance, the set of plots for REML just introduced are in practical terms the same as the MLE plots in Figure 6.83 and Figure 6.84. No further or more insightful conclusions can be obtained here.
6.4 Designs to estimate $\sigma^2$ for the \{1\}3-stage(2Ci) structure

Figure 6.85 REML $D_Y$ triangle plots for $\sigma^2$ in the \{1\}3-stage(2Ci) structure with interaction.

Figure 6.86 REML $D_Y$ robustness plot for $\sigma^2$ in the \{1\}3-stage(2Ci) structure with interaction.
6.4 Designs to estimate $\sigma^2$ for the \{1\}{3-stage(2Ci)} structure (interaction model)

![A-optimality triangle plots](image)

**Figure 6.87** REML $A_V$- triangle plots for $\sigma^2$ in the \{1\}{3-stage(2Ci)} structure with interaction.

![Efficiency plots](image)

**Figure 6.88** REML $A_V$- robustness plot for $\sigma^2$ in the \{1\}{3-stage(2Ci)} structure with interaction.
6.4 Designs to estimate $\sigma^2$ for the \{1\}3-stage(2Ci) structure

The recommendation in this case is to use $A_V$-optimality due to the higher and more uniform self efficiency of the designs, which in turn makes a better definition, in most cases, of the available regions of optimality.

6.4.6 $N=32$, MLE vs REML and pseudo-Bayesian optimality

Example 27 Total lack of knowledge for $\sigma^2$

The bar plots for MLE $D_V$- and $A_V$-optimality are presented in Figure 6.89 and the corresponding relative efficiency plots in Figure 6.90. Both criteria give the same result in the ordering of the designs with $d_2$ being the optimum. Design $d_2$ puts equal weight at each of the first two variability stages, which is not unreasonable for the given condition on $\sigma^2$. Designs $d_3$ and $d_1$ put the maximum weight at stage two and stage one, respectively. These two are about 96% relative efficient under both criteria of optimality. In general, the efficiency plots are highly reminiscent of each other. In this case, $D_V$- and $A_V$-optimality are equivalent.

The bar plots for REML $D_V$- and $A_V$-optimality are presented in Figure 6.91 and the corresponding relative efficiency plots in Figure 6.92 on page 233.
6 Design construction algorithms

6.4 Designs to estimate $\sigma^2$ for the \{1\}\{3-stage(2Ci)\} structure

As it can be seen from the two figures, REML favours the pair ($d_1$, $d_3$) and MLE prefers ($d_1$, $d_3$). However, in terms of relative efficiency both designs are equally good. The same happens with the pair ($d_4$, $d_5$). Design $d_2$ is the optimum. Therefore, the differences between MLE and REML are negligible.

Example 28 $\sigma_1^2 > \sigma_2^2 > \sigma_\epsilon^2$

The bar plots for MLE $D_V$- and $A_V$-optimality are presented in Figure 6.93 and the corresponding relative efficiency plots in Figure 6.94.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$\{3-stage(2Ci)\} structure

Similarly as in Example 25, pseudo-Bayesian MLE $\mathcal{D}_V$-optimality for the present case, does not give convincing results. Design $d_2$ is reported as optimum, which can be reasonable. The second best design, $d_3$, puts the maximum effort in the second stage, which is not the intuitive result. The relative efficiency plot shows that $d_3$ and $d_1$ cannot be considered to be equally good. In general, this criterion does not choose the design that puts more resources in the variability stage with more uncertainty. $\mathcal{D}_V$-optimality does not seem sensitive to the restriction imposed on $\sigma^2$. 

Figure 6.92  REML pseudo-Bayesian $\mathcal{D}_V$- and $\mathcal{A}_V$-optimality efficiency plots when $\sigma^2$ is unknown in the $\{1\}$\{3-stage(2Ci)\} structure with interaction.

Figure 6.93  MLE pseudo-Bayesian $\mathcal{D}_V$- and $\mathcal{A}_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the $\{1\}$\{3-stage(2Ci)\} structure with interaction.
6.4 Designs to estimate $\sigma^2$ for the $\{1\}$-$\{3\text{-stage}(2\text{Ci})\}$ structure

6 Design construction algorithms

(Interaction model)

On the other hand, as in Example 25, MLE $A_V$-optimality results in a design ordering that is much more intuitive. Here $d_1$ is chosen as optimum as expected. This ordering gives priority to the designs that put more weight in the first variability stage, where $\sigma_1^2$ is largest. Moreover, the relative efficiency of the competing designs is progressively smaller ranging from 0.94 to 0.69.

For the case of REML $D_V$- and $A_V$-optimality, bar plots are shown in Figure 6.95 and the corresponding relative efficiency plots in Figure 6.96.

REML $D_V$-optimality chooses $d_2$ as the optimum design, as in the MLE case. The following pair of designs is $(d_1, d_3)$ which is acceptable. The relative efficiencies of this pair is different, therefore having $d_1$ as the second best design is a good result. For $A_V$-optimality, the REML version is the same as the MLE version, this giving a completely expected result.

In this case, using REML improves noticeably the performance compared to the MLE $D_V$- criterion. It still chooses $d_2$ as optimum and gives $d_3$ and $d_1$ a very similar relative efficiency. REML $A_V$-optimality is no different than MLE, giving the expected results. It also shows good ability to discriminate between the competing designs.
6.5 Designs to estimate $\eta$ in the $\{1\}$-$\{3\}$-stage($2C$) structure

Consider that for the system under investigation the variance ratios are informative and of interest to the objectives of the study. An optimum design for the ratios can be obtained using the algorithm described in this chapter. This means that different information matrices and methods of estimation can be used. The information matrices used in the following examples are presented in Chapter 5.
In the algorithm, the grid of ratios is obtained by first generating a random vector for $\sigma^2$. Another two or three random vectors (depending on the model) are generated and each one gets divided by the $\sigma^2$ vector first obtained. Note that the parameter space for the ratios does not depend on a specific value of $\sigma^2$. Therefore, designs for the variance ratios can be evaluated without specific knowledge of $\sigma^2$.

For the present case, there are only two variance ratios. A rectangular plot can be used that conveys all the relevant information about the patterns of the variability with respect to the candidate designs, in a similar way as the triangular plots did in the previous examples. These rectangular plots have all the features that were described for the triangular plots in previous sections, but the interpretation is slightly different. Here, an optimum design is chosen according to the different proportions or weight of, e.g. $\eta_1$ to $\eta_2$, independently of the value of $\sigma^2$. A specific feature of the rectangular plots is that they make it possible to identify at which difference in the ratios one factor takes more relevance or provides more information about the variability of the model, and thus how the optimum design changes accordingly. The effect of the cases where one ratio increases much more than the other, and also when both ratios change in the same amount can also be easily identified.

### 6.5.1 N=16 and MLE vs REML

The list of candidate designs to be compared appears in Table 6.1 on page 165.

**Example 29** $\mathcal{D}_V$-optimality

The MLE $\mathcal{D}_V$-optimality plot can be found in Figure 6.97. The associated robustness plot is shown in Figure 6.98.

It can be seen that the three competing designs get to establish an optimality
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the $\{1\}$-{3-stage(2C)} structure

region. The symmetry line that was described previously for the triangular plots is also found here. Obviously, in this case the line corresponds to the condition $\eta_1 = \eta_2$. The mirror effect is not present here. An optimum design for a specific pair of ration of interest is found directly from the plot. As an example, consider the situation when $\eta_1 \geq 0.75$, $\eta_2 \geq 0.75$ and also that $\eta_1 > \eta_2$. According to the plot, design $d_2$ is $D_v$-optimum. If $\eta_1 \gg \eta_2$ then $d_1$ is optimum. The optimum designs for any other condition can be found similarly. In particular $d_3$ is optimum when $\eta_1 < 0.1$ and $\eta_2 < 0.1$.

Two additional interesting cases can be easily identified. When $\eta_\alpha \to \infty$ and


\( \eta_3 \to \infty \) the condition is equivalent to having a fixed effects model in blocks, which as mentioned before \( d_1 \) and \( d_2 \) are both optimum under the two criteria. When \( \eta_1 \to 0 \) and \( \eta_2 \to 0 \) design \( d_3 \) is \( \mathcal{D}_V \)-optimum and corresponds to the case when the most predominant variance component is \( \sigma_x^2 \).

The robustness plot in Figure 6.98 is used and interpreted as in previous sections. The self efficiency of Region 1 and Region 2 drop significantly across the regions, but the corresponding design is never outperformed. Region 3 shows a much higher self efficiency.

![Figure 6.99](image1.png)

**Figure 6.99** REML \( \mathcal{D}_V \)-plot for \( \eta \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

![Figure 6.100](image2.png)

**Figure 6.100** REML \( \mathcal{D}_V \)-robustness plot for \( \eta \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

The set of plots for the REML \( \mathcal{D}_V \)-optimality can be seen in Figure 6.99 and
Figure 6.100, respectively.

Note that the line of symmetry remains, but Region 1 and Region 2 are no longer spread in two sections. Region 1 corresponds to the condition $\eta_1 > \eta_2$, and for Region 2, $\eta_2 > \eta_1$. These conditions do not apply for the, comparatively, very small Region 3. Self efficiencies of Region 1 and Region 2 are almost the same as for the MLE case. For REML, relative efficiencies are very close to one across the two regions. This suggests that $d_1$ and $d_2$ are equivalent in most of the parameter space. Region 3 is highly self efficiency in a very small portion, where both variance ratios are very small.

**Example 30** $A_V$-optimality

![Figure 6.101](image.png)

**Figure 6.101** MLE $A_V$-plot for $\eta$ in the $\{1\} \{3\text{-stage}(2C)\}$ structure.

The MLE $A_V$-optimality plot is shown in Figure 6.101 and the robustness plot in Figure 6.102.

The optimality plot is very similar to the REML $D_V$-optimal one in Figure 6.99. The main difference is Region 3 which is larger here and of triangular shape rather
than diamond like. The self efficiency plots show that under the MLE $A_V$-optimality criterion the respective performances in Region 1 and Region 2 are slightly better, therefore giving a lower relative efficiency. The performance in Region 3 is high and more uniform than in the other two cases already reviewed.

For the REML $A_V$-optimality criterion the corresponding plots are shown in Figure 6.103 and Figure 6.104 for the optimality plot and the robustness plot, respectively.

The MLE and REML $A_V$-plots are almost the same. In the REML plot, Region 3 is diamond shaped and is triangularly shaped in the MLE plot. The
6.5 Designs to estimate \( \eta \) in the \( \{1\}\{3\text{-stage(2C)}\} \) structure

Figure 6.104  REML \( \mathcal{A}_V \)-robustness plot for \( \eta \) in the \( \{1\}\{3\text{-stage(2C)}\} \) structure.

Self efficiency is better for REML with much lower relative efficiencies, which is desired.

6.5.2 N=16, MLE vs REML and pseudo-Bayesian optimality

Example 31  Total lack of knowledge for \( \eta \)

The MLE \( D_V \)- and \( \mathcal{A}_V \)-optimality bar plots can be found next to each other in Figure 6.105. The corresponding relative efficiency plots are shown in Figure 6.106.

The MLE \( D_V \)-optimality criterion provides no useful information for choosing an optimum design. The relative efficiency of \( d_2 \) with respect to \( d_1 \) is greater than 0.999 and that of \( d_3 \) is greater than 0.996. Under the \( \mathcal{A}_V \)-criterion \( d_2 \) is chosen as the optimum. The efficiency plot reveals that \( d_1 \) is 0.998 relative efficient. The two designs are practically equivalent. Design \( d_3 \) is far from optimum.

The REML optimality plots can be seen in Figure 6.107 and the efficiency plots in Figure 6.108 on page 243.

The two types of optimality criteria coincide in that \( d_3 \) is far from optimum.

241
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the $\{1\}{3\text{-stage(2C)}\}$ structure

This can be seen in the optimality plot as well as in the relative efficiency plots. $D_V$-optimality better discriminate between $d_2$ and $d_1$ and chooses the former as the optimum. The $A_V$- criterion prefers $d_1$ as the optimum design, but $d_2$ is reported to be 0.98 relative efficient.

**Example 32** $\eta_1 > \eta_2$

When the selected method of estimation is MLE, Figure 6.109 presents the optimality plots with the relative efficiency plots being shown in Figure 6.110.
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the $\{1\}{3\text{-stage}(2C)}$ structure

Figure 6.107 REML $D_V$- and $A_V$-optimality plots when $\eta$ is unknown in the $\{1\}{3\text{-stage}(2C)}$ structure.

Figure 6.108 REML $D_V$- and $A_V$-optimality efficiency plots when $\eta$ is unknown in the $\{1\}{3\text{-stage}(2C)}$ structure.

Under MLE and $D_V$-optimality, design $d_2$ is optimum. However, the relative efficiency of $d_1$ is 0.98 as can be read from the efficiency plot. Therefore, the $D_V$-criterion does not discriminate confidently. In the case of $A_V$-optimality, design $d_1$ is optimum with $d_2$ having a relative efficiency of 0.93. This number is preferable than the 0.98 of the $D_V$- criterion. In neither case $d_3$ is competitive.

For the REML case, the optimality plots are given in Figure 6.111 whereas the relative efficiency plots are given in Figure 6.112 on page 245.

It is clear that both optimality criteria provide the same ordering of the
6.5 Designs to estimate $\eta$ in the \{1\}{3-stage(2C)} structure

Candidate designs, which is $(d_1, d_2, d_3)$. $D_V$-optimality gives $d_2$ a relative efficiency value of 0.968, whereas this is 0.82 for $A_V$-optimality which is preferable. For $d_3$ the same observations as before apply.

6.5.3 N=24 and MLE vs REML

The list of candidates for $N = 24$ contains nine designs; these are presented in Table 6.2 on page 165.
6.5 Designs to estimate $\eta$ in the $\{1\}{\{3\text{-stage}(2C)\}}$ structure

Example 33 $D_V$-optimality

Optimality plots are given in Figure 6.113. The respective robustness plots can be seen in Figure 6.114 on page 247.

As it has been the case in all examples before, MLE $D_V$-optimality shows a symmetry line in the optimality plots. In this case, also the complementary and mirror effects, as discussed previously, can be seen. Additionally, in Figure 6.113 only four of the nine candidates have an associated region of optimality. When $\eta_2 > \eta_1$, as the value of $\eta_2$ increases, the optimum design changes progressively in
the order \((d_4, d_3, d_2, d_1)\). When the condition is reversed and \(\eta_1 > \eta_2\) the order is also reversed giving \((d_1, d_2, d_3, d_4)\). The self efficiency in each region is low, as depicted in Figure 6.114. The relative efficiency of the contiguous design of any given region is high and it decreases progressively according to the aforementioned ordering.

![D-optimality](image)

**Figure 6.113** MLE \(D_V\)-plot for \(\eta\) in the \(\{1\}\{3\text{-stage}(2C)\}\) structure.

The REML \(D_V\)-optimality plot is shown in Figure 6.115 and the robustness plot in Figure 6.116.

The REML case here and the MLE case above show a significantly different result. As before, there is a symmetry line and only \(d_1\) to \(d_4\) have a region of optimality. However, \(d_2\) dominates when \(\eta_1 > \eta_2\) and \(d_3\) dominates when \(\eta_2 > \eta_1\). For the extreme cases of \(\eta_1 >> \eta_2\) and \(\eta_2 >> \eta_1\), designs \(d_1\) and \(d_4\), respectively, have a small region of optimality. The self efficiency is better with REML and it does not go lower than 0.2 in any region. Also, the relative efficiencies are lower in general with only the largest neighbour design having high efficiency across the region of optimality.
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure

Figure 6.114 MLE $D_V$-robustness plot for $\eta$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure.

Figure 6.115 REML $D_V$-plot for $\eta$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure.

Example 34 $A_V$-optimality

The results for MLE $A_V$-optimality can be seen in Figure 6.117 as an optimality plot with the corresponding robustness plot in Figure 6.118 on page 249.
6.5 Designs to estimate $\eta$ in the \{1\}\{3-stage(2C)\} structure

![Diagram showing optimality plots for different regions.]

**Figure 6.116** REML $D_V$-robustness plot for $\eta$ in the \{1\}\{3-stage(2C)\} structure.

The optimality plot shows the symmetry effects and the complementary effect. Again, only four designs have a region of optimality. Compared to the $D_V$-criterion examples above, the $A_V$-regions are more evenly spread. Nonetheless, $d_1$ and $d_4$ cover the majority of the plot. The self efficiency curves, shown in Figure 6.118 on the following page, reveal that the efficiency in any given region is comparable with all other previous examples, which is in general low. The relative efficiencies, on the other hand, are quite high with respect to neighbour designs.

For REML $A_V$-optimality the optimality plots are presented in Figure 6.119, and the robustness plot is given in Figure 6.120 on page 250.

When using REML, the optimality plot have very uniform optimality regions and, this time, six designs have an optimality region. The complementary effect
6.5 Designs to estimate $\eta$ in the $\{1\}{3\text{-stage}(2C)}$ structure

Figure 6.117 MLE $A_V$-plot for $\eta$ in the $\{1\}{3\text{-stage}(2C)}$ structure.

Figure 6.118 MLE $A_V$-robustness plot for $\eta$ in the $\{1\}{3\text{-stage}(2C)}$ structure.

can be observed. Designs $d_5$ and $d_6$ have rather small regions that correspond to the respective scenarios where one variance ratio is much larger than the other one, similar to the conditions given in reference to Figure 6.115 on page 247.
The self efficiencies in Region 1 to Region 4 are some of the largest found so far. The relative efficiencies in most regions are lower than in the MLE case above also with larger differences among regions.

Figure 6.119  REML $A_V$- plot for $\eta$ in the $\{1\}$\{3-stage(2C)\} structure.

Figure 6.120  REML $A_V$- robustness plot for $\eta$ in the $\{1\}$\{3-stage(2C)\} structure.
6.5 Designs to estimate $\eta$ in the \{1\}{3-stage(2C)} structure

### 6.5.4 N=24, MLE vs REML and pseudo-Bayesian optimality

**Example 35** Total lack of knowledge for $\eta$

The MLE $D_V$- and $A_V$-optimality plots are shown in Figure 6.121. The corresponding relative efficiency plots are shown in Figure 6.122 on the next page.

First note that only four designs are compared from the available nine. For $D_V$-optimality design $d_3$ is optimum, but the relative efficiency of $d_2$ is 0.98. The same can be observed with the pair $(d_4, d_1)$. The same behaviour can be seen under the $A_V$- criterion, but the design pairs are reversed.

In this example, there is no clear distinction as to which optimality criteria under MLE is better. Given the high relative efficiency of the design pairs, the experimenter may decide based on cost or ease of use of a particular design.

![MLE $D_V$- and $A_V$-optimality plots for when $\eta$ is unknown in the \{1\}{3-stage(2C)} structure.](image)

When considering REML, the respective optimality plots are given in Figure 6.123, whilst the relative efficiency plots appear in Figure 6.124. The obvious difference is that $A_V$-optimality has six designs not only four as the $D_V$- criterion. However, only the pair $(d_2, d_3)$ is worth considering, and the final conclusion is the same as in the MLE case above.
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the $\{1\}{3\text{-stage}(2C)}$ structure

For the condition of total lack of knowledge for $\eta$, there is no clear indicator that one combination of method of estimation and optimality criteria gives better results. The recommendation is then to choose based on practical matters.

**Example 36** $\eta_1 > \eta_2$

The MLE $D_V$- and $A_V$-optimality plots can be seen in Figure 6.125. The plots in Figure 6.126 on page 254 show the associated relative efficiencies.

For the present conditions, $D_V$-optimality is not really able to provide an optimum design with good comparative performance. This was already seen for
6 Design construction algorithms

6.5 Designs to estimate \( \eta \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure

- **Figure 6.124** REML \( D_V \)- and \( A_V \)-optimality efficiency plots when \( \eta \) is unknown in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

- **Figure 6.125** MLE \( D_V \)- and \( A_V \)-optimality plots when \( \eta_1 > \eta_2 \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

\( N = 16 \). Therefore, the sample size in this case do not give better information. On \( A_V \)-optimality it is clearer that the pair \((d_1, d_2)\) is better than the other designs.

The relative efficiency of \( d_2 \) with respect to the optimum \( d_1 \) is more than 0.99. Therefore this is an equivalent pair of designs.

The REML optimality results for the \( D_V \)- and \( A_V \)-criteria are shown in Figure 6.127 with the corresponding efficiency plots given in Figure 6.128.

The \( A_V \)- criterion, as seen before, considers six designs instead of only four. These additional two designs have low relative efficiency. So that it makes no real
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the $\{1\}_{3\text{-stage(2C)}}$ structure

![MLE D_V- and A_V-optimality efficiency plots when $\eta_1 > \eta_2$ in the $\{1\}_{3\text{-stage(2C)}}$ structure.](image1)

![REML D_V- and A_V-optimality plots when $\eta_1 > \eta_2$ in the $\{1\}_{3\text{-stage(2C)}}$ structure.](image2)

Figure 6.126

Figure 6.127

The pair $(d_2, d_1)$ is again almost equivalent with $d_1$ having a relative efficiency of 0.98. The same phenomenon is observed with the $D_V$- criterion although the pair in question is $(d_2, d_3)$.

6.5.5 $N=32$ and MLE vs REML

The candidate list of designs to be considered when the total number of observations is $N = 32$ has six different designs. It is shown in Table 6.3 on page 165.
Design construction algorithms

6.5 Designs to estimate $\eta$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure

**Figure 6.128**  REML $D_V$- and $A_V$-optimality efficiency plots when $\eta_1 > \eta_2$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure.

**Example 37** $D_V$-optimality

The MLE $D_V$-optimality plot is given in Figure 6.129 with the robustness plots being shown in Figure 6.130. First, note in the optimality plot that only three out of the six candidate designs are optimum for the range of $\eta_1$ and $\eta_2$. The symmetry line is present and the complementary effect can be observed. The self efficiencies in all regions are low, but Region 2 has its associated relative efficiencies much lower than the other two regions. This means that $d_2$ is the best performer in its region.

**Figure 6.129**  MLE $D_V$- plot for $\eta$ in the $\{1\}\{3\text{-stage}(2C)\}$ structure.
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the \{1\}3-stage(2C) structure

![Figure 6.130](image1.png)

Figure 6.130  MLE $D_V$-robustness plot for $\eta$ in the \{1\}3-stage(2C) structure.

The REML $D_V$-optimality plot is shown in Figure 6.131. Its corresponding robustness plot is shown in Figure 6.132. Similarly to the MLE case, only $d_1$ to $d_3$ are being considered by the optimality criterion $D_V$. Here, only for the conditions $\eta_1 >> \eta_2$ and $\eta_2 >> \eta_1$ designs $d_1$ and $d_3$, respectively, have an associated region of optimality. Everywhere else in the parameter space, $d_2$ is optimum. As with MLE, given that the associated relative efficiencies in Region 2 are never close to one, then $d_2$ performs strongly in its region compared to the others.

![Figure 6.131](image2.png)

Figure 6.131  REML $D_V$-plot for $\eta$ in the \{1\}3-stage(2C) structure.

256
6 Design construction algorithms

6.5 Designs to estimate $\eta$ in the \{1\}{3-stage(2C)} structure

Figure 6.132 MLE $D_V$-robustness plot for $\eta$ in the \{1\}{3-stage(2C)} structure.

Example 38 $A_V$-optimality

The MLE $A_V$-optimality plot appears in Figure 6.133. The associated robustness plot is shown in Figure 6.134 on the following page.

Figure 6.133 MLE $A_V$-plot for $\eta$ in the \{1\}{3-stage(2C)} structure.

The optimality plot shows that $d_6$ does not have a region of optimality. The plot is still symmetric and it has the complementary effect. Depending on how larger $\eta_1$ is with respect to $\eta_2$ the optimum design changes progressively as $(d_4, d_1, d_2)$. Conversely, if $\eta_2$ is larger the progression is $(d_5, d_2, d_3)$. Looking at the robustness plot, in Region 1 to Region 3, design $d_3$ is the best performer in its region based on both, self efficiency and the associated relative efficiencies. Elsewhere, the
highest self efficiency is seen in Region 4 followed by Region 5. However, these are the less relevant since they only apply in a very small portion of the parameter space where some values of $\eta$ are extreme.

For the REML $A_V$-optimality plots refer to Figure 6.135, and to Figure 6.136.
for the corresponding robustness plot. The interpretation of this plot is the same as for the MLE version in Figure 6.133. The small regions here elongate slightly toward larger values of $\eta_1$ and $\eta_2$, respectively. No other differences are apparent in the optimality plot. In the robustness plot it is easy to see that from Region 1 to Region 3, Region 2 is the best performer under REML, in this case mainly due to the relative efficiencies in that region.

Comparing the $A_V$- and $D_V$- criteria, $A_V$-optimality is preferred because it looks to have more consistent results. Regarding the method of estimation, there is no clear indication that one is better.
6.5 Designs to estimate \( \eta \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure

6.5.6 \( N=32 \), MLE vs REML and pseudo-Bayesian optimality

Example 39 Total lack of knowledge for \( \eta \)

The optimality plots can be found in Figure 6.137 and the relative efficiency plot in Figure 6.138.

![Figure 6.137](image1.png)

**Figure 6.137** MLE \( D_\nu \)- and \( A_\nu \)-optimality plots when \( \eta \) is unknown in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

![Figure 6.138](image2.png)

**Figure 6.138** MLE \( D_\nu \)- and \( A_\nu \)-optimality efficiency plots when \( \eta \) is unknown in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

By inspecting the set of plots it is clear that both criteria identify design \( d_2 \) as the optimum. Designs \( (d_1, d_3) \) are equivalent in terms of relative efficiency under
both criteria, but still far from the optimum.

The optimality plots for the REML $D_V$- and $A_V$- criteria are found in Figure 6.139 and the corresponding relative efficiency plots are given in Figure 6.140.

Both figures reveal that the two optimality criteria give the design ordering $(d_2, d_3, d_1)$. Design $d_2$ is clearly the optimum.

![Figure 6.139](image1.png)

**Figure 6.139** REML $D_V$- and $A_V$-optimality plots when $\eta$ is unknown in the $\{1\}\{3$-stage(2C)\}$ structure.

![Figure 6.140](image2.png)

**Figure 6.140** REML $D_V$- and $A_V$-optimality efficiency plots when $\eta$ is unknown in the $\{1\}\{3$-stage(2C)\}$ structure.

For this condition on $\eta$ (lack of knowledge), both criteria and both methods of estimation clearly coincide on $d_2$ as being the optimum design.
Example 40 \( \eta_1 > \eta_2 \)

The optimality plots for MLE \( D_V \)- and \( A_V \)-optimality can be seen in Figure 6.141 and the relative efficiency plots in Figure 6.142.

![Figure 6.141](image1)

**Figure 6.141** MLE \( D_V \)- and \( A_V \)-optimality plots when \( \eta_1 > \eta_2 \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

![Figure 6.142](image2)

**Figure 6.142** MLE \( D_V \)- and \( A_V \)-optimality efficiency plots when \( \eta_1 > \eta_2 \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.

The set of plots for \( D_V \)-optimality show that the three competing designs have very high relative efficiency. For example, \( d_2 \) being the optimum, the relative efficiencies of \( d_3 \) and \( d_1 \) are 0.996 and 0.98, respectively. Therefore, at the practical level, they are equivalent. A similar phenomenon was seen previously. The
conclusion is that MLE $D_V$-optimality is not able to detect small differences in either $\eta$ or indirectly changes in $\sigma^2$.

On the other hand, for $A_V$-optimality, design $d_1$ is optimum with $d_2$ having a relative efficiency value greater than 0.99. Again, these can be considered equivalent in relative efficiency terms. The other designs in the plot, i.e. ($d_3$, $d_4$, $d_5$), do not compete with relative efficiencies lower than 0.88.

The REML optimality plots are shown in Figure 6.143 with the associated relative efficiency plots in Figure 6.144. For $D_V$-optimality, design $d_2$ is optimum.

![Figure 6.143](image1)

**Figure 6.143** REML $D_V$- and $A_V$-optimality plots when $\eta_1 > \eta_2$ in the \{1\}3-stage(2C) structure.

![Figure 6.144](image2)

**Figure 6.144** REML $D_V$- and $A_V$-optimality efficiency plots when $\eta_1 > \eta_2$ in the \{1\}3-stage(2C) structure.
with $d_1$ and $d_3$ far from optimal having relative efficiencies of 0.88 and 0.86, respectively. Under $A_V$-optimality $d_2$ is also optimum with $d_1$ having relative efficiency of 0.98. Compare to other cases above, this figure is low and thus $d_1$ is not equivalent to $d_2$. The other designs have much lower relative efficiencies as can be seen in the respective plot.

6.6 Designs to estimate $\eta$ in the \{1\}{3-stage(2Ci)} structure (with interaction)

The information matrices used in the following examples are presented in Chapter 5. In the algorithm, the variance ratios are obtained by first creating the grid of ratios as described in Section 6.5. Then, the corresponding values of each ratio are used to compute the interaction value.

In the present case, there are two variance ratios and their interaction accounting for a total of $r = 3$ variance components. Thus, a triangular plot can be used to present the results.

6.6.1 N=16 and MLE vs REML

The list of candidate designs to be compared appears in Table 6.1 on page 165.

Example 41 $D_V$-optimality

The MLE $D_V$-optimality plot can be found in Figure 6.145. The associated robustness plot is shown in Figure 6.146.

It can be seen that the three competing designs get to establish an optimality region. The symmetry line that was described previously for the triangular plots is also found here. Again, the line corresponds to the condition $\eta_1 = \eta_2$. There is a complementary effect.
6.6 Designs to estimate $\eta$ in the \{1\}-3-stage(2 Ci) structure (with interaction)

Figure 6.145 MLE $D_V$-plot for $\eta$ in the \{1\}-3-stage(2 Ci) structure with interaction.

Figure 6.146 MLE $D_V$-robustness plot for $\eta$ in the \{1\}-3-stage(2 Ci) structure with interaction.

Region 3 is very small. It only extends to values of $\eta_3 < 0.03$, almost equivalent of no having the interaction term. The robustness plot in Figure 6.146 is used and interpreted as in previous sections. The self efficiency of Region 1 and Region 2 drop significantly across the regions, but even though the relative efficiency of the complementary design is high, the optimum design is never outperformed. Region 3 shows the same self efficiency performance as the other regions, but the relative efficiencies are very low, which is not surprising.

The set of plots for the REML $D_V$-optimality can be seen in Figure 6.147 and
6.6 Designs to estimate $\eta$ in the \{1\}\{3-stage(2\text{Ci})\} structure (with interaction)

Figure 6.148, respectively. Note that the line of symmetry remain, but Region 3 is no longer present. Region 1 corresponds to the condition $\eta_1 > \eta_2$, and for Region 2, $\eta_2 > \eta_1$. Self efficiencies of Region 1 and Region 2 are higher than in for the MLE case since they do not go lower than 0.2. The relative efficiencies are almost the same as in the MLE case after considering the change of scale due to the absence of $d_3$.

Figure 6.147  REML $D_Y$- plot for $\eta$ in the \{1\}\{3-stage(2\text{Ci})\} structure with interaction.

Figure 6.148  REML $D_Y$- robustness plot for $\eta$ in the \{1\}\{3-stage(2\text{Ci})\} structure with interaction.
6.6 Designs to estimate $\eta$ in the \{1\}{3-stage(2Ci)} structure (with interaction)

The efficiencies in both methods of estimation are very similar. MLE $\mathcal{D}_V$-optimality appear to be more sensitive to small changes of the $\eta_1$ and $\eta_2$ values.

Example 42 $\mathcal{A}_V$-optimality

![MLE $\mathcal{A}_V$-optimality plot](image1)

**Figure 6.149** MLE $\mathcal{A}_V$-plot for $\eta$ in the \{1\}{3-stage(2Ci)} structure with interaction.

![MLE $\mathcal{A}_V$-robustness plot](image2)

**Figure 6.150** MLE $\mathcal{A}_V$-robustness plot for $\eta$ in the \{1\}{3-stage(2Ci)} structure with interaction.

The MLE $\mathcal{A}_V$-optimality plot is shown in Figure 6.149 and the robustness plot in Figure 6.150. The optimality plot is very similar to the REML $\mathcal{D}_V$-optimality
in Figure 6.147. In this case, design $d_3$ is optimum at some locations within the triangular plot. However, they are widely spread and do not really constitute a region of optimality. Efficiencies can still be computed for $d_3$ and they are shown in the robustness plot as usual. For Region 1 and Region 2, the interpretation is exactly the same as above for the REML $D_V$-optimality triangle.

In the robustness plot it is easy to see that Region 1 and Region 2 have very similar performance with Region 2 showing a slightly higher minimum self efficiency. As seen before, the relative efficiency of the complementary design is high. It is difficult to draw conclusions from the Region 3 panel. The self efficiency plot can suggest that the points are spread equidistantly from each other. Therefore, the self efficiency appears relatively high. The relative efficiency plot in turn, shows the effects of the spreading of points. At specific places where $d_3$ is optimum, the relative efficiency drops significantly, but then it becomes high again quickly. At some points, $d_2$ outperforms $d_3$. Most likely this happens in the middle of two equidistant points where $d_3$ is optimum.

For the REML $A_V$-optimality criterion the corresponding plots are shown in Figure 6.151 and Figure 6.152 for the optimality plot and the robustness plot,
6.6 Designs to estimate $\eta$ in the \{1\}-\{3-stage(2Ci)\} structure (with interaction)

Figure 6.152  REML $A_V$- robustness plot for $\eta$ in the \{1\}-\{3-stage(2Ci)\} structure with interaction.

respectively.

The MLE and REML $A_V$- plots are almost the same. In the REML plot, only Region 1 and Region 2 exist. The self efficiency is better for REML, and also the relative efficiencies are lower, which is desired.

6.6.2 N=16, MLE vs REML and pseudo-Bayesian optimality

Example 43  Total lack of knowledge for $\eta$

The MLE $D_V$- and $A_V$-optimality bar plots can be found next to each other in Figure 6.153. The corresponding relative efficiency plots are shown in Figure 6.154.

These plots are the same as those in Figure 6.105 on page 242 for the case without interaction. Naturally, the interpretation is the same as therein.

The REML optimality plots can be seen in Figure 6.155 and the efficiency plots in Figure 6.156 on page 271.

The two types of optimality criteria do not consider $d_3$. $D_V$-optimality roughly
6.6 Designs to estimate $\eta$ in the \{1\}{3-stage(2Ci)} structure (with interaction)

Figure 6.153 MLE $D_V$- and $A_V$-optimality plots when $\eta$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.

Figure 6.154 MLE $D_V$- and $A_V$-optimality efficiency plots when $\eta$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.

discriminate between $d_2$ and $d_1$ and chooses the former as the optimum. The $A_V$-criterion prefers $d_1$ as the optimum design, but $d_2$ is reported to be 0.9993 relative efficient. Thus, for the $A_V$-criterion, the two designs are equivalent.

Example 44 $\eta_1 > \eta_2$

When the selected method of estimation is MLE, Figure 6.157 presents the optimality plots with the relative efficiency plots being shown in Figure 6.158.

Under MLE and $D_V$-optimality, design $d_2$ is optimum. However, the relative
6.6 Designs to estimate $\eta$ in the \{1\}{3-stage(2Ci)} structure (with interaction)

Figure 6.155 REML $D_V$- and $A_V$-optimality plots when $\eta$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.

Figure 6.156 REML $D_V$- and $A_V$-optimality efficiency plots when $\eta$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.

efficiency of $d_1$ is 0.999 as can be seen in the relative efficiency plot. Therefore, the $D_V$- criterion does not discriminate confidently. In the case of $A_V$-optimality, design $d_1$ is optimum with $d_2$ having a relative efficiency of 0.98. This number not much encouraging than the 0.98 of the $D_V$- criterion. In neither case $d_3$ is competitive.

For the REML case, the optimality plots are given in Figure 6.159 whereas the relative efficiency plots are given in Figure 6.160 on page 273.

As in the previous example, under REML $d_3$ is not considered by neither
6 Design construction algorithms

6.6 Designs to estimate $\eta$ in the $\{1\}{\{3\text{-stage}(2Ci)\}}$ structure (with interaction)

It is clear that both optimality criteria provide the same ordering of the candidate designs, which is $(d_1, d_2)$. $D_V$-optimality gives $d_2$ a relative efficiency value of 0.972, whereas this is 0.93 for $A_V$-optimality which is slightly preferable.

6.6.3 $N=24$ and MLE vs REML

The list of candidates for $N = 24$ contains nine designs; these are presented in Table 6.2 on page 165.
6.6 Designs to estimate $\eta$ in the \{1\}3-stage(2Ci) structure (with interaction)

Example 45 $D_V$-optimality

For the MLE $D_V$-optimality criterion, optimality plots are given in Figure 6.161. The respective robustness plots can be seen in Figure 6.162 on page 275.

The symmetry line, that has been characteristic of MLE $D_V$-optimality in all examples, appears here and also the complementary effect. Additionally, in Figure 6.161 only the first seven of the nine candidates have an associated region of optimality, although the envelopes of Region 5 to Region 7 are not very well
The self efficiency in each region gets to be low mainly in Region 1 and Region 4. In these regions, the relative efficiency of $d_2$ and $d_1$, respectively, is very close to one across each region. These two facts suggest that $d_1$ and $d_4$ may not be reliable to use. Region 7 has the highest self efficiency, with $d_5$ and $d_6$ showing high relative efficiency. Given the low self efficiency in Region 5 and Region 6, $d_7$ may be a good choice when $\eta_3 < 0.09$. In Region 2 and Region 3, the corresponding complementary design has high relative efficiency across the region.

![Figure 6.161](image)

**Figure 6.161** MLE $D_V$-plot for $\eta$ in the $\{1\}$-3-stage($2Ci$) structure with interaction.

The REML $D_V$-optimality plot is shown in Figure 6.163 and the robustness plot in Figure 6.164. The REML case here and the MLE case above show a significantly different result. As before, there is a symmetry line and only $d_1$ to $d_4$ have a region of optimality. The self efficiency with REML is very similar in all regions. In relative efficiency, the neighbour design, as expected, has the highest value but is never better than the optimum design in the region.

The overall performance of REML $D_V$-optimality is more consistent and clearer than the MLE version, therefore, the former option is preferable.
6.6 Designs to estimate $\eta$ in the \{1\{3-stage(2Ci)\} structure (with interaction)

Figure 6.162  MLE $D_{V}$- robustness plot for $\eta$ in the \{1\{3-stage(2Ci)\} structure with interaction.

Example 46 $A_{V}$-optimality

The results for MLE $A_{V}$-optimality can be seen in Figure 6.165 as an optimality plot with the corresponding robustness plot in Figure 6.166 on page 277. The optimality plot shows the symmetry effect and the complementary effect. In this case five designs have a region of optimality. Compared to the MLE $D_{V}$- criterion above, the envelope for Region 5 is not well outlined. The self efficiency curves, reveal that the performance in Region 1 to Region 4 is almost the same. The relative efficiencies are high for the neighbour designs. In this case, note that
6.6 Designs to estimate $\eta$ in the \{1\}3-stage(2Ci) structure (with interaction)

Figure 6.163  REML $D_V$- plot for $\eta$ in the \{1\}3-stage(2Ci) structure with interaction.

Figure 6.164  REML $D_V$- robustness plot for $\eta$ in the \{1\}3-stage(2Ci) structure with interaction.

the relative efficiency of $d_5$ in these four regions is not drastically lower. This is relevant given that Region 5 is difficult to identify. Thus, given the competitive self efficiency of $d_5$, it is possible to discard it without significant loss.
6.6 Designs to estimate $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure (with interaction)

For REML $A_V$-optimality the optimality plots are presented in Figure 6.167, and the robustness plot is given in Figure 6.168 on the next page. When using REML, the optimality plot shows only four regions. In the previous example, the

![MLE $A_V$-plot for $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure with interaction.](image)

**Figure 6.165** MLE $A_V$-plot for $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure with interaction.

![MLE $A_V$-robustness plot for $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure with interaction.](image)

**Figure 6.166** MLE $A_V$-robustness plot for $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure with interaction.
6 Design construction algorithms

6.6 Designs to estimate $\eta$ in the $\{1\}{3\text{-stage}(2Ci)}$ structure (with interaction)

Remarks made when comparing the MLE and REML $D_V$-optimality plots are also applicable in the present case. The REML method is a more stable version of the
MLE one.

The self efficiencies noticeable larger for the REML $\mathcal{A}_V$- criterion than for the
MLE version above. The relative efficiencies in most regions are lower than in the
MLE case above also with larger differences among regions with the neighbour
designs having the largest values. 

6.6.4 $N=24$, MLE vs REML and pseudo-Bayesian

optimality

Example 47 Total lack of knowledge for $\eta$

The MLE $\mathcal{D}_V$- and $\mathcal{A}_V$-optimality plots are shown in Figure 6.169. The corre-
sponding relative efficiency plots are shown in Figure 6.170. In the MLE $\mathcal{D}_V$-
criterion seven designs can be compared, whereas for the $\mathcal{A}_V$- criterion only five
are considered. For $\mathcal{D}_V$-optimality design $d_7$ is optimum. It is followed by $d_6$
and $d_5$. For these, the relative efficiency is never lower than 0.99. The relative
efficiency of $d_3$ and $d_2$ is just above 0.98. Even for $d_4$ and $d_1$ this is never lower
than 0.925.
6.6 Designs to estimate $\eta$ in the \{1\}{3-stage(2Ci)} structure (with interaction)

In contrast, for $A_V$-optimality it is clearer that $d_2$ is optimum with $d_3$ being equivalent with a relative efficiency of almost one, but both are reasonably apart from the other three designs.

When considering REML, the respective optimality plots are given in Figure 6.171, whilst the relative efficiency plots appear in Figure 6.172.

![Figure 6.170](image1.png)

**Figure 6.170** MLE $D_V$- and $A_V$-optimality efficiency plots when $\eta$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.

![Figure 6.171](image2.png)

**Figure 6.171** REML $D_V$- and $A_V$-optimality plots when $\eta$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.

In this case, both criteria prefer only $d_1$ to $d_4$. However, only the pair $(d_2, d_3)$ is worth considering, given that the relative efficiencies of $d_1$ and $d_4$ are low. For the two criteria the pair $(d_2, d_3)$ is practically equivalent in terms of relative
efficiency. Consequently, the final conclusion is the same as in the MLE case above.

\[ \eta_1 > \eta_2 \]

The MLE \( D_V \) - and \( A_V \)-optimality plots can be seen in Figure 6.173. The plots in Figure 6.174 on the following page show the associated relative efficiencies.

In the \( D_V \)-optimality plot seven designs are being compared. Design \( d_2 \) is optimum, but \( d_3 \) has a relative efficiency of almost one. Thus, these two designs
Design construction algorithms

6.6 Designs to estimate $\eta$ in the $\{1\}3$-stage(2Ci) structure (with interaction)

Figure 6.174  MLE $D_V$- and $A_V$-optimality efficiency plots when $\eta_1 > \eta_2$ in the $\{1\}3$-stage(2Ci) structure with interaction.

can be considered equivalent. The other four designs are not competitive based on the relative efficiency. In the case of the $A_V$- criterion, five designs are compared and $d_2$ is the optimum. However, $d_3$ and $d_1$ have virtually the same relative efficiency of 0.98. This can be seen as the triplet $(d_2, d_3, d_1)$ being equivalent. The other two designs are far from the optimum.

The REML optimality results for the $D_V$- and $A_V$-criteria are shown in Figure 6.175 with the corresponding efficiency plots given in Figure 6.176. Both criteria under REML consider only designs $d_1$ to $d_4$. In the two cases, the ordering
6.6 Designs to estimate $\eta$ in the $\{1\}{3\text{-stage}(2Ci)}$ structure (with interaction)

Figure 6.176  REML $D_V$- and $A_V$-optimality efficiency plots when $\eta_1 > \eta_2$ in the $\{1\}{3\text{-stage}(2Ci)}$ structure with interaction.

of designs is the same i.e. $(d_2, d_3, d_1, d_4)$. Similarly, the relative efficiency of $d_3$ is 0.97 for the $A_V$- criterion and 0.97 for the $D_V$- criterion. Design $d_1$ has higher relative efficiency in the $A_V$- case than for $D_V$-optimality. However, this value is comparatively far from $d_3$. The two criteria give almost equal results.  

6.6.5  N=32 and MLE vs REML

The candidate list of designs to be considered when the total number of observations is $N = 32$ has six different designs. It is shown in Table 6.3 on page 165.

Example 49 $D_V$-optimality

The MLE $D_V$-optimality plot is given in Figure 6.177 with the robustness plots being shown in Figure 6.178. Five out of the six candidate designs are being compared in this case. As was seen in the MLE $D_V$-optimality case of Section 6.6.3, not all the designs have a well defined region. In the present case, those are $d_4$ and $d_5$. Based on the better self efficiency of $d_4$ and the high relative efficiency of $d_5$ in Region 4, $d_5$ may be discarded and use $d_4$ instead. For designs $(d_1, d_2, d_3)$, the self efficiency drops noticeably, but in terms of relative efficiencies, no other
6.6 Designs to estimate $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure (with interaction)

Figure 6.177  MLE $D_V$- plot for $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure with interaction.

Figure 6.178  MLE $D_V$- robustness plot for $\eta$ in the $\{1\}\{3\text{-stage}(2Ci)\}$ structure with interaction.

design is better in their respective regions of optimality.

The REML $D_V$-optimality plot is shown in Figure 6.179. Its corresponding
robustness plot is shown in Figure 6.180. In this case only $d_1$ to $d_3$ are being compared. The self efficiencies are higher than in the MLE case above. Also, the relative efficiencies are noticeably lower in each region, which is desirable. As it is expected, in each region the highest relative efficiency is given by the neighbour design.

Figure 6.179  REML $D_V$- plot for $\eta$ in the \{1\}3-stage(2Ci) structure with interaction.

Figure 6.180  MLE $D_V$- robustness plot for $\eta$ in the \{1\}3-stage(2Ci) structure with interaction.
6.6 Designs to estimate \( \eta \) in the \( \{1\}\{3\text{-stage}(2C_i)\} \) structure (with interaction)

Example 50 \( \mathcal{A}_V \)-optimality

The MLE \( \mathcal{A}_V \)-optimality plot appears in Figure 6.181. The associated robustness plot is shown in Figure 6.182 on the following page.

![MLE \( \mathcal{A}_V \)-optimality plot](image)

Figure 6.181 MLE \( \mathcal{A}_V \)-plot for \( \eta \) in the \( \{1\}\{3\text{-stage}(2C_i)\} \) structure with interaction.

The triangular plot shows four designs. Design \( d_4 \) does not have a defined region of optimality. As seen before, the points at which \( d_4 \) is optimum are widely spread. Based on efficiencies, there is not enough evidence to not consider this design. In Region 1, Region 2 and Region 3, the expected behaviour is observed.

For the REML \( \mathcal{A}_V \)-optimality plots refer to Figure 6.183, and to Figure 6.184 for the corresponding robustness plot. The same differences between the MLE and the REML methods as those seen in the \( \mathcal{D}_V \)-optimality just above are directly applicable here.

6.6.6 \( N=32 \), MLE vs REML and pseudo-Bayesian optimality

Example 51 Total lack of knowledge for \( \eta \)

The optimality plots resulting from using the MLE information matrix can be found
6.6 Designs to estimate $\eta$ in the \( \{1\}\{3\text{-stage}(2Ci)\} \) structure (with interaction)

Figure 6.182 MLE $A_V$-robustness plot for $\eta$ in the \( \{1\}\{3\text{-stage}(2Ci)\} \) structure with interaction.

Figure 6.183 REML $A_V$-plot for $\eta$ in the \( \{1\}\{3\text{-stage}(2Ci)\} \) structure with interaction.

in Figure 6.185 and the relative efficiency plot in Figure 6.186. By inspecting the set of plots it is clear that both criteria identify design $d_2$ as the optimum. For $A_V$-optimality, designs $(d_1, d_3)$ are equivalent in terms of relative efficiency, but in
6.6 Designs to estimate $\eta$ in the \{1\}{3-stage(2Ci)} structure (with interaction)

this case they are far from the optimum. Under $D_V$-optimality, the same happens with designs $(d_3, d_4)$. Other designs are not competitive.

The optimality plots for the REML $D_V$- and $A_V$- criteria are found in Figure 6.187 and the corresponding relative efficiency plots are given in Figure 6.188. Both optimality plots reveal that the two optimality criteria give the design ordering $(d_2, d_3, d_1)$. The relative efficiency plots verify that design $d_2$ is clearly the optimum.
6.6 Designs to estimate \( \eta \) in the \( \{1\}\{3\text{-stage}(2Ci)\} \) structure (with interaction)

The optimality plots for MLE \( D_V \)- and \( A_V \)-optimality can be seen in Figure 6.189 and the relative efficiency plots in Figure 6.190.

As in the previous example, both optimality plots reveal that the two optimality criteria give the design ordering \((d_2, d_3, d_1)\). For \( D_V \)-optimality, \( d_2 \) is clearly the optimum. Under \( A_V \)-optimality, \( d_1 \) is 0.98 relative efficient, thus they may be considered equivalent.

The REML optimality plots are shown in Figure 6.191 with the associated
6 Design construction algorithms

6.6 Designs to estimate $\eta$ in the \{1\}{3-stage(2Ci)} structure (with interaction)

![Graph](image1)

**Figure 6.188** REML $D_V$- and $A_V$-optimality efficiency plots when $\eta$ is unknown in the \{1\}{3-stage(2Ci)} structure with interaction.

![Graph](image2)

**Figure 6.189** MLE $D_V$- and $A_V$-optimality plots when $\eta_1 > \eta_2$ in the \{1\}{3-stage(2Ci)} structure with interaction.

relative efficiency plots in Figure 6.192 on page 292.

For $D_V$-optimality, design $d_2$ is optimum with $d_1$ and $d_3$ far from optimal having relative efficiencies of 0.87 and 0.84, respectively. Under $A_V$-optimality $d_2$ is also optimum with $d_1$ having relative efficiency of 0.93. Compared to other cases above, this figure is low and thus $d_1$ is not equivalent to $d_2$.

For the given restriction on $\eta$, in this case $D_V$-optimality performs well and consistently.
6.7 The \{1\}\{3-stage(2N)\} structure to estimate $\sigma^2$

The information matrix for the case were the random factors are nested and the interest is in estimating $\sigma^2$, are given in Chapter 5. In this case, there are $r = 3$ variance components. Therefore, the results can be presented using triangular plots for $D_V$- and $A_V$-optimality.
6 Design construction algorithms 6.7 The \{1\}{3-stage(2N)} structure to estimate $\sigma^2$

Figure 6.192  REML $\mathcal{D}_V$- and $\mathcal{A}_V$-optimality efficiency plots when $\eta_1 > \eta_2$ in the \{1\}{3-stage(2Ci)} structure with interaction.

6.7.1 N=16 and MLE vs REML

The list of candidate designs to be compared appears in Table 6.1 on page 165.

Example 53 $\mathcal{D}_V$-optimality

The MLE $\mathcal{D}_V$-optimality plot can be found in Figure 6.193. The associated robustness plot is shown in Figure 6.194.

The symmetry that could be appreciated in the crossed cases before is not present in the nested case shown here. The nested structure by definition gives different weights to the different factors. In other words, by construction there is an asymmetry of the relative contributions of the factors to the variability of the design.

In addition to the loss of the symmetry line, the complementary and mirror effects seen before do not appear in the nested cases. Again, this is expected as mentioned in the above paragraph. Apart from these observations, the plots can be used in an identical manner.

From the triangle plot, it can be seen that the three competing designs get to establish an optimality region. Region 2 covers most of the area of the triangle. In general, it represents large values of $\sigma^2$. The small values are covered by Region 3.
Region 2 corresponds to large values of $\sigma_2^2$. No designs correspond exclusively to specific values of $\sigma_2^2$.

![MLE $D_V$- plot for $\sigma^2$ in the $\{1\}$\{3-stage(2N)\} structure.](image)

**Figure 6.193** MLE $D_V$- plot for $\sigma^2$ in the $\{1\}$\{3-stage(2N)\} structure.

The self efficiency curves are similar to those of the crossed cases reviewed before. In the present case, Region 2 has a noticeably higher values than the other two. All relative efficiencies are poor.

The set of plots for the REML $D_V$-optimality can be seen in Figure 6.195 and Figure 6.196, respectively. The REML triangle is very similar to the MLE triangle above. The difference is the proportional area or each optimality region. In the REML case, Region 1 is given preference. However, the location say, of the
regions is the same, only their size changed. One obvious consequence of this is, for example, that Region 3 is valid only for values of $\sigma^2 > 0.15$.

The self efficiencies are also very similar to the MLE version, with the difference that it is noticeably higher in Region 3. The relative efficiencies are also low.

Figure 6.195  REML $D_{Y^*}$ plot for $\sigma^2$ in the $\{1\}$-$\{3\}$-$\{3\}$-$\{2N\}$ structure.

Figure 6.196  REML $D_{Y^*}$ robustness plot for $\sigma^2$ in the $\{1\}$-$\{3\}$-$\{3\}$-$\{2N\}$ structure.
Example 54 $A_V$-optimality

The $A_V$-optimality plot is shown in Figure 6.197 and the robustness plot in Figure 6.198.

The MLE $A_V$-optimality plot is shown in Figure 6.197 and the robustness plot in Figure 6.198.

The optimality plot is very similar to the MLE $D_V$-optimality in Figure 6.195. In this case, design $d_3$ only covers values of $\sigma^2 > 0.58$. Preference is given to $d_2$ also in the present case.

In the robustness plot the differences are more pronounced. The self efficiencies are significantly higher, with the minimum at around 0.96. The relative efficiencies
are low in general in all the regions.

![A-optimality: reml](image)

**Figure 6.199** REML $A_V$-plot for $\sigma^2$ in the $\{1\{3-stage(2N)\}$ structure.

![A-optimality](image)

**Figure 6.200** REML $A_V$-robustness plot for $\sigma^2$ in the $\{1\{3-stage(2N)\}$ structure.

For the REML $A_V$-optimality criterion the corresponding plots are shown in Figure 6.199 and Figure 6.200 for the optimality plot and the robustness plot, respectively.

The MLE and REML $A_V$-plots are very similar. In the REML plot, Region 3 starts from $\sigma_i^2 > 0.8$. Region 1 is accordingly larger. The self efficiencies are almost the same as for the MLE case. The relative efficiencies are kept low.

Given the close similarities between the two criteria and methods of estimation, a general recommendation can not be given. A decision has to be made based on
the specific situation at hand.

6.7.2 \( N = 16 \), MLE vs REML and pseudo-Bayesian optimality

Example 55 Total lack of knowledge for \( \sigma^2 \)

The MLE \( D_V \)- and \( A_V \)-optimality bar plots can be found next to each other in Figure 6.201. The corresponding relative efficiency plots are shown in Figure 6.202.

The two criteria choose the design ordering \((d_2, d_1, d_3)\). In both cases, the relative efficiency of \(d_1\) with respect to \(d_2\) is very high, being no lower than 0.98. Thus, \(d_2\) and \(d_1\) can be seen as equivalent. Design \(d_3\) is far from optimum in either criteria.

The REML optimality plots can be seen in Figure 6.203 and the efficiency plots in Figure 6.204. The REML \( D_V \)-optimality plot roughly discriminate between \(d_2\) and \(d_1\) and chooses the former as the optimum. The \( A_V \)- criterion prefers \(d_2\) as the optimum design, but \(d_1\) is reported to be 0.98 relative efficient. Thus, for the \( A_V \)- criterion, the two designs are also equivalent.

\[ \square \]
Example 56 \( \sigma_1^2 > \sigma_2^2 > \sigma_i^2 \)

When the selected method of estimation is MLE, Figure 6.205 presents the optimality plots with the relative efficiency plots being shown in Figure 6.206.

For both criteria design \( d_1 \) is optimum. However, in this case, \( D_- \)-optimality cannot really distinguish between \( d_1 \) and \( d_2 \) with the latter having a relative efficiency greater than 0.99. In the case of \( A_- \)-optimality, \( d_1 \) is definitely the optimum.

For the REML case, the optimality plots are given in Figure 6.207 whereas
Figure 6.204  REML $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the $\{1\}$($3$-stage($2N$)) structure.

Figure 6.205  MLE $D_V$- and $A_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the $\{1\}$($3$-stage($2N$)) structure.

the relative efficiency plots are given in Figure 6.208 on page 301.

As in the previous example, under REML $d_1$ is the optimum design with $d_2$ having a low relative efficiency of 0.77. $D_V$-optimality still is not able to confidently descriminate between the three designs.

6.7.3  N=24 and MLE vs REML

The list of candidates for $N = 24$ contains nine designs; these are presented in Table 6.2 on page 165.
6 Design construction algorithms

6.7 The \( \{1\}{\{3}\text{-stage}(2N)} \) structure to estimate \( \sigma^2 \)

![D-optimality](image1)

**Figure 6.206** MLE \( D_V \)- and \( A_V \)-optimality efficiency plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}{\{3}\text{-stage}(2N)} \) structure.

![A-optimality](image2)

**Figure 6.207** REML \( D_V \)- and \( A_V \)-optimality plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}{\{3}\text{-stage}(2N)} \) structure.

**Example 57** \( D_V \)-optimality

For the MLE \( D_V \)-optimality criterion, optimality plots are given in Figure 6.209. The respective robustness plots can be seen in Figure 6.210 on page 302.

The nine designs from the candidate list can establish a region of optimality. Regions 1 to 4 are the largest. Regions 5 to 9 are mainly sensitive to the values of \( \sigma^2 \). Region 5 and Region 7 mainly cover small values and Region 6 and Region 8 cover large values. Region 9 covers all the range of \( \sigma^2 \).
6 Design construction algorithms 6.7 The \{1\}\{3-stage(2N)\} structure to estimate \(\sigma^2\)

\[\sigma_1^2 > \sigma_2^2 > \sigma_3^2\] in the \{1\}\{3-stage(2N)\} structure.

Figure 6.208  REML \(D_V\) and \(A_V\)-optimality efficiency plots when \(\sigma_1^2 > \sigma_2^2 > \sigma_3^2\) in the \{1\}\{3-stage(2N)\} structure.

The robustness plots are similar to some of the ones presented for the crossed model. The main difference here is the absence of the mirror and complementary effect. The neighbour effect is still seen here in some cases. In Figure 6.210 the optimum design is never outperformed in its region.

Figure 6.209  MLE \(D_V\) plot for \(\sigma^2\) in the \{1\}\{3-stage(2N)\} structure.

The REML \(D_V\)-optimality plot is shown in Figure 6.211 and the robustness plot in Figure 6.212.

The REML case here in comparison to the MLE version above show less sensitivity for smaller values of \(\sigma_1^2\). Region 7 and Region 6 are larger under REML. Consequently, Region 6 and Region 8 are smaller. Also, Region 4 was reduced. In
6 Design construction algorithms 6.7 The \( \{1\}\{3\text{-stage}(2N)\} \) structure to estimate \( \sigma^2 \)

**Figure 6.210** MLE \( D_V \)-robustness plot for \( \sigma^2 \) in the \( \{1\}\{3\text{-stage}(2N)\} \) structure.

**Figure 6.211** REML \( D_V \)-plot for \( \sigma^2 \) in the \( \{1\}\{3\text{-stage}(2N)\} \) structure.
terms of efficiencies, the same remarks as in the MLE case above can be made here.

As it was noticed before, it is not possible to recommend a criterion and method of estimation in this case. The decision has to be made considering the specifics of the experimental situation.

**Example 58 $\mathcal{A}_V$-optimality**

The results for MLE $\mathcal{A}_V$-optimality can be seen in Figure 6.213 as an optimality
plot with the corresponding robustness plot in Figure 6.214 on the next page.

![MLE A-plot for σ² in the {1}{3-stage(2N)} structure.](image)

**Figure 6.213** MLE A-plot for σ² in the {1}{3-stage(2N)} structure.

The optimality plot do not present any specific feature that could be described in detail. On the other hand, the robustness plots compared to the D-optimality version, show a much higher and stable self efficiency in all regions. The relative efficiencies low in most cases. Except possibly in Region 2 and Region 3, where the complementary design has a relative efficiency of almost one across all the region.

For REML A-optimality the optimality plots are presented in Figure 6.215, and the robustness plot is given in Figure 6.216 on page 307.

### 6.7.4 N=24, MLE vs REML and pseudo-Bayesian optimality

**Example 59** Total lack of knowledge for σ²

The MLE D- and A-optimality plots are shown in Figure 6.217. The corresponding relative efficiency plots are shown in Figure 6.218 on page 308.

In the MLE D- criterion design d₃ is optimum, but it is not possible to discriminate among the nine designs. The relative efficiency plot shows that all efficiencies are very close to unity, even for the “worst” design.
In contrast, for $\mathcal{A}_V$-optimality it is clearer that $d_2$ is optimum with $d_3$ and perhaps $d_1$ being equivalent with a relative efficiency of almost one. However, all other designs are clearly away from these three.

When considering REML, the respective optimality plots are given in Figure 6.219, whilst the relative efficiency plots appear in Figure 6.220. Under REML, the impossibility of the $\mathcal{D}_V$- criterion to discriminate the designs remains. For $\mathcal{A}_V$-optimality $d_1$ is optimum with $d_2$ being equivalent in terms of relative efficiency. The other designs are not competitive in comparison.
6 Design construction algorithms

6.7 The \{1\}\{3-stage(2N)\} structure to estimate \(\sigma^2\)

These examples showed that the \(\mathcal{D}_V\)-criterion is not capable of effectively discriminate the designs.

**Example 60** \(\sigma_1^2 > \sigma_2^2 > \sigma_t^2\)

The MLE \(\mathcal{D}_V\) and \(\mathcal{A}_V\)-optimality plots can be seen in Figure 6.221. The plots in Figure 6.222 on page 310 show the associated relative efficiencies.

The REML optimality results for the \(\mathcal{D}_V\)- and \(\mathcal{A}_V\)-criteria are shown in Figure 6.223 on page 311 with the corresponding efficiency plots given in Figure 6.224.

The interpretation of the results for this case when \(\sigma_1^2 > \sigma_2^2 > \sigma_t^2\) is the same as that of the previous example when \(\sigma^2\) is unknown.

**6.7.5 N=32 and MLE vs REML**

The candidate list of designs to be considered when the total number of observations is \(N = 32\) has six different designs. It is shown in Table 6.3 on page 165.

**Example 61** \(\mathcal{D}_V\)-optimality

The MLE \(\mathcal{D}_V\)-optimality plot is given in Figure 6.225 with the robustness plots...
The six candidate designs have a region of optimality in this case. The overall arrangement of the region in the triangle is virtually the same as in the Section 6.7.3. The characteristics of the robustness plots are also the same as therein.

The REML $\mathcal{D}_V$-optimality plot is shown in Figure 6.227. Its corresponding robustness plot is shown in Figure 6.228 on page 313.

In this case, both methods of estimation seem to give equally good results.
6 Design construction algorithms 6.7 The \(\{1\}\{3\text{-stage}(2N)\}\) structure to estimate \(\sigma^2\)

![Figure 6.217](image1)

**Figure 6.217** MLE \(D_V\) and \(A_V\)-optimality plots for when \(\sigma^2\) is unknown in the \(\{1\}\{3\text{-stage}(2N)\}\) structure.

![Figure 6.218](image2)

**Figure 6.218** MLE \(D_V\) and \(A_V\)-optimality efficiency plots when \(\sigma^2\) is unknown in the \(\{1\}\{3\text{-stage}(2N)\}\) structure.

**Example 62** \(A_V\)-optimality

The MLE \(A_V\)-optimality plot appears in Figure 6.229. The associated robustness plot is shown in Figure 6.230 on page 314.

The triangular plot shows four designs. Design \(d_4\) does not have a defined region of optimality. As seen before, the points at which \(d_4\) is optimum are widely spread. Based on efficiencies, there is not enough evidence to not consider this design. In Region 1 to Region 3, the expected behaviour is observed.

For the REML \(A_V\)-optimality plots refer to Figure 6.231, and to Figure 6.232 on page 314.
Figure 6.219  REML $D_V$- and $A_V$-optimality plots when $\sigma^2$ is unknown in the \{1\}{3-stage(2N)} structure.

Figure 6.220  REML $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the \{1\}{3-stage(2N)} structure.

for the corresponding robustness plot.

The same differences between the MLE ad the REML methods as those seen in the $D_V$-optimality just above are directly applicable here.
6 Design construction algorithms 6.7 The \( \{1\}\{3\text{-stage}(2N)\} \) structure to estimate \( \sigma^2 \)

![MLE D- and A-optimality plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}\{3\text{-stage}(2N)\} \) structure.](image1)

![MLE D- and A-optimality efficiency plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}\{3\text{-stage}(2N)\} \) structure.](image2)

**Figure 6.221** MLE \( D \)- and \( A \)-optimality plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}\{3\text{-stage}(2N)\} \) structure.

**Figure 6.222** MLE \( D \)- and \( A \)-optimality efficiency plots when \( \sigma_1^2 > \sigma_2^2 > \sigma_3^2 \) in the \( \{1\}\{3\text{-stage}(2N)\} \) structure.

6.7.6 \( n=32 \), MLE vs REML and pseudo-Bayesian optimality

**Example 63** Total lack of knowledge for \( \sigma^2 \)

The optimality plots resulting from using the MLE information matrix can be found in Figure 6.233 and the relative efficiency plot in Figure 6.234.

By inspecting the set of plots it is clear that both criteria identify design \( d_2 \) as the optimum. However, for \( D \)-optimality the relative efficiency plots show
6 Design construction algorithms

6.7 The \( \{1\}\{3\text{-stage}(2N)\} \) structure to estimate \( \sigma^2 \)

that this criterion cannot reliably single out the designs. This behaviour has been described in previous examples. The behaviour exhibited by the \( A_V \)- criterion has also been seen before. In this case there is good discrimination of the majority of the designs, although \( d_2 \) has a high relative efficiency of almost unity. Therefore, designs \((d_1, d_2)\) are equivalent in terms of relative efficiency.

The optimality plots for the REML \( D_V \)- and \( A_V \)- criteria are found in Figure 6.235 and the corresponding relative efficiency plots are given in Figure 6.236.

In this instance, there is no difference between the description of the plots here
6 Design construction algorithms 6.7 The \( \{1\}\{3\text{-stage}(2N)\} \) structure to estimate \( \sigma^2 \)

and those for the MLE case above. If any difference exists, it should be that for \( \mathcal{A}_V \)-optimality the optimum design is \( d_1 \), but \( d_2 \) still has high relative efficiency. Thus, the design pair \( (d_1, d_2) \) remains equivalent.
6.7 The \{1\}\{3-stage(2N)\} structure to estimate $\sigma^2$

Figure 6.227  REML $D_V$- plot for $\sigma^2$ in the \{1\}\{3-stage(2N)\} structure.

Figure 6.228  REML $D_V$- robustness plot for $\sigma^2$ in the \{1\}\{3-stage(2N)\} structure.

Example 64  $\sigma^2_1 > \sigma^2_2 > \sigma^2_\epsilon$

The optimality plots for MLE $D_V$- and $A_V$-optimality can be seen in Figure 6.237 and the relative efficiency plots in Figure 6.238. Both optimality plots reveal
that the two optimality criteria have design $d_1$ as optimum, followed in relative efficiency by $d_2$. However, as in the previous example, the same phenomenon happens with $D_V$-optimality. It cannot reliably distinguished between designs.
6 Design construction algorithms

6.7 The $\{1\}$\{3-stage(2N)\} structure to estimate $\sigma^2$

Figure 6.231 REML $A_V$- plot for $\sigma^2$ in the $\{1\}$\{3-stage(2N)\} structure.

Also, the same type of behaviour is repeated by $A_V$-optimality. However, in this example $d_1$ is not equalled in relative efficiency by any other design.

The REML optimality plots are shown in Figure 6.239 with the associated relative efficiency plots in Figure 6.240 on page 319.
6 Design construction algorithms 6.7 The \{1\}{\text{3-stage(2N)}}\} structure to estimate $\sigma^2$

Figure 6.233  MLE $D_V$- and $A_V$-optimality plots when $\sigma^2$ is unknown in the \{1\}{\text{3-stage(2N)}}\} structure.

Figure 6.234  MLE $D_V$- and $A_V$-optimality efficiency plots when $\sigma^2$ is unknown in the \{1\}{\text{3-stage(2N)}}\} structure.

As in the previous example for the condition of lack of knowledge on $\sigma^2$, in the present case, there is no difference between the description of the plots here and those for the MLE case above. The conclusions drawn are the same.  ■
6.8 Optimum designs using the Dispersion-Mean model

The dispersion-mean model was described in Section 3.5. It transforms the linear mixed model so that the variability structure can be expressed as a linear combination of the variance components. Using the least squares principle, this model can be used to find optimum designs for the variance components. Using the same model formulation, designs for different methods of estimation can be
Figure 6.237 MLE $D_V$- and $A_V$-optimality plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the \{1\}\{3-stage(2N)\} structure.

Figure 6.238 MLE $D_V$- and $A_V$-optimality efficiency plots when $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$ in the \{1\}\{3-stage(2N)\} structure.

obtained. These equivalencies are described \textit{e.g.} in Malley (1986) and Searle et al. (1992).

\subsection{6.8.1 Ordinary least squares}

Using OLS to estimate the dispersion-mean model yields MINQUEO estimates. These correspond to a model that only contains the random error with a preliminary value of one, \textit{i.e.} $\sigma_0^2 = 1$ and $\sigma_i^2 = 0$.

This simplified modelling is a straightforward approach to estimate the random
error with minimum variance and unbiasedly (Searle et al., 1992). Several different simplifications like the above, i.e. giving a priori values to subsets of variance components, are described by (Rao, 1997) as different forms of MINQUE and MIVQUE estimators.

The OLS information matrix is of the form

\[ M_y = X^T X, \] 

(6.4)
6 Design construction algorithms  6.8 Optimum designs using the Dispersion-Mean model

as was explained in Chapter 3. Using \( \mathbf{X} = \mathbf{1}_N \), \( \mathbf{X} \) is constructed according to equation (3.52).

The optimum designs found using this model specification and estimation method, are those that put the highest possible experimental effort at the observational stage. In other words, the highest number of replicates of observational treatments is required. This is expected since the only variance component involved is the random error. However, this gives the opportunity to evaluate how the efficiency of a design can be negatively affected. The ideas are presented by constructing designs for different possible experimental conditions.

**Designs for \( N = 12 \)**

Computed optimum designs for different number of observations are presented next. The candidate design list for \( N = 12 \) is presented in Table 6.4. Designs for the same models as used in previous sections are evaluated here.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

The list consists of only three designs. Recall that the designs are balanced. However, in this circumstance, it is reasonable to think of the symmetry, in terms of the weight, given to the different stages. For example, \( d_3 \) shows symmetry in the first and second variability stages, whereas \( d_2 \) is symmetric in the second and third stages. In this context, \( d_2 \) shows no symmetry.

The following plots show the relative efficiency of the candidate designs. The \( D_V \)- and \( A_V \)-optimality criteria are shown. The models correspond to the structures \{1\}{3-stage(2C)}, \{1\}{3-stage(2Ci)} and \{1\}{3-stage(2N)} to estimate \( \sigma^2 \).

320
Example 65 The \{1\}\{3-stage(2C)\} structure

The relative efficiency plots for the \{1\}\{3-stage(2C)\} structure are shown in Figure 6.241, with $D_V$-optimality on the left and $A_V$-optimality on the right.

![Relative efficiency plots for $D_V$- and $A_V$-optimality using the Dispersion-mean OLS method with $N = 12$ in the \{1\}\{3-stage(2C)\} structure.](image)

The two plots show the same pattern. Designs $d_1$ and $d_2$ are complementary. This reflects the symmetry between the first and second variability stages, \textit{i.e.} 3 and 2, respectively. Therefore, they show the same relative efficiency with respect to $d_3$ which is the optimum. Design $d_3$ puts most of the experimental resources in the observational stage. This is expected. The same behaviour is observed in both optimality criteria. The efficiency loss is around 22\% for $D_V$-optimality and 16\% for $A_V$-optimality.

The interpretation is regarding much is lost when miss specifying the initial values of the variance components. As expected, it is the difference between the variances (relative size) of the two stages that determine the efficiency loss rather than the specific value assigned to the variance components. Given that $d_1$ and $d_2$ are complementary, this difference is the same for the two designs and therefore the efficiency plot shows a line of the same height for the two designs. The simplicity of this example helps understand the idea behind the comparisons.
Example 66 The \{1\}\{3-stage(2Ci)\} structure

The relative efficiency plots for the crossed model with interaction, denoted by \{1\}\{3-stage(2Ci)\}, are shown in Figure 6.242. The pattern is the same as in the crossed case in Figure 6.241. However, the efficiency loss is higher here in the interaction case. For $D_V$-optimality is 41% and 20% for $A_V$-optimality. This suggests that when there is interaction the sensitivity to the preliminary values is much higher. It seems critical for $D_V$-optimality since the loss is doubled with respect to the case without interaction.

Example 67 The \{1\}\{3-stage(2N)\} structure

The relative efficiency plots for the nested model, denoted by \{1\}\{3-stage(2N)\}, are shown in Figure 6.243 on the following page.

Unlike the previous examples, in this case the efficiency loss is different for $d_1$ and $d_2$. Accordingly, there is a much higher loss when miss specifying the value of $\sigma_1^2$. The loss for poor specification of $\sigma_2^2$ is similar to that in the crossed structure with interaction.

It is possible to compare the efficiency of contiguous symmetric designs to calculate the *efficiency loss factor*. For the nested case in this example, under


Figure 6.243 Relative efficiency plots for $D_V$- and $A_V$-optimality using the Dispersion-mean OLS method with $N = 12$ in the $\{1\}{3\text{-stage}(2C)}$ structure.

$D_V$-optimality, the respective loss between $d_1$ and $d_2$ is around 1.5 times higher for $d_1$. In other words, the efficiency loss factor from $d_2$ to $d_1$, represented as $d_2 \rightarrow d_1$, is 1.5. For the criterion of $A_V$-optimality, the efficiency loss factor between the two designs is 1.15. As in the previous examples, $D_V$-optimality is more sensitive to the misspecification of the variance components values.

**Designs for $N = 24$**

The candidate design list for $N = 24$ has nine entries and can be consulted in Table 6.2 on page 165. In this case the pairs $(d_1, d_4)$, $(d_2, d_3)$, $(d_5, d_6)$ and $(d_7, d_8)$ are symmetric in the first and second variability stages. Design $d_9$ is the optimum with the highest experimental weight in the observational stage.

**Example 68** The $\{1\}{3\text{-stage}(2C)}$ and $\{1\}{3\text{-stage}(2Ci)}$ structures

The relative efficiency plots for the crossed model, without and with interaction, are shown in Figure 6.244 and Figure 6.245, respectively.

The pattern of the two plots in each of the figures is the same. The range of the vertical axis is the main difference. For $D_V$-optimality, the minimum efficiency value is 0.45 in the crossed case, and 0.20 when interaction is present. The
corresponding values for $A_V$-optimality are 0.76 and 0.56. Clearly, they are much lower with the $D_V$-criterion.

Additionally, the pairs of contiguous designs with the same symmetry that yield the same relative efficiency values are easily identified. Therefore, the efficiency loss factor can be found pairwise. These results are presented in Table 6.5 on the following page for the two structures in this example.

**Example 69** The $\{1\}\{3\text{-stage}(2N)\}$ structure

The relative efficiency plots for the nested variability structure are shown in Figure 6.246. In this case, under both criteria of optimality, neither design has
The symmetry of the designs does not appear to play a role, unlike in the crossed structures. The corresponding efficiency loss factors are reported in Table 6.6 on the next page.

### Table 6.5  Efficiency loss factor between symmetric design pairs with $N = 24$ in the $\{1\}\{3$-stage$(2C)\}$ and $\{1\}\{3$-stage$(2Ci)\}$ structures

<table>
<thead>
<tr>
<th>Pairs</th>
<th>$D_V$-</th>
<th>$A_V$-</th>
<th>$D_V$-</th>
<th>$A_V$-</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(d_2, d_3) \rightarrow (d_1, d_4)$</td>
<td>1.27</td>
<td>1.098</td>
<td>1.44</td>
<td>1.10</td>
</tr>
<tr>
<td>$(d_5, d_6) \rightarrow (d_2, d_3)$</td>
<td>1.19</td>
<td>1.062</td>
<td>1.67</td>
<td>1.27</td>
</tr>
<tr>
<td>$(d_7, d_8) \rightarrow (d_5, d_6)$</td>
<td>1.25</td>
<td>1.073</td>
<td>1.58</td>
<td>1.15</td>
</tr>
</tbody>
</table>

**Figure 6.246** Relative efficiency plots for $D_V$- and $A_V$-optimality using the Dispersion-mean OLS method with $N = 24$ in the $\{1\}\{3$-stage$(2N)\}$ structure.

Designs for $N = 48$

An 18 elements list of candidate designs is generated for the models being reviewed when $N = 48$. The designs are shown in Table 6.7.

In this case, there are eight symmetrical pairs of designs in the first and second variability stages. The range of experimental weight varies from two to twelve. It
Table 6.6  Efficiency loss factor between design with $N = 24$ in the 
\$\{1\}\{3\text{-stage}(2N)\}$ structure

<table>
<thead>
<tr>
<th>Efficiency loss factor</th>
<th>$D_V$-optimality</th>
<th>$A_V$-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair Factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_2 \rightarrow d_1$</td>
<td>1.8</td>
<td>1.14</td>
</tr>
<tr>
<td>$d_3 \rightarrow d_2$</td>
<td>1.33</td>
<td>1.05</td>
</tr>
<tr>
<td>$d_5 \rightarrow d_3$</td>
<td>1.12</td>
<td>1.03</td>
</tr>
<tr>
<td>$d_4 \rightarrow d_5$</td>
<td>1.11</td>
<td>1.20</td>
</tr>
<tr>
<td>$d_7 \rightarrow d_4$</td>
<td>1.6</td>
<td>1.18</td>
</tr>
<tr>
<td>$d_6 \rightarrow d_7$</td>
<td>1.13</td>
<td>1.03</td>
</tr>
<tr>
<td>$d_8 \rightarrow d_6$</td>
<td>1.33</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Table 6.7  Candidate designs list generated by the algorithm for $N = 48$

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_5$</th>
<th>$d_6$</th>
<th>$d_7$</th>
<th>$d_8$</th>
<th>$d_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>12</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$d_{10}$</td>
<td>$d_{11}$</td>
<td>$d_{12}$</td>
<td>$d_{13}$</td>
<td>$d_{14}$</td>
<td>$d_{15}$</td>
<td>$d_{16}$</td>
<td>$d_{17}$</td>
<td>$d_{18}$</td>
</tr>
<tr>
<td>a</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>n</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

is clearly the widest of the three $N$ choices presented in this section. This gives the opportunity to evaluate how the degree of asymmetry affects the efficiency. Design $d_{18}$ is the optimum with the highest replication of the observational stage.

**Example 70**  The $\{1\}\{3\text{-stage}(2C)\}$ structure

For the crossed model without interaction, i.e. $\{1\}\{3\text{-stage}(2C)\}$, the relative efficiency plots are given in Figure 6.247 on the following page.

The pairs of symmetric designs that have the same relative efficiency can be easily identified. Between the $D_V$- and $A_V$-optimality plots, the order of the design pairs is the same except for $(d_3, d_4)$ with respect to $(d_7, d_9)$. Design $d_8$ does not have a pair. The lowest relative efficiency value is 0.24 for the $D_V$-optimality
Relative efficiency plots for $D_V$- and $A_V$-optimality using the Dispersion-mean OLS method with $N = 48$ in the $\{1\}\{3\text{-stage(2C)}\}$ structure.

Table 6.8  Efficiency loss factor between symmetric design pairs with $N = 48$ in the $\{1\}\{3\text{-stage(2C)}\}$ structure

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Factor</th>
<th>Pairs</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(d_2, d_5) \rightarrow (d_1, d_6)$</td>
<td>1.38</td>
<td>$(d_2, d_5) \rightarrow (d_1, d_6)$</td>
<td>1.14</td>
</tr>
<tr>
<td>$(d_3, d_4) \rightarrow (d_2, d_3)$</td>
<td>1.10</td>
<td>$(d_7, d_9) \rightarrow (d_2, d_5)$</td>
<td>1.03</td>
</tr>
<tr>
<td>$(d_7, d_9) \rightarrow (d_3, d_4)$</td>
<td>1.05</td>
<td>$(d_3, d_4) \rightarrow (d_7, d_9)$</td>
<td>1.01</td>
</tr>
<tr>
<td>$(d_{10}, d_{13}) \rightarrow (d_7, d_9)$</td>
<td>1.33</td>
<td>$(d_{10}, d_{13}) \rightarrow (d_3, d_4)$</td>
<td>1.07</td>
</tr>
<tr>
<td>$d_8 \rightarrow (d_{10}, d_{13})$</td>
<td>1.01</td>
<td>$d_8 \rightarrow (d_{10}, d_{13})$</td>
<td>1.01</td>
</tr>
<tr>
<td>$(d_{11}, d_{12}) \rightarrow d_8$</td>
<td>1.21</td>
<td>$(d_{11}, d_{12}) \rightarrow d_8$</td>
<td>1.04</td>
</tr>
<tr>
<td>$(d_{14}, d_{15}) \rightarrow (d_{11}, d_{12})$</td>
<td>1.15</td>
<td>$(d_{14}, d_{15}) \rightarrow (d_{11}, d_{12})$</td>
<td>1.03</td>
</tr>
<tr>
<td>$(d_{16}, d_{17}) \rightarrow (d_{14}, d_{15})$</td>
<td>1.21</td>
<td>$(d_{16}, d_{17}) \rightarrow (d_{14}, d_{15})$</td>
<td>1.04</td>
</tr>
</tbody>
</table>

As in the previous examples, the progression of the performance of the designs is quantified by the efficiency loss factor. This is shown in Table 6.8. As it happens in previous examples, $D_V$-optimality is much more sensitive to the symmetry changes. $A_V$-optimality is visibly more robust. With the many designs under comparison here, it is clear by means of the relative efficiency plots and the efficiency loss factor, that the more symmetric the design, the better the performance.
Example 71  The \{1\}\{3-stage(2Ci)\} structure

For this structure, the relative efficiency plots are shown in Figure 6.248 and the efficiency loss factors are reported in Table 6.9.

![Figure 6.248](image)

Relative efficiency plots for \(D_V\)- and \(A_V\)-optimality using the Dispersion-mean OLS method with \(N = 48\) in the \{1\}\{3-stage(2Ci)\} structure.

The plots of the crossed structure with interaction are very similar to those of the case without interaction from Figure 6.248. However the minimum relative efficiency in the present example is much lower for both criteria. These are 0.05 and 0.42, respectively, for \(D_V\)- and \(A_V\)-optimality. The pairs of symmetric designs are the same for both criteria.

Table 6.9  Efficiency loss factor between symmetric design pairs with \(N = 48\) in the \{1\}\{3-stage(2Ci)\} structure

<table>
<thead>
<tr>
<th>Pairs</th>
<th>(D_V)</th>
<th>(A_V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((d_2, d_5) \rightarrow (d_1, d_6))</td>
<td>1.62</td>
<td>1.13</td>
</tr>
<tr>
<td>((d_3, d_4) \rightarrow (d_2, d_5))</td>
<td>1.15</td>
<td>1.03</td>
</tr>
<tr>
<td>((d_7, d_9) \rightarrow (d_5, d_4))</td>
<td>1.47</td>
<td>1.25</td>
</tr>
<tr>
<td>((d_{10}, d_{13}) \rightarrow (d_7, d_9))</td>
<td>1.65</td>
<td>1.14</td>
</tr>
<tr>
<td>(d_8 \rightarrow (d_{10}, d_{13}))</td>
<td>1.10</td>
<td>1.06</td>
</tr>
<tr>
<td>((d_{11}, d_{12}) \rightarrow d_8)</td>
<td>1.44</td>
<td>1.09</td>
</tr>
<tr>
<td>((d_{14}, d_{15}) \rightarrow (d_{11}, d_{12}))</td>
<td>1.41</td>
<td>1.10</td>
</tr>
<tr>
<td>((d_{16}, d_{17}) \rightarrow (d_{14}, d_{15}))</td>
<td>1.47</td>
<td>1.09</td>
</tr>
</tbody>
</table>

328
The efficiency loss is large in this case. The values of the efficiency loss factor make this clear. Based on these results, the symmetry of the design when the model contains an interaction term is very important.

**Example 72** The \( \{1\}\{3\text{-stage}(2N)\} \) structure

The general conclusions obtained from the results in this structure are very similar to those for the case with \( N = 24 \). The two obvious differences are the number of designs under comparison here with \( N = 48 \), and the lowest relative efficiencies. In the former example, these efficiencies are 0.12 and 0.45 for \( \mathcal{D}_V \)- and \( \mathcal{A}_V \)-optimality, respectively, and in the present example the corresponding values are 0.02 and 0.35. The efficiency loss factor results are provided in Table 6.10 on the next page and the relative efficiency plots are shown in Figure 6.249.

![Figure 6.249](image)

**Figure 6.249** Relative efficiency plots for \( \mathcal{D}_V \)- and \( \mathcal{A}_V \)-optimality using the Dispersion-mean OLS method with \( N = 48 \) in the \( \{1\}\{3\text{-stage}(2N)\} \) structure.

### 6.8.2 Generalised least squares

Using GLS to estimate the dispersion-mean model allows a number of possibilities. These possibilities basically mean that equivalent methods of estimation that can be implemented using the same model formulation. The references given in Section 6.8 and Section 6.8.1 give account of the equivalencies.
6 Design construction algorithms 6.8 Optimum designs using the Dispersion-Mean model

Table 6.10 Efficiency loss factor between design with $N = 48$ in the $\{1\}\{3\text{-stage}(2N)\}$ structure

<table>
<thead>
<tr>
<th>Efficiency loss factor</th>
<th>$D_N$-optimality</th>
<th>$A_N$-optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair Factor</td>
<td>Pair Factor</td>
<td>Pair Factor</td>
</tr>
<tr>
<td>$d_2 \to d_4$</td>
<td>1.91</td>
<td>$d_{11} \to d_8$</td>
</tr>
<tr>
<td>$d_3 \to d_2$</td>
<td>1.43</td>
<td>$d_9 \to d_{11}$</td>
</tr>
<tr>
<td>$d_7 \to d_3$</td>
<td>1.05</td>
<td>$d_{14} \to d_9$</td>
</tr>
<tr>
<td>$d_4 \to d_7$</td>
<td>1.43</td>
<td>$d_{12} \to d_{14}$</td>
</tr>
<tr>
<td>$d_5 \to d_4$</td>
<td>1.24</td>
<td>$d_{13} \to d_{12}$</td>
</tr>
<tr>
<td>$d_{10} \to d_5$</td>
<td>1.07</td>
<td>$d_{16} \to d_{13}$</td>
</tr>
<tr>
<td>$d_6 \to d_{10}$</td>
<td>1.10</td>
<td>$d_{15} \to d_{16}$</td>
</tr>
<tr>
<td>$d_8 \to d_6$</td>
<td>1.23</td>
<td>$d_{17} \to d_{15}$</td>
</tr>
</tbody>
</table>

Herein, the main interest is to use the dispersion mean model to find REML equivalent designs and also to find designs for models with non normal responses. This is done by taking advantage of the fact that the variance-covariance matrix of the dispersion-mean model includes kurtosis values $\gamma_i$. This is specified in equation (3.60). Since the non normality is introduced via the kurtosis value in (3.60), no radical changes to the formulation are needed to specify different distributions. It is assumed that the data is continuous but non normal. From the known distributions than can be clearly specified by their kurtosis value, the choices made for this section are the logistic and the uniform. For the former, $\gamma = 1.2$ and for the later $\gamma = -1.2$. These where chosen to see the outcome in from to extreme conditions.

The dispersion-mean model is singular and its size is $N^2$. The procedure is very computationally intensive. It requires several hours to compute results for a few hundred preliminary $\sigma_0^2$ vectors. The $\gamma \neq 0$ case is approximately three times slower than the $\gamma = 0$ case. For practicality, 1000 $\sigma_0^2$ vectors were used in with the former, and 500 vectors for the latter case. The computed optimum designs are obtained from the information matrix given by (3.61). To avoid excessive
redundancy at this point, the optimum designs were computed for the case \( N = 24 \). This a very good representative example with nine different candidate designs. The list of candidate designs is given in Table 6.2 on page 165.

**GLS with zero kurtosis**

The condition \( \gamma = 0 \) yields REML equivalent results. The condition implies normality of the observations. Note that when \( \gamma = 0 \) expression (3.60) shortens. This fact is what makes the computational effort approximately three times faster than the case with \( \gamma \neq 0 \).

**Example 73** The \{1\}{3-stage(2C)} structure

The REML results for this structure with \( N = 24 \) were presented in Section 6.3.5. Triangle and robustness plots are shown in Figures 6.19 and 6.20, respectively for \( \mathcal{D}_V \)-optimality. For \( \mathcal{A}_V \)-optimality, these appear in Figures 6.23 and 6.24. These were based on a very fine search grid of 50,000 \( \sigma^2_0 \) vectors.

As mentioned above, for the GLS method with the dispersion-mean model, 1000 \( \sigma^2_0 \) vectors were used in the search grid of the algorithm. The resulting criterion values between the two approaches, for the same set of \( \sigma^2_0 \), are identical to four decimal places and a very high percentage of values are identical to five decimal places.

The generated triangle plots are much less dense than the direct REML versions. The optimality regions are easily perceived, though. As expected, the plots look in general the same as those in Section 6.3.5 and are omitted here. Only the robustness plots are presented. In this dispersion-mean GLS example, they are shown in Figure 6.250 for \( \mathcal{D}_V \)-optimality and in Figure 6.251 on page 333 for \( \mathcal{A}_V \)-optimality.

The large regions of optimality, *i.e.* Region 1 to Region 4, have a good density of grid points. The similarities with the direct REML version are clear. The size
of the region, the efficiency drop, the shape of the efficiency curves and, more importantly, the ordering of the relative efficiency curves in each panel are virtually the same. In contrast, for the smaller regions, sometimes there are too few points and the resulting curves needed to be interpolated to give a reasonable plotting area. This is the reason for the slight differences. However, the general shape of the curves and the ordering is still maintained and it is easy to identify the similarities with the previously reported plots. The equivalency of the results is
clear.

**Figure 6.251**  GLS with $\gamma = 0$ $A_V$-robustness plot for $\sigma^2$ in the $\{1\}{3\text{-stage}(2C)}$ structure.

**Example 74** The $\{1\}{3\text{-stage}(2Ci)}$ and $\{1\}{3\text{-stage}(2N)}$ structures

The direct REML results for the crossed structure with interaction were presented in Section 6.4.3. Triangle and robustness plots are shown in Figures 6.69 and 6.70, respectively for $D_V$-optimality. For $A_V$-optimality, these appear in Figures 6.71 and 6.72.
The dispersion-mean GLS plots for the \{1\}\{3-stage(2Ci)\} structure are shown in Figure 6.252 and Figure 6.253 respectively for the $D_V$- and $A_V$- criteria of optimality.

![Diagram of GLS plots for different regions with $D_V$- robustness plot for $\sigma^2$ in the \{1\}\{3-stage(2Ci)\} structure.]

In the case of the nested structure, the direct REML results were presented in Section 6.7.3. Triangle and robustness plots are shown in Figures 6.211 and 6.212, respectively for $D_V$-optimality. For $A_V$-optimality, these appear in Figures 6.215 and 6.216. For the nested dispersion-mean GLS model, the robustness
6 Design construction algorithms  6.8 Optimum designs using the Dispersion-Mean model

plots are first shown for $D_V$-optimality in Figure 6.254 on the following page and subsequently for $A_V$-optimality in Figure 6.255 on page 337.

The remarks made in the previous example regarding the construction of the plots based on the dispersion-mean model are also applicable for the two structures in the current example. Consequently, the plots are presented without further discussion. The equivalency of the results with those of the direct REML versions is evident.
In order to find optimum designs for non normal responses using the dispersion-mean model, only the specification of the kurtosis value of the selected distribution is required. The full expression for the variance-covariance matrix, as given in (3.60), is then used in the GLS information matrix (3.61).

In this example a kurtosis values of $\gamma = 1.2$ is used. This value corresponds to the logistic distribution. During the set up period, the user specifies if the

---

**Figure 6.254**  GLS with $\gamma = 0 \ D_V$- robustness plot for $\sigma^2$ in the $\{1\}\{3\text{-stage}(2N)\}$ structure.

**GLS with positive kurtosis**

In order to find optimum designs for non normal responses using the dispersion-mean model, only the specification of the kurtosis value of the selected distribution is required. The full expression for the variance-covariance matrix, as given in (3.60), is then used in the GLS information matrix (3.61).

In this example a kurtosis values of $\gamma = 1.2$ is used. This value corresponds to the logistic distribution. During the set up period, the user specifies if the
data is non normal. Accordingly, the user is prompted for a kurtosis value if appropriate. The algorithm then uses the corresponding expressions to compute the GLS information matrix and then chooses an optimum design. As mentioned before, in this case, 500 $\sigma_0^2$ vectors in the search grid were used.

Triangle plots can be generated afterwards. Although the density is very low, in general the regions of optimality can be observed. However, the shapes of the established regions are very similar to those of the normal REML results. The

**Figure 6.255** GLS with $\gamma = 0$ $A_V$-robustness plot for $\sigma^2$ in the \{1\}3-stage(2N) structure.
size of the optimality regions may differ slightly. In these circumstances, the triangle plots can be disregarded. The robustness plots can be used to evaluate the competing designs.

**Example 75** The \( \{1\}\{3\text{-stage}(2C)\} \) structure

The robustness plots based on the dispersion-mean GLS method for the crossed model are shown in Figure 6.256 for the \( D_V \)-optimality criterion.

![Figure 6.256](image)

**Figure 6.256** GLS with \( \gamma = 1.2 D_V \)-robustness plot for \( \sigma^2 \) in the \( \{1\}\{3\text{-stage}(2C)\} \) structure.
In general, the plots resemble those of the REML case. The main differences are the size of the regions. In this logistic version, Region 2 and Region 3 are smaller by approximately 0.16. Region 5 to Region 9 are also smaller, each in different amount, where Region 7 has the largest reduction by a factor of 1/32. This may be more a consequence of the relatively small grid. Naturally, no design is outperformed in its own region. As seen before, the closet challenger is the complementary design, if available.

**Figure 6.257** GLS with $\gamma = 1.2$ $A_V$-robustness plot for $\sigma^2$ in the $\{1\\}{3\text{-stage}(2C)}$ structure.
For the $A_V$-optimality criterion, the robustness plots are shown in Figure 6.257. Relative to the REML case, the self efficiency of Region 2 and Region 3 is about 10% lower in the current example. The rest of the plot does not show significant changes.

Example 76 The $\{1\}\{3\text{-stage}(2\text{Cl})\}$ structure

The corresponding robustness plots for the crossed structure with interaction are shown in Figure 6.258 on the next page when the chosen criteria is $D_V$-optimality.

Unlike the previous model, the results here are not very similar to those of the REML version. In this case, a slightly higher number of changes can be noted. Region 6 has a significantly higher self efficiency, but the size is half. Region 9 has significantly lower self efficiency but has different behaviour since it is kept very close to unity for half the region that is much larger here. This is almost the same in Region 7.

Comparison of the model without interaction in Figure 6.256 and the present one with interaction in Figure 6.258 there are noticeable differences. For this case, Region 1 is significantly different as well as Region 4. There is a dramatic expansion in the size of regions five, six and nine. Region 8 has much higher self efficiency. This strongly suggests that the chosen distribution has big impact when the model contains an interaction term.

The results under $A_V$-optimality are shown in Figure 6.259 on page 342. This plot has its own merits and is not comparable to other plots. It is clear that Region 2 is significantly larger than all the rest. Region 1 is the second largest but is half the size. The self efficiency is high in all regions, and no design is outperformed in its region.
Figure 6.258  GLS with $\gamma = 1.2 \, D_V$-robustness plot for $\sigma^2$ in the $\{1\} \{3\text{-stage}(2\text{Ci})\}$ structure.

**Example 77** The $\{1\} \{3\text{-stage}(2\text{N})\}$ structure

Figure 6.260 on page 343 shows the robustness plot for the nested structure. This one is remarkably similar to the normal REML version in Figure 6.212 on page 303. These plots correspond to $D_V$-optimality. No further description seems necessary.

For $A_V$-optimality the robustness plots appear in Figure 6.261 on page 344. The similarities with the normal REML version in Figure 6.216 on page 307 are also significant.
These similarities between the non normal case and the normal case suggest that the nested structure is not very sensitive to the considered departure from normality.

---

**GLS with negative kurtosis**

In this example a kurtosis values of $\gamma = -1.2$ is used. This value corresponds with the uniform distribution. As before, 500 $\sigma_0^2$ vectors in the search grid were used.
Figure 6.260 GLS with $\gamma = 1.2$ $D_V$-robustness plot for $\sigma^2$ in the $\{1\}\{3\text{-stage}(2N)\}$ structure.

Example 78 The $\{1\}\{3\text{-stage}(2C)\}$ structure

The robustness plots based on the dispersion-mean GLS method for the crossed model are shown in Figure 6.262 on page 345 for the $D_V$-optimality criterion.

The similarities with the normal REML version are also noticeable here. Region sizes, lowest self efficiencies and ordering of relative efficiencies per regions are almost the same.

For the $A_V$-optimality criterion, the robustness plots are shown in Figure 6.263.
6 Design construction algorithms 6.8 Optimum designs using the Dispersion-Mean model

Figure 6.261 GLS with $\gamma = 1.2$ $A_V$-robustness plot for $\sigma^2$ in the $\{1\}$-3-stage(2N) structure.

Relative to the REML case in Figure 6.24 on page 188, only a few self efficiencies appear slightly lower. There is no significant difference to warrant discussion.

Example 79 The $\{1\}$-3-stage(2Ci) structure

The corresponding robustness plots for the crossed structure with interaction are shown in Figure 6.264 on page 347 when the chosen criteria is $D_V$-optimality.

In this case there is no similarity with respect to the normal REML case. By
Figure 6.262  GLS with $\gamma = -1.2$ $D_V$- robustness plot for $\sigma^2$ in the $\{1\}$-$\{3\text{-stage}(2C)\}$ structure.

comparison with the model without interaction in Figure 6.262 and the present one with interaction in Figure 6.264 there differences are also significant. This reinforces the conclusion that the chosen distribution has big impact when the model contains an interaction term.

The results under $A_V$-optimality are shown in Figure 6.265 on page 348. The findings are similar to those in the case of positive kurtosis. Namely, there is no point of comparison with any other plot and the present case has to be evaluated
Example 80 The \{1\}{3-stage(2N)} structure

Figure 6.266 on page 349 shows the robustness plot for the nested structure. In this case, the similarities with the normal REML version in Figure 6.212 on page 303 seem stronger than those in the case with positive kurtosis. These plots correspond to \( D_V \)-optimality.
For $A_V$-optimality the robustness plots appear in Figure 6.267 on page 350. The similarities with the normal REML version in Figure 6.216 on page 307 are also significant.

Once more, due to the strong similarities, the conclusion the nested structure is not very sensitive to the departure of normality seems to remain valid.

---

Figure 6.264 GLS with $\gamma = -1.2$ $D_V$-robustness plot for $\sigma^2$ in the $\{1\}{3\text{-stage}(2Ci)}$ structure.
6.9 Summary

The theory of construction algorithms for optimum designs is used extensively in this chapter. The implementation in the R language of a search algorithm for the variance components is described. Results are presented for several cases. Among others, these include REML estimation, the crossed model with interaction and the dispersion-mean model. Most of the results of the algorithm are provided

Figure 6.265  GLS with $\gamma = -1.2$ $A_V$-robustness plot for $\sigma^2$ in the $\{1\}{3\text{-stage}(2Ci)}$ structure.

6.9 Summary

The theory of construction algorithms for optimum designs is used extensively in this chapter. The implementation in the R language of a search algorithm for the variance components is described. Results are presented for several cases. Among others, these include REML estimation, the crossed model with interaction and the dispersion-mean model. Most of the results of the algorithm are provided.
Figure 6.266  GLS with $\gamma = -1.2$ $D_Y$-robustness plot for $\sigma^2$ in the $\{1\} \{3\text{-stage}(2N)\}$ structure.

graphically. Triangular and rectangular plots are created when using the classical criteria depending if the variability structure is 3-stage or 2-stage, and rank ordered plots are used for the pseudo-Bayesian criteria. This algorithm and corresponding results are the bases of a paper published in the journal Computational Statistics and Data Analysis (Loeza-Serrano and Donev, 2014). Additionally, robustness plots were introduced to assess the performance of a design along its associated region of optimality.

349
Figure 6.267  GLS with $\gamma = -1.2 A_V$-robustness plot for $\sigma^2$ in the $\{1\}\{3\text{-stage(2N)}\}$ structure.

The faculty of visually comparing competing designs given by generated graphics provides a powerful tool for the experimenter. The plots give the opportunity to chose a design, compare alternative designs and assess the robustness of each one. The latter also provides the capability of finding a trade-off between the cost, statistical and practical considerations for the design. Results for designs for variance components that convey so much information in two plots had not been seen before.
Chapter 7

Discussion and conclusions

7.1 Discussion

7.1.1 Unbalanced cases

The presentation given in Chapters 2, 3 and 4 concerning MLE, REML and DM is valid for balanced and unbalanced data. However, the closed form expressions of information matrices found in Chapter 5 are only valid for balanced cases. This discussion presents an initial encounter with a possible way of extending the results in the thesis to unbalanced designs. These results are the initial effort. Further work is required to develop the complete framework.

Evidence that balanced designs are optimal for variance components can be found e.g. in Mukerjee and Huda (1988), and considering different levels of balancedness, Ankenman et al. (2003). The latter found that the most balanced design is optimum.

In spite of this, extensions to unbalanced data are important since some experimental conditions cannot be balanced. In the unbalanced case, closed form expressions of information matrices can not be obtained. The general expressions and methods given in the early chapters have been used to find optimum designs.
The algorithms presented in Chapter 6 have been modified to reflect this. The computational cost is increased noticeably. Depending on the complexity of the problem, this extra burden can be justified.

There is no unified approach to construct optimum designs for variance components for unbalanced experimental situations. In this case, the information matrices of suitable models are dependent on the number of observations defining the design structure, regardless of the chosen method of estimation. Therefore, one approach may work better than another for a given experiment. When a balanced design is not a feasible option, the alternative design will be one preserving as much balance as possible.

One possible implementation to find such unbalanced designs is developed here. Assume a design structure with $v$ stages and total number of observations $N$. Using $w$, the value of $N$ is factorized into its divisors. The resulting set represent the possible number of largest structural units in the top stage of the structure. For the next stage, the combinations $\binom{N/w}{w}$ are obtained. Each element of a combination represents the number of smaller structural units in each of the $w$ larger units. The process is applied throughout the structure of the design. The restriction for the minimum block size is two. This procedure finds a list of candidate unbalanced designs with the characteristic that at each stage, a high level of balancedness is maintained. This class of designs will be called least unbalanced designs.

The candidate list of least unbalanced designs may be short. However, this approach set the foundation for further extensions of class. These extensions should aim at enlarging the number of candidates while being clear of on the level of balancedness desired.

A modified version of the algorithm in Section 6.2.3 is needed to find an optimum design. As mentioned before, the general form for the MLE information
matrix given in equation (3.29) is used for the computations. The results can still be reported graphically, as shown in Chapter 6.

**Example 81 (Unbalanced \{1\}\{3-stage(2)\}) structure, N = 24** To demonstrate the use of the methodology, the \{1\}\{3-stage(2)\}) structure was used in the crossed and nested cases. The number of observations is \(N = 24\). The examples shown here are similar to those used in Chapter 6, and were chosen to demonstrate the effect of the unbalancedness.

For the given number of observations and a 3-stage structure, there can only be four least unbalanced designs. The list of candidate designs is shown in Table 7.1. Consider design \(d_1\) and note that it has two large structural units in Stage 1. At Stage 2, one of the two large units will consist of four small structural units, and the second large unit have only three small units. This is represented as \((4, 3)\). At stage 3, each of the four small units will have three observational units, and each of the three small units will have four observational units, respectively. This is indicated as \((3, 4)\). All designs are interpreted in the same fashion.

**Table 7.1** Least unbalanced candidate designs

<table>
<thead>
<tr>
<th></th>
<th>(d_1)</th>
<th>(d_2)</th>
<th>(d_3)</th>
<th>(d_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Stage 2</td>
<td>(4,3)</td>
<td>(6,2)</td>
<td>(4,2,2)</td>
<td>(3,3,2)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>(3,4)</td>
<td>(3,3)</td>
<td>(3,3,3)</td>
<td>(3,3,3)</td>
</tr>
</tbody>
</table>

For the cases of \(A_V\)- and \(D_V\)-optimality triangular plots were obtained. When the structure is crossed, the result is depicted in Figure 7.1. When the interest is in pseudo-Bayesian optimality, for the condition of total lack of knowledge for \(\sigma^2\), the output is shown in Figure 7.2. For the nested cases, the corresponding results can be consulted, respectively in Figure 7.3 and Figure 7.4 on page 355.

These plots convey the same information as those introduced in the previous chapters. Exploring the \(A_V\)-optimality triangle on the left of Figure 7.1, it is
Discussion and conclusions

7.1 Discussion

Figure 7.1 $A_V$- and $D_V$-optimality plots for $\sigma^2$ in the unbalanced $\{1\}$\{$3$-stage$(2C)$\} structure.

![Image](image1.png)

Figure 7.2 Pseudo-Bayesian $A_V$- and $D_V$-optimality plots when $\sigma^2$ is unknown for the unbalanced $\{1\}$\{$3$-stage$(2C)$\} structure.

![Image](image2.png)

interesting that $d_1$ is only optimum when $\sigma^2_\epsilon$ is large and the other two variance components are comparatively small. Design $d_4$ seem to be most sensitive to $\sigma^2_2$, whereas $d_3$ and $d_4$ depend much more on $\sigma^2_1$. On the right hand side, the $D_V$-optimality triangle shows that $d_1$ is optimum only when $\sigma^2_1$ has no influence. When $\sigma^2_1$ is approximately reciprocal to $\sigma^2_\epsilon$, design $d_2$ is best. There is no dominating condition for the other two designs.
When using the pseudo-Bayesian criteria of optimality with the crossed structure, it is clear from Figure 7.2 that $d_3$ is both $A_V$- and $D_V$-optimum. Designs $d_2$ and $d_4$ are almost equally good, but much more inefficient than $d_3$.

![Figure 7.3](image1.png)  
**Figure 7.3** $A_V$- and $D_V$-optimality plots for $\sigma^2$ in the unbalanced $\{1\}\{3\text{-stage(2N)}\}$ structure.

![Figure 7.4](image2.png)  
**Figure 7.4** Pseudo-Bayesian $A_V$- and $D_V$-optimality plots when $\sigma^2$ is unknown for the unbalanced $\{1\}\{3\text{-stage(2N)}\}$ structure.

For the nested structure, note from Figure 7.3 that $d_3$ is never optimum since it does not appear on either triangle. Another interesting feature is that the two
criteria of optimality choose the same designs for approximately the same range of values of $\sigma^2$. The only exemption is that for $D_{V}$-optimum, $d_1$ is more sensible to the range when $0.3 < \sigma_1^2 < 0.5$ and $\sigma_2^2 < 0.2$, which is a very small area.

The pseudo-Bayesian plots on Figure 7.4 show that $d_4$ and $d_3$ are almost equally optimum, however $d_3$ uses fewer resources and it may be preferable at a minimum efficiency lost. This is the case for both, $D_{V}$- and $A_{V}$-optimality.

In comparison to the results obtained in the balanced case (Chapter 6), it is clear that, even from the generation of the candidate list, the unbalanced case is more complicated. The interpretation of the value of any $\sigma_i^2$ is not straightforward, since the unbalancedness hides its relative weight. This is a reason why finding least unbalanced designs is useful.

It is informative to explore the behaviour of the optimality criterion across the design region $\mathcal{Z}$. Consider the case where $\sigma_2^2 > 0.5$, i.e. when it dominates over $\sigma_1^2$ and $\sigma_2^2$. This is the region above the horizontal line shown on the triangular plots in Figure 7.1 on page 354.

In this case, the $A_{V}$-optimality criterion chooses $d_2$, the design with more replication in stage 2, although most unbalanced. When $\sigma_1^2$ is almost as large as $\sigma_2^2$, $d_3$ is chosen, which reflects the higher weight of stage 1, but keeps the most unbalanced replication in stage 2. For the balanced case, in the same region, $d_8$ is chosen except when $\sigma_1^2$ is almost as large as $\sigma_2^2$, in which case $d_2$ is preferred. The latter is the same behaviour as in the unbalanced case and it is expected. For $D_{V}$-optimality, the right plot in Figure 7.1 shows that $d_3$ is best, except when $0.05 < \sigma_1^2 < 0.1$ where the less unbalanced $d_4$ is chosen. However, the design with most weight on stage 2, i.e. $d_2$, is only best when $\sigma_2^2$ is not dominant, which is unexpected. In contrast, the balanced case is more predictable choosing the designs with more replication in the stage of higher variability.

For the nested structure, consider the case where $\sigma_c^2 > 0.5$, which corresponds
to the lower left part of the triangle delimited by the diagonal line shown in Figure 7.3. The least unbalanced design of the candidate list, \(d_1\), is best when \(\sigma_1^2\) is very large. Design \(d_2\), that puts most weight in stage 2, is indeed optimum as \(\sigma_2^2\) becomes larger than \(\sigma_1^2\). Similarly, when \(\sigma_1^2\) is larger than \(\sigma_2^2\), a design with more weight in stage 1 is better, \(i.e.\ d_3\). However, note that since \(\sigma_2^2\) dominates, stage 2 is most balanced. This applies for both criteria. For the balanced case, the experimental effort is shifted progressively as the corresponding variance component dominates, which is expected.

What is observed in general, is that in the unbalanced case the task is to identify the stage with largest variance to direct the experimental effort. Afterwards, the amount of balancedness is dictated by the value of \(\sigma_2^2\). However, it is not clear at which point one variance component starts to dominate. In the balanced case, only the first step is required and there is no difficulty to identify the region of dominance of a particular variance component.

The example also shows the usefulness of the algorithm based on general expressions for the MLE information matrix. It may happen that in practice none of the candidate designs is feasible, but they can serve as a benchmark to assess the efficiency of more highly unbalanced designs. With respect to the balanced case, the computational cost is about four to six times higher for the unbalanced cases. The impact that a relatively small amount of unbalancedness has on the optimality of designs for variance components is significant. This reiterates the importance of further studies on the topic.

7.2 Conclusions

The most advantageous way to summarise the results contained in the thesis is by grouping them. When taken collectively, they provide a unified framework that
experimenters can use consistently in the different experimental phases of a study. This framework is especially strong when used for the modelling of experimental conditions and the design of an optimum experiment using hierarchical models with variance components.

The first group of results proposes a general but highly detailed way to define the linear mixed effects model. The extension considers the explicit definition of all the elements needed to construct a model, even if they do not become part of the final model equation. One key aspect of the proposed formulation involves declaring the random part as a functional that individually determines the form of the design matrices for each random regressor; these are not limited to be indicator matrices since they can include fixed regressors. Further, the model is strictly divided into the \textit{treatment structure} and the \textit{variability structure}. This allows the separate definition of the structures but using one single rationale. The rationale is that of combining with relatively no restrictions, simple design structures called \textit{factor layouts}. As a whole, this approach provides a strong link between the identification and definition of the required structure of the experiment to be executed, and the statistical model that can represent that experiment accurately. Moreover, it has been conceived to promote the habit of thinking concurrently about the relationships between the planning and execution of an experiment and the data analysis to be performed. The approach is compared with the use of Hasse diagrams.

In a second group of results different forms of information matrices for variance components were obtained. MLE closed form expressions for balanced designs for random models with 3-stage variability structure. This was done for the nested layout case when the information matrix is a function of the vector of variance components $\sigma^2$. For the crossed, crossed with interaction and nested layouts expressions were obtained when the information matrix is a function of the vector $\sigma^2$. 

358
of the ratios of the variances $\eta$. For the case of the ratios, a general expression based on REML is provided. A most valuable outcome is the understanding of the process of deriving the matrices. Since the applications demanding the use of variance components models are quite broad, it is most important to be able to construct different matrices in different circumstances as demanded, whenever possible.

The third group of results consist of the definition of algorithms that represent the methodologies from the other two groups and that are able to construct the optimum designs, with their subsequent implementation. The main result involves a collection of algorithms to find optimum designs for the variance components. They can be seen as a one single algorithmic tool. This algorithm implements a number of options including designs when the interest is in either $\sigma^2$ or $\eta$. It also works with classical $\mathcal{D}_v$- and $\mathcal{A}_v$-optimality criteria or pseudo-Bayesian. An important achievement is the graphical presentation of the results. The proposed plots convey enough details for the experimenter to make an informed decision about the design to implement in practice. A number of these results form the basis of a paper accepted for publication by the journal Computational Statistics and Data Analysis (Loeza-Serrano and Donev, 2014). Furthermore, the algorithms can potentially be combined to provide a scalable tool to define an experimental phase and improving results with sequential experimentation.

Sequential experimentation when variance components are involved, is at a glance a simple extension of the work. However, it is not difficult to identify key complications. For instance, the most efficient way (theoretically and in practice) to handle the construction of designs for the fixed and the random parameters jointly is not yet determined, even for relatively simple mixed models. Also, the selection of the optimality criterion has to be considered carefully. A natural choice for the optimality criterion for variance components seems to be $\mathcal{V}$-optimality.
also known as $I$-optimality. There are difficulties with its implementation due to the integration of the nonlinear functions of the variances. There is a subtle antagonism in the requirements for an optimum design for the mean response and those for the variance components. How the find the balance is an active research area. On the practical side, the algorithms that were created are a solid base for the generation of an R package to construct optimum designs for a large variety of situations and even to perform analyses of the data obtained. This can also be investigated in the future.

As mentioned in the discussion, this work can be naturally extended to general cases of unbalanced designs. The initial development of an approach for unbalanced cases was presented. Even further progress could be made to find optimum designs for the variance components using generalised linear models with random effects. Another natural extension is to find optimum designs for general functions of the variance components, not only for the ratios. Developments in these fronts would be highly valuable but they require different mathematical tools and hence, can be seen as independent research projects for the future, rather than direct extensions of the present work.

The industrial internship that is reported in an appendix, added a very positive dimension to the work and the spirit of the research. It is significant that the results presented here were put into practice. Equally valuable is the experience given by challenges that the statistician faces when implementing the methodologies.
Appendix A

Mathematics results

The results shown here can be consulted in Harville (2008), Searle et al. (1992), Malley (1986), Rao (1973), Rao (1997), among others.

A.1 Matrix Algebra

Derivative of quadratic forms

A quadratic form of the random vector \( y \) is \( y^T A y \).

\[
\frac{\partial (x^T A x)}{\partial x} = (A + A^T)x
\]  

(A.1)

which for a symmetric matrix \( A \) becomes

\[
\frac{\partial (x^T A x)}{\partial x} = 2A x.
\]  

(A.2)

Derivative of a matrix whose elements are functions of \( x \)

For \( F \) representing a matrix of functions of a vector \( x = (x_1, \ldots, x_m) \) of \( m \) variables and \( a \) being a constant

\[
\frac{\partial (a F)}{\partial x_j} = a \frac{\partial F}{\partial x_j},
\]  

(A.3)
and the derivative of the inverse is

$$\frac{\partial F^{-1}}{\partial x_j} = -F^{-1} \frac{\partial F}{\partial x_j} F^{-1}. \quad (A.4)$$

**Derivative of the trace**

For $F$ as specified before

$$\frac{\partial}{\partial x_j} \text{tr}(F) = \text{tr} \left( \frac{\partial F}{\partial x_j} \right). \quad (A.5)$$

**Derivatives of determinants**

For $F$ as specified above and also nonsingular

$$\frac{\partial |F|}{\partial x_j} = |F| \text{tr} \left[ F^{-1} \frac{\partial F}{\partial x_j} \right], \quad (A.6)$$

$$\frac{\partial \log(|F|)}{\partial x_j} = \frac{1}{|F|} \text{tr} \left( F^{-1} \frac{\partial F}{\partial x_j} \right). \quad (A.7)$$

**Some properties of Kronecker products**

$$AB \otimes KN = (A \otimes K)(B \otimes N) \quad (A.8)$$

### A.2 Statistics

**Quadratic forms**

Consider the random variable $y$ with a probability distribution such that $E(y) = X\theta$ and $\text{var}(y) = V$. Then it follows that

$$E(y^T A y) = \text{tr}(AV) + \theta^T X^T A X \theta. \quad (A.9)$$
If specifically $y \sim N(\mathbf{X}\theta, \mathbf{V})$, then

$$\text{var}(y^T \mathbf{A} y) = 2\text{tr}\left[ (\mathbf{A}\mathbf{V})^2 \right] + 4(\mathbf{X}\theta)^T \mathbf{A} \mathbf{V} \mathbf{A} \mathbf{X} \theta. \quad (A.10)$$

Also,

$$E((y - \mathbf{X}\theta)^T \mathbf{A}^{-1} (y - \mathbf{X}\theta)) = \text{tr}(\mathbf{A}\mathbf{V}), \quad (A.11)$$

where the matrix $\mathbf{A}$ has no random elements.

A mean squares is denoted by $MS$, which is a quadratic of linear combinations of $y$. Let the corresponding number of degrees of freedom be denoted by $f$ and the expectation as $EMS$. Under normality assumptions,

$$\frac{fMS}{EMS} \sim \chi^2_f. \quad (A.12)$$

It then follows that

$$\text{var}(MS) = \frac{2EMS^2}{f} = E(MS^2) - EMS^2. \quad (A.13)$$

Therefore,

$$E(MS^2) = \frac{(f + 2)\text{var}(MS)}{2} \quad (A.14)$$

and an unbiased estimator for $\text{var}(MS)$ can be written as

$$\text{vâr}(MS) = \frac{2MS^2}{f + 2}. \quad (A.15)$$
Maximum likelihood: parameter transformation

Let $\psi = g(\theta)$ for $g(\cdot)$ being a monotone function. Then, the score function of $\psi$ is

$$H = \frac{\partial \theta_i}{\partial \psi_j}, \quad (A.16)$$

where $i, j = 1, \ldots, r$. The information matrix after the transformation is

$$M(\psi) = E \left( \frac{\partial l}{\partial \psi} \frac{\partial l}{\partial \psi^T} \right), \quad (A.17)$$

where $l$ is the log-likelihood function. Further

$$\frac{\partial l}{\partial \psi} = \sum_i \frac{\partial \theta_i}{\partial \psi_j} \frac{\partial l}{\partial \theta_i} = H \frac{\partial l}{\partial \theta}. \quad (A.18)$$

Consequently,

$$M(\psi) = HE \left( \frac{\partial l}{\partial \theta} \frac{\partial l}{\partial \theta^T} \right) H^T = HM(\theta)H^T. \quad (A.19)$$

### A.3 vec operator and permutation matrix

The vec operator

The vec operator linearises a matrix $A$ with $m$ columns by creating a column vector $mm \times 1$ by placing each column of $A$ one under the other starting with the first column.

The vec-permutation matrix

The vec-permutation matrix of size $mm \times mm$ is denoted by $S_m$. One interpretation is that $S_m$ is partitioned as a $m \times m$ matrix of $m \times m$ submatrices. These submatrices have the $(i, j)$th element as zero and only the $(j, i)$th is one. For
example,
\[
S_2 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}.
\]

(A.20)

Properties

\[
\text{vec}(ABC) = (C^T \otimes A)\text{vec}B.
\]

(A.21)

\[
\text{tr}(AB) = (\text{vec}A^T)^T \text{vec}B.
\]

(A.22)

For \( t \) being a column vector,
\[
(t \otimes t) = \text{vec}(tt^T).
\]

(A.23)

\[
S_n(A_{n \times k} \otimes B_{n \times k})S_k = A_{n \times k} \otimes B_{n \times k},
\]

(A.24)

\[
S_n(A_{n \times k} \otimes B_{n \times k}) = (A_{n \times k} \otimes B_{n \times k})S_k.
\]

(A.25)
Industrial internship

Experimental designs for
production process optimisation

B.1 Introduction

A four month internship was done with an automotive products company. The project was funded by the research grant Shorter KTP No 10000651, which was jointly sponsored by EPSRC and the industrial partner Federal-Mogul Friction Products Limited.

Federal-Mogul is a global manufacturer of automotive components. The Friction Products Group supplies brake pads and brake linings to most vehicle manufacturers, with six production sites within the EU alone. The design centre in the UK determines the mixtures of raw material used to produce the brake pads. The design team tests the new mixtures by making prototype parts. Mixtures change due to differences in raw material batches and mixers. The mixing and production processes may also vary for different production lots. The production process compacts (press) the mix batch into individual parts that are then baked
in lots. The individual parts are labeled to show mix batch and production lot.

The characteristics of the final product are related to the mixture, but the settings of the manufacturing process are equally influential to achieve consistent performance and quality. Therefore, the design team has to determine the mixture and the manufacturing process settings for the products to meet the specified customer requirements.

B.2 Description of the project

B.2.1 Justification

Appropriate manufacture of brake pads in Federal-Mogul this is a two fold challenge. First, the right mixture has to be designed and the correct process settings are needed so that the product has the required properties. Second, the design team must provide a procedure to launch a new product into production, so that the engineering teams at the different manufacturing locations are able to find the exact settings for their production lines, while keeping quality levels consistent across all sites.

Common methods used in industry for general improvement, such as Six Sigma, have been employed by the company. Results have been positive, but further progress has not been possible. Every improvement step is increasingly difficult. Unconventional tools are needed. Attention has now been given to methods for identifying and understanding the sources of variation during the development and transfer of new products into production.

In the first instance, it has been recognised that the experimental units get to be split. For example, one batch of raw material being split to create several mixture lots. Also, mixture can be split and be pressed and baked in different equipment. This involves a hierarchical structure (see Chapter 4). Additionally, production
process limitations must be considered. Standard Split-Plot design of experiments accommodate hierarchies created by partitioning sampled units, however there is not an established procedure to incorporate the process restrictions into the Split-Plot structure. Designs for multistage variability have to be found for this application.

Secondly, process studies have shown that product variability has complex behaviour. The overall variation is different for different mix batches, different equipment and different manufacturing sites. Therefore, it is of much interest to learn about how to perform experiments to estimate the variance components more accurately.

B.2.2 Objectives

The aim of this project was to provide efficient alternatives to conventional design of experiments for process optimisation and to propose the use of designs for variance components estimation. The project was composed of four general stages:

Stage 1
Review of past and current practice on statistical investigations in Federal-Mogul.

Stage 2
Determination of the features required for experimental designs for improving the process.

Stage 3
Proposal, execution and analysis of improved experimental designs for the process.

Stage 4
Run experiments for the precise estimation of variance components.

The specification of the linear mixed model, as given in Section 3.3, was used. The theory of Split-Plot experimental designs using optimality criteria was the base line for the process optimisation experiments. Balanced designs for estimating the
variance components for the \( \{1\}\{3\text{-stage}(2C)\} \) structure were used (see Chapter 5 and Chapter 6).

### B.2.3 Outcomes

**Stage 1** Observations:
- Temperature and Batch are key input factors.
- There is a nuisance factor related to the ageing of the mix.
- Previous designs had been set up mainly as screening designs, but the main variables had already been identified. Consequently, experimental data is not ideal for process optimisation.
- Blocking was not advantageously used. The runs were not properly randomised, particularly, Temperature levels were changed sequentially, suggesting that Temperature is a HC factor.

**Stage 2** Key requirements:
- Estimability of full quadratic model on seven process factors. Mixture designs were not covered in this project.
- Block effects for Temperature and Batch have to be included.
- Ageing (of the mix) effects need to be assessed.

**Stage 3** Proposal of an improved design featuring:
- Efficient blocking by Temperature levels that are reset twice per day for better randomisation.
- Batch effect included in the model and used a the whole-plot factor.
- A design structured to account for potential linear time trends (ageing effects).

**Stage 4** Assessment of pilot experiments for variance components and comparison with a proposed balanced design.
The proposed Split-Plot design is executed over three days. The whole-plot factor is Batch. A large mix batch is prepared for the three days. Thus, there are three whole-plots. It was considered that Batch have a 1-factor(1C) layout with random effects. At the split-plot level, the configuration is a randomised block design (CRBD). Temperature with three levels ($T_1, T_2, T_3$) is the blocking factor. The within block treatments come from completely randomised layout of six process factors. Three blocks can be run in a day. The blocks are randomised per day according to the rule $T_1 > T_2 < T_3$, i.e. only one cooling cycle is allowed.

The working shift and the temperature settle time only allowed to run a maximum of 30 observational units. These were spread by assigning two block size of 13 and one of size four. The 13 treatments come from a 6-factor(6C) layout of the split-plot factors and the remaining four are centre points. Centre points were requested by the industrial planer. This implies that there are 30 experimental runs in one day and a total of 90 for the whole experiment. With $N = 90$ a full second order model in six split-plot factors, one or two whole-plot (blocking) factors, a random batch effect and the overall mean can be estimated. This setting gives the flexibility of running an experiment with $N = 60$ observations simply by using only two days instead of three.

According to the outcomes from Stage 3, there is an ageing effect of the mix. Consequently, the CRBD at the split-plot level was arranged in a patterned structure that balances the design for linear trend effects. The construction of optimum designs for linear and quadratic trend is described in Atkinson and Donev (1996). Table 1 from that publication can be used in the current situation by considering each block as a treatment. If two das are ran, then part (a) of that table gives the proper order to use. For three days, part (b) is used.

After running a few experiments following the description above, it was confirmed that Temperature should contribute fixed as well as random effects since the
predictive capabilities of the model were significantly increased. This was assessed using JMP’s profiler feature. Also, since it was confirmed that by chemistry the ageing cannot be avoided or controlled, the configuration for the time trend was kept. How these designs have been incorporated in the current practice in Federal-Mogul is described in Section B.3.1.

Some experiments having the estimation of variance components in mind were carried out by Federal-Mogul across different locations. There was no time to work together with the company to design these experiments. The results of the largest one were provided by the company and a study was done to compare the executed design with a balanced design that would have potentially been suggested. This exercise is described in Section B.3.2.

B.3 Final outcome

This section describes how the work done have impacted the current operation in Federal-Mogul.

B.3.1 Design with trend consideration for process optimisation

The operation now has two stages. First, the operational range of the process factors, which varies across products and factories, are mapped. This is done using a small CRB design $14 < N < 20$ blocked by Temperature using alternating levels. This is the same configuration as the split-plot stage. Only the maximum and minimum levels of all process factors are needed. Then randomised trials spanning the given ranges are allocated to the blocks. Block sizes from three to five are used. For this experiment, only one mix batch is used, and all the runs have to be completed in one day. At the second stage, the improved Split-Plot
design is used to optimise process settings. Obviously, the useful range of the factor levels comes from the first stage.

This procedure is applied by the engineering team in the first place to find the extreme operational ranges and construct prototype parts relevant for testing compliance with the customer requirements. Subsequently, the manufacturing teams are given the safety windows as a starting point, and then run both stages of the procedure to find the corresponding process settings that would work for them, using the improved Split-Plot design to fine tune the settings.

### B.3.2 Variance study

Data from an experiment intended to estimate variance components was provided. The goal here is to compare the design that was used, with an optimum balanced design for variance components that is proposed. The provided data were used with the R code developed to compare designs for variance components that has been described in Section 6.2.

The original experiment and the data

The data provided was collected in an experiment involving two manufacturing locations. It comprises \( N = 945 \) observations. The response is the compression rate. The variability of the response due to the effects of random batches of mix and random presses is what the company is interested in. There were a total of nine batches of mix and eight presses used in the experiment. In one location the number of observations was \( n_1 = 799 \) obtained using six different batches of mix and six presses. At the second location \( n_2 = 146 \) with seven batches and two presses.

The company suggested that for this evaluation the location should be ignored. Consequently, it will be considered that during the experiment nine batches and
eight presses were used. A suitable model for the situation of interest is:

\[ y = \mu + Z_1 \alpha_1 + Z_2 \beta_2 + \epsilon, \quad (B.1) \]

as was defined in equation (5.8). Random interaction effects were not considered since it was not possible to agree on their interpretation. The design has \( \ell_\alpha = 9 \) random batch effects denoted by \( \alpha \), and \( \ell_\beta = 8 \) random press effects denoted by \( \beta \). In total, there are three variance components in this model: \( \sigma_1^2 \), \( \sigma_2^2 \) and \( \sigma_\epsilon^2 \), corresponding to the batch of mix, press, and random error. The matrices \( Z_1 \) and \( Z_2 \) have the following form:

\[
\begin{array}{cccccccccc}
B1 & B2 & B3 & B4 & B5 & B6 & B7 & B8 & B9 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
6 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
7 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
9 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
10 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
11 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
12 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
13 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
14 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
15 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
16 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
17 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
18 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
P1 & P2 & P3 & P4 & P5 & P6 & P7 & P8 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
6 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
9 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
10 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
11 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
12 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
13 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
14 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
15 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
16 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
17 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
18 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

These matrices show the variability structure of the design that was used. The random effects \( \sigma_1^2 \) and \( \sigma_2^2 \) are crossed. However, the data is not balanced because every batch of mix was not used the same number of times in the same press. Additionally, this representation only shows one replicate of each design point; each row has to be repeated the corresponding number of times to form the full
matrices (945 rows) required for subsequent computations.

Using estimated values of each variance, provided by Federal-Mogul, we proceeded to compute the information matrix of the variance components for the original experiment according to equation (5.9) using the model in (B.1). The variance values are $\sigma_1^2 = 8.3$, $\sigma_2^2 = 10$, $\sigma_\epsilon^2 = 7.3$. These values have to be rescaled into a $[0, 1]$ range, so that each value represents the proportion of variability they contribute toward the total. The scaled values are:

Scaled preliminary variances:

[1] 0.3242188 0.3906250 0.2851562

The structure of the original experiment (FM) gives the following result:

FM original experiment Information matrix:

\[
\begin{bmatrix}
se & s1 & s2 \\
se & 5719.1943195 & 1.973761 & 0.7721072 \\
s1 & 1.9737614 & 32.430056 & 1.7038359 \\
s2 & 0.7721072 & 1.703836 & 20.1288338
\end{bmatrix}
\]

This information matrix was used to assess the relative efficiency of the optimum design that is going to be proposed. The relative efficiency is an indication of how much information is available using the proposed optimum design in comparison to the information generated by the original design. A value greater than one implies that the optimum design is better.

The proposed optimised design

The procedure to choose an optimum design begins with the creation of a list of candidate designs based on the required total number of observation $N$. The resulting list is going to be shortened in order to reflect practical limitations in the experimental conditions. These constraints are:
1. the number of batches of mix ($n_1$) has to be the smallest possible since they are expensive,

2. the number of available presses ($n_2$) should be no more than 21, since that is the maximum available,

3. the number of replicates ($n$) should be no more than 52, so that one condition takes at most one day of experimentation.

An even number of observations is going to be used to choose the optimum design so that the experimental plan can be divided easily in two locations if necessary. With these constraints, a list of candidate designs based on $N = 944$ (rather than 945) observations gives no designs to compare. By using $N = 936$ instead, the resulting list contains 20 designs, which provide more relevant combinations to consider. Note that the list of 20 candidate designs includes one design (number five) with nine batches and eight presses as in the original experiment, which results in 13 replicates. This particular design $d = (9, 8, 13)$ can be used as a benchmark. The list of candidate designs is as follows.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>[2,]</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>[3,]</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>[4,]</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>[5,]</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>[6,]</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>[7,]</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>[8,]</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>[9,]</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>[10,]</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>
The second step consists of creating a grid of values for $\sigma^2$ vectors, so that most of the design region can be evaluated. The grid consists of 30 thousand random vectors. The third step performs the actual computations of information matrices.

![Figure B.1](image)

**Figure B.1** $D$- and $A$-locally optimum designs. Design 18 is $D$-optimum and $A$-optimum for the preliminary values shown as the pink dot.

The optimality criteria and design comparisons. Two optimality criteria are evaluated: $D$-optimality and $A$-optimality. The results are reported in a triangular plot in Figure B.1, as those explained in Section 6.2, showing the range of different $\sigma^2$
values where a given design is optimum. The pink dot inside the triangle depicts the vector of preliminary estimated values that was provided by Federal-Mogul which is \( \sigma_1^2 = 8.3, \sigma_2^2 = 10, \sigma_\epsilon^2 = 7.3 \).

The results can be seen numerically as

A and D optimum design:

\[
\begin{array}{ccc}
n_1 & n_2 & n \\
9 & 13 & 8 \\
\end{array}
\]

Information matrix Optimum designs:

\[
\begin{bmatrix}
se & s1 & s2 \\
se & 5626.3424169 & 0.3732504 & 0.5445516 \\
s1 & 0.3732504 & 38.8180433 & 0.9699479 \\
s2 & 0.5445516 & 0.9699479 & 39.2077123 \\
\end{bmatrix}
\]

Relative D efficiency:

[1] 1.320492

Relative A efficiency:

[1] 0.8595715

The reported optimum design indicates that \( n_1 = 9 \) different batches of mix and \( n_2 = 13 \) different presses have to be used with \( n = 8 \) pads having to made at each batch-press combination. The relative \( \mathcal{D} \)-efficiency being larger than one indicates that under the \( \mathcal{D} \)-optimality criterion the chosen design is extracting more information than the original experiment. In addition, the optimum design is \( \mathcal{A} \)- and \( \mathcal{D} \)- optimum.

A more practical way to appreciate the results can be stated as follows. Since an optimum design has been found to provide more information than the original
design using roughly the same resources, is there another design that provides the same information as the original design using less resources?

To answer the question, comparing the variances of the variance estimates is helpful. The original design provides these variance values (scaled to $10^3$ for easier comparison), where smaller is better, as:

**FM variance of estimates:**

<table>
<thead>
<tr>
<th></th>
<th>se</th>
<th>s1</th>
<th>s2</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>0.1748542</td>
<td>30.9739496</td>
<td>49.9020967</td>
</tr>
</tbody>
</table>

and the optimum design $D_1 = (9, 13, 8)$ gives

**Optimum Design_1 variance of estimates:**

<table>
<thead>
<tr>
<th></th>
<th>se</th>
<th>s1</th>
<th>s2</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>0.1777357</td>
<td>25.7771654</td>
<td>25.5209950</td>
</tr>
</tbody>
</table>

Hence, the variance of the estimate for $\sigma^2_1$ is roughly the same for both designs. This is expected since the number of observations is almost the same. The variance for $\sigma^2_1$ provided by the optimum design is about 17% lower than the original, and the variance for $\sigma^2_2$ is 49% lower, *i.e.* almost half of the original design.

Note that the variance of the estimate for $\sigma^2_1$ is extremely low, about 200 times smaller than the other two variances. However, the interest of this experiment is in estimating as precisely as possible the random effects of batches and of presses, for which good estimates of $\sigma^2_1$ and $\sigma^2_2$ are needed. This suggests that the experimental resources can be allocated differently in order to:

1. use the same number of observations to increase the precision on $\sigma^2_1$ and $\sigma^2_2$,
2. keep the same precision on $\sigma^2_1$ and $\sigma^2_2$ as in the original design by reducing the number of observations, or
3. get similar precision on both $\sigma^2_1$ and $\sigma^2_2$. 

378
Point number one is already achieved by the optimum design $D_1$. To cover point two, consider the design $D_2 = (8, 7, 2)$ which gives the following variances:

**Design_2 variance of estimates:**

\[
\begin{array}{ccc}
    se & s1 & s2 \\
    1.659461 & 33.086273 & 52.641160 \\
\end{array}
\]

which are only 6% for $\sigma_1^2$ and 5% higher for $\sigma_2^2$ than the original design but using only about 12% of the observations (112). For point number three, consider the design $D_3 = (9, 13, 4)$ that uses about half the number of observations than the original Federal-Mogul design, but the estimates of the variance of the estimated variance components are about twice smaller:

**Design_3 variance of estimates:**

\[
\begin{array}{ccc}
    se & s1 & s2 \\
    0.3638214 & 26.2045529 & 26.0315611 \\
\end{array}
\]

This comparison shows that the original design used by Federal-Mogul was rather wasteful. Better estimates of the variance components can be obtained with much less resources using an optimum balanced design for the variance components.

B.4 Conclusion

Process optimisation

This project identified a number of issues. As a result, sensible improvements have been made. The new experimental designs have important features that were weak or lacking in the previous practice. Additional sources of variability
were identified. Now, they are dealt with more carefully. For instance, they are modelled explicitly in the statistical analysis yielding better results.

The new experimental designs are considerably better than those that were in place in Federal-Mogul before. These new designs fully exploit the presence of blocking variables, such as temperature, and are randomised more reliably. Also, they are large enough for good model estimation.

This optimisation approach has allowed the process cycle times to be much shorter. This is an important contribution because it will translate into significant financial gains over the lifecycle of a product.

**Experiments for variance components**

Three experimental designs were compared to the original experiment. This study showed that the original design can be improved in terms of balancedness and in the number of replicates that are required for good estimation of the variance components. This can make the experiment easier to run in a shorter period of time, thus potentially improving the reliability and reducing costs. Remarkably, the same performance could have been achieved by using only 12% of the experimental resources used in the original design.

**General remarks**

The industrial internship project proved to be very useful for all people involved. The industrial partner was able to achieve improvements they could not find before. From an academic point of view, it was useful to familiarise with challenges faced by practitioners in the field.

More importantly, it was seen that there is methodology needed in practice that is not fully covered in research. Additionally, very few commercial packages can be fully customised to help in designing complex experiments, as those used
here, and analysing the results. Conventional methods are not sufficient.

\section*{B.5 Summary}

A four months industrial internship was done during the course of the PhD program. This opportunity was very rewarding for all parties involved. It was possible to put the research reported in this thesis into practice. However, no new research results were obtained from the internship. This was not the main objective and there was not enough time.

Solutions requiring the application of well known statistical tools were suggested. Improvements were achieved by the hosting company. For more complicated problems, it was shown that the methodology presented in this thesis is relevant and can be used in industry.

Recommendations based on the results of the application of the methodology were provided. These could improve the operations of the company by significantly reducing the amount of experimental resources needed.
Bibliography


JMP Software. SAS Institute, Inc., Cary, N.C. USA.


387


SAS software. SAS Institute, Inc., Cary, N.C. USA.


Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance thay
give towards a proper choice of the distribution of observations. Biometrika, 12:1–85.


