

# PRECISION OF THE PATH OF STEEPEST ASCENT IN RESPONSE SURFACE METHODOLOGY

Ewa M. Sztendur  
B.Sc., Victoria University (1997)



THIS THESIS IS PRESENTED IN FULFILMENT OF  
THE REQUIREMENTS OF THE DEGREE OF  
DOCTOR OF PHILOSOPHY

SCHOOL OF COMPUTER SCIENCE AND MATHEMATICS  
FACULTY OF HEALTH, ENGINEERING AND SCIENCE

VICTORIA UNIVERSITY  
December 2005

© Copyright 2005

by

Ewa M. Sztendur

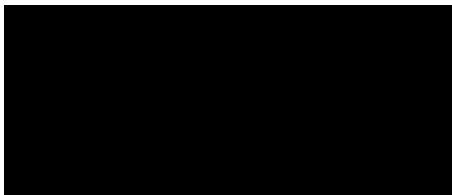
The author hereby grants to Victoria University  
permission to reproduce and distribute copies of this thesis  
document in whole or in part.

FTS THESIS  
519.57 SZT  
30001008800254  
Sztendur, Ewa M  
Precision of the path of  
steepest ascent in response  
surface methodology

# Declaration

I, Ewa Malgorzata Sztendur, declare that the PhD thesis entitled “Precision of the Path of Steepest Ascent in Response Surface Methodology” is no more than 100,000 words in length, exclusive of tables, appendices, references, and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

I also hereby declare that this thesis is written in accordance with the University’s Policy with respect to the Use of Project Reports and Higher Degree Theses.



Signature

23/02/06

Date

# Dedication

I dedicate this work to my family with whom my success is shared.

# Abstract

This thesis provides some extensions to the existing method of determining the precision of the path of steepest ascent in response surface methodology to cover situations with correlated and heteroscedastic responses, including the important class of generalised linear models. It is shown how the eigenvalues of a certain matrix can be used to express the proportion of included directions in the confidence cone for the path of steepest ascent as an integral, which can then be computed using numerical integration. In addition, some tight inequalities for the proportion of included directions are derived for the two, three, and four dimensional cases. For generalised linear models, methods are developed using the Wald approach and profile likelihood confidence regions approach, and bootstrap methods are used to improve the accuracy of the calculations.

Regions where quadratic (second order) models are needed occur in many instances. Canonical analysis and ridge analysis are techniques used for examining higher dimensional quadratic response surfaces. Ridge analysis corresponds to the path of steepest ascent for second order response surfaces. In the thesis, a method for determining the precision of an estimated canonical axis is developed, and a method for the precision of the path of steepest ascent for ridge analysis is proposed.

The thesis also concerns the question of how to augment the existing design if the precision of the path of steepest ascent is too low. It is shown how to augment the experimental design with additional runs to maximise the proportion of directions excluded from the confidence cone. An important aspect of power, in this context, is designing an experiment so that the proportion of directions excluded from the confidence cone for the path of steepest ascent is sufficiently

high. A method is described that can be useful for the determination of the sample size needed to achieve the required precision. Further, methods for generalised linear models when the data has a Binomial or Poisson distribution are also considered.

In addition to the above, the thesis describes some solutions for cases when the experimental variables are constrained. If a linear constraint is included, the calculated path can be modified by projecting it onto the constraint plane once the constraint plane is reached. It is shown how the precision of the constrained path of steepest ascent can be easily determined.

# Acknowledgements

I would like to thank my supervisor, Dr Neil Diamond, whose tolerance and help has been invaluable. I consider myself extremely fortunate for having had the opportunity to work with a person of Dr Diamond's calibre. I thank Dr Diamond for sharing his time and knowledge as well as for his constant encouragement and guidance.

During the course of "writing up" this thesis, Dr Diamond left Victoria University. Knowing how busy he was, I would like to thank him for his responsible and caring attitude towards completion of my thesis.

A warm "Thank you" to Dr Lutfar Khan for agreeing to become my supervisor and for his help after Dr Diamond left Victoria University.

I would like to thank the School of Computer Science and Mathematics at Victoria University for providing an excellent learning environment and friendly atmosphere. I am particularly indebted to Associate Professor Neil Barnett for his support and for providing me with a research scholarship during my studies.

I thank Associate Professor Pietro Cerone for reading and suggesting ideas in presenting the thesis.

I am very grateful to all members of the academic, technical and administrative staff for their always helpful and friendly assistance. In particular, I would like to thank Iwona Miliszewska, my compatriot, for her encouragement and wise advice.



I would like to thank Florica Șt. Cîrstea and George Hanna for their friendship. and all the chats and laughs that made my work at the School very enjoyable.

I would like to thank Professor Radmila Kiprianova for allowing me to use her data and also for her encouragement.

Many heartfelt thanks go to my family: my husband, Dariusz, and my sons, Michal and Maciek. I am very grateful for their understanding, love, patience and support.

# Author's Publications During the Candidature

Sztendur, E.M. and N.T. Diamond (1999), Confidence interval calculations for the path of steepest ascent in response surface methodology, *in* 'The Proceedings of the Collaborative Research Forum', School of Communications and Informatics, Victoria University of Technology, Melbourne, pp. 129–132.

Sztendur, E.M. and N.T. Diamond (2001), Inequalities for the precision of the path of steepest ascent in response surface methodology, *in* Y. Cho, J. Kim and S. Dragomir, eds, 'Inequality Theory and Applications', Vol. 1, Nova Science Publishers, Inc., New York, pp. 301–307.

Sztendur, E.M. and N.T. Diamond (2002), 'Extensions to confidence region calculations for the path of steepest ascent', *Journal of Quality Technology* **34**, 289–296.

# Notation

The density function of a random variable  $X_1$  with a  $t$ -distribution with  $\nu_1$  degrees of freedom is written as  $t(x_1, \nu_1)$ . The distribution function is written as  $T(x_1, \nu_1)$ , while the inverse function where

$$\Pr(X_1 \leq x_1) = p_1$$

is written as  $T_{p_1}(\nu_1)$ .

The distribution function of a random variable  $X_2$  with a  $\chi^2$ -distribution with  $\nu_2$  degrees of freedom is written as  $\chi^2(x_2, \nu_2)$ , while the inverse function where

$$\Pr(X_2 \leq x_2) = p_2$$

is written as  $\chi^2_{p_2}(\nu_2)$ .

The distribution function of a random variable  $X_3$  with an  $F$ -distribution with  $\nu_3$  and  $\nu_4$  degrees of freedom is written as  $F(x_3, \nu_3, \nu_4)$ , while the inverse function where

$$\Pr(X_3 \leq x_3) = p_3$$

is written as  $F_{p_3}(\nu_3, \nu_4)$ .

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Precision of the Path of Steepest Ascent . . . . .	6
1.3	Ridge Analysis . . . . .	7
1.4	Nonlinear Experimental Design . . . . .	12
1.5	Further Developments in RSM . . . . .	14
1.6	Method of Steepest Ascent for GLMs . . . . .	14
1.7	Constrained Experimental Regions . . . . .	15
1.8	Thesis Outline . . . . .	18
<b>2</b>	<b>Non-Orthogonal Designs with Heterogeneous Errors</b>	<b>19</b>
2.1	Introduction . . . . .	19
2.2	Derivation of the Existing Method . . . . .	21
2.3	An Example . . . . .	23
2.4	Extension to Non-Orthogonal Designs with Heterogeneous Errors	27
2.4.1	Continuation of the Example . . . . .	30
2.5	Inequalities for Percentage Included . . . . .	32
2.6	Other Approaches to Evaluating the Percentage Included. . . . .	40
2.7	Computations . . . . .	41
2.8	Conclusion . . . . .	42
<b>3</b>	<b>Application to Generalised Linear Models</b>	<b>44</b>
3.1	Introduction . . . . .	44
3.2	Wald Approach . . . . .	46

3.3	Likelihood Based Confidence Regions . . . . .	50
3.4	Bootstrap Approach . . . . .	55
3.4.1	Parametric Bootstrap . . . . .	55
3.4.2	Nonparametric Bootstrap . . . . .	56
3.5	Conclusion . . . . .	61
<b>4</b>	<b>Power Considerations</b>	<b>62</b>
4.1	Introduction . . . . .	62
4.2	Power Analysis for Homoscedastic Data . . . . .	63
4.3	Distribution of the Estimated Precision for Homoscedastic Data. .	65
4.4	Power Analysis for Heteroscedastic Data . . . . .	69
4.5	Improving the Precision of the Path . . . . .	72
4.6	Best Designs to Estimate the Path for GLMs . . . . .	76
4.6.1	Introduction . . . . .	76
4.6.2	Poisson Data . . . . .	76
4.6.3	Binomial Data . . . . .	80
4.7	Conclusion . . . . .	85
<b>5</b>	<b>Constraints and Second Order Models</b>	<b>87</b>
5.1	Introduction . . . . .	87
5.2	Linear Input Constraints . . . . .	88
5.2.1	An Example . . . . .	89
5.3	Second Order Models . . . . .	91
5.4	Examples . . . . .	92
5.4.1	Two Dimensional Example . . . . .	92
5.4.2	Three Dimensional Example . . . . .	97
5.4.3	Four Dimensional Example . . . . .	101
5.4.4	Five Dimensional Example . . . . .	103
5.5	Uncertainty of Canonical Direction . . . . .	107
5.5.1	Two Dimensional Example . . . . .	107
5.5.2	Three Dimensional Example . . . . .	110
5.5.3	Four Dimensional Example . . . . .	111

5.5.4	Five Dimensional Example . . . . .	112
5.6	Precision of Ridge Analysis . . . . .	113
5.7	Conclusion . . . . .	114
<b>6</b>	<b>Discussion and Further Work</b>	<b>116</b>
6.1	Introduction . . . . .	116
6.2	Homogeneous Case . . . . .	117
6.3	Heterogeneous Case . . . . .	118
6.4	Augmenting . . . . .	119
6.5	Constraints and Second Order Models . . . . .	120
6.6	Limitations . . . . .	120
6.6.1	Correct Model Form . . . . .	120
6.6.2	Correct Link . . . . .	121
6.6.3	Overdispersion . . . . .	121
6.6.4	Known Constraint Functions . . . . .	121
6.7	Future Work . . . . .	122
6.7.1	Consideration of Bias and Variance . . . . .	122
6.7.2	GLMs with Varying Dispersion . . . . .	122
6.7.3	Comparison of D-Optimal and Minimum Proportion Designs	123
6.7.4	Allocation of Observations for Binomial Data . . . . .	123
6.7.5	Bayesian Designs . . . . .	124
6.7.6	Application to Designs with Randomisation Restrictions .	124
6.7.7	Multiple Responses . . . . .	124
6.7.8	Estimated Constraint Functions . . . . .	125
6.7.9	GLMs and Second Order Models . . . . .	125
6.7.10	Canonical Analysis and Ridge Analysis . . . . .	125
	<b>Appendices</b>	<b>126</b>
<b>A</b>	<b>Splus commands for Chapter 2</b>	<b>127</b>
A.1	Data set used in chapters 2 and 3 . . . . .	127
A.2	Analysis using transformed response and original design . . . . .	127

A.3	Analysis using transformed response and modified design . . . . .	128
A.4	Gaussian quadrature functions . . . . .	128
A.5	perinc . . . . .	130
<b>B</b>	<b>Splus commands for Chapter 3</b>	<b>133</b>
B.1	Analysis using GLMs and original design . . . . .	133
B.2	Analysis using GLMs and modified design . . . . .	133
B.3	perincglm . . . . .	134
B.4	profpath . . . . .	136
B.5	contourprop . . . . .	138
B.6	boot3 . . . . .	139
B.7	boot4 . . . . .	140
<b>C</b>	<b>Splus commands for Chapter 4</b>	<b>141</b>
C.1	powerfn . . . . .	141
C.2	approxp . . . . .	142
C.3	heterofn . . . . .	143
C.4	heteropowerlik . . . . .	144
C.5	augdet . . . . .	145
C.6	augfun . . . . .	145
C.7	Commands to determine best set of augmenting runs . . . . .	147
C.8	poisfund . . . . .	148
C.9	poisfunp . . . . .	148
C.10	poissim . . . . .	149
C.11	bcbfn3 . . . . .	150
C.12	binfund . . . . .	151
C.13	binomfunp . . . . .	152
C.14	binsim . . . . .	152
C.15	bcbfn4 . . . . .	154
<b>D</b>	<b>Splus commands for Chapter 5</b>	<b>155</b>
D.1	Commands for linear constraints . . . . .	155

D.2 Two dimensional example . . . . .	156
D.3 Three dimensional example . . . . .	156
D.4 Four dimensional example . . . . .	157
D.5 Five dimensional example . . . . .	158
D.6 ridgeplot . . . . .	158
D.7 maxplot . . . . .	159
D.8 canonical . . . . .	161
D.9 dlr . . . . .	162
D.10 Direct fitting of canonical form-two dimensions . . . . .	163
D.11 Direct fitting of canonical form-three dimensions . . . . .	164
D.12 Direct fitting of canonical form-four dimensions . . . . .	165
D.13 Calculation of precision of canonical axis-three dimensions . . . . .	167
D.14 Calculation of precision of canonical axis-four dimensions . . . . .	167
<b>Bibliography</b>	<b>169</b>



# List of Tables

2.1	Part of the data for the car grille opening panels experiment. . . .	24
2.2	Summary of regression results using original design. . . . .	25
2.3	Summary of regression results using modified design. . . . .	31
2.4	Percentage of directions excluded from the 95% confidence cone, calculated using Gaussian quadrature. . . . .	32
3.1	Summary of fitted GLM model with Poisson link and original design.	47
3.2	Summary of fitted GLM model with Poisson link and modified design. . . . .	49
3.3	Percentage of directions excluded from the 95% confidence cone, using Wald and profile likelihood methods. . . . .	55
4.1	Mean and standard deviation of proportion of directions included in the 95% confidence cone for the path of steepest ascent versus the effect size. . . . .	73
4.2	Best sets of augmenting runs and predicted percentage of directions included in the 95% confidence cone using D-optimality. . . . .	75
4.3	Best sets of augmenting runs and predicted percentage included in the 95% confidence cone using direct minimisation, assuming no overdispersion. . . . .	75
4.4	Best sets of augmenting runs and predicted percentage included in the 95% confidence cone using direct minimisation, assuming overdispersion. . . . .	76
4.5	D-optimal and minimum proportion designs for Poisson experi- ments with 4 runs constrained between $-1$ and $1$ . . . . .	77
4.6	Ranges for the factors for the Poisson simulation study. . . . .	78

4.7 D-optimal and minimum proportion designs for Poisson experiments with 5 runs constrained between  $-1$  and  $1$ . . . . . 80

4.8 D-optimal and minimum proportion designs for Poisson experiments with 6 runs constrained between  $-1$  and  $1$ . . . . . 81

4.9 D-optimal and minimum proportion designs for Binomial experiments with 4 runs of 50 trials constrained between  $-1$  and  $1$ . . . . 83

4.10 Ranges for the factors for the Binomial simulation study. . . . . 84

5.1 Summary of results of fitting canonical models using non-linear least squares to the data given by Box et al. (1978, p.519). . . . . 96

5.2 Symmetrical composite design type  $B_5$  for intrinsic viscosity model. 106

# List of Figures

2.1	Confidence cone for the path of steepest ascent. . . . .	20
2.2	95% confidence cone for the path of steepest ascent when $B$ is fixed, involving $D$ and $F$ only, for the example given in Table 2.1.	26
2.3	Proportion of directions included in the 95% confidence cone for the path of steepest ascent for various values of $a_1$ and $a_2$ . . . . .	35
2.4	Percentage error of upper bound for the proportion of directions included in the 95% confidence cone. . . . .	36
3.1	95% profile likelihood confidence regions for the path of steepest ascent. . . . .	53
3.2	Comparison of bootstrap likelihood ratio statistics and theoretical quantiles assuming $\chi^2(2)$ distribution. . . . .	57
3.3	Comparison of bootstrap likelihood ratio statistics and theoretical quantiles assuming $F(2, 10)$ distribution. . . . .	59
3.4	95% profile likelihood region for path of steepest ascent based on bootstrap quantiles. . . . .	60
4.1	Power graph for the path of steepest ascent for 5 factors and 8 runs for the homogeneous case. . . . .	66
4.2	Power graph for the path of steepest ascent for 5 factors and 16 runs for the homogeneous case. . . . .	67
4.3	Power graph for the path of steepest ascent for 5 factors and 32 runs for the homogeneous case. . . . .	68
4.4	Distribution of the estimated precision for homoscedastic data. . .	69

4.5	Values of the expected proportion of directions included in the 95% confidence cone for the path of steepest ascent as a function of $\theta$ and $\phi$ . . . . .	71
4.6	Boxplots of proportion of directions included in the 95% confidence cone for the path of steepest ascent versus the effect size. . . . .	72
4.7	D-optimal and minimum proportion designs for Poisson experiments with 4 runs constrained between $-1$ and $1$ . . . . .	78
4.8	Histogram of efficiencies of the D-optimal design relative to the minimum proportion design for the Poisson simulation study using a 100 run orthogonal array. . . . .	79
4.9	D-optimal and minimum proportion designs for Poisson experiments with 5 runs constrained between $-1$ and $1$ . . . . .	81
4.10	D-optimal and minimum proportion designs for Poisson experiments with 6 runs constrained between $-1$ and $1$ . . . . .	82
4.11	D-optimal and minimum proportion designs for Binomial experiments with 4 runs of 50 trials constrained between $-1$ and $1$ . . . . .	83
4.12	Histogram of efficiencies of the D-optimal design relative to the minimum proportion design for the Binomial simulation study using a 100 run orthogonal array. . . . .	85
5.1	Ridge plot for the two factor chemical example problem given by Box et al. (1978, p. 519). . . . .	94
5.2	Maximum ridge co-ordinates and predicted maximum for the two factor chemical example given by Box et al. (1978, p. 519). . . . .	95
5.3	Ridge plot for the three factor reactor study example given by Box and Draper (1987, p. 362). . . . .	98
5.4	Maximum ridge co-ordinates and predicted maximum for the three factor reactor study example given by Box and Draper (1987, p. 362). . . . .	99
5.5	Ridge plot for the four factor helicopter example given by Box and Liu (1999). . . . .	104
5.6	Optimum ridge co-ordinates and predicted maximum for the four factor helicopter example given by Box and Liu (1999). . . . .	105

5.7 Ridge plot for the five factor polymilization acrylamide experiment  
given by Kiprianova and Markovska (1993). . . . . 108

5.8 Optimum ridge co-ordinates and predicted minimum for the five  
factor polymilization acrylamide experiment given by Kiprianova  
and Markovska (1993) . . . . . 109

# Chapter 1

## Introduction

### 1.1 Background

With the increased focus on experimental design, response surface methods have received considerable attention in recent decades. This interest has grown from the need for quality and precision in industry. Statistically designed experiments are one of the most powerful tools in statistical analysis as they can greatly increase the efficiency of experimenters. The aim of much experimentation is to find out how a number of experimental variables affect a response, and to find the combination of conditions that provides the highest response, as well as to understand the relationship over a region of interest (see, for example Box and Liu, 1999; Box, 1999).

Response surface methodology (RSM) originated with the work of Box and Wilson (1951), who were at the time involved in industrial research with ICI in the United Kingdom. There are many situations for which RSM has proved to be a very useful tool. Hill and Hunter (1966) illustrated chemical and processing applications of canonical analysis and use of multiple responses. Mead and Pike (1975) investigated the extent to which RSM had been used in applied research and gave examples from biological applications. Myers, Khuri and Carter (1989) summarised the developments in RSM that had occurred since the review of Hill and Hunter (1966), while a more recent summary of the current status of RSM

and some indication of possible developments was given by Myers (1999).

The exploration of an experimental region using response surface methods revolves around the assumption that the expected response,  $E(y)$ , is a function of controllable variables  $x_1, x_2, \dots, x_k$ ; where the  $x_i$ 's are suitably scaled and centred linear transformations of the independent variables.

Classical RSM can be divided into three stages:

### First Order Design

Often, the initial estimate of the optimum operating conditions for the system will be far from the true optimum. In such circumstances, the objective of the experimenter is to move rapidly to the general vicinity of the optimum. When far from the optimum, it is usual to assume, using Taylor series arguments, that locally a first-order model

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (1.1)$$

is an adequate approximation to the true surface over the experimental region. It is also usually assumed that the experimental errors  $\varepsilon$  are independent and identically distributed, although this is sometimes not a good approximation.

### Experimentation Along The Path of Steepest Ascent.

The method of steepest ascent given by Box and Wilson (1951) is a procedure for moving sequentially in the direction of the maximum increase in the response. This direction, called the path of steepest ascent, is given by

$$x_2 = \frac{b_2}{b_1} x_1, \dots, x_k = \frac{b_k}{b_1} x_1, \quad (1.2)$$

where  $b_1, \dots, b_k$  are the estimates of  $\beta_1, \dots, \beta_k$  respectively. Experiments are conducted on the path of steepest ascent until no further increase in response is observed. After that, a new first order model is used, with possibly a wider range since as the response surface is “climbed” the first order effects usually become smaller relative to the random error, and then another path of steepest ascent is computed. This process is continued until lack of fit of the first order model

indicates that the experimenter is close to the optimum and a second order model is fitted to account for the curvature and interaction in the response surface. In order that the coefficients of the second order model can be estimated the design needs to be expanded by adding extra points.

Brooks and Mickey (1961) examined alternative strategies for steepest ascent experiments, while Myers and Khuri (1979) investigated formalising a stopping rule to determine when a new set of experiments should be performed.

The direction of steepest ascent given in equation (1.2) is not invariant to scale changes of the independent variables. Box and Draper (1987, pp. 199–202) argue that this dependence is reasonable, since it combines the experimenters prior knowledge of the response surface with the experimental data. To overcome the scale dependence of the path of steepest ascent, Kleijnen, den Hertog and Angün (2004) developed an adapted path of steepest ascent (ASA). The ASA path starts from the design point with minimal prediction variance, and chooses as the next point for experimentation the point that maximises the lower bound of a  $100(1 - \alpha)\%$  confidence interval for the predicted response. The ASA path takes into account the variance-covariance matrix of the regression coefficients to modify the direction of the path of steepest ascent, and also provides an appropriate step size. When the design is orthogonal, the ASA path is equivalent to the path of steepest ascent given by equation (1.2).

## Second Order Experimentation

After the path of steepest ascent phase is completed, a resolution V design augmented by axial points and a centre point, forming a central composite design, is used, and then a second order model

$$y = \beta_0 + \sum_{i=1}^k \beta_{ii}x_i^2 + \sum_{1 \leq i < j \leq k} \beta_{ij}x_i x_j + \varepsilon \quad (1.3)$$

is fitted to the data (Box, Hunter and Hunter, 1978, p. 319). Central composite designs require five levels for each of the variables. Alternatively, more economical designs that require only three levels can be used (Box and Behnken, 1960). If



the fitted second order model is written in matrix notation, then

$$\hat{y} = b_0 + \mathbf{x}^T \mathbf{b} + \mathbf{x}^T \mathbf{B} \mathbf{x}, \quad (1.4)$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} & \cdots & \frac{1}{2}b_{1k} \\ \frac{1}{2}b_{12} & b_{22} & \cdots & \frac{1}{2}b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}b_{1k} & \frac{1}{2}b_{2k} & \cdots & b_{kk} \end{bmatrix}.$$

In equation (1.4),  $\mathbf{b}$  is a  $(k \times 1)$  vector of the first-order regression coefficients and  $\mathbf{B}$  is a  $(k \times k)$  symmetric matrix whose diagonal elements are the pure quadratic coefficients, and whose off-diagonal elements are one-half the mixed quadratic coefficients. Differentiating  $\hat{y}$  in equation (1.4) with respect to the vector  $\mathbf{x}$  and equating it to zero results in

$$\frac{\partial \hat{y}}{\partial \mathbf{x}} = \mathbf{b} + 2\mathbf{B}\mathbf{x} = 0.$$

Therefore, the stationary point is

$$\mathbf{x}_S = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b}, \quad (1.5)$$

and the predicted response at the stationary point is

$$\hat{y}_S = \hat{\beta}_0 + \frac{1}{2}\mathbf{x}_S^T \mathbf{b}. \quad (1.6)$$

To interpret the fitted surface the method of canonical analysis was developed by Box and Wilson (1951), allowing the determination of whether the region near the stationary point of the surface is a maximum, minimum, ridge, rising ridge, or saddle point. In canonical analysis the model is transformed into a new coordinate system. The axes of the system are rotated until they are parallel to the axes of the fitted response surface. This rotation removes all cross-product terms and is called the *A* canonical form.

If the  $i$ th normalised eigenvector of  $\mathbf{B}$  is denoted by  $\mathbf{d}_i$ , with corresponding eigenvalues  $\lambda_i$ , then new variables  $\mathbf{z}$  can be defined as

$$\mathbf{z} = \mathbf{D}^T \mathbf{x},$$

where

$$\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_k].$$

In this rotated co-ordinate system

$$\hat{y} = b_0 + \mathbf{z}^T \hat{\boldsymbol{\phi}} + \mathbf{z}^T \hat{\mathbf{A}} \mathbf{z}, \quad (1.7)$$

and

$$\begin{aligned} \hat{\mathbf{A}} &= \mathbf{D}^T \mathbf{B} \mathbf{D} \\ &= \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k). \end{aligned}$$

If needed, first order terms are removed as well by changing the origin, resulting in the  $B$  canonical form given by

$$\hat{y} = \hat{y}_S + \lambda_1 \tilde{z}_1^2 + \lambda_2 \tilde{z}_2^2 + \dots + \lambda_k \tilde{z}_k^2, \quad (1.8)$$

where  $\tilde{z}_i$  are formed using

$$\tilde{z}_i = z_i - \bar{z}_i, \quad i = 1, \dots, k,$$

and  $\bar{z}_i$  is the average value of  $z_i$  over the experimental runs. The size and sign of the eigenvalues determine the nature of the fitted response surface. If the eigenvalues are all positive the stationary point is a minimum. On the other hand, if the eigenvalues are all negative the stationary point is a maximum. If the eigenvalues have different signs the stationary point is a saddle point, while if any of the eigenvalues are zero the response surface is a ridge.

Of particular importance are stationary ridges, where the optimum conditions are given by a line, plane, or hyperplane, corresponding to equation (1.8) with one or more of the eigenvalues equal to zero; and rising ridges, where maximum improvement in the response can only be achieved by changing two or more variables, while only sub-optimal improvement can be made by changing just one variable. A rising ridge is given by equation (1.7) with at least one of the eigenvalues zero, but the corresponding first order coefficient not equal to zero.

With real data, of course, an eigenvalue will not be exactly zero. However, if one or more of the eigenvalues are small relative to the others and of about the

same size as their standard error, then there is an indication that there may be approximately a linear or planar maximum, rather than a point maximum. As discussed by Box and Draper (1987, pp. 329–330), this is of considerable importance since it means that there may be a whole range of alternative optimum conditions which may allow a primary response to be optimised with satisfactory levels for secondary responses.

## 1.2 Precision of the Path of Steepest Ascent

As indicated above, the calculation of the path of steepest ascent is an important part of response surface methodology. In practice, experimenters need to know whether the direction of steepest ascent has been determined precisely enough to proceed. Box (1955) and Box and Draper (1987, pp. 190–194) discussed a methodology for computing a confidence region or confidence cone (or hypercone) for the direction of steepest ascent. If the path of steepest ascent has been determined precisely enough, the proportion of possible directions included in the confidence cone will be small. One limitation of the method given by Box (1955) and Box and Draper (1987, pp. 190–194) is that it only applies where the response is a continuous, uncorrelated, normally distributed variable, with homogeneous variance.

After an initial first-order response surface has been fitted, the method will allow an assessment of whether the path of steepest ascent has been determined precisely enough. If it has, further experiments along the path are conducted, and the location of subsequent experiments are found.

Brooks and Mickey (1961) examined how many experiments were needed to optimally determine the path of steepest ascent. In the case of uncorrelated and homogeneous errors, they defined the error to be  $\theta$ , the angle between the true path and the estimated path. They also examine  $\frac{S \cos \theta}{t}$ , which they termed the improvement per unit of experimental effort, where  $t$  is the number of trials and  $S$  is the size of the step to be taken. They show that the maximum value is given where  $t$  is one more than the number of factors.

On the other hand, if the path is not determined precisely enough additional experiments need to be conducted. There has not been any specific work on what experiments need to be added in this case, although the literature on the design of experiments for nonlinear models, discussed in section 1.4, is relevant since the path of steepest ascent is defined in terms of ratios of the regression coefficients, that is, a nonlinear function of the parameters.

One restriction of the current method is that the response must be uncorrelated and homoscedastic. An extension to cover situations when this is not the case would be very useful, and, as shown below, can be used in a number of situations.

### 1.3 Ridge Analysis

An alternative to canonical analysis was developed by A.E. Hoerl (1959) and further advocated by R.W. Hoerl (1985), giving the path of steepest ascent for second order models, resulting in a curve and not a straight line as in first order models. Ridge analysis is a technique for examining higher-dimensional quadratic response surfaces. Ridge analysis is employed in RSM to graphically illustrate the behaviour of response surfaces, and to locate overall and local optimum regions on the hyperspheres  $\mathbf{x}^T \mathbf{x} = R^2$ .

A second-order response surface in  $k$  independent variables can be represented as equation (1.4). If a sphere of radius  $R$  is drawn in the  $\mathbf{x}$ -coordinate space with the point  $(0, 0, \dots, 0)$  as the origin, then somewhere on the sphere there will be a maximum  $\hat{y}(R)$ . The corresponding vector of  $\mathbf{x}$ -coordinates, considered as a function of  $R$ , is defined as the maximum ridge, and gives the path of steepest ascent from the origin. Similarly, the minimum ridge gives the path of steepest descent. A plot of  $\hat{y}(R)$  for  $R$  values from 0 to  $C$ , where  $C$  is the maximum radius to be considered, allows the maximum attainable response to be determined.

Using Lagrange multipliers, equation (1.4) is differentiated with respect to the vector  $\mathbf{x}$ . To obtain the constrained stationary points  $\partial \hat{y} / \partial \mathbf{x}$  is set equal to

0. This results in

$$\mathbf{x}_S = -\frac{1}{2}(\mathbf{B} - \mu\mathbf{I})^{-1}\mathbf{b}. \quad (1.9)$$

Therefore, for a fixed  $\mu$ ,  $\mathbf{x}_S$  is a stationary point on  $R^2 = \mathbf{x}^T\mathbf{x}$ . Whether a particular stationary point is a maximum or a minimum is determined by  $\mu$ . Draper (1963) shows that if maximisation is required, only values of  $\mu$  greater than the largest eigenvalue ( $\lambda_k$ ) of the  $\mathbf{B}$  matrix should be substituted into equation (1.9). On the other hand, if  $\mu$  is smaller than the smallest eigenvalue ( $\lambda_1$ ) of  $\mathbf{B}$ , the solution of equation (1.9) will result in a minimum.

Box and Hunter (1954) provided a method for estimating a confidence region about the stationary point of a second order response surface. The set of first derivatives of  $\hat{y}$  with respect to  $\mathbf{x}$  is given by

$$\begin{aligned} b_1 + 2b_{11}x_1 + b_{12}x_2 + \dots + b_{1k}x_k &= \xi_1, \\ b_2 + b_{12}x_1 + 2b_{22}x_2 + \dots + b_{2k}x_k &= \xi_2, \\ &\vdots \\ b_k + b_{1k}x_1 + b_{2k}x_2 + \dots + 2b_{kk}x_k &= \xi_k, \end{aligned}$$

where at the stationary point  $\xi_1, \dots, \xi_k$  are equal to zero. For a fixed set of the  $\mathbf{x}$ 's,

$$\frac{\boldsymbol{\xi}^T \mathbf{V}^{-1} \boldsymbol{\xi}}{ks^2} \sim F(k, \nu),$$

where  $\mathbf{V}\sigma^2$  is the variance-covariance matrix of the  $\xi$ 's and  $s^2$  is an estimate of the experimental variance  $\sigma^2$  based on  $\nu$  degrees of freedom. Hence, the probability of the inequality

$$\frac{\boldsymbol{\xi}^T \mathbf{V}^{-1} \boldsymbol{\xi}}{ks^2} \leq F_{1-\alpha}(k, \nu)$$

is  $1 - \alpha$ , so the boundary of a  $100(1 - \alpha)\%$  confidence region is given by the values of  $\mathbf{x}$  such that

$$\boldsymbol{\xi}^T \mathbf{V}^{-1} \boldsymbol{\xi} = s^2 k F_{1-\alpha}(k, \nu). \quad (1.10)$$

Box and Hunter (1957) showed that when a rotatable second order design is used equation (1.10) reduces to a much more convenient non-matrix expression. A software tool for the computation and display of these confidence regions was given by del Castillo and Cahaya (2001).

Peterson, Cahaya and del Castillo (2002) show that a drawback of the confidence region method of Box and Hunter (1954) is that the confidence region for a maximum may consist of two disjoint regions: one region corresponding to parameter values associated with a maximum response and one region corresponding to parameter values associated with a saddle point. They give an approach, discussed later, that overcomes this problem.

Carter, Chinchilli, Myers and Campbell (1986) provide a method for computing conservative confidence limits of the eigenvalues of the matrix of quadratic and cross-product term,  $\mathbf{B}$ , and the mean response at a constrained optima. In addition, they show how to construct conservative confidence intervals for a ridge analysis of a response surface. They use a result discussed by Rao (1973, p. 240) and Spjøtvoll (1972) that

$$\Pr[\inf_{\boldsymbol{\beta} \in C} g(\boldsymbol{\beta}) \leq g(\boldsymbol{\beta}) \leq \sup_{\boldsymbol{\beta} \in C} g(\boldsymbol{\beta})] \leq 1 - \alpha,$$

where  $\boldsymbol{\beta}$  is a parameter vector and  $C$  is the  $100(1 - \alpha)\%$  confidence interval for  $\boldsymbol{\beta}$ . The eigenvalues of  $\mathbf{B}$  are a function of  $\boldsymbol{\beta}$ , and by setting  $g_i(\boldsymbol{\beta})$  equal to the  $i$ th eigenvalue

$$\min_{\boldsymbol{\beta} \in C} g_i(\boldsymbol{\beta}), \max_{\boldsymbol{\beta} \in C} g_i(\boldsymbol{\beta}) \quad i = 1, \dots, k$$

a  $100(1 - \alpha)\%$  confidence interval for the  $i$ th eigenvalue is obtained, while a similar bound, with  $g_i(\boldsymbol{\beta})$  being the dot product of  $(\mathbf{B} - \mu\mathbf{I})^{-1}$  and  $\frac{1}{2}\mathbf{b}$ , gives a confidence interval for the  $i$ th co-ordinate of the ridge analysis solution. To identify points in  $C$  where

$$C = \left\{ \boldsymbol{\beta} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) (\mathbf{X}^T \mathbf{X})^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) / ps^2 \leq F_{1-\alpha}(p, n - p) \right\},$$

a polar transformation is made where  $C$  is transformed to  $C^*$ , a  $p$ -dimensional hypersphere of radius  $r = (ps^2 F_{1-\alpha}(p, n - p))^{\frac{1}{2}}$ . A large number of points in  $C^*$  are generated, the corresponding points in  $C$  computed, and the minimum and maximum values of the relevant  $g_i(\boldsymbol{\beta})$  give the confidence interval.

While the method developed by Carter et al. (1986) represents an important extension to ridge analysis, their results, however, presented a number of problems. Apart from being conservative, the method requires a considerable

amount of computation. Carter, Chinchilli and Campbell (1990) developed a more computationally convenient method based on large sample theory to compute an approximate  $100(1 - \alpha)\%$  confidence region for the eigenvalues of  $\mathbf{B}$ . The method used was called the *delta method*, and corresponds to using a Taylor's series expansion to approximate a function of a random variable. In general, if a vector of random variables  $\mathbf{w}$  has mean  $\mu$  with covariance matrix  $\Sigma$ , then a vector-valued function  $g(\mathbf{w})$  is asymptotically distributed with mean  $g(\mu)$  and the covariance is given by  $\mathbf{U}^T \Sigma \mathbf{U}$ , where

$$\mathbf{U} = \frac{\partial g}{\partial \mathbf{w}}.$$

Specifically, the eigenvalues of  $\mathbf{B}$  are functions of the second-order coefficients, and since the means and variances of the second-order coefficients can be estimated, it is possible to find an approximate  $100(1 - \alpha)\%$  confidence region for the eigenvalues  $\lambda$  given by

$$\hat{C}_\lambda = \left\{ \lambda \in R^k : (\hat{\lambda} - \lambda)^T (\hat{\mathbf{H}}^T \mathbf{V} \hat{\mathbf{H}})^{-1} (\hat{\lambda} - \lambda) \leq \chi_{1-\alpha}^2(k) \right\},$$

where  $\hat{\lambda}$  are the estimated eigenvalues,  $\mathbf{H}$  is a function of the eigenvectors, and  $\mathbf{V}$  is a function of the variance-covariance matrix of the parameters. An equivalent confidence interval can be obtained for an individual eigenvalue. If the errors are homogeneous, the interval above should be replaced by

$$\hat{C}_\lambda = \left\{ \lambda \in R^k : (\hat{\lambda} - \lambda)^T (\hat{\mathbf{H}}^T \mathbf{V} \hat{\mathbf{H}})^{-1} (\hat{\lambda} - \lambda) \leq ks^2 F_{1-\alpha}(k, n - p) \right\},$$

which gives better coverage in small samples.

Bisgaard and Ankenman (1996) gave a simpler method to that developed by Carter et al. (1990) since the latter method requires a significant amount of vector and matrix manipulation. The simpler method, which is called the *double linear regression* method, only requires the fitting of two linear regressions where a full quadratic model is fitted. First, a standard quadratic regression model is fitted, and the canonical axes are determined. After that, the co-ordinate system is rotated so that it coincides with the canonical axes, and another quadratic regression model is fitted. Bisgaard and Ankenman (1996) showed that the quadratic

terms of the second regression are equivalent to the eigenvalues of the original  $\mathbf{B}$  matrix, and therefore a confidence region for the quadratic terms can be used to give a confidence region for the eigenvalues.

Ankenman (2003) extended the double linear regression method to obtain confidence intervals for the first order coefficients, overcoming a problem with the double linear regression method, which can seriously underestimate the standard error of the estimates. He gives a parameterisation for the  $A$  canonical form of the quadratic response surface, which provides good estimates and standard errors of the first order coefficient estimates. He also shows how the  $A$  canonical form can be directly fitted using non-linear least squares, an approach which was first illustrated by Box and Draper (1987, p. 359).

Box and Draper (1987) gave two examples where they directly fitted a canonical form. In the first example, (pp. 355–360) they discussed a three factor chemical experiment which gave rise to an approximate stationary planar ridge. They fitted the model

$$y = \hat{y}_S - \lambda(\alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \alpha_{10})^2 + \varepsilon,$$

where  $\hat{y}_S$  is the estimated response on the fitted maximal plane; while the values of  $\alpha_{11}$ ,  $\alpha_{12}$  and  $\alpha_{13}$  are the direction cosines of a line perpendicular to the maximal plane;  $\alpha_{10}$  measures the distance in design units of the nearest point on the fitted plane to the design origin; and  $\lambda$  measures the quadratic fall-off in a direction perpendicular to the fitted maximal plane. In the second example (pp. 360–368), they discussed a small reactor study, again involving three factors, which this time gives rise to a rising ridge surface. They fitted the model

$$y = \hat{y}_{S'} - \lambda_1(\alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \alpha_{10})^2 + \lambda_2(\alpha_{21}x_1 + \alpha_{22}x_2 + \alpha_{23}x_3) + \varepsilon,$$

where  $\hat{y}_{S'}$  is the estimated response at  $S'$ , the point nearest to the origin on the rising planar surface;  $\alpha_{11}$ ,  $\alpha_{12}$  and  $\alpha_{13}$  are the direction cosines of lines perpendicular to the plane;  $\alpha_{10}$  measures the shortest distance of the planar ridge from the design origin;  $\lambda_1$  measures the rate of quadratic fall-off as we move away from the plane of the ridge;  $\alpha_{21}$ ,  $\alpha_{22}$  and  $\alpha_{23}$  are the direction cosines of the line of



steepest ascent up the planar ridge; and  $\lambda_2$  is the linear rate of increase of yield as we move up the ridge.

## 1.4 Nonlinear Experimental Design

Methods for designing experiments for non-linear situations have a long history after the early work of Box and Lucas (1959). In particular, methods for minimising the volume of the confidence region of the estimated parameters, called D-optimal designs, are often used. These methods, originally developed by Kiefer (1958, 1959, 1961*a,b*, 1962*a,b*) and Kiefer and Wolfowitz (1959, 1969), have been studied and extended by many authors (see, for example, Covey-Crump and Silvey, 1970; Wynn, 1970, 1972; Box and Draper, 1987; DuMouchel and Jones, 1994; Atkinson and Haines, 1996).

It is often desirable to augment the existing design with additional runs in order to achieve the desired precision of the parameter estimates. When this is the case, it is desired to choose a criterion for judging whether one set of added experimental observations is superior to another set of observations. The most common criterion is to maximise the determinant  $|\mathbf{X}^T \mathbf{X}|$ , where  $\mathbf{X}$  is the design matrix. Methods for augmenting the existing data in response surface experiments have been presented by Dykstra (1966, 1971) and Hebble and Mitchell (1972). Gaylor and Merrill (1968) were concerned with augmenting existing data in multiple regression. Box and Hunter (1965) used the  $|\mathbf{X}^T \mathbf{X}|$  criterion to determine the next design point for nonlinear models. Hill and Hunter (1974) described a method for situations where it is desired to estimate one subset of parameters more precisely than another. The procedure takes into account prior knowledge, and uses a Bayesian development to obtain a modified criterion for a subset of the parameters of interest. The criterion is then maximised with respect to the independent variables.

Jia and Myers (1998) and Myers (1999) discuss experimental design approaches for Binomial and Poisson data. They show that for the Binomial case with two

explanatory variables,  $x_1$  and  $x_2$ , and a logistic model

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)},$$

the D-optimal design is to place four points on a parallelogram with two points having predicted  $p$  values of 0.227 and two points having predicted  $p$  values of 0.773 (called  $ED_{22.7}$  and  $ED_{77.3}$ ). As with a non-linear design it is necessary to have guesses of the parameters in order to determine the best design.

An alternative, but potentially more robust procedure, is to use Bayesian Designs (Chaloner and Verdinelli, 1995), where a design is chosen to minimise the function

$$\int_{\boldsymbol{\beta}} R(\boldsymbol{\delta}, \boldsymbol{\beta}) \pi(\boldsymbol{\beta}) d\boldsymbol{\beta},$$

where  $\boldsymbol{\delta}$  denotes the design,  $\boldsymbol{\beta}$  the parameter vector with prior probability  $\pi(\boldsymbol{\beta})$  and  $R(\boldsymbol{\delta}, \boldsymbol{\beta})$  denotes a criterion which we wish to maximise. Bayes D-optimal design corresponds to

$$R(\boldsymbol{\delta}, \boldsymbol{\beta}) = \ln |\mathbf{I}|,$$

where  $\mathbf{I}$  is the information matrix generated by the logistic model. Bayes designs generally give more levels of the experimental variables, and are thus more robust to the guesses of the experimenter.

In any case, once experiments are conducted, all the parameters can be estimated, and new sets of experiments can be designed in the light of the information generated.

Similar considerations apply to models where the data follows a Poisson distribution. For example, assume it is expected that the mean count is

$$\lambda = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2).$$

As discussed by Myers (1999), the D-optimal design is given by (0,0), the control, and two points on the “ $EC_{13.5}$ ” contour, where the expected count is 13.5% of the expected count at the control.

## 1.5 Further Developments in RSM

RSM as presented by Box and Wilson (1951) was only a beginning that led to further development of these ideas by other scientists. Over the next decades, RSM and other design techniques spread throughout the chemical and process industries, as well as research and development work. The tools and techniques of RSM are also useful in engineering, quality technology, physics, biology, agriculture, psychology, and statistics.

In the last 20 years, great attention has been drawn to RSM. The number of problems that can be solved with RSM has increased. This may be due to developments in nonlinear optimisation, Bayesian experimental designs, nonparametric regression, estimation using GLMs, mixed model analysis, and many others (Myers et al., 1989).

The current status and future directions of response surface methodology was reviewed by Myers (1999). One of the most important points made both by Myers and the discussants to the paper was the increasing importance of generalised linear models, and the use of multiple responses in response surface methodology.

## 1.6 Method of Steepest Ascent for GLMs

Situations in which a response takes the form of a count, proportion, or is exponentially distributed arise frequently in practice. In those situations, the least squares assumptions, of normal errors and homogeneous variance, are not satisfied. Generalised linear models, introduced by Nelder and Wedderburn (1972), have become the basic tool for statistical analysis of such data. These models, which include the standard regression models as special cases, are estimated using iteratively reweighted least squares.

Applications of GLMs were expanded following the publication of a book on the topic by McCullagh and Nelder (1989) and the availability of easily used software (for example Aitkin et al., 1989). Further developments were due to the increasing emphasis on quality improvement. Nelder and Lee (1991) and Grego (1993) showed their use in “Taguchi” type experiments, while Hamada and Nelder

(1997) gave some examples of the use of GLMs in quality improvement, and Myers and Montgomery (1997) presented a tutorial on the topic.

This interest in GLMs prompted Myers (1999) and others to urge the greater use of GLMs in RSM. Some applications were given by Lewis, Montgomery and Myers (2001*b*). However, no work on a method for determining the precision of the path of steepest ascent for GLMs has as yet been presented.

## 1.7 Constrained Experimental Regions

There are some problems in which it is not possible to proceed very far in the direction of steepest ascent due to constrained experimental variables. Box and Draper (1987, pp.194–199) show how the direction of steepest ascent can be modified so that a linear constraint is not violated. However, an adaptation of the method to determine the precision of the path of steepest ascent to cover constrained regions has not been presented in the literature.

An important case of the constrained optimisation problem is for dual response systems, where there is a primary response which we desire to maximise (or minimise) and a secondary response or responses which we desire to take on certain values. Myers and Carter (1973) studied this problem. They considered the case where the fitted primary response  $\hat{y}_p$  is given by

$$\hat{y}_p = b_0 + \mathbf{x}^T \mathbf{b} + \mathbf{x}^T \mathbf{B} \mathbf{x},$$

while the constraint response is given by

$$\hat{y}_s = c_0 + \mathbf{x}^T \mathbf{c} + \mathbf{x}^T \mathbf{C} \mathbf{x}.$$

Myers and Carter (1973) found conditions where it is possible to optimise  $\hat{y}_p$  subject to  $\hat{y}_s = k$ , where  $k$  is some desirable or acceptable value of the constraint response. Using Lagrangian multipliers, they optimised

$$L = b_0 + \mathbf{b}^T \mathbf{x} + \mathbf{x}^T \mathbf{B} \mathbf{x} - \mu(\mathbf{c}_0 + \mathbf{c}^T \mathbf{x} + \mathbf{x}^T \mathbf{C} \mathbf{x} - k)$$

by setting the derivative of  $L$  with respect to  $\mathbf{x}$  equal to zero, and obtained

$$(\mathbf{B} - \mu \mathbf{C}) \mathbf{x} = \frac{1}{2}(\mu \mathbf{c} - \mathbf{b}).$$

The nature of the stationary point depends on the matrix of second partial derivatives

$$\mathbf{M}(\mathbf{x}) = 2(\mathbf{B} - \mu\mathbf{C}).$$

To give a maximum,  $\mathbf{M}(\mathbf{x})$  must be negative definite; while to give a minimum,  $\mathbf{M}(\mathbf{x})$  must be positive definite. The approach adopted by Myers and Carter (1973) is similar to that used in ridge analysis. They selected values of  $\mu$  which make  $\mathbf{M}(\mathbf{x})$  negative (positive) definite, assuming a maximum (minimum) is required. This can always be done if  $\mathbf{C}$  is definite, but only in some cases if  $\mathbf{C}$  is indefinite.

Once the constrained optimum is obtained a confidence region can be calculated using the method developed by Stablein, Carter and Wampler (1983), an extension of the method given by Box and Hunter (1954). Stablein et al. (1983) consider the case where it is desired to optimise a primary response, which can be approximated by a quadratic function, but there are  $m(\geq 1)$  secondary constraint functions  $g_i(\mathbf{x}) = c_i$ , which are also assumed to be quadratic. In particular, for ridge analysis there is one constraint function given by

$$g_i(\mathbf{x}) = \mathbf{x}^T \mathbf{x} = R^2.$$

In Peterson (1993) and Peterson et al. (2002) the standard quadratic model used in ridge analysis is generalised to a parametrically linear model for the response surface given by

$$y = \beta_0 + z(\mathbf{x})^T \boldsymbol{\theta} + \varepsilon,$$

where  $z(\mathbf{x})$  is a vector-valued function of  $\mathbf{x}$  and  $\boldsymbol{\theta}$  is a vector of regression coefficients. In addition, they extend the type of constraint that can be used. Peterson (1993) provides conservative or approximate confidence intervals about the optimal mean response as the radius is varied, called a “guidance band”, which assists the experimenter in determining the optimal settings for the experimental variables. One of the advantages of the approach given by Peterson (1993) and Peterson et al. (2002) is that the radius constraint is dealt with directly, and hence there is no need to introduce an unknown Lagrange multiplier.

Gilmour and Draper (2003) gave confidence regions for ridge analysis solutions. They modified the approach of Stablein et al. (1983), who had, in effect, fixed the Lagrange multiplier  $\mu$  and varied the radius of solutions, by fixing the radius and varying over  $\mu$ . For two dimensions, they gave a plot of the boundaries of the confidence intervals for varying radii, and suggested a suitable plot for three dimensions. They also tabulated the percentage of the surface area of a circle or sphere covered by the confidence intervals at different radii.

Peterson, Cahaya and del Castillo (2004) demonstrated that the method given by Peterson et al. (2002) is an alternative to that used by Gilmour and Draper (2003). They gave examples where the proportion of the circle or sphere that is covered by the confidence intervals or regions is less than when the approach of Gilmour and Draper (2003) is used. However, as Gilmour and Draper (2004) point out, this may be because the  $100(1 - \alpha)\%$  confidence regions for the maximum of a response surface used in Gilmour and Draper (2003) are nonuniform, that is, they include the true maximum with a probability of at least  $1 - \alpha$ , over all possible values of the parameters, while all other methods in the literature, including the methods given by Peterson (1993) and Peterson et al. (2002), do not have this guaranteed coverage.

A more direct approach to solving dual response systems is to set-up the optimisation as a nonlinear programming problem following del Castillo and Montgomery (1993), who used the generalised reduced gradient algorithm to find the value of the Lagrange multiplier,  $\mu$ , and hence the optimal levels of the experimental variables. Similarly, del Castillo (1996) shows how to optimise a primary response subject to a number of secondary response variable constraints using a combination of nonlinear programming, confidence regions for the stationary points of quadratic responses (Box and Hunter, 1954), and confidence cones for the directions of maximum improvement for linear responses (Box, 1955; Box and Draper, 1987, pp. 190–194).

## 1.8 Thesis Outline

In Chapter 2, we give a derivation of the current method, and generalise the current method to data with non-orthogonal designs and heterogeneous errors, where the regression coefficients are heteroscedastic. We also show how to give the percentage of directions included in the confidence cone for the path of steepest ascent as an integral, which can be exactly evaluated when  $k = 2$ , but requires numerical integration for higher values of  $k$ . For  $k = 3$  and  $k = 4$ , relatively tight inequalities can be derived, but in general numerical integration can be used.

In Chapter 3, we apply the new method to GLMs. These are important in many practical situations. First, we use Wald methods for both when there is no overdispersion and when there is overdispersion. Second, we use profile likelihood methods which give more exact results. We also give an easy method for numerically evaluating the percentage included. In addition, we propose Bootstrap methods to give results that do not rely on asymptotic approximations.

In Chapter 4, we look at power. We first give the power for homogeneous situations, and show that the power depends only on the size of the coefficient vector, the variance, and the sample size. For heterogeneous situations, the power will be different depending on the direction of the path, and hence the least precise direction can be used prior to the experiment. For GLMs, best sets of experiments are achieved in order to determine the minimum percentage included, for both the cases when the data follows the Binomial or Poisson distributions. In addition, we look at augmenting experiments when the path has not been determined precisely enough.

In Chapter 5, we look at cases where there is a constraint. The important case of ridge regression is also examined. A method for determining the precision of a canonical axis is developed, and applied to a number of examples.

# Chapter 2

## Non-Orthogonal Designs with Heterogeneous Errors

### 2.1 Introduction

In the initial stage of a response surface study, a first order design is used to fit a linear model

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon, \quad (2.1)$$

where  $y$  is the response and  $x_1, \dots, x_k$  are the coded levels of the experimental factors. If the linear model fits well, the path of steepest ascent giving the maximum predicted increase in the response can be calculated. Experimentation is conducted along the path of steepest ascent until curvature is detected.

If curvature is present, a second order experiment is conducted and a second order model

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \beta_{ij} x_i x_j + \varepsilon \quad (2.2)$$

is fitted and the results are summarised using contour diagrams and canonical analysis (see, for example, Box and Draper, 1987, chapters 10 and 11).

Box (1955) and Box and Draper (1987, pp. 190–194) presented a method for determining whether the path of steepest ascent has been determined precisely



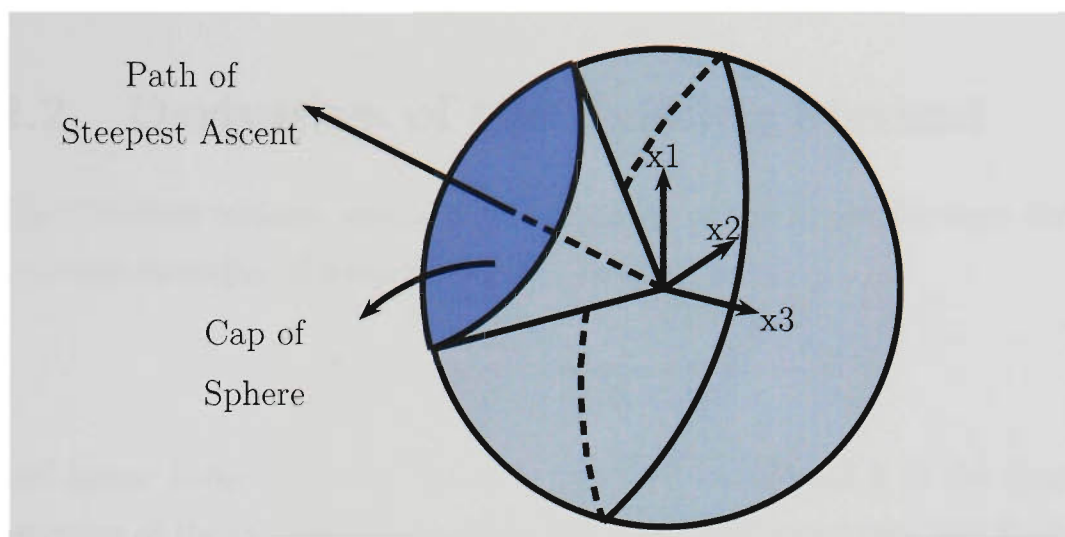


Figure 2.1: Confidence cone for the path of steepest ascent. The proportion of directions included in the confidence cone gives a measure of the precision of the path of steepest ascent, and is measured by taking the ratio of the surface area of the cap of the sphere within the confidence cone to the surface area of the sphere.

enough (see also Myers and Montgomery, 1995, pp. 194–198). The method gives the fraction of all possible directions that are included in the confidence ‘cone’ around the fitted path of steepest ascent, illustrated in Figure 2.1.

If this proportion is small, then we can say that the path has been determined precisely enough, and experiments along the path can be conducted. Otherwise, further experiments to improve the precision of the path, or to determine whether the surface is essentially flat, need to be run. This method applies to orthogonal designs with uncorrelated and homogeneous errors. However, in practice, there are many situations where the design will not be orthogonal, nor the errors homogeneous. One example would be in generalised linear models, which are being increasingly used.

The purpose of this chapter is to give a derivation of the current method, and to develop a method that would apply more generally. An example will be worked out in detail. In addition, some inequalities for the precision of the path will be derived and applied.

Some of the material covered in this chapter has appeared in Sztendur and

Diamond (1999, 2001, 2002).

## 2.2 Derivation of the Existing Method

The direction cosines, which give the cosines of the angles between the  $i$ th axis and true direction of steepest ascent, are given by

$$\delta_i = \frac{\beta_i}{\sqrt{\beta_1^2 + \cdots + \beta_k^2}} \quad i = 1, \dots, k.$$

and hence  $E(b_i) = \gamma\delta_i$ , where  $\gamma = \sqrt{\beta_1^2 + \cdots + \beta_k^2}$  and  $b_i$  is the least squares estimate of the  $i$ th regression coefficient. Box (1955) and Box and Draper (1987, pp. 190–194) show how a  $100(1 - \alpha)\%$  confidence region for the  $\delta_i$  can be determined by considering the estimated regression co-efficients  $b_i$  as the response, the  $\delta_i$ 's as levels of a single predictor variable, and  $\gamma$  as a regression co-efficient to be estimated. The confidence region is given by the  $\delta$ 's which just fail to make the residual mean square significant compared to the variances of the estimated regression coefficients, at some desired level  $\alpha$ . The region is given by

$$\frac{\left\{ \sum b_i^2 - \left( \sum_{i=1}^k b_i \delta_i \right)^2 / \sum \delta_i^2 \right\}}{(k-1)s_b^2} \leq F_{1-\alpha}(k-1, \nu_b), \quad (2.3)$$

where the estimated variance of  $b_i$  is  $s_b^2$ , based on  $\nu_b$  degrees of freedom.

A good indication of whether the direction has been determined precisely enough is provided by the magnitude of the solid angle of the 'confidence' cone about the estimated path of steepest ascent vector, which is, as shown below, a function of the  $t$ -distribution.

The first step in developing the current method is to find the surface area of a  $k$ -dimensional sphere of radius  $R$ ,  $SA(k, R)$ . The formula is given by

$$SA(k, R) = \frac{2\pi^{k/2} R^{k-1}}{\Gamma(k/2)}, \quad (2.4)$$

which can be proved using induction on  $k$ . For  $k = 2$ , the equation (2.4) is correct since  $SA(2, R) = 2\pi R$ , the circumference of a circle of radius  $R$ . Assume equation

(2.4) is true for  $k$ . then for  $k + 1$  we have

$$\begin{aligned} SA(k + 1, R) &= 2 \int_0^R SA(k, \sqrt{R^2 - r^2}) \frac{R}{\sqrt{R^2 - r^2}} dr \\ &= \frac{4\pi^{k/2} R}{\Gamma(\frac{k}{2})} \int_0^R (R^2 - r^2)^{\frac{k-2}{2}} dr. \end{aligned} \quad (2.5)$$

Since,

$$\int_0^a x^m (a^n - x^n)^p dx = \frac{a^{m+1+np} \Gamma(\frac{m+1}{n}) \Gamma(p+1)}{n \Gamma(\frac{m+1}{n} + p + 1)}$$

(Spiegel (1997), p.95), and if we let  $p = \frac{k-2}{2}$ ,  $m = 0$ ,  $a = R$ , and  $n = 2$ . equation (2.5) becomes

$$\begin{aligned} SA(k + 1, R) &= \frac{4\pi^{\frac{k}{2}} R}{\Gamma(\frac{k}{2})} R^{1+k-2} \frac{\Gamma(\frac{k}{2}) \Gamma(\frac{k}{2})}{2\Gamma(\frac{1}{2} + \frac{k-2}{2} + 1)} \\ &= \frac{2\pi^{\frac{k+1}{2}} R^k \Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}) \Gamma(\frac{k+1}{2})} \\ &= \frac{2\pi^{\frac{k+1}{2}} R^k}{\Gamma(\frac{k+1}{2})}, \end{aligned}$$

and hence, by induction, equation (2.4) is true for all positive integers  $k$ .

Without loss of generality, take  $R$  to be 1. The surface area of the  $k$ -dimensional cap as a proportion of the surface area of the  $k$ -dimensional sphere,  $\gamma$ , can be calculated from the surface area of the  $(k - 1)$ -dimensional sphere and is given, for some  $r_0$ , by

$$\begin{aligned} \gamma &= \frac{\int_{r_0}^1 SA(k - 1, \sqrt{1 - r^2}) \frac{1}{\sqrt{1 - r^2}} dr}{SA(k, 1)} \\ &= c \int_{r_0}^1 (1 - r^2)^{\frac{k-3}{2}} dr, \end{aligned} \quad (2.6)$$

where

$$c = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k-1}{2})}.$$

Making the substitution

$$t = f(r) = (k - 1)^{\frac{1}{2}} \frac{r}{\sqrt{1 - r^2}},$$

we obtain

$$\frac{1}{\sqrt{\pi(k-1)}} \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k-1}{2})} \int_{f(r_0)}^{\infty} \left(1 + \frac{t^2}{k-1}\right)^{-\frac{k}{2}} dt, \quad (2.7)$$

which can be seen to be the area of the tail of the  $t$ -distribution with  $(k-1)$  degrees of freedom. Specifically, if we denote the distribution function of the  $t$ -distribution with  $\nu$  degrees of freedom by  $T(t, \nu)$ , then  $\gamma$ , the surface area of the cap as a proportion of the total surface area of the unit hypersphere, can be expressed as a function of the radius of the projection onto the  $(k-1)$ -dimensional plane through the origin and orthogonal to the path of steepest ascent, and is given by

$$\gamma = 1 - T\left(\sqrt{k-1} \frac{\sqrt{1-r^2}}{r}, k-1\right), \quad (2.8)$$

which is easily evaluated using standard statistical software such as Splus or Minitab.

## 2.3 An Example

A very interesting example of a fractional factorial experiment was conducted by Hsieh and Goodwin (1986) of the Chrysler corporation. The experiment was concerned with determining how to reduce the number of defects in the finish of sheet-molded grille opening panels. The factors in the experiment were Mold Cycle ( $A$ ), Viscosity ( $B$ ), Mold Temperature ( $C$ ), Mold Pressure ( $D$ ), Weight ( $E$ ), Priming ( $F$ ), Thickening Process ( $G$ ), Glass Type ( $H$ ), and Cutting Pattern ( $J$ ). The response was the observed number of defects,  $\hat{c}$ . Part of the data is given in the left hand panel of Table 2.1. The nonsignificant factors are not included in the table. Bisgaard and Fuller (1994–95) re-analysed the experiment using the transformed response

$$y = \frac{\sqrt{\hat{c}} + \sqrt{\hat{c} + 1}}{2}$$

with ordinary least squares, and showed that the important factors were  $D$  and  $F$ . In addition, there was a significant interaction string given by  $BG + CJ + EH$ .

Table 2.1: Part of the data for the car grille opening panels experiment. Left Panel: The original design matrix for the significant factors, the response and the transformed response. Right Panel: The modified design matrix, allowing the fitting of the “conditional” effects of  $B$  at both levels of  $G$ .

Run	$B$	$D$	$F$	$G$	$\hat{c}$	$\frac{(\sqrt{\hat{c}}+\sqrt{\hat{c}+1})}{2}$	$B_{G+}$ $(\frac{B+BG}{2})$	$B_{G-}$ $(\frac{B-BG}{2})$	$D$	$F$	$G$
1	−	−	−	+	56	7.52	−	0	−	−	+
2	−	−	−	−	17	4.18	0	−	−	−	−
3	+	−	+	+	2	1.57	+	0	−	+	+
4	+	−	+	−	4	2.12	0	+	−	+	−
5	−	−	+	−	3	1.87	0	−	−	+	−
6	−	−	+	+	4	2.12	−	0	−	+	+
7	+	−	−	−	50	7.12	0	+	−	−	−
8	+	−	−	+	2	1.57	+	0	−	−	+
9	−	+	+	+	1	1.21	−	0	+	+	+
10	−	+	+	−	0	0.50	0	−	+	+	−
11	+	+	−	+	3	1.87	+	0	+	−	+
12	+	+	−	−	12	3.54	0	+	+	−	−
13	−	+	−	−	3	1.87	0	−	+	−	−
14	−	+	−	+	4	2.12	−	0	+	−	+
15	+	+	+	−	0	0.50	0	+	+	+	−
16	+	+	+	+	0	0.50	+	0	+	+	+

On the basis of the data alone it is impossible to distinguish between the three possible interactions  $BG$ ,  $CJ$ , or  $EH$ . Assume that it was decided that  $BG$  was the important interaction. (A similar development would also apply to  $CJ$  and  $EH$ ). Using the data given in Appendix A.1 and Splus commands given in Appendix A.2, a model was fitted using the explanatory variables  $B$ ,  $D$ ,  $F$ ,  $G$ , and  $BG$ . Details of the fitted model are given in Table 2.2. The estimated transformed response is given by

$$\hat{y} = 2.5112 - 0.9975x_D - 1.2125x_F - 0.1625x_B - 0.2013x_G - 0.7700x_Bx_G. \tag{2.9}$$

Note that  $BG$  involves a quantitative variable ( $B$ ) and a qualitative variable ( $G$ ). If it is desired to reduce the number of defects even further by using the path of steepest descent, then if  $B$  is fixed, the path of steepest descent is given

Coefficients	Value	Std. Error	t value	Pr(>  t )
(Intercept)	2.5112	0.3304	7.6009	0.0000
<i>D</i>	-0.9975	0.3304	-3.0192	0.0129
<i>F</i>	-1.2125	0.3304	-3.6699	0.0043
<i>B</i>	-0.1625	0.3304	-0.4918	0.6334
<i>G</i>	-0.2013	0.3304	-0.6091	0.5560
<i>BG</i>	-0.7700	0.3304	-2.3306	0.0420

Residual standard error: 1.322 on 10 degrees of freedom.  
Multiple R-squared: 0.7411  
F-statistic: 5.726 on 5 and 10 degrees of freedom, the p-value is 0.009481

Table 2.2: Summary of regression results using the original design.

by

$$x_F = 1.2155x_D,$$

irrespective of which levels of *B* and *G* are chosen.

Using equation (2.3) with the values of  $b_1(D) = -0.9975$ ,  $b_2(F) = -1.2125$ ,  $s_b = 0.3304$  and  $k = 2$ , the 95% confidence region is given by

$$\frac{\{2.465162 - (-0.9975\delta_1 - 1.2125\delta_2)^2/(\delta_1^2 + \delta_2^2)\}}{0.3304^2} \leq 4.96403, \tag{2.10}$$

which can be written as a quadratic function of  $\delta_2$  as

$$0.453049\delta_2^2 - (2.418937\delta_1)\delta_2 + 0.9281988\delta_1^2 \leq 0,$$

with solution

$$\delta_2 \leq (2.6696 \pm 2.2535)\delta_1.$$

Hence, the 95% confidence cone for the path of steepest descent extends from  $\delta_2 = 0.4161\delta_1$  to  $\delta_2 = 4.9231\delta_1$ , as shown in Figure 2.2. In other words, the angle,  $\theta_1$ , between the true path of steepest ascent and the  $x_1$  axis extends from

$$\theta_1 = \tan^{-1}(0.4161) = 22.6^\circ$$

to

$$\theta_1 = \tan^{-1}(4.9231) = 78.5^\circ.$$

Note that the confidence cone is in the first quadrant as both regression coefficients are negative and we want to decrease the response.

The proportion of directions included in the confidence cone is the arc length of the cone divided by the circumference of the circle, and is given by

$$\frac{\tan^{-1}(4.9231) - \tan^{-1}(0.4161)}{2\pi} = 0.1554,$$

and hence the 95% confidence region excludes about 84.5% of possible directions.

Alternatively, the projection of the cone onto the line orthogonal to the path of steepest ascent has radius 0.469, and hence the percentage included in the confidence cone, using equation (2.9), is given by

$$1 - T\left(\frac{\sqrt{1 - .469^2}}{.469}, 1\right) = 0.155,$$

that is, the same result as before.

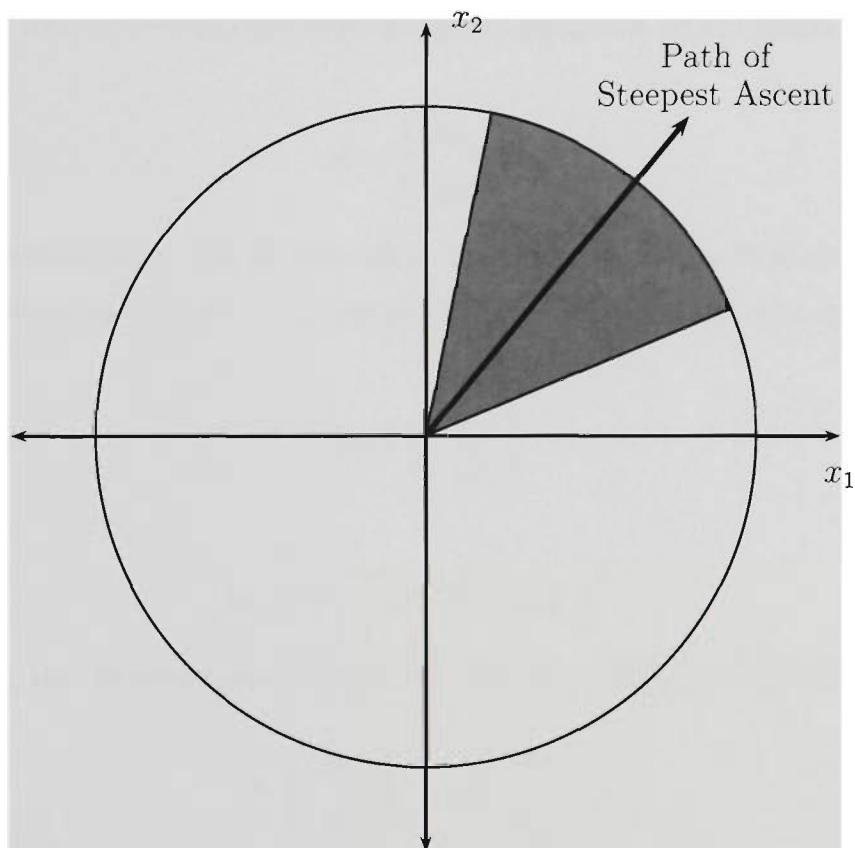


Figure 2.2: 95% confidence cone for the path of steepest ascent when  $B$  is fixed, involving  $D$  and  $F$  only, for the example given in Table 2.1.

## 2.4 Extension to Non-Orthogonal Designs with Heterogeneous Errors

Assume that the data  $\mathbf{y}$  has variance-covariance matrix  $\mathbf{V}\sigma^2$  with design matrix, expanded to model form, given by  $\mathbf{X}$ , where  $\mathbf{X}^T\mathbf{X}$  is not necessarily proportional to  $\mathbf{I}$ . Then, the generalised least square estimate of the regression coefficients is

$$\mathbf{b} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y},$$

with variance-covariance matrix

$$V(\mathbf{b}) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\sigma^2. \quad (2.11)$$

(See, for example, Rao, 1973, p. 230). In practice, the covariance matrix is unknown, so it must be estimated and therefore equation (2.11) holds only asymptotically. Set

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix},$$

where  $\mathbf{b}_1$  corresponds to the  $k_1$  variables considered in the path of steepest ascent and  $\mathbf{b}_2$  corresponds to the other variables, including at least the intercept. For the example,

$$\mathbf{b}_1 = \begin{pmatrix} \hat{D} \\ \hat{F} \end{pmatrix}$$

and

$$\mathbf{b}_2 = \begin{pmatrix} \hat{\mu} & \hat{B} & \hat{G} & \hat{BG} \end{pmatrix}^T.$$

Following the development in Box and Draper (1987, pp. 190–194), we have

$$E \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \gamma \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix},$$

but here with

$$V(\mathbf{b}) = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\sigma^2 = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{12}^T & \mathbf{G}_{22} \end{pmatrix} \sigma^2.$$



Then,

$$\frac{\mathbf{b}_1^T \mathbf{G} \mathbf{b}_1 - \boldsymbol{\delta}_1^T \mathbf{G} \mathbf{b}_1 (\boldsymbol{\delta}_1^T \mathbf{G} \boldsymbol{\delta}_1)^{-1} \mathbf{b}_1^T \mathbf{G} \boldsymbol{\delta}_1}{(k_1 - 1)s^2} \leq F_{1-\alpha}(k_1 - 1, \nu_b) \quad (2.12)$$

gives a  $100(1 - \alpha)\%$  confidence region for the elements of  $\boldsymbol{\delta}_1$  with  $\mathbf{G} = \mathbf{G}_{11}^{-1}$ . and  $s^2$  is the estimate of  $\sigma^2$ , based on  $\nu_b$  degrees of freedom. For the case where  $\mathbf{V} = \sigma^2 \mathbf{I}$ , equation (2.12) reduces to equation (2.3).

It can be shown that boundary of the the cap can be projected onto a  $(k_1 - 1)$ -dimensional hyperplane passing through the origin, providing a hyper-ellipsoid with lengths of the semi-axes given by

$$a_1 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_2}}, \dots, a_{k_1-1} = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_{k_1}}}, \quad (2.13)$$

where  $\lambda_1, \dots, \lambda_{k_1}$  are the eigenvalues of

$$\mathbf{H} = \mathbf{G} \mathbf{b}_1 \mathbf{b}_1^T \mathbf{G} - c \mathbf{G}, \quad (2.14)$$

with

$$c = \mathbf{b}_1^T \mathbf{G} \mathbf{b}_1 - (k_1 - 1)s^2 F_{1-\alpha}(k_1 - 1, \nu_b). \quad (2.15)$$

To see this, the boundary of the cap is given by the elements of  $\boldsymbol{\delta}_1$  satisfying equation (2.12), and hence

$$\boldsymbol{\delta}_1^T \mathbf{G} \mathbf{b}_1 \mathbf{b}_1^T \mathbf{G} \boldsymbol{\delta}_1 = c \boldsymbol{\delta}_1^T \mathbf{G} \boldsymbol{\delta}_1$$

with  $c$  as in equation (2.15). This implies that

$$\boldsymbol{\delta}_1^T (\mathbf{G} \mathbf{b}_1 \mathbf{b}_1^T \mathbf{G} - c \mathbf{G}) \boldsymbol{\delta}_1 = 0.$$

Let  $\lambda_1, \dots, \lambda_{k_1}$ ; and  $X_1, \dots, X_{k_1}$ ; be the eigenvalues and the eigenvectors, respectively, of  $\mathbf{G} \mathbf{b}_1 \mathbf{b}_1^T \mathbf{G} - c \mathbf{G}$ , and hence

$$\lambda_1 X_1^2 + \dots + \lambda_{k_1} X_{k_1}^2 = 0.$$

However,

$$X_1^2 = 1 - X_2^2 - \dots - X_{k_1}^2,$$

and hence

$$\lambda_1(1 - X_2^2 - \dots - X_{k_1}^2) + \lambda_2^2 X_2^2 + \dots + \lambda_{k_1}^2 X_{k_1}^2 = 0.$$

which implies

$$\lambda_1 = (\lambda_1 - \lambda_2)X_2^2 + \dots + (\lambda_1 - \lambda_{k_1})X_{k_1}^2,$$

a  $(k_1 - 1)$ -dimensional hyperellipsoid, with lengths of the semiaxes given by equation (2.13).

The surface area of the cap is given by

$$\int \dots \int_{\frac{x_2^2}{a_1^2} + \dots + \frac{x_{k_1}^2}{a_{k_1-1}^2} \leq 1} \frac{1}{\sqrt{1 - X_2^2 - X_3^2 - \dots - X_{k_1}^2}} dX_2 \dots dX_{k_1}. \quad (2.16)$$

For  $k = 2$ , we get

$$\begin{aligned} SA &= \int_{\frac{x_2^2}{a_1^2} \leq 1} \frac{1}{\sqrt{1 - X_2^2}} dX_2 \\ &= \int_{-a_1}^{a_1} \frac{1}{\sqrt{1 - X_2^2}} dX_2 \\ &= 2 \sin^{-1} a_1. \end{aligned}$$

In the example, when considering a fixed value of  $B$ , the eigenvalues of the  $\mathbf{H}$  matrix are given by 138.8249 and  $-492.2567$ . Hence, the value of  $a_1 = 0.4690$ . The surface area of the cap is then

$$2 \sin^{-1}(0.4690) = 0.9764.$$

Therefore, the proportion of directions included in the confidence cone is

$$\frac{0.9764}{2\pi} = 0.1555,$$

agreeing with the results in section 2.3.

Applying a transformation from rectangular to polar coordinates (see, for

example, Carter et al. (1986)) to equation (2.16). we get

$$\begin{aligned} X_2 &= \rho a_1 \sin \theta_1, \\ X_3 &= \rho a_2 \cos \theta_1 \sin \theta_2, \\ &\vdots \\ X_{k-1} &= \rho a_{k-2} \cos \theta_1 \cos \theta_2 \dots \cos \theta_{k-3} \sin \theta_{k-2}, \\ X_k &= \rho a_{k-1} \cos \theta_1 \cos \theta_2 \dots \cos \theta_{k-3} \cos \theta_{k-2}, \end{aligned}$$

where

$$\begin{aligned} 0 &\leq \rho \leq 1, \\ -\frac{\pi}{2} &\leq \theta_i \leq \frac{\pi}{2} \quad i = 1, \dots, k-3, \\ -\pi &\leq \theta_{k-2} \leq \pi. \end{aligned}$$

The Jacobian of the transformation is

$$g(r, \theta_1, \dots, \theta_{k-2}) = \rho^{k-2} (\cos \theta_1)^{k-3} \dots \cos \theta_{k-3} \prod_{i=1}^{k-1} a_i.$$

Therefore, the integral given by equation (2.16) is

$$\int_{\rho=0}^1 \int_{\theta_1=-\frac{\pi}{2}}^{\frac{\pi}{2}} \dots \int_{\theta_{k-3}=-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{\theta_{k-2}=-\pi}^{\pi} \frac{g(r, \theta_1, \dots, \theta_{k-2})}{h(r, \theta_1, \dots, \theta_{k-2})} \prod_{i=1}^{k-1} d\theta_i dr, \quad (2.17)$$

where

$$h(r, \theta_1, \dots, \theta_{k-2}) = \sqrt{1 - \sum_{i=1}^{k-1} \rho^2 a_i^2 \sin^2 \theta_i \prod_{j \leq i} \cos^2 \theta_j - \rho^2 a_{k-1}^2 \prod_{j \leq i} \cos^2 \theta_j}.$$

### 2.4.1 Continuation of the Example

In the example introduced in Section 2.3, we calculated the path of steepest ascent for the homogeneous case, and the method given by Box (1955) and Box and Draper (1987, pp. 190–194) was appropriate for that situation.

If  $B$  can be varied, then one or other of the levels of  $G$  needs to be chosen. If, for example, the high level of  $G$  is chosen, then the estimated transformed response is given by

$$\hat{y} = 2.3099 - 0.9975x_D - 1.2125x_F - 0.9325x_{B_{G+}},$$

Coefficients	Value	Std. Error	t value	Pr(>  t )
(Intercept)	2.5112	0.3304	7.6009	0.0000
$D$	-0.9975	0.3304	-3.0192	0.0129
$F$	-1.2125	0.3304	-3.6699	0.0043
$B_{G+}$	-0.9325	0.4672	-1.9958	0.0739
$B_{G-}$	0.6075	0.4672	1.3002	0.2277
$G$	-0.2013	0.3304	-0.6091	0.5560

Residual standard error: 1.322 on 10 degrees of freedom.  
 Multiple R-squared: 0.7411  
 F-statistic: 5.726 on 5 and 10 degrees of freedom, the p-value is 0.009481

Table 2.3: Summary of regression results using modified design.

where the coefficients of  $x_{B_{G+}}$  denotes the effect of  $B$  at the high level of  $G$ . Note that the standard errors of the coefficients of  $x_D$  and  $x_F$  are both 0.3304 while the standard error of the coefficient of  $x_{B_{G+}}$  is 0.4672, since the latter coefficient is the sum of  $x_B$  and  $x_{BG}$  coefficients. The path of steepest ascent is given by

$$x_F = 1.2155x_D, \quad x_{B_{G+}} = 0.9348x_D.$$

An alternative approach, which turns out to be more convenient, is to rewrite the design matrix as in the right hand panel of Table 2.1. The results of refitting the model, using the Splus commands given in Appendix A.3, are reported in Table 2.3, and as can be seen the standard error of the effect of  $B$  at high  $G$  is given directly. This method of calculating the “conditional” effects is due to Ankenman and Bisgaard (1995), and is equivalent to the “Trans-factor” technique developed by Taguchi (1987).

In this case, the eigenvalues of  $\mathbf{H}$  are

$$\lambda_1 = 204.6044, \quad \lambda_2 = -287.5042, \quad \lambda_3 = -513.0886,$$

leading to  $a_1 = 0.6450$  and  $a_2 = 0.5341$ . The surface area of the cap is given by

$$\begin{aligned} & \iint_{\frac{x_2^2}{a_1^2} + \frac{x_3^2}{a_2^2} \leq 1} \frac{1}{\sqrt{1 - X_2^2 - X_3^2}} dX_2 dX_3 \\ &= \int_{r=0}^1 \int_{\theta_1=-\pi}^{\pi} \frac{ra_1^2}{\sqrt{1 - r^2a_1^2 \sin^2 \theta_1 - r^2a_2^2 \cos^2 \theta_1}} d\theta_1 dr. \end{aligned}$$

$n$ -point	Percentage Excluded
2	90.31%
3	90.58%
4	90.44%
5	90.46%
6	90.47%
7	90.46%
8	90.46%
10	90.46%

Table 2.4: Percentage of directions excluded from the 95% confidence cone, calculated using Gaussian quadrature.

This integral can be evaluated using numerical integration. Splus commands for computing the proportion of directions included in the confidence cone using Gaussian quadrature (see, for example, Stroud, 1971) are given in Appendix A.4. For example, if we choose to use a 3-point Gaussian product formula, the estimated integral is given by

$$\frac{1}{4}2\pi \sum_{i=1}^3 \sum_{j=1}^3 w_i w_j f\left(\frac{1}{2} + \frac{1}{2}t_i, \pi t_j\right),$$

where

$$f(r, \theta_1) = \frac{ra_1a_2}{\sqrt{1 - r^2a_1^2 \sin^2 \theta_1 - r^2a_2^2 \cos^2 \theta_1}},$$

and  $t_1 = -0.77459667$ ,  $t_2 = 0$ ,  $t_3 = 0.77459667$ ,  $w_1 = 0.55555555$ ,  $w_2 = 0.88888888$  and  $w_3 = 0.55555555$ . The result is 1.1839. Since the surface area of the sphere is  $4\pi = 12.5663$ , the proportion included is 0.0942.

Higher order Gaussian quadrature can be used to obtain more accurate results. For example, a 10-point product Gaussian quadrature gives 0.0954 as the proportion included and so the 95% confidence region excludes 90.46% of possible directions.

## 2.5 Inequalities for Percentage Included

In the example above, the projection of the cap onto the  $(k - 1)$ -dimensional plane orthogonal to the path of steepest ascent is ellipsoidal. In this section,

some inequalities for the surface area of the cap will be developed.

For three dimensions, very simple bounds can be found by determining the major and minor axes,  $a_1$  and  $a_2$  respectively, of the ellipse. Since the surface area of the hypersphere above the ellipse must be greater than the surface area above a circle of radius  $a_2$  and less than the surface area of the hypersphere above a circle of radius  $a_1$  we have the inequality

$$1 - T(\sqrt{k-1} \frac{\sqrt{1-a_2^2}}{a_2}, k-1) \leq \gamma \leq 1 - T(\sqrt{k-1} \frac{\sqrt{1-a_1^2}}{a_1}, k-1).$$

For the example, with  $\alpha = 0.05$ ,  $a_1 = 0.6448$  and  $a_2 = 0.5339$ , the proportion included in the confidence cone is

$$0.0772 \leq \gamma \leq 0.1178,$$

which may be precise enough for practical purposes. For higher dimensions, if the axes of the projected ellipsoid are  $a_1 \geq a_2 \geq \dots \geq a_{k-1}$ , then

$$1 - T(\sqrt{k-1} \frac{\sqrt{1-a_{k-1}^2}}{a_{k-1}}, k-1) \leq \gamma \leq 1 - T(\sqrt{k-1} \frac{\sqrt{1-a_1^2}}{a_1}, k-1).$$

A sharper bound for  $k=3$  can be found by the following development:

$$SA = \iint_{\frac{x_2^2}{a_1^2} + \frac{x_3^2}{a_2^2} \leq 1} \frac{1}{\sqrt{1-X_2^2-X_3^2}} dX_2 dX_3.$$

Let

$$X_2 = ra_1 \sin \theta_1,$$

$$X_3 = ra_2 \cos \theta_1,$$

then,

$$SA = \int_{r=0}^1 \int_{\theta_1=-\pi}^{\pi} f(r, \theta_1) d\theta_1 dr,$$

where

$$f(r, \theta_1) = \frac{ra_1a_2}{\sqrt{1-r^2a_1^2 \sin^2 \theta_1 - r^2a_2^2 \cos^2 \theta_1}}.$$

Since

$$\min_{\theta_1 \in (-\pi, \pi)} f(r, \theta_1) = \frac{ra_1a_2}{\sqrt{1-r^2a_2^2}},$$

and

$$\max_{\theta_1 \in (-\pi, \pi)} f(r, \theta_1) = \frac{ra_1a_2}{\sqrt{1 - r^2a_1^2}},$$

we have

$$\int_0^1 \frac{2\pi ra_1a_2}{\sqrt{1 - r^2a_1^2}} dr \leq SA \leq \int_0^1 \frac{2\pi ra_1a_2}{\sqrt{1 - r^2a_2^2}} dr.$$

and hence,

$$\frac{2\pi a_1}{a_2} \left(1 - \sqrt{1 - a_2^2}\right) \leq SA \leq \frac{2\pi a_2}{a_1} \left(1 - \sqrt{1 - a_1^2}\right). \quad (2.18)$$

The surface area of a  $k$ -dimensional hypersphere of radius  $R$  is given in equation (2.4), and thus, we have

$$\gamma \geq \frac{\frac{2\pi a_1}{a_2} \left(1 - \sqrt{1 - a_2^2}\right)}{\frac{2\pi^{k/2} R^{k-1}}{\Gamma(k/2)}}, \quad (2.19)$$

and

$$\gamma \leq \frac{\frac{2\pi a_2}{a_1} \left(1 - \sqrt{1 - a_1^2}\right)}{\frac{2\pi^{k/2} R^{k-1}}{\Gamma(k/2)}}. \quad (2.20)$$

Therefore,

$$0.0933 \leq \gamma \leq .0976.$$

Note that when  $a_1 = a_2 = r$ , we obtain  $SA = 2\pi(1 - \sqrt{1 - r^2})$ , which can be also found by using equation (2.6) and the expression for the distribution function of the  $t$ -distribution.

When  $a_1$  and  $a_2$  are very similar, the upper bound and actual result are very close, but when  $a_1$  and  $a_2$  are quite different, there is a bigger deviation. Figure 2.3 shows the percentage included for various values of  $a_1$  and  $a_2$ . Figure 2.4 shows the percentage error of the upper bound given in equation (2.18) compared to the actual result computed using 32-point Gaussian quadrature.

For  $k = 4$ ,

$$SA = \iiint_{\frac{x_2^2}{a_1^2} + \frac{x_3^2}{a_2^2} + \frac{x_4^2}{a_3^2} \leq 1} \frac{1}{\sqrt{1 - X_2^2 - X_3^2 - X_4^2}} dX_2 dX_3 dX_4,$$

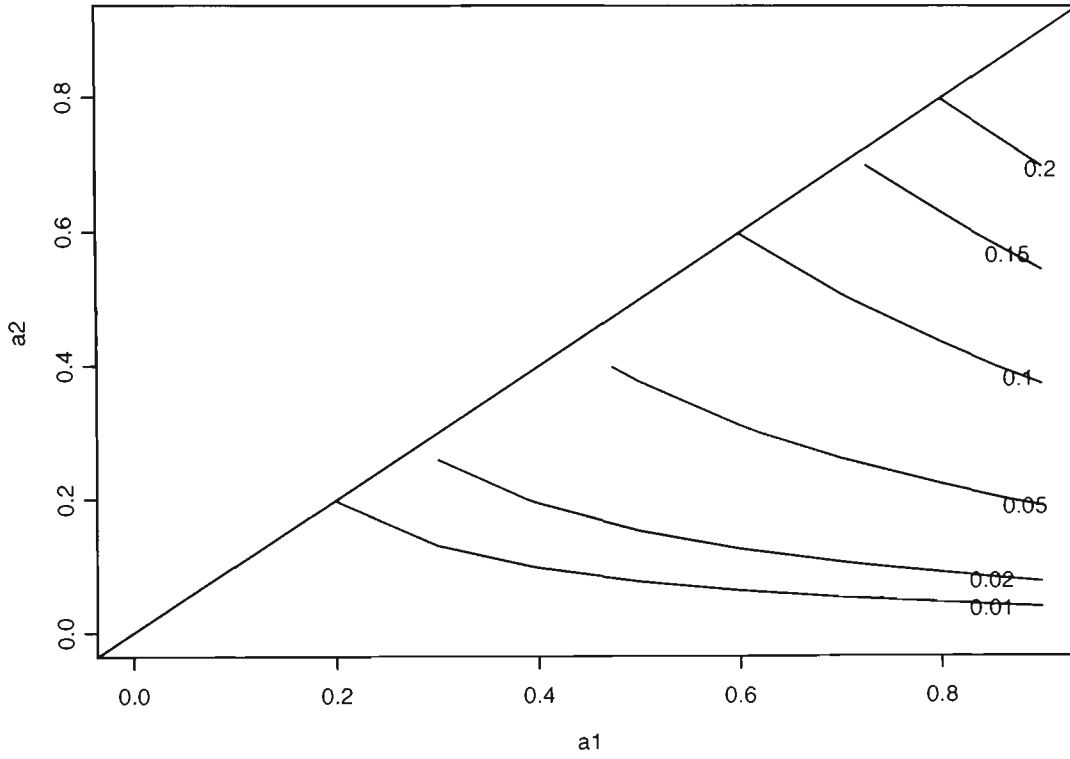


Figure 2.3: Proportion of directions included in the 95% confidence cone for the path of steepest ascent for various values of  $a_1$  and  $a_2$ , which are the major and minor axes of the projected ellipse.

that is, if  $X_1^2 + X_2^2 + X_3^2 + X_4^2 = 1$ , we have

$$SA = \iiint \frac{1}{X_1} dX_2 dX_3 dX_4.$$

Converting into “ellipsoidal” co-ordinates (see, for example, Spiegel, 1997, p. 50),

$$\begin{aligned} X_2 &= ra_1 \cos \theta_1 \sin \theta_2, \\ X_3 &= ra_2 \sin \theta_1 \sin \theta_2, \\ X_4 &= ra_3 \cos \theta_2, \end{aligned}$$

and since the Jacobian of the transformation is  $r^2 a_1 a_2 a_3 \sin \theta_2$ ,

$$SA = \int_{r=0}^1 \int_{\theta_1=0}^{2\pi} \int_{\theta_2=0}^{\pi} \frac{r^2 a_1 a_2 a_3 \sin \theta_2}{1 - r^2 a_1^2 \cos^2 \theta_1 \sin^2 \theta_2 - r^2 a_2^2 \sin^2 \theta_1 \sin^2 \theta_2 - r^2 a_3^2 \cos^2 \theta_2} d\theta_2 d\theta_1 dr.$$



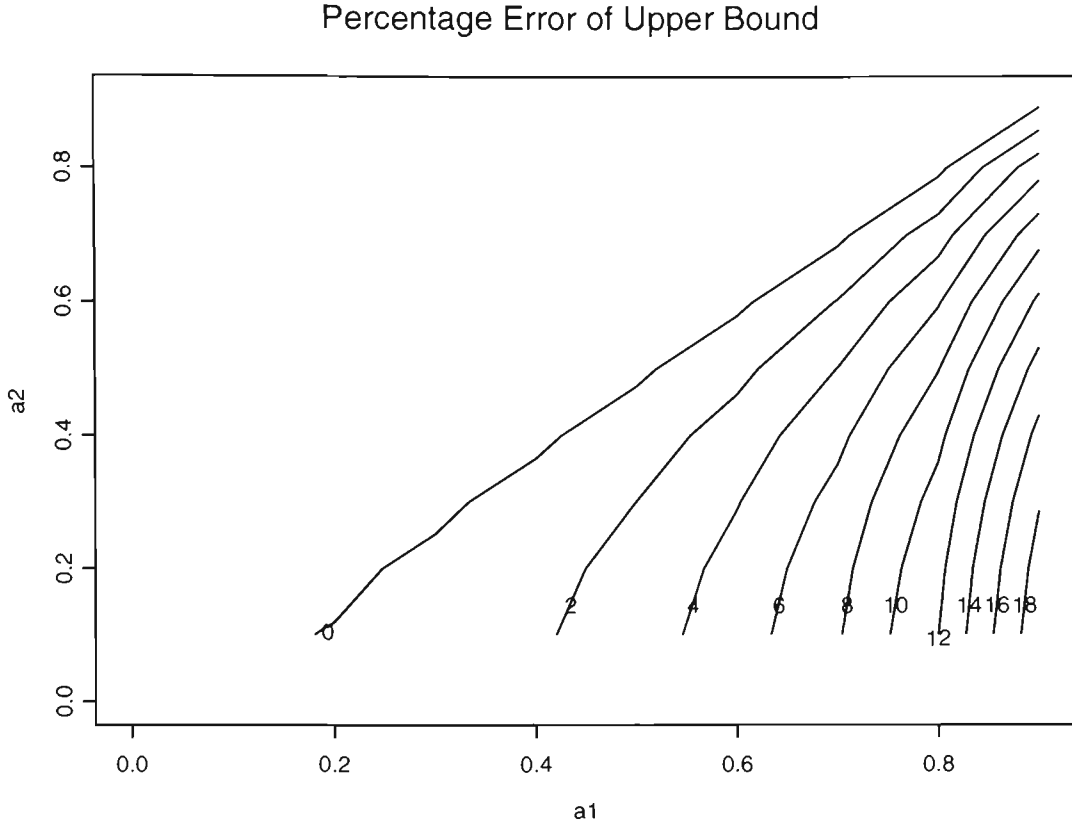


Figure 2.4: Percentage error of upper bound for the proportion of directions included in the 95% confidence cone given in equation (2.18) compared to the actual result computed using 32-point Gaussian quadrature.

The integrand can be re-written as

$$\frac{r^2 a_1 a_2 a_3 \sin \theta_2}{1 - r^2 a_2^2 \sin^2 \theta_2 - (r^2 a_1^2 - r^2 a_2^2) \cos^2 \theta_1 \sin^2 \theta_2 - r^2 a_3^2 \cos^2 \theta_2},$$

and since its maximum is

$$\frac{r a_1 a_2 a_3 \sin \theta_2}{\sqrt{1 - r^2 a_1^2 \sin^2 \theta_2 - r^2 a_3^2 \cos^2 \theta_2}},$$

when  $\cos^2 \theta_1 = 1$ , and its minimum is

$$\frac{r a_1 a_2 a_3 \sin \theta_2}{\sqrt{1 - r^2 a_2^2 \sin^2 \theta_2 - r^2 a_3^2 \cos^2 \theta_2}},$$

when  $\cos^2 \theta_1 = 0$ , then

$$L \leq SA \leq U,$$

where

$$\begin{aligned} L &= 2\pi \int_{r=0}^1 \int_{\theta_2=0}^{\pi} \frac{r^2 a_1 a_2 a_3 \sin \theta_2}{\sqrt{1 - r^2 a_2^2 \sin^2 \theta_2 - r^2 a_3^2 \cos^2 \theta_2}} d\theta_2 dr, \\ U &= 2\pi \int_{r=0}^1 \int_{\theta_2=0}^{\pi} \frac{r^2 a_1 a_2 a_3 \sin \theta_2}{\sqrt{1 - r^2 a_1^2 \sin^2 \theta_2 - r^2 a_3^2 \cos^2 \theta_2}} d\theta_2 dr. \end{aligned}$$

We can evaluate  $L$ , and in a similar fashion  $U$ . Make the substitution  $u = \cos \theta_2$ , and hence

$$\begin{aligned} L &= 2\pi \int_{r=0}^1 \int_{u=-1}^1 \frac{-r^2 a_1 a_2 a_3 \sqrt{1-u^2}}{\sqrt{1 - r^2 a_2^2 (1-u^2) - r^2 a_3^2 u^2}} \frac{1}{\sqrt{1-u^2}} du dr \\ &= 2\pi \int_{r=0}^1 \int_{u=-1}^1 \frac{r^2 a_1 a_2 a_3}{\sqrt{1 - r^2 a_2^2 + (r^2 a_2^2 - r^2 a_3^2) u^2}} du dr \\ &= \frac{2\pi}{\sqrt{a_2^2 a_3^2}} \int_{r=0}^1 \int_{u=-1}^1 \frac{r a_1 a_2 a_3}{\sqrt{\frac{1-r^2 a_2^2}{r^2 a_2^2 - r^2 a_3^2} + u^2}} du dr \\ &= \frac{2\pi}{\sqrt{a_2^2 a_3^2}} \int_{r=0}^1 \left[ r a_1 a_2 a_3 \sinh^{-1} \left( r u \sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}} \right) \right]_{u=-1}^1 dr \\ &= \frac{2\pi}{\sqrt{a_2^2 a_3^2}} \int_{r=0}^1 r a_1 a_2 a_3 \left\{ \sinh^{-1} \left( r \sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}} \right) - \sinh^{-1} \left( -r \sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}} \right) \right\} dr \\ &= \frac{4\pi a_1 a_2 a_3}{\sqrt{a_2^2 a_3^2}} \int_{r=0}^1 r \left( \sinh^{-1} r \sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}} \right) dr. \end{aligned}$$

Again, make the substitution

$$x = r \sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}},$$

and so

$$\begin{aligned}
L &= \frac{4\pi a_1 a_2 a_3}{\sqrt{a_2^2 - a_3^2}} \int_{x=0}^{\sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}}} \frac{x}{\sqrt{a_2^2 x^2 + (a_2^2 - a_3^2)}} \sinh^{-1}(x) \frac{(a_2^2 - a_3^2)}{(a_2^2 x^2 + (a_2^2 - a_3^2))^{\frac{3}{2}}} dx \\
&= 4\pi a_1 a_2 a_3 \sqrt{a_2^2 - a_3^2} \int_{x=0}^{\sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}}} \frac{x}{(a_2^2 x^2 + (a_2^2 - a_3^2))^2} \sinh^{-1}(x) dx \\
&= 4\pi a_1 a_2 a_3 \sqrt{a_2^2 - a_3^2} \left\{ \left[ \frac{-\sinh^{-1}(x)}{2a_2^2} \left[ \frac{1}{a_2^2 x^2 + (a_2^2 - a_3^2)} \right] \right]_0^{\sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}}} \right. \\
&\quad \left. + \frac{1}{2a_2^2} \int_{x=0}^{\sqrt{\frac{a_2^2 - a_3^2}{1 - r^2 a_2^2}}} \frac{1}{\sqrt{1+x^2}} \frac{1}{(a_2^2 x^2 + (a_2^2 - a_3^2))} dx \right\} \\
&= 4\pi a_1 a_2 a_3 \sqrt{a_2^2 - a_3^2} \left\{ \frac{-(1 - a_2^2)}{2a_2^2 (a_2^2 - a_3^2)} \sinh^{-1} \left( \sqrt{\frac{a_2^2 - a_3^2}{1 - a_2^2}} \right) \right. \\
&\quad \left. + \frac{1}{4a_2^2} \int_{w=0}^{\frac{a_2^2 - a_3^2}{1 - 2a_2^2}} \frac{1}{\sqrt{w+1}} \frac{1}{\sqrt{w}} \frac{1}{\sqrt{a_2^2 w + (a_2^2 - a_3^2)}} dw \right\},
\end{aligned}$$

after making the substitution  $w = x^2$ .

Maple<sup>TM</sup> (Monogan et al., 2000) gives the later integral as

$$\frac{\tanh^{-1} \left\{ \frac{\frac{1}{2} \frac{-a_3^2 + a_2^2 w + a_2^2 - 2wa_3^2}{a_2^2 a_3 \sqrt{\frac{-a_2^2 + a_3^2}{a_2^4}} \sqrt{w(w+1)}}}{a_2^2} \right\} \sqrt{w} \sqrt{1+w}}{a^3 \sqrt{\frac{-a_2^2 + a_3^2}{a_2^4}} \sqrt{w(w+1)} a_2^2}.$$

Cancelling out, we get

$$\begin{aligned}
&\frac{\tanh^{-1} \left\{ \frac{(a_2^2 - a_3^2) + w(a_2^2 - 2a_3^2)}{2a_3 \sqrt{a_3^2 - a_2^2} \sqrt{w(w+1)}} \right\}}{a^3 \sqrt{a_2^2 - a_3^2}} \\
&= \frac{\tanh^{-1} \left\{ \frac{(a_2^2 - a_3^2) + w(a_2^2 - 2a_3^2)}{2a_3 i \sqrt{a_2^2 - a_3^2} \sqrt{w(w+1)}} \right\}}{a^3 i \sqrt{a_2^2 - a_3^2}} \\
&= \frac{1}{a_3 \sqrt{a_2^2 - a_3^2}} \tan^{-1} \left\{ \frac{(a_2^2 - a_3^2) + w(a_2^2 - 2a_3^2)}{2a_3 \sqrt{a_2^2 - a_3^2} \sqrt{w(w+1)}} \right\}.
\end{aligned}$$

Hence,

$$L = 4\pi a_1 a_2 a_3 \sqrt{a_2^2 - a_3^2} \left\{ \frac{-(1 - a_2^2)}{2a_2^2 (a_2^2 - a_3^2)} \sinh^{-1} \left( \sqrt{\frac{a_2^2 - a_3^2}{1 - a_2^2}} \right) - \left[ \frac{1}{a_3 \sqrt{a_2^2 - a_3^2}} \frac{1}{4a_2^2} \tan^{-1} \left\{ \frac{(a_2^2 - a_3^2) + w(a_2^2 - 2a_3^2)}{2a_3 \sqrt{a_2^2 - a_3^2} \sqrt{w(w+1)}} \right\} \right]_0^{\frac{a_2^2 - a_3^2}{1 - a_2^2}} \right\}.$$

Simplifying the square brackets, we obtain

$$\begin{aligned} & \frac{1}{4a_2^2 a_3 \sqrt{a_2^2 - a_3^2}} \tan^{-1} \left\{ \frac{(a_2^2 - a_3^2) + \frac{a_2^2 - a_3^2}{1 - a_2^2} (a_2^2 - 2a_3^2)}{2a_3^2 \sqrt{a_2^2 - a_3^2} \sqrt{\frac{(a_2^2 - a_3^2)}{(1 - a_2^2)} \left\{ \frac{(a_2^2 - a_3^2)}{(1 - a_2^2)} + 1 \right\}}} \right\} \\ &= \frac{1}{4a_2^2 a_3 \sqrt{a_2^2 - a_3^2}} \left[ \tan^{-1} \left\{ \frac{1 - 2a_3^2}{2a_3 \sqrt{1 - a_3^2}} \right\} - \frac{\pi}{2} \right], \end{aligned}$$

and therefore

$$\begin{aligned} L &= 4\pi a_1 a_2 a_3 \sqrt{a_2^2 - a_3^2} \left[ \frac{1}{4a_2^2 a_3 \sqrt{a_2^2 - a_3^2}} \left( \frac{\pi}{2} - \tan^{-1} \left\{ \frac{1 - a_3^2}{2a_3 \sqrt{1 - a_3^2}} \right\} \right) - \frac{(1 - a_2^2)}{2a_2^2 (a_2^2 - a_3^2)} \sinh^{-1} \left( \sqrt{\frac{a_2^2 - a_3^2}{1 - a_2^2}} \right) \right] \\ &= \frac{\pi a_1 a_3}{a_2} \left[ \frac{1}{a_3} \cot^{-1} \left( \frac{1 - 2a_3^2}{2a_3 \sqrt{1 - a_3^2}} \right) - \frac{2(1 - a_2^2)}{\sqrt{a_2^2 - a_3^2}} \sinh^{-1} \left( \sqrt{\frac{a_2^2 - a_3^2}{1 - a_2^2}} \right) \right]. \end{aligned} \tag{2.21}$$

Similarly, for  $U$ , replacing  $a_2$  for  $a_1$  in the result for  $L$ , we obtain

$$U = \frac{\pi a_2 a_3}{a_1} \left[ \frac{1}{a_3} \cot^{-1} \left( \frac{1 - 2a_3^2}{2a_3 \sqrt{1 - a_3^2}} \right) - \frac{2(1 - a_1^2)}{\sqrt{a_1^2 - a_3^2}} \sinh^{-1} \left( \sqrt{\frac{a_1^2 - a_3^2}{1 - a_1^2}} \right) \right]. \tag{2.22}$$

For example, with  $a_1 = 0.6$ ,  $a_2 = 0.5$ , and  $a_3 = 0.4$ , we obtain  $L = 0.5405$  and  $U = 0.5561$ , and hence

$$0.0274 \leq \gamma \leq 0.0281.$$

## 2.6 Other Approaches to Evaluating the Percentage Included.

For homoscedastic data, equation (2.8) gives an expression for the surface area of the cap of the sphere within the confidence cone relative to the surface area of the sphere. Consider the case where  $k = 3$ . If we divide the circle given by the projection of the cap onto the plane orthogonal to the path of steepest ascent into equal sectors each with angle  $\theta$ , then the relative surface area on the sphere above a sector is equal to

$$\frac{\theta}{2\pi} \left[ 1 - T \left( \sqrt{2} \frac{\sqrt{1-r^2}}{r}, 2 \right) \right].$$

For heteroscedastic data, the projected confidence cone is ellipsoidal and for  $k = 3$  is an ellipse. We can also divide the ellipse into sectors. As the number of sectors increases the relative surface area above a sector becomes closer and closer to the relative surface area above a circle, and is given by

$$\frac{d\theta}{2\pi} \left[ 1 - T \left( \sqrt{2} \frac{\sqrt{1-r^2(\theta)}}{r(\theta)}, 2 \right) \right],$$

where

$$r(\theta) = \frac{1}{\sqrt{\frac{\cos^2 \theta}{a_1^2} + \frac{\sin^2 \theta}{a_2^2}}}.$$

The total surface area is then given by

$$\begin{aligned} & \frac{1}{2\pi} \int_0^{2\pi} 1 - T \left( \sqrt{2} \frac{\sqrt{1-r^2(\theta)}}{r(\theta)} \right) d\theta \\ &= 1 - \frac{1}{2\pi} \int_0^{2\pi} T \left( \sqrt{2} \frac{\sqrt{1-r^2(\theta)}}{r(\theta)} \right) d\theta. \end{aligned} \quad (2.23)$$

This method reduces a double integration into a univariate integration. Using the example, with  $a_1 = 0.6450$  and  $a_2 = 0.5341$ , we obtain  $\gamma = 0.0954337$ .

Although Gaussian quadrature or a similar method to that presented in this section could be applied for any number of dimensions, it is very easy to use

Monte-Carlo simulation to give an approximate result. The method is quite simple. For  $k$  dimensions, for each simulation:

1. Generate  $k$  standard normal random numbers:  $x_1, x_2, \dots, x_k$ .
2. Reduce the random numbers so that they fall on the surface of the sphere using

$$z_i = \frac{x_i}{\sqrt{\sum_{i=1}^k x_i^2}} \quad i = 1, \dots, k. \quad (2.24)$$

3. The proportion of simulations, where

$$\sum \frac{z_i^2}{a_i^2} \leq 1 \quad i = 1, \dots, k-1 \quad (2.25)$$

$$\text{and} \quad z_k > 0, \quad (2.26)$$

estimates the proportion of directions included in the  $100(1 - \alpha)\%$  confidence cone, where the  $a_1, \dots, a_k$  are generated from the  $\mathbf{H}$  matrix given by equation (2.14) using equation (2.13). The sample size can be chosen to make the estimated proportion sufficiently precise.

## 2.7 Computations

An Splus function, `perinc`, is given in Appendix A.5 and does the necessary calculations. For two dimensions, it determines the exact proportion of directions included in the confidence cone. For three dimensions, it uses numerical integration using equation (2.23). For four dimensions, it computes upper and lower bounds for the proportion of directions included in the confidence cone using the formulae developed in section 2.5, while Monte-Carlo estimation is used for four or higher dimensions.

## 2.8 Conclusion

In this chapter, the method given by Box (1955) and Box and Draper (1987, pp. 190–194) was examined. The method determines whether the path of steepest ascent has been determined precisely enough by giving the fraction of possible directions included in the confidence cone. An alternative derivation of the method is given when the design is orthogonal and the data is homoscedastic. The fraction of all possible directions included in the confidence cone is shown to be a function of the tail area of the  $t$ -distribution.

The method was applied to a fractional factorial example. In the example, the response was a count. The response was transformed using a modified square root transformation, and two main effects as well as a significant interaction string was found to be important. Each of the interactions in the interaction string involved a quantitative variable and a qualitative variable. Assuming that one of the interactions in the interaction string could be identified as real, the analysis depends on whether the quantitative variable is fixed or variable. If the quantitative variable is fixed, the path of steepest ascent only depends on the two significant main effects and the fraction of directions within the confidence cone can be determined using the current method. On the other hand, if the quantitative variable can be varied, then the path of steepest ascent depends on which of the two levels of the qualitative variable is chosen. One problem here is that the regression coefficients are now heteroscedastic – the effect of the quantitative variable at either of the levels of the qualitative variable has a higher variance than the other two effects. Without modification, the method of Box (1955) and Box and Draper (1987, pp. 190–194) does not apply in this case.

A modification of the method for heteroscedastic regression coefficients was given in this chapter. It was shown how the confidence region on the sphere can be projected onto the hyperplane passing through the origin orthogonal to the path of steepest ascent, providing a hyperellipsoid. Methods for determining the lengths of the axes of the hyperellipsoid are given, and this enables an expression for the surface area of the cap. When there are two regression coefficients an exact expression can be given. For higher dimensions, a transformation onto a

rectangular set of coordinates is convenient. Gaussian quadrature can be used to determine the surface area of the cap. Alternatively, for three dimensions we can express the surface area as a univariate integral involving the cumulative distribution function of the  $t$ -distribution.

In addition, some inequalities for the surface area of the cap have been established when  $k = 3$  and  $k = 4$ . These inequalities are quite sharp, and therefore provide an acceptable solution without the need for numerical integration. For higher dimensions, it is very easy to determine the proportion of directions included in the confidence cone using Monte-Carlo methods.

In the next chapter, the results for heterogeneous errors will be used for the important class of generalised linear models. Even more accurate results can be obtained using profile likelihood based confidence regions.



# Chapter 3

## Application to Generalised Linear Models

### 3.1 Introduction

As an alternative to applying results for normally distributed data to a transformed response, we can use generalised linear models (GLMs), where the mean of a non-normally distributed response is modelled as a function of a linear combination of the explanatory variables.

In linear models, the assumptions are:

1. The observations  $y_i \sim N(\mu, \sigma^2)$ ,
2.  $E(y_i) = \mu_i = \mathbf{x}'_i \boldsymbol{\beta}$ , where  $\boldsymbol{\beta}$  is a vector of parameters,
3. The  $y_i$ 's are independent.

In generalised linear models, assumptions 1 and 2 are widened:

1. The  $y_i$ 's have a distribution from the exponential family of distributions (which includes the normal, binomial, Poisson, multinomial, gamma, etc.).

A random variable  $y$  belongs to the exponential family if its density function or probability function has the form

$$f(y_i, \theta_i, \phi) = \exp \{ r(\phi) [y_i \theta_i - g(\theta_i)] + h(y_i, \phi) \}, \quad i = 1, 2, \dots, n,$$

where  $\phi$  is the scale parameter and  $\theta_i$  is the natural location parameter.

2. (a) There is a linear predictor  $\eta = \mathbf{x}'\boldsymbol{\beta}$ .
- (b) There is a link function  $g()$  linking  $\mu_i = E(y_i)$  with the linear predictor  $\eta_i : g(\mu_i) = \eta_i$ .

The choice of the link function determines the nature of the model used. The choice of distribution and link is an important part of modelling using generalised linear models. Myers and Montgomery (1997) present canonical links and models for some members of the exponential family:

Distribution	Location Parameter	Link Function	Model
Normal	$\theta = \mu$	$\mu = \mathbf{x}'\boldsymbol{\beta}$ (Identity)	$\mu = \mathbf{x}'\boldsymbol{\beta}$
Poisson	$\theta = \ln(\mu)$	$\ln(\mu) = \mathbf{x}'\boldsymbol{\beta}$ (log)	$\mu = e^{\mathbf{x}'\boldsymbol{\beta}}$
Binomial	$\theta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\ln\left(\frac{\mu}{1-\mu}\right) = \mathbf{x}'\boldsymbol{\beta}$ (logit)	$\mu = \frac{1}{1+e^{-\mathbf{x}'\boldsymbol{\beta}}}$
Exponential	$\theta = 1/\mu$	$1/\mu = \mathbf{x}'\boldsymbol{\beta}$ (reciprocal)	$\mu = \frac{1}{\mathbf{x}'\boldsymbol{\beta}}$

3. The  $y_i$ s are independent.

Generalised linear models are usually fitted using iteratively re-weighted least squares, which is equivalent to maximum likelihood estimation. The asymptotic variance-covariance matrix of the coefficients in the linear predictor is given by

$$V(\hat{\boldsymbol{\beta}}) = \frac{(\mathbf{X}^T \boldsymbol{\Delta} \mathbf{U} \boldsymbol{\Delta} \mathbf{X})^{-1}}{[r(\phi)]^2}, \tag{3.1}$$

where  $\mathbf{U}$  is a diagonal matrix with the  $i$ th diagonal element consisting of the estimated variance of the  $i$ th data point,  $r(\phi)$  is the function of the scale parameter in the density function of the exponential family of distributions (for Binomial and Poisson data,  $r(\phi) = 1$ , while for Normal data  $r(\phi) = \frac{1}{\sigma^2}$ ), and  $\boldsymbol{\Delta}$  is a diagonal

matrix with  $\delta_i$ , the partial derivative of the location parameter with respect to the linear predictor, on the  $i$ th diagonal. When the canonical link is used then  $\Delta = \mathbf{I}$ . For details see Myers and Montgomery (1997).

The generalisation of equation (2.3) depends on whether there is overdispersion (Myers and Montgomery, 1997) present in the data. Overdispersion corresponds to an inflation of the natural variance expected of the distribution, and is common in many applications. To determine whether overdispersion is present, the residual deviance

$$\lambda(\boldsymbol{\beta}) = -2 \ln \left[ \frac{L(\boldsymbol{\beta})}{L(\mu)} \right]$$

is calculated, where  $L(\boldsymbol{\beta})$  is the maximum likelihood for the fitted model and  $L(\mu)$  is the maximum likelihood associated with the model given by

$$y_i = \mu_i + \epsilon_i, \quad i = 1, \dots, n.$$

If the residual deviance divided by the degrees of freedom is much greater than the scale parameter expected from the distribution then we say that overdispersion is present.

In this chapter, the results of Chapter 2 will be extended to cover the case of generalised linear models. The example used in Chapter 2 will be analysed using GLMs rather than analysing the transformed response. The Wald and profile likelihood approaches will be used to give a confidence cone for the path of steepest ascent, and the proportion of directions included in the confidence cone will be calculated. In addition, bootstrap methods will be used to check the appropriateness of the confidence cones.

Part of the material covered in this chapter has been published in Sztendur and Diamond (2002).

## 3.2 Wald Approach

Assuming a linear approximation is appropriate and no overdispersion is present, we can use a similar development as in section 2.4 to obtain a  $100(1 - \alpha)\%$

Coefficients	Value	Std. Error	t value
(Intercept)	1.0446786	0.1824753	5.725040
$D$	-0.8958797	0.1125891	-7.957071
$F$	-1.1756876	0.1398239	-8.408342
$B$	-0.2935723	0.1161399	-2.527747
$G$	-0.3012060	0.1161399	-2.593476
$BG$	-0.8206526	0.1161399	-7.066068

(Dispersion Parameter for Poisson family taken to be 1 )

Null Deviance: 313.1613 on 15 degrees of freedom

Residual Deviance: 16.60859 on 10 degrees of freedom

Table 3.1: Summary of fitted GLM model with Poisson link and original design.

confidence region for the  $\delta$ 's as

$$\mathbf{b}^T \mathbf{K} \mathbf{b} - \boldsymbol{\delta}^T \mathbf{K} \mathbf{b} (\boldsymbol{\delta}^T \mathbf{K} \boldsymbol{\delta})^{-1} \mathbf{b}^T \mathbf{K} \boldsymbol{\delta} \leq \chi^2_{1-\alpha}(k-1), \quad (3.2)$$

with

$$V(\mathbf{b}) = V \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \frac{(\mathbf{X}^T \boldsymbol{\Delta} \mathbf{U} \boldsymbol{\Delta} \mathbf{X})^{-1}}{[r(\phi)]^2} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{12}^T & \mathbf{K}_{22} \end{pmatrix}, \quad (3.3)$$

where  $\mathbf{K} = \mathbf{K}_{11}^{-1}$ ,  $\mathbf{b}_1$  corresponds to the  $k_1$  variables in the path of steepest ascent, and  $\mathbf{b}_2$  corresponds to the other variables, including at least the intercept.

An approximate  $100(1-\alpha)\%$  confidence region for the  $\delta$ 's when overdispersion is considered is

$$\frac{\nu_b [\mathbf{b}_1^T \mathbf{K} \mathbf{b}_1 - \boldsymbol{\delta}^T \mathbf{K} \mathbf{b}_1 (\boldsymbol{\delta}^T \mathbf{K} \boldsymbol{\delta})^{-1} \mathbf{b}_1^T \mathbf{K} \boldsymbol{\delta}]}{(k-1)\lambda(\boldsymbol{\beta})} \leq F_{1-\alpha}(k-1, \nu_b), \quad (3.4)$$

where  $\mathbf{K}$  is as above and  $\lambda(\boldsymbol{\beta})$  is the residual deviance.

As in section 2.3, a model was fitted using the Splus commands given in Appendix B.1 with defects as the response variable and the factors  $D$ ,  $F$ ,  $B$ ,  $G$ , and  $BG$  as explanatory variables. In this case, however, the generalised linear model with Poisson link was used rather than using a transformed response. The results are given in Table 3.1.

If  $B$  is fixed, the linear predictor is given by

$$\text{constant} - 0.8959x_D - 1.1759x_F, \quad (3.5)$$

where the standard errors of the coefficients are 0.1126 and 0.1398 respectively. The path of steepest ascent is given by

$$x_F = 1.3125x_D.$$

Using the method given in Chapter 2. the eigenvalues of

$$\mathbf{H} = \mathbf{K}\mathbf{b}_1\mathbf{b}_1^T\mathbf{K} - c\mathbf{K}, \quad (3.6)$$

with

$$c = \mathbf{b}_1^T\mathbf{K}\mathbf{b}_1 - \chi^2_{0.95}(1) \quad (3.7)$$

were calculated as

$$\lambda_1 = 235.9301, \quad \lambda_2 = -8552.2471,$$

leading to

$$a_1 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_2}} = 0.1638,$$

and hence the proportion of directions included in the confidence cone is

$$\frac{2 \sin^{-1} a_1}{2\pi} = 0.0524,$$

which is considerably less than using the transformation used in Chapter 2. The `Splus` function, `perincglm`, given in Appendix B.3, was used to compute the above result and the results that follow in this section.

When we have to take overdispersion into account, we need the eigenvalues of  $\mathbf{H}$ , given in equation (3.6), but with

$$c = \mathbf{b}_1^T\mathbf{K}\mathbf{b}_1 - \frac{(k-1)\lambda(\beta)F_{1-\alpha}(k-1, \nu_b)}{\nu_b}, \quad (3.8)$$

where  $\alpha = 0.05$ ,  $k = 2$  and  $\nu_b = 10$ . The eigenvalues were calculated as

$$\lambda_1 = 507.1564, \quad \lambda_2 = -8250.7876.$$

Therefore,  $a_1 = 0.2406$ , and hence the proportion of directions included in the approximate confidence cone is 0.0774.

Coefficients	Value	Std. Error	t value
(Intercept)	1.0446786	0.1824753	5.725040
$D$	-0.8958797	0.1125891	-7.957071
$F$	-1.1756876	0.1398239	-8.408342
$B_{G+}$	-1.1142248	0.1982336	-5.620768
$B_{G-}$	0.5270803	0.1210676	4.353602
$G$	-0.3012060	0.1161399	-2.593476

(Dispersion Parameter for Poisson family taken to be 1 )

Null Deviance: 313.1613 on 15 degrees of freedom

Residual Deviance: 16.60859 on 10 degrees of freedom

Table 3.2: Summary of fitted GLM model with Poisson link and modified design.

As in section 2.4.1, it might be desired to reduce the number of defects by allowing  $B$  to vary. After fitting a generalised linear model to the data given in Table 2.1, using the conditional effect method with the Splus commands given in Appendix B.2, the linear predictor at the high level of  $G$  is given by

$$0.7434 - 0.8959x_D - 1.1759x_F - 1.1142x_{B_{G+}},$$

where the standard errors of the coefficients are 0.1126, 0.1398, and 0.1982 respectively. See the summary of results in Table 3.2.

The path of steepest descent is given by

$$x_F = 1.3125x_D, \quad x_{B_{G+}} = 1.2437x_D.$$

Using the method given in Chapter 2, the eigenvalues of  $\mathbf{H}$  in equation (3.6) with

$$c = \mathbf{b}_1^T \mathbf{K} \mathbf{b}_1 - \chi_{0.95}^2(2)$$

were calculated as

$$\lambda_1 = 291.8115, \quad \lambda_2 = -5042.3440, \quad \lambda_3 = -10652.3330,$$

leading to

$$a_1 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_2}} = 0.2339,$$

and

$$a_2 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_3}} = 0.1633.$$

To consider overdispersion, the eigenvalues of  $\mathbf{H}$  in equation (3.6) were again calculated but with

$$c = \mathbf{b}_1^T \mathbf{K} \mathbf{b}_1 - 2\lambda(\boldsymbol{\beta}) F_{0.95}(2, 10)/10.$$

The eigenvalues are

$$\lambda_1 = 670.131, \quad \lambda_2 = -4766.639, \quad \lambda_3 = -10118.936,$$

giving

$$a_1 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_2}} = 0.3511$$

and

$$a_2 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_3}} = 0.2492.$$

Numerical integration, using `perincglm` given in Appendix B.3, shows that the 95% confidence region excludes about 97.76% of possible directions considering the overdispersion and 99.04% when the possible overdispersion is ignored. Note that when using the transformed response in Chapter 2, the 95% confidence region excluded about 90.46%. This shows the increased effectiveness by using the generalised linear model approach rather than simply transforming the response. This agrees with the results of Myers and Montgomery (1997) and Lewis et al. (2001*a,b*) who showed that using GLMs instead of transforming the responses usually results in shorter and more realistic confidence intervals for predicted values. Note that shorter confidence intervals or shorter confidence regions are only better when they have the correct coverage.

### 3.3 Likelihood Based Confidence Regions

The Wald approach used in the previous section is based on a quadratic approximation to the log likelihood function at the maximum likelihood estimator. A more accurate result could be obtained by examining the log likelihood directly. In this section a profile likelihood based confidence region will be determined and illustrated using the example.

For the example, the linear predictor

$$\beta_0 + \beta_1 D + \beta_2 F + \beta_3 BG_+ + \beta_4 BG_- + \beta_5 G \quad (3.9)$$

can be written as

$$\beta_0 + \delta(\gamma_1 D + \gamma_2 F + \gamma_3 BG_+) + \beta_4 BG_- + \beta_5 G, \quad (3.10)$$

where

$$\delta = \sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2},$$

and

$$\gamma_i = \frac{\beta_i}{\delta} \quad i = 1, 2, 3.$$

Note that  $\gamma_1^2 + \gamma_2^2 + \gamma_3^2 = 1$ . As far as the path of steepest ascent is concerned,  $\beta_0$ ,  $\beta_4$ , and  $\beta_5$  are nuisance parameters. It is convenient to apply a Householder transformation, so that the path is vertical, that is, proportional to  $(0, 0, 1)^T$  in the new co-ordinate system. To do this rewrite the model as

$$\begin{aligned} & \begin{pmatrix} 1 & BG_- & G \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_4 \\ \beta_5 \end{pmatrix} + \delta \begin{pmatrix} D & F & BG_+ \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & BG_- & G \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_4 \\ \beta_5 \end{pmatrix} + \delta \begin{pmatrix} D & F & BG_+ \end{pmatrix} \mathbf{H}_u \mathbf{H}_u \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & BG_- & G \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_4 \\ \beta_5 \end{pmatrix} + \delta \begin{pmatrix} D & F & BG_+ \end{pmatrix} \mathbf{H}_u \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix}. \end{aligned}$$

$\mathbf{H}_u$  is the Householder transformation (see, for example, Seber, 1977, p. 312), such that

$$\mathbf{H}_u \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \end{pmatrix}.$$

$\mathbf{H}_u$  can be determined as  $\mathbf{H}_u = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T$ , where

$$\mathbf{u} = \frac{\mathbf{e}_3 - \|\mathbf{e}_3\|\hat{\gamma}}{\|\mathbf{e}_3 - \|\mathbf{e}_3\|\hat{\gamma}\|},$$



and

$$\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \hat{\gamma}_i = \begin{pmatrix} \hat{\gamma}_{i1} \\ \hat{\gamma}_{i2} \\ \hat{\gamma}_{i3} \end{pmatrix}.$$

Note that  $\zeta_1^2 + \zeta_2^2 + \zeta_3^2 = 1$  since  $\mathbf{H}_u$  is orthogonal.

Define  $\boldsymbol{\theta}_1 = (\zeta_1, \zeta_2)$  and  $\boldsymbol{\theta}_2 = (\delta, \beta_0, \beta_1, \beta_5)$ . The profile likelihood for  $\boldsymbol{\theta}_1$  is defined as

$$R(\boldsymbol{\theta}_1) = \max_{\boldsymbol{\theta}_2} \left[ \frac{L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{L(\hat{\boldsymbol{\theta}})} \right].$$

For details see Meeker and Escobar (1995).

Now,

$$\begin{aligned} -2 \log R(\boldsymbol{\theta}_1) &= \max_{\boldsymbol{\theta}_2} \left[ -2 \log L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + 2 \log L(\hat{\boldsymbol{\theta}}) \right] \\ &= D_{M_0} - D_M \\ &= D_{M_0} - 16.60859, \end{aligned}$$

where  $D_{M_0}$  is the model deviance with  $\boldsymbol{\theta}_1$  given and  $D_M$  is the deviance with  $\boldsymbol{\theta}_1$  unrestricted. Note that if overdispersion is ignored,  $\phi$  is known, and

$$\frac{D_{M_0} - D_M}{\phi} \sim \chi^2(p - q), \quad (3.11)$$

where  $p$  is the number of regressors for model  $M$ , including the intercept; and  $q$  is the number of parameters with  $\boldsymbol{\theta}_1$  given.

A similar procedure can be applied when overdispersion is present. In this case,  $\phi$  is not known and it is customary to use (see, for example, Venables and Ripley, 1999) the approximate result

$$\frac{D_{M_0} - D_M}{\hat{\phi}(p - q)} \sim F(p - q, n - p). \quad (3.12)$$

A grid of values of  $\zeta_1$  and  $\zeta_2$  was generated and  $\zeta_3$  was calculated as

$$\zeta_3 = \sqrt{1 - \zeta_1^2 - \zeta_2^2},$$

corresponding to the upper hemisphere. For each point in the grid the deviance  $D_{M_0}$  was determined and a contour was drawn at the level where

$$D_{M_0} = D_M + \chi_{0.95}^2(2) = 16.608 + 5.991 = 22.599,$$

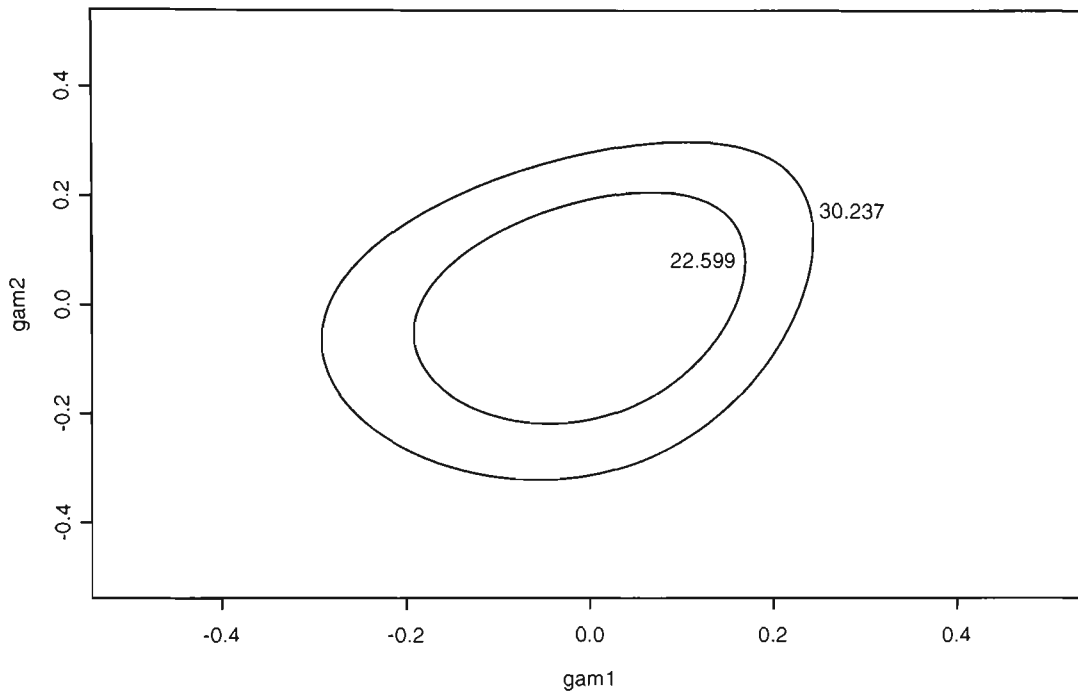


Figure 3.1: 95% profile likelihood regions for the path of steepest ascent. The inner contour is based on  $\chi^2$  quantiles and gives the region for the case when overdispersion is ignored. The outer contour is based on theoretical quantiles of the  $F(2, 10)$  distribution and gives the region for the case when overdispersion is considered.

when not considering the overdispersion. If overdispersion is considered, the contour is drawn at the level where

$$D_{M_0} = D_M + 2(1.6608)F_{0.95}(2, 10) = 30.236.$$

A function, `profpath`, to compute the profile likelihood regions is given in Appendix B.4.

The contour diagrams given in Figure 3.1 represent profile likelihood regions. The inner contour gives the 95% confidence region for possible directions if overdispersion is ignored. The outer contour gives the 95% region for possible directions if overdispersion is considered. To obtain the surface area of the cap of the sphere above the interior of the contour, a method similar to the method in section 2.6 was used. The Splus function, `profpath`, given in Appendix B.4,

gives a set of points on the contour. For each point polar co-ordinates

$$\theta = \tan^{-1} \left( \frac{\zeta_2}{\zeta_1} \right) \text{ and } r(\theta) = \sqrt{\zeta_1^2 + \zeta_2^2}$$

can be computed. The surface area above the segment between two consecutive points  $(r(\theta_i), \theta_i)$  and  $(r(\theta_{i+1}), \theta_{i+1})$  is approximately

$$\frac{(\theta_{i+1} - \theta_i)}{2\pi} \left[ 1 - \frac{1}{2} \left( T \left( \sqrt{2} \frac{\sqrt{1 - r(\theta_i)^2}}{r(\theta_i)}, 2 \right) + T \left( \sqrt{2} \frac{\sqrt{1 - r(\theta_{i+1})^2}}{r(\theta_{i+1})}, 2 \right) \right) \right].$$

and by summing all these values we can get an approximation to the surface area of the cap which will converge to the true value as the number of points on the contour increases. The calculations were done using the Splus function `contourprop` given in Appendix B.5.

The calculated surface area for the outer contour in Figure 3.1 was 0.0209 and the corresponding percentage excluded 97.91%. The surface area corresponding to the inner contour was 0.0095 and the percentage excluded was 99.05%, which is slightly larger than the value obtained using the Wald approach (97.76% and 99.04% respectively). Table 3.3 summarises these results.

For dimensions higher than three, Monte-Carlo simulation is convenient. For each simulation:

1. Generate  $k$  standard normal random numbers:  $x_1, x_2, \dots, x_k$ .
2. Reduce the random numbers so that they fall on the surface of the sphere using

$$z_i = \frac{x_i}{\sqrt{\sum_{i=1}^k x_i^2}} \quad i = 1, \dots, k. \quad (3.13)$$

3. For  $k = 3$ , use GLMs to fit the model (3.10), and the analogous model for  $k > 3$ , using

$$\gamma_i = z_i, \quad i = 1, \dots, k - 1,$$

where  $z_k > 0$  and determine the deviance.

$k$	Approach	No Overdispersion % Excluded	Overdispersion % Excluded
2	GLM - Wald	92.26	92.26
3	GLM - Wald	99.04	97.76
	GLM - P.L.	99.05	97.91

Table 3.3: Percentage of directions excluded from the 95% confidence cone, using Wald and profile likelihood methods.

4. The proportion of simulations where the calculated deviance is less than

$$D_M + \chi^2_{0.95}(p - q),$$

ignoring the overdispersion, or

$$D_M + \hat{\phi}(p - q)F_{0.95}(p - q, n - p),$$

considering the overdispersion, is an estimate of the proportion of directions included in the 95% confidence cone for the path of steepest ascent. The sample size is chosen to make the estimated proportion sufficiently precise.

### 3.4 Bootstrap Approach

In the previous section likelihood based confidence regions were generated for both the case with overdispersion and without overdispersion. These confidence regions were based on distributional assumptions that are only approximate. For instance, Davison and Hinkley (1989) assert that it is customary to assume that the scaled deviance (equation (3.11)) is approximately  $\chi^2$  and that equation (3.12) has an  $F$ -distribution, but the later has “little theoretical justification”.

#### 3.4.1 Parametric Bootstrap

##### No overdispersion

One way to check these results is to use a parametric bootstrap, (Efron, 1979; Efron and Tibshirani, 1993). In a parametric bootstrap observations are generated from the fitted model. For each of these sets of observations, denoted by

$y_1^*, y_2^*, \dots, y_n^*$  the GLM model is fitted resulting in estimates  $\theta_1^*, \theta_2^*, \dots, \theta_n^*$  and the likelihood ratio statistics, given by

$$R(\hat{\theta}) - R(\hat{\theta}_i^*), \quad i = 1, \dots, n.$$

are calculated.

This procedure was applied to the data given in Table 2.1 using 999 bootstrap samples and the function, `boot3`, given in Appendix B.6. Figure 3.2 gives a comparison of the theoretical quantiles of  $\chi^2(2)$  and the bootstrap likelihood ratio statistics. The graph shows that the theoretical quantiles are approximately correct since the bootstrap likelihood ratio statistics fall close to a straight line up to the 95th percentile. In fact, the nominal 95% confidence region appears to be in fact a 94% confidence region. Correcting this, the appropriate contour level should be drawn at 23.045 rather than 22.599. The proportion of directions included in the confidence region is 1.02% rather than 0.95%.

Note that Gleser and Hwang (1987) have shown that the Bootstrap confidence regions used here and in the remainder of this chapter do not guarantee the required coverage. However, as pointed out by Di Cicco and Efron (1996), such confidence intervals and confidence regions give much better results than standard intervals and regions.

## Overdispersion

The parametric bootstrap above fails to account for any possible overdispersion. To overcome this a parametric bootstrap from a fitted *negative binomial distribution* could be used (see, for example, Venables and Ripley, 1999, p. 233).

### 3.4.2 Nonparametric Bootstrap

Part of the problem with using the parametric bootstrap is that the bootstrap samples may not have the statistical properties of the original data. This can be overcome by using the nonparametric bootstrap. Davison and Hinkley (1989) give three methods of generating the simulated responses: using Pearson residuals, residuals on the linear predictor scale, and deviance residuals. Below the Pearson

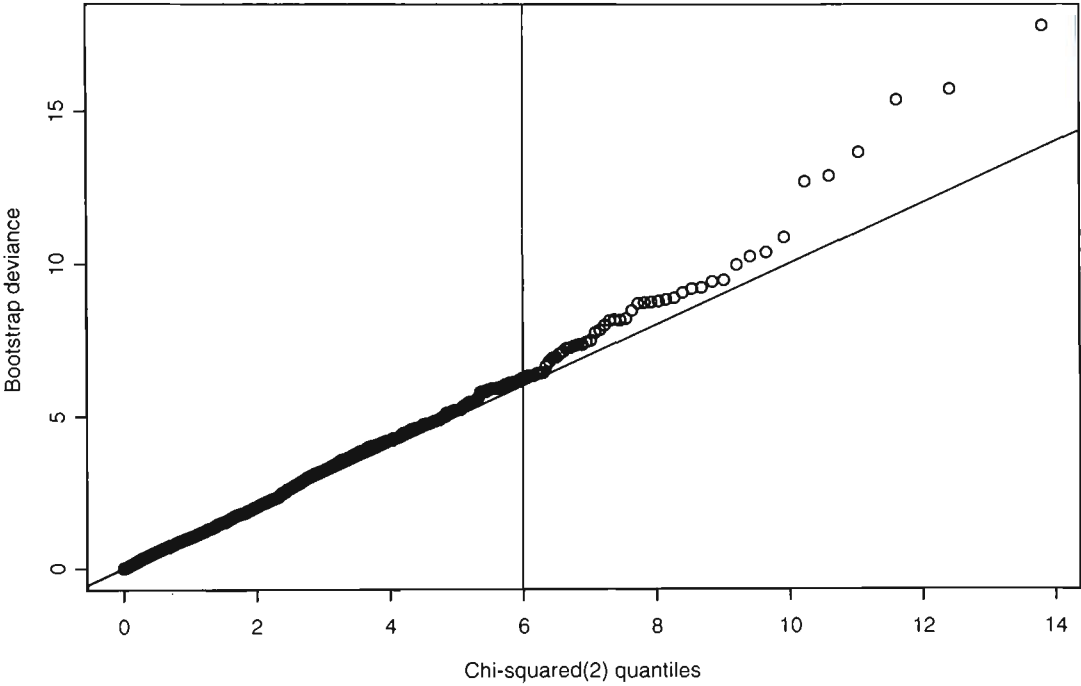


Figure 3.2: Comparison of bootstrap likelihood ratio statistics and theoretical quantiles assuming  $\chi^2(2)$  distribution. Given the linearity of the plot up to the 95th percentile there is strong evidence that the asymptotic distribution is appropriate.

residuals are used; the other residuals should give similar results. As Davison and Hinkley (1989) point out, all three methods have an obvious drawback in that they can give negative or noninteger values of  $y^*$ , even if the original data are counts. The simple fix is to round the values to the nearest appropriate value.

### Pearson Residuals

The Pearson residuals are defined to be

$$r_{p_j} = \frac{y_j - \hat{\mu}_j}{\{\hat{\phi}V(\hat{\mu}_j)(1 - h_j)\}^{\frac{1}{2}}}, \quad (3.14)$$

where  $h_j$  is the  $j$ th diagonal element of the hat matrix  $\mathbf{H}$  given by

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}. \quad (3.15)$$

Simulated responses are given by

$$y_j^* = \hat{\mu}_j + \left\{ \hat{\phi}V(\hat{\mu}_j)(1 - h_j) \right\}^{\frac{1}{2}} \varepsilon_j^*, \quad j = 1, \dots, n,$$

where

$$\varepsilon_1^*, \dots, \varepsilon_n^*$$

is a random sample with replacement of size  $n$  from the mean-adjusted, standardized Pearson residuals  $r_{P_j} - \bar{r}_P$ , with  $r_{P_j}$  defined in equation (3.14). Figure 3.3 gives a comparison of the theoretical quantiles of  $F(2, 10)$  and the bootstrap likelihood ratio statistics, using the function, `boot4`, given in Appendix B.7. The graph shows that the theoretical quantiles are not appropriate since the plot of the bootstrap likelihood ratio statistics against the  $F(2, 10)$  quantiles does not coincide with the line on the graph. The appropriate 95% quantile is 6.27 and not  $F_{0.95}(2, 10) = 4.10$ .

The likelihood confidence region was based on using the value 4.10. Examination of the bootstrap simulations shows that the confidence region is actually a 84.2% confidence region rather than a 95% one. Correcting this, Figure 3.2 was redrawn with a contour at a level of 37.44 as in Figure 3.4. The corresponding surface area is calculated using numerical integration and equals 0.0310, implying that the percentage excluded is 96.9%.

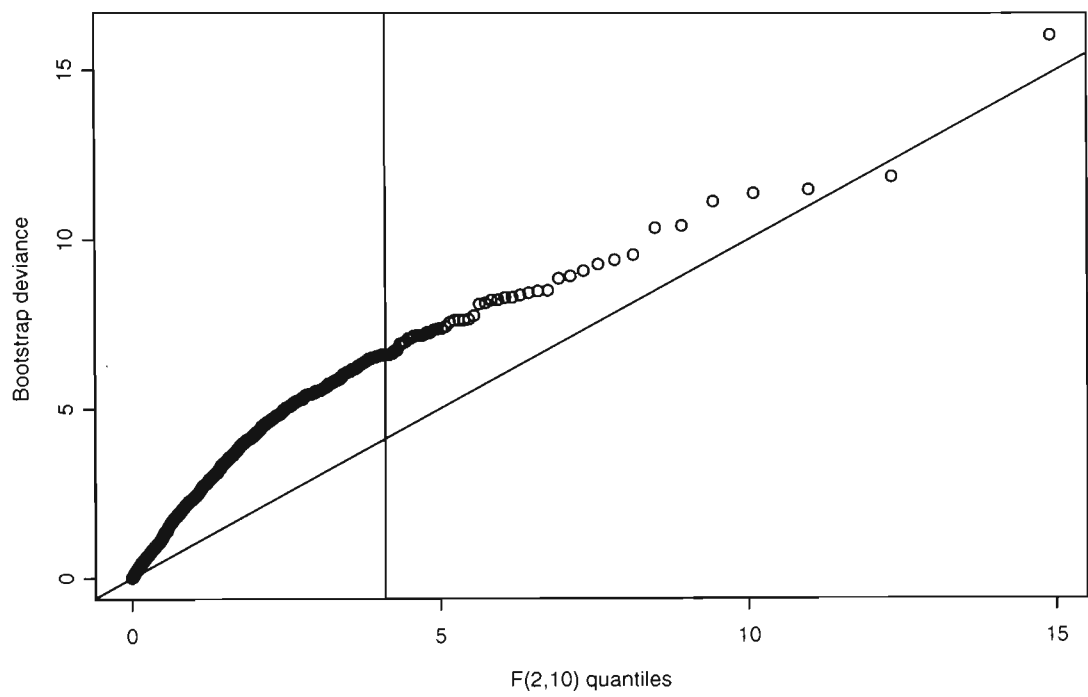


Figure 3.3: Comparison of bootstrap likelihood ratio statistics and theoretical quantiles assuming  $F(2, 10)$  distribution. Given the curvature of the plot there is strong evidence that the asymptotic distribution gives misleading results.



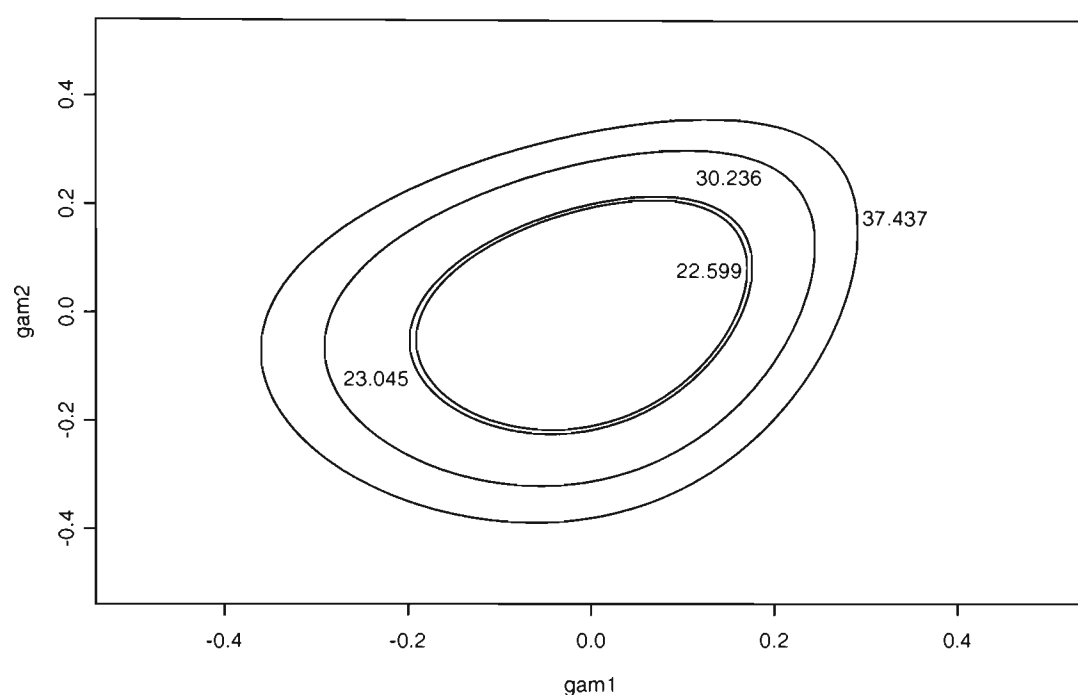


Figure 3.4: 95% profile likelihood regions for path of steepest ascent based on theoretical and bootstrap quantiles. The two inner contours give the regions ignoring overdispersion while the two outer quantiles take overdispersion into account. For each set the inner contour is based on theoretical quantiles, while the outer contour uses bootstrap quantiles to ensure the required coverage.

### 3.5 Conclusion

In Chapter 2 the fractional factorial design conducted by Hsieh and Goodwin (1986) was analysed. The experiment had a count response and a transformation was used so that the response was normally distributed. An alternative, and usually better method, is to use generalised linear models, which have been dealt with in this chapter.

Using a linear approximation with no overdispersion the approximate variance-covariance matrix for the estimated parameters can be used to determine a confidence region similar to that given in Chapter 2. Generally, these confidence regions are smaller when using the GLM model rather than using transformed responses. They can also be extended to cases where there is overdispersion.

Even better confidence regions can be found using profile likelihoods. This was applied to the data set for both the no overdispersion and overdispersion cases. When  $k = 3$ , a method using the contour module in Splus was given.

Both the Wald and likelihood based confidence region methods are based on approximate asymptotic results, which may be inaccurate. Improvements can be made using bootstrap methods, both parametric and nonparametric.

Parametric bootstrap methods were applied using Poisson generated data for when there is no overdispersion. The bootstrap results showed when the asymptotic distributions are inappropriate, and what corrections should be made.

Even better nonparametric bootstrap checks can be made. These show, for this example, that the likelihood based confidence region has a significance level of only approximately 84.2%. The bootstrap method allows for this to be corrected.

# Chapter 4

## Power Considerations

### 4.1 Introduction

In the previous chapters, it has been shown how useful it is to measure the precision of the path of steepest ascent. If the path is determined precisely enough, then experimentation to better operating conditions can be achieved by following the path. In both chapters 2 and 3, the focus has been on determining the percentage of directions included in the confidence region for the path of steepest ascent.

However, if the path has not been determined precisely enough, it might not be clear as to what to do—either follow a poorly determined path, or do further experimentation to more clearly determine the path. If a poorly determined path is followed, it may be that we are led to conditions which are far from the optimum. In this chapter, examination will centre on how many observations are needed to answer this question with sufficient precision, and how to appropriately augment the experimental design in cases where the precision is insufficient. A further application will be to determine the best set of experimental runs when the response is the proportion of defectives or the number of defects.

An important aspect of designing a first order experiment in response surface methodology is to be able to determine the path with sufficient precision. In this context, a sensitive experiment is one that is sufficiently powerful, so that

there is a high probability that the proportion of directions included in the confidence cone for the path of steepest ascent is sufficiently small, and hence further experimentation along the path is worthwhile. This can be achieved by using a sufficiently large sample size (adding to the number of observations in each treatment condition). The question of how many experiments are required is an important one for experimenters.

This chapter describes methods that can be useful for determination of the sample size needed to achieve the required percentage of directions excluded from the confidence cone for the the path of steepest ascent. Initially, for normally distributed responses, the case of homogeneous errors will be dealt with, followed by the case of heterogeneous errors.

Power considerations also are important for data which are modelled using generalised linear models. Two cases will be dealt with—when we have a Binomial response, and when the data follows a Poisson distribution.

Finally, sometimes even with the best design, we might find ourselves requiring the path to be determined more precisely after the experiment. In these situations, we want to be able to augment the experiment with extra runs. Methods for determining the number and location of these experiments will be developed and illustrated using the example that has been used in chapters 2 and 3.

## 4.2 Power Analysis for Homoscedastic Data

In section 2.2, the proportion of directions included in the confidence cone for the path of steepest ascent,  $\gamma$ , was given by

$$\begin{aligned} & \frac{1}{\sqrt{\pi(k-1)}} \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k-1}{2})} \int_{f(r_0)}^{\infty} \left(1 + \frac{t^2}{k-1}\right)^{-\frac{k}{2}} dt \\ & = 1 - T(f(r_0), k-1), \end{aligned}$$

where  $T(x, k-1)$  is the cumulative distribution function of the  $t$ -distribution with  $k-1$  degrees of freedom,

$$f(r_0) = \sqrt{k-1} \frac{r_0}{\sqrt{1-r_0^2}}, \quad (4.1)$$

and  $r_0$  is the distance from the centre of the sphere to the plane given by

$$\frac{\left\{ \sum b_i^2 - \left( \sum_{i=1}^k b_i \delta_i \right)^2 / \sum \delta_i^2 \right\}}{(k-1)s_b^2} \leq F_{1-\alpha}(k-1, \nu_b). \quad (4.2)$$

Without loss of generality, rotate the axes so that in the new co-ordinate system we have estimated regression coefficients  $c_1, \dots, c_{k-1} = 0$  and  $c_k = \sqrt{b_1^2 + b_2^2 + \dots + b_k^2}$ .

In this co-ordinate system we have

$$\begin{aligned} \frac{\sum b_i^2 - \delta_k^2 \sum \beta_i^2}{(k-1)s_b^2} &\leq F_{1-\alpha}(k-1, \nu_b) \\ \frac{(1 - \delta_k^2) \sum b_i^2}{(k-1)s_b^2} \delta_k^2 &\leq F_{1-\alpha}(k-1, \nu_b) \\ \delta_k^2 &\geq 1 - \frac{F_{1-\alpha}(k-1, \nu_b)(k-1)s_b^2}{\sum b_i^2}, \end{aligned}$$

and hence

$$r_0 = \sqrt{1 - \frac{F_{1-\alpha}(k-1, \nu_b)(k-1)s_b^2}{\sum b_i^2}}.$$

Therefore,

$$\begin{aligned} f(r_0) &= \frac{\sqrt{k-1} \sqrt{1 - \frac{F_{1-\alpha}(k-1, \nu_b)(k-1)s_b^2}{\sum b_i^2}}}{\sqrt{\frac{F_{1-\alpha}(k-1, \nu_b)(k-1)s_b^2}{\sum b_i^2}}} \\ &= \sqrt{\frac{\sum b_i^2}{F_{1-\alpha}(k-1, \nu_b)s_b^2} - (k-1)}. \end{aligned}$$

The probability of including less than a certain proportion of directions included in the confidence cone,  $p$ , is given by

$$\begin{aligned} \Pr(1 - T(f(r_0), k-1) \leq p) &= \Pr(T(f(r_0), k-1) \geq 1-p) \\ &= \Pr(f(r_0) \geq T_{1-p}(k-1)) \\ &= \Pr\left(\frac{\sum b_i^2}{s_b^2} \geq F_{1-\alpha}(k-1, \nu_b) \{[T_{1-p}(k-1)]^2 + (k-1)\}\right). \end{aligned}$$

But note

$$\frac{\sum b_i^2}{ks_b^2} \sim F(k-1, \nu_b; \lambda),$$

that is, a noncentral  $F$  with noncentrality parameter,  $\lambda$ , equal to  $\frac{N \sum \beta_i^2}{\sigma^2}$ , where  $N$  is the number of runs in the experiment.

Hence,

$$\text{Power} = 1 - F \left( \frac{F_{1-\alpha}(k-1, \nu_b) \{ [T_{1-p}(k-1)]^2 + (k-1) \}}{k}, k-1, \nu_b; \frac{N \sum \beta_i^2}{\sigma^2} \right),$$

where  $F(x, \nu_1, \nu_2; \lambda)$  is the cumulative distribution of the noncentral  $F$  distribution with  $\nu_1, \nu_2$  degrees of freedom and noncentrality parameter  $\lambda$ . Power graphs, using the Splus function `powerfn` given in Appendix C.1, have been prepared in terms of

$$\frac{\lambda}{N} = \frac{\sum \beta_i^2}{\sigma^2},$$

denoted as “ssl”, the standardised squared lengths of regression coefficients. Figure 4.1 is the power graph for  $N = 8$ . If the ratio of the squared lengths of the regression coefficients vector to the variance is 2, then we have only a 9% chance of excluding more than 98% of directions. Figure 4.2 gives the power graph for  $N = 16$ . We now have a 50% chance of excluding more than 98% of directions for the same ratio. Figure 4.3 presents the power graph for  $N = 32$ . The chance of excluding more than 98% of directions increases to 98% now.

### 4.3 Distribution of the Estimated Precision for Homoscedastic Data.

For homoscedastic data, the distribution of  $\gamma$  can be evaluated since  $\gamma$  is a function of a random variable  $X$ , where  $X$  has a noncentral  $F$  distribution.

Specifically,

$$\gamma = 1 - T \left( \sqrt{\frac{X}{F_{1-\alpha}(k-1, \nu_b)}} - (k-1), k-1 \right) = h(X), \quad X \geq (k-1)F_{1-\alpha}(k-1, \nu_b),$$

where  $X \sim F(k-1, \nu_b, \lambda)$  with  $\lambda = \frac{\sum \beta_i^2}{N}$ , and therefore

$$X = F_{1-\alpha}(k-1, \nu_b) \{ (k-1) + [T_{1-\gamma}(k-1)]^2 \}.$$

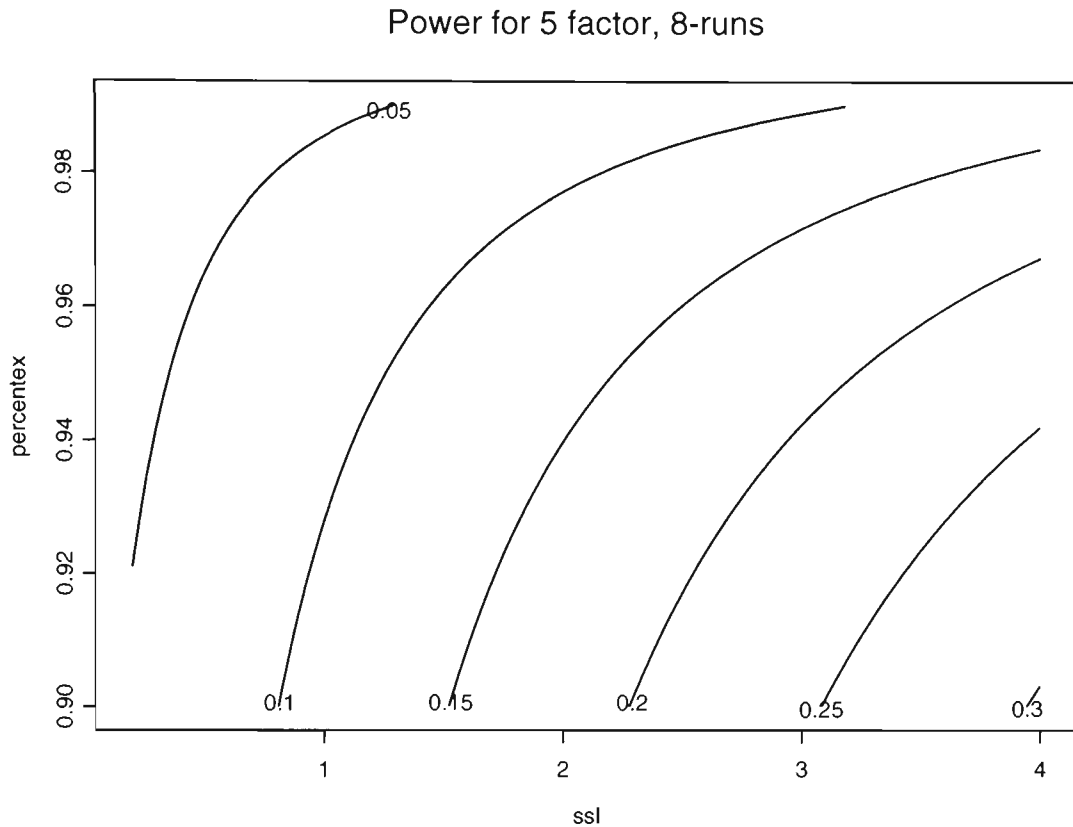


Figure 4.1: Power graph for the path of steepest ascent for 5 factors and 8 runs for the homogeneous case. The graph gives the probability (given by the contour levels) of excluding at least a certain proportion of directions (vertical axis) of steepest ascent vs. the standardised squared lengths (horizontal axis).

Since  $\gamma$  is a function of  $X$ , we can write down the density of  $\gamma$  as

$$f(\gamma) = g \left( F_{1-\alpha}(k-1) \left\{ (k+1) + [T_{1-\gamma}(k-1)]^2 \right\} \right) \left| \frac{dX}{d\gamma} \right|,$$

where  $g$  is the density of the noncentral  $F$  distribution.

Now,

$$\begin{aligned} \frac{dX}{d\gamma} &= -2F_{1-\alpha}(k-1, \nu_b) \frac{d}{d\gamma} (T_{1-\gamma}(k-1)) \\ &= 2F_{1-\alpha}(k-1) \frac{1}{T'(T_{1-\gamma}(k-1))} \\ &= \frac{-2F_{1-\alpha}(k-1, \nu_b)}{t(T_{1-\gamma}(k-1), k-1)}, \end{aligned}$$

and hence the density function of  $\gamma$  can be evaluated using software such as Splus.

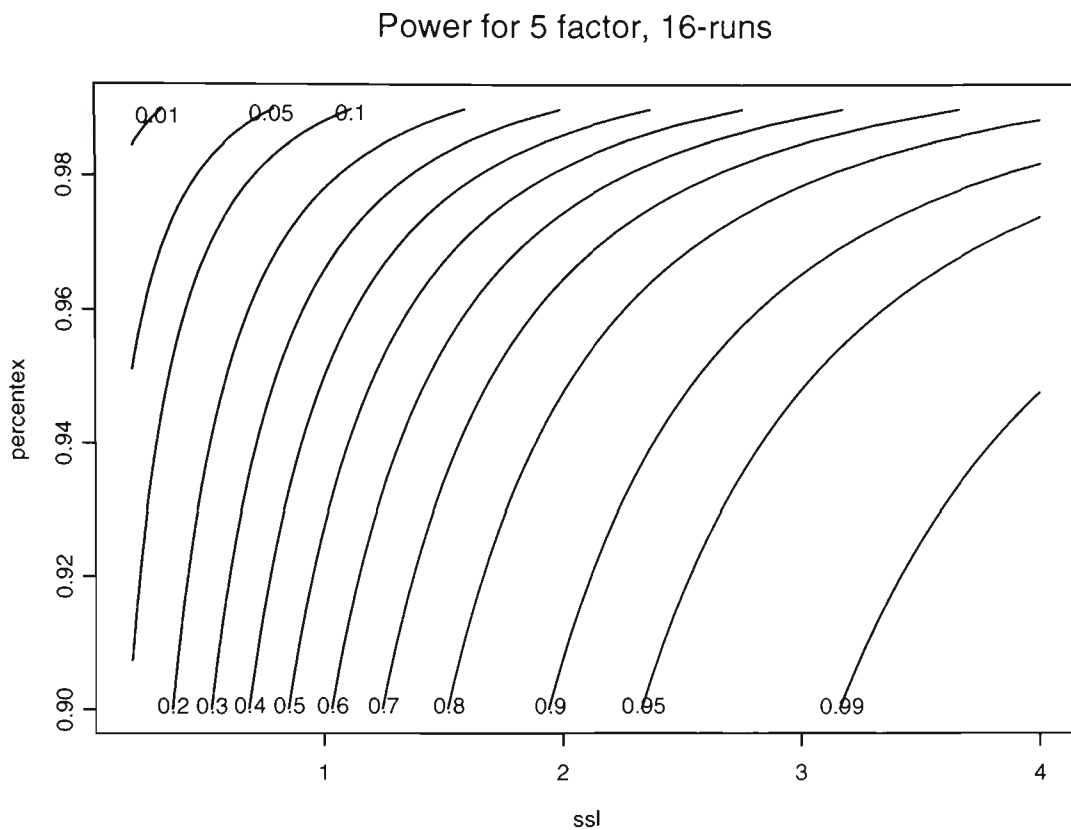


Figure 4.2: Power graph for the path of steepest ascent for 5 factors and 16 runs for the homogeneous case. The graph gives the probability (given by the contour levels) of excluding at least a certain proportion of directions (vertical axis) of steepest ascent vs. the standardised squared lengths (horizontal axis).

For example, if  $k = 2$ ,  $\nu_b = 10$ , and  $\lambda = 30$  the density function in Figure 4.4, computed using `pden` in Appendix B.5, is obtained.

Note, the density is improper since the percentage included can only be computed if

$$\frac{\sum b_i^2}{s_b^2} \geq (k-1)F_{1-\alpha}(k-1, \nu_b).$$

Approximate means and variances can be evaluated as

$$\begin{aligned} E(p) &\simeq f(E(X)) \\ V(p) &\simeq V(X) \left[ f'(E(X)) \right]^2, \end{aligned}$$

where the means and variances of  $X$  are given by (see, for example, Evans, Hasting



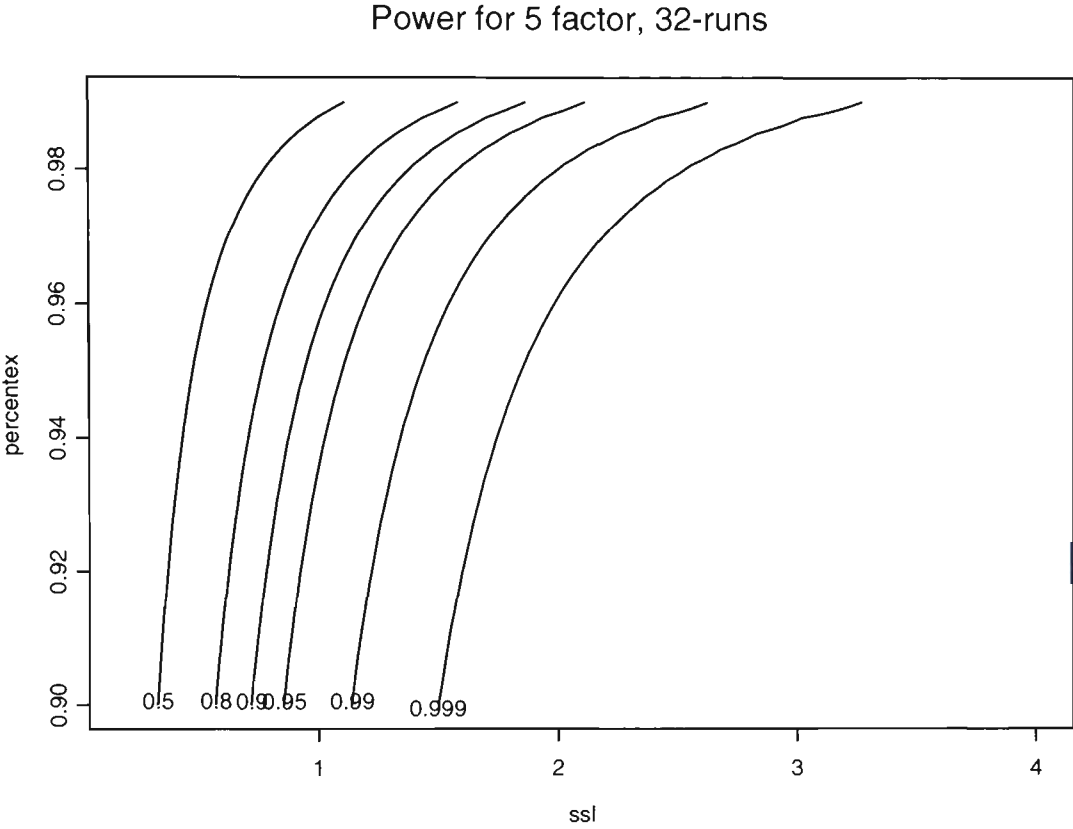


Figure 4.3: Power graph for the path of steepest ascent for 5 factors and 32 runs for the homogeneous case. The graph gives the probability (given by the contour levels) of excluding at least a certain proportion of directions (vertical axis) of steepest ascent vs. the standardised squared lengths (horizontal axis).

and Peacock, 1993, p. 73)

$$E(X) = \left[ \frac{(\lambda + (k - 1))\nu_b}{(k - 1)(\nu_b - 2)} \right],$$
$$V(X) = \frac{2\nu_b^2 [(k - 1)^2 + (2\lambda + (\nu_b - 2))(k - 1) + \lambda(\lambda + 2\nu_b - 4)]}{(k - 1)^2(\nu_b - 4)(\nu_b - 2)^2}.$$

Using `approxp`, given in Appendix C.2, for the above example, the approximate mean of  $\gamma$  is 0.117, while the approximate standard deviation is 0.043.

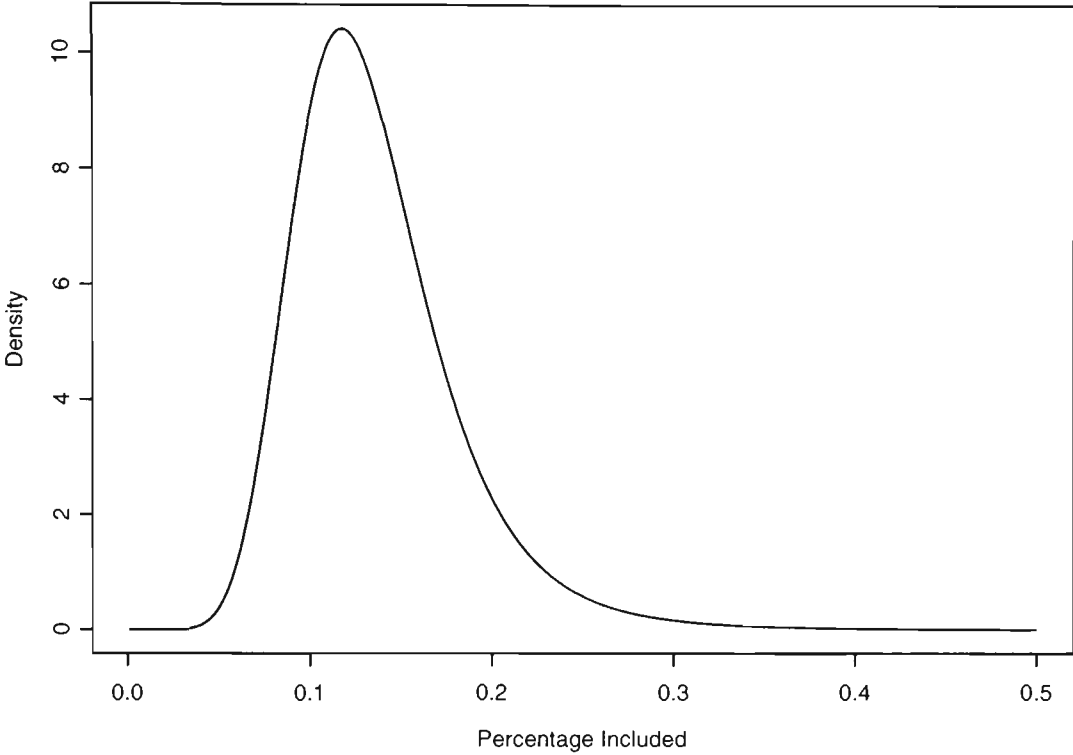


Figure 4.4: Distribution of the estimated precision for homoscedastic data for  $k = 2$ ,  $\nu_b = 10$  and  $\lambda = 30$ .

### 4.4 Power Analysis for Heteroscedastic Data

In the previous section, methods for determining the power of the precision of the path of steepest ascent were developed for the case when the errors are homogeneous. However, as shown in Chapter 2, it is often the case that the errors are heterogeneous, or the design is not orthogonal. This is also the case for GLMs, leading to heteroscedastic estimates of the regression coefficients.

If the data is homoscedastic, the expected proportion of directions included in the confidence cone for the path of steepest ascent does not depend on the direction of the path of steepest ascent. However, if the data is heteroscedastic, the expected proportion depends both on the orientation of the path relative to the experimental design and also on the orientation of the heterogeneity response

surface relative to the path of steepest ascent.

For example, assume that a  $2^2$  factorial design with levels  $\pm 1$  is to be run, where the mean response is expected to be

$$E(y) = 5 \cos \theta X_1 + 5 \sin \theta X_2,$$

where  $\theta$  describes the orientation of the mean response surface relative to the design. Similarly, assume that the responses are expected to be independent but heteroscedastic with variance given by

$$V(y) = \exp(1.1529 \cos \phi X_1 + 1.1529 \sin \phi X_1).$$

In this case  $\phi$  represents the orientation of the variance function relative to the original design. The “average variance” is 1, while the constant 1.1529 is chosen to represent a 10-fold increase in the variance over the experimental region.

For various values of  $\theta$  and  $\phi$  the expected proportion included was calculated using `heterofn`, given in Appendix C.3, and plotted in Figure 4.5. The conclusions are:

1. The minimum expected percentage included in the confidence cone occurs when the mean response surface is aligned with both the design and the variance response surface.
2. The worst case occurs when the mean response surface is aligned with the design but orthogonal to the variance response surface.

To explore this further, a simulation study of 1000 data sets was undertaken, when the model was

$$y = \beta X_1 + \varepsilon_1,$$

with

$$V(\varepsilon_1) = \exp(1.1529 X_1),$$

representing a process where the variance model is aligned to the mean model, and also with

$$V(\varepsilon_1) = \exp(1.1529 X_2),$$

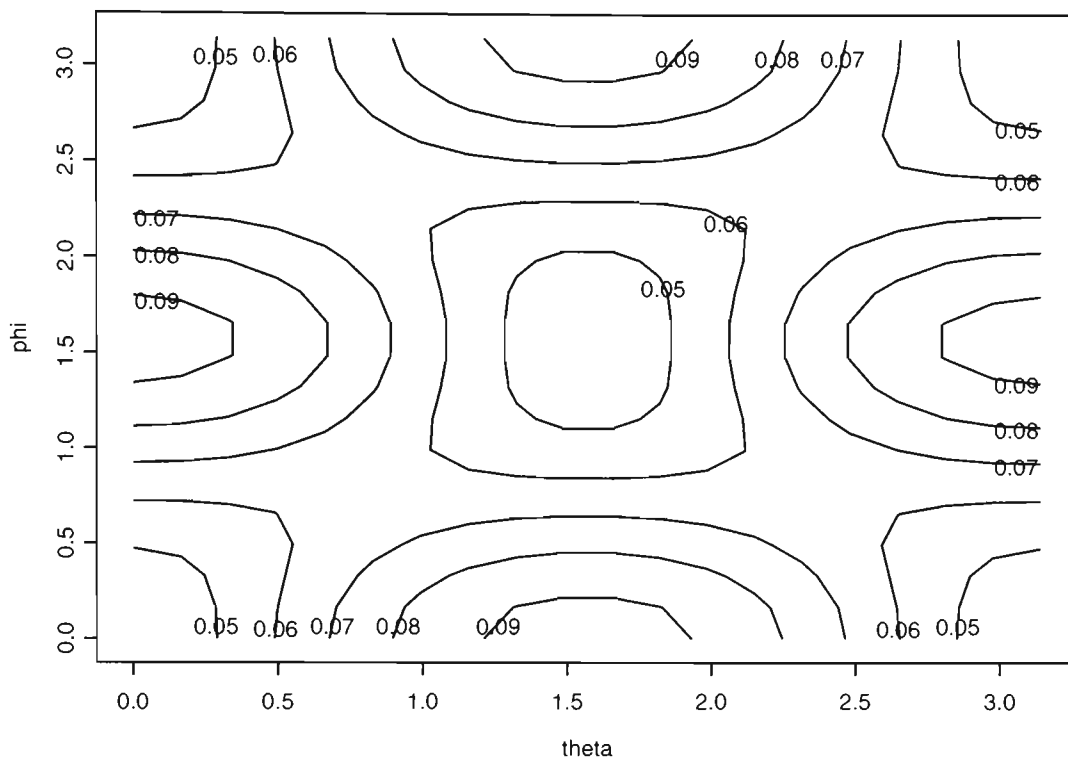


Figure 4.5: Values of the expected proportion of directions included in the 95% confidence cone for the path of steepest ascent as a function of  $\theta$ , the orientation of the mean response surface relative to the expected design, and  $\phi$ , the orientation of the variance response surface relative to the design.

representing a process where the variance model is orthogonal to the mean model. Values of  $\beta = 1, 2, 3$  and 4 were used. In each simulation the percentage of directions included in the confidence cone for the path of steepest ascent was calculated, if it was possible to calculate it. These simulations were done using `heteropowerlik`, given in Appendix C.4, and are illustrated in Figure 4.6 and summarised in Table 4.1.

Examination of the results shows that, generally, the median proportion of directions included in the 95% confidence cone for the path of steepest ascent is smaller when the variance and mean response models are aligned than when they are orthogonal. However, it also appears that the distribution has more skewness

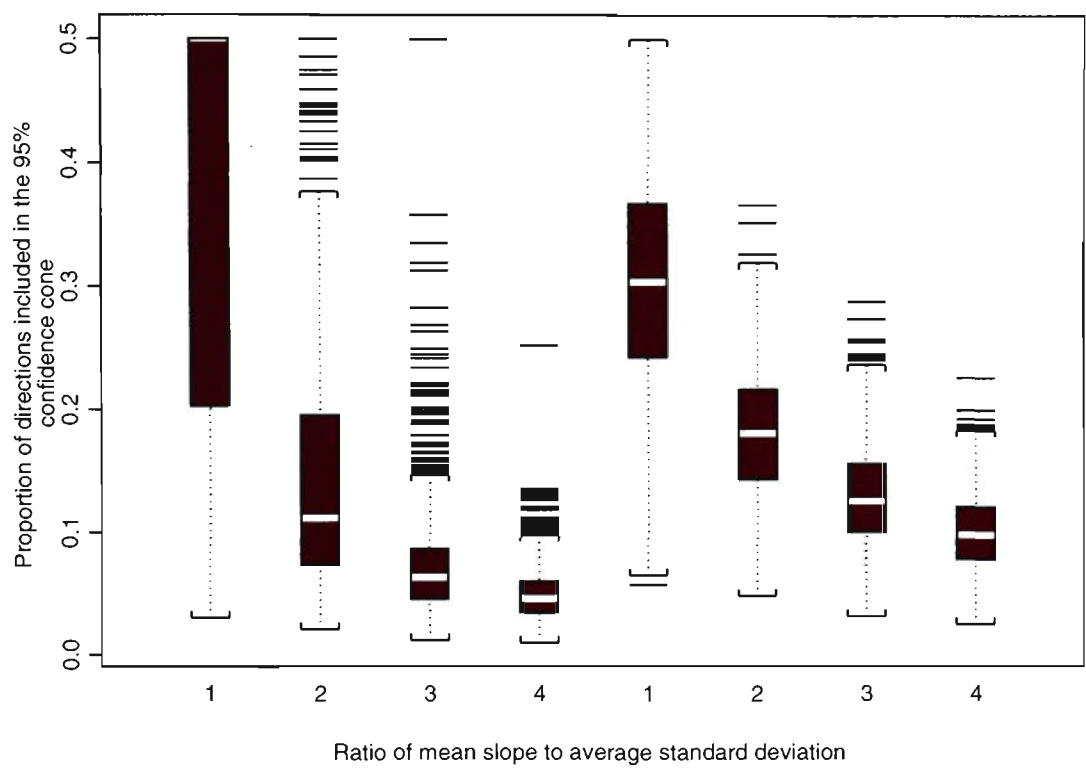


Figure 4.6: Boxplots of proportion of directions included in the 95% confidence cone for the path of steepest ascent versus the effect size, the ratio of the regression coefficients relative to the standard deviation. The four boxplots on the left correspond to the case where the variance model is aligned to the mean model, while the four boxplots on the right correspond to the case where the variance model is orthogonal to the mean model.

for aligned models than for orthogonal models.

## 4.5 Improving the Precision of the Path

When the path of steepest ascent has not been estimated precisely enough, the experimental design needs to be augmented with additional runs. In this section, a criterion will be developed when we want to add a number of points to the design in order to improve the path of steepest ascent. The most common criterion for the homogeneous case is to maximise the determinant of the  $\mathbf{X}^T\mathbf{X}$  matrix,

Effect Size	Aligned			Orthogonal		
	Mean	St.dev.	Emp.prob.	Mean	St.dev.	Emp.prob.
1	0.2036	0.1121	0.3750	0.2927	0.0814	0.9420
2	0.1283	0.0792	0.8570	0.1860	0.0562	1.0000
3	0.0757	0.0448	0.9900	0.1313	0.0401	1.0000
4	0.0503	0.0219	1.0000	0.1008	0.0301	1.0000

Table 4.1: Mean and standard deviation of proportion of directions included in the 95% confidence cone for the path of steepest ascent versus the effect size, the ratio of the regression coefficients relative to the standard deviation. The empirical probability column gives the proportion of simulations where the proportion of directions included in the confidence cone was able to be calculated.

(see, for example, Gaylor and Merrill, 1968; Dykstra, 1971). One reason for the popularity of the determinant criterion, no longer of critical importance, is that this criterion was easier to compute than other criteria. In the present situation, the determinant criterion is likely to be useful since the determinant is proportional to the volume of the confidence region, and thus, related to the surface area of the confidence cone.

When only a subset of the parameters are of interest, the criterion to be maximised is altered, following Hill and Hunter (1974), to

$$\Psi = |\mathbf{X}_{11} - \mathbf{X}_{12}\mathbf{X}_{22}^{-1}\mathbf{X}_{21}|, \tag{4.3}$$

where the  $\mathbf{X}^T\mathbf{X}$  is partitioned as

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{bmatrix},$$

and  $\mathbf{X}_{11}$  corresponds to the parameters of interest. A more general extension to account for heterogeneous errors had been given by Box (1971) who used the criterion

$$\Psi = |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^T|, \tag{4.4}$$

and

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

which equals  $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})$  for heteroscedastic data and  $(\mathbf{X}^T \Delta \mathbf{U} \Delta \mathbf{X})$  for generalised linear models.

With the example of chapters 2 and 3 we found that the percentage of directions included in the confidence region for the path of steepest ascent was 0.96%, when analysing the data using generalised linear models and Wald confidence regions, and ignoring any possible overdispersion; and 2.24% when overdispersion needs to be considered. The path of steepest ascent is determined quite precisely and in most cases it would not be necessary to increase the level of precision. However, if we desire to do so, additional experimental runs will be required.

Consider the case when one additional run is required. As in chapters 2 and 3, we assume that the high level of  $G$  is to be chosen so that  $BG_- \equiv 0$  and  $G \equiv 1$ . The criterion  $\Psi$  in equation (4.4) was used where  $\mathbf{A} = \mathbf{X}^T \Delta \mathbf{U} \Delta \mathbf{X}$  and  $\mathbf{A}_{11}$  consists of the rows and columns of  $\mathbf{A}$  corresponding to  $D$ ,  $F$ , and  $BG_+$ .

The above criterion is a function of three parameters, the level of  $D$ , the level of  $F$  and the level of  $B \equiv BG_+$ . Direct maximisation of  $\Psi$ , with all factors over the range  $[-1, 1]$ , using `augdet` given in Appendix C.5 and the commands given in Appendix C.7, gave  $D = -1$ ,  $F = -1$ , and  $B = 1$ . At these values the expected value of the percentage included is 0.026%, a considerable improvement.

An alternative is to directly minimise the expected percentage of directions included in the confidence region. The best set of conditions, found using `augfun` and the commands given in Appendix C.7, is  $D = -1$ ,  $F = -1$ , and  $B = 0.945$  with the same expectation of 0.026%. There is no advantage in directly minimising the percentage of directions included in the confidence region. Simulations show that by randomly choosing a point, there is only approximately a 0.7% chance of the expectation being less than 0.03%, and approximately a 35% chance of the expectation being greater than 0.039%.

When overdispersion is considered, the best set of conditions to minimise  $\Psi$ , given by  $D = -1$ ,  $F = -1$ , and  $B = 1$ , results in an expected percentage of directions included in the confidence region for the path of steepest ascent of 0.105%, and these sets of conditions correspond to the smallest possible value

$m$	Best Set of Runs			No Overdispersion	Overdispersion
	$D$	$F$	$B_{G+}$	Predicted % Included	Predicted % Included
1	-1	-1	1	0.026	0.105
2	1	-1	-1	0.020	0.066
	-1	0.9	-1		
3	-1	-1	1	0.016	0.043
	-1	0.87	-1		
	-1	-1	1		

Table 4.2: Best sets of augmenting runs and predicted percentage of directions included in the 95% confidence cone using D-optimality.

$m$	Best Set of Runs			No Overdispersion
	$D$	$F$	$B_{G+}$	Predicted % Included
1	-1	-1	0.99	0.026
2	-1	-1	0.95	0.020
	-1	0.64	-1	
3	-1	0.62	-1	0.016
	-1	-1	1	
	-1	-1	1	

Table 4.3: Best sets of augmenting runs and predicted percentage included in the 95% confidence cone using direct minimisation, assuming no overdispersion.

of the expected percentage. Hence, for one augmenting run, the D-optimal augmenting design is just as good as directly minimising the percentage of directions included in the confidence cone.

If further improvements are required, then extra experiments can be run. Table 4.2 gives the minimum percentage included by adding one to three runs when using the criterion method. Similarly, Table 4.3 gives the results obtained from using direct minimisation of the percentage included ignoring any possible overdispersion; while Table 4.4 gives the results of directly minimising the percentage included considering overdispersion. Again, for this example, the D-optimal design is just as good as directly minimising the percentage included in the confidence cone.



$m$	Best Set of Runs			Overdispersion
	$D$	$F$	$B_{G+}$	Predicted % Included
1	-1	-1	1	0.105
2	-1	-1	0.95	0.066
	1	-1	-1	
3	-1	-1	0.93	0.043
	-1	0.62	-1	
	1	-1	-1	

Table 4.4: Best sets of augmenting runs and predicted percentage included in the 95% confidence cone using direct minimisation, assuming overdispersion.

## 4.6 Best Designs to Estimate the Path for GLMs

### 4.6.1 Introduction

In the previous section, the best experimental points to improve the precision of the path of steepest ascent were found. These results are useful, as it would be preferable to refine the estimated path rather than follow one which was poorly defined. Complimentary results are given in this section where the design giving the maximum precision of the path of steepest ascent is found, for both Poisson and Binomial data. It should be noted that these results are probably less useful than those in the previous section since, in practice, many other design criteria would be just as important or more important than the precision of the path of steepest ascent, such as estimation of the regression parameters, check for lack of fit, design balance, etc.

### 4.6.2 Poisson Data

For Poisson data, the asymptotic variance-covariance matrix of the parameter estimates is given by equation (3.1) with  $r(\phi) = 1$ . A D-optimal design is one where  $\Psi$  given in equation (4.4) is maximised with

$$\mathbf{A} = \mathbf{X}^T \Delta \mathbf{U} \Delta \mathbf{X} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

Run	$X_1$	$X_2$	Run	$X_1$	$X_2$
1	-1	1	1	-1	1
2	1	-1	2	-1	1
3	1	-1	3	1	-1
4	1	1	4	1	0.76

Table 4.5: D-optimal and minimum proportion designs for Poisson experiments with 4 runs constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$  for the example considered in section 4.6.1. Right Panel: Optimal design to minimise the expected proportion of directions included in the 95% confidence cone(See Figure 4.7).

(following Box (1971), to account for the heterogeneous errors).

Take, as an example, an experiment where we have two factors,  $X_1$  and  $X_2$ , constrained to be between  $-1$  and  $1$ , and a factorial experiment with  $X_1$  and  $X_2$  set at  $\pm 1$  is conducted, and the response is assumed to follow a Poisson distribution with mean given by

$$\lambda = 15 \exp \left( \frac{1}{\sqrt{5}} X_1 + \frac{2}{\sqrt{5}} X_2 \right),$$

and hence the expected response at the design origin is 15, the expected size of the regression coefficients vector is 1, and the  $X_2$  coefficient is twice that of the  $X_1$  coefficient. The expected proportion of directions included in the confidence cone for the path of steepest ascent is 7.70%.

The D-optimal design for this situation can be obtained by direct maximisation of  $\Psi$  given in equation (4.4). The D-optimal design, found using `poisfund` given in Appendix C.8 and `poissim` given in Appendix C.10, is given in the left hand panel of Table 4.5. With this design, the expected value included in the 95% confidence cone for the path of steepest ascent, found using `poisfunp` given in Appendix C.9 , is 7.01%, slightly better than using a simple factorial.

Rather than maximising  $\Psi$ , we could minimise the percent included in the confidence cone. The result, found using `poisfunp` given in Appendix C.9 and `poissim` given in Appendix C.10, is given in the right hand panel of Table 4.5. With this design, the expected value included in the 95% confidence cone for the

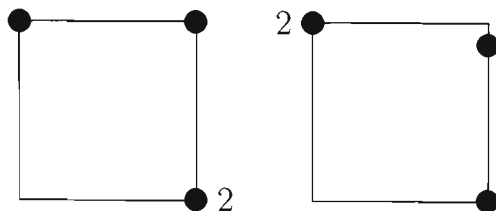


Figure 4.7: D-optimal and minimum proportion designs for Poisson experiments with 4 runs constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$  for the example considered in section 4.6.1. Right Panel: Optimal design to minimise the expected proportion of directions included in the 95% confidence cone.

Factor	Range
$\lambda_0$	$5 \longleftrightarrow 25$
$\sqrt{\beta_1^2 + \beta_2^2}$	$0.5 \longleftrightarrow 2.5$
$\tan^{-1}\left(\frac{\beta_1}{\beta_2}\right)$	$0 \longleftrightarrow \frac{\pi}{4}$

Table 4.6: Ranges for the factors for the Poisson simulation study.

path of steepest ascent is 6.62%, slightly better than using the D-optimal design. The two designs are illustrated in Figure 4.7.

For this case, the D-optimal design is almost as good as the design with the minimum expected percentage included in the confidence cone. To see how general this result is, a simulation study was undertaken. In the study, the response follows a Poisson distribution with mean given by

$$\lambda = \lambda_0 \exp(\beta_1 X_1 + \beta_2 X_2),$$

where the factors studied were:

- $\lambda_0$ , the expected mean at the design origin;
- $\sqrt{\beta_1^2 + \beta_2^2}$ , the “size” of the regression coefficients vector; and
- $\tan^{-1}\left(\frac{\beta_1}{\beta_2}\right)$ , the “orientation” of the regression coefficients vector.

A 100 run orthogonal array (Sloane, 2005) was used, and each factor was varied over the ranges given in Table 4.6 with ten equally spaced levels per factor. For

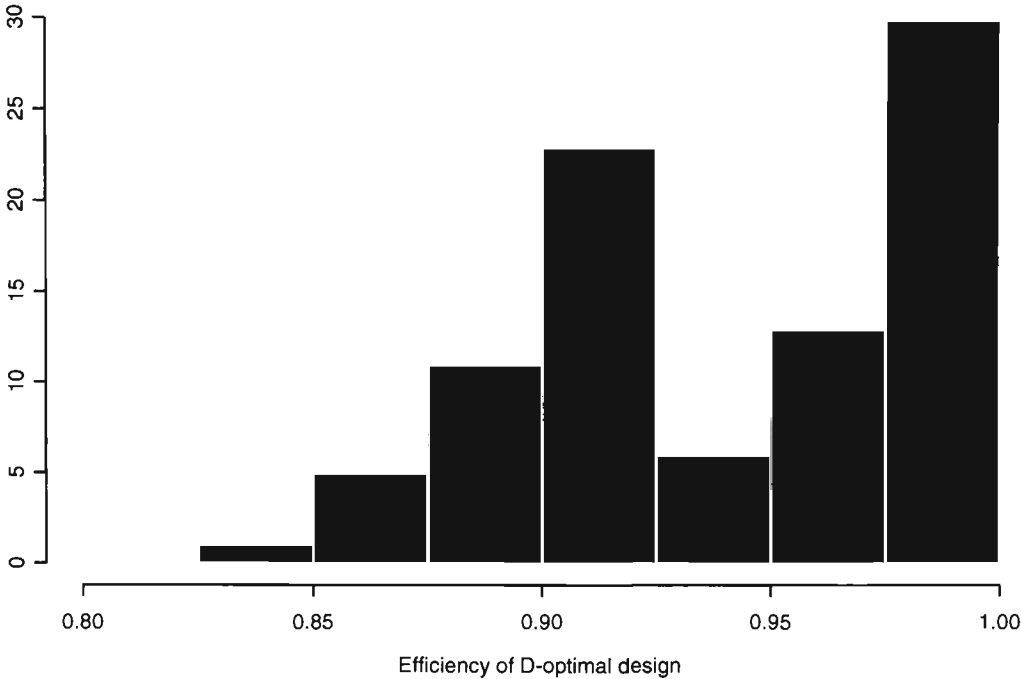


Figure 4.8: Histogram of efficiencies of the D-optimal design relative to the minimum proportion design for the Poisson simulation study using a 100 run orthogonal array. The efficiency is defined to be the ratio of the expected proportion of directions included in the 95% confidence cone for the path of steepest ascent for the minimum proportion design relative to the D-optimal design.

each run, the D-optimal design was determined and also the design that gave the minimum expected proportion of directions included in the 95% confidence cone for the path of steepest ascent (or equivalently, descent, since a minimum response is usually desired with a Poisson response — note, this is the case even though the variance increases with the mean with Poisson data). For both designs, the values of  $X_1$  and  $X_2$  were constrained to be between  $-1$  and  $1$ , and the expected proportion of directions included in the confidence cone was recorded. Finally, the ratio of expected proportion for the minimum proportion design relative to the D-optimal design was calculated as a measure of efficiency of the D-optimal design to estimate the path of steepest ascent. A histogram of the results, calculated

Run	$X_1$	$X_2$	Run	$X_1$	$X_2$
1	-1	1	1	-1	1
2	-1	1	2	-1	1
3	1	-1	3	1	-0.77
4	1	-1	4	1	-0.77
5	1	1	5	1	1

Table 4.7: D-optimal and minimum proportion designs for Poisson experiments with 5 runs constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$  for the example considered in section 4.6.1. Right Panel: Optimal design to minimise the expected proportion of directions included in the 95% confidence cone (See Figure 4.9).

using `bcbfn3` given in Appendix C.11, is given in Figure 4.8. The histogram shows that 75% of the simulations have an efficiency of over 90%, indicating that over the range covered by the factors in the simulation study that there is very little lost in using the D-optimal design.

Irrespective of whether D-optimal or minimum proportion designs are used, if further improvements are required, then larger designs can be employed. Table 4.7 shows optimal designs for 5-run experiments. These experiments are illustrated in Figure 4.9. The expected percentage included now is 5.65%, when using the D-optimal design; and 5.63%, when using direct minimisation of the percentage included in the confidence cone. Similarly, Table 4.8 and Figure 4.10 show optimal designs for 6-run experiments. This time, the expected percentage included, when using the criterion method, is 5.30%; and using the minimum percentage included, the result is 5.08%. We can see that direct minimisation of the percentage included gives marginally better results than direct maximisation of the criterion.

4.6.3 Binomial Data

Similar considerations apply for Binomial data as for Poisson data. Take, as an example, an experiment where we have two factors constrained to be between

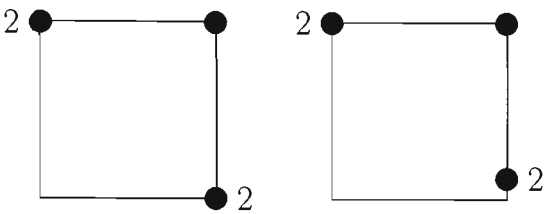


Figure 4.9: D-optimal and minimum proportion designs for Poisson experiments with 5 runs constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$  for the example considered in section 4.6.1. Right Panel: Optimal design to minimise the expected proportion of directions included in the 95% confidence cone.

Run	$X_1$	$X_2$	Run	$X_1$	$X_2$
1	1	-1	1	-1	1
2	1	-1	2	-1	1
3	1	-1	3	-1	1
4	-1	1	4	1	-0.75
5	-1	1	5	1	-0.75
6	1	1	6	1	0.97

Table 4.8: D-optimal and minimum proportion designs for Poisson experiments with 6 runs constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$  for the example considered in section 4.6.1. Right Panel: Optimal design to minimise the expected proportion of directions included in the 95% confidence cone (See Figure 4.10).

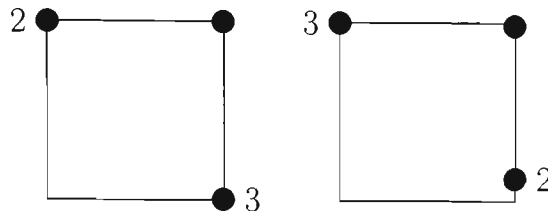


Figure 4.10: D-optimal and minimum proportion designs for Poisson experiments with 6 runs constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$  for the example considered in section 4.6.1. Right Panel: Optimal design to minimise the expected proportion of directions included in the 95% confidence cone.

$-1$  and  $1$  with a response which we assume follows a Binomial distribution. The probability of success is expected to be

$$p = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}, \quad (4.5)$$

where

$$\beta_0 = \log\left(\frac{0.6}{1-0.6}\right), \quad \beta_1 = \sqrt{5}, \quad \text{and} \quad \beta_2 = 2\sqrt{5},$$

and hence the expected probability at the design origin is  $0.6$ , the expected size of the regression coefficients vector is  $5$ , and the  $X_2$  coefficient is twice that of the  $X_1$  coefficient. If a factorial experiment is conducted, with  $X_1$  and  $X_2$  set at  $\pm 1$  and with a sample size of  $50$  at each design point, then the proportion included in the confidence cone for the path of steepest ascent is  $14.02\%$ .

The D-optimal design for this situation can be obtained by direct maximisation of  $\Psi$  given in equation (4.4). The D-optimal design, found using `binfund` given in Appendix C.12 and `binsim` given in Appendix C.14, is given in the left hand panel of Table 4.9. With this design, the expected value included in the 95% confidence cone for the path of steepest ascent, found using `binomfunp` given in Appendix C.13, is  $2.13\%$ , considerably better than using a simple factorial. Note, the D-optimal design consists of two points on the  $ED_{17.6}$  contour and two points on  $ED_{82.4}$ . The contour levels for the D-Optimal design are slightly different to those of Jia and Myers (1998) and Myers (1999), since in this section the focus

Run	$X_1$	$X_2$	Run	$X_1$	$X_2$
1	-1	0.06	1	-1	0.23
2	-1	0.75	2	-1	0.58
3	1	-0.94	3	1	-0.76
4	1	-0.25	4	1	-0.41

Table 4.9: D-optimal and minimum proportion designs for Binomial experiments with 4 runs of 50 trials constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$ . Right Panel: Optimal design to minimise the expected percentage included (See Figure 4.11).

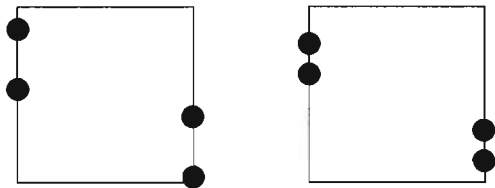


Figure 4.11: D-optimal and minimum proportion designs for Binomial experiments with 4 runs of 50 trials constrained between  $-1$  and  $1$ . Left Panel: D-optimal design for the regression coefficients of  $X_1$  and  $X_2$ . Right Panel: Optimal design to minimise the expected percentage included.



Factor	Range
$p_0$	$0.2 \longleftrightarrow 0.8$
$\sqrt{\beta_1^2 + \beta_2^2}$	$1 \longleftrightarrow 10$
$\tan^{-1}\left(\frac{\beta_1}{\beta_2}\right)$	$0 \longleftrightarrow \frac{\pi}{4}$
$n$	$20 \longleftrightarrow 80$

Table 4.10: Ranges for the factors for the Binomial simulation study.

has been on the regression coefficients  $\beta_1$  and  $\beta_2$  of equation (4.5), whereas Jia and Myers (1998) and Myers (1999) gave D-optimal designs for  $\beta_1$ ,  $\beta_2$  and the intercept.

Rather than maximising  $\Psi$ , we could minimise the percent included in the confidence cone. The result, found using `binomfunp` given in Appendix C.13 and `binsim` given in Appendix C.14, is given in the right hand panel of Table 4.9. With this design, the expected value included in the 95% confidence cone for the path of steepest ascent is 1.84%, showing that the D-optimal design has an efficiency of 86.4% relative to the minimum proportion design.

To examine the efficiency of the D-optimal design relative to the minimum proportion design over a wider range of conditions, a similar simulation study to that done for the Poisson case was conducted. In the study, the response follows a Binomial distribution where the probability of success is given by

$$p = \frac{\exp\left(\log\left(\frac{p_0}{1-p_0}\right) + \beta_1 X_1 + \beta_2 X_2\right)}{1 + \exp\left(\log\left(\frac{p_0}{1-p_0}\right) + \beta_1 X_1 + \beta_2 X_2\right)},$$

and the factors studied were:

- $p_0$ , the expected probability of success at the design origin;
- $\sqrt{\beta_1^2 + \beta_2^2}$ , the “size” of the regression coefficients vector;
- $\tan^{-1}\left(\frac{\beta_1}{\beta_2}\right)$ , the “orientation” of the regression coefficients vector; and
- $n$ , the sample size for each binomial trial.

Again, a 100 run orthogonal array was used, and each of the four factors was varied over the ranges given in Table 4.10, with ten equally spaced levels per factor. A histogram of the efficiencies of the minimum proportion design relative to

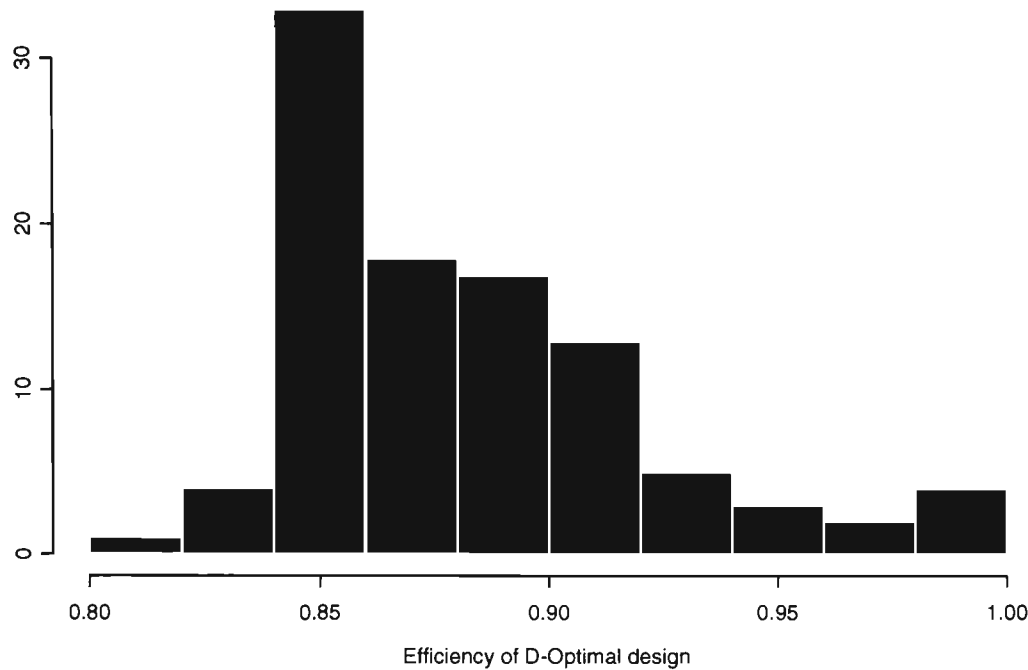


Figure 4.12: Histogram of efficiencies of the D-optimal design relative to the minimum proportion design for the Binomial simulation study using a 100 run orthogonal array. The efficiency is defined to be the ratio of the expected proportion of directions included in the 95% confidence cone for the path of steepest ascent for the minimum proportion design relative to the D-optimal design.

the D-optimal design, calculated using `bcbn4` given in Appendix C.15, is given in Figure 4.12. The histogram shows that 83% of the simulations have an efficiency of over 85%, indicating that the D-optimal design provides relatively good results over the range covered by the factors in the simulation study.

## 4.7 Conclusion

In this chapter methods for determining how precisely the path will be measured prior to the experiment have been determined, and ways to improve the experiment have been given in cases where the path has not been determined precisely

enough.

For homoscedastic, normally distributed data, the power can be determined using the non-central  $F$ -distribution and depends on the sample size and the expected size of the regression coefficient vector relative to experimental variance. Tradeoffs between the desired sample size and the expected power can be made if some guesses of these quantities can be made.

When the data is expected to be heteroscedastic, there are some directions that will be determined more precisely than other directions. The sample size required thus depends on the direction of the path expected, unlike the homogeneous case, where the power is the same, irrespective of the direction of the path. To be conservative, the least precise path direction could be used in the design of the experiment.

When the path has not been determined precisely enough, it is possible to easily design experiments that will improve the precision with the minimal number of additional experimental runs.

For Binomial and Poisson data, it is shown how to determine the best set of experiments in order to determine the path with maximum precision. These results are compared to methods for determining D-optimal experimental designs. In many cases the D-optimal designs are almost as efficient as the minimum proportion designs.

# Chapter 5

## Constraints and Second Order Models

### 5.1 Introduction

Previous chapters of the thesis have been concerned with determining the precision of the path of steepest ascent when the predictor is linear. The original method developed by Box (1955) and Box and Draper (1987, pp. 190–194) has been extended to cover heterogenous errors and non-orthogonal designs. An important aspect of this work is that the precision of the path can be conveniently determined when the data can be analysed using GLMs.

Quite often, besides a primary response of interest, there are one or more secondary responses that also need to be considered. It is not enough to choose operating conditions which maximise the primary response without regard to these secondary responses. Often, there are restrictions on the experimental variables, or outside an operating range undesirable effects occur, or such conditions are simply impossible. In addition, often the response will not be a function of a linear predictor, and some generalisation is required to make the predictor quadratic in the experimental variables.

These issues are dealt with in this chapter. Specifically, linear constraints are examined, and methods for determining the precision of the path subject to the

constraint are given.

When a response is a quadratic function of the experimental variables, canonical or ridge analysis is appropriate. The path coordinates can be viewed as the conditions giving rise to the maximum response subject to the constraint that the solution lies on the sphere of a particular radius, and by varying the radius the ridge analysis solution is given. The ridge analysis solution describes the path of steepest ascent and results in a curve rather than a line, as in the linear case.

## 5.2 Linear Input Constraints

Sometimes a complication arises, when seeking the optimum, if a constraint is included. The path of steepest ascent may lead to an optimal operating condition,  $\mathbf{x}_S$ , which falls in a region of inoperability, since it violates the constraint. In these cases, a modified direction of steepest ascent is found, which allows an optimal solution without violating the constraint (Box and Draper, 1987, pp. 194–199).

If the region of operability is bounded by a plane

$$a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k, \quad (5.1)$$

then the modified path can be viewed as a projection onto the constraint plane.

The vector  $\mathbf{a} = (a_1, \dots, a_k)^T$  is orthogonal to the plane given by equation (5.1). To project the path onto the line of which  $\mathbf{a}$  is an element, the projection matrix is given by

$$\mathbf{P} = \mathbf{a}(\mathbf{a}^T \mathbf{a})^{-1} \mathbf{a}^T,$$

and hence, to project the path onto the constraint plane the projection matrix

$$\mathbf{P}_1 = (\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{a}(\mathbf{a}^T \mathbf{a})^{-1} \mathbf{a}^T$$

is used. The projected path of steepest ascent onto the constraint plane is given by

$$\mathbf{c}_0 + \mathbf{P}_1 \mathbf{b}_1,$$

where  $\mathbf{b}_1$  is the regression coefficient vector defining the path and  $\mathbf{c}_0$  gives the co-ordinates of the intersection of the unconstrained path of steepest ascent and

the constraint plane. As shown below, the methods of Chapter 2 can be used to determine the precision of the constrained path of steepest ascent.

### 5.2.1 An Example

In the example given by Box and Draper (1987, p. 196), the activity of a certain chemical mixture depends on the proportion of three ingredients A ( $x_1$ ), B ( $x_2$ ) and C ( $x_3$ ). The relationship is estimated by

$$\hat{y} = 12.2 + 3.8x_1 + 2.6x_2 + 1.9x_3,$$

and hence the vector representing the path of steepest ascent is  $\mathbf{b}_1 = (3.8, 2.6, 1.9)^T$ .

To find the confidence region for the path the method given in Chapter 2 is used. Since there are three variables ( $k=3$ ), the confidence region takes the form of a cone. The projection of the cap of the cone onto the line orthogonal to the path of steepest ascent has semi-axes given by

$$a_1 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_2}}, \quad a_2 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_3}},$$

where  $\lambda_1 = 47.9317$ ,  $\lambda_2 = \lambda_3 = -1539.9083$  are the eigenvalues of

$$\mathbf{H} = \mathbf{G}\mathbf{b}_1\mathbf{b}_1^T\mathbf{G} - c\mathbf{G}, \quad (5.2)$$

and

$$c = \mathbf{b}_1^T\mathbf{G}\mathbf{b}_1 - \chi^2_{0.95}(2), \quad (5.3)$$

with the assumption that  $\sigma^2 = 1$ , say, from previous experimentation.

Since  $\lambda_2 = \lambda_3$ , then  $a_1 = a_2 = 0.1742$ , and the proportion included in the confidence cone before the constraint is encountered is thus calculated, using the commands in Appendix D.1, as

$$1 - T\left(\frac{\sqrt{1 - 0.1742^2}}{0.1742}, 2\right) = .0076.$$

### Confidence Region for the Projected Path

In the above example, a constraint is imposed on the experimental variables. The linear constraint is

$$-30 + 5x_1 + 10x_2 + 10x_3 = 0.$$

If we let

$$\mathbf{P} = \mathbf{a} (\mathbf{a}^T \mathbf{a})^{-1} \mathbf{a}^T,$$

then the projection matrix is

$$\mathbf{P}_1 = \mathbf{I} - \mathbf{P}, \quad (5.4)$$

and the projected path of steepest ascent is given by

$$\mathbf{P}_1 \mathbf{b}_1. \quad (5.5)$$

For the example,  $\mathbf{a}^T = (5, 10, 10)$ , and hence

$$\mathbf{P}_1 = \frac{1}{9} \begin{pmatrix} 8 & -2 & -2 \\ -2 & 5 & -4 \\ -2 & -4 & 5 \end{pmatrix}.$$

Therefore,

$$\mathbf{P}_1 \mathbf{b}_1 = \begin{pmatrix} 2.3778 \\ -0.2444 \\ -0.9444 \end{pmatrix},$$

agreeing with the results of Box and Draper (1987, p. 196).

Since the original path of steepest ascent is projected onto the constraint plane, it is convenient to use a Householder transformation to find a new co-ordinate system. The Householder transformation,  $\mathbf{H}_u$ , is chosen so that the vector orthogonal to the constraint plane,  $\mathbf{a}$ , is taken to  $\|\mathbf{a}\|\mathbf{e}_3 = (0, 0, \|\mathbf{a}\|)^T$ .  $\mathbf{H}_u$  is given by  $\mathbf{I} - 2\mathbf{u}\mathbf{u}^T$ , where the unit vector is

$$\mathbf{u} = \frac{\mathbf{a} - \|\mathbf{a}\|\mathbf{e}_3}{\|\mathbf{a} - \|\mathbf{a}\|\mathbf{e}_3\|}.$$

For the current example, the path in the new co-ordinate system is given by

$$\mathbf{b}_n = \mathbf{H}_u \mathbf{P}_1 \mathbf{b} = (1.433, -2.133, 0)^T.$$

The  $\mathbf{G}$  and  $\mathbf{H}$  matrices, defined in equations (2.12) and (2.14) respectively, are given by

$$\mathbf{G} = 8\mathbf{H}_u\mathbf{P}_1\mathbf{P}_1\mathbf{H}_u$$

and

$$\mathbf{H} = \mathbf{G}\mathbf{b}_n\mathbf{b}_n^T\mathbf{G} - (\mathbf{b}_n^T\mathbf{G}\mathbf{b}_n - \chi_{0.95}^2(1))\mathbf{G}.$$

The eigenvalues of  $\mathbf{H}$  are  $\lambda_1 = 30.7317$  and  $\lambda_2 = -392.0239$ . Therefore,

$$a_1 = \sqrt{\frac{\lambda_1}{\lambda_1 - \lambda_2}} = 0.2696,$$

and the proportion included in the confidence cone, calculated using the commands in Appendix D.1, is

$$1 - T\left(\frac{\sqrt{1 - a_1^2}}{a_1}, 1\right) = 0.0869.$$

Box and Draper (1987, p. 198) suggest that the modified path should only be viewed as one worthy of further exploration because:

1. the original path of steepest ascent is often subject to fairly large errors,
2. the linear approximation is probably inaccurate at points not close to the region of the design, and
3. in some cases the constraint is not known exactly *a priori*, but must be estimated.

The same considerations apply to the precision of the modified path, as expressed by the proportion of directions included in the confidence cone.

It should be noted that constraints on both the inputs and outputs are studied by Angün et al. (2002).

### 5.3 Second Order Models

As discussed in Chapter 1, response surface methodology is a collection of techniques that can be adapted to the needs of the experimenter. Initially, first order



designs and models, including the calculation of the path of steepest ascent, work well. Eventually, however, it is necessary to use second order models, and the corresponding designs that allow such models to be fitted, in order to provide information about the stationary points of experimental systems.

Once a second order model is fitted to the data, canonical analysis can be used to better understand the nature of the system that is being investigated. Often the fitted model will reveal a ridge system, which may be stationary or rising. Exploration and exploitation of these ridge systems is an important, but probably neglected, part of response surface methodology.

Canonical analysis can be complemented by ridge analysis, due to Hoerl (1959). In ridge analysis, the path of steepest ascent for second order models is calculated. For second order models, the path is a curve and not a line, as it is for first order models. A good discussion of the relationship between ridge analysis and canonical analysis was given by Box and Liu (1999). They discussed an example where the surface was a saddle giving two paths of steepest ascent and two paths of steepest descent. In the example, where the centre of the contour system was close to the design origin, the paths converged very rapidly onto the axes of the canonical variables established using canonical analysis.

In the remainder of this chapter, four examples of second order models are studied. The nature of the fitted surface is determined using canonical analysis and ridge analysis. The main focus of the chapter is a determination of the precision of the canonical axis which needs to be followed in order to maximise (or minimise) the response.

## 5.4 Examples

### 5.4.1 Two Dimensional Example

The first example was discussed by Box et al. (1978, p. 519). It concerns a chemical example with two input variables, time and temperature; and the response was the conversion rate. The input variables were coded and initially a  $2^2$  factorial was run with two centre points. Based on the results, a path of steepest

ascent was calculated, and experimentation was conducted along the path. Near the best set of conditions on the path an additional  $2^2$  factorial was done. This factorial experiment showed a significant lack of fit of the first order model. An additional set of experiments was carried out. This additional set of experiments contained the star points of a central composite design as well as a number of centre points. After the additional runs, a second degree equation was fitted to the data resulting in the model

$$\hat{y} = 87.3750 - 1.3837x_1 + 0.3620x_2 - 2.1437x_1^2 - 3.0937x_2^2 - 4.8750x_1x_2. \quad (5.6)$$

Details of the data and Splus commands are given in Appendix D.2. For this example,

$$\mathbf{B} = \begin{bmatrix} -2.14375 & -2.43750 \\ -2.43750 & -3.09375 \end{bmatrix}.$$

The eigenvalues of  $\mathbf{B}$  in ascending order are

$$\lambda_1 = -5.102, \quad \lambda_2 = -0.135.$$

Figure 5.1 gives the ridge plot for this example. The plot was produced using the function `ridgeplot` given in Appendix D.6. The importance of the graph was emphasised by Hoerl (1959). The graph gives the dependence of the radius of the maximum response against the value of the Lagrangian multiplier,  $\mu$ . The poles where the radius tends to infinity correspond to the eigenvalues of the matrix  $\mathbf{B}$ . Since a maximum is required, the coordinates of the maximum response can be found by considering  $\mu$  values greater than the maximum eigenvalue of  $\mathbf{B}$ , that is  $\mu$  values greater than  $-0.135$ . The co-ordinates corresponding to the maximum response as well as the maximum responses are given in Figure 5.2. The plot was produced using the function `maxplot` given in Appendix D.7. The co-ordinates are obtained by solving equation (1.9) with values of  $\mu$  greater than  $-0.135$ .

It is also very worthwhile to perform a canonical analysis of the fitted model given by equation (5.6). Since the eigenvalues are both negative we are dealing with a maximum. The optimum conditions are at the stationary point given by

$$\mathbf{x}_S = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b} = (-3.7370, -3.0028)^T.$$

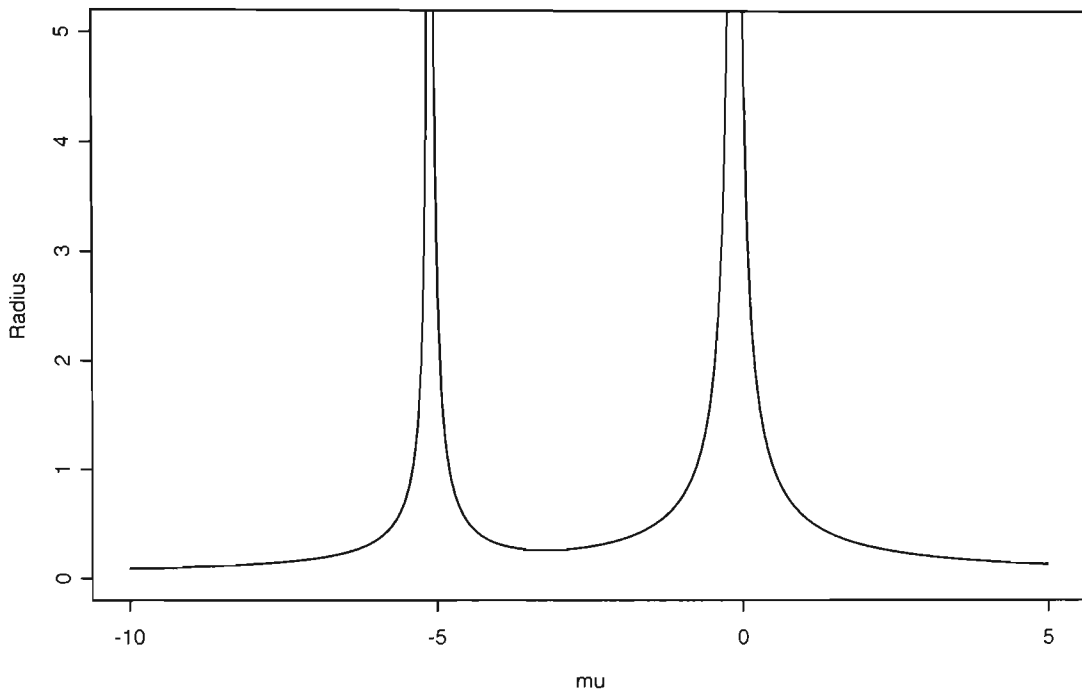


Figure 5.1: Ridge plot for the two factor chemical example given by Box et al. (1978, p.519). The graph gives the dependence of radius of the minimum response against the value of the Lagrange multiplier  $\mu$ .

at a distance of 4.79 experimental units to the centre of the experimental region. Since the stationary point is remote from the centre of the experimental region, the  $A$  canonical form given by

$$\hat{y} = 87.3750 + 1.2981X_1 + 0.6005X_2 - 0.1354X_1^2 - 5.1021X_2^2 \quad (5.7)$$

is used, where

$$\begin{aligned} X_1 &= -0.772x_1 + 0.636x_2, \\ X_2 &= -0.636x_1 - 0.772x_2. \end{aligned} \quad (5.8)$$

A function, `canonical`, to perform the canonical analysis is given in Appendix D.8.

Since the magnitude of one of the eigenvalues is quite small, some sort of ridge system is suggested. Using the double linear regression function, `dlr`, given in

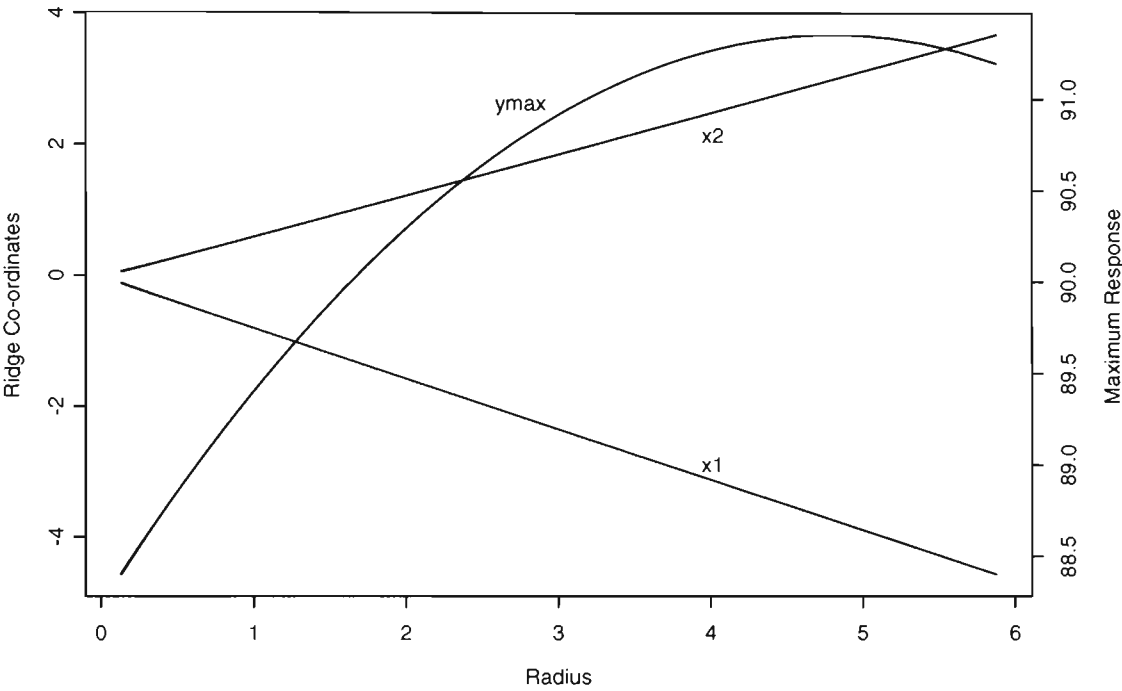


Figure 5.2: Maximum ridge co-ordinates and predicted maximum for the two factor chemical example given by Box et al. (1978, p. 519)

Model	Residual degrees of freedom	Residual sum of squares
Full Model	5	15.42
Rising Ridge( $\lambda_2 = 0$ )	6	15.44
Stationary Ridge( $\lambda_2 = \phi_2 = 0$ )	7	28.92

Table 5.1: Summary of results of fitting canonical models using non-linear least squares to the data given by Box et al. (1978, p. 519).

Appendix D.9, the design is rotated so that the standard error of the eigenvalues are the same as the standard errors of the pure quadratic terms in equation (5.6), and since these are 0.694, there is strong evidence that there is in fact a ridge system.

To test whether this is true, the canonical form given by equation (5.7) was fitted using the method of Ankenman (2003). The full quadratic equation is equivalent to the model

$$\begin{aligned} y = & \; b_0 + (x_1, x_2)\mathbf{D} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} + \\ & \; (x_1, x_2)\mathbf{D} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{D}^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \varepsilon, \end{aligned}$$

where

$$\mathbf{D} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Since it is suspected that there is a rising ridge, two additional models were fitted to the data. A rising ridge was fitted by setting the value of  $\lambda_2 = 0$ . Similarly, a stationary ridge model was fitted to the data by setting  $\lambda_2 = \phi_2 = 0$ . Table 5.1 gives a summary of the results, obtained using the Splus commands given in Appendix D.10.

The results clearly establish that a ridge model is appropriate. Testing for a rising ridge, with  $H_0 : \lambda_2 = 0$  and  $H_1 : \lambda_2 \neq 0$ ; the  $F$  statistic is 5.16, with a significance probability of 0.063, and hence, at the 10% level at least, a rising ridge is a reasonable model for the response surface. The estimated value of

$\theta = 2.452$  radians. The rising ridge equation is given by

$$\hat{y} = 87.267 + 1.298X_1 - 0.601X_2 - 5.075X_2^2.$$

where

$$\begin{pmatrix} X_1 & X_2 \end{pmatrix} = \begin{pmatrix} \cos \hat{\theta} & \sin \hat{\theta} \\ -\sin \hat{\theta} & \cos \hat{\theta} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

with  $\hat{\theta} = 2.452$ , leading to

$$X_1 = -0.772x_1 + 0.635x_2,$$

$$X_2 = -0.636x_1 - 0.772x_2,$$

which are very similar to the results given by equation (5.8).

For increased conversion rates experimentation should be performed along the ridge described by the  $X_1$  axis.

### 5.4.2 Three Dimensional Example

Another case considered was the example from Box and Draper (1987. p. 360). In this example, an experiment was conducted sequentially in four blocks. The aim was to maximise the concentration of a desired product. The three factors were flow rate, concentration of catalyst, and temperature.

A second degree equation fitted to the 24 runs was

$$\begin{aligned} \hat{y} = & 51.7958 + 0.7446x_1 + 4.8133x_2 + 8.0125x_3 - 3.8333x_1^2 + 1.2167x_2^2 - 6.2583x_3^2 \\ & + 0.3750x_1x_2 + 10.3500x_1x_3 - 2.8250x_2x_3. \end{aligned} \quad (5.9)$$

Details of the data and Splus commands are given in Appendix D.3. For this example,

$$\mathbf{B} = \begin{bmatrix} -3.833333 & 0.187500 & 5.175000 \\ 0.187500 & 1.216667 & -1.412500 \\ 5.175000 & -1.412500 & -6.258333 \end{bmatrix}.$$

The eigenvalues of  $\mathbf{B}$ , in ascending order, are

$$\lambda_1 = -10.489, \quad \lambda_2 = -0.097, \quad \lambda_3 = 1.71.$$

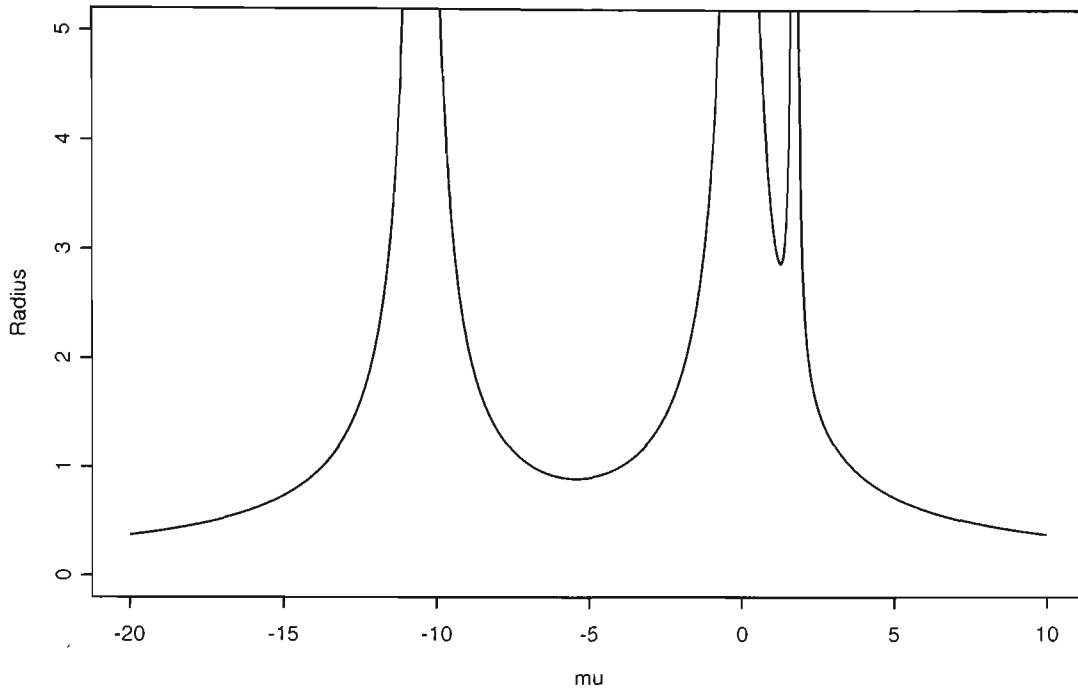


Figure 5.3: Ridge plot for the three factor reactor study example given by Box and Draper (1987, p.362). The graph gives the dependence of radius of the minimum response against the value of the Lagrange multiplier  $\mu$ .

Following the same procedure as for the two factor example, a ridge plot is obtained (Figure 5.3). Since a maximum is desired, only values of  $\mu$  greater than  $\lambda_3$  need to be considered. Considering these values of  $\mu$  the co-ordinates of the maximum response are plotted against the radius in Figure 5.4. In addition, the predicted maximum response has been superimposed on the graph.

The stationary conditions are at

$$\mathbf{x}_S = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b} = (25.7673, 15.4756, 18.4542)^T,$$

at a distance of 35.2705 design units from the design origin, and since the stationary point is far from the centre of the experimental region, the  $A$  canonical form

$$\begin{aligned} \hat{y} = & 51.796 + 1.249X_1 - 6.808X_2 - 6.326X_3 \\ & + 1.711X_1^2 - 0.097X_2^2 - 10.489X_3^2 \end{aligned} \quad (5.10)$$

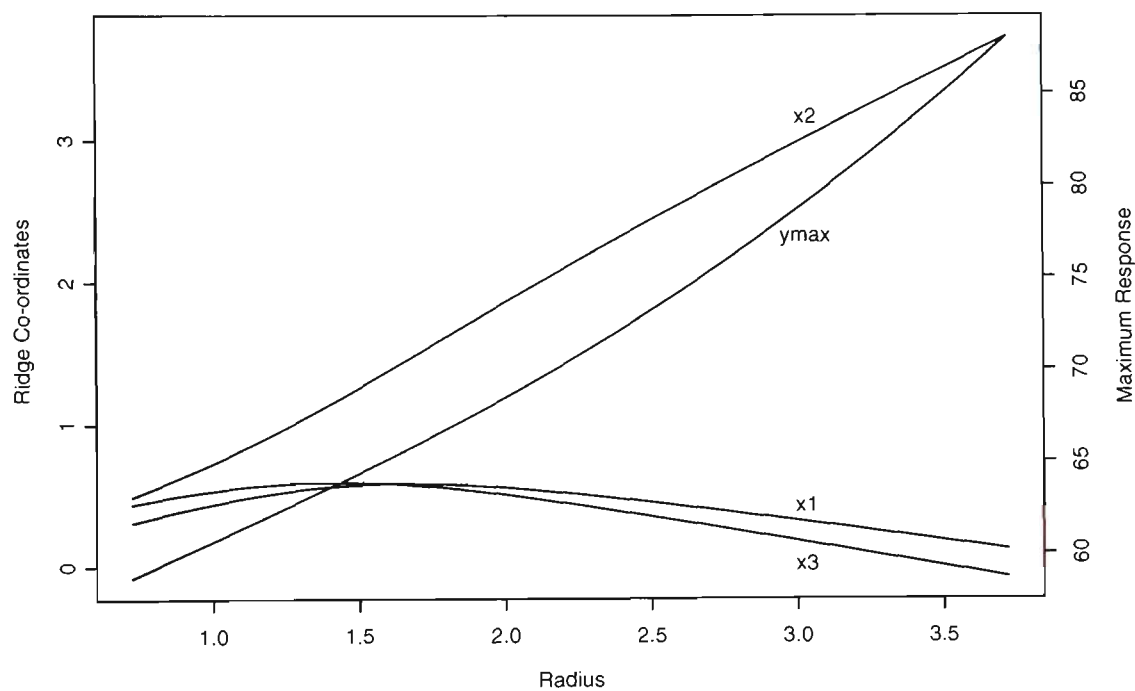


Figure 5.4: Maximum ridge co-ordinates and predicted maximum for the three factor reactor study example given by Box and Draper (1987, p. 362).



is used, where

$$X_1 = -0.297x_1 + 0.888x_2 - 0.350x_3. \quad (5.11)$$

$$X_2 = -0.733x_1 - 0.447x_2 - 0.513x_3. \quad (5.12)$$

$$X_3 = 0.612x_1 - 0.104x_2 - 0.784x_3. \quad (5.13)$$

The design is orthogonal, and the standard error of the quadratic terms in equation (5.9) is 0.543, suggesting a ridge system.

Following the method developed by Ankenman (2003), the model

$$y = b_0 + \mathbf{x}^T \mathbf{D}(\boldsymbol{\theta}) \boldsymbol{\phi} + \mathbf{x}^T \mathbf{D}(\boldsymbol{\theta}) \boldsymbol{\Lambda} \mathbf{D}^T(\boldsymbol{\theta}) \mathbf{x} + \varepsilon$$

was fitted, using the Splus commands given in Appendix D.11, where

$$\mathbf{D}^T(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{23} & \sin \theta_{23} \\ 0 & -\sin \theta_{23} & \cos \theta_{23} \end{pmatrix} \begin{pmatrix} \cos \theta_{13} & 0 & \sin \theta_{13} \\ 0 & 1 & 0 \\ -\sin \theta_{13} & 0 & \cos \theta_{13} \end{pmatrix} \begin{pmatrix} \cos \theta_{12} & \sin \theta_{12} & 0 \\ -\sin \theta_{12} & \cos \theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Explicitly, for 3 dimensions we have

$$\begin{aligned} z_1 &= \cos \theta_{12} \cos \theta_{13} x_1 \\ &\quad + \cos \theta_{13} \sin \theta_{13} x_2 + \sin \theta_{13} x_3, \end{aligned} \quad (5.14)$$

$$\begin{aligned} z_2 &= (-\cos \theta_{12} \sin \theta_{23} \sin \theta_{13} - \sin \theta_{12} \cos \theta_{23}) x_1 \\ &\quad + (-\sin \theta_{12} \sin \theta_{13} \sin \theta_{23} + \cos \theta_{12} \cos \theta_{23}) x_2 \\ &\quad + \sin \theta_{23} \cos \theta_{13} x_3, \end{aligned} \quad (5.15)$$

$$\begin{aligned} z_3 &= (-\cos \theta_{12} \cos \theta_{23} \sin \theta_{13} + \sin \theta_{12} \sin \theta_{23}) x_1 \\ &\quad + (-\sin \theta_{12} \cos \theta_{23} \sin \theta_{13} - \sin \theta_{23} \cos \theta_{12}) x_2 \\ &\quad + \cos \theta_{13} \cos \theta_{23} x_3. \end{aligned} \quad (5.16)$$

The estimates of  $\theta_{12}$ ,  $\theta_{13}$  and  $\theta_{23}$  were 1.893, 2.784 and 0.580 respectively, and substituting into equations (5.14–5.16) we obtain the same result as equations (5.11–5.13), except that the sign of  $X_1$  is reversed.

Following Box and Draper (1987, p.365) and Ankenman (2003), the terms associated with  $X_1$  and the quadratic terms associated with  $X_2$  are dropped and

the model is again fitted using non-linear least squares. The estimates of  $\theta_{12}$ ,  $\theta_{13}$  and  $\theta_{23}$  are now 2.062, 2.689 and 0.513 respectively, with the variance-covariance matrix given by

$$\mathbf{V}^* = \begin{pmatrix} 0.00982 & -0.00254 & -0.00429 \\ -0.00254 & 0.00201 & 0.00111 \\ -0.00429 & 0.00111 & 0.00336 \end{pmatrix}. \quad (5.17)$$

Substituting the estimated values of  $\theta_{12}$ ,  $\theta_{13}$  and  $\theta_{23}$  into equations (5.14–5.16), we obtain

$$\hat{y} = 54.340 - 6.921Z_1 - 6.326Z_2 - 10.812Z_2^2, \quad (5.18)$$

where

$$Z_1 = -0.667x_1 - 0.600x_2 - 0.441x_3,$$

$$Z_2 = 0.612x_1 - 0.104x_2 - 0.784x_3,$$

indicating a rising ridge in the  $Z_1$  direction. These are very similar to the results given by Box and Draper (1987, p.360) who obtained

$$\hat{y} = 54.07 - 10.812(Z_1 - 0.301)^2 + 6.923(Z_2),$$

where

$$Z_1 = -0.612x_1 + 0.104x_2 + 0.784x_3,$$

$$Z_2 = 0.683x_1 + 0.597x_2 + 0.420x_3,$$

by directly fitting the canonical form using non-linear regression.

### 5.4.3 Four Dimensional Example

The next example considered was a four factor example presented in Box and Liu (1999). The example involved an experiment in which it is required to maximise flight times of a paper helicopter. The factors considered were: wing area,  $x_1$ , wing length,  $x_2$ , body width,  $x_3$ , and body length,  $x_4$ . The second degree equation

was fitted to the data and the resulting model was

$$\begin{aligned}
 \hat{y} = & 371.3250 - 0.0833x_1 - 5.0833x_2 + 0.2500x_3 - 6.0833x_4 \\
 & -2.0375x_1^2 - 1.6625x_2^2 - 2.5375x_3^2 - 0.1625x_4^2 \\
 & -2.8750x_1x_2 - 3.7500x_1x_3 + 4.3750x_1x_4 + 4.6250x_2x_3 \\
 & -1.5000x_2x_4 - 2.1250x_3x_4.
 \end{aligned} \tag{5.19}$$

Details of the data and Splus commands are given in Appendix D.4. For this example,

$$\mathbf{B} = \begin{bmatrix} -2.0375 & -1.4375 & -1.8750 & 2.1875 \\ -1.4375 & -1.6625 & 2.3125 & -0.7500 \\ -1.8750 & 2.3125 & -2.5375 & -1.0625 \\ 2.1875 & -0.7500 & -1.0625 & -0.1625 \end{bmatrix}.$$

The eigenvalues of  $\mathbf{B}$ , in ascending order, are

$$\lambda_1 = -4.6520, \quad \lambda_2 = -3.8079, \quad \lambda_3 = -1.1983, \quad \lambda_4 = 3.2582.$$

Following the same procedure as for the two factor example, a ridge plot is obtained and is given in Figure 5.5. The co-ordinates of the maximum response are plotted against the radius in Figure 5.6. In addition, the predicted maximum response has been superimposed on the graph.

The stationary conditions are at

$$\mathbf{x}_S = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b} = (0.8607, -0.3307, -0.8395, -0.1161)^T,$$

a distance of only 1.2524 units from the origin. Since the eigenvalues are of different signs the ridge system is a saddle point.

The analysis is carried out in terms of the  $A$  canonical form

$$\begin{aligned}
 \hat{y} = & 371.3250 - 5.9140X_1 + 1.3331X_2 - 4.8371X_3 - 1.6632X_4 \\
 & + 3.2582X_1^2 - 1.1983X_2^2 - 3.8079X_3^2 - 4.6520X_4^2,
 \end{aligned} \tag{5.20}$$

where

$$X_1 = 0.518x_1 - 0.450x_2 - 0.452x_3 + 0.571x_4. \quad (5.21)$$

$$X_2 = -0.041x_1 - 0.582x_2 - 0.376x_3 - 0.720x_4. \quad (5.22)$$

$$X_3 = -0.761x_1 - 0.506x_2 + 0.122x_3 + 0.388x_4, \quad (5.23)$$

$$X_4 = 0.389x_1 - 0.451x_2 + 0.800x_3 - 0.076x_4. \quad (5.24)$$

The design is rotatable, and hence, the standard errors of the eigenvalues are the same as the standard error of the quadratic terms, and equal to 0.6039. One of the key findings of the Box and Liu (1999) experiment was that it was possible to *increase* the flight times of a helicopter by conducting further experiments in either direction along the  $X_1$  axis since the coefficient of  $X_1^2$  is significant and positive.

#### 5.4.4 Five Dimensional Example

In the five-factor batch polymerization of an acrylamide experiment (Kiprianova and Markovska, 1993, p.83), the input factors were  $x_1$ , initiator concentration (%);  $x_2$ , monomer concentration ( $\text{mol dm}^{-3}$ );  $x_3$ , temperature of polymerization (K);  $x_4$ , pH of initial reaction mixture and  $x_5$ , residence time (S). The experimental design chosen was a  $B_5$  symmetrical composite design.

The responses were the intrinsic viscosity and polydispersity coefficient. The 26-run experimental design and results are given in Table 5.2. For the intrinsic viscosity coefficient the full second order model was fitted and is given by

$$\begin{aligned} \hat{y} = & 2.7310 - 0.2850x_1 + 0.0962x_2 - 0.2265x_3 - 0.0561x_4 \\ & + 0.3656x_5 - 0.0332x_1^2 - 0.1807x_2^2 - 0.0477x_3^2 \\ & - 0.0112x_4^2 + 0.4673x_5^2 + 0.2569x_1x_2 - 0.1175x_1x_3 \\ & - 0.0904x_1x_5 + 0.2260x_2x_3 + 0.3421x_2x_4 + 0.2781x_2x_5 \\ & + 0.2517x_3x_4 + 0.1231x_4x_5. \end{aligned} \quad (5.25)$$

Details of the data and Splus commands are given in Appendix D.5. For this

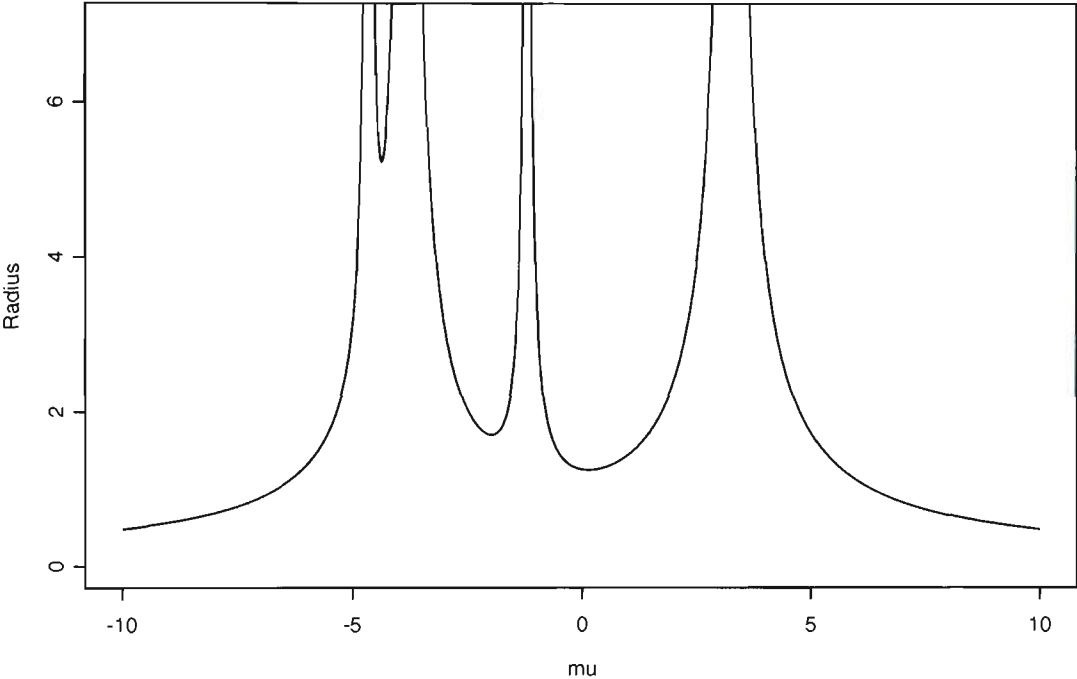


Figure 5.5: Ridge plot for the four factor helicopter example given by Box and Liu (1999). The graph gives the dependence of radius of the maximum response against the value of the Lagrange multiplier  $\mu$ .

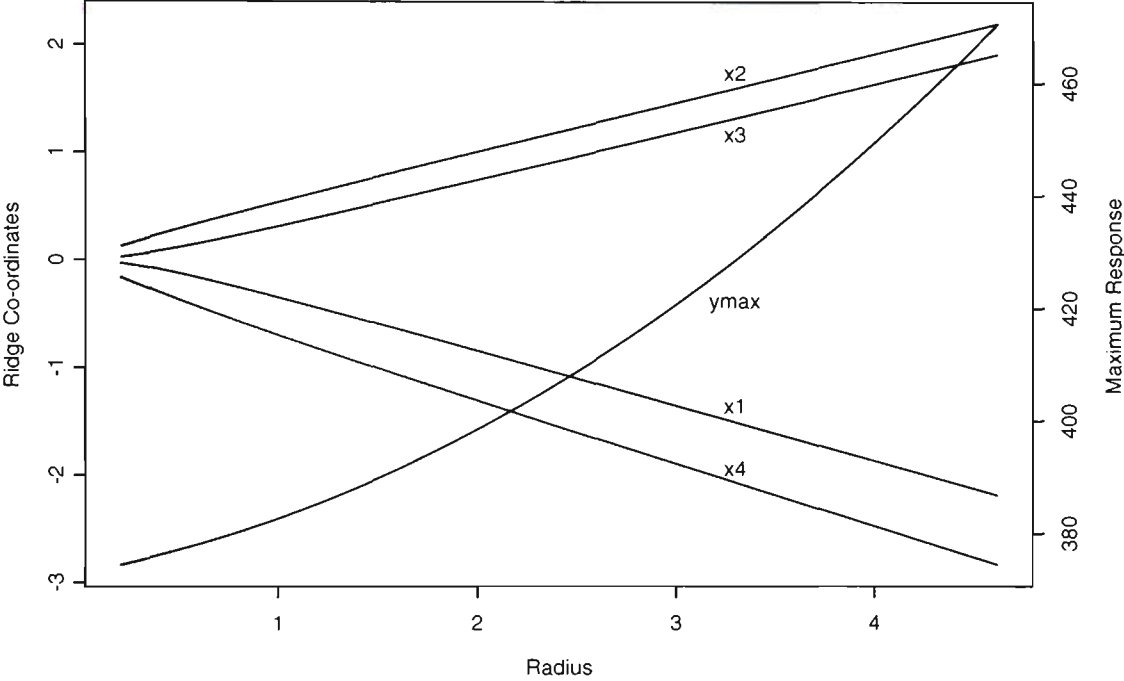


Figure 5.6: Optimum ridge co-ordinates and predicted maximum for the four factor helicopter example given by Box and Liu (1999).

Run	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y$	Run	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y$
1	1	1	1	1	1	4.016	14	-1	1	-1	-1	-1	2.275
2	-1	1	1	1	-1	2.753	15	1	-1	-1	-1	-1	3.681
3	1	-1	1	1	-1	1.495	16	-1	-1	-1	-1	1	4.540
4	-1	-1	1	1	1	3.161	17	1	0	0	0	0	2.383
5	1	1	-1	1	-1	2.407	18	-1	0	0	0	0	3.017
6	-1	1	-1	1	1	3.765	19	0	1	0	0	0	3.014
7	1	-1	-1	1	1	2.407	20	0	-1	0	0	0	2.091
8	-1	-1	-1	1	-1	2.737	21	0	0	1	0	0	2.552
9	1	1	1	-1	-1	1.879	22	0	0	-1	0	0	2.819
10	-1	1	1	-1	1	3.216	23	0	0	0	1	0	2.897
11	1	-1	1	-1	1	1.774	24	0	0	0	-1	0	2.547
12	-1	-1	1	-1	-1	3.222	25	0	0	0	0	1	3.519
13	1	1	-1	-1	1	3.514	26	0	0	0	0	-1	2.882

Table 5.2: Symmetrical composite design type  $B_5$  for intrinsic viscosity model.

example,

$$\mathbf{B} = \begin{bmatrix} -0.03321875 & 0.1284375 & -0.05875000 & 0.00981250 & -0.0451875 \\ 0.12843750 & -0.1807187 & 0.11300000 & 0.17106250 & 0.1390625 \\ -0.05875000 & 0.1130000 & -0.04771875 & 0.12587500 & -0.0096250 \\ 0.00981250 & 0.1710625 & 0.12587500 & -0.01121875 & 0.0615625 \\ -0.04518750 & 0.1390625 & -0.00962500 & 0.06156250 & 0.4672813 \end{bmatrix}.$$

Ridge analysis was used to obtain the direction of improvement, that is, minimum viscosity. The eigenvalues of  $\mathbf{B}$  in ascending order are

$$\begin{aligned} \lambda_1 &= -0.3581658, \quad \lambda_2 = -0.1760145, \quad \lambda_3 = 0.0302720, \\ \lambda_4 &= 0.1805130, \quad \lambda_5 = 0.5178010. \end{aligned}$$

Figure 5.7 gives a plot of  $\mu$  versus  $R$ . For this problem, a minimum is required, and therefore the co-ordinates of the minimum response can be found using values of  $\mu < -0.35816577$ .

Figure 5.8 gives a plot of the coordinates  $X_1, \dots, X_5$  and the corresponding value of  $R$  as well as the predicted minimum response. Since the eigenvalues are

of different signs we have a saddle point. The stationary point is at

$$\mathbf{x}_S = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b} = (1.6182, 0.9126, -0.8136, 0.4824, -0.5178)^T.$$

at a distance of 2.1657 units from the centre of the experiment. The  $A$  canonical form is

$$\begin{aligned}\hat{y} = & 2.7310 + 0.3449X_1 + 0.2801X_2 - 0.1166X_3 \\ & + 0.1485X_4 - 0.2133X_5 + 0.5178X_1^2 + 0.1805X_2^2 \\ & + 0.0303X_3^2 - 0.1760X_4^2 - 0.3580X_5^2,\end{aligned}\quad (5.26)$$

where

$$\begin{aligned}X_1 &= -0.025x_1 - 0.248x_2 + 0.0828x_3 + 0.209x_4 + 0.942x_5, \\ X_1 &= -0.204x_1 - 0.427x_2 - 0.526x_3 - 0.6426x_4 + 0.295x_5, \\ X_3 &= 0.866x_1 + 0.242x_2 - 0.4297x_3 - 0.077x_4 + 0.014x_5, \\ X_4 &= -0.241x_1 - 0.118x_2 - 0.679x_3 + 0.680x_4 - 0.067x_5, \\ X_5 &= 0.387x_1 - 0.827x_2 + 0.267x_3 + 0.274x_4 + 0.143x_5.\end{aligned}$$

The design is not rotatable, so the double linear regression method was used to determine the standard errors of the eigenvalues. They are 0.1674, 0.0829, 0.1412, 0.1131, and 0.1255 respectively. Hence, in order to achieve lower viscosities, experimentation should be conducted in either direction along the  $X_5$  axis.

## 5.5 Uncertainty of Canonical Direction

The four examples considered in this chapter show how useful canonical analysis is. For the two-dimensional example and the three-dimensional example, the canonical analysis revealed that the response surfaces could be described as rising ridges, while for the four-dimensional and five-dimensional examples, the response surfaces are saddles. In all cases, the canonical analysis indicates a direction which will lead to more optimal conditions. In this section the uncertainty of the indicated direction will be examined.

### 5.5.1 Two Dimensional Example

For the two dimensional example, the uncertainty in the orientation of this direction is given by the standard error of the estimate of  $\theta$ , which was 0.0793, based



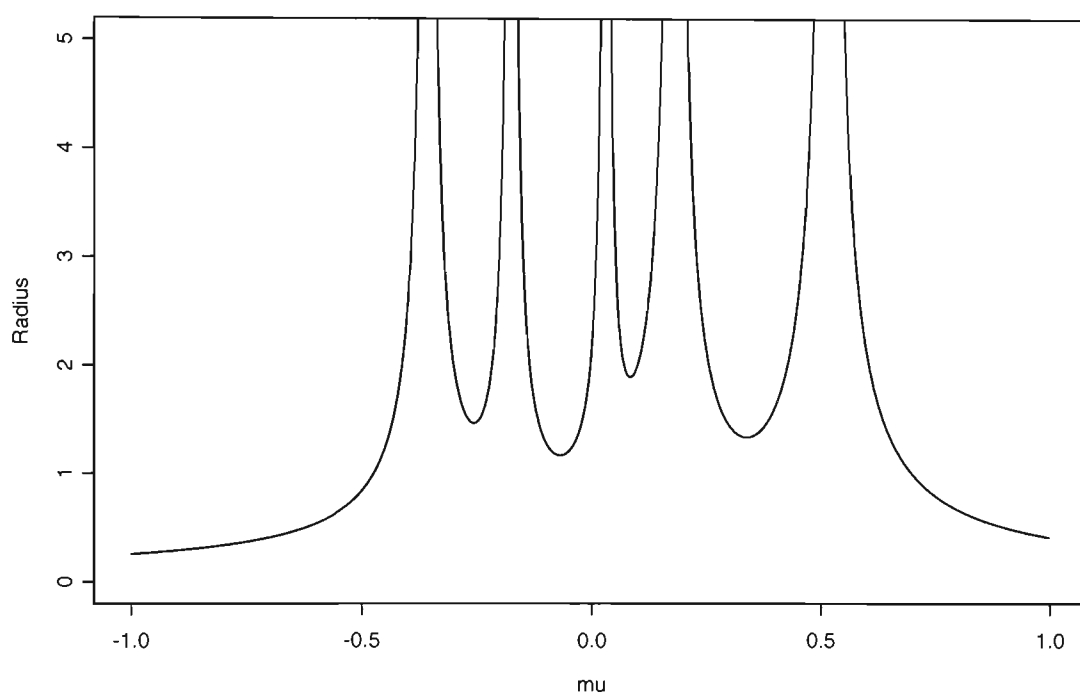


Figure 5.7: Ridge plot for the five factor polymilization acrylamide experiment given by Kiprianova and Markovska (1993). The graph gives the dependence of radius of the minimum response against the value of the Lagrange multiplier  $\mu$ .

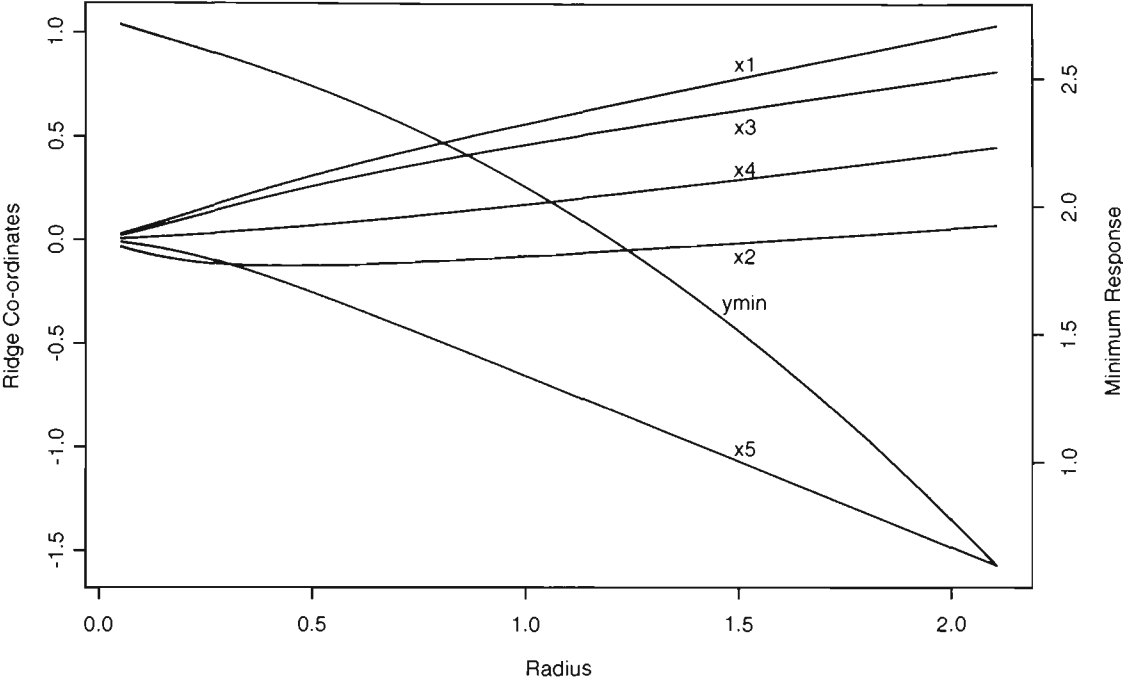


Figure 5.8: Optimum ridge co-ordinates and predicted minimum for the five factor polymilization acrylamide experiment given by Kiprianova and Markovska (1993).

on 6 degrees of freedom. A 95% confidence interval for the true value of  $\theta$  based on the  $t$ -distribution extends from 2.256 to 2.646 radians and includes 6.2% of possible directions.

### 5.5.2 Three Dimensional Example

For the three dimensional example, equation (5.18) indicates that there is a rising ridge and therefore experimentation should be conducted along the  $Z_1$  axis in order to increase the response. The estimated direction cosines of  $Z_1$  are functions of the estimated values of  $\theta_{12}$ ,  $\theta_{13}$  and  $\theta_{23}$  and the *multivariate delta method* could be used to give a confidence cone for the true direction cosines. However, a simpler approximate method was adopted. First, the appropriate Householder transformation,  $\mathbf{H}_u$ , to take the direction cosines of  $Z_1$  to  $(0, 0, 1)^T$ , was determined. Then, a large number (say  $N = 10,000$ ) of observations with mean  $(2.062, 2.689, 0.513)^T$  and variance  $\mathbf{V}^*$ , given in equation (5.17), was generated. For each observation equations (5.14–5.16) were used to convert the simulated value of  $(\theta_1, \theta_2, \theta_3)^T$  into a simulated direction vector. The simulated direction vectors were then pre-multiplied by  $\mathbf{H}_u$  and the first two co-ordinates were taken, which corresponds to projecting the simulated direction vector onto the plane orthogonal to the path of steepest ascent. With  $N = 10,000$  random samples the variance-covariance matrix of the first two co-ordinates was calculated to be

$$\mathbf{V}^{**} = \begin{pmatrix} 0.00180 & -0.00144 \\ -0.00144 & 0.00789 \end{pmatrix}.$$

Since

$$(\boldsymbol{\delta} - \hat{\boldsymbol{\delta}})^T (\mathbf{V}^{**})^{-1} (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}) \sim \chi^2(2),$$

where  $\hat{\boldsymbol{\delta}} = (0, 0)$  and  $\boldsymbol{\delta}$  is the projection of the direction cosine vector onto the plane orthogonal to the the path of steepest ascent, an approximate 95% joint confidence region for  $\delta_1$  and  $\delta_2$  can be generated as an ellipse with semi-axes

$$a_1 = \sqrt{\frac{\chi_2^2(0.95)}{\lambda_1}} \quad \text{and} \quad a_2 = \sqrt{\frac{\chi_2^2(0.95)}{\lambda_2}},$$

where  $\lambda_1 = 677.2966$  and  $\lambda_2 = 121.8066$  are the eigenvalues of  $(\mathbf{V}^{**})^{-1}$ . The computed surface area above the projected region corresponds to a fraction of 0.00523 and hence excludes 99.5% of possible directions. Splus commands that were used are given in Appendix D.13.

### 5.5.3 Four Dimensional Example

Again, following the method developed by Ankenman (2003), the model

$$y = b_0 + \mathbf{x}^T \mathbf{D}(\boldsymbol{\theta}) \boldsymbol{\phi} + \mathbf{x}^T \mathbf{D}(\boldsymbol{\theta}) \boldsymbol{\Lambda} \mathbf{D}^T(\boldsymbol{\theta}) \mathbf{x} + \varepsilon$$

was fitted, using the Splus commands given in Appendix D.12, where now  $\mathbf{D}$  is the product of six matrices involving the rotation angles  $\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}$  and  $\theta_{34}$ . The estimated values of these angles are 5.567, 5.701, 0.607, 3.396, 5.214 and 1.763 radians respectively, with variance-covariance matrix given by

$$\mathbf{V}^* = \begin{pmatrix} 0.0086 & 0.0006 & 0.0027 & 0.0034 & 0.0040 & -0.0015 \\ 0.0006 & 0.0037 & 0.0007 & 0.0039 & 0.0008 & 0.0041 \\ 0.0027 & 0.0007 & 0.0066 & 0.0016 & 0.0014 & 0.0001 \\ 0.0034 & 0.0039 & 0.0016 & 0.0616 & -0.0016 & 0.0532 \\ 0.0040 & 0.0008 & 0.0014 & -0.0016 & 0.0239 & -0.0034 \\ -0.0015 & 0.0041 & 0.0001 & 0.0532 & -0.0034 & 0.2618 \end{pmatrix}.$$

Using the estimated  $\theta$  values we get the same canonical analysis equations as in equations (5.21–5.24).

To determine the uncertainty involved with the  $X_1$  axis, a similar method as for the three dimensional case is used, with the Splus commands given in Appendix D.14. The estimated direction cosines of  $X_1$  are functions of the estimated values of  $\theta_{12}, \theta_{13}, \theta_{14}, \theta_{23}, \theta_{24}$  and  $\theta_{34}$ . First, the appropriate Householder transformation,  $\mathbf{H}_{\mathbf{u}}$ , was determined to take the direction cosines of  $X_1$  to  $(0, 0, 0, 1)^T$ . A large number (say  $N = 10,000$ ) of observations with mean  $(2.062, 2.689, 0.513)^T$  and variance  $\mathbf{V}^*$  was then generated. For each observation the simulated value of the  $\boldsymbol{\theta}$  vector was transformed into a simulated direction vector. The simulated direction vectors were pre-multiplied by  $\mathbf{H}_{\mathbf{u}}$ , and the first three co-ordinates were

taken, which corresponds to projecting the simulated direction vector onto the hyperplane orthogonal to the path of steepest ascent. With  $N = 10,000$  random samples the variance-covariance matrix of the first three co-ordinates was calculated to be

$$\mathbf{V}^{**} = \begin{pmatrix} 0.0067 & -0.0007 & -0.0017 \\ -0.0007 & 0.0031 & 0.0005 \\ -0.0017 & 0.0005 & 0.0032 \end{pmatrix}.$$

Since

$$(\boldsymbol{\delta} - \hat{\boldsymbol{\delta}})^T (\mathbf{V}^{**})^{-1} (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}) \sim \chi^2(3),$$

where  $\hat{\boldsymbol{\delta}} = (0, 0, 0)$  and  $\boldsymbol{\delta}$  is the projection of the direction cosine vector onto the plane orthogonal to the the path of steepest ascent, we can generate a 95% confidence region for  $\delta_1, \delta_2$ , and  $\delta_3$  which is an ellipsoid with semi-axes

$$a_1 = \sqrt{\frac{\chi_3^2(0.95)}{\lambda_1}}, \quad a_2 = \sqrt{\frac{\chi_3^2(0.95)}{\lambda_2}}, \quad \text{and} \quad a_3 = \sqrt{\frac{\chi_3^2(0.95)}{\lambda_3}},$$

where  $\lambda_1 = 416.566$ ,  $\lambda_2 = 334.877$  and  $\lambda_3 = 131.777$  are the eigenvalues of  $(\mathbf{V}^{**})^{-1}$ , which gives  $a_1 = 0.2132$ ,  $a_2 = 0.1338$  and  $a_3 = 0.1199$ . Using the inequalities (2.21–2.22), the upper and lower bounds for the the computed surface area as a proportion of the total surface area of the sphere above the projected region are 0.072 and 0.117. However, since for a maximum or saddle point the sign of the canonical axis is arbitrary, the computed surface area should be compared to the surface area of the four dimensional hemisphere and hence both bounds need to be multiplied by two, showing that the 95% confidence cone for the  $X_1$  canonical axis includes between 14.4% and 23.4% of possible directions. There is quite a large amount of uncertainty in the direction.

#### 5.5.4 Five Dimensional Example

A similar approach can be used for the five dimensional example. The rotation matrix  $\mathbf{D}$  involves 10 rotation angles. Note that since the response surface is a saddle, as for the four dimensional case, when computing the proportion of directions included in the confidence cone for the  $X_5$  axis, the appropriate comparison

to the surface area of the cap associated with the confidence cone for the canonical direction is the surface area of the five dimensional hemisphere and not the surface area of the five dimensional sphere.

## 5.6 Precision of Ridge Analysis

In the previous section, the uncertainty associated with a direction for further experimentation, identified by canonical analysis, was examined. It should be noted that the ridge co-ordinates usually converge quite quickly onto the canonical axes and hence the results in the previous section can be used to give the uncertainty associated with a ridge analysis solution.

If desired, before the ridge co-ordinates converge onto the canonical axes, an alternative assessment of uncertainty can be undertaken. One way would be to use either of the methods developed by Peterson et al. (2002) or Gilmour and Draper (2003). A helpful way to view these results would be to plot the maximum ridge co-ordinates and predicted maximum as in Figure 5.2, but with an additional vertical axis giving the proportion of the surface area of the circle or sphere covered by the confidence region for the co-ordinates corresponding to the optimal conditions.

An additional method would be to use the results of Chapter 2 to determine the precision of the best direction on the path of steepest ascent. The method is discussed below.

Points on the path of steepest ascent for second-order models are given by the solution of

$$\mathbf{x}_S = -\frac{1}{2}(\mathbf{B} - \mu\mathbf{I})^{-1}\mathbf{b}$$

for various values of  $\mu$  greater than the largest eigenvalue of  $B$ . At these points, the gradient of the response at  $\mathbf{x}_S$

$$\hat{y} = b_0 + \mathbf{b}^T\mathbf{x}_S + \mathbf{x}_S^T\mathbf{B}\mathbf{x}_S$$

can be determined as

$$\left. \frac{d\hat{y}}{dx} \right|_{x_s} = \mathbf{b}^T + 2\mathbf{B}\mathbf{x}_s \quad (5.27)$$

$$= \mathbf{A} \begin{pmatrix} \mathbf{b} \\ \mathbf{b}_Q \end{pmatrix}, \quad (5.28)$$

where

$$\mathbf{b}_Q^T = (b_0 \ b_{11} \ b_{22} \ \dots \ b_{kk} \ b_{12} \ \dots \ b_{k-1,k}).$$

The variance-covariance matrix of the gradient is given by

$$\mathbf{A}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{A}^T \sigma^2. \quad (5.29)$$

In general, equation (5.29) will be heterogeneous and non-orthogonal. To determine the precision of the path as described by the gradient direction, the method of Chapter 2 for heterogeneous regression parameters can be applied using  $\mathbf{G} = \mathbf{A}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{A}^T$ . For a two dimensional case, for example,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 2x_{10} & 0 & x_{20} \\ 0 & 0 & 1 & 0 & 2x_{20} & x_{10} \end{pmatrix},$$

where  $(x_{10}, x_{20})$  is a point on the path. For each radius the percentage included in the confidence cone can be calculated, and a plot of the percentages can be overlaid on the ridge co-ordinates graph.

## 5.7 Conclusion

In this chapter, methods for determining the precision of the path of steepest ascent subject to constraints have been developed.

When there is a linear constraint the original path needs to be projected onto the constraint plane. Even if the errors are homogeneous and the design is orthogonal, the projected regression coefficients will have a non-orthogonal variance-covariance matrix. However, using the method developed in Chapter 2, it is easy to determine the precision of the projected path.

For a quadratic response, canonical analysis and ridge analysis are appropriate. As the radius increases the ridge analysis solution tends to asymptote to a canonical axis. A method for determining the precision of the canonical axes has been presented, using the method developed by Ankenman (2003) to fit the canonical second order model.

The methods developed by Peterson et al. (2002) and Gilmour and Draper (2003) are also relevant. In addition, another method, based on the results of Chapter 2 has been suggested.



# Chapter 6

## Discussion and Further Work

### 6.1 Introduction

The objective of this research was to extend some methods used in response surface methodology, with the main focus on the precision of the path of steepest ascent. The method of steepest ascent is important in experimental work as, based on an initial experimental design, it provides the experimenter with a direction for future experiments. The method for determining the precision of the path of steepest ascent developed by Box (1955) and Box and Draper (1987, pp. 190–194) is particularly useful. However, in its present form, it could only be applied to orthogonal designs with uncorrelated and homogeneous errors. In many instances, including GLMs, the above assumptions are not met.

In this thesis, a useful extension to the existing method is presented, so that it can be applied more generally. In fact, this extension turns out to be useful in a variety of situations. In addition, some techniques are proposed for augmenting existing designs. These techniques may be particularly useful in situations when the path of steepest ascent has not been determined precisely enough. Some techniques that can be useful for determining the precision of canonical analysis are also developed.

A possible limitation of the method of steepest ascent occurs when the experimental variables are constrained. In this thesis, analogous methods are included

for the precision of the path of steepest ascent when a linear constraint is encountered.

## 6.2 Homogeneous Case

The method developed by Box (1955) and Box and Draper (1987, pp. 190–194) is described in Chapter 2. An alternative derivation of the method is also presented. This method determines the precision of the path of steepest ascent by giving the proportion of possible directions included in the confidence cone. It is shown that the proportion of all possible directions included in the confidence cone is a function of the tail area of the  $t$ -distribution.

The utility of the current method is investigated by applying it to a fractional factorial example. The response in this example is a count (of defects); therefore, the response needs to be transformed to be able to apply the existing method. In the example, there was a significant interaction between a qualitative and quantitative variable. If the quantitative variable in the interaction cannot be varied, the regression coefficients that determine the path of steepest ascent are homogeneous, and therefore the precision of the path of steepest ascent can be determined by the current method. On the other hand, if the quantitative variable in the interaction can be varied, then one of the two levels of the qualitative variable needs to be chosen. In this case, the regression coefficients that determine the path of steepest ascent are heterogeneous, and therefore the current method cannot be used without modification.

A modification to the existing method is applied by projecting the confidence region, which is a cap of a sphere, onto the hyperplane passing through the origin and perpendicular to the path of steepest ascent. This projection provides a hyperellipsoid, with the semi-axes having equal lengths for the homogeneous case, and different lengths for the heterogeneous case. Methods are given for obtaining the lengths of the axes of the hyperellipsoid. An extension to the current method

can be accomplished by considering the structure of the variance-covariance matrix. For situations with two regression coefficients, an exact expression is obtained. This method, applied to the homogeneous case considered above, gives the same results as the existing method. To apply the new method for higher dimensions, however, a transformation onto a rectangular set of coordinates is used before applying numerical integration. To determine the surface area of the cap, Gaussian quadrature or Monte-Carlo integration is used.

Some inequalities are developed for the precision of the path of steepest ascent. These give sharp bounds which, for some practical purposes, can be precise enough without the need for numerical integration.

### 6.3 Heterogeneous Case

In Chapter 3, the method developed in Chapter 2 is used again, but this time with generalised linear models applied to the original count response, rather than OLS with a transformed response. Two approaches for the determination of confidence regions are studied. First, the Wald approach is applied for situations where overdispersion can be ignored and when it must be considered. The results show the increased effectiveness when the new method is used directly with the count response, rather than using a transformed response, since the size of the confidence region is smaller, and hence the percentage included in the confidence cone is also smaller.

A second approach considered is the application of profile likelihoods. Again, this method can be used for both the no overdispersion and overdispersion situations. It is shown that the profile likelihood method gives similar results to the Wald approach for the example considered, but there may be examples where it will outperform the Wald approach.

The confidence regions obtained using the Wald and profile likelihood methods are based on approximate asymptotic results, which may be inaccurate. Therefore, some parametric and nonparametric bootstrap checks are proposed. The bootstrap results show when the asymptotic distributions are inappropriate, and

what adjustments should be made in order to get accurate results.

## 6.4 Augmenting

In Chapter 4, the focus is on some design aspects of determining the path of steepest ascent with sufficient precision. For the homogeneous case, the precision of the path can be determined prior to the experiment using standard power techniques. A formula is derived involving the distribution function of the noncentral  $F$ -distribution and the inverse of the  $t$ -distribution. Not surprisingly, the power depends on the sample size, the experimental error, and the magnitude of the regression coefficients vector.

When the data is heteroscedastic, the precision of the path depends also on the direction of the path. There are some directions that are determined more precisely than other directions. If the magnitude of the path, but not the direction, can be predicted, then a conservative power analysis can be conducted by assuming that the path is in the direction of minimum precision.

An important case when heterogeneity applies is for GLMs. Two important cases are for the Poisson model, often used for modelling counts, and the Binomial model, used for modelling proportions. The question considered is how to design experiments for GLMs, where the objective is to best determine the direction for further experimentation. Using an example with two levels, it is shown that considerable improvements over the standard factorial designs can be achieved by choosing the design points to minimise the proportion included in the confidence cone for the path.

If the path has not been determined precisely enough, further experiments might be required. Methods are given for determining the nature and number of these experiments, using either a D-optimal design, or by directly minimising the proportion of directions included in the confidence cone for the path of steepest ascent. In many cases the D-optimal designs give quite good results.

## 6.5 Constraints and Second Order Models

The results of chapters 2, 3, and 4 apply to unconstrained situations. However, in many practical situations constraints must be considered. These constraints may arise as restrictions on the explanatory/experimental variables, because of a need to keep secondary responses at desirable levels, or a need to keep the solution within a certain distance from the centre of the experimental region.

For a linear constraint, the optimum path is followed until the constraint is reached, and once it is reached, the path is projected onto the constraint plane. The variance-covariance of the projected path coefficients can be easily determined, and the method developed in Chapter 2 can be applied. An example is given that shows that the projected path may have much less precision than the original path once the constraint plane is reached, and hence further experiments may be required.

Four examples are considered involving two to five dimensions. For each of the examples, a canonical analysis and ridge analysis was conducted. As the radius increases, the ridge solution converges quickly to the canonical axes. A useful method for determining confidence cones for the canonical axes was developed and applied to the examples.

## 6.6 Limitations

Although the results in this thesis should prove useful to experimenters, there are a number of limitations that should be borne in mind. Some of these limitations are discussed in this section.

### 6.6.1 Correct Model Form

In response surface methodology, we generally use linear and quadratic regression models. It is not suggested that these models are exactly correct, but it is hoped that they are sufficiently accurate to make the predictions based on these models useful. In general, the accuracy of the model decreases the further from the

centre of the experimental region. No consideration of the bias involved has been considered, and hence the further we are along the path, the less accurate and less useful the results will become.

### 6.6.2 Correct Link

With GLMs, as well as choosing the distribution and the linear predictor, the analyst must specify the link function, the function which links the mean of the distribution to the linear predictor. In cases where the link is incorrectly specified the confidence regions may be inaccurate. The results in this thesis assume that the link is correctly specified, and if not, the effect on the confidence region is relatively minor.

### 6.6.3 Overdispersion

When using GLMs, the decision of whether to include overdispersion is a critical one as far as the confidence region is concerned. If there is actually overdispersion but it is ignored, the confidence region will likely be smaller than it should be, and hence the path might be considered to be determined precisely enough when, in reality, further experimentation might be the best course of action. On the other hand, if there is no overdispersion, and it is adjusted for, the confidence regions will be generally inflated in size, and further experiments might be conducted when the most fruitful action would be to do experiments along the path.

### 6.6.4 Known Constraint Functions

In Chapter 5, it is shown how to determine the precision of the modified path when there is a constraint. It is assumed that the constraint is known. However, sometimes the constraint is not known, but is estimated from previous experiments, or from the current experiment using one or more secondary responses. The effect of estimating the constraint function is not considered in this thesis. It is expected that including the uncertainty in the constraint function will increase

the size of the confidence region, and hence the proportion of possible directions included in the confidence cone.

## 6.7 Future Work

There are a number of areas that warrant further research:

### 6.7.1 Consideration of Bias and Variance

As we follow the path, we get further and further away from the centre of the experimental region, and the level of the lack of fit of the response surface will increase. However, at this stage, the methods developed ignore this lack of fit, and only consider the variance. Obtaining a more accurate measure of uncertainty, by considering both bias and variance, would strengthen the applicability of the results. For that reason, the results are only useful within a limited spherical region.

### 6.7.2 GLMs with Varying Dispersion

One important motivation for the use of experimental design is the desire to improve the quality of products and processes. The idea of using experimental designs to minimise the variability subject to the mean being on target is a very useful one, and has been promoted by Taguchi and others.

In GLMs there is a relationship between the mean and the variance function. If desired, an overdispersion factor can be used; however, there are some practical cases where both the mean and the variability can depend on the experimental variables. Smyth (1989) gives a model where both the mean and variance can be modelled using GLMs.

If it is desired to minimise the variability subject to the estimate being on target, the results of the previous chapter can be adopted. An extension to considering the impact of the estimation of the secondary response would be to consider the case when the primary and secondary responses are calculated

using GLMs. In this case, the primary response would be the variance, while the secondary response would be the mean.

### 6.7.3 Comparison of D-Optimal and Minimum Proportion Designs

In Chapter 4, simulations using an orthogonal array were used to show, over the ranges studied, that D-optimal designs were relatively efficient compared to the minimum proportion designs. This is not surprising since a D-optimal design gives the minimum sized confidence region for the regression coefficients. However, some of the simulations showed that the minimum proportion designs can provide much more efficient estimators of the path of steepest ascent in some situations. Further examination and conduct of simulations studies similar to those in Chapter 4 are desirable. In particular, the literature on computer experiments (see, for example, Sacks, Welch, Mitchell and Wynn, 1989) could be used to determine the efficiency of the D-optimal design relative to the minimum proportion design, depending on the expected response at the centre of the experimental region, the expected size and orientation of the regression coefficients vector, and (for Binomial data) the sample size. Using this methodology, a best unbiased linear predictor relating these parameters to the performance of designs could be developed and used in the planning of future experiments. Extensions to higher order dimensions should also be done.

### 6.7.4 Allocation of Observations for Binomial Data

In Chapter 4, methods were considered for determining the best locations of experimental observations when the data follows a Binomial or Poisson distribution, and the path is desired with maximum precision. It is shown that if estimates of the response surface are available, designs can be found that offer a considerable improvement over the standard factorial designs.

For Binomial data, only equal sample sizes were considered. Even more improvement could be obtained by allowing the number of observations at each set



of conditions to be different.

### 6.7.5 Bayesian Designs

In Chapter 4, the designs are based on assuming particular values of the regression coefficients. If we replace these estimates by prior distributions, different optimum designs will be selected. A study of these Bayesian designs for the path of steepest ascent would be extremely useful.

### 6.7.6 Application to Designs with Randomisation Restrictions

Frequently, one or more of the experimental variables are harder to vary than some of the others. In these cases, split-plot designs are used (see, for example, Box and Jones, 1992). With these designs, the estimated parameters are heterogeneous, and therefore the method for determining the precision of the path of steepest ascent developed in Chapter 2 could be applied. For second order designs, Letsinger et al. (1996) determined how best a Central Composite Design should be conducted to allow for these randomisation restrictions. However, no work to date has examined how path of steepest ascent experiments should be conducted with randomisation restrictions.

### 6.7.7 Multiple Responses

In most practical cases there is more than one response. One approach presented by Derringer and Suich (1980) is to develop a desirability function which transforms the multiresponse problem into a single response function by means of mathematical transformations. This method has been extensively studied by Harrington (1965); del Castillo and Montgomery (1993), and del Castillo et al. (1996). However, as Khuri (1999) points out, the known methods only apply to continuous responses, and extensions to GLMs is of vital importance. Not much has appeared in the literature on the effectiveness of the method for the precision of the path of steepest ascent when there are multiple responses and GLMs.

### 6.7.8 Estimated Constraint Functions

As indicated previously, the constraints considered in Chapter 5 are assumed to be known. However, in practice, these constraints are often estimated. While it is possible to determine the path giving the maximum primary response subject to the estimated secondary response being less than a certain value, say  $k$ , it might be more appropriate to optimise the primary response subject to there being a high probability that the actual secondary response is less than  $k$ . This extension would have wide application in a variety of practical situations.

### 6.7.9 GLMs and Second Order Models

The examples studied in Chapter 5 each had a continuous response, which was assumed to follow a normal distribution. It should be possible to extend the method of Ankenman (2003) to cover GLMs, and to determine the precision of a canonical axis using a similar method to that given in section 5.5.

### 6.7.10 Canonical Analysis and Ridge Analysis

Canonical analysis and ridge analysis are complementary techniques for studying the nature of fitted second order response surfaces. The direction of steepest ascent determined by ridge analysis usually converges quite rapidly onto a canonical axis found using canonical analysis. The method of determining the precision of the canonical axis, as given in section 5.5, should be useful for both canonical analysis and ridge analysis. It would also be useful to compare the results of the method with the methods developed by Peterson et al. (2002) and Gilmour and Draper (2003) as applied to ridge analysis.

# Appendices

# Appendix A

## Splus commands for Chapter 2

### A.1 Data set used in chapters 2 and 3

```
B_c(-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1)
D_c(-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1)
FF_c(-1,-1,1,1,1,1,-1,-1,1,1,-1,-1,-1,-1,1,1)
G_c(1,-1,1,-1,-1,1,-1,1,1,-1,1,-1,-1,1,-1,1)
defects_c(56,17,2,4,3,4,50,2,1,0,3,12,3,4,0,0)
trresp_c(7.52,4.18,1.57,2.12,1.87,2.12,7.12,1.57,
1.21,0.50,1.87,3.54,1.87,2.12,0.50,0.50) BGplus_(B+B*G)/2
BGminus_(B-B*G)/2
chap23ex_as.data.frame(cbind(B,D,FF,G,BGplus,BGminus,
defects,trresp))
```

### A.2 Analysis using transformed response and original design

The following command was used to analyse data in Table 2.1 using the transformed response (left panel).

```
ch24leftfit_lm(trresp~D+FF+B+G+B:G,data=chap23ex)
```

## A.3 Analysis using transformed response and modified design

The following command was used to analyse data in Table 2.1 using the transformed response (right panel).

```
ch24rightfit_lm(trresp~D+FF+BGplus+BGminus+G,data=chap23ex)
```

## A.4 Gaussian quadrature functions

The following functions were used to obtain Gaussian quadrature for three dimensional data.

```
gauss1_function(r, th, a = 0.6450, b = 0.5341)
{
  (r * a * b)/sqrt(1 - r^2 * a^2 * (cos(th))^2 - r^2 * b^2 *
    (sin(th))^2)
}

gauss2_function(n)
{
  if(n == 2){
    absc <- c( - sqrt(1/3), sqrt(1/3))}
  else if(n == 3) {
    absc <- c(-0.77459667, 0, 0.77459667)}
  else if(n == 4) {
    absc <- c(-0.86113631, -0.33998104, 0.33998104,
      0.86113631)}
  else if(n == 5) {
    absc <- c(-0.90617985, -0.53846931, 0, 0.53846931,
      0.90617985)}
  else if(n == 6) {
    absc <- c(-0.93246951, -0.66120939, -0.23861918,
      0.23861918,0.66120939, 0.93246951)}
  else if(n == 7) {
```

```
    absc <- c(-0.94910791, -0.74153119, -0.40584515,
              0, 0.40584515, 0.74153119, 0.94910791)}
else if(n == 8) {
    absc <- c(-0.96028986, -0.79666648, -0.52553241,
              -0.18343464, 0.18343464, 0.52553241, 0.79666648,
              0.96028986)}
else if(n == 10) {
    absc <- c(-0.97390653, -0.86506337, -0.67940957,
              -0.43339539, -0.14887434, 0.14887434, 0.43339539,
              0.67940957, 0.86506337, 0.97390653)}
if(n == 2) {
    weight <- c(1, 1)
}
else if(n == 3) {
    weight <- c(0.55555555, 0.88888888, 0.55555555)
}
else if(n == 4) {
    weight <- c(0.34785485, 0.65214515, 0.65214515,
              0.34785485)}
else if(n == 5) {
    weight <- c(0.23692689, 0.47862867, 0.56888889,
              0.47862867, 0.23692689)}
else if(n == 6) {
    weight <- c(0.17132449, 0.36076157, 0.46791393,
              0.46791393, 0.36076157, 0.17132449)}
else if(n == 7) {
    weight <- c(0.12948497, 0.27970539, 0.38183005,
              0.41795918, 0.38183005, 0.27970539, 0.12948497)}
else if(n == 8) {
    weight <- c(0.10122854, 0.22238103, 0.31370665,
              0.36268378, 0.36268378, 0.31370665, 0.22238103,
              0.10122854)}
else if(n == 10) {
    weight <- c(0.06667134, 0.14945135, 0.21908636,
              0.26926672, 0.29552422, 0.29552422, 0.26926672,
              0.21908636, 0.14945135, 0.06667134)}
else if(n == 32) {
    weight <- c(0.00701814576495, 0.0162774265831,
              0.0253910098329, 0.0342745478477,
              0.0428359896785, 0.0509978738117,
              0.0586839394615, 0.0658220603578,
```

```

                                0.0723456094297, 0.078193695762,
                                0.083311711103, 0.0876518688047,
                                0.0911736454878, 0.0938441590423,
                                0.0956384754512, 0.0965398415811,
                                0.0956384754512, 0.0938441590423,
                                0.0911736454878, 0.0876518688047,
                                0.083311711103, 0.078193695762,
                                0.0723456094297, 0.0965398415811,
                                0.0658220603578, 0.0586839394615,
                                0.0509978738117, 0.0428359896785,
                                0.0342745478477, 0.0253910098329,
                                0.0162774265831, 0.00701814576495)
    }
    weightmat <- matrix(0, n, n)
    abscmat <- matrix(0, n, n)
    weightmat <- outer(weight, weight, "*")
    abscmat <- outer(0.5 + 0.5 * absc, pi + pi * absc, "gauss1")
    sa <- (2 * pi^(1.5))/gamma(1.5)
    (2 * pi)/(4 * sa) * sum(weightmat * abscmat)
}

```

## A.5 perinc

The following function takes a linear model object and determines the semiaxes of the ellipsoid formed when projecting the confidence cone for the path of steepest ascent onto the plane orthogonal to the path. For two dimensions, it determines the exact proportion included in the confidence cone, and approximates it using numerical integration, or Monte-Carlo integration for higher dimensions, as well as giving upper and lower bounds.

```

perinc_function(lmobject, which=-1, mcssize=1000)
{
    b1_lmobject$coefficients[which]
    Gmat_solve(summary(lmobject)$cov.unscaled[which, which])
    df_lmobject$df.residual
}

```

```

sigma_summary(lmobject)$sigma
cval_t(b1)%*%Gmat%*%b1-(length(b1)-1)*sigma^2*qf(0.95,
length(b1)-1,df)
Hmat_Gmat%*%cbind(b1)%*%rbind(b1)%*%Gmat-cval[1,1]*Gmat
eigs_-sort(-eigen(Hmat)$values)
print(eigs)
a_sqrt(eigs[1]/(eigs[1]-eigs[-1]))
print(a)
if (length(b1)==2)
{
  pinc1_asin(a[1])/pi
  cat("Percent Included=",round(1000*100*pinc1)/1000,
      "Percent Excluded=",round(1000*100*(1-pinc1))/1000,"\n")
}
else if (length(b1)==3)
{
  lowerbound_a[1]/(2*a[2])*(1-sqrt(1-a[2]^2))
  upperbound_a[2]/(2*a[1])*(1-sqrt(1-a[1]^2))
  ff_function(theta,a1,a2){
r_1/sqrt(cos(theta)^2/a1^2+sin(theta)^2/a2^2)
pt(sqrt(2)*sqrt(1-r^2)/r,2)}
pinc1_1-integrate(ff,lower=0,upper=2*pi,a1=a[1],
a2=a[2])$integral/(2*pi)
cat("Lower Bound=",round(1000*100*lowerbound)/1000,
    "Upper Bound=",round(1000*100*(upperbound))/1000,"\n")
cat("Percent Included=",round(1000*100*pinc1)/1000,
    "Percent Excluded=",round(1000*100*(1-pinc1))/1000,
    "\n")
}
else if (length(b1)==4)
{
lowerbd4d_(((pi*a[1]*a[3])/(a[2]))*((1/a[3])*
atan(1/((1-2*a[3]^2)/(2*a[3]*
sqrt(1-a[3]^2)))))-(2*(1-a[2]^2))/
(sqrt(a[2]^2-a[3]^2)))*
asinh(sqrt((a[2]^2-a[3]^2)/(1-a[2]^2))))/(2*pi^2)
upperbd4d_(((pi*a[2]*a[3])/(a[1]))*((1/a[3])*atan(1/
((1-2*a[3]^2)/(2*a[3]*sqrt(1-a[3]^2)))))-(2*
(1-a[1]^2))/sqrt(a[1]^2-a[3]^2))*asinh(sqrt(
(a[1]^2-a[3]^2)/(1-a[1]^2))))/(2*pi^2)
x_matrix(rnorm(4*mcssize),mcssize,4)

```



```

    z_x/sqrt(apply(x^2,1,"sum"))
    pinc1_nrow(z[apply(z[,-4]^2/a^2,1,"sum") <=1 &
    z[,4]>0,] )/mcssize
    cat("Lower Bound=",round(1000*100*lowerbound)/1000,
    "Upper Bound=",round(1000*100*(upperbound))/1000,"\n")
    cat("Approximate Percent Included=",round(1000*100*
    pinc1)/1000,"Approximate Percent Excluded=",
    round(1000*100*(1-pinc1))/1000,"\n")
  }
else
{
  lowerbd4d_1-pt(sqrt(length(b1)-1)*sqrt(1-
a[length(b1)-1]^2)/a[length(b1)-1],length(b1)-1)
  upperbd4d_1-pt(sqrt(length(b1)-1)*sqrt(1-a[1]^2)/
  a[1],length(b1)-1)
  x_matrix(rnorm(mcssize*length(b1)),mcssize,
  length(b1))
  z_x/sqrt(apply(x^2,1,"sum"))
  pinc1_nrow(z[apply(z[,-length(b1)]^2/a^2,1,
  "sum") <=1 & z[,length(b1)]>0,] )/mcssize
  cat("Lower Bound=",round(1000*100*lowerbound)/1000,
  "Upper Bound=",round(1000*100*(upperbound))/1000,"\n")
  cat("Approximate Percent Included=",round(
  1000*100*pinc1)/1000,"Approximate Percent Excluded=",
  round(1000*100*(1-pinc1))/1000,"\n")
}
pinc1
}

```

### Examples of Use:

```

perinc(ch24leftfit,which=2:3)
perinc(ch24rightfit,which=2:4)

```

# Appendix B

## Splus commands for Chapter 3

### B.1 Analysis using GLMs and original design

The following command was used to analyse data in Table 2.1 using GLMs (left panel).

```
ch31fit_glm(defects~D+FF+B+G+B:G,family=poisson,data=chap23ex,  
family=poisson)  
summary(ch31fit,cor=F)
```

### B.2 Analysis using GLMs and modified design

The following command was used to analyse data in Table 2.1 using GLMs (right panel).

```
ch32fit_glm(defects~D+FF+BGplus+BGminus+G,family=poisson,  
data=chap23ex)  
summary(ch32fit,cor=F)
```

### B.3 perincglm

The following function takes a generalised linear model object and determines the semiaxes of the ellipsoid formed when projecting the confidence cone for the path of steepest ascent onto the plane orthogonal to the path. For two dimensions, it determines the exact proportion included in the confidence cone and approximates it using numerical integration, or Monte-Carlo integration for higher dimensions, as well as giving upper and lower bounds.

```
perincglm_function(glmobject, which=-1, overdispersion=F){
  b1_glmobject$coefficients[which]
  Gmat_solve(summary(glmobject)$cov.unscaled[which, which])
  df_glmobject$df.residual
  if (overdispersion==T){
    cval_t(b1)%*%Gmat%*%b1-(length(b1)-1)*glmobject$deviance*
    qf(0.95, length(b1)-1, df)/df}
  else {cval_t(b1)%*%Gmat%*%b1-qchisq(0.95, length(b1)-1)}
  Hmat_Gmat%*%cbind(b1)%*%rbind(b1)%*%Gmat-cval[1, 1]*Gmat
  eigs_-sort(-eigen(Hmat)$values)
  print(eigs)
  a_sqrt(eigs[1]/(eigs[1]-eigs[-1]))
  print(a)
  if (length(b1)==2)
    {pinc1_asin(a[1])/pi
    cat("Percent Included=", round(1000*100*pinc1)/1000,
      "Percent Excluded=", round(1000*100*(1-pinc1))/1000,
      "\n")}
  else if (length(b1)==3)
    {lowerbound_a[1]/(2*a[2])*(1-sqrt(1-a[2]^2))
    upperbound_a[2]/(2*a[1])*(1-sqrt(1-a[1]^2))
    ff_function(theta, a1, a2){
      r_1/sqrt(cos(theta)^2/a1^2+sin(theta)^2/a2^2)
      pt(sqrt(2)*sqrt(1-r^2)/r, 2)
    }
    pinc1_1-integrate(ff, lower=0, upper=2*pi, a1=a[1],
      a2=a[2])$integral/(2*pi)
    cat("Lower Bound=", round(1000*100*lowerbound)/1000,
      "Upper Bound=", round(1000*100*(upperbound))/1000,
```

```

"\n")
cat("Percent Included=",round(1000*100*pinc1)/1000,
"Percent Excluded=",round(1000*100*(1-pinc1))/1000,
"\n")
}
else if (length(b1)==4)
{lowerbd4d_(((pi*a[1]*a[3])/(a[2]))*((1/a[3])*
atan(1/((1-2*a[3]^2)/(2*a[3]*sqrt(1-a[3]^2))))-
(2*(1-a[2]^2))/(sqrt(a[2]^2-a[3]^2))*
asinh(sqrt((a[2]^2-a[3]^2)/(1-a[2]^2)))))/(2*pi^2)
upperbd4d_(((pi*a[2]*a[3])/(a[1]))*((1/a[3])*
atan(1/((1-2*a[3]^2)/(2*a[3]*sqrt(1-a[3]^2))))-(2*
(1-a[1]^2))/(sqrt(a[1]^2-a[3]^2))*asinh(sqrt
((a[1]^2-a[3]^2)/(1-a[1]^2)))))/(2*pi^2)
x_matrix(rnorm(4*mcssize),mcssize,4)
z_x/sqrt(apply(x^2,1,"sum"))
pinc1_nrow(z[apply(z[,4]^2/a^2,1,"sum") <=1 &
z[,4]>0,] )/mcssize
cat("Lower Bound=",round(1000*100*lowerbound)/
1000,"Upper Bound=",
round(1000*100*(upperbound))/1000,"\n")
cat("Approximate Percent Included=",round
(1000*100*pinc1)/1000,
"Approximate Percent Excluded=",
round(1000*100*(1-pinc1))/1000,"\n")
}
else {
lowerbd4d_1-pt(sqrt(length(b1)-1)*
sqrt(1-a[length(b1)-1]^2)/a[length(b1)-1],
length(b1)-1)
upperbd4d_1-pt(sqrt(length(b1)-1)*sqrt(1-a[1]^2)/
a[1],length(b1)-1)
x_matrix(rnorm(mcssize*length(b1)),mcssize,
length(b1))
z_x/sqrt(apply(x^2,1,"sum"))
pinc1_nrow(z[apply(z[,-length(b1)]^2/a^2,1,
"sum") <=1 & z[,length(b1)]>0,] )/mcssize
cat("Lower Bound=",round(1000*100*
lowerbound)/1000,"Upper Bound=",round(1000*100*
(upperbound))/1000,"\n")
cat("Approximate Percent Included=",

```

```

round(1000*100*pinc1)/1000,
"Approximate Percent Excluded=",
round(1000*100*(1-pinc1))/1000,"\n")
}
pinc1}
}

```

#### Examples of Use:

```

perincglm(ch31fit,which=2:3)
perincglm(ch31fit,which=2:3,overdispersion=T)
perincglm(ch32fit,which=2:4)
perincglm(ch32fit,which=2:4,overdispersion=T)

```

## B.4 profpath

The following function takes a generalised model object and computes profile likelihood confidence regions for the path of steepest ascent.

```

profpath_function(dataframe,which,others,response,
lev=c(test$deviance+qchisq(.95,2),test$deviance+
(test$deviance/test$df.residual)*(length(which)-1)*
qf(.95,length(which)-1,test$df.residual)))
{
  whichothers_c(which,others)
  y_as.vector(dataframe[,response])
  x_as.matrix((dataframe[,whichothers]))
  test_glm(y~x,family=poisson)
  test$deviance
  e3_c(0,0,1)
  gamma_test$coefficients[2:(length(which)+1)]
  u_gamma=sqrt(t(gamma)%*%gamma)*e3

```

```

u_u/sqrt(t(u)%*%u)
Hu3d_diag(3)-2*u%*%t(u)
gam1_seq(-.5,.5,.01)
gam2_gam1
dev_matrix(0,length(gam1),length(gam2))
gam3_dev
for (i in 1:length(gam1))
{
  for (j in 1:length(gam2))
  {
    cat(i,j,"\n")
    gam3[i,j]_sqrt(1-gam1[i]^2-gam2[j]^2)
    gamvec_c(gam1[i],gam2[j],gam3[i,j])
    dev[i,j]_glm(y~I(x[,1:(length(which))])%*%(Hu3d)%*%
    gamvec)+x[-(1:(length(which)))],
    family=poisson)$deviance
  }
}
lev_c(test$deviance+qchisq(.95,2),test$deviance+
(test$deviance/test$df.residual)*(length(which)-1)*
qf(.95,length(which)-1,test$df.residual))
profcont_contour(gam1,gam2,dev,levels=lev,save=T,plotit=T)
}

profcont_profpath(chap23ex,which=c(2,3,5),others=c(4,6),
response=7)
profcont1_profpath(chap23ex,which=c(2,3,5),others=c(4,6),
response=7,lev=c(22.599,23.045,30.236,37.437))
profcont1[[4]]

```

#### Examples of Use:

```
profpath(chap23ex,which=c(2,3,5),others=c(4,6), response=7)
```

## B.5 contourprop

The following function takes a contour object and computes the approximate proportion of directions included in the confidence region for the path of steepest ascent for three dimensional models.

```
contourprop_function(contourobject,which)
{
  x1c_contourobject[[which]]$x
  x2c_contourobject[[which]]$y
  x1c_x1c[!is.na(x1c)]
  x2c_x2c[!is.na(x2c)]
  th_atan(x1c,x2c)
  ra_ra[order(th)]
  th_th[order(th)]
  ra_sqrt(x1c^2+x2c^2)
  parsum_rep(0,length(th))
  for (i in 2:length(th))
  {
    parsum[i]_parsum[i-1]+(th[i]-th[i-1])/(2*pi)*
      (1-0.5*(pt(sqrt(2)*sqrt(1-ra[i-1]^2)/ra[i-1],2)+
        pt(sqrt(2)*sqrt(1-ra[i]^2)/ra[i],2)))
  }
  parsum[length(th)]
}
```

### Examples of Use:

```
contourprop(profcont,2)
```

## B.6 boot3

The following function takes a linear model object and performs parametric bootstrap simulations in order to determine correct contour level to get the appropriate coverage.

```
boot3_function(dataframe,which,others,response) {
  whichothers_c(which,others)
  y_as.vector(dataframe[,response])
  x_as.matrix((dataframe[,whichothers]))
  test_glm(y~x,family=poisson)
  x1_as.matrix((dataframe[,which]))
  x2_as.matrix((dataframe[,others]))
  y_rpois(test$residuals,test$fitted.values)
  deviance1_glm(y~x,family=poisson)$deviance
  deviance2_glm(y~I(x1%*%test$coefficients[2:(length(which)
+1)])+x2,family=poisson)$deviance
  deviance2-deviance1
} bootres3_rep(0,999) for (i in 1:999) {
  cat(i,"\n")
  bootres3[i]_boot3(chap23ex,which=c(2,3,5),others=c(4,6),
  response=7)
}
plot(qchisq(seq(.001,.999,.001),2),sort(bootres3),
xlab="Chi-squared(2) quantiles",ylab="Bootstrap deviance")
abline(a=0,b=1)
abline(v=qchisq(.95,2))
sort(bootres3)[950]
qchisq(seq(.001,.99,.001),2)
```



## B.7 boot4

The following function takes a linear model object and performs nonparametric bootstrap simulations in order to determine correct contour level to get the appropriate coverage.

```
boot4_function(dataframe,which,others,response)
{
  whichothers_c(which,others)
  y_as.vector(dataframe[,response])
  x_as.matrix((dataframe[,whichothers]))
  test_glm(y~x,family=poisson,dispersion=0)
  pr_resid(test,type="pearson")
  prstand_pr/summary(test)$dispersion
  mapr_prstand=mean(prstand)
  rs_sample(mapr,length(y))
  muhat_predict(test,type="response")
  y_round(muhat+sqrt(summary(test)$dispersion*muhat)*rs)
  y[y<0]_0
  x1_as.matrix((dataframe[,which]))
  x2_as.matrix((dataframe[,others]))
  deviance1_glm(y~x,family=poisson)$deviance
  deviance2_glm(y~I(x1%*%test$coefficients[2:(length(which)+1)])
  +x2,family=poisson)$deviance
  deviance2-deviance1
}
bootres4_rep(0,999)
for (i in 1:999)
{
  cat(i,"\n")
  bootres4[i]_boot4(chap23ex,which=c(2,3,5),others=c(4,6),
  response=7)
} plot(qf(seq(.001,.999,.001),2,10),sort(bootres4), xlab="F(2,10)
quantiles",ylab="Bootstrap deviance")
abline(a=0,b=1)
abline(v=qf(.95,2,10) )
```

# Appendix C

## Splus commands for Chapter 4

### C.1 powerfn

Power function for proportion of directions included in  $100\alpha\%$  confidence cone for the path of steepest ascent. The arguments are:

$k$	dimension
$1 - \alpha$	confidence level
nub	residual degrees of freedom
$N$	sample size
$p$	proportion of directions
sigmabetasq	$\sum_i \beta_i^2$
sigmasq	$\sigma^2$

```
powerfn_function(k,alpha,N,nub,p,sigmabetasq,sigmasq){
  1-pf(qf(1-alpha,k-1,nub)*
    (qt(1-p,k-1)^2+(k-1))/k,k-1,nub,N*sigmabetasq/sigmasq)}
```

Examples of Use:

```
powerfn(5,.05,32,26,.02,2,1)
```

## pden

Density function for percentage included for homoscedastic data. Note that the function uses numerical derivatives since Splus does not give density of noncentral  $F$ .

```
pden_function(gamma,k,nub,ncp,h=.000001)
{
  X_qf(.95,k-1,nub)*(qt(1-gamma,k-1)^2+(k-1))
  gX_(pf(X+h,k-1,nub,ncp)-pf(X,k-1,nub,ncp))/
  h-2*gX*qf(.95,k-1,nub)*(-1*qt(1-gamma,k-1))/
  dt(qt(1-gamma,k-1),k-1)
}
```

## C.2 approxp

Approximate means and standard deviations for percent included for homoscedastic data.

```
approxp_function(k,nub,lambda)
{
  approxmean_((lambda+k-1)*nub)/((k-1)*(nub-2))
  approxvar_2*nub^2*((k-1)^2+(2*lambda+nub-2)*(k-1)+lambda*(
  lambda+2*nub-4))/((k-1)^2*(nub-4)*(nub-2)^2)
  pmean_pt(sqrt(approxmean/qf(.95,k-1,nub)-(k-1)),k-1)
  psd_sqrt(approxvar)*dt(sqrt(approxmean/qf(.95,k-1,nub)-(k-1)),
  k-1)/( 2*qf(.95,k-1,nub)*sqrt(approxmean/
  qf(.95,k-1,nub)-(k-1)))
  list(pmean,psd)
}
```

Examples of Use:

```
approxp(2,10,30)
```

### C.3 heterofn

Function to give expected proportion of directions included in 95% confidence cone for the path of steepest ascent, when the angle between the  $2^2$  design and the mean function is  $\theta$ , and the angle between the design and the variance function is  $\phi$ . The magnitude of the mean coefficients is  $\text{meannorm}$ , while the magnitude of the variance coefficients is  $\text{heteronorm}$ .

```
heterofn_function(theta, phi, meannorm, heteronorm)
{
  xdes <- matrix(c(1, 1, 1, 1, -1, 1, -1, 1, -1, -1, 1, 1), 4, 3)
  vmat <- diag(exp(heteronorm * cos(phi) * xdes[, 2] +
    heteronorm * sin(phi) * xdes[, 3]))
  bvmat <- solve(t(xdes) %*% solve(vmat) %*% xdes)
  gmat <- solve(bvmat[2:3, 2:3])
  bb <- c(meannorm * cos(theta), meannorm * sin(theta))
  hmat <- gmat %*% bb %*% t(bb) %*% gmat - as.vector(t(bb) %*%
    gmat %*% bb - qchisq(0.95, 1)) * gmat
  evals <- rev(sort(eigen(hmat)$values))
  a1 <- sqrt(evals[1]/(evals[1] - evals[2]))
  asin(a1)/pi
}
```

Examples of Use:

```
heterofn(pi/4, pi/6, 5, 1.51293)
```

## C.4 heteropowerlik

Function to give proportion of directions included in the 95% confidence cone for the path of steepest ascent with two dimensions, when only the first factor has a non-zero effect, the angle between the  $2^3$  design and the mean function is theta, and the angle between the design and the variance function is phi. The magnitude of the mean coefficients is meannorm. while the magnitude of the variance coefficients is heteronorm. The proportion of directions is determined using profile likelihood techniques.

```
heteropowerlik_function(bnorm, theta, heteronorm, phi)
{
  xdes<-matrix(c(1, 1, 1, 1, 1, 1, 1, 1, -1, 1, -1, 1, -1, 1,
    -1, 1, -1, -1, 1, 1, -1, -1, 1, 1), 8, 3)
  vmat<-(exp(heteronorm*cos(phi)*xdes[,2]+heteronorm*
    sin(phi)*xdes[,3]))
  vmat<-diag(vmat)
  wt<-1/diag(vmat)
  y<-bnorm*cos(theta)*xdes[,2]+bnorm*sin(theta)*
    xdes[,3]+rnorm(8)/sqrt(wt)
  genlinmod <- summary(glm(y ~ xdes - 1, weights = wt))
  bbl <- genlinmod$coefficients[2:3]
  disp <- genlinmod$dispersion
  list(bbl,disp)
  e2_c(0,1)
  u_bbl=sqrt(t(bbl)%*%bbl)*e2
  u_u/sqrt(t(u)%*%u)
  Hu2d_diag(2)-2*u%*%t(u)
  gam1_seq(-1,1,.01)
  dev_rep(0,length(gam1))
  gam2_dev
  for (i in 1:length(gam1))
  {
    gam2[i]_sqrt(1-gam1[i]^2)
    gamvec_c(gam1[i],gam2[i])
    dev[i]_glm(y~I(xdes[,2:3]\%*\%(Hu2d)\%*\%gamvec),weights=wt)$
    deviance
  }
}
```

```

plot(gam1,dev-(disp*qf(0.95,1,5)+disp))
abline(h=0)
a1_approx(dev[1:101],seq(-1,0,.01),disp*qf(0.95,1,5)+disp)$y
a2_approx(dev[101:201],seq(0,1,.01),disp*qf(0.95,1,5)+disp)$y
pinc_(asin(a2)-asin(a1))/(2*pi)
pinc
}

```

### Examples of Use:

```
heteropowerlik(4,0,1.51293,pi/6)
```

## C.5 augdet

Determinant criteria for augmenting runs for the example in chapters 2 and 3.

```

augdet_function(x)
{
  xmat_rbind(cbind(Dphd,Fphd,BGplus,rep(1,16),Gphd),
    cbind(matrix(x,length(x)/3,3),matrix(1,length(x)/3,2)))
  umat_diag(exp(as.vector(xmat%%ch32fit$
    coefficients[c(2:4,1,6)])))
  varmat_t(xmat)%%umat%%xmat
  crit_varmat[1:3,1:3]-varmat[1:3,4:5]%%solve(varmat[4:5,4:5])%%
  varmat[4:5,1:3]
  det(crit)
}

```

## C.6 augfun

Minimum proportion of directions included in 95% confidence cone criteria for augmenting runs for the example in chapters 2 and 3.

```

augfun_function(x,overdispersion=F)
{
  xmat_rbind(cbind(Dphd,Fphd,BGplus,rep(1,16),Gphd),
    cbind(matrix(x,length(x)/3,3),matrix(1,length(x)/3,2)))
  umat_diag(exp(as.vector(xmat%%ch32fit$
    coefficients[c(2:4,1,6)])))
  varmat_solve(t(xmat)%%umat%%xmat)
  G11p_varmat[1:3,1:3]
  Gmatp_solve(G11p)
  b1p_ch32fit$coefficients[2:4]
  if (overdispersion==T)
  {
    df_10+length(x)/3
    cval_t(b1p)%%Gmatp%%b1p-(length(b1p)-1)*ch32fit$deviance*
      qf(0.95,length(b1p)-1,df)/df
  }
  else
  {
    cval_t(b1p)%%Gmatp%%b1p-qchisq(0.95,length(b1p)-1)
  }
  Hmatp_Gmatp%%b1p%%t(b1p)%%Gmatp-as.vector(cval)*Gmatp
  evals_rev(sort(eigen(Hmatp)$values))
  a1_evals[1]/(evals[1]-evals[2])
  a2_evals[1]/(evals[1]-evals[3])
  a_c(a1,a2)
  ff_function(theta,a1,a2)
  {
    r_1/sqrt(cos(theta)^2/a1^2+sin(theta)^2/a2^2)
    pt(sqrt(2)*sqrt(1-r^2)/r,2)
  }
  pinc1_1-integrate(ff,lower=0,upper=2*pi,
    a1=a[1],a2=a[2])$integral/(2*pi)

  pinc1
}

```

## C.7 Commands to determine best set of augmenting runs

Commands to determine best set of augmenting runs with determinant criteria and minimum percentage included criteria, with and without overdispersion. for 1, 2 and 3 runs.

```
yd1_nlminb(start=runif(3,-1,+1),objective=augdet,
lower=-1,upper=1)
yd1$parameters
augfun(yd1$parameters)
augfun(yd1$parameters,overdispersion=T)
yf1_nlminb(start=runif(3,-1,+1),objective=augfun,
lower=-1,upper=1)
yf1$parameters
augfun(yf1$parameters)
yf1o_nlminb(start=runif(3,-1,+1),objective=augfun,
lower=-1,upper=1,
overdispersion=T)
yf1o$parameters
augfun(yf1o$parameters,overdispersion=T)
yd2_nlminb(start=runif(6,-1,+1),objective=augdet,
lower=-1,upper=1)
yd2$parameters
augfun(yd2$parameters)
augfun(yd2$parameters,overdispersion=T)
yf2_nlminb(start=runif(6,-1,+1),objective=augfun,
lower=-1,upper=1)
yf2$parameters
augfun(yf2$parameters)
yf2o_nlminb(start=runif(6,-1,+1),objective=augfun,
lower=-1,upper=1,
overdispersion=T)
yf2o$parameters
augfun(yf2o$parameters,overdispersion=T)
yd3_nlminb(start=runif(9,-1,+1),objective=augdet,
lower=-1,upper=1)
yd3$parameters
augfun(yd3$parameters)
```



```

augfun(yd3$parameters, overdispersion=T)
yf3_nlmminb(start=runif(9,-1,+1), objective=augfun,
lower=-1, upper=1)
yf3$parameters
augfun(yf3$parameters)
yf3o_nlmminb(start=runif(9,-1,+1), objective=augfun,
lower=-1, upper=1,
overdispersion=T)
yf3o$parameters
augfun(yf3o$parameters, overdispersion=T)

```

## C.8 poisfund

Function to give the negative of the determinant criteria for a Poisson experiment for the example in section 4.6.1.

```

poisfund_function(x, a, b, cc)
{
  xmat <- matrix(c(rep(1, length(x)/2), x), length(x)/2, 3)
  lp <- a + b * xmat[, 2] + cc * xmat[, 3]
  lambdap <- exp(lp)
  umat <- diag(lambdap)
  varmat <- t(xmat) %*% umat %*% xmat
  crit <- varmat[2:3, 2:3] - varmat[2:3, 1, drop=F] %*%
    varmat[1, 2:3, drop = F]/varmat[1, 1]
  -det(crit)}

```

## C.9 poisfunp

Function to give the percentage of directions included in the confidence region for the path of steepest ascent for the example in section 4.6.1.

```

poisfunp_function(x, a, b, cc, nn)
  {xmat <- matrix(c(rep(1, length(x)/2), x),
    length(x)/2, 3)
  lp <- a + b * xmat[, 2] + cc * xmat[, 3]
  lambdap <- exp(lp)
  umat <- diag(lambdap)
  varmat <- solve(t(xmat)%*%umat%*%xmat)
  G11p <- varmat[2:3, 2:3]
  Gmatp <- solve(G11p)
  b1p <- c(b, cc)
  ccalc <- t(b1p)%*% Gmatp%*% b1p-qchisq(0.95,1)
  Hmatp <- Gmatp %*% b1p%*%t(b1p)%*%
    Gmatp-as.vector(ccalc)*Gmatp
  evals <- rev(sort(eigen(Hmatp)$values))
  a1 <- sqrt(evals[1]/(evals[1] - evals[2]))
  pinc <- (2 * asin(a1))/(2 * pi)
  pinc}

```

## C.10 poissim

Function to compare D-optimal and minimum percentage designs for Poisson data.

```

poissim_function(aval, bval, ccval, nruns)
{  min1 <- 1
  min2 <- 1
  min3 <- 1
  for(ii in 1:50) {
    cat(ii)
    ydval <- nlminb(start = runif(2 * nruns, -1,1),
      objective = poisfund, lower = -1,
      upper = 1, a = aval, b = bval,
      cc = ccval)
    cat("ydval$parameters", "\n")
    print(matrix(ydval$parameters, nruns, 2))
  }
}

```

```

print(poisfunpred(ydval$parameters,
a = aval, b = bval, cc = ccval))
cat("pinc for ydval$parameters", "\n")
print(poisfunp(ydval$parameters, a = aval,
b = bval, cc = ccval, nn = nnval))
current <- poisfund(ydval$parameters,
a = aval, b = bval, cc = ccval)
cat(ii, min1, current, "\n")
min1 <- min(min1, current, na.rm = T)
if(abs(current - min1) < 1e-005 | ii == 1) {
cat("current D-optimal", ii, "\n")
min3 <- poisfunp(ydval$parameters,
a = aval, b = bval, cc = ccval)}
ypval <- nlminb(ydval$parameters,
objective = poisfunp, lower = -1, upper = 1,
a = aval, b = bval, cc = ccval,
trace = T)
print(matrix(ypval$parameters, nruns, 2))
print(poisfunpred(ypval$parameters,
a = aval, b = bval, cc = ccval))
cat("pinc for ypval$parameters", "\n")
print(poisfunp(ypval$parameters, a = aval,
b = bval, cc = ccval, nn = nnval))
min2 <- min(min2, poisfunp(ypval$parameters,
a = aval, b = bval, cc = ccval), na.rm = T)
}
cat(min3, min2, min2/min3, "\n")
min2/min3
}

```

## C.11 bcfn3

Function to determine the efficiency of the D-optimal design relative to the minimum proportion designs for Poisson data using the 100 run orthogonal array given by Sloane (2005), and saved in matrix “orth”.

```
bcfn3_function(lambdahat, u, v){
  a <- log(lambdahat)
  cc <- u/sqrt(1 + tan(v)^2)
  b <- cc * tan(v)
  poissim(a, b, cc,4)}
```

#### Examples of Use:

```
pres_rep(NA,100)
for (i in 1:100){
  cat(i,"\n")
res_try(bcfn3(5+20/9*orth[i,1],0.5+2/9*orth[i,2],
  pi/36*orth[i,3]))
if(!is.null(res)){pres[i]_res}}
hist(pres)
```

## C.12 binfund

Function to give the negative of the determinant criteria for a Binomial experiment for the example in section 4.6.2.

```
binfund_function(x, a, b, cc, nn)
{
  xmat <- matrix(c(rep(1, length(x)/2), x), length(x)/2, 3)
  lp <- a + b * xmat[, 2] + cc * xmat[, 3]
  pp <- exp(lp)/(1 + exp(lp))
  umat <- diag(nn * pp * (1 - pp))
  varmat <- t(xmat) %*% umat %*% xmat
  crit <- varmat[2:3, 2:3] - varmat[2:3, 1, drop
= F] %*% varmat[1, 2:3, drop = F]/varmat[1, 1]
  - det(crit)}
```

### C.13 binomfunp

Function to give the percentage of directions included in the confidence region for the path of steepest ascent for the example in section 4.6.2.

```
binomfunp_function(x, a, b, cc, nn)
{
  xmat <- matrix(c(rep(1, length(x)/2), x), length(x)/2, 3)
  lp <- a + b * xmat[, 2] + cc * xmat[, 3]
  pp <- exp(lp)/(1 + exp(lp))
  umat <- diag(nn * pp * (1 - pp))
  varmat <- solve(t(xmat) %*% umat %*% xmat)
  G11p <- varmat[2:3, 2:3]
  Gmatp <- solve(G11p)
  b1p <- c(b, cc)
  ccalc <- t(b1p) %*% Gmatp %*% b1p - qchisq(0.95, 1)
  Hmatp <- Gmatp %*% b1p %*% t(b1p) %*% Gmatp
  - as.vector(ccalc) * Gmatp
  evals <- rev(sort(eigen(Hmatp)$values))
  a1 <- sqrt(evals[1]/(evals[1] - evals[2]))
  pinc <- (2 * asin(a1))/(2 * pi)
  pinc}
```

### C.14 binsim

Function to compare D-optimal and minimum percentage designs for Binomial data.

```
binsim_function(aval, bval, ccval, nnval, nruns)
{
  min1 <- 1
  min2 <- 1
  min3 <- 1
  for(ii in 1:50) {
    cat(ii)
    ydval <- nlminb(start = runif(2 * nruns, -1,1),
      objective = binfund, lower = -1, upper =1,
```

```

    a = aval, b = bval, cc= ccval, nn = nnval)
cat("ydval$parameters", "\n")
print(matrix(ydval$parameters, nruns, 2))
print(binomfunpred(ydval$parameters,
  a =aval,b = bval,cc = ccval, nn = nnval))
cat("pinc for ydval$parameters", "\n")
print(binomfunp(ydval$parameters,
  a = aval, b = bval, cc =ccval, nn = nnval))
current <- binfund(ydval$parameters,
  a = aval, b = bval, cc =cval, nn = nnval)
cat(ii, min1, current, "\n")
min1 <- min(min1, current, na.rm = T)
if(abs(current - min1) < 1e-005 | ii == 1) {
  cat("current D-optimal", ii, "\n")
  min3 <- binomfunp(ydval$parameters,
    a = aval, b = bval,cc = ccval, nn = nnval)}
ypval <- nlminb(ydval$parameters,
objective = binomfunp, lower=-1,upper = 1,
  a = aval, b = bval, cc = ccval, nn = nnval,
  trace = T)
print(matrix(ypval$parameters, nruns, 2))
print(binomfunpred(ypval$parameters,
  a = aval, b = bval, cc = ccval, nn = nnval))
cat("pinc for ypval$parameters", "\n")
print(binomfunp(ypval$parameters,
  a = aval, b = bval, cc = ccval, nn = nnval))
min2 <- min(min2, binomfunp(ypval$parameters,
  a = aval, b = bval, cc = ccval, nn=nnval),na.rm = T)}
cat(min3, min2, min2/min3, "\n")
min2/min3}}

```

#### Examples of Use:

```
binsim(log(0.6/(1-0.6)),sqrt(5),2*sqrt(5),50,4)
```

## C.15 bcf4

Function to determine the efficiency of D-optimal design relative to the minimum proportion designs for Binomial data using the 100 run orthogonal array given by Sloane (2005), and saved in matrix “orth”.

```
> bcf4_function(phat, u, v, nn)
{  a <- log(phat/(1 - phat))
   cc <- u/sqrt(1 + tan(v)^2)
   b <- cc * tan(v)
   binsim4(a, b, cc, nn)}
```

### Examples of Use:

```
bres_rep(NA,100)
for (i in 1:100){
  cat(i,"\n")
  bres_try(bcf4(0.2+0.6/9*orth[i,1],1+orth[i,2],
    pi/36*orth[i,3],20+60/9*orth[i,4]))
  if(!is.null(res)){bres[i]_res}}
hist(bres)
```

# Appendix D

## Splus commands for Chapter 5

### D.1 Commands for linear constraints

The following commands were used for the example in section 5.2. First, the proportion of directions included in the 95% confidence cone for the unconstrained path of steepest ascent was calculated.

```
bvec_as.vector(c(3.8,2.6,1.9))
xmat_matrix(1,8,4)
for (i in 2:4){xmat[,i]_rep(c(-1,1),each=2^(i-2),length.out=8)}
Gmat_solve(solve(t(xmat)%*%xmat))[2:4,2:4]
Hmat_Gmat%*%bvec%*%t(bvec)%*%Gmat-as.numeric(t(bvec)%*%Gmat%*%
bvec-qchisq(.95,2))*Gmat
evals_rev( sort(eigen(Hmat)$values))
a1_sqrt(evals[1]/(evals[1]-evals[2]))
1-pt(sqrt(2)*sqrt(1-a1^2)/a1,2)
```

Second, the equation of the projected path was computed.

```
avec_c(5,10,10)
pmat_avec%*%solve(t(avec)%*%avec)%*%t(avec)
(diag(3)-pmat)%*%bvec
```



Finally, the proportion of directions included in the 95% confidence cone for the constrained path of steepest ascent was calculated.

```
e3vec_c(0,0,1)
uvec_avec=sqrt(as.numeric(t(avec)%*%avec)) *e3vec
uvec_uvec/sqrt(as.numeric(t(uvec)%*%uvec))
Hu_diag(3)-2*uvec%*%t(uvec)
bnew_(Hu%*(diag(3)-pmat) %*%bvec)[1:2]
Gmat_(Hu%*(diag(3)-pmat)%*(8*diag(3))%*%
(diag(3)-pmat)%*%Hu)[1:2,1:2]
Hmat_Gmat%*%bnew%*%t(bnew)%*%Gmat-as.numeric(t(bnew)%*%Gmat)%*%
bnew-qchisq(.95,1))*Gmat
evals_rev( sort(eigen(Hmat)$values))
a1_sqrt(evals[1]/(evals[1]-evals[2]))
1-pt(sqrt(1-a1^2)/a1,1)
```

## D.2 Two dimensional example

Two dimensional example given in Box et al. (1978, p.519).

```
x1_c(-1,1,-1,1,0,0,-sqrt(2),sqrt(2),0,0,0,0)
x2_c(-1,-1,1,1,0,0,0,0,-sqrt(2),sqrt(2),0,0)
y2d_c(78.8,84.5,91.2,77.4,89.7,86.8,
      83.3,81.2,81.2,79.5,87.0,86.0)
twoddata_as.data.frame(cbind(x1,x2,y2d))
twodfit_lm(y2d~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=twoddata,y=T)
summary(twodfit,cor=F)
```

## D.3 Three dimensional example

Three dimensional example given Box and Draper (1987, p.360).

```

x1_c(-1,1,-1,1,0,0,-1,1,-1,1,0,0,-sqrt(2),sqrt(2),0,0,0,0,
-sqrt(2),sqrt(2),0,0,0,0)
x2_c(-1,-1,1,1,0,0,-1,-1,1,1,0,0,0,0,-sqrt(2),sqrt(2),0,0,
0,0,-sqrt(2),sqrt(2),0,0)
x3_c(1,-1,-1,1,0,0,-1,1,1,-1,0,0,0,0,0,-sqrt(2),sqrt(2),
0,0,0,0,-sqrt(2),sqrt(2))
y3d_c(40,18.6,53.8,64.2,53.5,52.7,39.5,59.7,42.2,33.6,54.1,
51.0,43.0,43.9,47.0,62.8,25.6,49.7,39.2,46.3,44.9,58.1,27.0,50.7)
blocks_as.factor(c(rep(1,6),rep(2,6),rep(3,6),rep(4,6)))
threeddata_as.data.frame(cbind(blocks,x1,x2,x3,y3d))
threedfit_lm(y3d~as.factor(blocks)+x1+x2+x3+I(x1^2)+I(x2^2)+
I(x3^2)+I(x1*x2)+I(x1*x3)+I(x2*x3),data=threeddata,y=T)
summary(threedfit,cor=F)

```

## D.4 Four dimensional example

Four dimensional example given by Box and Liu (1999)

```

x1_c(-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,0,0,
-2,2,0,0,0,0,0,0,0,0,0,0)
x2_c(-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,-1,-1,1,1,0,0,
0,0,-2,2,0,0,0,0,0,0,0,0)
x3_c(-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1,0,0,
0,0,0,0,-2,2,0,0,0,0,0,0)
x4_c(-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1,0,0,
0,0,0,0,0,0,-2,2,0,0,0,0)
y4d_c(367,369,374,370,372,355,397,377,350,373,358,
363,344,355,370,362,377,375,361,364,355,373,
361,360,380,360,370,368,369,366)
block1_c(rep(1,18),rep(0,12))
block2_c(rep(0,18),rep(1,12))

blocks_as.factor(c(rep(1,18),rep(2,12)))
fourddata_as.data.frame(cbind(block1,block2,x1,x2,x3,x4,y4d))
fourdfit_lm(y4d~block1+block2+x1+x2+x3+x4+I(x1^2)+I(x2^2)+
I(x3^2)+I(x4^2)+I(x1*x2)+I(x1*x3)+I(x1*x4)+I(x2*x3)+

```

```
I(x2*x4)+I(x3*x4)-1,data=fourddata,y=T)
summary(fourdfit,cor=F)
```

## D.5 Five dimensional example

Five dimensional example given by Kiprianova and Markovska (1993, p.83).

```
x1_c(rep(c(1,-1),9),rep(0,8))
x2_c(rep(c(1,1,-1,-1),4),c(0,0,1,-1,0,0,0,0,0,0))
x3_c(rep(c(1,1,1,1,-1,-1,-1,-1),2),c(0,0,0,0,1,-1,0,0,0,0))
x4_c(rep(1,8),rep(-1,8),c(0,0,0,0,0,0,1,-1,0,0))
x5_c(1,-1,-1,1,-1,1,1,-1,-1,1,1,-1,1,-1,1,rep(0,8),1,-1)
y5d_c(4.016,2.753,1.495,3.161,2.407,3.765,2.407,2.737,
1.879,3.216,1.774,3.222,3.514,2.275,3.681,4.540,
2.383,3.017,3.014,2.091,2.552,2.819,2.897,2.547,3.519,2.882)
fiveddata_as.data.frame(cbind(x1,x2,x3,x4,x5,y5d))
fivedfit_lm(y5d~x1+x2+x3+x4+x5+I(x1^2)+I(x2^2)+I(x3^2)+I(x4^2)+
I(x5^2)+I(x1*x2)+I(x1*x3)+I(x1*x4)+I(x1*x5)+I(x2*x3)+I(x2*x4)+
I(x2*x5)+I(x3*x4)+I(x3*x5)+I(x4*x5),data=fiveddata)
summary(fivedfit,cor=F)
```

## D.6 ridgeplot

Function to give ridge plot.

```
ridgeplot_function(lmobject,nblocks=1,mulower,muupper,radupper)
{
  if (nblocks>1)
    betahat_summary(lmobject)$coefficients[,1][-(2:nblocks)]
  else
    betahat_summary(lmobject)$coefficients[,1]
  nn_round((sqrt(1+8*length(betahat))-3)/2)
  bvec_betahat[2:(nn+1)]
```

```

Bmat_matrix(0,nn,nn)
for (i in 1:nn)
{
  Bmat[i,i]_betahat[i+nn+1]
}
pos_2*nn+1
for (i in 1:(nn-1))
  {for (j in (i+1):nn){
    pos_pos+1
    cat(i,j,pos,"\n")
    Bmat[i,j]_0.5*betahat[pos]
    Bmat[j,i]_Bmat[i,j]}}
print(bvec)
print(Bmat)
muval_seq(mulower,muupper,length=1000)
radres_rep(0,1000)
for (i in 1:1000)
  {xs_-0.5*solve(Bmat-muval[i]*diag(nn))%*%bvec
   radres[i]_as.vector(sqrt(t(xs)%*%xs))}
plot(muval,radres,ylim=c(0,radupper),type="l",xlab="mu",
ylab="Radius"))

```

#### Examples of Use:

```

ridgeplot(twodfit,nblocks=1,mulower=-10,muupper=5,radupper=5)
ridgeplot(threedfit,nblocks=4,mulower=-20,muupper=10,radupper=5)
ridgeplot(fourdfit,nblocks=1,mulower=-10,muupper=5,radupper=5)
ridgeplot(fivedfit,nblocks=1,mulower=-1,muupper=1,radupper=5)

```

## D.7 maxplot

Function to give co-ordinates of maximum response versus radius and estimated maximum response.

```

maxplot_function(lmobject,mulower,muupper,nblocks=1){
  if (nblocks>1)
    { betahat_summary(lmobject)$coefficients[,1]
      [-(2:nblocks)]}
    else betahat_summary(lmobject)$coefficients[,1]
nn_round((sqrt(1+8*length(betahat))-3)/2)
bvec_betahat[2:(nn+1)]
Bmat_matrix(0,nn,nn)
for (i in 1:nn) {
  Bmat[i,i]_betahat[i+nn+1]}
pos_2*nn+1
for (i in 1:(nn-1))
  {for (j in (i+1):nn){
    pos_pos+1
    cat(i,j,pos,"\n")
    Bmat[i,j]_0.5*betahat[pos]
    Bmat[j,i]_Bmat[i,j]}}
print(bvec)
print(Bmat)
mu_seq(mulower,muupper,length=1000)
xs_matrix(0,1000,nn)
R_rep(0,1000)
ymax_rep(0,1000)
for (i in 1:length(mu)){
  xcoord_-0.5*solve(Bmat-mu[i]*diag(nn))%*%bvec
  xs[i,]_xcoord
  R[i]_sqrt(t(xcoord)%*%xcoord)
  ymax[i]_betahat[1]+t(xcoord)%*%bvec+t(xcoord)%*%
  Bmat%*%xcoord
  cat(i,ymax[i],"\n"))}
ratio_(ymax-min(ymax))/(max(ymax)-min(ymax))
cat(ratio,"\n")
oldpar_par()
par(mar=c(5.1,4.1,4.1,4.1))
ymaxrange_pretty(range(ymax))
matplot(R,cbind(xs,min(xs)+(max(xs)-min(xs))*
  (ymax-min(ymax))/(max(ymax)-min(ymax))),
  type="l",xlab="Radius",ylab="Ridge Co-ordinates",lty=1,col=1)
axis(side=4,at=min(xs)+(max(xs)-min(xs))*
  (ymaxrange-min(ymax))/(max(ymax)-min(ymax)),
  lab=ymaxrange,srt=90)
mtext("Maximum Response",side=4,line=3)
par(oldpar))

```

Examples of Use:

```

maxplot(twodfit,mulower=-0.025,muupper=5)
maxplot(threedfit,mulower=1.9,muupper=5,nblocks=4)
maxplot(fivedfit,mulower=-5,muupper=-0.41)
maxplot(fourdfit,mulower=3.9,muupper=20,nblocks=2)

```

## D.8 canonical

Function to perform canonical analysis on fitted second degree regression equation, give location of stationary point, and distance of stationary point from origin.

```

canonical_function(lmobject){
  betahat_summary(lmobject)$coefficients[,1]
  nn_round((sqrt(1+8*length(betahat))-3)/2)
  bvec_betahat[2:(nn+1)]
  Bmat_matrix(0,nn,nn)
  for (i in 1:nn){
    Bmat[i,i]_betahat[i+nn+1]}
  pos_2*nn+1
  for (i in 1:(nn-1)){
    for (j in (i+1):nn){
      pos_pos+1
      Bmat[i,j]_0.5*betahat[pos]
      Bmat[j,i]_Bmat[i,j]}}
  xs_-0.5*solve(Bmat)%*%bvec
  R_sqrt(t(xs)%*%xs)
  cat("Canonical Analysis","\n")
  print(eigen(Bmat))
  cat("Location of Stationary Point ")
  cat(xs,"\n")
  cat("Distance of Stationary Point from Origin= ")
  cat(R,"\n")}

```

Examples of Use:

```
canonical{twodfit,nblocks=1}
canonical(threedfit,nblocks=4)
canonical(fourdfit,nblocks=2)
canonical(fivedfit,nblocks=1)
```

**D.9 dlr**

Function to perform double linear regression.

```
dlr_function(lmobject,nblocks=1){
  if (nblocks>1)
    betahat_summary(lmobject)$coefficients[,1][-(2:nblocks)]
  else
    betahat_summary(lmobject)$coefficients[,1]
  nn_round((sqrt(1+8*length(betahat))-3)/2)
  bvec_betahat[2:(nn+1)]
  Bmat_matrix(0,nn,nn)
  for (i in 1:nn){
    Bmat[i,i]_betahat[i+nn+1]}
  pos_2*nn+1
  for (i in 1:(nn-1)){
    for (j in (i+1):nn){
      pos_pos+1
      Bmat[i,j]_0.5*betahat[pos]
      Bmat[j,i]_Bmat[i,j]}}
  x_model.matrix(lmobject)[,(nblocks+1):(nblocks+nn)]%*(
  eigen(Bmat)$vectors)
  xx_matrix(0,nrow(x),nblocks+2*nn+(nn-1)*nn/2)
  xx[,1:nblocks]_model.matrix(lmobject)[,1:nblocks]
  xx[, (nblocks+1):(nblocks+nn)]_x
  xx[, (nblocks+nn+1):(nblocks+2*nn)]_x^2
  newcol_nblocks+2*nn
  for (i in 1:(nn-1)){
    for (j in (i+1):nn){
      newcol_newcol+1
      cat(i,j,newcol,"\n")
```

```
xx[,newcol]_x[,i]*x[,j]}}
summary(lm(lmobject$y~xx-1),cor=F)}
```

#### Examples of Use:

```
dlr{twodfit,nblocks=1)
dlr(threedfit,nblocks=4)
dlr(fourdfit,nblocks=2)
dlr(fivedfit,nblocks=1)
```

## D.10 Direct fitting of canonical form-two dimensions

```
z12dfn <- function(x1, x2, theta)
{
  cos(theta) * x1 +sin(theta)* x2
}
z22dfn <- function(x1, x2, theta)
{
  - sin(theta) * x1 + cos(theta) * x2
}
test2dplin <- nls(y2d ~ cbind(rep(1, length(y2d)),
  z12dfn(x1, x2, theta), z22dfn(x1, x2, theta),
  I(z12dfn(x1, x2, theta)^2), I(z22dfn(x1, x2, theta)^2)),
  data = twoddata, start = list(theta = pi/4), algorithm =
  "plinear", trace = T)
summary(test2dplin)
```



## D.11 Direct fitting of canonical form-three dimensions

```

zfn_function(x1, x2, x3, theta12, theta13, theta23){
  test12 <- diag(3)
  test12[c(1, 2), c(1, 2)] <-matrix(c(cos(theta12),
    - sin(theta12), sin(theta12), cos(theta12)), 2, 2)
  test13 <- diag(3)
  test13[c(1, 3), c(1, 3)] <-matrix(c(cos(theta13),
    - sin(theta13), sin(theta13), cos(theta13)), 2, 2)
  test23 <- diag(3)
  test23[c(2, 3), c(2, 3)] <-matrix(c(cos(theta23),
    - sin(theta23), sin(theta23), cos(theta23)), 2, 2)
  test23 %*% test13 %*% test12}
testplin3_nls(y~cbind(block1,block2,block3,block4,
  (I(zfn(x1,x2,x3,theta12,theta13,theta23)[1,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[1,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[1,3]*x3)),
  (I(zfn(x1,x2,x3,theta12,theta13,theta23)[2,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[2,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[2,3]*x3)),
  (I(zfn(x1,x2,x3,theta12,theta13,theta23)[3,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,3]*x3)),
  (I(zfn(x1,x2,x3,theta12,theta13,theta23)[1,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[1,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[1,3]*x3)^2),
  (I(zfn(x1,x2,x3,theta12,theta13,theta23)[2,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[2,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[2,3]*x3)^2),
  (I(zfn(x1,x2,x3,theta12,theta13,theta23)[3,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,3]*x3)^2)),
  data=mydata,start=list(theta12=1.89,theta13=2.8,theta23=0.58),
  algorithm="plinear",trace=T)
testplin3s_nls(y~cbind(block1,block2,block3,block4,
  I(zfn(x1,x2,x3,theta12,theta13,theta23)[2,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[2,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[2,3]*x3),
  I(zfn(x1,x2,x3,theta12,theta13,theta23)[3,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,3]*x3),
  I(zfn(x1,x2,x3,theta12,theta13,theta23)[3,1]*x1+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,2]*x2+
  zfn(x1,x2,x3,theta12,theta13,theta23)[3,3]*x3)^2),
  data=threeddata,start=list(theta12=1.9,theta13=2.8,theta23=0.6),
  algorithm="plinear",trace=T)

```

## D.12 Direct fitting of canonical form-four dimensions

```

zfn_function(x1,x2,x3,x4,theta12,theta13,theta14,theta23,
  theta24,theta34)
{
  test12_diag(4)
  test12[c(1,2),c(1,2)]_matrix(c(cos(theta12),-sin(theta12),
    sin(theta12),cos(theta12)),2,2)
  test13_diag(4)
  test13[c(1,3),c(1,3)]_matrix(c(cos(theta13),-sin(theta13),
    sin(theta13),cos(theta13)),2,2)
  test14_diag(4)
  test14[c(1,4),c(1,4)]_matrix(c(cos(theta14),-sin(theta14),
    sin(theta14),cos(theta14)),2,2)
  test23_diag(4)
  test23[c(2,3),c(2,3)]_matrix(c(cos(theta23),-sin(theta23),
    sin(theta23),cos(theta23)),2,2)
  test24_diag(4)
  test24[c(2,4),c(2,4)]_matrix(c(cos(theta24),-sin(theta24),
    sin(theta24),cos(theta24)),2,2)
  test34_diag(4)
  test34[c(3,4),c(3,4)]_matrix(c(cos(theta34),-sin(theta34),
    sin(theta34),cos(theta34)),2,2)
  test34%%test24%%test23%%test14%%test13%%test12}
testplin4_nls(y4d~cbind(block1,block2,I(zfn(x1,x2,x3,
  x4,theta12,theta13,theta14,theta23,theta24,theta34)
  [1,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
  theta23,theta24,theta34)[1,2]*x2+zfn(x1,x2,x3,x4,
  theta12,theta13,theta14,theta23,theta24,theta34)
  [1,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
  theta23,theta24,theta34)[1,4]*x4),I(zfn(x1,x2,x3,x4,
  theta12,theta13,theta14,theta23,theta24,theta34)
  [2,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
  theta23,theta24,theta34)[2,2]*x2+zfn(x1,x2,x3,x4,
  theta12,theta13,theta14,theta23,theta24,theta34)
  [2,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
  theta23,theta24,theta34)[2,4]*x4),I(zfn(x1,x2,x3,x4,
  theta12,theta13,theta14,theta23,theta24,theta34)

```

```

[3,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[3,2]*x2+zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[3,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[3,4]*x4),I(zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[4,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[4,2]*x2+zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[4,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[4,4]*x4),I(zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[1,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[1,2]*x2+zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[1,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[1,4]*x4)^2,I(zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[2,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[2,2]*x2+zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[2,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[2,4]*x4)^2,I(zfn(x1,x2,x3,
x4,theta12,theta13,theta14,theta23,theta24,theta34)
[3,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[3,2]*x2+zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[3,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[3,4]*x4)^2,I(zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[4,1]*x1+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[4,2]*x2+zfn(x1,x2,x3,x4,
theta12,theta13,theta14,theta23,theta24,theta34)
[4,3]*x3+zfn(x1,x2,x3,x4,theta12,theta13,theta14,
theta23,theta24,theta34)[4,4]*x4)^2),data=fourddata,
start=list(theta12=pi/4,theta13=pi/4,theta14=pi/4,
theta23=pi/4,theta24=pi/4,theta34=pi/4),
algorithm="plinear",trace=T)
summary(testplin4)

```

### D.13 Calculation of precision of canonical axis-three dimensions

```

parms_summary(testplin3s)$parameters
e3_c(0,0,1)
dircos_as.vector(zfn(x1,x2,x3,parms[1],parms[2],parms[3]))[2,]
uvec_e3-dircos
uvec_uvec/sqrt(t(uvec)%*%uvec)
Hu_diag(3)-2*uvec%*%t(uvec)
Hu%*%dircos
projres_matrix(0,10000,3)
meanval_(testplin3s)$parameters[1:3]
covval_summary(testplin3s)$cov.unscaled[1:3,1:3]*
summary(testplin3s)$sigma^2
for (i in 1:10000){cat(i,"\n")
  parmsim_as.vector(rmvnorm(1,mean=meanval,
    cov=covval))
projres[i,]_as.vector(zfn(x1,x2,x3,parmsim[1],parmsim[2],
  parmsim[3]))[2,]
  projres[i,]_as.vector(Hu%*%as.vector(projres[i,]))}
plot(projres[,1:2])
a_sqrt(qchisq(.95,2) /eigen(solve(var(projres[,1:2])) )$values)
ff_function(theta,a1,a2){r_1/sqrt(cos(theta)^2/a1^2
+sin(theta)^2/a2^2)pt(sqrt(2)*sqrt(1-r^2)/r,2)}
1-integrate(ff,lower=0,upper=2*pi,a1=a[1],
a2=a[2])$integral/(2*pi)

```

### D.14 Calculation of precision of canonical axis-four dimensions

```

thetastar4d_summary(testplin4)$parameters[1:6]
vstar4d_summary(testplin4)$cov.unscaled[1:6,1:6]*
summary(testplin4)$sigma^2
yvec_zfn4d(x1,x2,x3,x4,thetastar4d[1],thetastar4d[2],

```

```

thetastar4d[3],thetastar4d[4],thetastar4d[5],thetastar4d[6])[1,]
evec_c(0,0,0,1)
uvec_yvec=evec
uvec_uvec/sqrt(t(uvec)%*%uvec)
Hu_diag(4)-2*uvec%*%t(uvec)
projres4_matrix(0,10000,4)
meanval_thetastar4d
covval_vstar4d
for (i in 1:10000){cat(i,"\n")
  parmsim4d_as.vector(rmvnorm(1,mean=meanval,
    cov=covval))
projvec_as.vector(zfn4d(x1,x2,x3,x4,parmsim4d[1],
  parmsim4d[2],parmsim4d[3],
  parmsim4d[4],parmsim4d[5],parmsim4d[6])[1,])
projres4[i,]_as.vector(Hu%*%projvec)}
plot(projres4[,1:2])
a_sqrt(qchisq(.95,2)/-sort(-eigen(solve(var(projres4[,1:3]))))
$values)
lowerbd4d_(((pi*a[1]*a[3])/(a[2]))*((1/a[3])*
  atan(1/((1-2*a[3]^2)/(2*a[3]*
    sqrt(1-a[3]^2)))))-(2*(1-a[2]^2))/
  (sqrt(a[2]^2-a[3]^2))*
  asinh(sqrt((a[2]^2-a[3]^2)/(1-a[2]^2))))/(2*pi^2)
upperbd4d_(((pi*a[2]*a[3])/(a[1]))*((1/a[3])*atan(1/
  ((1-2*a[3]^2)/(2*a[3]*sqrt(1-a[3]^2)))))-(2*
  (1-a[1]^2))/sqrt(a[1]^2-a[3]^2))*asinh(sqrt(
    (a[1]^2-a[3]^2)/(1-a[1]^2))))/(2*pi^2)

```

# Bibliography

- Aitkin, M., D. Anderson, B. Francis and J. Hinde (1989), *Statistical Modelling in GLIM*, Oxford University Press, London.
- Angün, E., G. Gurkan, D. den Hertog and J.P.C. Kleijnen (2002), Response surface methodology revisited, *in* E.Yucesan, C.Chen, J.Snowdon and J.Charnes, eds, 'Proceedings of the 2002 Winter Simulation Conference', pp. 377–383.
- Ankenman, B.E. (2003), 'Identifying rising ridge behavior in quadratic response surfaces', *IIE Transactions* **35**, 493–502.
- Ankenman, B.E. and S. Bisgaard (1995), *Personal Communication* .
- Atkinson, A.C. and L.M. Haines (1996), Designs for nonlinear and generalized linear models, *in* S.Gosh and C.R.Rao, eds, 'Handbook of Statistics', Elsevier, Amsterdam, pp. 437–475.
- Bisgaard, S. and B. Ankenman (1996), 'Standard errors for the eigenvalues in second-order response surface models', *Technometrics* **38**, 238–246.
- Bisgaard, S. and H.T. Fuller (1994–95), 'Analysis of factorial experiments with defects or defectives as the response', *Quality Engineering* **7**, 429–443.
- Box, G.E.P. (1955), 'Contribution to the discussion, Symposium on Interval Estimation', *Journal of the Royal Statistical Society, Series B* **16**, 211–212.
- Box, G.E.P. (1999), 'Statistics as a catalyst to learning by scientific method. Part II - a discussion', *Journal of Quality Technology* **31**, 16–29.

- Box, G.E.P. and D.W. Behnken (1960), 'Some new three-level designs for the study of quantitative variables', *Technometrics* **2**, 455-475.
- Box, G.E.P. and H.L. Lucas (1959), 'Design of experiments in nonlinear situations', *Biometrika* **46**, 77-90.
- Box, G.E.P. and J.S. Hunter (1954), 'A confidence region for the solution of a set of simultaneous equations with an application to experimental design'. *Biometrika* **41**, 190-199.
- Box, G.E.P. and J.S. Hunter (1957), 'Multi-factor experimental designs for exploring response surfaces', *The Annals of Mathematical Statistics* **28**, 195-241.
- Box, G.E.P. and K.B. Wilson (1951), 'On the experimental attainment of optimal conditions', *Journal of the Royal Statistical Society, Series B* **13**, 1-45.
- Box, G.E.P. and N.R. Draper (1987), *Empirical Model-Building and Response Surfaces*, John Wiley and Sons, New York.
- Box, G.E.P. and P.Y.T. Liu (1999), 'Statistics as a catalyst to learning by scientific method. Part I - an example', *Journal of Quality Technology* **31**, 1-15.
- Box, G.E.P. and S. Jones (1992), 'Split-plot designs for robust product experimentation', *Journal of Applied Statistics* **19**, 3-25.
- Box, G.E.P. and W.G. Hunter (1965), Sequential designs of experiments for non-linear models, in 'IBM Scientific Computing Symposium in Statistics', IBM, Yorktown Heights, NY., pp. 113-137.
- Box, G.E.P., W.G. Hunter and J.S. Hunter (1978), *Statistics for Experimenters*, John Wiley and Sons, New York.
- Box, M.J. (1971), 'An experimental design criterion for precise estimation of a subset of the parameters in a nonlinear model', *Biometrika* **58**, 149-153.
- Brooks, S.H. and M.R. Mickey (1961), 'Optimum estimation of gradient direction in steepest ascent experiments', *Biometrics* **17**, 48-56.

- Carter, W.H., Jr., V.M. Chinchilli and E.D. Campbell (1990), 'A large-sample confidence region useful in characterizing the stationary point of a quadratic response surface', *Technometrics* **32**, 425–435.
- Carter, W.H., Jr., V.M. Chinchilli, R.H. Myers and E.D. Campbell (1986), 'Confidence intervals and an improved ridge analysis of response surfaces', *Technometrics* **28**, 339–346.
- Chaloner, K. and I. Verdinelli (1995), 'Bayesian experimental design: A review', *Statistical Science* **10**, 273–304.
- Covey-Crump, P.A.K. and S.D. Silvey (1970), 'Optimal regression designs with previous observations', *Biometrika* **57**, 551–566.
- Davison, A. C. and D. V. Hinkley (1989), *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge, England.
- del Castillo, E. (1996), 'Multiresponse process optimization via constrained confidence regions', *Journal of Quality Technology* **28**, 61–70.
- del Castillo, E. and D. C. Montgomery (1993), 'A nonlinear programming solution to the dual response problem', *Journal of Quality Technology* **25**, 199–204.
- del Castillo, E., D. C. Montgomery and D. R. McCarville (1996), 'Modified desirability function for multiple response optimization', *Journal of Quality Technology* **28**, 337–345.
- del Castillo, E. and S. Cahaya (2001), 'A tool for computing confidence regions on the stationary point of a response surface', *The American Statistician* **55**, 358–365.
- Derringer, G. and R. Suich (1980), 'Simultaneous optimization of several response variables', *Journal of Quality Technology* **12**, 214–219.
- Di Cicco, T.J. and B. Efron (1996), 'Bootstrap confidence intervals', *Statistical Science* **11**, 189–228.



- Draper, N.R. (1963), 'Ridge analysis of response surfaces', *Technometrics* **5**, 469–479.
- DuMouchel, W. and B. Jones (1994), 'A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model', *Technometrics* **36**, 37–47.
- Dykstra, O. (Jr.) (1966), 'The orthogonalization of undesigned experiments', *Technometrics* **8**, 279–290.
- Dykstra, O. (Jr.) (1971), 'The augmenting of experimental data to maximize  $X^T X$ ', *Technometrics* **13**, 682–888.
- Efron, B. (1979), 'Bootstrap methods: Another look at the jackknife', *The Annals of Statistics* **7**, 1–26.
- Efron, B. and R.J. Tibshirani (1993), *An introduction to the bootstrap*, Chapman and Hall, New York.
- Evans, M., N. Hasting and B. Peacock (1993), *Statistical Distributions*, 2nd edn, John Wiley and Sons, New York.
- Gaylor, D.W. and J.A. Merrill (1968), 'Augmenting existing data in multiple regression', *Technometrics* **10**, 73–81.
- Gilmour, S.G. and N.R. Draper (2003), 'Confidence regions around the ridge of optimal response on fitted second-order response surfaces', *Technometrics* **45**, 333–339.
- Gilmour, S.G. and N.R. Draper (2004), 'Response', *Technometrics* **46**, 358.
- Gleser, L.J. and J.T. Hwang (1987), 'The non-existence of  $100(1 - \alpha)\%$  confidence sets of finite expected diameters in error-in-variable and related models', *The Annals of Statistics* **15**, 1351–1362.
- Grego, J. M. (1993), 'Generalized linear models and process variation', *Journal of Quality Technology* **25**, 288–295.

- Hamada, M. and J.A. Nelder (1997), 'Generalised linear models for quality-improvement experiments', *Journal of Quality Technology* **29**, 292–304.
- Harrington, Jr., E. C. (1965), 'The desirability function', *Industrial Quality Control* **21**, 494–498.
- Hebble, T.L. and J. Mitchell, T (1972), 'Repairing response surface designs'. *Technometrics* **14**, 767–779.
- Hill, W.J. and W.G. Hunter (1966), 'A review of response surface methodology: A literature survey', *Technometrics* **8**, 571–590.
- Hill, W.J. and W.G. Hunter (1974), 'Design of experiments for subset of parameters', *Technometrics* **16**, 425–434.
- Hoerl, A. E. (1959), 'Optimum solution of many variables equations', *Chemical Engineering Progress* **55**, 69–78.
- Hoerl, R. W. (1985), 'Ridge analysis 25 years later', *The American Statistician* **39**, 186–192.
- Hsieh, P.I. and D.E. Goodwin (1986), Sheet molded compound process improvement, in 'Fourth Symposium on Taguchi Methods', American Supplier Institute, Dearborn, MI, pp. 13–21.
- Jia, Y. and R.H. Myers (1998), 'Bayesian experimental designs for logistic regression models', *Technical Report No. 98-22, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA* .
- Khuri, A.I. (1999), 'Discussion', *Journal of Quality Technology* **31**, 58–60.
- Kiefer, J. (1958), 'On the nonrandomised optimality and the randomised nonoptimality of symmetrical designs', *The Annals of Mathematical Statistics* **29**, 675–699.
- Kiefer, J. (1959), 'Optimum experimental designs', *Journal of the Royal Statistical Society, Series B* **21**, 272–319.

- Kiefer, J. (1961a), 'Optimum designs in regression problems. II'. *The Annals of Mathematical Statistics* **32**, 298–325.
- Kiefer, J. (1961b), Optimum experimental designs V, with applications to systematic and rotatable designs, in 'Fourth Berkeley Symposium'. Vol. 1, pp. 381–405.
- Kiefer, J. (1962a), 'An extremum result', *Canadian Journal of Mathematics* **14**, 597–601.
- Kiefer, J. (1962b), 'Two more criteria equivalent to  $D$ -optimality to designs', *The Annals of Mathematical Statistics* **33**, 792–796.
- Kiefer, J. and J. Wolfowitz (1959), 'Optimum designs in regression problems', *The Annals of Mathematical Statistics* **30**, 271–294.
- Kiefer, J. and J. Wolfowitz (1969), 'The equivalence of two extremum problems', *Canadian Journal of Mathematics* **12**, 363–366.
- Kiprianova, R. and L. Markovska (1993), *Experimental Design and Analysis*, Forschungszentrum Jülich, GmbH.
- Kleijnen, J.P.C., D. den Hertog and E. Angün (2004), 'Response surface methodologies steepest ascent size revisited', *European Journal of Operational Research* **159**, 121–131.
- Letsinger, J.D., R.H. Myers and M. Lentner (1996), 'Response surface methods for bi-randomization structures', *Journal of Quality Technology* **28**, 381–397.
- Lewis, S.L., D.C. Montgomery and R.H. Myers (2001a), 'Confidence interval coverage for designed experiments analyzed with generalized linear models', *Journal of Quality Technology* **33**, 279.
- Lewis, S.L., D.C. Montgomery and R.H. Myers (2001b), 'Examples of designed experiments with nonnormal responses', *Journal of Quality Technology* **33**, 265–278.

- McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. 2nd edn. John Wiley and Sons, New York.
- Mead, R. and D.J. Pike (1975), 'A review of response surface methodology from a biometric viewpoint', *Biometrics* **31**, 803–851.
- Meeker, W.Q. and L.A. Escobar (1995), 'Teaching about approximate confidence regions based on maximum likelihood estimation', *The American Statistician* **49**, 48–53.
- Monogan, M.B, K.O. Geddes, K.M. Heal, G. Labahn, S.M. Vorkoetter and J. McCarron (2000), *Maple 6 Programming Guide*, Waterloo Maple Inc., Waterloo, Canada.
- Myers, R.H. (1999), 'Response surface methodology - current status and future directions', *Journal of Quality Technology* **31**, 30–44.
- Myers, R.H. and A.I. Khuri (1979), 'A new procedure for steepest ascent', *Communications in Statistics, A* **8**, 1359–1376.
- Myers, R.H., A.I. Khuri and W.H. Carter, Jr. (1989), 'Response surface methodology: 1966-1988', *Technometrics* **31**, 137–157.
- Myers, R.H. and D.C. Montgomery (1995), *Response Surface Mathodology: Process and Product Optimization Using Designed Experiments*, John Wiley and Sons, New York.
- Myers, R.H. and D.C. Montgomery (1997), 'A tutorial on generalised linear models', *Journal of Quality Technology* **29**, 274–291.
- Myers, R.H. and W.H. Carter, Jr. (1973), 'Response surface techniques for dual response systems', *Technometrics* **15**, 301–317.
- Nelder, J. A. and R. W. M. Wedderburn (1972), 'Generalized linear models', *Journal of the Royal Statistical Society, Series A* **132**, 370–384.

- Nelder, J. A. and Y. Lee (1991), 'Generalized linear models for the analysis of Taguchi-type experiments', *Applied Stochastic Models and Data Analysis* **7**, 107–120.
- Peterson, J.J. (1993), 'A general approach to ridge analysis with confidence intervals', *Technometrics* **35**, 204–214.
- Peterson, J.J., S. Cahaya and E. del Castillo (2002), 'A general approach to confidence regions for optimal factor levels of response surfaces'. *Biometrics* **58**, 422–431.
- Peterson, J.J., S. Cahaya and E. del Castillo (2004), 'Letter to editor'. *Technometrics* **46**, 355–357.
- Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, 2nd edn, Wiley, New York.
- Sacks, J., W.J. Welch, T.J. Mitchell and H.P. Wynn (1989), 'Design and analysis of computer experiments', *Statistical Science* **4**, 409–435.
- Seber, G.A.F. (1977), *Linear Regression Analysis*, John Wiley and Sons, New York.
- Sloane, N.J.A. (2005), 'A library of orthogonal arrays'.  
**URL:** [www.research.att.com/~njas/oadir/](http://www.research.att.com/~njas/oadir/)
- Smyth, G.K. (1989), 'Generalized linear models with varying dispersion', *Journal of the Royal Statistical Society, Series B* **51**, 47–60.
- Spiegel, M. R. (1997), *Schaum's Outline Series - Mathematical Handbook of Formulas and Tables*, McGraw-Hill, New York.
- Spjøtvoll, E. (1972), 'Multiple comparison of regression functions', *The Annals of Statistics* **43**, 1076–1088.
- Stablein, D.M., W.H. Carter, Jr. and G.L Wampler (1983), 'Confidence regions for constrained optima in response surface experiments', *Biometrics* **39**, 759–763.

- Stroud, A.H. (1971), *Approximate Calculation of Multiple Integrals*. Prentice-Hall, New Jersey.
- Sztendur, E.M. and N.T. Diamond (1999), Confidence interval calculations for the path of steepest ascent in response surface methodology, in 'The proceedings of the Collaborative Research Forum', School of Communications and Informatics, Victoria University of Technology, Melbourne, pp. 129–132.
- Sztendur, E.M. and N.T. Diamond (2001), Inequalities for the precision of the path of steepest ascent in response surface methodology, in Y.Cho, J.Kim and S.Dragomir, eds, 'Inequality Theory and Applications', Vol. 1, Nova Science Publishers, Inc., New York, pp. 301–307.
- Sztendur, E.M. and N.T. Diamond (2002), 'Extension to confidence region calculations for the path of steepest ascent', *Journal of Quality Technology* **34**, 289–296.
- Taguchi, G. (1987), *System of Experimental Design - Engineering Methods to Optimize Quality and Minimize Costs*, UNIPUB/Kraus International Publications and American Supplier Institute, New York, Michigan.
- Venables, W.N. and B.D. Ripley (1999), *Modern Applied Statistics with S-PLUS*, 3rd edn, Springer, New York.
- Wynn, H. P. (1970), 'The sequential generation of D-optimum experimental designs', *The Annals of Mathematical Statistics* **41**, 1655–1664.
- Wynn, H. P. (1972), 'Results in the theory and construction of D-optimum experimental designs', *Journal of the Royal Statistical Society, Series B* **34**, 133–147.