

Optimal designs for two-colour microarray experiments

Penny S. Sanchez

Discipline of Statistics

School of Mathematical Sciences

The University of Adelaide

February 2, 2010

Table of Contents

Declaration	iii
Acknowledgments	v
Publications arising from this thesis	vii
1 Introduction to biology and microarray technology	1
1.1 Introduction	1
1.2 Genes and DNA	2
1.3 Microarrays	3
1.3.1 Introduction	3
1.3.2 Steps involved in carrying out a two-colour microarray experiment	3
1.3.3 Changes in microarray technology	6
1.4 Normalization	7
1.5 Spot measurements	8
2 Review of statistical literature	9
2.1 Introduction	9
2.2 Principles of experimental design	10
2.3 Early designs proposed for microarray experiments	13

2.4	Approach based on Pareto optimality	14
2.4.1	Design problem	14
2.4.2	Background statistical knowledge	14
2.4.3	Pareto optimality	18
2.4.4	Factorial experiments	19
2.4.5	Time course experiments	20
2.4.6	Limitations on available mRNA	23
2.5	Thesis outline	24
3	Pareto optimality for contrasts	27
3.1	Introduction	27
3.2	Motivating examples	28
3.2.1	Leukaemogenesis experiment	28
3.2.2	Sphingosine kinase experiment	31
3.3	Pareto optimality for contrasts	34
3.3.1	Linear models and contrasts	34
3.3.2	Pareto optimality	36
3.3.3	Constraints	37
3.3.4	Penalty Approach	38
3.4	Applications	39
3.4.1	Leukaemogenesis experiment	39
3.4.2	Sphingosine kinase experiment	46
3.5	Dye allocation	52
3.6	Concluding comments	56

<i>TABLE OF CONTENTS</i>	5
4 Pareto simulated annealing for microarray experiments	59
4.1 Introduction	59
4.2 Simulated annealing for microarray experiments	61
4.2.1 Introduction	61
4.2.2 Strategic concepts	62
4.2.3 Algorithm	64
4.2.4 Tuning the parameters	65
4.3 Pareto simulated annealing	67
4.3.1 Introduction	67
4.3.2 Strategic concepts	68
4.3.3 Role of weights	69
4.4 Core Pareto simulated annealing algorithm	70
4.5 Quality measures	72
4.5.1 Comparison with exact set of Pareto optimal designs	72
4.5.2 Random search	73
4.6 Tuning parameters for Pareto simulated annealing	76
4.6.1 Introduction	76
4.6.2 Central composite experimental plan	77
4.6.3 Analysis	78
4.6.4 Parameter selection algorithm	78
4.6.5 Adaptive Pareto simulated annealing algorithm	81
4.7 Application to the leukaemogenesis experiment	85
4.7.1 Background information	85
4.7.2 Parameter selection algorithm: 36 slides	85

4.7.3	Comparison of Pareto simulated annealing to random search methods	89
4.7.4	Adaptive Pareto simulated annealing algorithm	90
4.8	Concluding comments	102
5	Further comments	103
5.1	Introduction	103
5.2	Technical replication	103
5.3	Complex experiments	108
5.4	Concluding comments	112
A	Computer programs	113
B	R analysis	115
B.1	Preliminary PSA 36 slides, plan 1	116
B.2	Preliminary PSA 36 slides, plan 2	118
B.3	Preliminary PSA 36 slides, plan 2 with alternative quality measures	120
B.4	Comparison of Pareto simulated annealing to random search methods	124
B.5	Adaptive PSA 100 slides, plan 1	125
B.6	Adaptive PSA 100 slides, plan 2	127
B.7	Adaptive PSA 160 slides	129
B.8	Adaptive PSA 160 slides for alternative quality measures	131
	Bibliography	135

My PhD research focuses on the recommendation of optimal designs for two-colour microarray experiments. Two-colour microarrays are a technology used to investigate the behaviour of many thousands of genes in a single experiment. This technology has created the potential for making significant advances in the field of bioinformatics. Careful statistical design is crucial to realize the full potential of microarray technology. My research has focused on the recommendation of designs that are optimal in terms of precision for effects that are of scientific interest, making the most effective use of available resources. Based on statistical efficiency, the optimality criterion used is Pareto optimality. A design is defined to be Pareto optimal if there is no other design that leads to equal or greater precision for each effect of scientific interest and strictly greater precision for at least one. My PhD thesis was submitted in June and key aspects of my research are summarised below.

Pareto optimality enables the recommendation of designs that are particularly efficient for the effects that are of scientific interest. I have developed methodology to cater for effects of interest that correspond to contrasts rather than solely considering parameters of the statistical linear model. My approach also caters for additional experimental considerations such as contrasts that are of equal scientific interest. During my PhD, I have provided advice regarding the design of two-colour microarray experiments aimed at discovering the genetic basis of medical conditions.

For large experiments, it is not feasible to examine all possible designs in an exhaustive search for Pareto optimal designs. I have adapted the multiple objective metaheuristic method of Pareto simulated annealing to the microarray context. The aim of Pareto simulated annealing is to generate an approximation to the set of Pareto optimal designs in a relatively short time. At each iteration, a sample of generating designs is used to explore the design space in an efficient way. This involves the setting of a number of Pareto simulated annealing parameters and the development of appropriate quality measures. I have developed algorithms to search systematically for the optimal values of

the tuning parameters based on Pareto simulated annealing and response surface methodology.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopy.

Signed,

Date:

Acknowledgments

There are many people I would like to thank for assisting me to achieve my potential during my PhD research. Firstly, I thank my PhD supervisors, Andrew Metcalfe and Gary Glonek, for their collective guidance and support. I thank Andrew for his enthusiasm, inspiration, dedication and being a wonderful role model. I thank Gary for his ingenuity, innovative thinking, attention to detail and high aspirations. I am grateful to Anna Tsykin for her additional support and advice in many areas including the biological aspects relevant to my research.

I am appreciative of the School of Mathematical Sciences for providing a supportive environment. I thank Liz Cousins for her encouragement and assistance during my PhD, particularly as Postgraduate Co-ordinator. I am also grateful to Charles Pearce for the wisdom and experience he has shared over many years. I thank Patty Solomon for assistance in the earlier stages of my PhD.

I am grateful to Telstra for additional support during my PhD. I thank Bob Richter and the team for our positive and motivating discussions, advice and support.

In terms of accessing materials in alternative formats, I have been grateful for support from the Disability Service at the University of Adelaide, postgraduate students Jason Whyte, Brian Webby and Josephine Varney as well as the statistical researchers I have communicated with in Australia and overseas.

I am appreciative for the financial grants and scholarships that I have received during my PhD, namely the Australian Postgraduate Award, June Opie Fellowship, Jean Gilmore Bursary,

International Biometrics Society Travel Award, AMSI/ICE-EM Winter School Travel Grant, AMSI Summer Symposium Travel Grant, AMATA Bursary as well as additional support from the School of Mathematical Sciences and Telstra.

Finally, I thank my family and friends that I hold dear to me for sharing this experience with me. I thank my husband Phil for his commitment, strength, advice, unconditional support and always believing in me. I also thank my parents, Val, Josie and Peggy-Anne for their encouragement and support.

Publications arising from this thesis

Bennett, P. S., Glonek G. F. V. and Solomon, P. J. (2005). Optimal Designs for Gene Expression Studies. *Statistical Solutions to Modern Problems: Proceedings of the 20th International Workshop on Statistical Modelling, Sydney, 10-15 July*. University of Western Sydney Press, Sydney.

Sanchez, P. S., and Glonek G. F. V. (2009). Optimal designs for two-colour microarray experiments. *Biostatistics 2009* 10(3), 561-574, 2009.

Chapter 1

Introduction to biology and microarray technology

1.1 Introduction

This chapter provides an introduction to the biology and microarray technology relevant for studying the statistical issues to be presented later in this thesis.

1.2 Genes and DNA

As described by Nguyen et al. (2002), the biological processes involved in measuring gene expression can be thought of as “information transfer processes”. Genes contain information that is crucial for the creation of protein, which, in turn, is required for cell life processes. A typical gene is made up of deoxyribonucleic acid (DNA) and is found in the nucleus of a cell. DNA is a sequence of nucleotide molecules arranged in a double helix structure. Each nucleotide molecule consists of a phosphate, a sugar and a nitrogen base. The nitrogen base can be of four different varieties, namely adenine (A), guanine (G), cytosine (C) and thymine (T). The structure of the various bases is such that A is attracted to (or pairs with) T and G pairs with C. The structure of the molecules and the resulting attractions of the base pairs is the reason why DNA sequences are in the form of a double helix. The helix structure in DNA can be broken to create single strands, known as complementary DNA, when heat is applied. This is necessary for microarray experiments as complementary DNA is more stable than DNA.

Complementary DNA (cDNA), used in microarray experiments is synthesized using messenger RNA as a template. Messenger RNA is derived from protein coding genes and provides the sequence information needed to synthesize proteins. The measurements obtained from a microarray experiment provide a measure of the level of gene activity, commonly called gene expression, for the genes under consideration. The various steps involved in a microarray experiment are described in the next section.

1.3 Microarrays

1.3.1 Introduction

Microarray experiments can be of various types, including high-density nylon membrane arrays, short oligonucleotide (Affymetrix) arrays and two-colour microarrays (Glonek and Solomon, 2004). The class of two-colour microarrays consists of either spotted long oligonucleotide arrays or cDNA arrays. In a two-colour microarray experiment, the activity of genes in *two* biological samples is compared on each microarray slide.

In practice, two-colour microarrays have been commonly used and two-colour microarray experiments have been the focus of a substantial amount of research. It has been the case that two-colour microarrays have been of relatively low cost. This has provided the potential for more slides to be used for a given budget, for exploratory work and in situations where biological variability is high compared to that introduced due to using the arrays.

This thesis focuses on two-colour microarray experiments. Therefore, for simplicity, the reference to microarray experiments in this thesis actually refers to two-colour microarray experiments unless stated otherwise.

1.3.2 Steps involved in carrying out a two-colour microarray experiment

A typical two-colour microarray experiment is carried out as follows (based on Nguyen et al. (2002)).

1. Preparation of slides.

In a microarray experiment, cDNA sequences, or probes, are printed onto the various glass slides to be used. The cDNA is chosen in accordance with the aims of the experiment and the consideration of which genes could be potentially important. Each probe, or spot, represents a different gene under study. Given that there are tens of thousands of spots on each slide,

this technology allows for the inspection of tens of thousands of genes simultaneously. The cDNA probes are single-stranded. This facilitates the binding of samples to probes during the experiment.

2. Preparation of samples.

Each sample, to be applied to the slides, is prepared for use by obtaining cDNA from mRNA via the process of reverse transcription. The samples consist of biological material such as cultured cells or tissue from any living organism.

3. Labeling.

On a given microarray slide, the gene activity of two samples is to be compared. In order to differentiate between the two samples, each is labeled with a different fluorescent dye. Often the two choices are red (cy5) and green (cy3).

4. Hybridization.

For each microarray slide, the two labeled samples to be compared are mixed and applied to the slide. On a given slide, the red and green samples compete to bind to the cDNA on each spot. This is known as hybridization. The more binding that occurs, the higher the gene expression. In particular, for each spot, hybridization explores the relative gene activity for the red and green samples. Following hybridization, each slide is washed to remove any unbound cDNA.

5. Scanning the slides.

Each slide is scanned in order to obtain information relating to how much of each sample bound to the cDNA probe during hybridization. In doing so, each slide is scanned at two wavelengths, one to allow for the measurement of the intensity of the red dye and the other for the green.

6. Image processing.

At the image processing stage, the raw data, obtained from scanning, is used to extract numerical data of interest. In particular, the intensity of each spot on a given slide is measured as the ratio of intensities, red and green, from the two scanned images. In order to do so, the two images obtained from scanning undergo the tasks of gridding, segmentation and information extraction. Gridding involves overlaying the two scanned images and marking the location of each spot on the slide. Segmentation takes place to determine the area for each spot. For the scanned images, pixels within the spot area are foreground and those outside are background. Information about the intensities is extracted and summarised for both the foreground and background.

1.3.3 Changes in microarray technology

The microarray technology described in this thesis was current at the time that the research presented in this thesis began. Although the technology continues to progress, the statistical design aspects that this thesis addresses for two-colour microarray experiments are still relevant. An example of a more recent experiment that uses two-colour microarray technology, where microRNA is the biological material being measured rather than mRNA, can be found in Gregory et al. (2008).

1.4 Normalization

In a microarray experiment, a source of bias, called technology bias, arises due to the use of microarray technology. As described in Smyth et al. (2003), examples of technology bias are

- the red intensity measurements are typically lower than the green for samples that are expected to have equal intensities, and
- this imbalance is not usually constant across spots within and between arrays, depending on variables such as the location of a spot on the slide and overall spot intensity.

This means that the intensity measurements obtained at the image processing stage must be adjusted in order to account for any such bias. This is known as normalization. Methods of normalization can be carried out within each array or between arrays. Further details about such methods can be found in Smyth et al. (2003). Following normalization, normalized intensity measurements form the dataset for the experiment. In this thesis, it is assumed that intensity measurements are actually normalized intensity measurements unless stated otherwise.

1.5 Spot measurements

The key measurements, for each spot, in a microarray experiment are as follows.

- The intensity of the red sample, R .
- The intensity of the green sample, G .
- The log ratio of intensities, M , given by

$$M = \log_2(R/G) = \log_2 R - \log_2 G. \quad (1.1)$$

- The average log intensity, A , given by

$$A = \frac{1}{2} \log_2(R.G) = \frac{1}{2}(\log_2 R + \log_2 G). \quad (1.2)$$

In particular, M and A are measured on the log base 2 scale. This is convenient in terms of the interpretation of the values obtained. For example, for M ,

- if the red and green intensities are equal, then $M = 0$,
- if the red intensity is twice that of the green, then $M = 1$,
- if the red intensity is four times that of the green, then $M = 2$, and so on.

In the next chapter, the concepts introduced so far will be extended to incorporate statistical considerations.

Chapter 2

Review of statistical literature

2.1 Introduction

Research in the area of microarray experiments is relatively modern and of great interest to scientists, including biologists and statisticians. Statistical research has involved looking at the design and analysis of microarray experiments. In recent years, research specific to the design has emerged from statisticians to aid the effectiveness of the experiments. In particular, the goal of a statistician is to design experiments that will maximise the precision of the effects that are of interest to biologists (Yang and Speed, 2003). In order to achieve this, many aspects of the design of microarray experiments must be considered.

In this chapter, a review of the research that has been carried out by statisticians involved with the design of microarray experiments is provided. Firstly, some of the accepted good practices for carrying out the experiments will be put forward. This will be followed by a closer look at the statistical designs proposed by researchers, indicating where further research is needed. Finally, the thesis outline is provided.

2.2 Principles of experimental design

Issues central to the design choice, for a microarray experiment, include the scientific and physical considerations (Yang and Speed, 2003). The design must be chosen so that it will be able to answer the scientific questions of interest. In doing so, the physical constraints, such as the amount of mRNA available, must be taken into account also.

In the literature, there are two main approaches to the analysis of two-color microarray experiments, the normalization based approach and the ANOVA approach. In the normalization based approach, such as in Yang and Speed (2002), log-expression ratios are calculated and then normalized to remove extraneous variability as a pre-processing step. Statistical analyses are then applied to the normalized data. The alternative strategy for analysis is the ANOVA approach such as in Churchill (2004). The ANOVA approach does not necessarily have a separate normalization step but accounts for certain extraneous effects through the inclusion of appropriate effects in an analysis of variance model. (That model also forms the basis of the design aspects considered such as in Kerr and Churchill (2001).)

Although the two approaches to analysis appear quite different, many of the differences are, in fact, superficial as pointed out in Kerr (2003). This is illustrated in the somewhat simplified example that follows. Suppose the log-intensity of wildtype and mutant samples are measured from n slides. Denote the log-intensity by Y_{ij} where i is 1 for wildtype and 2 for mutant and j represents the slide number. The ANOVA approach fits the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij} \quad (2.1)$$

where μ is the overall mean, α_i represents the treatment of interest, β_j represents the random effect of slide j and E_{ij} are the random errors. The analysis provides estimates of the α_i together with their standard errors and also estimates of the variance between slides and the variance of the

errors. The approach adopted in this thesis is to consider the differences for each slide given by

$$M_j = Y_{1j} - Y_{2j} = \alpha_1 - \alpha_2 + E_{1j} - E_{2j} \quad (2.2)$$

and to assume normalization takes place. Kerr (2003) demonstrates equivalences between model formulations in more complex cases.

In the remainder of this thesis, the normalization based approach is adopted. In particular, it is assumed that normalization can be performed so that the normalized log-ratios then form the input for the main statistical analysis. The discussion of design and optimality are then premised on simple statistical models for the normalized log-ratios.

In the design of microarray experiments, a key statistical issue is replication. The purpose of replication is to improve the precision of estimates as well as to provide the basis for statistical inference (Yang and Speed, 2003). As described by Yang and Speed (2003), there are various forms of replication as follows. One form of replication is to have duplicate spots in adjacent positions on the slide. The purpose of this is to gain information about the quality of the measurements. Another form of replication is to use technical replicates whereby the samples used in different hybridizations originate from the same biological extraction but are labeled independently. Biological replicates, however, involve using samples from different extractions; either separate extractions from the same individual or extractions from various individuals. Biological replication allows for a wider generalization of conclusions whereas technical replication helps to reduce the variability of the measurements obtained in the experiment (Speed and Yang, 2002). The question of which replicates to hybridize on the same slide is another key consideration regarding replication and will be discussed further in sections that follow.

A recommended statistical principle for carrying out a microarray experiment is the use of dye-swapping (Yang and Speed, 2002), (Yang and Speed, 2003), (Churchill, 2002). The practice of

dye-swapping means that a hybridization, involving a particular configuration with two samples, is carried out twice; once with certain dye assignments and the other with the dye assignments swapped. This is undertaken in order to account for any dye bias that may arise due to the systematic differences between the red and green intensities associated with the dyes.

2.3 Early designs proposed for microarray experiments

One of the earliest designs proposed for a microarray experiment was the common reference design. This involves comparing samples of interest via a common reference sample. Such a design consists of the common reference sample appearing on each of the available slides hybridized together with each other treatment, one treatment per available slide. Various studies of the common reference design, compared to other choices, have shown that it is not the most efficient design for estimating the effects of interest (Glonek and Solomon, 2004), (Kerr and Churchill, 2001).

Another method for choosing a design, proposed by Kerr and Churchill (2001), is based on using concepts from classical experimental design as follows. The factors considered are variety (or treatment), gene, array (or slide) and dye. Models are set up to take into account the main effects of the factors and various interactions. Using the classical approach, A-optimality is presented as the criterion for recommending designs. In doing so, an objective function is set up to minimize the standard error of the various effects in the model under consideration.

Using the criterion of A-optimality, Kerr and Churchill (2001) demonstrate that their approach is superior to the common reference design for estimating the precision of parameter estimates. Recall, however, that certain effects may be of primary interest, such as particular interactions, in a microarray experiment. The classical approach, taking all effects into account without any preference, is not set up to specifically minimize the standard error of particular effects. This begs the question as to whether a criterion for design choice could be employed to specifically minimize the standard error of the effects of primary interest. An approach to do so is the topic of the next section.

2.4 Approach based on Pareto optimality

In the following sections, the methodology developed in Glonek and Solomon (2004), to recommend optimal designs for microarray experiments, is reviewed.

2.4.1 Design problem

Recall that the microarray technique was introduced in Section 1.3.2. In particular, it was explained that two samples are competitively hybridized on each microarray slide available for the experiment. Glonek and Solomon (2004) introduce the design problem by indicating that, from the point of view of the biologists conducting such experiments, two key questions to be addressed are as follows.

- Which samples should be competitively hybridized together on the same slide?
- How many times should each slide be replicated?

From the point of view of a statistician, those questions are addressed by identifying the biological effects that are of interest, adopting an appropriate parameterisation and choosing a design that optimises for those effects. In order to do so, the notion of Pareto optimality for parameters is introduced as an optimality criterion. This relies on the statistical background provided in the next section.

2.4.2 Background statistical knowledge

In the development of the methodology to find optimal designs for microarray experiments, some background statistical information is required. The relevant notation and concepts are introduced in terms of a typical factorial microarray experiment.

Notation and parameterisation

Factorial experiments involve the examination of various factors, with varying levels, in order to determine how they affect the response variable of interest. In the case of a microarray experiment, the response is the log ratio of intensity measurements for a particular gene. Note that, although there are thousands of genes involved in the experiment for which the design applies, analysis is carried out separately for each gene.

As an example, consider a 2×2 factorial microarray experiment. Let the factors of interest be A and B with levels $(0, a)$ and $(0, b)$ respectively. Following Glonek and Solomon (2004), the parameterisation to be adopted is given in Table 2.1.

The log intensity for the baseline (00), where both factors are at the low level, is represented by μ . The simple effect for factor A, with factor B held at the low level, is represented by the parameter α . This is found by subtracting the log intensity for (00) from the log intensity for (a0). In a similar way, the simple effect for factor B, with factor A held at the low level, is represented by the parameter β . The interaction between factors A and B is represented by the parameter $\alpha\beta$. This can be described as the effect on the log intensity due to the change in the level of factor A as the level of factor B changes also and is given by

$$\alpha\beta = (ab - 0b) - (a0 - 00). \quad (2.3)$$

In the context of factorial microarray experiments, interaction is typically of primary interest. For the 2×2 experiment, notice that the equation for interaction given above involves more than two samples, requiring the log intensities for all four treatment combinations. However, recall that cDNA microarray experiments involve the competitive hybridization of two samples on each slide available for the experiment. It follows that the estimation of the parameter $\alpha\beta$ must be carried out using more than one slide. This indicates that the choice of which samples to hybridize together

on each slide will have an impact on estimation.

Table 2.1: Parameterisation for a 2×2 factorial microarray experiment

Factor levels for A and B	(Expected) log intensity
00	μ
a0	$\mu + \alpha$
0b	$\mu + \beta$
ab	$\mu + \alpha + \beta + (\alpha\beta)$

For a 2×2 factorial microarray experiment, the parameter estimation is carried out using a number of slides from the set of all possible configurations. The set of configurations consists of each possible allocation of two distinct samples, obtained from Table 2.1, to a single slide. One sample will be labeled with red dye (cy5) and the other green (cy3). Parameter estimation follows from finding the log ratio of the intensities of the samples labeled with red and green dye; $\log(R/G)$. Following the presentation from Glonek and Solomon (2004), the set of possible configurations and associated expected log intensity ratios is given in Table 2.2.

Equivalently, the expected log intensity ratios, $E(\mathbf{M})$, can be given in matrix form as follows.

$$E(\mathbf{M}) = A\boldsymbol{\beta} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \alpha\beta \end{pmatrix}$$

where A is called the configuration matrix and $\boldsymbol{\beta}$ represents the parameter vector.

Table 2.2: Configurations and associated log ratios

Configuration			Expected
	Green	Red	Log Ratio
1	00	$a0$	α
2	00	$0b$	β
3	00	ab	$\alpha + \beta + (\alpha\beta)$
4	$0b$	ab	$\alpha + (\alpha\beta)$
5	$a0$	ab	$\beta + (\alpha\beta)$
6	$a0$	$0b$	$\beta - \alpha$

A design is specified by the configurations chosen for each of the slides available in the experiment. The design matrix, X is then constructed from the corresponding rows of the configuration matrix A .

Estimation

In this thesis, the analysis of two-colour microarray experiments is considered to be based on linear models of the form

$$E(\mathbf{M}) = \boldsymbol{\eta} = X\boldsymbol{\beta} \text{ and } \text{var}(\mathbf{M}) = \sigma^2 I \quad (2.4)$$

where \mathbf{M} is the vector of log intensity-ratios from all slides in the experiment, $\boldsymbol{\eta}$ is the vector of expected values, X is the design matrix and $\boldsymbol{\beta}$ is the parameter vector. The assumption of independent errors corresponds to the use of independent biological replication throughout the experiment, and applies throughout this thesis unless stated otherwise. The linear model is applied separately for each gene.

In the microarray context, inference is usually carried out separately for each of the parameters

or contrasts of interest. Typically, separate t -tests or moderated t -tests are used (Smyth, 2004). In the context of this thesis a contrast is defined as any linear combination of the expected values in the linear model (2.4). This corresponds to the usual definition of a contrast because for any linear combination of treatment differences, the coefficients for the individual treatment means will sum to zero.

Once a particular design is chosen, with design matrix X and corresponding expected log ratios, $E(\mathbf{M})$, the microarray experiment is carried out. Each slide corresponds to a particular configuration and hybridization carried out. The corresponding ratio of observed log intensities for all hybridizations carried out is measured to be \mathbf{m} . Using the measurements obtained, the following estimates can be calculated (Glonek and Solomon, 2004).

1. The estimates of the parameters, $\hat{\boldsymbol{\beta}}$ are found, using least squares, to be

$$(X^T X)^{-1} X^T \mathbf{m}. \quad (2.5)$$

2. The standard error for the estimator of the i th parameter is given by

$$\sigma \sqrt{c_i}, \quad (2.6)$$

where c_i is the i th diagonal element of the matrix $(X^T X)^{-1}$ and σ is the standard deviation between slides for a given gene.

The calculation of standard errors takes into account the design matrix, which relies on the configurations selected for the experiment.

2.4.3 Pareto optimality

Glonek and Solomon (2004) introduce the notion of admissible designs, herein referred to as Pareto optimal designs, to be those that are Pareto optimal with respect to individual parameters. In that

context, a design with a total of n slides and design matrix X is said to be Pareto optimal if there exists no other design with n slides and design matrix X_* such that

$$c_i \geq c_i^*$$

for all i with strict inequality for at least one i , where c_i, c_i^* are respectively the diagonal elements of $(X^T X)^{-1}$ and $(X_*^T X_*)^{-1}$ (Glonck and Solomon, 2004).

The application of Pareto optimality for parameters, presented in Glonek and Solomon (2004), is discussed below.

2.4.4 Factorial experiments

Recall that the notation and parameterisation for the 2×2 experimental situation was introduced in Section 2.4.2 based on Glonek and Solomon (2004). As an example, Glonek and Solomon (2004) consider such an experiment, with 6 slides, wherein Pareto optimality is applied to the parameters α , β and $\alpha\beta$. The set of Pareto optimal designs was found to consist of 21 designs. One such Pareto optimal design is specified by the allocation of two slides to each of configurations 1 and 2 and one slide allocated to each of configurations 4 and 5 from Table 2.2. This design is among those designs that are most efficient for the estimation of the interaction parameter $\alpha\beta$, whereby $c_{\alpha\beta} = 0.75$. Note that $c_\alpha = c_\beta = 0.42$. However, the reference design, specified by the allocation of 2 slides to configurations 1, 2 and 3, is not Pareto optimal, whereby $c_{\alpha\beta} = 1.5$ and $c_\alpha = c_\beta = 0.5$.

The notion of Pareto optimality can be applied to other types of factorial microarray experiments. To illustrate, Glonek and Solomon (2004) consider a 2×3 factorial microarray experiment with 10 slides available. They consider a simplified problem in which only certain configurations, suggested by results from the 2×2 case, are permitted. A suitable parameterisation is adopted and Pareto optimality is applied to the parameters that are of interest in that experiment.

2.4.5 Time course experiments

Glonek and Solomon (2004) consider the problem of finding optimal designs in the context of three situations for relatively small time course experiments. It is emphasized that design choice is driven by optimising for parameters that correspond to the effects of biological interest. For each situation, the effects of interest are identified and an appropriate parameterisation is adopted. Then Pareto optimality is applied to the parameters that correspond to the effects of interest.

Finally, the set of Pareto optimal designs is inspected in order to select an appropriate design for the experiment. As an example, Glonek and Solomon (2004) consider a time course microarray experiment with four time points, summarised below.

The 6 possible configurations for the time course experiment are given in Table 2.3. Furthermore, it is assumed that 6 slides are available whereby the total number of possible designs is 462.

Note that the total number of possible designs, say T , is calculated using a combinatorial formula that considers the allocation of n slides to r configurations so that

$$T = \binom{n+r-1}{r-1}. \quad (2.7)$$

Thus, for the time course experiment with $n = 6$ slides and $r = 6$ configurations, using Equation (2.7) it is found that

$$T = \binom{6+6-1}{6-1} = 462. \quad (2.8)$$

Time points compared to a baseline

This situation considers the case where it is of biological interest to compare each time point to a given baseline in terms of differential expression. For the time course experiment with four time points and 6 slides, the differential expression between the following pairs of time points is of

Table 2.3: Configurations for the time course experiment with four time points

1	Time 0	Time 1
2	Time 0	Time 2
3	Time 0	Time 3
4	Time 1	Time 2
5	Time 1	Time 3
6	Time 2	Time 3

interest:

- Time 0 and Time 1,
- Time 0 and Time 2, and
- Time 0 and Time 3.

When Pareto optimality is applied, the set of Pareto optimal designs is of size 44. Following inspection of the set, the recommended optimal designs are those that estimate the effects of interest with equal precision. Those designs are given by Design 1 and 2 in Table 2.4.

Adjacent time points

In this situation, biological interest is in the differential expression between adjacent time points. For the time course experiment with four time points and 6 slides, the differential expression between the following pairs of time points is of interest:

- Time 0 and Time 1,
- Time 1 and Time 2, and

Table 2.4: Time course experiments.

	Replication					
	Configuration					
Design	1	2	3	4	5	6
1	1	1	1	1	1	1
2	2	2	2	0	0	0
3	2	0	0	2	0	2

- Time 2 and Time 3.

When Pareto optimality is applied to the 6 slides experiment, the set of Pareto optimal designs is of size 36. Following inspection of the set, the recommended optimal designs are those that estimate the effects of interest with equal precision. Those designs are given by Design 1 and 3 in Table 2.4.

Time profiles

The third situation is when it is of interest to look at time profiles. For the time course experiment with four time points and 6 slides, suppose the time points are equally spaced and that linear and quadratic effects are of interest. When Pareto optimality is applied, the set of Pareto optimal designs is of size 71. Following inspection of the set, the recommended design is that which estimates the linear and quadratic effects with equal precision. This is given by Design 1 in Table 2.4.

Based on relatively small numbers of slides available for the time course experiment with four time points, Glonek and Solomon (2004) suggest that designs that allocate equal numbers of slides to all possible configurations may prove to be useful for this situation.

2.4.6 Limitations on available mRNA

For an experiment with n slides, the consideration of all possible designs assumes no limitations on the amount of mRNA available. However, consider the situation where the amount of available mRNA is limited such that it places limitations on the number of replicates available for certain treatment combinations and configurations. Consequently, it is not possible to construct all of the designs that would be possible in the unrestricted case. The notion of Pareto optimality can be applied in the context of considering a limited amount of mRNA available as follows (Glonck and Solomon, 2004). Firstly, it is suggested that the set of all possible designs is reduced to consist only of the subset of designs that can be constructed given the available mRNA. Following this, Pareto optimality is applied to the subset of designs. This is illustrated in Glonck and Solomon (2004) for a 2×2 experiment.

2.5 Thesis outline

Statisticians can play a crucial role at the design stage of microarray experiments to ensure the most effective allocation of available resources. As introduced by Glonek and Solomon (2004), Pareto optimality enables the recommendation of designs that are particularly efficient for the effects of most interest to biologists. This is relevant in the microarray context where analysis is typically carried out separately for those effects.

In Chapter 3, the Pareto optimal approach is developed to allow for effects of interest that correspond to contrasts rather than solely considering parameters of the linear model. Furthermore, the approach presented will cater for additional experimental considerations such as contrasts that are of equal scientific interest. This amounts to partitioning all relevant contrasts into subsets of effects that are of equal importance. Based on the partitions, a penalty will be employed in order to recommend optimal designs. Further to this, the issue of gene-specific dye bias will be addressed. The approach will be illustrated in studies of leukaemia and breast cancer.

To cater for situations where it is not feasible to carry out an enumerative search for Pareto optimal designs, a guided search algorithm will be developed in Chapter 4. In particular, the multiple objective metaheuristic method of Pareto simulated annealing will be adapted to the microarray context. The aim of Pareto simulated annealing is to generate an approximation to the set of Pareto optimal designs in a relatively short time. At each iteration, a sample of generating designs is used to explore the design space in an efficient way. This involves the setting of a number of Pareto simulated annealing parameters and the development of appropriate quality measures to assess their performance in the microarray context. Algorithms will be developed to search systematically for the optimal values of the tuning parameters based on Pareto simulated annealing and response surface methodology. The algorithms will be demonstrated in the context

of a factorial microarray experimental situation.

In Chapter 5, issues related to technical replication and other complex experiments are discussed and final comments are provided.

Chapter 3

Pareto optimality for contrasts

3.1 Introduction

Pareto optimality enables the recommendation of designs that are particularly efficient for the effects of most interest to biologists. This is relevant in the microarray context where analysis is typically carried out separately for those effects. In this chapter, the notion of Pareto optimal designs is extended to allow for effects of interest that correspond to contrasts of interest rather than solely considering parameters of the linear model. The approach is further developed to cater for additional experimental considerations such as contrasts that are of equal scientific interest. This amounts to partitioning all relevant contrasts into subsets of effects that are of equal importance. Based on the partitions, a penalty is employed in order to recommend designs for complex and varied microarray experiments. Finally, the issue of gene-specific dye bias is addressed. The approach will be illustrated in studies of leukaemia and breast cancer. An abridged version of this chapter was published in *Biostatistics* (Sanchez and Glonek, 2009).

3.2 Motivating examples

This section introduces two examples of typical experimental situations in the microarray context. The examples provide the motivation to extend the notion of Pareto optimality for parameters only to incorporate additional contrasts and appropriate constraints.

3.2.1 Leukaemogenesis experiment

The Child Health Research Institute in Adelaide has conducted microarray experiments to investigate the genetic basis of leukaemia. As introduced in Glonek and Solomon (2004), the 2×2 factorial microarray experiment aimed at discovering novel genes expressed in the leukaemic cell line V449E will be considered. The two experimental factors are cell line and time. The cell lines under study are the control, $FI(\Delta)$, and the leukaemic cell line V449E. To allow for the detection of differential expression between the two cell lines, the relevant time points are 0 and 24 hours. Interaction between cell line and time is of primary interest in this experiment. This is typically the effect of primary interest in the microarray context. Effects relating to cell line and time are of subsequent interest.

The parameterisation adopted for this experiment is given in Table 3.1. The cell line $FI(\Delta)$ at time 0, with log mean expression or intensity level μ , is defined as the baseline intensity. Table 3.1 is used to construct the expected log ratios given in Table 3.2. This represents the six possible hybridizations that arise from taking pairs of the four experimental conditions or samples that can be applied to a given microarray slide. The corresponding graphical representation is given in Figure 3.1.

The three parameters of the linear model correspond to the interaction parameter ($\alpha\beta$), the cell line parameter (α) that represents the change in expression of the two cell lines at time 0 and the time parameter (β) that represents the behaviour of the cell line $FI(\Delta)$ over time. Each of the

Table 3.1: Expression of a given gene in the 2×2 factorial experiment

Experimental Condition	Log Intensity
00	μ
a0	$\mu + \alpha$
0b	$\mu + \beta$
ab	$\mu + \alpha + \beta + \alpha\beta$

three parameters correspond directly to biological effects that are of interest in this experiment.

The Pareto optimality criterion introduced in Glonek and Solomon (2004) was used to optimise explicitly for the parameters in the linear model for mean expression levels. In this chapter it is argued that such a formulation does not capture some important considerations for design. It is shown that these can be accommodated by the introduction of additional contrasts and the imposition of appropriate constraints.

As an illustration, consider the 2×2 experimental situation. When interest is focused solely on the parameters of the linear model, the original approach is appropriate. However it will typically be the case that additional contrasts are of interest to the researcher. Of particular relevance is that such contrasts arise further to the consideration of interaction. In particular, in cases where the interaction parameter for a certain gene is found to be statistically significant, it is generally desirable to understand the nature of the interaction.

For the leukaemogenesis experiment, an interaction is said to be present for genes for which the change in expression from 0 to 24 hours is different for the two cell lines. For such genes it is also of interest to describe the patterns of expression over time. One instance is an interaction that

Table 3.2: Expected log ratio $M = \log(R/G)$

Configuration			Expected
	Green	Red	Log Ratio
1	00	a0	α
2	00	0b	β
3	00	ab	$\alpha + \beta + \alpha\beta$
4	0b	ab	$\alpha + \alpha\beta$
5	a0	ab	$\beta + \alpha\beta$
6	a0	0b	$\beta - \alpha$

occurs for a certain gene when its expression remains constant over time for $FI(\Delta)$ but increases for $V449E$. Such patterns of expression are most naturally investigated by considering the differences in expression over time for the two cell lines separately.

As defined in Table 3.1, the change in expression for $FI(\Delta)$ corresponds to the parameter β . However the change in expression for $V449E$ does not occur as a separate parameter. Instead, it may be obtained as a linear combination of other parameters in the model, given by the contrast $\beta + \alpha\beta$. Therefore the time effects, represented by the parameter β and the additional contrast $\beta + \alpha\beta$, are of subsequent interest and must be allowed for in the Pareto optimal approach. In a similar way, it can be argued that the cell line effects are of subsequent interest for genes for which an interaction is present. Therefore those effects, represented by the parameter α and the contrast $\alpha + \alpha\beta$, must be allowed for also. Incorporating additional contrasts of interest can be accommodated by optimising for their variances as well as those of the underlying parameters.

Optimising for all of the contrasts of interest separately would be expected to lead to a large

increase in the number of Pareto optimal designs. In addition, many of those designs would not be desirable because of imbalance for contrasts that are considered to be of equal importance. Here, it is proposed that such designs can be eliminated by forming subsets consisting of contrasts of equal importance and imposing constraints so that the variances for contrasts within each subset are constant. For example, in the leukaemogenesis experiment, consider the two time effects β and $\beta + \alpha\beta$. It can be argued that these effects should be of equal importance as follows. Firstly, they are equivalent in the sense that they both represent the behaviour of a given cell line over time. Secondly, those types of simple effects for a given factor are typically of equal interest to biologists. Therefore it is relevant to consider the optimality of designs for which the effects are of equal precision, whereby

$$\text{var}(\hat{\beta}) = \text{var}(\hat{\beta} + \hat{\alpha}\beta). \quad (3.1)$$

Similarly, for cell line effects, the optimality of designs for which

$$\text{var}(\hat{\alpha}) = \text{var}(\hat{\alpha} + \hat{\alpha}\beta) \quad (3.2)$$

will be considered. The approach based on the notion of Pareto optimality presented in this chapter will therefore take into account all relevant contrasts and constraints in order to recommend optimal designs.

3.2.2 Sphingosine kinase experiment

The Hanson Institute in South Australia has conducted research to discover genes associated with the dysregulation of sphingosine kinase in the study of breast cancer. Sphingosine kinase is an enzyme that plays an important role in the regulation of cell proliferation, survival and cell death. There are various forms of sphingosine kinase that perform this role. Some forms act to stimulate cell growth whereas others inhibit proliferation. In a cell, when the levels of various forms of

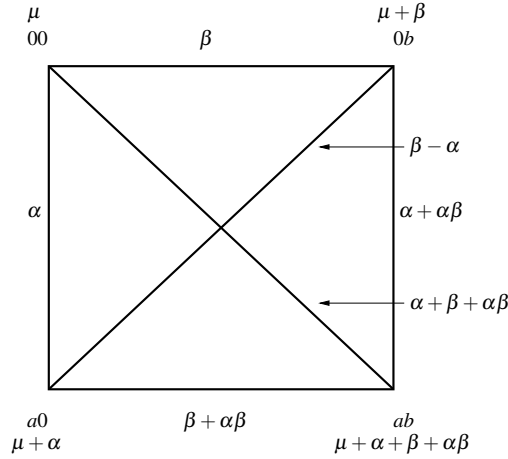


Figure 3.1: Parameterisation and hybridizations for 2×2 leukaemogenesis experiment.

sphingosine kinase are unbalanced, disregulation can lead to diseases such as cancer. In the study of breast cancer, estrogen is relevant also.

In particular, a 2×3 factorial microarray experiment to investigate the disregulation of sphingosine kinase in the study of breast cancer will be considered. The two factors are cell type and estrogen. Cell type is related to sphingosine kinase production and the three levels are the control (0), active mutant (a) and switched-off mutant (b). In addition, estrogen is required to stimulate a response and its levels are absent (0) and present (g). The various interactions between cell type and estrogen are of primary interest as they relate to the disregulation of sphingosine kinase. In addition, simple effects involved with cell types and estrogen are of secondary interest.

The parameterisation adopted is represented in Figure 3.2. The baseline parameter μ represents the control cell type in the absence of estrogen. The parameter γ represents the behaviour of the control in the absence and presence of estrogen, α the change in expression between the active and control cell types when estrogen is absent and β the change in expression between the switched-off and control cell types when estrogen is absent. The two parameters that relate to interactions are

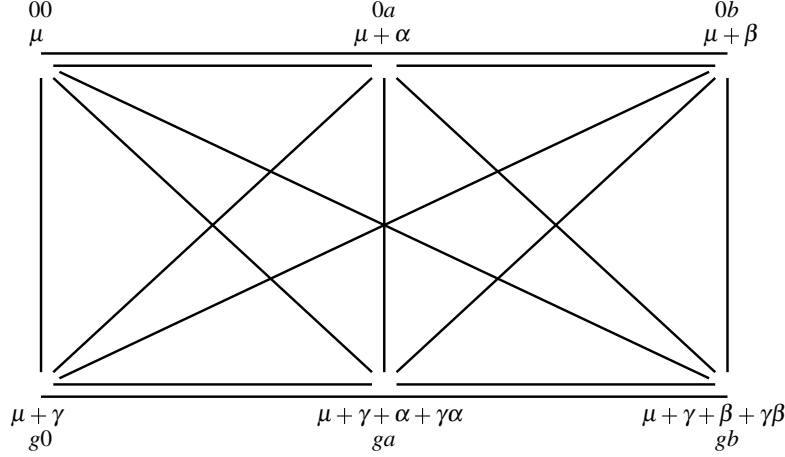


Figure 3.2: Parameterisation and hybridizations for 2×3 sphingosine kinase experiment.

$\gamma\alpha$, the change in expression between the active and control cell types in the absence and presence of estrogen, and $\gamma\beta$, the change in expression between the switched-off and control cell types in the absence and presence of estrogen. In terms of differences in log intensities, $\gamma\alpha$ corresponds to $(ga - 0a) - (g0 - 00)$ and $\gamma\beta$ to $(gb - 0b) - (g0 - 00)$.

As is typically the case in the microarray context, interactions are of primary interest in the sphingosine kinase experiment. However, not all interactions can be represented by parameters of the linear model. Recall that two interactions are represented by the parameters $\gamma\alpha$ and $\gamma\beta$. The third interaction, the change in expression between the active and switched-off cell types in the absence and presence of estrogen, corresponds to $(gb - 0b) - (ga - 0a)$. This can be shown to be represented by the contrast $\gamma\beta - \gamma\alpha$ rather than an independent parameter. Therefore interactions form a subset of contrasts that are not linearly independent and an approach that allows for contrasts is necessary. Secondly, in a similar way to the 2×2 experiment, the simple effects relating to cell type and estrogen can be partitioned into subsets of contrasts of secondary interest. To summarise, the model involves five parameters. However there are twelve contrasts of

interest and these are partitioned into three subsets;

- interactions $(\gamma\alpha, \gamma\beta, \gamma\beta - \gamma\alpha)$,
- estrogen $(\gamma, \gamma + \gamma\alpha, \gamma + \gamma\beta)$ and
- cell type $(\alpha, \beta, \alpha + \gamma\alpha, \beta + \gamma\beta, \beta - \alpha, \beta + \gamma\beta - \alpha - \gamma\alpha)$.

In addition, it is appropriate to develop constraints based on effects that are of equal importance.

In the sphingosine kinase experiment, the interactions are of equal importance in the following sense. Each represents the change in expression between two cell types in the absence and presence of estrogen. Furthermore, those effects are typically of equal interest to biologists conducting such experiments.

Similarly, within each of the subsets for estrogen and cell type, the contrasts are of equal importance. The identification of optimal designs for which variances of relevant contrasts are most constant within each of the subsets is desired and will be considered.

3.3 Pareto optimality for contrasts

Motivated by the two examples in Section 3.2, the specification and estimation of parameters and contrasts are now introduced.

3.3.1 Linear models and contrasts

In a two-colour microarray experiment, a hybridization or configuration consists of the allocation of two treatment combinations to a given slide. The expected log ratio is obtained as the difference in expected log intensity for the two treatment combinations. The set of configurations consists of all possible hybridizations for the experiment. The linear model for the set of configurations can

be represented by

$$\boldsymbol{\theta} = A\boldsymbol{\beta} \quad (3.3)$$

where $\boldsymbol{\theta}$ is the vector of expected log ratios, A is the configuration matrix and $\boldsymbol{\beta}$ is the parameter vector.

The linear model given in (3.3) can be parameterised in many different but equivalent ways depending on the choice of basis for the underlying model space \mathcal{M} . Potential ambiguity is avoided by considering contrasts that are linear functions of the expected log ratios for the set of configurations. Using the linear model (3.3), a set of contrasts, say $\boldsymbol{\gamma}$, can be represented by a set of linear functions of $\boldsymbol{\theta}$. Let the i -th contrast be given by

$$\gamma_i = \mathbf{b}_i^T \boldsymbol{\theta} \quad (3.4)$$

where \mathbf{b}_i is a suitable vector of coefficients. This specification can be shown to be unambiguous in the sense that the meaning of the contrasts does not depend on the parameterisation chosen.

For computational purposes, a parameterisation must be chosen and, assuming that one is chosen, the vector of contrasts can be expressed as

$$\boldsymbol{\gamma} = B\boldsymbol{\beta} \quad (3.5)$$

where B is an appropriate matrix.

In a particular experiment, the model matrix X is obtained by selecting the appropriate rows from A according to the hybridizations actually performed. The best linear unbiased estimate for the parameter vector is $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{M}$ with

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \text{ and } \text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (X^T X)^{-1},$$

see, for example, Searle (1971). The best linear unbiased estimate for the contrast vector is then

$$\hat{\boldsymbol{\gamma}} = B\hat{\boldsymbol{\beta}} = B(X^T X)^{-1} X^T \mathbf{M} \quad (3.6)$$

and the corresponding variance matrix is given by

$$\text{var}(\hat{\gamma}) = B \text{var}(\hat{\beta}) B^T = \sigma^2 B (X^T X)^{-1} B^T. \quad (3.7)$$

For example, for the leukaemogenesis experiment the contrasts are represented as

$$\gamma = \begin{pmatrix} \alpha \\ \beta \\ \alpha\beta \\ \alpha + \alpha\beta \\ \beta + \alpha\beta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \alpha\beta \end{pmatrix} = B\beta$$

3.3.2 Pareto optimality

Under the linear model $E(\mathbf{M}) = X\beta$, the variances of the parameter estimates are given by $\text{var}(\hat{\beta}_i) = \sigma^2 c_i$ where c_i is the i -th diagonal element of $(X^T X)^{-1}$. Glonek and Solomon (2004) defined Pareto optimal designs with respect to the diagonal elements of $(X^T X)^{-1}$. That is, with respect to the variances of the parameters in the linear model (2.4). In this thesis, that definition is extended to include the variances of the contrasts of interest, $\hat{\gamma}$ as follows.

Definition 1. *Pareto optimality*

A design with a total of n slides and design matrix X is said to be Pareto optimal for the contrasts $\gamma = B\beta$ if there exists no other design with n slides and design matrix X_ such that*

$$c_i^* \leq c_i, \quad (3.8)$$

for all i and

$$c_i^* < c_i \quad (3.9)$$

for at least one i . The variance coefficients c_i and c_i^* are the diagonal elements of $B(X^T X)^{-1} B^T$ and $B(X_*^T X_*)^{-1} B^T$ respectively.

3.3.3 Constraints

The inclusion of contrasts, in addition to the parameters of the linear model, is necessary to consider all effects of interest in the experiment. However, applying Pareto optimality to each relevant contrast separately would lead to a proliferation in the number of Pareto optimal designs. As described in Section 3.2, it is proposed that this difficulty is overcome by partitioning all relevant contrasts into subsets of effects that are of equal importance. The partitioning is based on both scientific interest and quantities that are equivalent in terms of their interpretation. For example, recall that for the leukaemogenesis experiment, the two subsets are those for cell line $(\alpha, \alpha + \alpha\beta)$ and time $(\beta, \beta + \alpha\beta)$. Thus efficient designs are sought for which variances are constant within the subsets.

Firstly, to assist with enforcing the constraints, using the traces of variance coefficients for each subset is proposed rather than applying Pareto optimality to contrasts separately. For the leukaemogenesis experiment, this would lead to

$$t_A = c_\alpha + c_{\alpha+\alpha\beta} \quad (3.10)$$

for cell line and

$$t_B = c_\beta + c_{\beta+\alpha\beta} \quad (3.11)$$

for time. Thus the objectives for the leukaemogenesis experiment are the traces, described above, as well as the variance coefficient for the single interaction term given by

$$t_{AB} = c_{\alpha\beta}. \quad (3.12)$$

Some rationale for constructing traces based on variances for each relevant subset of contrasts follows from what is found in the case of a single factor experiment. For a single factor, it is found that equality of variances for contrasts occurs when the design corresponds to that with minimum trace for the variance coefficients of those contrasts. However, in more general settings, minimising the trace does not guarantee equality of the variances and it is necessary to impose any such constraints explicitly.

A second measure to explore equality within subsets of equal importance is to develop constraints for the variances of relevant contrasts. For example, for the leukaemogenesis experiment, equality would be achieved for cell line and time if designs adhere to the constraints given in Equations (3.1) and (3.2). However, enforcing constraints exactly can be unduly restrictive for discrete optimisation problems. For example, there may be situations where none of the Pareto optimal designs satisfy the constraints. In extreme cases, it may happen that there are no designs at all that satisfy the constraints. The imposition of constraints may also be problematic from a computational perspective in situations where the set of all designs that satisfy a given set of constraints may be difficult to construct. As outlined in the next section, a flexible approach that avoids these difficulties is to incorporate a suitable penalty function based on the constraints.

3.3.4 Penalty Approach

In order to incorporate constraints based on subsets of equal importance, a suitable penalty function, D , is constructed based on the differences in variance coefficients of contrasts that appear in the constraints. The penalty approach takes the i -th penalized objective to be

$$t_i^{(D)} = (1 - w) * t_i + w * D \quad (3.13)$$

where $w \in [0, 1]$ is the weight associated with the penalty and t_i is the i -th objective. Then the Pareto optimal set is generated using the penalized objectives.

For example, for the leukaemogenesis experiment, the contrasts of equal interest correspond to the constraints given in equations (3.1) and (3.2) given in Section 3.2.1. Based on those constraints, the appropriate penalty function is given by

$$D = (c_\alpha - c_{\alpha+\alpha\beta})^2 + (c_\beta - c_{\beta+\alpha\beta})^2. \quad (3.14)$$

Then Equation (3.13) is applied to each of the objectives t_A , t_B and t_{AB} as outlined in Section 3.3.3 by substituting the value obtained from Equation (3.14). Then the Pareto optimal set is generated for the penalized objectives $t_A^{(D)}$, $t_B^{(D)}$ and $t_{AB}^{(D)}$.

An attractive feature of the penalty approach is its simplicity with regard to implementation. In addition, the Pareto optimal set generated has the potential to be substantially smaller than would arise from the original approach by Glonek and Solomon (2004). This is of practical importance when selecting an appropriate design. Furthermore, employing a penalty extends naturally to more complex applications for which metaheuristics may be required rather than the enumerative approach used in this chapter.

3.4 Applications

3.4.1 Leukaemogenesis experiment

Recall, from Section 3.2.1, that there are five contrasts of interest for the leukaemogenesis experiment. The interaction between cell line and time is of primary interest and forms the objective $t_{AB} = c_{\alpha\beta}$. Of subsequent interest are the two subsets of contrasts relating to cell line and time. As shown in Section 3.3.3, the traces of variance coefficients for contrasts within each of the subsets, t_A and t_B , form appropriate objectives. Thus the application of Pareto optimality involves the three objectives t_A , t_B and t_{AB} . Furthermore, within each of the subsets for cell line and time, the squared difference in variance coefficients for the two contrasts of equal interest form the basis for

the penalty function, given in Equation (3.14), to be employed in the penalty approach.

Suppose 16 slides are available for the experiment. Then a total of 20,349 designs are possible. Application of the penalty approach with $w \geq 0.975$ produces a set of 16 Pareto optimal designs. These are given by designs 1–16 in Table 3.3. Each of the 16 Pareto optimal designs satisfies the constraints in this situation. Furthermore, each appears in the larger Pareto optimal set of size 132 obtained when Pareto optimality is applied to the three objectives with no penalty.

The minimum variance coefficient for the interaction parameter within the constrained set of 16 designs is equal to the unconstrained minimum. This corresponds to the first 7 designs in Table 3.3. The first design, shown in Figure 3.3, is recommended as it is most efficient for interaction and balanced for all other contrasts of interest. Designs 2–7 would also prove to be useful in situations where the efficiency of subsets of contrasts are not of equal priority.

The approach based on penalised traces of variance coefficients of relevant contrasts presented in this chapter eliminates undesirable designs that would be obtained by applying the definition of Pareto optimal designs, given in Glonek and Solomon (2004), to parameters only. In particular, when there are 16 slides available, applying Pareto optimality to parameters leads to a total of 391 designs. Of those designs, some are useful but many are inefficient for the estimation of interaction and highly unbalanced within the subsets of equal importance. For example, Design 17 in Table 3.3, is highly unbalanced with respect to the variances of the cell line contrasts whereby $c_\alpha = 0.071$ and $c_{\alpha+\alpha\beta} = 1.071$ and would not usually be considered to be a useful design.

Based on three penalized objectives, the penalty approach controls how much emphasis is placed on the constraints for the subsets of equal importance. As the penalty weight increases, more emphasis is placed on obtaining designs that best satisfy the constraints. Application of the penalty approach with a relatively high weight results in a manageable number of efficient designs to consider for design selection. For example, when 16 slides are available, using a penalty weight

$w \geq 0.975$ leads to the set of 16 Pareto optimal designs that satisfy the constraints. Inspection of these designs suggests that application of the penalty does not result in the exclusion of any desirable designs when compared to the set of 132 designs obtained when no penalty is employed. The improvement in manageability due to using penalized objectives based on traces of variance coefficients for subsets for moderately sized experiments is further illustrated in Table 3.4.

Now compare the results presented here to what can be derived from the classical approach of partial confounding given in Yates (1935). In the microarray context, slides represent blocks of size 2. Under the classical approach, treatment means are usually parameterised in terms of a set of orthogonal contrasts. For the leukaemogenesis experiment, such contrasts represent the cell line main effect, α' , the time main effect, β' , and interaction, $\alpha\beta'$, together with a grand mean term. The classical contrasts can be expressed in terms of the parameterisation in Table 3.1 whereby

$$\alpha' = (\alpha + 0.5\alpha\beta)/2, \quad \beta' = (\beta + 0.5\alpha\beta)/2, \quad \text{and} \quad \alpha\beta' = \alpha\beta/4.$$

Consider the pairs of blocks

$$A = \{(00, a0), (0b, ab)\}, B = \{(00, 0b), (a0, ab)\} \text{ and } C = \{(00, a0), (a0, 0b)\}.$$

Within a single replicate of pair A , both the cell line main effect α' and interaction $\alpha\beta'$ are estimable. However the time main effect β' is completely confounded with the block effects. Similarly, the cell line main effect α' is completely confounded within B and the interaction $\alpha\beta'$ is completely confounded within C . Therefore any design consisting only of replicates from pairs A and B will be optimal for interaction and this is consistent with the results that have been obtained here. On the other hand, pair C is most efficient for the estimation of the classical main effects but does not contain information for interaction or any of the other contrasts that are of interest in this experiment. It follows that the diagonal hybridizations in pair C should not appear in a design optimal for any one of the contrasts α , $\alpha + \alpha\beta$, β , $\beta + \alpha\beta$ and $\alpha\beta$, that is, in designs 1, 2 and 3

in Table 3.3. Moreover, the Pareto Optimal designs which include diagonal hybridizations largely result in a loss in efficiency with respect to the contrasts of interest. For example, comparison of designs 1 and 16 in Table 3.3 shows a 33% increase in $c_{\alpha\beta}$ whereas there is an improvement in efficiency of only 2.7% for each of c_α , c_β , $c_{\alpha+\alpha\beta}$ and $c_{\beta+\alpha\beta}$. Thus employing designs with diagonal hybridizations is not recommended when interest is focused on the contrasts considered here.

In the illustration of this experimental situation, an even number of slides was considered to be available. When various odd numbers of slides are considered, it is found that enforcing the constraints would lead to designs that allocate slides to diagonal hybridizations. However, the weight in the penalty approach can be lowered to consider designs that employ no diagonals. Among the resulting designs, those that are recommended provide an improvement in the efficiency of interaction, despite their slight imbalance for other subsets of equal importance. These designs are simple variations of the types of designs obtained for even numbers of slides. For example, for 15 slides, any of the 4 designs obtained by removing one slide from any of the 4 configurations employed in Figure 3.3 are recommended.

A wide range of weights were investigated for the penalty approach for moderately sized experiments. When the penalty weight was relatively high, it was found that the Pareto optimal set obtained was most manageable and contained efficient designs. Based on experience, then, an effective strategy would be to have an initial weight close to 1. Subsequently the weight could be lowered until it is found that the Pareto optimal set provides no further designs that are considered to be suitable for recommendation.

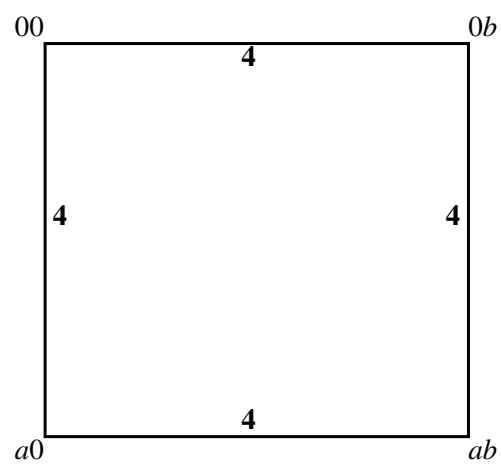


Figure 3.3: Recommended Pareto optimal design for leukaemogenesis experiment with 16 slides.

Table 3.3: Designs for leukaemogenesis experiment with 16 slides

Design	Replication for configurations (configurations as in Table 3.2)										
	1	2	3	4	5	6	c_α	c_β	$c_{\alpha\beta}$	$c_{\alpha+\alpha\beta}$	$c_{\beta+\alpha\beta}$
1	4	4	0	4	4	0	0.188	0.188	0.25	0.188	0.188
2	7	1	0	7	1	0	0.134	0.563	0.25	0.134	0.563
3	1	7	0	1	7	0	0.563	0.134	0.25	0.563	0.134
4	6	2	0	6	2	0	0.146	0.313	0.25	0.146	0.313
5	2	6	0	2	6	0	0.313	0.146	0.25	0.313	0.146
6	5	3	0	5	3	0	0.163	0.229	0.25	0.163	0.229
7	3	5	0	3	5	0	0.229	0.163	0.25	0.229	0.163
8	6	1	1	6	1	1	0.143	0.321	0.286	0.143	0.321
9	1	6	1	1	6	1	0.321	0.143	0.286	0.321	0.143
10	5	2	1	5	2	1	0.155	0.238	0.286	0.155	0.238
11	2	5	1	2	5	1	0.238	0.155	0.286	0.238	0.155
12	4	3	1	4	3	1	0.171	0.196	0.285	0.171	0.196
13	3	4	1	3	4	1	0.196	0.171	0.286	0.196	0.171
14	4	2	2	4	2	2	0.167	0.208	0.333	0.167	0.208
15	2	4	2	2	4	2	0.208	0.167	0.333	0.208	0.167
16	3	3	2	3	3	2	0.183	0.183	0.333	0.183	0.183
17	14	1	0	1	0	0	0.071	1	1.071	1	2.071

Table 3.4: Manageability for leukaemogenesis experiment

Number of slides	Total designs	Pareto optimal ($w = 0$)	Penalty approach	Associated weight
8	1,287	29	5	0.875
12	6,188	77	9	0.94
16	20,349	132	16	0.975
20	53,130	202	24	0.985
24	118,755	320	32	0.99

3.4.2 Sphingosine kinase experiment

In Section 3.2.2, twelve contrasts of interest were identified for the sphingosine kinase experiment. In particular, all relevant contrasts were partitioned into three subsets, namely those for interactions, cell type and estrogen. Within each subset, the contrasts represent effects of equal importance. Based on Section 3.3.3, three objectives are constructed corresponding to the traces of variance coefficients for contrasts within each of the subsets. The penalty function, to be applied in the penalty approach, is formed in a similar way to that developed for the leukaemogenesis experiment. For each subset, the sum of squared differences in variance coefficients over all pairs of contrasts is calculated. Then the penalty function is based on the total sum of squares over all subsets such that each is represented with equal emphasis.

Suppose 15 slides are available for the experiment. Then there are 77,558,760 possible designs. Application of the penalty approach with $w \geq 0.8$ produces a set of 8 designs given by designs 1–8 in Table 3.5. Each of the 8 designs satisfy the constraints in this situation. Furthermore, the first 7 designs are Pareto optimal if the constraint is removed.

Design 1, also given in Figure 3.4, provides minimum variances for interactions within the constrained set. This design allocates relatively more slides to hybridizations involving effects for estrogen compared to those for cell type. However Design 2, given in Figure 3.5, allocates relatively fewer slides to hybridizations involving estrogen. Such a design places greater emphasis on the cell type effects but also leads to a loss of efficiency for the interactions. This is further illustrated by inspecting other designs in Table 3.5. Thus, unlike the balance between effects for factors observed for the leukaemogenesis experiment, it is found that a preference for hybridizations involving the two-level factor estrogen contributes towards higher precision for interactions. Note however, it may not be advisable to recommend designs that allocate almost all available slides to those hybridizations for larger experiments. Such designs would result in increasingly poor precision

for cell type effects, relative to those for estrogen and interactions, as the number of available slides increases. For a given experiment, the penalty approach allows for exploring the compromise for the subsets of effects. This facilitates the selection of an appropriate design.

As was the case for the leukaemogenesis experiment, it is found that diagonal hybridizations are not efficient for the estimation of interactions. Consequently, designs that do not employ diagonals are recommended compared to some well-known alternative designs. Firstly, consider the all-pairs Design 7 given in Table 3.5. This would be the only design obtained under a classical approach such as D-optimality. Now compare Design 7, that employs 6 diagonals, to Design 1 that employs none. Design 1 is highly efficient, providing a 33% reduction in variances for interactions over the all-pairs design. Now consider the reference design, 9, whereby all samples are compared to the baseline. This design is highly unbalanced within the three subsets of equal importance and employs 6 diagonal hybridizations. Furthermore, the reference design is not found to be Pareto optimal when Pareto optimality is applied to the three objectives. Design 1 provides an improvement in the efficiency of interactions by over 56% over the reference design. The examples demonstrate the advantages of the Pareto optimal approach for design selection presented in this chapter compared to reference designs and classical alternatives.

Now consider Pareto optimality for parameters only as presented in Glonek and Solomon (2004). Applying that approach to the present experiment, Design 10 is found to be Pareto optimal and it deems Design 1 not to be. Although Design 10 is efficient for parameters, it is highly unbalanced within the subsets of equal importance. However, Design 1 is balanced for each subset. Furthermore, Design 1 has a trace for variances of all interactions that is lower than that for Design 10. It is the approach that optimises for appropriate subsets of contrasts that finds Design 1 to be Pareto optimal, facilitating its recommendation. This illustrates that the approach presented in this chapter is preferable to optimising for parameters only.

Application of the penalty approach with a high weight for 15 slides leads to optimal designs that provide balance within the subsets of equal importance. A wider examination of this experimental situation indicates that this is typically the case when the number of available slides is a multiple of 3. When the number of slides is not a multiple of 3, however, it can be found that no designs lead to balance within each of the subsets. Of the possible designs, those that provide most balance are inefficient for interactions as they primarily employ diagonal hybridizations. These are the designs obtained when the penalty approach is applied with a high weight. Lowering the weight leads to consideration of designs that employ no diagonals and are efficient for interactions. For example, when 16 slides are available, the 3 designs obtained by adding one slide to any of the vertical configurations in Figure 3.4 are recommended.

This experimental situation has been considered for moderate numbers of slides available. Compared to the total number of designs possible, applying Pareto optimality to the three objectives leads to a manageable number of efficient designs. The dramatic fall in the number of designs can be observed in Table 3.6. After exploring the range of weights possible, the following strategy for design selection is suggested. Initially, set the weight close to 1. Subsequently, lower the weight until the Pareto optimal set obtained provides no further designs that would be appropriate for recommendation.

Table 3.5: Designs for sphingosine kinase experiment with 15 slides. The horizontal, vertical and diagonal configurations correspond to those given in Figure 3.4. The variance coefficient, or range of variance coefficients, of contrasts within subsets is presented.

Design	Interactions	Estrogen	Cell type	Verticals	Horizontals	Diagonals
1	0.444	0.259	0.444	3 3 3	1 1 1 1 1 1	0 0 0 0 0 0
2	0.5	0.5	0.292	1 1 1	2 2 2 2 2 2	0 0 0 0 0 0
3	0.5	0.233	0.833	4 4 4	0 0 0 0 0 0	1 0 1 0 1 0
4	0.5	0.233	0.833	4 4 4	0 0 0 0 0 0	0 1 0 1 0 1
5	0.546	0.293	0.364	2 2 2	1 1 1 1 1 1	1 0 1 0 1 0
6	0.546	0.293	0.364	2 2 2	1 1 1 1 1 1	0 1 0 1 0 1
7	0.667	0.333	0.333	1 1 1	1 1 1 1 1 1	1 1 1 1 1 1
8	0.571	0.257	0.476	3 3 3	0 0 0 0 0 0	1 1 1 1 1 1
9	1–1.333	0.333–0.667	0.333–0.667	3 0 0	3 0 3 0 0 0	3 3 0 0 0 0
10	0.411–0.571	0.171–0.289	0.603–1.143	5 3 3	1 1 1 1 0 0	0 0 0 0 0 0

Table 3.6: Manageability for sphingosine kinase experiment

Number of slides	Total designs	Pareto optimal ($w = 0$)	Penalty approach	Associated weight
9	817,190	145	4	0.275
12	9,657,700	306	6	0.675
15	77,558,760	836	8	0.8
18	471,435,600	1,474	10	0.875

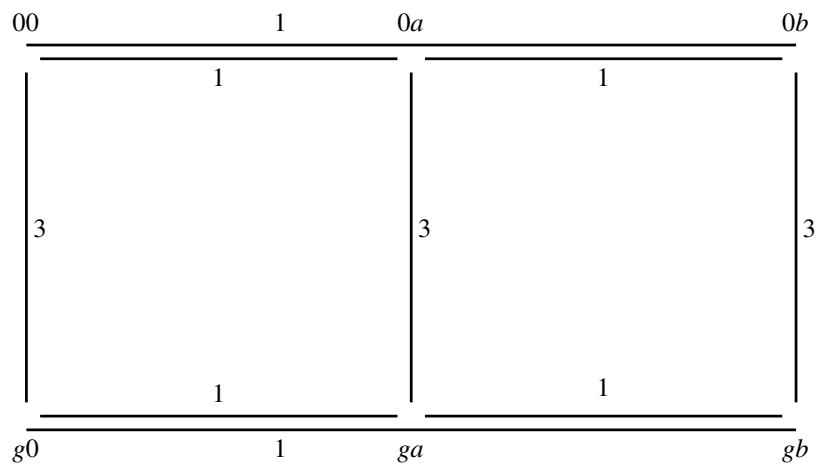


Figure 3.4: Recommended Pareto optimal design for sphingosine kinase experiment with 15 slides.

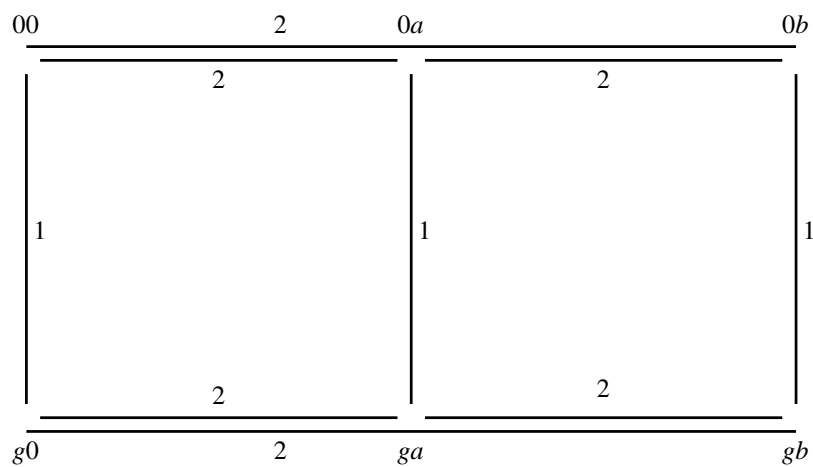


Figure 3.5: Another example of a Pareto optimal design for the sphingosine kinase experiment with 15 slides.

3.5 Dye allocation

In two-colour microarray experiments, dye bias may occur in a variety of contexts. Systematic dye biases are those for which one channel appears consistently at a higher intensity for all genes on the array. Such dye bias can also depend on the overall spot intensity and the spatial location within an array but can typically be removed by normalization, see for example Yang et al. (2002). Gene-specific or feature-specific dye bias occurs when the bias is different for different genes, indicating the presence of dye by gene interactions. This type of dye bias cannot be removed by normalization.

As explained in Dobbin et al. (2005) and Martin-Magniette et al. (2005), if gene-specific dye bias is not taken into account during the design and analysis of the experiment, it can affect conclusions regarding differential expression. In terms of design selection, Dobbin et al. (2005) and Martin-Magniette et al. (2005) suggest that a variety of considerations be taken into account. Designs that employ dye-swaps can effectively remove gene-specific dye bias. However it is not clear that they would necessarily be the most efficient with respect to estimation of the contrasts of interest. Indeed, it may happen in some cases that dye-swapped experiments are not feasible to carry out. Another approach, suggested by Tsai et al. (2006), is to seek designs that are A-optimal for the parameters of interest in the context of a linear model that includes a dye effect.

The approach to statistical design presented in this chapter is focused primarily on achieving high efficiency for the contrasts of scientific interest. In this light, a possible strategy is to choose a design based solely on considerations of efficiency. A suitable dye allocation can then be imposed on the chosen design to minimise the effect of any dye biases that may be present.

Suppose that the design, X , has been selected for the experiment on the basis of Pareto optimality. Recall that the rows of X correspond to the hybridizations as specified by the configuration matrix (3.3). This formulation assumes a given dye orientation for each possible hybridization.

Reversal of the dye allocation for a given slide is achieved by multiplication of the corresponding row of X by -1 . This operation does not affect $X^T X$ hence reversals of dye allocation do not impact on efficiency.

Now suppose that dye effects, arising from gene-specific dye bias, are incorporated into the linear model introduced in (2.4) so that

$$E(\mathbf{M}) = \boldsymbol{\eta} + \delta \mathbf{1} = X\boldsymbol{\beta} + \delta \mathbf{1} \quad (3.15)$$

where δ represents the dye effect for the gene under study. The notion of a balanced dye allocation is defined as follows.

Definition 2. *A dye allocation for a design with n slides and design matrix X is said to be balanced with respect to the contrasts $B\boldsymbol{\beta}$ provided that*

$$B(X^T X)^{-1} X^T \mathbf{1} = \mathbf{0}.$$

This definition is one of orthogonality and has two important properties. Firstly, if dye bias is present, the ordinary least squares estimate $\hat{\boldsymbol{\gamma}} = B(X^T X)^{-1} X^T \mathbf{M}$ will be unbiased provided that the dye allocation is balanced. Hence the design is robust in the sense that failure to include a term for dye bias will not lead to biased estimates for the contrasts of interest. Note however, fitting such a model will result in an inflated residual variance for genes where a dye bias is present.

Secondly, if the model including dye bias (3.15) is used, then

$$\text{var}(\hat{\boldsymbol{\gamma}}) = \sigma^2 B \left((X^T X)^{-1} + \frac{1}{n - \mathbf{1}^T X (X^T X)^{-1} X^T \mathbf{1}} (X^T X)^{-1} X^T \mathbf{1} \mathbf{1}^T X (X^T X)^{-1} \right) B^T. \quad (3.16)$$

If the dye allocation is balanced, this reduces to $\sigma^2 B(X^T X)^{-1} B^T$. Hence there is no loss of information associated with the estimation of a gene-specific dye bias however it results in a loss of a degree of freedom.

When the number of contrasts exceeds the number of parameters and B is a matrix of full column rank, the definition of balance reduces to the usual orthogonality condition, $X^T \mathbf{1} = \mathbf{0}$. Moreover, if the design consists only of even numbers of replicates of each configuration, a dye-swapped allocation can be used and will automatically satisfy this condition.

The approach to statistical design presented here is to adopt the linear model given in (2.4) and to seek designs that are Pareto optimal with respect to the contrasts of scientific interest. Following design selection, a suitable dye allocation is sought to minimise the effect of any gene-specific dye bias in model (3.15). An alternative approach is to assume that the dye bias is always present and that the optimal allocation occurs as part of the overall optimisation. This approach has been presented in Tsai et al. (2006) in the context of seeking A-optimal designs for a single factor experiment. In the present framework, the alternative approach can be implemented by seeking designs that are Pareto optimal with respect to the diagonal elements of (3.16). The two approaches will produce the same designs for a given experimental situation and set of contrasts of interest when a balanced dye allocation can be found.

If a completely balanced allocation does not exist for a given optimal design X then the two approaches may produce different designs. In particular, there may be designs X_* that are Pareto optimal with respect to the diagonal elements of (3.16) that may not be Pareto optimal with respect to the diagonal elements of (3.7). In such cases, X_* would be superior to X when the term for dye bias is included in the model but inferior if it is excluded. Based on experience, both approaches are typically expected to produce the same or very similar optimal designs.

If a completely balanced allocation is not possible, then a dye allocation can be sought to minimise the effects of dye bias with respect to the contrasts of greatest interest under the present approach as follows. Let X be a design matrix with given dye orientation and let

$$\mathbf{d} = B(X^T X)^{-1} X^T \mathbf{v} \quad (3.17)$$

where \mathbf{v} is the dye orientation vector with elements that have the value 1 if the dye orientation is the same as that given in X and -1 if the dye allocation is reversed. An orientation vector \mathbf{v} can then be sought either to optimise explicitly for the contrasts of greatest interest, or alternatively to optimise an overall measure of balance such as $D = \sum_i |d_i|$.

For an experiment with n slides available, there are potentially 2^{n-1} different dye allocations to be checked. Since dye-swapped replicates do not contribute to \mathbf{d} , a simple strategy is to enforce all possible dye-swaps for replicate slides and then minimise with respect to the remaining slides that cannot be dye-swapped.

To illustrate, consider the sphingosine kinase experiment with 15 slides and the design in Figure 3.4. For this design, each of the 6 horizontal configurations appear once and each of the 3 vertical configurations appear three times. Thus the problem is one of dye allocation for the 6 horizontal slides and the remaining 3 vertical slides following dye-swapping. Enumeration of all possible dye allocations for the remaining slides yields 8 allocations that achieve the minimal value $D = 1$. In each of those 8 allocations, it is found that the remaining vertical configurations must have the same dye orientation. The patterns of allocation for the horizontal configurations are given in Table 3.7.

Furthermore, from Equation (3.17), suppose $X^T \mathbf{v}$ is partitioned such that $X^T \mathbf{v} = X_H^T \mathbf{v}_H + X_V^T \mathbf{v}_V$ for leftover horizontal (H) and vertical (V) replicates respectively. For each of the 8 allocations, it is found that $X_H^T \mathbf{v}_H = 0$, indicating that the horizontals provide balance. On the other hand, the verticals introduce a minimal amount of gene-specific dye bias in this case. Note that dye bias would not be present in cases where an even number of replicates are allocated to the vertical replicates instead. More generally, then, for the sphingosine kinase experiment, no gene-specific dye bias is present for designs that satisfy both the conditions that

- the variance is constant within each of the subsets of biological interest, and

Table 3.7: Dye allocation for the remaining horizontals for the 2×3 sphingosine kinase experiment.

Configurations with the same sign have the same dye orientation and those with different signs have the opposite dye orientation.

Configuration	Allocation 1	Allocation 2
00 vs 0a	-1	-1
g0 vs ga	-1	1
00 vs 0b	1	1
g0 vs gb	1	-1
0a vs 0b	-1	-1
ga vs gb	-1	1

- vertical replicates are dye-swapped.

3.6 Concluding comments

The first step in a well-designed microarray experiment is to clearly state the scientific objectives of the study and to formulate specific questions to be answered. Following this, the objectives or questions should be put in order of priority. The next step is to formulate a conceptual probabilistic model and frame the questions in the context of this model. In general, the questions will be expressed as hypotheses about parameters and combinations of parameters. In this chapter, linear combinations of parameters, referred to as contrasts, have been considered. The design problem is to optimise the precision of estimators of contrasts subject to the available experimental resources.

The criterion of statistical efficiency based on Pareto optimality in Glonek and Solomon (2004), introduced for parameters only, has been extended in this chapter in order to optimise for contrasts

of scientific interest in the linear model. This extension is necessary because it is often the case that more contrasts are of interest than can be expressed as parameters of the linear model. For example, in the 2×3 sphingosine kinase experiment, it was illustrated that the contrasts for the interactions, which are typically of primary interest in the factorial microarray context, can not all be expressed as parameters of the linear model. In addition, the development that considers contrasts allows for the optimisation of effects regardless of whether or not they have been expressed as parameters in the parameterisation adopted. In practice, a parameterisation that is convenient can be adopted and Pareto optimality is applied based on the efficiency of the contrasts of interest.

In this chapter, the formation of subsets composed of contrasts of equal importance has been considered. A given subset may arise from effects that are of equal scientific interest or from quantities that may be interpreted equivalently. For example, in the sphingosine kinase experiment, the subsets consisted of effects for cell type, estrogen and interactions. It has been argued that effects within a given subset should ideally be estimated with equal precision. To implement this, it is appropriate to consider suitably constructed objective functions that are penalised to ensure equality. This reduces the problem of design selection to looking at the set consisting of only the most appropriate efficient designs. It has been found that this approach leads to relatively small sets of efficient designs to be considered.

The methodology for design selection developed here consists of identifying and prioritizing relevant contrasts, developing constraints and applying Pareto optimality. The issue of dye allocation is addressed following design selection.

To illustrate the methodology developed in this chapter, a number of factorial microarray experimental situations were considered. For those experiments, it was feasible to consider all possible designs in an exhaustive search for Pareto optimal designs. For larger experiments, it may not be feasible to carry out an exhaustive search for Pareto optimal designs. This presents the problem of

exploring the design space in an efficient way based on a guided search algorithm. The development of a metaheuristic method to generate a good approximation to the set of Pareto optimal designs when an enumerative approach is not feasible is the topic of the next chapter.

Chapter 4

Pareto simulated annealing for microarray experiments

4.1 Introduction

This chapter introduces Pareto simulated annealing, (PSA), a multiple objective metaheuristic method used to solve multiple objective combinatorial optimisation problems. In the context of microarray experiments, the aim of Pareto simulated annealing is to find a good approximation to the exact set of Pareto optimal designs in a relatively short time. This is relevant for microarray experiments for which it is infeasible to carry out an exhaustive search for the exact set of Pareto optimal designs.

The core Pareto simulated annealing algorithm, to be presented in the microarray context, is an adaptation of the general PSA concepts given in Czyzak and Jaszkievicz (1998) and Jaszkievicz (2001). Pareto simulated annealing is an extension of single objective simulated annealing. Thus the method of simulated annealing is presented in the microarray context prior to the presentation of Pareto simulated annealing. Following this, the core Pareto simulated annealing algorithm is

expanded on with the use of response surface methodology to develop an algorithm for finding optimal designs for large microarray experiments. The algorithms developed are illustrated in the context of a factorial microarray experimental situation.

4.2 Simulated annealing for microarray experiments

4.2.1 Introduction

Consider a microarray experiment with n slides and r configurations. A design, given by $\mathbf{d} = (d_1, d_2, \dots, d_r)$, consists of the number of slides allocated to each configuration such that

- $d_1, d_2, \dots, d_r \geq 0$,
- $d_1 + d_2 + \dots + d_r = n$.

For a design \mathbf{d} , the corresponding variance objective function is given by $f(\mathbf{d})$. The goal is to minimize $f(\mathbf{d})$ using simulated annealing.

Simulated annealing provides a search strategy to find a design that minimizes the value of the variance objective function. At the start, an initial design is chosen to be the current design. During the search, the current design is compared to a nearby, or neighbourhood, design. If the variance objective function is decreased, the new design is accepted to replace the current design. If the variance objective function increases, the new design may be accepted according to a specified probability function. At the beginning, almost all neighbourhood designs are accepted. This facilitates exploration of the design space. Gradually, the acceptance of neighbourhood designs becomes more selective in terms of a reduction in the probability that a generating design will move to a design that leads to a deterioration in the value of the variance objective function. Near the end of the search, only neighbourhood solutions that do not lead to a deterioration in the value of the variance objective function are accepted.

In the sections that follow, the general simulated annealing procedure, given in Pirlot (1996), is adapted to the microarray context.

4.2.2 Strategic concepts

Methods for randomly selecting starting designs

Denote a design by $\mathbf{d} = (d_1, d_2, \dots, d_r)$ where d_i is the number of slides allocated to configuration i . In order to randomly generate \mathbf{d} , the following can be carried out until all slides have been allocated to configurations:

- choose a configuration at random,
- choose a number between 1 and the number of slides that still need to be allocated and
- add that number of slides to the total for the randomly chosen configuration.

Let the above method be called *RD1*.

An alternative method, say *RD2*, to randomly generate a design would be to do the following until all slides have been allocated to configurations:

- choose a configuration at random and
- add one slide to the total for the randomly chosen configuration.

The methods given above are simple in terms of ease of implementation, with *RD1* expected to require fewer iterations to carry out. However it should be noted that each possible design will not have an equal probability of being selected, *RD2* generates designs according to the multinomial distribution.

Neighbourhood

Define the neighbourhood of \mathbf{d} to be any design obtained by removing a slide from one configuration and re-allocating it to a different configuration. Such a neighbourhood structure allows for the possibility of moving from a given design to any other design over time.

Acceptance probability

Define the acceptance probability to be the probability of moving from the current design \mathbf{d} to neighbourhood design \mathbf{y} whereby

$$P(\mathbf{d}, \mathbf{y}, T) = \min\{1, \exp((f(\mathbf{d}) - f(\mathbf{y}))/T)\} \quad (4.1)$$

where T is the temperature. This means that the current design will always move to a neighbourhood design that does not lead to a deterioration in the value of the variance objective function. Otherwise, the current design moves to the neighbourhood design with a probability less than one. The temperature parameter, T , affects the acceptance probability.

Cooling schedule

In order to achieve the simulated annealing strategy described in Section 4.2.1, the temperature parameter is progressively decreased according to a cooling schedule as follows.

- Select an initial temperature T_0 , number of steps at each temperature, l , and cooling rate κ (where $0 < \kappa < 1$)
- For each temperature, T_k , do the following
 - Maintain the temperature for l steps
 - After l steps, decrease the temperature. For example, using cooling rate κ ,

$$T_k = \kappa^k T_0 \quad (4.2)$$

- Stop when the final temperature is reached

Stopping rule

The stopping rule determines when the search ends. One example of a stopping rule is that the search ends when the minimum value achieved for the objective, given by $f(\mathbf{d}^*)$, does not improve by

at least ϵ_1 % after k_1 consecutive series of l steps. Another example is to stop when the proportion of accepted moves is less than ϵ_2 % for k_1 consecutive series of l steps. A third alternative is to stop after k_3 consecutive series of l steps, that is, after a certain number of designs, say n_D , have been visited.

4.2.3 Algorithm

The following algorithm for simulated annealing, based on the general algorithm given in Pirlot (1996), is adapted to the microarray context as follows.

During the algorithm, let the current design be \mathbf{d} and the corresponding value of the variance objective function be $f(\mathbf{d})$. Let the best design obtained be \mathbf{d}^* and the corresponding minimum value of the variance objective function be $f(\mathbf{d}^*)$.

Initial conditions

Various initial conditions are set as follows.

1. Select an initial design \mathbf{d} .
2. Set the initial best design and minimum variance objective function to be $\mathbf{d}^* = \mathbf{d}$ and $f(\mathbf{d}^*) = f(\mathbf{d})$.
3. Set the initial temperature such that $T = T_0$.

Iterative steps

Repeat the following until the stopping condition is fulfilled.

1. Randomly select a neighbourhood design \mathbf{y} and calculate its corresponding value for the variance objective function $f(\mathbf{y})$.

2. Always accept a neighbourhood design that does not lead to a deterioration in the value of the variance objective function. That is, if $f(\mathbf{y}) \leq f(\mathbf{d})$, accept \mathbf{y} to be the new current design \mathbf{d} .
3. If the neighbourhood design provides an improvement in the value of the variance objective function, update the best design and corresponding value of the variance objective function. That is, if $f(\mathbf{y}) < f(\mathbf{d}^*)$, then $\mathbf{d}^* = \mathbf{y}$ and $f(\mathbf{d}^*) = f(\mathbf{y})$.
4. A neighbourhood design that leads to a deterioration in the value of the variance objective function is accepted with probability $P(\mathbf{d}, \mathbf{y}, T)$ (acceptance probability) as follows.
 - Randomly select a number $0 \leq p \leq 1$.
 - If $p \leq P(\mathbf{d}, \mathbf{y}, T)$, then accept the neighbourhood design \mathbf{y} to be the new current design \mathbf{d} .
5. If the condition for changing the temperature is fulfilled, decrease the temperature T such that it becomes κT , $0 < \kappa < 1$.

4.2.4 Tuning the parameters

The values of the parameters must be selected in order to carry out the simulated annealing algorithm. Although the choice of parameter values is arbitrary, some general advice for setting them is provided in what follows, based on Pirlot (1996).

Initial temperature

Setting the initial temperature affects the acceptance probability. The initial temperature could be set to be relatively high so that it corresponds to a relatively high acceptance probability such as 0.9. This could facilitate the exploration of the design space. However, the effectiveness of this

depends on the application and it could be the case that lower initial temperatures may provide better results.

Cooling rate and number of steps

The cooling rate and number of steps are correlated in terms of the speed at which the cooling schedule is carried out. The larger the value for the cooling rate κ and the number of steps at each temperature l , the more slowly the cooling schedule is carried out and vice versa. The choice of values depends on what is feasible computationally and the application so the best advice is to experiment.

Stopping criterion

The parameters for the % improvement in the best value obtained for the variance objective function and the proportion of accepted moves stopping rules described in Section 4.2.2, ϵ_1 and ϵ_2 , are typically low such as 5 %. The choice of the number of designs visited depends on what is feasible for a given application.

4.3 Pareto simulated annealing

4.3.1 Introduction

The general Pareto Simulated annealing concepts given in Czyzak and Jaszkievicz (1998) and Jaszkievicz (2001) are presented here, in the microarray context, by expanding on concepts introduced for single objective simulated annealing presented earlier in this chapter. In particular, for Pareto simulated annealing, the concept of the neighbourhood and cooling schedule from simulated annealing are preserved and the acceptance probability is modified to take multiple objectives into account. An important feature of Pareto simulated annealing is that the search explores the design space using a set of designs, called *generating designs*, rather than a single design as is the case for simulated annealing.

The aim of Pareto simulated annealing is to find a good approximation to the exact set of Pareto optimal designs in a relatively short time. In order to do so, a guided search is carried out in the design space using a sample of generating designs. At each iteration of the search, a given generating design is compared to a nearby, or neighbourhood, design. At the beginning of the search, the generating designs move most freely around the design space. The search becomes gradually more selective over time in terms of a reduction in the probability that a generating design will move to a design that provides no improvement in any of the variance objective functions. In addition to the use of generating designs, a set of designs, called the *potentially Pareto optimal set*, is maintained during the search. Another aspect of the guided search is the information exchange that takes place among generating designs. During the search, generating designs are compared to each other and are influenced to repel each other to be dispersed throughout the variance objective function space.

4.3.2 Strategic concepts

Recall that simulated annealing was introduced in Section 4.2. For Pareto simulated annealing, the same notation is adopted and the structure of the neighbourhood is preserved.

In what follows, concepts for Pareto simulated annealing that are additional to those for the simulated annealing framework are introduced.

Generating designs

The generating designs are a set of designs used to explore the design space. During the search, individual generating designs may be replaced by neighbouring designs.

Multiple objective function

Recall that a single objective is optimised in the case of simulated annealing. Pareto simulated annealing is concerned with optimising for multiple objectives since the variance objective function is multi-dimensional. For example, for a generating design \mathbf{d} , the corresponding variance objective function, $f(\mathbf{d})$, is p -dimensional when optimising for p objectives.

Potentially Pareto optimal set

The potentially Pareto optimal set is the set of designs obtained by carrying out Pareto simulated annealing. It is maintained during the search so that it is updated at each iteration, that is, each time a new design is visited.

Acceptance probability

The set of generating designs consists of one or more individual designs. Moves are made by individual generating designs one at a time during the search. The acceptance probability is a measure

of the probability of moving from the current generating design \mathbf{d} to a randomly selected neighbourhood design \mathbf{y} . Taking into account multiple objectives, given by variance objective functions, the following two acceptance probability functions are given based on Czyzak and Jaszekiewicz (1998).

$$P(\mathbf{d}, \mathbf{y}, T, \Lambda) = \min\{1, \exp(\max_i \{\lambda_i(f_i(\mathbf{d}) - f_i(\mathbf{y}))/T\})\} \quad (4.3)$$

and

$$P(\mathbf{d}, \mathbf{y}, T, \Lambda) = \min\{1, \exp(\sum_{i=1}^p \lambda_i(f_i(\mathbf{d}) - f_i(\mathbf{y}))/T)\} \quad (4.4)$$

where T is the temperature and $\Lambda = \{\lambda_i\}$ is the set of weights used to influence the dispersion of generating designs.

For each of the acceptance probability functions, given above, the following holds. If a neighbourhood design does at least as well as the generating design for all variance objective functions, then the generating design will be replaced by the neighbourhood design. If the neighbourhood design is less efficient for all objectives, then the probability of moving to the neighbourhood design will always be less than one. In all other cases, the probability of moving to a neighbourhood design:

- is 1 when using the acceptance probability function given by Equation (4.3) since there is at least one objective that the neighbourhood design is doing better on than the generating design, and
- depends on the weighted sum of differences for each variance objective function when using the acceptance probability function given by Equation (4.4).

4.3.3 Role of weights

The role of the weights, Λ , is to disperse the generating designs throughout the variance objective space. This is implemented by adjusting the weights used in the acceptance probability function

based on information exchange between generating designs so that they act to repel each other in the variance objective space. Firstly, for each generating design, the Euclidean distance is used to determine which of the other generating designs is closest to it as follows. Consider a given generating design \mathbf{d}^i with corresponding variance objective function $f(\mathbf{d}^i)$. Consider the j -th generating design \mathbf{d}^j with corresponding variance objective function $f(\mathbf{d}^j)$ in the set of generating designs such that $i \neq j$. The distance between \mathbf{d}^i and \mathbf{d}^j is taken to be $S = \sum_{k=1}^p (f(\mathbf{d}^i)_k - f(\mathbf{d}^j)_k)^2$. The closest generating design to \mathbf{d}^i is taken to be the generating design \mathbf{d}^j for which S is minimal and which is not dominated by \mathbf{d}^i . Then the weights for each objective, to be applied to the acceptance probability function, are adjusted such that

- for the objectives that the given generating design is doing well on compared to the closest generating design, the weights are increased so that the given generating design is more likely to make a move that continues to improve on those objectives, and
- for the objectives that the given generating design is not doing well on compared to the closest generating design, the weights are decreased so that the given generating design is less likely to be geared towards making a move that improves on those objectives.

If the given generating design \mathbf{d}^i dominates all other designs in the set of generating designs, a closest generating design does not exist and the weights of the given generating design are randomly adjusted. That is, for each of the objectives, each of the associated weights are adjusted separately to increase or decrease with equal probability.

4.4 Core Pareto simulated annealing algorithm

Based on the procedure for Pareto simulated annealing presented in Czyzak and Jaszkievicz (1998) and Jaszkievicz (2001), the general algorithm has been adapted to the microarray context. Thus

the following core Pareto simulated annealing algorithm is proposed.

Initialisation

1. Select an initial set of generating designs $\{\mathbf{d}\}$.
2. For each generating design \mathbf{d} , generate an initial weight vector such that $\lambda_i^{\mathbf{d}} = 1/p$ where p is the number of variance objective functions.
3. Initialise the set of potentially Pareto optimal designs, PP , to be those designs that are Pareto optimal among the initial set of generating designs.
4. Set the initial temperature such that $T = T_0$.

Iterative Steps

For each generating design, \mathbf{d} , do the following until the stopping condition is fulfilled.

1. Construct a neighbourhood design \mathbf{y} as follows. Randomly select a configuration in $\{\mathbf{d}\}$ that has at least one slide allocated to it. Remove one slide from that configuration and re-allocate it to a different configuration that is selected at random.
2. If \mathbf{y} is not dominated by \mathbf{d} , update the set PP with \mathbf{y} .
3. Select the closest generating design, in the objective space, \mathbf{d}^o that is Pareto optimal with respect to \mathbf{d} . If there is no such generating design, adjust the weights such that $\lambda_i^{\mathbf{d}} = \lambda_i^{\mathbf{d}}\alpha$ or $\lambda_i^{\mathbf{d}} = \lambda_i^{\mathbf{d}}/\alpha$, each with probability equal to 0.5. Otherwise adjust the weights such that

$$\lambda_i^{\mathbf{d}} = \begin{cases} \lambda_i^{\mathbf{d}}\alpha, & \text{if } f_i(\mathbf{d}) < f_i(\mathbf{d}^o) \\ \lambda_i^{\mathbf{d}}/\alpha, & \text{if } f_i(\mathbf{d}) \geq f_i(\mathbf{d}^o) \end{cases}$$

(α is greater than, but close to, 1.)

4. Normalize the weights such that $\sum_i \lambda_i^{\mathbf{d}} = 1$.
5. Accept the neighbourhood design, to replace the generating design, with probability $P(\mathbf{d}, \mathbf{y}, T, \Lambda^{\mathbf{d}})$.
6. If the condition for changing the temperature is fulfilled, decrease the temperature T such that it becomes κT , $0 < \kappa < 1$.

4.5 Quality measures

To evaluate of the performance of Pareto simulated annealing, the following quality measures are proposed.

4.5.1 Comparison with exact set of Pareto optimal designs

The quality measures presented in this section arise from considering Czyzak and Jaszkievicz (1998). In all cases, it is assumed that the exact set of Pareto optimal designs is given.

Number of designs missed

Quality measure Q_m is based on the number of designs that appear in the exact set R of Pareto optimal designs but not in the potentially Pareto optimal set PP obtained by applying Pareto simulated annealing. This constitutes the number of designs missed given by

$$Q_m = \text{card}\{R\} - (\text{card}\{PP \cap R\}). \quad (4.5)$$

A similar measure is the proportion of designs missed whereby

$$Q_p = \frac{\text{card}\{R\} - (\text{card}\{M \cap R\})}{\text{card}\{R\}}. \quad (4.6)$$

In practice, it is convenient to use the empirical logit whereby

$$Q_l = \log(Q_m + 0.5) / (\text{card}\{R\} + 0.5 - Q_m) \quad (4.7)$$

Average distance

Quality measure Q_a is based on the average distance of designs in the exact set to the closest design in the set of potentially Pareto optimal designs as follows. For a given design, say \mathbf{v} , in the exact set, the closeness of a design \mathbf{u} in the set of potentially Pareto optimal designs is given by

$$c(\mathbf{u}, \mathbf{v}) = \max_{i=1,2,\dots,p} \{0, w_i(f_i(\mathbf{u}) - f_i(\mathbf{v}))\} \quad (4.8)$$

The weight $w_i = 1/\Delta_i$ where Δ_i is the range for the i -th objective in the exact set. The closest design is that \mathbf{u} for which $c(\mathbf{u}, \mathbf{v})$ is minimized. Following such measurements for all designs in the exact set, Q_a is the average of the distances such that

$$Q_a = \frac{1}{\text{card}\{R\}} \sum_{\mathbf{v} \in R} \{ \min_{\mathbf{u} \in PP} \{c(\mathbf{u}, \mathbf{v})\} \}. \quad (4.9)$$

Worst case

Quality measure Q_w presents the worst case scenario as follows. For each design in the exact set, the closest design in the set of potentially Pareto optimal designs is found using Equation (4.8). After all designs are considered, the worst case is returned such that

$$Q_w = \max_{\mathbf{v} \in R} \{ \min_{\mathbf{u} \in PP} \{c(\mathbf{u}, \mathbf{v})\} \}. \quad (4.10)$$

4.5.2 Random search

The effectiveness of Pareto simulated annealing, in an absolute sense, can also be investigated by its comparison to a random search. In this section, methods to carry out a random search are developed and their comparison to Pareto simulated annealing is introduced.

Methods for carrying out a random search

To carry out a random search for Pareto optimal designs, a method for selecting starting designs is chosen. In addition, subsequent designs must be generated at each iteration.

Suppose *RD1* or *RD2*, introduced in Section 4.2.2, is used to generate a given number of starting designs. It is possible to proceed with a random search by using the same method chosen for the selection of starting designs to construct designs at each iteration of the search. Alternatively, however, following a method chosen for selecting starting designs, neighbourhood designs could be constructed at each iteration by removing a slide from a configuration chosen at random and re-allocating it to another randomly chosen configuration.

The methods, described above, are given by

- *M1*: use *RD1* to select starting designs and subsequent designs at each iteration;
- *M2*: use *RD2* to select starting designs and subsequent designs at each iteration;
- *M3*: use *RD1* to select starting designs and subsequently select neighbourhood designs at each iteration; and
- *M4*: use *RD2* to select starting designs and subsequently select neighbourhood designs at each iteration; the motivation is that, with one starting design, it can be thought of as defining a Markov chain on the space of all designs whose stationary distribution is uniform over all designs.

Regardless of the method chosen, during the random search, the set of potentially Pareto optimal designs is updated at each iteration. This continues until the stopping condition is fulfilled.

Comparison of Pareto simulated annealing to a random search

Comparison of Pareto simulated annealing to random search methods can be carried out as follows. Suppose Pareto simulated annealing is carried out for a given number of visits. Then a random search can be carried out whereby the same number of designs are to be visited as in the case of Pareto simulated annealing. Suppose the exact set of Pareto optimal designs is available. Then the quality measures introduced in Section 4.5.1 can be calculated for the potentially Pareto optimal sets obtained from Pareto simulated annealing and random searches, both using the exact set as the reference set. The resulting measures can be compared to each other. If the quality measure is taken to be the number of Pareto optimal designs missed, then the performance of the Pareto simulated annealing algorithm can simply be assessed by comparison to the hypergeometric distribution from a set of randomly chosen designs. An illustration will be given later in this chapter.

4.6 Tuning parameters for Pareto simulated annealing

4.6.1 Introduction

The core Pareto simulated annealing algorithm, presented in Section 4.4, involves a number of tuning parameters that affect its performance. In this section, algorithms to search systematically for the optimal values of the tuning parameters are developed. To achieve this, the performance of the core Pareto simulated annealing algorithm is systematically evaluated with the use of response surface methodology (see, for example, Box et al. (1978)). This is feasible in simple examples for which the exact Pareto optimal set R is known and can be used to inform the choice of parameters in larger problems. Following this, an adaptive algorithm for the selection of tuning parameters is developed in the practical case when the exact set is not known.

The tuning parameters of the core algorithm to be considered are summarised as follows.

Number of generating designs (NG): this is the number of designs used to explore the design space such that each visits successive designs in its neighbourhood.

Initial temperature (IT): this is the temperature that is set at the beginning of the algorithm.

Cooling rate (CR): this is the rate, $\kappa \in (0, 1]$, at which the temperature T is decreased such that it becomes κT following the completion of a temperature level.

Repulsion coefficient (RC): this is the multiplicative factor, $\alpha \geq 1$, used to adjust weights to allow for the dispersion of generating designs.

Acceptance rule (AR): this is the rule used to calculate the probability of moving from the current generating design to a randomly selected neighbourhood design. Rule 0 and rule 1, given by equations (4.3) and (4.4) respectively, are considered.

The number of designs, initial temperature and cooling rate are quantitative parameters. The consideration of the repulsion coefficient is to assign two values, one to indicate no repulsion ($\alpha = 1$) and the other to indicate repulsion (choose $\alpha > 1$). The acceptance rule has two values corresponding to the two options proposed for the acceptance probability function.

4.6.2 Central composite experimental plan

In order to find the optimal values for the tuning parameters, a sequence of experiments is conducted. Each such experiment is defined by a central composite design. The quantitative parameters, number of generating designs, initial temperature and cooling rate, are each assigned 5 values, corresponding to very low, low, medium, high and very high. The values for each quantitative parameter are typically set so that the interval between the low and medium levels is equal to that for medium and high. Furthermore, these intervals are typically twice the size of that for the interval between very low and low and that for high and very high.

The central composite design is constructed as follows. Firstly, for the quantitative parameters, form the setting combinations consisting of:

- all combinations of low and high levels for the quantitative parameters,
- the combination corresponding to each quantitative parameter set to the medium level,
- the combinations arising from setting each quantitative parameter in turn to very low while the others are set to medium and
- the combinations arising from setting each quantitative parameter in turn to very high while the others are set to medium.

To consider the repulsion coefficient and acceptance rule also, each of the setting combinations formed for the quantitative parameters is carried out in the absence and presence of repulsion and

for both acceptance rules. All setting combinations constitute the central composite design.

A single experiment then consists of a specified number of replicates of the central composite design.

4.6.3 Analysis

The analysis of a single experiment is performed using multiple linear regression. Each tuning parameter is represented by a variable x . For the quantitative variables, number of generating designs, initial temperature and cooling rate, the values of x are -1.5 , -1 , 0 , 1 and 1.5 . For the repulsion coefficient and acceptance rule, the values of x are -1 and 1 .

The quality measure, Q , forms the response variable. Examples of such measures were presented in Section 4.5.1 for the case where the exact set is known. Later in this section, practical cases where the exact set is not known are catered for.

The linear model is defined to be

$$ML : E(Q) = \beta_0 + \sum_{i=1}^k \beta_i x_i$$

and the quadratic model is defined to be

$$MQ : E(Q) = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \gamma_i x_i^2 + \sum_{i < j} \gamma_{ij} x_i x_j.$$

In what follows, algorithms for finding optimal values for the tuning parameters are proposed based on the core algorithm and use of response surface methodology.

4.6.4 Parameter selection algorithm

For cases where the exact set R is known, the following algorithm for the selection of suitable value for the Pareto simulated annealing tuning parameters is proposed.

Initialisation

1. Set the parameter values for the first experiment.
2. Using the appropriate combinations of parameter values, form the first composite design.
3. Perform a number of replicate runs of the core Pareto simulated annealing algorithm for each combination of parameters in the first central composite design.
4. Calculate the quality measure Q for each run.
5. Fit the linear and quadratic models given by ML and MQ respectively.
6. Determine whether the quadratic model provides a substantial improvement over the linear model. A formal test to do so is given as follows. Calculate the F-statistic $\frac{(R_{MQ}^2 - R_{ML}^2)/(\nu_{ML} - \nu_{MQ})}{(1 - R_{MQ}^2)/\nu_{MQ}}$. where ν_{ML} and ν_{MQ} are the degrees of freedom for ML and MQ respectively. Compare the calculated value with $F_{\nu_{MQ} - \nu_{ML}, \nu_{MQ}}$. Based on the critical point of F at the 10% level of significance, accept the quadratic model if the calculated value of F exceeds the critical point.
7. If the quadratic model provides no substantial improvement over the linear model, apply the method of steepest descent. A formal specification is to adjust the tuning parameter values in proportion to their estimated coefficients. Thus, if the coefficients of the k PSA tuning parameters, x_i , are β_i and x_1 is changed by Δ then set

$$x_j = \frac{\beta_j}{\beta_1} \Delta \quad (4.11)$$

for $j = 2, \dots, k$. The value of Δ is typically taken so that the move is 1 in terms of the coded units. That is:

$$\Delta = (1 + \beta_1^2 + \dots + \beta_k^2)^{-0.5}. \quad (4.12)$$

In practice, there may be restrictions on the feasible range of values for the tuning parameters and an approximation to the estimated direction of steepest descent is used. If, however, the quadratic model is accepted, proceed to the optimisation step.

Iterative Steps

The following steps are repeated until it is determined that the quadratic model is appropriate.

1. Using the method of steepest descent, set the parameter values for the current experiment.
2. Using the parameter values, form the current central composite design.
3. Perform a number of replicate runs of the core Pareto simulated annealing algorithm for each combination of parameters in the current central composite design.
4. Calculate the quality measure Q for each run.
5. Fit the linear and quadratic models given by ML and MQ respectively.
6. Determine whether the quadratic model provides a substantial improvement over the linear model.
7. If the quadratic model provides no substantial improvement over the linear model, apply the method of steepest descent. If, however, the quadratic model is accepted, proceed to the optimisation step.

Optimisation step

Following the determination of an appropriate quadratic model, the optimal values for the parameters can be estimated as follows.

1. Choose x_1, x_2, \dots, x_k to find the least value of the quadratic function

$$\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i + \sum_{i=1}^k \hat{\gamma}_i x_i^2 + \sum_{i < j} \hat{\gamma}_{ij} x_i x_j,$$

within the specified domain, where $\{\hat{\beta}_0, \hat{\beta}_i, \hat{\gamma}_i, \hat{\gamma}_{ij}\}$ are the least squares estimates obtained from the final experiment. If this least value lies on the boundary, it is not a minimum (unless a minimum lies precisely on the boundary). If it is not a minimum, use it to define the direction of steepest descent and return to Iterative Step 7.

2. Calculate the corresponding values for the tuning parameters on the original scale.

For cases where the exact set is known, this algorithm can be applied to obtain insight into the importance and effect of each of the tuning parameters. Moreover, it provides the basis for the development and testing of the adaptive algorithm proposed in the next section. Applications are presented later in this chapter.

4.6.5 Adaptive Pareto simulated annealing algorithm

In practical cases, it is not feasible to determine the exact set of Pareto optimal designs. The parameter selection algorithm can be modified to form an adaptive algorithm that caters for such cases. In particular, the exact set R is replaced by a suitably constructed reference set U that is defined by cumulatively combining all of the potentially Pareto optimal designs from all experiments as the iterations proceed. At each iteration, the reference set U is used in place of R for the calculation of quality measures as described in Section 4.5.1.

The adaptive Pareto simulated annealing algorithm, in which the tuning parameters are adaptively modified, is defined as follows.

Initialisation

1. Set the parameter values for the first experiment.
2. Using the appropriate combinations of parameter values, form the first composite design.
3. Perform a number of replicate runs of the core Pareto simulated annealing algorithm for each combination of parameters in the first central composite design.
4. Construct U to consist of the designs that are Pareto optimal among all of the potentially Pareto optimal (PP) sets obtained from all of the runs performed. This constitutes the reference set.
5. Calculate the quality measure Q for each run.
6. Fit the linear and quadratic models given by ML and MQ respectively. Calculate the F-statistic $\frac{(R^2_{MQ} - R^2_{ML})/(\nu_{ML} - \nu_{MQ})}{(1 - R^2_{MQ})/\nu_{MQ}}$, where ν_{ML} and ν_{MQ} are the degrees of freedom for ML and MQ respectively. Compare the calculated value with $F_{\nu_{MQ} - \nu_{ML}, \nu_{MQ}}$. Based on the critical point of F at the 10% level of significance, accept the quadratic model if the calculated value of F exceeds the critical point.
7. If the quadratic model provides no substantial improvement over the linear model, apply the method of steepest descent. A formal specification is to adjust the tuning parameter values in proportion to their estimated coefficients. Thus, if the coefficients of the k PSA tuning parameters, x_i , are β_i and x_1 is changed by Δ then set

$$x_j = \frac{\beta_j}{\beta_1} \Delta \quad (4.13)$$

for $j = 2, \dots, k$. The value of Δ is typically taken so that the move is 1 in terms of the coded units. That is:

$$\Delta = (1 + \beta_1^2 + \dots + \beta_k^2)^{-0.5}. \quad (4.14)$$

In practice, there may be restrictions on the feasible range of values for the tuning parameters and an approximation to the estimated direction of steepest ascent is used. If, however, the quadratic model is accepted, proceed to the optimisation step.

Iterative Steps

The following steps are repeated until it is determined that the quadratic model is appropriate.

1. Using the method of steepest descent, set the parameter values for the current experiment.
2. Using the parameter values, form the current central composite design.
3. Perform a number of replicate runs of the core Pareto simulated annealing algorithm for each combination of parameters in the current central composite design.
4. Update U to consist of only those designs that are Pareto optimal among all potentially Pareto optimal (PP) sets obtained from the runs just performed and U from the previous step.
5. Calculate the quality measure Q , with respect to the current reference set U , for each run.
6. Fit the linear and quadratic models given by ML and MQ respectively.
7. Determine whether the quadratic model provides a substantial improvement over the linear model.
8. If the quadratic model provides no substantial improvement over the linear model, apply the method of steepest descent. If, however, the quadratic model is accepted, proceed to the optimisation step.

Optimisation step

Following the determination of an appropriate quadratic model, the optimal values for the parameters can be estimated as follows.

1. Choose x_1, x_2, \dots, x_k to minimize the quadratic function

$$\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i + \sum_{i=1}^k \hat{\gamma}_i x_i^2 + \sum_{i < j} \hat{\gamma}_{ij} x_i x_j$$

where $\{\hat{\beta}_0, \hat{\beta}_i, \hat{\gamma}_i, \hat{\gamma}_{ij}\}$ are the least squares estimates obtained from the final experiment. If this least value lies on the boundary, it is not a minimum (unless a minimum lies precisely on the boundary). If it is not a minimum, use it to define the direction of steepest descent and return to Iterative Step 8.

2. Calculate the corresponding values for the tuning parameters on the original scale.
3. Perform a number of replicate runs at the optimal values of the tuning parameters.
4. Take the final estimate, U , for the Pareto optimal set to be those designs that are Pareto optimal among all potentially Pareto optimal (*PP*) sets obtained from the runs at the optimum and U from the final iterative step.

In the next section, the algorithms developed above are applied.

4.7 Application to the leukaemogenesis experiment

In this section, the algorithms, developed earlier in this chapter, are applied to the leukaemogenesis experiment.

4.7.1 Background information

Recall, from Section 3.3.4, that the objectives to which Pareto optimality was applied for the exact method was based on the penalised objectives for each subset of interest, given by t_A , t_B and t_{AB} . It follows that it is appropriate to use those penalised objectives in the applications of Pareto simulated annealing presented here for this experimental situation. For all of the cases considered, the objectives are penalised with penalty weight 0.9999.

For the implementation of Pareto simulated annealing, it is necessary to select starting designs. Recall that methods for randomly selecting starting designs were given in Section 4.2.2. For convenience, the method *RD1* is used to randomly select starting designs in the illustrations of Pareto simulated annealing presented here.

4.7.2 Parameter selection algorithm: 36 slides

As an illustration, consider the case where 36 slides are available for the leukaemogenesis experiment. In this situation, there are 749,398 possible designs. It is feasible to carry out the exact method according to the penalty approach and the Pareto optimal set obtained consists of 63 designs. This constitutes the reference set.

Now suppose that the parameter selection algorithm is applied to the experiment. A central composite experimental plan for the PSA parameters is adopted as follows. The number of generating designs, initial temperature and cooling rate consist of 5 levels each, corresponding to very low, low, medium, high and very high. The values for those parameters are given in Table 4.1.

For a given acceptance rule, the method of setting the initial temperature in Table 4.1 is carried out as follows. Let the proportion of moves be the number of times a generating design is replaced by a neighbourhood design divided by the total number of visits at a given temperature. The initial temperatures were chosen to correspond to approximate proportions of moves: 30%, 40%, 60%, 80% and 90%. These proportions depend on acceptance probabilities and vary with generating designs. The correspondences between initial temperature and proportion of moves were estimated from pilot runs. Note that this was the method used for setting initial temperatures in all applications presented in this chapter.

The number of steps is fixed to be 600. The values for the repulsion coefficient and acceptance rule used each consist of 2 levels; (1, 1.05) and (rule 0, rule 1). Note that rule 0 and rule 1 are given by Equations (4.3) and (4.4) respectively. Thus the experimental plan consists of 15 setting combinations to vary the number of generating designs, initial temperature and cooling rate for a given level of the repulsion coefficient and acceptance rule.

The 15 combinations are carried out for each combination of the repulsion coefficient and acceptance rule, with a total of 60 settings. The number of replicates for each setting is 4 thus there are 240 runs in total. The stopping criterion is the completion of the temperature level that results in having visited at least 40,000 designs, which is 5.333% of the design space.

Table 4.1: Values for parameters that have 5 levels for the first experimental plan for the leukemogenesis experiment with 36 slides

NG	1	4	10	16	19
Initial T rule 0	0.00002	0.00003	0.000085	0.0004	0.002
Initial T rule 1	0.00005	0.00009	0.00035	0.0013	0.007
CR	0.4	0.5	0.7	0.9	1

The quality measure used is based on the number of Pareto optimal designs, M , in the exact set missed by the application of the particular PSA setting.

In particular, the empirical logit, introduced in Section 4.5.1, is used such that

$$Y = \log(M + 0.5)/(63.5 - M) \quad (4.15)$$

given that there are 63 designs in the exact set. The results from the first PSA experimental plan were analysed using linear models in R . A transcript of the R analysis is given in Appendix B.1. The adjusted R-squared for the linear model is 34% compared to 60% for the quadratic model. The linear model is statistically significant, inasmuch as there is overwhelming evidence against a null hypothesis that all the coefficients except the intercept are 0, but the quadratic model is a statistically significant improvement. The F-ratio for testing the null hypothesis that the coefficients of all the quadratic and cross-product terms are 0 is 12.67. When compared with the quantiles of an F-distribution with 13 and 221 degrees of freedom, for the numerator and denominator respectively, there is evidence to reject this null hypothesis at the 0.6×10^{-5} level. However, the least value of the fitted quadratic model, subject to the constraints, occurs on a boundary. In the coded units, the least value of the empirical logit is -1.941 when the number of generating designs, initial temperature, cooling rate, repulsion coefficient and acceptance rule are set at 0.13, -1.5 , 1.5, -1 and 1 respectively. The linear model suggests that an increase in the number of generating designs and the cooling rate and a reduction in the initial temperature provide improvements.

Taking the estimates and feasible values for the coefficients into account, the parameter settings from the first experimental plan are modified to the values given in Table 4.2. This forms the second central composite experimental plan. Note that the initial temperatures were chosen to correspond to approximate proportions of moves: 10%, 20%, 40%, 60% and 70%.

In addition, it is found that the model from the first experimental plan suggests that the low level

Table 4.2: Values for parameters that have 5 levels after applying steepest descent for the leukemogenesis experiment with 36 slides

NG	10	13	19	25	28
Initial T rule 1	0.00001	0.00003	0.00009	0.00035	0.0007
CR	0.7	0.75	0.85	0.95	1

of repulsion, which corresponds to no repulsion, is better and that acceptance rule 1 is better. Thus the second experimental plan consists of the 15 setting combinations from the modified parameter settings for the number of generating designs, initial temperature and cooling rate and applies no repulsion and uses acceptance rule 1. Furthermore, to maintain a total of 240 runs as carried out in the first plan, each setting in the second plan has 16 replicates. The stopping rule and number of steps from the first plan are preserved.

The results from the second plan were analysed in R and a transcript of the R analysis is given in Appendix B.2. The linear model is only just statistically significant and has an adjusted R-squared of 3%. Moving to the quadratic model, the adjusted R-squared increases to 63% and it is statistically significant. The quadratic model implies being near the minimum of a quadratic surface. Thus no further iterations are required and the quadratic model is used to find the co-ordinates of the minimum.

Using Solver in Excel, the co-ordinates of the minimum of the quadratic surface were found and the values for each of the parameters are given in Table 4.3.

The logit response and associated analysis presented is appropriate for the current application. In addition, recall that the average distance and worst case quality measures were introduced in Section 4.5.1 as alternatives. Using the second experimental plan for the PSA settings, the

Table 4.3: Estimated optimal parameter values for the empirical logit quality measure for the leukaemogenesis experiment with 36 slides

Parameter	Estimated value
NG	28
Initial T	0.00005
CR	1

current application, that is, the second experimental plan, is analysed with those alternative quality measures. The associated analysis in R is given in Appendix B.3. For the average distance quality measure, the optimal values for the tuning parameters were found to be $NG = 28$, $\text{Initial T} = 0.00015$ and $CR = 1$. For the worst case distance quality measure, the optimal values for the tuning parameters were found to be $NG = 10$, $\text{Initial T} = 0.00018$ and $CR = 1$.

4.7.3 Comparison of Pareto simulated annealing to random search methods

Recall that methods for carrying out a random search, namely $M1$, $M2$, $M3$ and $M4$, were introduced in Section 4.5.2. The effectiveness of Pareto simulated annealing will be investigated by comparing its performance to that of the random search methods.

As an illustration, the second central composite design used to apply Pareto simulated annealing to the leukaemogenesis experiment with 36 slides, from Section 4.7.2 with parameter values given in Table 4.2, is compared to each of the random search methods. As was the case for the application of Pareto simulated annealing, the number of replicates for each of the random search methods is 240 and the stopping condition is based on visiting 40,000 designs. For methods $M1$ and $M2$, the stopping condition is to stop when 40,000 designs have been visited. For $M3$ and $M4$, let there be 19 starting designs and a convenient stopping condition is to stop when 45,866 designs have

been visited. As was the case for the application of Pareto simulated annealing, the response is the empirical logit of the number of Pareto optimal designs in the exact set that are missed by the method, given by Equation (4.15).

The results for Pareto simulated annealing and the random search methods were analysed using a one way analysis of variance and the associated R output is given in Appendix B.4. In summary, $M1$, $M2$, $M3$ and $M4$ missed an average of 58.4, 54.2, 59.6 and 59.1 designs out of the 63 in the exact set. The associated standard deviations were 1.5, 1.7, 2.1 and 2.2. For a sample of 40,000 designs chosen at random without replacement, it follows from the hypergeometric distribution that the expected number of missed designs is 59.64 and the standard deviation is 1.78. In contrast, PSA missed an average of 21 designs with a standard deviation of 10.8. Based on the analysis, it is found that Pareto simulated annealing substantially outperforms each of the random search methods. This vindicates the use of Pareto simulated annealing.

4.7.4 Adaptive Pareto simulated annealing algorithm

In this section, the adaptive Pareto simulated annealing algorithm is applied.

Adaptive algorithm: 36 slides

For the leukaemogenesis experiment with 36 slides, given in Section 4.7.2, the first composite experimental plan for the parameter selection algorithm was given in Table 4.1. Suppose the reference set U is constructed to consist of those designs that are Pareto optimal among the 240 PP sets obtained from the first plan. It is found that the reference set U constructed in this way is identical to the set of all 63 designs in R obtained from the exact method. Therefore the adaptive algorithm, that uses U , would lead to the same results and analysis as the application of the parameter selection algorithm in this case.

Adaptive algorithm: 100 slides

Now suppose 100 slides are available for the experiment instead of 36. In doing so, the aim is to choose an experiment large enough to illustrate the use of the adaptive algorithm, wherein U is constructed and updated, but small enough to allow for its comparison to exact results. When 100 slides are available, there are 96,560,646 possible designs and the exact set consists of 389 Pareto optimal designs, obtained by enumeration. To illustrate the exact set graphically, the efficiency was calculated for each component of the variance objective function, relative to the overall minimum for that component. Figure 4.1 provides a 3-dimensional view of the exact set. The horizontal axis is the relative efficiency of the subset of effects for cell line, the “into the paper” axis is the relative efficiency of the subset of effects for time and the vertical axis is the relative efficiency of the interaction. Each point represents the three efficiencies for a design in the exact Pareto optimal set.

The most striking feature is that if the efficiency of interaction is reduced, the increase in efficiencies of cell line and time is only slight. In addition to interaction being of primary scientific interest, this further supports the choice of a design that is efficient for the interaction. Figures 4.2 and 4.3 are respectively the top and front views of Figure 4.1. The top view emphasises the fact that only very small improvements in efficiency for the cell line and time effects can be obtained by trading off efficiency for the interaction term. The front view provides a convenient graphical representation that will be used to display the results of the PSA algorithm in the remainder of this chapter.

To investigate its behaviour, the adaptive Pareto simulated annealing algorithm is now applied without reference to the exact set. As this illustration is for testing purposes, the search begins with settings including the number of generating designs, initial temperature and cooling rate, that are not necessarily expected to perform well. The parameter values for the first experimental plan

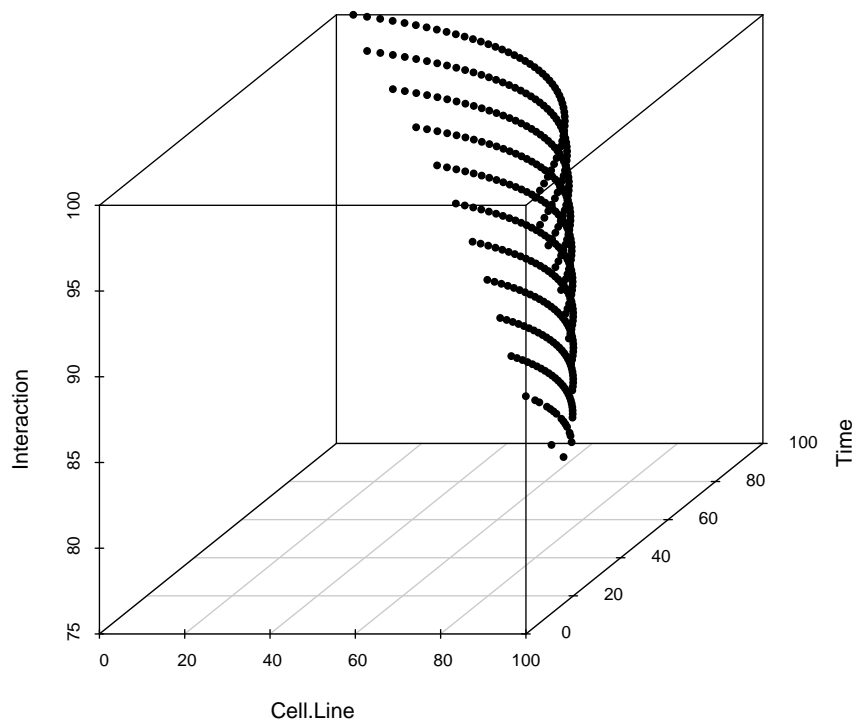


Figure 4.1: Relative efficiencies of subsets for 2×2 leukaemogenesis experiment with 100 slides for the exact set.

for this experiment are given in Table 4.4. A central composite experimental plan with 15 setting combinations is used. Each setting combination is replicated 4 times with a total of 60 runs. No repulsion is applied and acceptance rule 1 is used. The initial temperatures were chosen to correspond to approximate proportions of moves: 82.5%, 85%, 90%, 95% and 97.5%. The number of steps is 600. The quality measure used is the empirical logit of the number of Pareto optimal designs in the exact set missed by the application of the particular PSA setting, relative to the current reference set, U . The stopping criterion is the completion of a temperature level that results

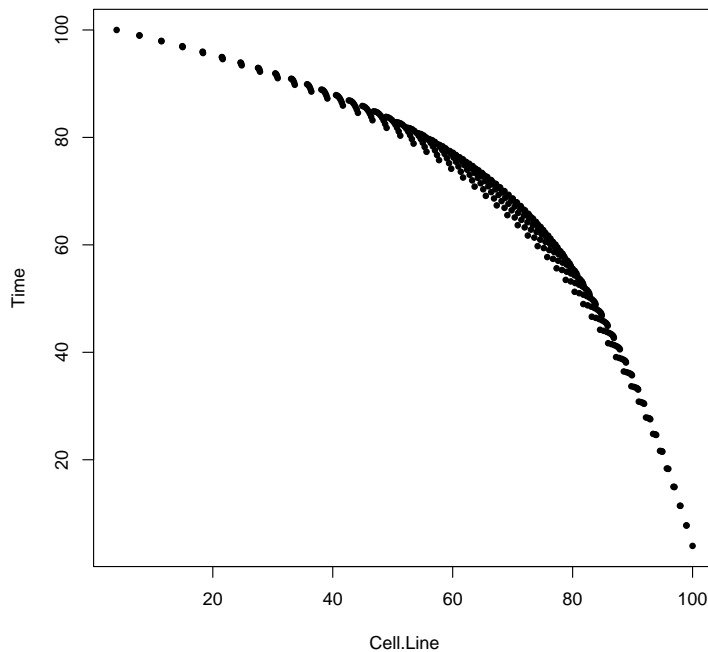


Figure 4.2: Top view for 2×2 leukaemogenesis experiment with 100 slides for the exact set.

in having visited at least 50,000 designs.

The resultant reference set U_1 consists of 371 potentially Pareto optimal designs and was constructed by finding all designs that are Pareto optimal among the 60 PP sets from the 60 PSA runs. The front view of U_1 , represented by relative efficiencies is shown in Figure 4.4. Comparison to the exact set shown in Figure 4.3 shows the same general form, but with a number of the Pareto optimal designs missed.

The results from the first PSA experimental plan, using U_1 as the reference set, were analysed using linear models in R and the associated output is given in Appendix B.5. For the linear model, the Adjusted R-squared is 59.7% and for the quadratic model the Adjusted R-squared is 59.2%. There is no suggestion that including quadratic terms improves the fit. The linear model indicates

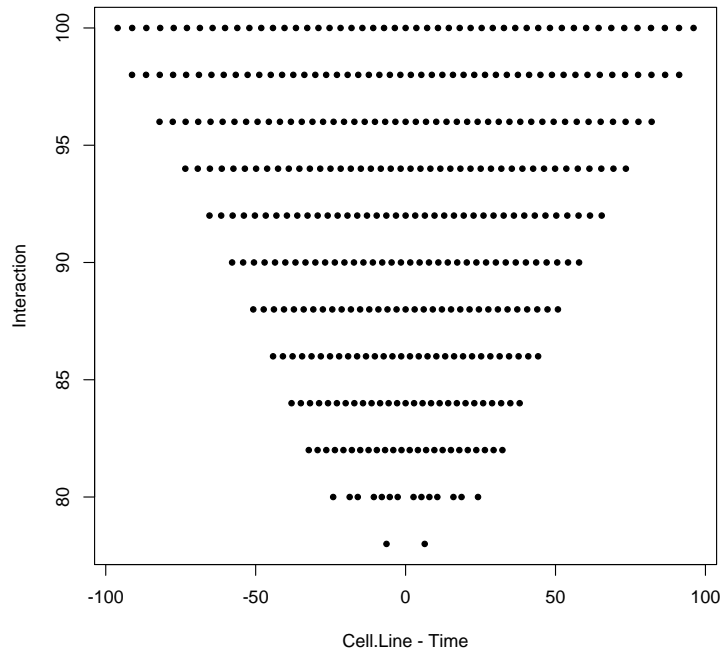


Figure 4.3: Front view for 2×2 leukaemogenesis experiment with 100 slides for the exact set.

that the algorithm will be improved if the initial temperature and cooling rate are decreased. Thus the method of steepest descent is applied, taking into account the estimates and feasible values of the coefficients for those parameters.

The second experimental plan consisted of a similar 60 runs to the first plan but used the parameter values given in Table 4.5. The initial temperatures were chosen to correspond to approximate proportions of moves: 70%, 72.5%, 77.5%, 82.5% and 85%.

The updated reference set U_2 consists of 381 potentially Pareto optimal designs and was constructed by finding all designs that are Pareto optimal among the 60 PP sets from the 60 PSA runs of the second plan together with the reference set from the first plan. The results from the second PSA experimental plan, using the updated set U_2 as the reference set, were analysed using linear

Table 4.4: Values for parameters that have 5 levels for the first experimental plan for the leukaemogenesis experiment with 100 slides

NG	1	3	7	11	13
Initial T rule 1	300	600	1200	2500	6000
CR	0.4	0.5	0.7	0.9	1

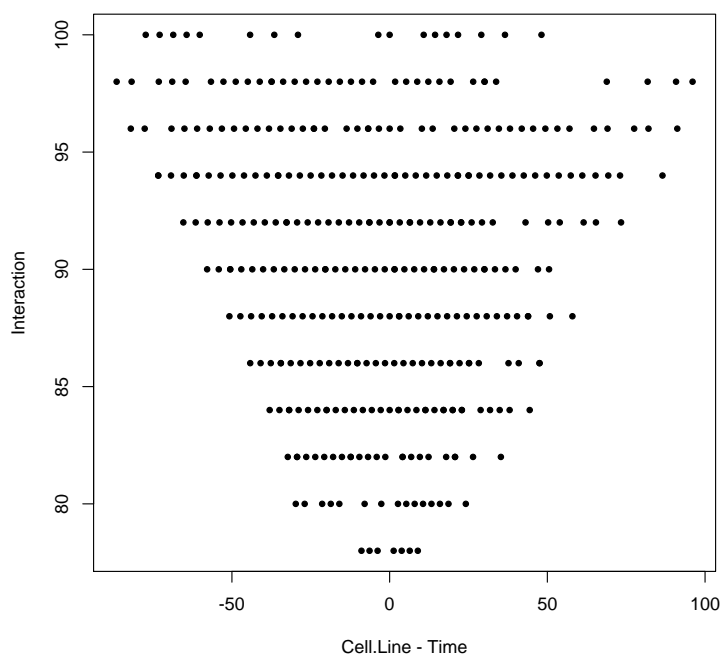


Figure 4.4: Front view for 2×2 leukaemogenesis experiment with 100 slides, adaptive PSA Plan 1.

models in R and the associated output is given in Appendix B.6. For this plan, the linear model has an adjusted R-squared of 24% whereas the quadratic model has an adjusted R-squared of 50%. The quadratic model thus implies being near the minimum of a quadratic surface. Thus no further

Table 4.5: Values for parameters that have 5 levels for the second experimental plan for the leukemogenesis experiment with 100 slides

NG	1	3	7	11	13
Initial T rule 1	25	40	90	300	600
CR	0.1	0.2	0.4	0.6	0.7

iterations are required and the optimisation step is carried out as follows.

The quadratic model is used to find the co-ordinates of the minimum using Solver in Excel. The optimum values of the PSA parameters within the region explored are found to be 11 for the number of generating designs, 25 for initial temperature and 0.6620 for the cooling rate. The algorithm was carried out from those optimum values for 120 runs. Then U_2 was updated from the set of 381 designs to find those Pareto optimal among the 381 designs as well as the 120 PP sets from the runs at the optimum. This results in the same 389 Pareto optimal designs obtained by the exact approach. In fact, the 120 runs at the optimum contain the exact Pareto optimal set of 389 designs. Thus, in this example, the adaptive algorithm is completely successful in finding both a suitable set of tuning parameters and recovering the exact Pareto optimal set.

Now consider the improvement obtained by applying the method of steepest descent from the first to the second experimental plan in terms of inspecting the data. Recall that application of the penalty approach is feasible for the exact method for this situation and results in obtaining 389 Pareto optimal designs. For the first composite experimental plan, it is found that 187 of the 389 designs are missed. The first plan also produced an additional 169 designs that are not Pareto optimal as they do not appear in the set obtained by the exact method.

The second composite experimental plan provides a substantial improvement as follows. For the

second plan, there are only 36 designs that were missed out of the set of 389 Pareto optimal designs obtained by the exact method and only an additional 28 designs that are not Pareto optimal. The improvement apparent in the second central composite design can also be seen by comparing the front view of U_2 shown Figure 4.5 to that of U_1 shown in Figure 4.4.

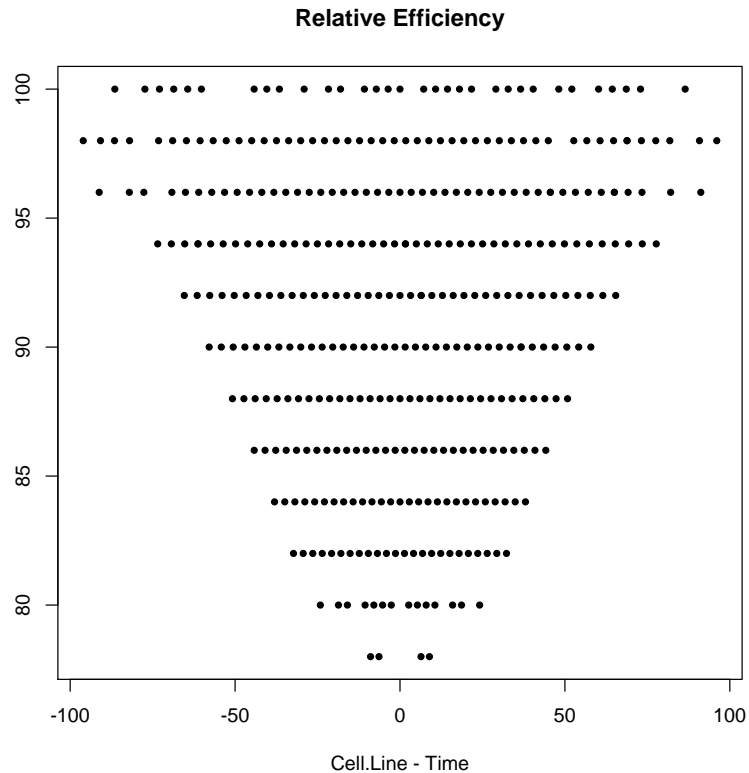


Figure 4.5: Front view for 2×2 leukaemogenesis experiment with 100 slides, adaptive PSA Plan 2.

Adaptive algorithm: 160 slides

Now suppose 160 slides are available for the 2×2 experimental situation. Having a large number of slides, for example, may be of interest in a clinical experiment where many patients are available and

one factor is a treatment with two levels and the other is time with two time points. It is infeasible timewise to find the set of all Pareto optimal designs when 160 slides are available. Therefore the adaptive algorithm is applied as follows.

Firstly, a central composite experimental plan is adopted arising from the parameter values given in Table 4.6. The initial temperatures were chosen to correspond to approximate proportions of moves: 42.5%, 50%, 65%, 80% and 87.5%.

Table 4.6: Values for parameters that have 5 levels for the experimental plan for the leukaemogenesis experiment with 160 slides

NG	10	20	40	60	70
Initial T	0.5	5	40	400	1200
CR	0.4	0.5	0.7	0.9	1

The central composite experimental plan consists of 15 setting combinations. Each setting combination for the central composite experimental plan was replicated 4 times, with a total of 60 runs. The number of steps is fixed to be 60 and the stopping criterion for each run is the completion of the temperature level that results in having visited at least 100,000 designs. The quality measure used is the empirical logit of the number of designs missed by the application of the particular PSA setting, taking the reference set to be U .

After the first central composite experimental plan was executed, the reference set U was constructed to be the designs deemed to be Pareto optimal among the 60 PP sets obtained from the runs. For this experiment, U was found to consist of 940 designs.

The results were analysed in R and the associated output is given in Appendix B.7. The linear model is not statistically significant ($P=0.22$). For the quadratic model, the adjusted R^2 is 54%.

In particular, note that all three interaction terms are statistically highly significant.

The quadratic model implies being near the minimum of a quadratic surface. Thus no further iterations are required and the optimisation step is carried out as follows.

The quadratic model is used to find the co-ordinates of the minimum using Solver in Excel. The optimum values of the PSA parameters within the region explored were found to be 70 for the number of generating designs, 0.5 for initial temperature and 0.6211 for the cooling rate.

The algorithm was carried out from the optimum values for 60 runs. Then U was updated from the set of 940 designs to find those Pareto optimal among those 940 designs as well as the 60 PP sets from the runs at the optimum. This leads to an updated U consisting of a set of 924 designs obtained by adding 172 designs and removing 188 designs. Of the 172 designs added, 156 of those are of the form that would be expected to be found based on earlier work, such as the leukaemogenesis experiment with 16 slides displayed in Table 3.3. The form of such designs expected to be found in the Pareto optimal set is those that allocate equal numbers of slides to configurations 1 and 4, configurations 2 and 5 and configurations 3 and 6. Each of those designs satisfy the constraints for the subsets of interest. The updated set U consists of 50 designs that provide minimum variance for interaction of 0.025. Each of those 50 designs satisfy the constraints. During the updating of U , due to carrying out 60 runs at the optimum, 10 such designs were added, including the design with 40 slides allocated to each of configurations 1, 2, 4 and 5. This design is highly desirable as it is among those in U with minimum variance of 0.025 for interaction and provides equal variance for cell line and time effects of 0.01875.

The improvement obtained in the optimisation step can also be seen by comparing Figures 4.6 and 4.7 that show respectively the front views for the reference set after the first adaptive step and then after the final optimisation. Note that in both cases, the relative efficiencies are with respect to the final reference set, U . This comparison shows an obvious improvement but the gaps

in Figure 4.7 suggest that a number of Pareto optimal designs have still been missed.

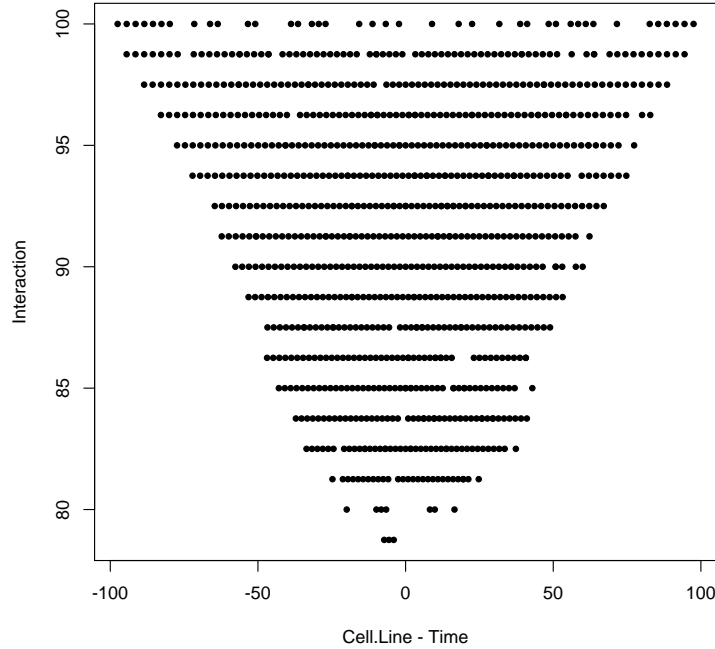


Figure 4.6: Front view for 2×2 leukaemogenesis experiment with 160 slides, adaptive PSA Plan 1.

Now consider fitting models for the first PSA plan for 160 slides using the average and worst case quality measures that were defined in Section 4.5.1 except that U is the reference set as described in Section 4.6.5. The analysis for both measures was carried out in R and the associated transcript is in Appendix B.8.

When either the average distance or worst case quality measure is used in place of the logit response, the optimum values of the PSA parameters for the number of generating designs and the initial temperature remain at the upper end and lower end of their ranges, 70 and 0.5 respectively. The optimum cooling rate is 0.64 for the average distance and 0.70 for the worst case quality

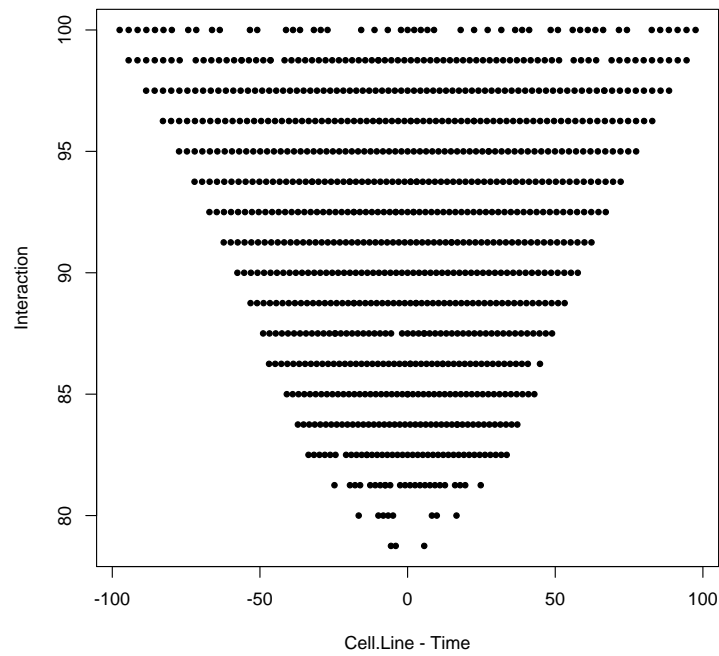


Figure 4.7: Front view for 2×2 leukaemogenesis experiment with 160 slides, adaptive PSA optimisation step.

measures, relatively near 0.62 found for the logit response quality measure. In this application, at least, the choice of quality measure has little effect on the optimum values of the PSA parameters.

4.8 Concluding comments

In this chapter, the metaheuristic method of Pareto simulated annealing was adapted to the microarray context. Further to this, algorithms were developed for tuning the Pareto simulated annealing parameters, based on the core algorithm as well as response surface methodology. Both the preliminary and adaptive algorithms were demonstrated to be effective strategies when illustrated for the 2×2 experimental situation.

The adaptive algorithm can be incorporated into the methodology for finding optimal designs as follows. As described earlier in this thesis, for a microarray experiment with clearly defined objectives and questions to be answered, an appropriate model can be formulated and relevant contrasts and constraints can be used to form suitable objective functions. This process is applicable regardless of whether it is feasible to carry out an enumerative search. When it is feasible to carry out an enumerative search, Pareto optimality can be applied to optimise the objective functions as detailed in Chapter 3. The adaptive algorithm developed in this chapter can be used to generate an approximation to the exact set of Pareto optimal designs thus catering for experiments where it is infeasible to carry out an enumerative search.

Chapter 5

Further comments

5.1 Introduction

Recall, from Chapter 2, technical replicates were described to be those for which samples used in different hybridizations originate from the same biological extraction but are labeled independently. Biological replicates, however, are those for which samples from different extractions are used. In the next section, an outline of the methodology that allows for the presence of both biological and technical replicates is provided. Following this, some further aspects of complex experiments are discussed and final comments are given. Abridged versions of Section 5.2. and Section 5.3. are presented in Sanchez and Glonek (2009).

5.2 Technical replication

The methodology developed earlier in this thesis has assumed the errors associated with the log-ratios are uncorrelated. This is the case when only biological replicates are permitted. If the same biological sample is applied to more than one slide, that is if technical replication is permitted, errors for slides that share biological material will be correlated. Thus the least squares analysis

and implied variance calculations do not apply. The methodology developed can accommodate technical replication by the adoption of a suitable linear mixed model.

Suppose r_k independent biological samples are obtained from population k for $k = 1, 2, \dots, K$. Then the total number of independent biological samples is given by $r_1 + r_2 + \dots + r_K = r$. The linear mixed model for the vector of log-ratios, \mathbf{M} , is given by

$$\mathbf{M} = X\boldsymbol{\beta} + Z\mathbf{b} + \mathbf{e} \quad (5.1)$$

where X and $\boldsymbol{\beta}$ are as given previously, \mathbf{b} is an r -dimensional vector of random effects corresponding to the r biological samples, Z is a $n \times r$ matrix indicating the allocation of biological samples to slides and \mathbf{e} is the experimental error.

In what follows, it is assumed that each of the random effects b_k are independent with $E(b_k) = 0$ and $\text{var}(b_k) = \sigma_b^2$ and also that $E(e_i) = 0$ and $\text{var}(e_i) = \sigma_e^2$ independently of the b_k . The biological variance is σ_b^2 and the experimental error variance is σ_e^2 . The allocation matrix Z is defined by

$$z_{ik} = \begin{cases} -1 & \text{if sample } k \text{ is used on the green channel of slide } i \\ 1 & \text{if sample } k \text{ is used on the red channel of slide } i \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for a given biological sample in column k of Z , the presence of technical replicates is indicated by two or more non-zero entries appearing in that column. Under the assumptions of the linear mixed model, it follows that

$$E(\mathbf{M}) = X\boldsymbol{\beta} \text{ and } \text{var}(\mathbf{M}) = \sigma_b^2 Z Z^T + \sigma_e^2 I.$$

As a simple illustration of the notation, consider the 2×2 example with 3 slides composed of a single hybridization for each of configurations 1, 2 and 3 from Table 3.2. This design represents hybridizing the baseline with each other treatment combination. Suppose that two independent

biological samples are available for the baseline (00) so that $r_1 = 2$. In particular, suppose that the first biological sample appears as technical replicates in the hybridizations utilizing configurations 1 and 3 and the second is used solely for the hybridization utilizing configuration 2. Let all other treatment combinations, $(a0)$, $(0b)$, and (ab) , consist of independent biological samples whereby $(r_2 = r_3 = r_4 = 1)$. Then

- the vector of log-ratios, $\mathbf{M} = (M_1, M_2, M_3)^T$,
- the design matrix

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

- the parameter vector, $\beta = (\alpha, \beta, \alpha\beta)^T$,
- the allocation matrix,

$$Z = \begin{pmatrix} -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix},$$

- the vector of random effects, $\mathbf{b} = (b(00, 1), b(00, 2), b(a0, 1), b(0b, 1), b(ab, 1))^T$,
- and the errors, $\mathbf{e} = (e_1, e_2, e_3)^T$.

In this simple illustration, the introduction of a technical replicate introduces a correlation between slides 1 and 3 resulting in

$$ZZ^T = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

rather than the identity matrix obtained when using only independent biological samples.

Various analyses are available for the mixed model (5.1), see for example Corbeil and Searle (1976). However, for the purpose of theoretical calculations of efficiency, consider the variance matrix

$$\text{var}(\hat{\gamma}) = \sigma_e^2 C \quad (5.2)$$

where

$$C = B(X^T V^{-1} X)^{-1} B^T \text{ and } V = (\sigma_b^2 / \sigma_e^2) Z Z^T + I. \quad (5.3)$$

A similar approach has been taken by Tsai et al. (2006) who considered A-optimal designs for a single factor experiment. In the present framework, it is natural to extend Definition 1 (see Section 3.3.2) to seek designs that are Pareto optimal based on the diagonal elements of C . The optimisation is over all design matrices X and allocation matrices Z compatible with the available biological material and slides.

To illustrate the optimisation over Z , consider the 2×2 factorial experiment and a design composed of a single slide for each of the configurations (00 vs $a0$), (00 vs $0b$) and (00 vs ab) given in Table 3.2. Suppose that three independent biological samples, S_1 , S_2 , S_3 , of the baseline condition 00 are available. Since all three slides include the baseline condition, there are 5 distinct allocations. For the purpose of numerical illustration, assume $\sigma_b^2 / \sigma_e^2 = 0.25$ (based on Cui and Churchill (2003)). Table 5.1 shows the variance coefficients for each of the five allocations. The first allocation, with the same biological sample applied to all three slides, is superior to the other four allocations in this case.

More generally, for a given experiment, the required optimisation is over all possible allocations of technical replicates for each possible design. The number of allocations to be considered grows rapidly as the size of the experiment increases so enumerative methods will be feasible only for very small problems. In realistic situations, an approximate search strategy will be needed such as Pareto simulated annealing.

Table 5.1: Variance coefficients under different allocations of the baseline biological samples.

Baseline sample applied							
(00 vs a0)	(00 vs 0b)	(00 vs ab)	c_α	c_β	$c_{\alpha\beta}$	$c_{\alpha+\alpha\beta}$	$c_{\beta+\alpha\beta}$
S_1	S_1	S_1	1.5	1.5	4	2.5	2.5
S_1	S_2	S_1	1.5	1.5	4	3	2.5
S_1	S_2	S_2	1.5	1.5	4	2.5	3
S_1	S_1	S_2	1.5	1.5	5	3	3
S_1	S_2	S_3	1.5	1.5	4.5	3	3

The linear mixed model accommodates technical replication under various assumptions. The following highlights some scientific issues that should be considered further regarding the use of the model. It should be recognised that the additive form of the model and that the variance ratio is known and is constant are simplified assumptions, constituting limitations of the approach. The calculations presented have focused purely on theoretical efficiency and do not take into account the practical questions of whether it would be possible to estimate the variances needed to analyse such an experiment. The results in Table 5.1 show that improvements in efficiency can, in principle, be achieved by careful allocation of technical replicates. However they are not intended as practical experimental designs. The lack of biological variation in the designs in Table 5.1 would make it impossible to estimate the biological variability. However it is conjectured that this would not be the case if more slides were available.

5.3 Complex experiments

Optimal experimental designs are motivated by considerations of statistical efficiency and as such provide the maximum information for a given amount of experimental effort. However, as pointed out by Wit et al. (2005), certain non-optimal designs, most notably the common reference design continue to be popular. In that design, all samples are hybridised against material from a common reference sample that is of no scientific interest in its own right and merely serves as a common standard for comparison. It has been shown that such designs are less efficient than optimal experimental designs (Kerr and Churchill, 2001), (Wit et al., 2005). Nevertheless, in the light of their popularity, it is worth considering the question of whether there are situations when the use of common reference designs can be justified.

Possible reasons for using a common reference design may include extensibility, robustness to failed hybridizations, elimination of dye effects and simplicity. Although a detailed examination of all of these issues is beyond the scope of this thesis, note that they can also be addressed in the context of optimal designs, see for example Wit et al. (2005). In what follows, the focus is on comparing the efficiency of optimal designs to common reference designs in the context of complex experiments.

The potential benefits of using a common reference design are generally not formulated in terms of statistical efficiency. Therefore the decision of whether to use an optimal design or a common reference design is a trade-off between statistical efficiency and other considerations. A situation in which the use of a common reference design may be justified is when optimal designs offer only small advantages in terms of statistical efficiency.

One context in which optimal designs might be expected to provide only modest gains in efficiency is in complex experiments where the number of contrasts of interest is substantially

greater than the available number of slides. However, as is shown in the following examples, it is possible to achieve substantial gains in this context.

Firstly, consider a single factor experiment with r experimental conditions so that there are $r(r-1)/2$ pairwise comparisons of interest. Suppose two biological samples are available under each condition and that $2r$ slides are available. This material could be used to produce two independent replicates of a common reference design. Alternatively, the same amount of biological material and available slides could be used to construct a connected interwoven loop design, a modification of the interwoven loop design in Wit et al. (2005), that utilizes technical replicates within each loop. Assuming, as in Section 5.2, that $\sigma_e^2 = 4\sigma_b^2$ it can be checked that the variance for any pairwise comparison in the common reference design is $1.25\sigma_e^2$. For $r = 5$ the maximum variance of any pairwise comparison from the connected interwoven loop is $0.68\sigma_e^2$. For $r = 9$ and $r = 12$, the maxima are respectively $0.93\sigma_e^2$ and $0.97\sigma_e^2$. For $r = 30$, the average variance for all 435 pairwise comparisons in the connected interwoven loop design is $0.997\sigma_e^2$ and the maximum variance for a pairwise comparison is $1.103\sigma_e^2$. Moreover, if not all comparisons are of equal interest, as might be the case in a time course experiment, then an optimal design tailored to the contrasts of interest could be expected to provide an even greater improvement over a common reference design. Thus, although the benefit in terms of statistical efficiency decreases as r increases, significant improvements are nevertheless possible for even large values of r .

For factorial experiments, the number of contrasts potentially of interest also increases rapidly with the number of levels of the factors. However, unlike the single factor experiment, it is not necessarily the case that the relative advantage of an optimal design diminishes as the number of treatment combinations in the experiment increases. To illustrate, consider a $2 \times r$ experiment and let μ_{ij} be the mean log expression for treatment combination (i, j) . Suppose primary interest is on the $\binom{r}{2}$ interaction contrasts $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'}$ for $i = 1, 2, \dots, r$; $j = 1, 2$. If two biological

samples are available at each treatment combination and a total of $4r$ slides are available then two replicates of a common reference design can be made and each interaction contrast can be estimated with variance $2\sigma_b^2 + 2\sigma_e^2$. On the other hand, the same material can be used to make two replicates of each “vertical” configuration $(i, 2)$ vs $(i, 1)$ and a single replicate of a loop design with respect to i at $j = 1$ and at $j = 2$. For this design, it can be checked that each interaction contrast can be estimated with variance no greater than $2\sigma_b^2 + \sigma_e^2$. Assuming $\sigma_e^2 = 4\sigma_b^2$, the reduction in variance compared to the common reference design will be at least 40% irrespective of r .

Now consider the scenario where the cost of slides is relatively low compared to the cost of biological samples. Then it may be feasible to increase the number of available slides and employ further technical replication based on a fixed amount of available biological material. As an example, consider the single factor experiment, with $r = 9$ and two biological samples available under each condition as previously. Now suppose it is feasible to produce twice as many slides. Then technical replicates of the common reference design and connected interwoven loop design can be used to accommodate the increase in number of slides to be produced from 18 to 36. Then the variance for any pairwise comparison in the common reference design is $0.75\sigma_e^2$ whereas the maximum variance of any pairwise comparison in the connected interwoven loop design is $0.417\sigma_e^2$. Thus the connected interwoven loop design provides a 44.4% reduction in the maximum variance of any pairwise comparison over the common reference design when 36 slides are available compared to 25.6% for 18 slides. This indicates that the connected interwoven loop design can provide an even greater improvement over the common reference design when the available amount of biological material is fixed and additional slides are used appropriately.

The preceding examples suggest that suitably chosen optimal designs can offer significant improvements in efficiency over common reference designs in a range of situations including experiments with large numbers of treatments. In this light, the approach of identifying and prioritizing

the scientifically relevant contrasts and then choosing a Pareto optimal design is well suited to experimental design for large and complex microarray experiments. An important practical issue to consider for complex experiments is that it is not feasible to examine all possible designs in an exhaustive search for Pareto optimal designs. Therefore an appropriate strategy for such problems would involve using a guided search algorithm such as Pareto simulated annealing.

5.4 Concluding comments

In this chapter, the design methodology developed earlier in this thesis has been expanded to provide an outline for dealing with microarray experiments that involve the use of both biological and technical replicates. In practice, it is expected that this methodology would be integrated into the overall strategy of design selection as given below.

1. Clearly state the objectives and questions to be answered for the microarray experiment to be carried out.
2. Formulate the appropriate statistical model. In cases where all biological replicates or all technical replicates are used in the experiment, the linear model given in (2.4) can be adopted. The mixed model given by Equation (5.1) is adopted when a mixture of biological and technical replicates is used.
3. Develop appropriate contrasts and constraints to construct objective functions.
4. Optimise the objective functions by applying Pareto optimality for cases where it is feasible to carry out an enumerative search. Otherwise perform adaptive Pareto simulated annealing.
5. Inspect the set of designs obtained from the optimisation of objectives and select an appropriate design.

Appendix A

Computer programs

Major parts of the code developed as part of this thesis have been provided on a CD that is enclosed with the hard copy of this thesis. The computer program concerned with Sanchez and Glonek (2009) can also be found as supplementary material at <http://biostatistics.oxfordjournals.org/cgi/content/abstract/10/3/561>.

Appendix B

R analysis

The R analysis associated with Chapter 4 is provided as follows.

B.1 Preliminary PSA 36 slides, plan 1

```
> y<-log((miss+0.5)/(63.5-miss))
> #Linear model:

Call:
lm(formula = y ~ designs + temp + cool + repel + ap)

Residuals:
    Min       1Q   Median       3Q      Max
-2.2365 -0.6103 -0.0605  0.5854  3.2329

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.04623    0.05627   0.821  0.41220
designs      -0.47120    0.06164  -7.644 5.40e-13 ***
temp         0.44854    0.06164   7.277 5.14e-12 ***
cool        -0.12908    0.06164  -2.094  0.03733 *
repel        0.16187    0.05627   2.877  0.00439 **
ap          -0.09430    0.05627  -1.676  0.09509 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8717 on 234 degrees of freedom
Multiple R-squared:  0.3515,    Adjusted R-squared:  0.3377
F-statistic: 25.37 on 5 and 234 DF,  p-value: < 2.2e-16
```

```
> #Quadratic model:
```

```
Call:
```

```
lm(formula = y ~ (designs + temp + cool + repel + ap)^2 + I(designs^2) +  
    I(temp^2) + I(cool^2))
```

```
Residuals:
```

	Min	1Q	Median	3Q	Max
	-1.445763	-0.448362	-0.002658	0.390807	2.606314

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.4476963	0.1510245	-2.964	0.003365	**
designs	-0.4712011	0.0480198	-9.813	< 2e-16	***
temp	0.4485380	0.0480198	9.341	< 2e-16	***
cool	-0.1290829	0.0480198	-2.688	0.007733	**
repel	0.1618736	0.0438359	3.693	0.000280	***
ap	-0.0943043	0.0438359	-2.151	0.032537	*
I(designs^2)	0.2971061	0.0723861	4.104	5.70e-05	***
I(temp^2)	0.2963381	0.0723861	4.094	5.95e-05	***
I(cool^2)	-0.0007375	0.0723861	-0.010	0.991880	
designs:temp	0.1012307	0.0600248	1.686	0.093114	.
designs:cool	0.3231378	0.0600248	5.383	1.86e-07	***
designs:repel	0.0527508	0.0480198	1.099	0.273172	
designs:ap	-0.1143109	0.0480198	-2.380	0.018139	*
temp:cool	0.4018140	0.0600248	6.694	1.76e-10	***
temp:repel	0.0534931	0.0480198	1.114	0.266499	
temp:ap	-0.1468815	0.0480198	-3.059	0.002497	**
cool:repel	0.1307536	0.0480198	2.723	0.006988	**
cool:ap	-0.1484250	0.0480198	-3.091	0.002252	**
repel:ap	-0.2013784	0.0438359	-4.594	7.31e-06	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.6791 on 221 degrees of freedom
```

```
Multiple R-squared: 0.6283,    Adjusted R-squared: 0.5981
```

```
F-statistic: 20.76 on 18 and 221 DF,  p-value: < 2.2e-16
```

B.2 Preliminary PSA 36 slides, plan 2

```
> #There are 63 Pareto optimal designs
> y<-log((miss+0.5)/(63.5-miss))
> psa2.l<-lm(y ~ designs + temp + cool)
> psa2.q<-lm(y ~ (designs + temp + cool)^2 + I(designs^2) + I(temp^2) + I(cool^2))
>
> #Summary of linear model
> summary(psa2.l)
```

Call:

```
lm(formula = y ~ designs + temp + cool)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.09614	-0.62847	0.01473	0.49650	2.16437

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.78642	0.05311	-14.808	< 2e-16 ***
designs	-0.06353	0.05818	-1.092	0.27597
temp	-0.16691	0.05818	-2.869	0.00449 **
cool	-0.02179	0.05818	-0.375	0.70834

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8228 on 236 degrees of freedom

Multiple R-Squared: 0.03894, Adjusted R-squared: 0.02673

F-statistic: 3.188 on 3 and 236 DF, p-value: 0.02449

```
> #Summary of quadratic model
> summary(psa2.q)
```

Call:

```
lm(formula = y ~ (designs + temp + cool)^2 + I(designs^2) + I(temp^2) +
    I(cool^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.373289	-0.284817	0.003137	0.290039	1.567284

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.21441	0.11294	-10.753	< 2e-16	***
designs	-0.06353	0.03591	-1.769	0.0782	.
temp	-0.16691	0.03591	-4.648	5.65e-06	***
cool	-0.02179	0.03591	-0.607	0.5446	
I(designs^2)	-0.09129	0.05413	-1.686	0.0931	.
I(temp^2)	0.75795	0.05413	14.002	< 2e-16	***
I(cool^2)	-0.15307	0.05413	-2.828	0.0051	**
designs:temp	0.19574	0.04489	4.361	1.96e-05	***
designs:cool	-0.02745	0.04489	-0.611	0.5415	
temp:cool	0.24093	0.04489	5.367	1.95e-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5078 on 230 degrees of freedom

Multiple R-Squared: 0.6432, Adjusted R-squared: 0.6292

F-statistic: 46.06 on 9 and 230 DF, p-value: < 2.2e-16

B.3 Preliminary PSA 36 slides, plan 2 with alternative quality measures

```
> #Average distance
> #Linear model for ave dist
> summary(psa.al)
```

```
Call:
lm(formula = average ~ designs + temp + cool)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-0.010758 -0.005414 -0.002916  0.001998  0.068182
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0057555  0.0007126   8.077 3.41e-14 ***
designs       -0.0004642  0.0007806  -0.595   0.553
temp        -0.0053863  0.0007806  -6.900 4.74e-11 ***
cool        -0.0006343  0.0007806  -0.813   0.417
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.01104 on 236 degrees of freedom
Multiple R-Squared: 0.1708, Adjusted R-squared: 0.1603
F-statistic: 16.21 on 3 and 236 DF, p-value: 1.299e-09
```

B.3. PRELIMINARY PSA 36 SLIDES, PLAN 2 WITH ALTERNATIVE QUALITY MEASURES121

```
> #Quadratic model for average distance
> summary(psa.aq)
```

Call:

```
lm(formula = average ~ (designs + temp + cool)^2 + I(designs^2) +
    I(temp^2) + I(cool^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.020368	-0.005356	0.001014	0.002482	0.054781

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.0062843	0.0020690	3.037	0.00266	**
designs	-0.0004642	0.0006579	-0.706	0.48109	
temp	-0.0053863	0.0006579	-8.188	1.83e-14	***
cool	-0.0006343	0.0006579	-0.964	0.33594	
I(designs^2)	-0.0031333	0.0009917	-3.160	0.00179	**
I(temp^2)	0.0057207	0.0009917	5.769	2.56e-08	***
I(cool^2)	-0.0032220	0.0009917	-3.249	0.00133	**
designs:temp	0.0008843	0.0008223	1.075	0.28333	
designs:cool	-0.0002324	0.0008223	-0.283	0.77770	
temp:cool	0.0012086	0.0008223	1.470	0.14302	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009304 on 230 degrees of freedom

Multiple R-Squared: 0.4261, Adjusted R-squared: 0.4036

F-statistic: 18.97 on 9 and 230 DF, p-value: < 2.2e-16

```
> #Worst case
> #Linear model for worst case
> summary(psa.wl)
```

```
Call:
lm(formula = worst ~ designs + temp + cool)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-0.053632 -0.024895 -0.006583  0.009119  0.149900
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.039442   0.002465  16.002 < 2e-16 ***
designs       -0.001154   0.002700  -0.427   0.670
temp        -0.019531   0.002700  -7.233 6.54e-12 ***
cool        -0.004144   0.002700  -1.535   0.126
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.03819 on 236 degrees of freedom
Multiple R-Squared: 0.1886, Adjusted R-squared: 0.1783
F-statistic: 18.29 on 3 and 236 DF, p-value: 1.057e-10
```


B.3. PRELIMINARY PSA 36 SLIDES, PLAN 2 WITH ALTERNATIVE QUALITY MEASURES123

```
> #Quadratic model for worst case
> summary(psa.wq)
```

Call:

```
lm(formula = worst ~ (designs + temp + cool)^2 + I(designs^2) +
    I(temp^2) + I(cool^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-8.160e-02	-1.364e-02	-9.705e-05	9.438e-03	1.520e-01

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.037742	0.006983	5.405	1.62e-07	***
designs	-0.001154	0.002220	-0.520	0.60389	
temp	-0.019531	0.002220	-8.796	3.40e-16	***
cool	-0.004144	0.002220	-1.866	0.06330	.
I(designs^2)	-0.009317	0.003347	-2.784	0.00582	**
I(temp^2)	0.021621	0.003347	6.460	6.18e-10	***
I(cool^2)	-0.010263	0.003347	-3.066	0.00243	**
designs:temp	0.004782	0.002776	1.723	0.08625	.
designs:cool	-0.000515	0.002776	-0.186	0.85297	
temp:cool	0.007749	0.002776	2.792	0.00568	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0314 on 230 degrees of freedom

Multiple R-Squared: 0.4653, Adjusted R-squared: 0.4444

F-statistic: 22.24 on 9 and 230 DF, p-value: < 2.2e-16

B.4 Comparison of Pareto simulated annealing to random search methods

```
> y<-log((miss+0.5)/(63.5-miss))
> summary(psarands.lm)
```

```
Call:
lm(formula = y ~ M)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-2.06346 -0.31663 -0.03852  0.30422  2.41473
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.78642     0.03734  -21.06  <2e-16 ***
M1           3.27665     0.05281   62.05  <2e-16 ***
M2           2.57172     0.05281   48.70  <2e-16 ***
M3           3.68495     0.05281   69.78  <2e-16 ***
M4           3.52705     0.05281   66.79  <2e-16 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.5785 on 1195 degrees of freedom
Multiple R-Squared: 0.8474, Adjusted R-squared: 0.8469
F-statistic: 1659 on 4 and 1195 DF, p-value: < 2.2e-16
```

B.5 Adaptive PSA 100 slides, plan 1

```
> y=log((miss+0.5)/(371.5-miss))
> #Linear model
> r100smhu_lm=lm(formula = y ~ designs + temp + cool)
> summary(r100smhu_lm)
```

Call:

```
lm(formula = y ~ designs + temp + cool)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.8629	-0.6628	0.0826	0.6920	1.6889

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.5617	0.1222	37.335	< 2e-16 ***
designs	0.2401	0.1338	1.794	0.0782 .
temp	0.6501	0.1338	4.857	9.97e-06 ***
cool	1.0677	0.1338	7.977	8.43e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9464 on 56 degrees of freedom

Multiple R-Squared: 0.6176, Adjusted R-squared: 0.5971

F-statistic: 30.15 on 3 and 56 DF, p-value: 9.82e-12

```
> #Quadratic model
> r100smhu_q=lm(formula = y ~ (designs + temp + cool)^2 + I(designs^2) +
+ I(temp^2) + I(cool^2))
> summary(r100smhu_q)
```

Call:

```
lm(formula = y ~ (designs + temp + cool)^2 + I(designs^2) + I(temp^2) +
    I(cool^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.02961	-0.62504	0.02317	0.90528	1.73193

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.3742	0.4233	10.334	5.25e-14	***
designs	0.2401	0.1346	1.784	0.0805	.
temp	0.6501	0.1346	4.830	1.33e-05	***
cool	1.0677	0.1346	7.933	2.11e-10	***
I(designs^2)	0.2471	0.2029	1.218	0.2290	
I(temp^2)	0.0980	0.2029	0.483	0.6312	
I(cool^2)	-0.1202	0.2029	-0.592	0.5563	
designs:temp	0.0481	0.1682	0.286	0.7761	
designs:cool	-0.1244	0.1682	-0.740	0.4630	
temp:cool	0.2056	0.1682	1.222	0.2275	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9516 on 50 degrees of freedom

Multiple R-Squared: 0.6548, Adjusted R-squared: 0.5927

F-statistic: 10.54 on 9 and 50 DF, p-value: 6.262e-09

>

B.6 Adaptive PSA 100 slides, plan 2

```
> #Linear model
> y=log((miss+0.5)/(381.5-miss))
> r100smh2allu_lm=lm(formula = y ~ designs + temp + cool)
>
> summary(r100smh2allu_lm)
```

Call:

```
lm(formula = y ~ designs + temp + cool)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.4573	-0.6999	-0.3148	0.4325	3.6859

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.6832	0.1457	25.272	< 2e-16 ***
designs	-0.6879	0.1596	-4.309	6.7e-05 ***
temp	0.1650	0.1596	1.033	0.306
cool	-0.2089	0.1596	-1.308	0.196

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.129 on 56 degrees of freedom

Multiple R-Squared: 0.276, Adjusted R-squared: 0.2372

F-statistic: 7.115 on 3 and 56 DF, p-value: 0.0003927

```
> #Quadratic model
> r100smh2allu_q=lm(formula = y ~ (designs + temp + cool)^2 + I(designs^2) +
+ I(temp^2) + I(cool^2))
>
> summary(r100smh2allu_q)
```

Call:

```
lm(formula = y ~ (designs + temp + cool)^2 + I(designs^2) + I(temp^2) +
    I(cool^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.1690	-0.4486	-0.1018	0.4214	2.3656

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.0525	0.4107	7.432	1.27e-09	***
designs	-0.6879	0.1306	-5.267	2.94e-06	***
temp	0.1650	0.1306	1.263	0.21236	
cool	-0.2089	0.1306	-1.600	0.11601	
I(designs^2)	0.5859	0.1969	2.976	0.00449	**
I(temp^2)	-0.1175	0.1969	-0.597	0.55320	
I(cool^2)	0.2884	0.1969	1.465	0.14913	
designs:temp	0.4437	0.1632	2.718	0.00900	**
designs:cool	0.2429	0.1632	1.488	0.14301	
temp:cool	0.5075	0.1632	3.109	0.00310	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9235 on 50 degrees of freedom

Multiple R-Squared: 0.5674, Adjusted R-squared: 0.4895

F-statistic: 7.287 on 9 and 50 DF, p-value: 1.105e-06

B.7 Adaptive PSA 160 slides

```
> y<-log((miss+0.5)/(940.5-miss))
> #Linear model
> r160f2l<-lm(y ~ designs + temp + cool)
> summary(r160f2l)
```

Call:

```
lm(formula = y ~ designs + temp + cool)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.09306	-0.93588	-0.09024	0.61093	3.08992

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.0296	0.1735	23.220	<2e-16 ***
designs	-0.1181	0.1901	-0.621	0.5370
temp	0.2174	0.1901	1.144	0.2576
cool	0.3207	0.1901	1.687	0.0971 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.344 on 56 degrees of freedom

Multiple R-Squared: 0.075, Adjusted R-squared: 0.02544

F-statistic: 1.513 on 3 and 56 DF, p-value: 0.221

```
> #Quadratic model
> r160f2q<-lm(y ~ (designs + temp + cool)^2 + I(designs^2) + I(temp^2) +I(cool^2))
> summary(r160f2q)
```

Call:

```
lm(formula = y ~ (designs + temp + cool)^2 + I(designs^2) + I(temp^2) +
    I(cool^2))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.4672	-0.6351	-0.1392	0.6743	2.0779

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.5895	0.4089	8.779	1.06e-11	***
designs	-0.1181	0.1300	-0.908	0.368088	
temp	0.2174	0.1300	1.673	0.100670	
cool	0.3207	0.1300	2.467	0.017095	*
I(designs^2)	0.0610	0.1960	0.311	0.756907	
I(temp^2)	-0.1512	0.1960	-0.771	0.444107	
I(cool^2)	0.6183	0.1960	3.155	0.002718	**
designs:temp	0.6023	0.1625	3.706	0.000527	***
designs:cool	0.7767	0.1625	4.780	1.58e-05	***
temp:cool	0.6652	0.1625	4.093	0.000155	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9193 on 50 degrees of freedom

Multiple R-Squared: 0.6137, Adjusted R-squared: 0.5442

F-statistic: 8.827 on 9 and 50 DF, p-value: 8.404e-08

B.8 Adaptive PSA 160 slides for alternative quality measures

```
> #Average distance
> #Linear model
> r160f2al<-lm(average ~ designs + temp + cool)
> summary(r160f2al)
```

Call:

```
lm(formula = average ~ designs + temp + cool)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.027010	-0.015047	-0.007048	0.006209	0.092335

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0190283	0.0032126	5.923	2.03e-07 ***
designs	-0.0006006	0.0035192	-0.171	0.865
temp	0.0040297	0.0035192	1.145	0.257
cool	0.0048855	0.0035192	1.388	0.171

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02488 on 56 degrees of freedom

Multiple R-Squared: 0.05513, Adjusted R-squared: 0.004511

F-statistic: 1.089 on 3 and 56 DF, p-value: 0.3613

```
> # Average distance measure
> # Quadratic model
> summary(r160f2aq)
```

Call:

```
lm(formula = average ~ (designs + temp + cool)^2 + I(designs^2) +
    I(temp^2) + I(cool^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.042639	-0.010235	-0.001050	0.005684	0.048347

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0043233	0.0075417	0.573	0.56904
designs	-0.0006006	0.0023979	-0.250	0.80325
temp	0.0040297	0.0023979	1.680	0.09910 .
cool	0.0048855	0.0023979	2.037	0.04692 *
I(designs^2)	0.0053483	0.0036147	1.480	0.14526
I(temp^2)	0.0025259	0.0036147	0.699	0.48792
I(cool^2)	0.0097718	0.0036147	2.703	0.00935 **
designs:temp	0.0127256	0.0029974	4.245	9.45e-05 ***
designs:cool	0.0146138	0.0029974	4.875	1.14e-05 ***
temp:cool	0.0137074	0.0029974	4.573	3.19e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01696 on 50 degrees of freedom

Multiple R-Squared: 0.6083, Adjusted R-squared: 0.5378

F-statistic: 8.628 on 9 and 50 DF, p-value: 1.157e-07

```
> # Worst case distance measure
> # Linear model
> r160f2wl<-lm(worst ~ designs + temp + cool)
```

```
summary(r160f2wl)
```

```
Call:
```

```
lm(formula = worst ~ designs + temp + cool)
```

```
Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.10546	-0.05873	-0.02800	0.03673	0.24928

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.090673	0.010647	8.516	1.10e-11	***
designs	-0.006983	0.011663	-0.599	0.552	
temp	0.010552	0.011663	0.905	0.369	
cool	0.011803	0.011663	1.012	0.316	

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.08247 on 56 degrees of freedom
```

```
Multiple R-Squared: 0.03782, Adjusted R-squared: -0.01373
```

```
F-statistic: 0.7337 on 3 and 56 DF, p-value: 0.5363
```

```
> #Worst case distance measure
> # Quadratic model
> summary(r160f2wq)
```

Call:

```
lm(formula = worst ~ (designs + temp + cool)^2 + I(designs^2) +
    I(temp^2) + I(cool^2))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.150838	-0.033369	-0.003733	0.018047	0.137701

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.037505	0.025747	1.457	0.151456
designs	-0.006983	0.008187	-0.853	0.397749
temp	0.010552	0.008187	1.289	0.203353
cool	0.011803	0.008187	1.442	0.155590
I(designs^2)	0.019724	0.012341	1.598	0.116279
I(temp^2)	0.006770	0.012341	0.549	0.585737
I(cool^2)	0.037308	0.012341	3.023	0.003939 **
designs:temp	0.042949	0.010233	4.197	0.000111 ***
designs:cool	0.039759	0.010233	3.885	0.000301 ***
temp:cool	0.046470	0.010233	4.541	3.55e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05789 on 50 degrees of freedom

Multiple R-Squared: 0.5768, Adjusted R-squared: 0.5006

F-statistic: 7.571 on 9 and 50 DF, p-value: 6.746e-07

Bibliography

- Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978). *Statistics for Experimenters*. New York: J. Wiley & Sons.
- Churchill, G. A. . (2004). Using anova to analyze microarray data. *BioTechniques* 37(2), 173–177.
- Churchill, G. A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics Supplement* 32, 490–495.
- Corbeil, R. R. and S. R. Searle (1976). Restricted maximum likelihood (reml) estimation of variance components in the mixed model. *Technometrics* 18(1), 31–38.
- Cui, X. Q. and G. A. Churchill (2003). *How many mice and how many arrays? Replication in mouse cDNA microarray experiments*. In K. F. Johnson and S. M. Lin (Ed.), *Methods of Microarray Data Analysis III*. Kluwer Academic Publishers, Norwell, MA.
- Czyzak, P. and A. Jaszkievicz (1998). Pareto simulated annealing-a metaheuristic technique for multiple objective combinatorial optimisation. *Journal of Multi-Criteria Decision Analysis* 7, 34–47.
- Dobbin, K. K., E. S. Kawasaki, D. W. Petersen, and R. M. Simon (2005, May). Characterizing dye bias in microarray experiments. *Bioinformatics* 21(10), 2430–2437.

- Glonek, G. F. V. and P. J. Solomon (2004). Factorial and time course designs for cDNA microarray experiments. *Biostatistics* 5, 89–111.
- Gregory, P. A., A. G. Bert, E. L. Paterson, S. C. Barry, A. Tsykin, G. Farshid, M. A. Vadas, Y. Khew-Goodall, and G. J. Goodall (2008). The mir-200 family and mir-205 regulate epithelial to mesenchymal transition by targeting zeb1 and sip1. *Nature cell biology* 10(5), 593–601.
- Jaszkiewicz, A. (2001). *Multiple Objective Metaheuristic Algorithms for Combinatorial Optimization*. Ph. D. thesis, Poznan University of Technology, Poznan.
- Kerr, M. and G. A. Churchill (2001). Experimental design for gene expression microarrays. *Biostatistics* 2, 183–201.
- Kerr, M. K. . (2003). Linear models for microarray data analysis: Hidden similarities and differences. *Journal of Computational Biology* 10, 891–901.
- Martin-Magniette, M.-L., J. Aubert, E. Cabannes, and J.-J. Daudin (2005, May). Evaluation of the gene-specific dye bias in cDNA microarray experiments. *Bioinformatics* 21(9), 1995–2000.
- Nguyen, D., A. A., N. Wang, and R. Carroll (2002). DNA microarray experiments: biological and technical aspects. *Biometrics* 58, 701–717.
- Pirlot, M. (1996). General local search methods. *European Journal of Operational Research* 92, 493–511.
- Sanchez, P. S. and G. F. V. Glonek (2009). Optimal designs for two-colour microarray experiments. *Biostatistics* 10(3), 561–74.
- Searle, S. (1971). *Linear Models*. New York: J. Wiley & Sons.

- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 1(3).
- Smyth, G. K., Y. H. Yang, and T. Speed (2003). Statistical issues in cDNA microarray data analysis. *Methods in Molecular Biology* 224, 111–136.
- Speed, T. P. and Y. H. Yang (2002). Direct versus indirect designs for cDNA microarray experiments. *Sankhya: the Indian Journal of Statistics* 64, 706–720.
- Tsai, S. F., C. T. Liao, and F. S. Chai (2006). Statistical designs for two-color microarray experiments involving technical replication. *Computational Statistics & Data Analysis* 51(3), 2078–2090.
- Wit, E., A. Nobile, and R. Khanin (2005). Near-optimal designs for dual channel microarray studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54(5), 817–830.
- Yang, Y., S. Dudoit, P. Luu, D. Lin, V. Peng, J. Ngai, and T. P. Speed (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* 30(4), e15.
- Yang, Y. H. and T. P. Speed (2002). Design issues for cDNA microarray experiments. *Nature Reviews Genetics* 3, 579–588.
- Yang, Y. H. and T. P. Speed (2003). *Design and analysis of comparative microarray experiments*. In T. P. Speed (Ed.), *Statistical analysis of gene expression microarray data*. Boca Raton: CRC Press.
- Yates, F. (1935). Complex experiments. *Royal Statistical Society* 2, 181–247.