ALGORITHMS FOR THE CONSTRUCTION OF CONSTRAINED AND UNCONSTRAINED OPTIMAL DESIGNS

BY

ABDULHADI M. ALAHMADI

A thesis presented to the

University of Glasgow

Faculty of Science

Department of Statistics

for the degree of Doctor of Philosophy

July 1993

SUMMARY

The aim of this thesis is to explore algorithms for solving the following constrained optimisation problems :

Problem(a):

" Maximise a criterion $\phi(p)$ subject to the constraints $\sum_{i=1}^{J} p_i = 1;\ 0 \leq p_i \leq 1.$"

In particular we consider cases in which $\phi(p) = \Psi\{M(p)\}$ where $\Psi\{\bullet\}$ is a standard design criterion (e.g. c- or D-optimality) and $M(p) = VPV'$ $= \sum_{i=1}^{J} p_i \underline{v}_i \underline{v}_i^t$, V is $k \times J$ design matrix and $P = Diag(p_1, p_2, p_3, \ldots, p_J)$.

We also consider the following non-standard design problems:

Problem(b):

"Maximise $\phi_c(p) = -\left[\underline{a}' M^{-1}(p) \underline{b}\right]^2$ or $\phi_\rho(p) = -\dfrac{\left[\underline{a}' M^{-1}(p) \underline{b}\right]^2}{\left[\underline{a}' M^{-1}(p) \underline{a}\right]\left[\underline{b}' M^{-1}(p) \underline{b}\right]}$

subject to the constraints $\sum_{i=1}^{J} p_i = 1, 0 \leq p_i \leq 1$" where $\underline{a}, \underline{b} \in \Re^k$.

We call these two criteria the covariance and the correlation criterion respectively.

Problem(c):

"Maximise a standard design criterion, $\phi(p)$, subject to a zero covariance condition and $\sum_{i=1}^{J} p_i = 1, 0 \leq p_i \leq 1.$"

Chapter 1 provides an introduction to the area of optimum experimental design for the linear regression design problem with parameter vector $\underline{\theta}$. This problem seeks to obtain a best inference for all or some of the components of $\underline{\theta}$ by making the dispersion matrix of their estimates small in some sense. In this chapter we summarise the main criteria used for this purpose.

Chapter 2 studies a class of multiplicative algorithms of the form $p_i^{(r+1)} = p_i^{(r)} f(d_i,\delta) / \sum_{j=1}^{J} p_j^{(r)} f(d_j,\delta)$, indexed by a function $f(d,\delta)$ which depends on the derivatives of the criterion $\phi(p)$ and a free parameter δ for solving **problem(a)**. The performance of the algorithm is investigated in constructing D-optimal designs under optimal choices of the parameter δ, and in constructing c-optimal designs starting from difficult initial designs, using an optimal and fixed value of the parameter δ. The work for this chapter has appeared in Torsney and Alahmadi (1992).

Chapter 3 considers the covariance and correlation criterion of **problem(b)**. The only property we know of these criteria is homogeneity in the weights p of degree -2 and zero respectively. This type of criterion differs from the standard optimality criteria such as c-, D- and A-optimality criteria . It may have negative first partial derivatives. An explicit solution has been found for the optimal weights and the optimal value for the covariance criterion when the number of design points equals the number of parameters i.e $J = k$, while in the case when $J > k$ we have explored a new version of the above algorithm for dealing with this type of problem .

Chapters 4 and 5 are concerned with the solution of **problem (c).**

In Chapter 4 we consider the case when the number of design points equals the number of the unknown parameters $\underline{\theta}$. In this case we find a class of designs which guarantees zero covariance. Zero covariance is guaranteed under a transformation of the design weights $\underline{p}$ to two or three sets of variables each of which forms a probability vector. We wish to maximise standard design criteria with respect to these weights . This yields an extension of **problem (a)** of Chapter 2 to that of maximising a criterion with respect to two probability vectors and we use a natural extension of the algorithm used for that problem . For the above mentioned results the efficiencies of the restricted optimal design under the zero covariance constraint relevant to the unrestricted optimal design has been calculated .

Chapter 5 considers the case when the number of design points exceeds the number of the parameters. Using a Lagrangian approach, the problem is transformed to one of simultaneous maximisation of two functions of the same probability vector each of which is maximised at the same value of this vector and have a common maximum of zero. This yields another extension of **problem(a).**

Chapter 6 summarises the results obtained in the preceding chapters as well as giving an indication of future work that could be done.

Acknowledgements

This thesis is a result of research of several years, during which I have been helped in various ways by many people to whom I owe a debt of gratitude.

First of all, I should like to acknowledge my supervisor, *Dr Ben Torsney,* for suggesting the problems and for all his advice, support, patience and encouragement. I am particularly grateful to him for giving me the opportunity to meet and communicate with all *visitors* in the same speciality, whose ideas and methods have informed and inspired me, both through their written work and through many enjoyable and stimulating discussions. These colleagues, many of them my seniors, are sources of a great deal of the knowledge I have.

Secondly, thanks are due to the ex-head of department; *Professor Titterington,* and to the current head of department, *Professor Ian Ford,* for providing me with the opportunity to study at the University of Glasgow and for providing the necessary facilities. These were manifested by organising sensible courses and providing required references. Thanks are also due to all *members of staff* in the Department of Statistics, for providing me with every possible assistance in a very friendly environment and also for their excellent way of teaching and invaluable advice and encouragement. I would also like to express my appreciation to all my *friends and fellow research students* for making my stay here very enjoyable.

Thirdly, I should like to acknowledge the financial support of *King Abdul Aziz University* in the Kingdom of Saudi Arabia.

Finally, thanks are due to my *wife* and my *children* without whose unlimited patience this thesis would not have been possible.

To my mother and my brothers

Table of Contents

Title Page i

Summary ii

Acknowledgements v

Table of Contents vii

CHAPTER 1 Motivation 2

1.1 Optimum Linear Regression Design 2

1.2 Discretizing the Design Space 5

1.3 Design Criteria and Their Properties 9

1.3.1 Case 1 9
D-optimality criterion 9
A-optimality criterion 11
G-optimality criterion 12
E-optimality criterion 13

1.3.2 Case 2 14
D_A-optimally criterion 15
Linear optimality criterion 17
E_A-optimality criterion 17

CHAPTER 2 Further Development of Algorithms for Constructing Optimising Distributions 20

2.1 Introduction 20

2.2 Optimality Conditions 22

2.2.1 Directional Derivatives 22

2.2.2 Conditions for Local Optimality 23

2.3 A class of Algorithm 24

2.4 Properties of Proposed Algorithm 25

2.4.1 General Properties 25

2.4.2 Properties of Specific Cases 28

2.5 Construction of Optimal Designs: Empirical Results on Convergence 33

2.5.1 Construction of Optimal Designs 33

2.5.2 The Convergence of the Algorithm 34

2.6 Tables 38

CHAPTER 3 Covariance and Correlation Criterion 43

3.1 Introduction 43

3.2 Simplifying the Covariance Criterion in the Case of *kxk* Design Matrix 45

3.2.1 Explicit Solution 46

3.3 The Invariance of the Covariance and Correlation Criterion Under Specific Choices of $\underline{a}$ and $\underline{b}$ 51

3.4 Algorithm 52

3.5 Examples and Discussions 53

3.6 The Convergence of the Algorithm 60

CHAPTER 4 Maximise A Concave Criterion Subject To Zero Covariance 72

4.1 Introduction 72

4.2 Study Different Cases for Design Satisfying Zero Covariance 73

4.2.1 Case 1 73

4.2.2 Equivalence of the c- and the Linear Optimality Criteria 76

4.2.3 Case 2 76

4.2.4 Properties of the Transformation 78

4.2.5 Case 3 80

4.3 Algorithm 83

4.4 Examples and Discussions 86

4.5 Efficiencies of Constraint Optimal Design 95

CHAPTER 5 Optimal Design Subject To Zero Covariance(General Case) 117

5.1 Introduction 117

5.2 Quadratic Regression Model 117

5.3 General Problem (GP) 120

5.4 Proposed Algorithm 124

5.5 Examples and Discussions 126

Chapter 6 Conclusion and Further Work 130

6.1 Conclusion 130

6.2 Further Work 134

Appendix 1 137

References 141

CHAPTER ONE

Motivation

1.1 Optimum Linear Regression Design

1.2 Discretizing The Design Space

1.3 Design Criteria And Their Properties

CHAPTER ONE

MOTIVATION

1.1 Optimum Linear Regression Design

The concept of optimum regression design arises when an observable univariate variable y has probability model $p(y/\underline{u}(\underline{x}),\underline{\theta},\sigma)$ which depends on:

(1) A column vector $\underline{\theta}=(\theta_1,\theta_2,\ldots,\theta_k)'$ of parameters which are fixed but unknown to the experimenter. The true value of $\underline{\theta}$ is known to belong to a set $\Theta \in R^k$.

(2) The quantity $\underline{x}$ represents a vector of control variables. It can be chosen by the experimenter, its value being restricted to the space χ, where χ will be often a closed compact set of Euclidean space of some dimension called the design space. Typically it will be continuous but can be discrete.

(3) σ is a nuisance parameter; this also is fixed and unknown but is not of primary interest.

(4) The vector $\underline{u}=\underline{u}(\underline{x})=\{ f_1(\underline{x}), f_2(\underline{x}),\ldots,f_k(\underline{x}) \}'$ where the functions $f_i(\underline{x})$, $i=1,2,\ldots,k$ are of known bounded form.

The regression is linear when the model is linear in the unknown parameters $\underline{\theta}$ but not necessarily linear in $\underline{x}$. So the linear model will be of the form:

$$E(y / \underline{u}, \underline{\theta}, \sigma) = \underline{u}' \underline{\theta} \tag{1.1.1}$$

Now our aim is to select, say n, values of x from the design space χ to obtain as good an inference as possible for all or some of the parameters $\underline{\theta}$. The basic problem is: at what x-values should these observations be taken?.

Such a selection of x-values from the design space χ is termed a design or regression design. We wish to choose them optimally i.e. use an optimal design.

Suppose the model (1.1.1) is true. Let y_i denote the observation obtained at x_i so that

$$E(y_i) = \underline{u}_i' \underline{\theta} \quad , \quad \underline{u}_i = \{ f_1(\underline{x}_i), \ldots, f_k(\underline{x}_i) \}' \quad , i = 1, 2, \ldots, n$$

and suppose the observations $y_1, y_2, \ldots, y_n$ are taken to be uncorrelated and of equal variance σ^2. The y_i's then satisfy the standard linear model:

$$E(Y) = X\underline{\theta} \quad , \quad D(Y) = \sigma^2 I_n \tag{1.1.2}$$

where $Y = (y_1, y_2, \ldots, y_n)$, X is the $n \times k$ matrix whose $(i, j)th$ element is $f_j(x_i)$, I_n is the $n \times n$ identity matrix and $D(Y)$ denotes the dispersion matrix of Y.

The least squares normal equations for the model (1.1.2) are of the form:

$$(X'X)\hat{\underline{\theta}} = X'Y \tag{1.1.3}$$

So if the full parameter system $\underline{\theta} \in \Theta$ is of interest, then the selection of $\underline{x}$ must at least insure that the matrix $(X'X)$ is non-singular, in which case the unique solution for (1.1.3) is given by:

$$\hat{\underline{\theta}} = (X'X)^{-1} X'Y \tag{1.1.4}$$

with $E(\hat{\underline{\theta}}) = \theta$ and $D(\hat{\underline{\theta}}) = \sigma^2 (X'X)^{-1}$, where $E(\hat{\underline{\theta}})$, $D(\hat{\underline{\theta}})$ are the expectation vector and the dispersion matrix of $\hat{\underline{\theta}}$ respectively.

Clearly the dispersion matrix of $\hat{\underline{\theta}}$ does not depend on $\underline{\theta}$ and only depends proportionately on the parameter σ^2. We have to select $\underline{x}$ which makes the matrix $D(\hat{\underline{\theta}})$ as small as possible, namely an $\underline{x}$ which makes the $k \times k$ matrix $(X'X)$ large in some sense.

1.2 Discretizing The Design Space

Recall from the previous section that the linear model of formula (1.1.1) is

$$E(y/\underline{u},\underline{\theta},\sigma)=\underline{u}'\underline{\theta} \tag{1.2.1}$$

where $\underline{u}=\{f_1(\underline{x}),f_2(\underline{x}),\dots,f_k(\underline{x})\}'$. Clearly choosing a vector $\underline{x}$ in the design space χ is equivalent to choosing a k-vector $\underline{v}$ in the closed bounded k-dimensional space $V=\underline{f}(\chi)$, where $\underline{f}$ is the vector valued function $(f_1,f_2,\dots,f_k)'$. Thus there is no loss of generality and considerable notational convenience in replacing (1.2.1) by

$$E(y/\underline{v},\underline{\theta},\sigma)=\underline{v}'\underline{\theta} \tag{1.2.2}$$

where $\underline{v}\in V=\{\underline{v}:\underline{v}=[f_1(\underline{x}),f_2(\underline{x}),\dots,f_k(\underline{x})]',\underline{x}\in\chi\}$. So from now on we will refer to V as the design space. Typically this design space is continuous but we can assume that V is discrete. An explanation for this will be given later on in this section.

Suppose now that the discrete design space V consists of J distinct vectors $\underline{v}_1,\underline{v}_2,\dots,\underline{v}_J$. Then the basic linear model is

$$E(y/\underline{v},\underline{\theta},\sigma)=\underline{v}'\underline{\theta} \quad ,\underline{v}\in V=\{\underline{v}_1,\underline{v}_2,\dots,\underline{v}_J\} \tag{1.2.3}$$

Suppose we can take n observations. We must decide how many of these, say n_j (non negative integer), to take at v_j, $\sum_{j=1}^{J}n_j=n$. Given these choices the matrix $(X'X)$ can be expressed in the form

$$X'X=M(\underline{n}) \tag{1.2.4}$$

where $$M(\underline{n}) = \sum_{j=1}^{J} n_j \underline{v}_j \underline{v}_j^t = V\ N\ V^t \qquad (1.2.5)$$

and $V = \left(\underline{v}_1 \mid \underline{v}_2 \mid \dots \mid \underline{v}_J\right)$, $N = diag\left(n_1, n_2, \dots, n_J\right)$

We now want to choose $\underline{n}$ to make the matrix $M(\underline{n})$ big in some sense. Given that the n_js must be integer this is an integer programming problem and in the design context is described as an exact design problem.

However

$$M(\underline{n}) = nM(p) \qquad (1.2.6)$$

where $$M(p) = \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^t \qquad (1.2.7)$$

$$= V\ P\ V^t \qquad (1.2.8)$$

and $P = diag\ (p_1, p_2, \dots, p_J)$, $p_j = {}^{n_j}\!/_{n}$. Note that $p_j \geq 0$, $\sum_{j=1}^{J} p_j = 1$.

Our problem becomes that of choosing p to make the matrix $M(p)$ large subject to $p_j = {}^{n_j}\!/_{n}$. Relaxing the latter to $p_j \geq 0$ and $\sum_{j=1}^{J} p_j = 1$ yields an approximate design problem.

Note we can view p as defining a probability distribution on V to yield

$$M(p) = E_p[\underline{v}\ \underline{v}^t\ ] \qquad (1.2.9)$$

where $P(\underline{v} = \underline{v}_j\ ) = p_j$.

Definition (1.2.1) Design Measure:

The collection of variables of the form

$$\begin{pmatrix} \underline{v}_1 \ , \underline{v}_2 \ , \ldots , \underline{v}_J \\ p_1 \ , p_2 \ , \ldots , p_J \end{pmatrix}$$

is called a design, while p is called the design measure, in that it assigns weight or probability p_j to the vector $\underline{v}_j \in V$, where

$$\sum_{j=1}^{J} p_j = 1 \ , \ 0 \le p_j \le 1 \ , \ j = 1, 2, \ldots , J .$$

Definition (1.2.2) Support Of A Design Measure:

The support of the design measure p in the design space V is defined to be those vertices $\underline{v}_j$ with non-zero weighting under p. We denote it by $Supp(p)$. Often there will be an optimal design, say p^* such that $Supp(p^*)$ is a strict subset of V. Notationally

$$Supp(p) = \{\underline{v}_j \in V : \mathrm{p_j} > 0 \} \ , j = 1, 2, \ldots , J$$

Properties Of The Matrix *M(p)*:

(a) The matrix $M(p)$ is a non negative definite symmetric matrix. The symmetry of this matrix follows from its definition (1.2.6), and the nonnegativeness of the appropriate quadratic form is easy to verify:

$$\underline{x}' M(p) \underline{x} = \underline{x}' E_p \left[\underline{v} \, \underline{v}' \right] \underline{x} = E_p \left[\underline{x}' \underline{v} \, \underline{v}' \underline{x} \right]$$

$$= E_p \left[(\underline{x}' \underline{v})^2 \right] \ \ge 0$$

(b) Let $M = \{M(p) : p$ is any probability measure on $V\}$. The set M is the convex hull of the set $\{\underline{v}\,\underline{v}' : \underline{v} \in V\}$. Note that if p_v is the probability measure that puts unit weight at the point $\underline{v} \in V$, then $M(p_v) = \underline{v}\,\underline{v}'$, see Silvey (1980).

The third property is given by the following theorem.

Theorem (1.2.1) (Caratheodory's Theorem) :(see Fedorov 1972)

Each point M of the convex hull M^* of any subset μ of n-dimensional space can be represented in the form:

$$M = \sum_{j=1}^{n+1} \alpha_j u_j \ , \quad \sum_{j=1}^{n+1} \alpha_j = 1 \ , \ \alpha_j \geq 0 \ , \ u_j \in \mu \ , \ j = 1,2,\ldots,n+1$$

If M is a boundary point of the set M^* then α_{n+1} can be set to zero.

To see the importance of the above theorem we note that each $M \in M$ has at least one representation of the form:

$$M = \sum_{l=1}^{L} p_l \underline{v}_l \underline{v}_l'$$

where $\underline{v}_l \in V$, $l = 1,2,\ldots, L$ and $L \leq \{[k(k+1)/2]+1\}$. Also by the same theorem if M is a boundary point of M, the inequality involving L can be strengthened to $L \leq \{[k(k+1)/2]\}$.

Thus we have that any continuous design measure and in particular any continuous optimal design measure can be replaced by at least one finite discrete probability distribution, and so we have an explanation for having initially assumed V discrete.

1.3 Design Criteria And Their Properties

As we have seen, it may be possible to obtain a best inference for all or some of the unknown parameters $\underline{\theta} \in \Theta$ by making the matrix $M(p)$ large in some sense. So we consider various ways in which we might wish to make the matrix $M(p)$ large, namely by maximizing some real valued function $\phi(p) = \varphi\{M(p)\}$. Note that the function ϕ is called the criterion function, and in turn, the criterion defined by the function ϕ is usually called ϕ-optimality. A design maximizing $\phi(p)$ is called a ϕ-optimal design.

1.3.1 case(1):

In this section we consider the case when we are interested in all the unknown parameters $\underline{\theta} \in \Theta$ of the linear model (1.1.1); that is all the parameters $\theta_1, \theta_2, \ldots, \theta_k$ are important. We want to make the matrix $M(p)$ large in some sense. The matrix $M(p)$ must therefore be non-singular and hence positive definite. We shall consider four criteria.

(I) The D-optimality Criterion

The D-optimality criterion is defined by the criterion function:

$$\phi_1(p) = \log\{\det[M(p)]\} = -\log\{\det[M(p)]^{-1}\} \qquad (1.3.1)$$

If we assume normality of the errors in the linear model (1.1.1), then the general form of the joint confidence region for the vector of unknown parameters $\underline{\theta} \in \Theta$ is described by an ellipsoid of the form :

$$\{ (\underline{\theta} - \hat{\underline{\theta}})' M(p) (\underline{\theta} - \hat{\underline{\theta}}) \leq constant \},$$

where $\hat{\underline{\theta}}$ is the least squares estimate or the maximum likelihood estimate of $\underline{\theta}$. The D-optimality criterion chooses $M(p)$ to make the volume of the above ellipsoid as small as possible because it is the case that this volume is proportional to $\{det[M(p)]\}^{1/2}$. The value of $log\{det[M(p)]\}$ is finite if and only if $M(p)$ is non-singular i.e. when all the unknown parameters are estimatable. This is the celebrated criterion of D-optimality, the most extensively studied of all design criteria; see Ford(1976), Shah and Sinha(1989), Atkinson and Donev(1992).

Properties of $\phi_1(p) = \varphi_1\{M(p)\}$:

(a) φ_1 is an increasing function over the set of positive definite symmetric matrices. That is for $M_1, M_2 \in \mathcal{M}$ then $\varphi_1(M_1 + M_2) \geq \varphi_1(M_1)$ where $\mathcal{M}$ is the set of all non negative definite symmetric matrices.

(b) φ_1 is a concave function of the positive definite symmetric matrices. That is for every $\alpha \in (o, 1)$ and $M_1, M_2 \in \mathcal{M}$

$$\varphi_1[\alpha M_1 + (1-\alpha)M_2] \geq \alpha\,\varphi_1(M_1) + (1-\alpha)\varphi_1(M_2), \quad 0 \leq \alpha \leq 1.$$

Note that: if $M_1, M_2 \in \mathcal{M}_+$ then we can say that φ_1 is strictly concave, where $\mathcal{M}_+ = \{M \in \mathcal{M} : \det(M) \neq 0\}$; see Ford(1976).

(c) ϕ_1 is differentiable whenever it is finite, and the first derivative is:

$$\frac{\partial \phi_1}{\partial p_j} = \underline{v}_j^t M^{-1}(p)\underline{v}_j \qquad (1.3.2)$$

(d) φ_1 is invariant under non-singular linear transformation of $\underline{v} \in V$.

The invariance property of this criterion can be easily seen to follow from formula (1.2.6) for $M(p)$. Suppose $V=\{v_1, v_2,\dots, v_J\}$ is transformed to $\omega=\{w_1, w_2,\dots, w_J\}$ under the linear transformation $w_j = Hv_j$, where H is a $k\times k$ matrix. Then a design assigning weight p_j to w_j has design matrix

$$M_w(p) = W\ P W^t = H\ V\ P\ V^t H^t$$

where V, W are respectively $k\times J$ matrices whose jth column is v_j, w_j.

Then $$\varphi_1\left[M_w(p)\right] = \log\{\ \det\left[M_w(p)\right]\ \}$$

$$= \log\{\ \det\left[H\ V\ P\ V^t\ H^t\right]\ \}$$

$$= \log\{\ \det(V\ P\ V^t)\times\det(H)^2\ \}$$

$$= \log\{\ \det\left[M(p)\right]\ \} + \log\left\{\left[\det(H)\right]^2\right\}$$

$$=\varphi_1\left[M(p)\right] + \text{constant}$$

(II) The A-optimality Criterion:

The criterion of A-optimality is defined by the criterion function:

$$\phi_2(p) = -Trace\left[M^{-1}(p)\right] \qquad (1.3.3)$$

Hence an A-optimum design seeks to minimise the sum of the variances of the least squares estimators or their average variance (A for average) but does not take correlations between these estimates into account. This criterion was considered by Elfving (1952) and Chernoff(1953).

Properties of $\phi_2(p) = \varphi_2\{M(p)\}$:

(a) φ_2 is an increasing function over the set of positive definite symmetric matrices.

(b) φ_2 is a concave on M and strictly concave on M_+; (see Ford(1976)).

(c) ϕ_2 is differentiable whenever it is finite, and the first derivative is

$$\frac{\partial \phi_2}{\partial p_j} = \underline{v}_j^t M^{-2}(p)\underline{v}_j \tag{1.3.4}$$

(III) The G-optimality Criterion:

The G-optimality criterion is defined by the criterion function:

$$\phi_3(p) = -\underset{\underline{v}\in V}{Max}\, \underline{v}^t M^{-1}(p)\underline{v} \tag{1.3.5}$$

This criterion seeks to minimise the maximum of $\underline{v}^t M^{-1}(p)\underline{v}$ which is proportional to the variance of $\underline{v}^t\hat{\theta}$. Kiefer and Wolfowitz(1960) prove the equivalence of this criterion and the D-optimality criterion.

Properties of $\phi_3(p) = \varphi_3\{M(p)\}$:

(a) φ_3 is an increasing function over the set of positive definite symmetric matrices.

(b) φ_3 is concave on M and strictly concave on M_+; (see Ford (1976)).

(c) φ_3 is invariant under a non-singular transformation of $\underline{v} \in V$. To see this consider the same linear transformation mentioned in this section for D-optimality. Then

$$\varphi_3(M_w(p)) = -\underset{w \in \omega}{Max}\, \underline{w}^t M_w^{-1}(p)\underline{w},$$

$$= -\underset{H\underline{v} \in \omega}{Max}(H\underline{v})^t (HVPV^t H^t)^{-1}(H\underline{v}),$$

$$= -\underset{\underline{v} \in H^{-1}\omega}{Max}\, \underline{v}^t H^t (H^t)^{-1}(V\ P\ V^t)^{-1} H^{-1} H\underline{v},$$

$$= -\underset{\underline{v} \in V}{Max}\, \underline{v}^t M(p)^{-1}\underline{v},$$

$$= \varphi_3(M\ (p)).$$

(d) Suppose that uniquely $\underline{v}_i^t M^{-1}(p)\underline{v}_i = \underset{t}{Max}\, \underline{v}_t^t M^{-1}(p)\underline{v}_t$, then ϕ_3 has unique partial derivatives corresponding to positive weights, namely

$$\frac{\partial \phi_3}{\partial p_j} = \left[\underline{v}_j^t M^{-1}(p)\underline{v}_i\right]^2$$

Otherwise ϕ_3 is not differentiable.

(IV) The E-optimality Criterion:

The E-optimality criterion is defined by the criterion function:

$$\phi_4(p) = -\lambda_{max}\left[M(p)\right]^{-1} = -\lambda_{max}^{-1} \quad (1.3.6)$$

where $\lambda_{max}\left[M(p)\right]^{-1}$ denotes the largest eigenvalue of $M^{-1}(p)$ (see Kiefer (1974)).

Properties of $\phi_4(p) = \varphi_4\{M(p)\}$:

(a) φ_4 is an increasing function over the set of positive definite symmetric matrices.

(b) φ_4 is a concave function over M and strictly concave on M_+.

(c) Let $(\lambda_1, \lambda_2, \ldots, \lambda_j)$, λ_{max} denote the eigenvalues and the maximum eigenvalue of $M(p)$ respectively. If λ_{max} is unique then ϕ_4 has unique partial derivatives corresponding to positive weights. Otherwise ϕ_4 is not differentiable.

1.3.2 case(2):

In this case we shall assume that the experimenter is interested only in some of the unknown parameters or some linear combinations of the parameters of the linear model (1.1.1).

Suppose we are interested in s linear combinations of the parameters $\theta_1, \theta_2, \ldots, \theta_k$, namely those s linear combinations which are elements of the vector $A'\underline{\theta}$, where A' is an $s \times k$ matrix of rank $s \leq k$. In particular when $A = \left[I_s : 0\right]$ where I_s is the $s \times s$ identity matrix and 0 is the $s \times (k-s)$ zero matrix, then in this case we are interested only in estimating the first s parameters $\theta_1, \theta_2, \ldots, \theta_s$ of $\underline{\theta} \in \Theta$.

Now if $M(p) \in M_+$ then the variance matrix of the least-squares estimator of $A'\underline{\theta}$ is proportional to the matrix $A'M^{-1}(p)A$. But if $M(p) \notin M_+$, i.e. if $M(p)$ is singular, then the basic requirement for estimating the vector $A'\underline{\theta}$ is that the range space (column space) of A is in the range space of $M(p)$ which results in the invariance of the matrix $A'M^{-}(p)A$ to the choice of generalised inverse $M^{-}(p)$ of $M(p)$ (see Graybill (1983), Theorem(6.6.9)).

Note that a generalised inverse of a matrix M is defined as any matrix M^{-} satisfying the condition $M\ M^{-}\ M = M$. This generalised inverse exists for each matrix M, but it is not unique except when $M(p)$ is a square non-singular matrix, in which case $M^{-} = M^{-1}$ uniquely. A particular example is when $M^{-} = M^{+}$, where M^{+} is the Moore-Penrose generalised inverse [some authors call it the pseudo inverse of the p-inverse (see Seber(1977))] which satisfies three more conditions, namely: $M^{+}M\ M^{+} = M^{+}$ and symmetry of $\left(M^{+}M\right)$ and $\left(M\ M^{+}\right)$.

So a good design will be one which makes the matrix $A'M^{-}(p)A$ as small as possible among $M(p)$, since the variance matrix of the least-squares estimator of $A'\underline{\theta}$ is proportional to $A'M^{-}(p)A$. For this purpose we will consider three alternative criteria.

(I) The D_A-optimality criterion :

Sibson(1974) defined the D_A-optimality criterion by the criterion function:

$$\phi_5(p) = -\log det\left[A'M^{-}(p)A\right] \tag{1.3.7}$$

Properties of $\phi_5(p) = \varphi_5\{M(p)\}$:

(a) φ_5 is an increasing function over the set of positive definite symmetric matrices.

(b) φ_5 is concave on M and strictly concave on M_{+} .

(c) ϕ_5 has unique partial derivatives corresponding to positive weights, namely

$$\frac{\partial \phi_5}{\partial p_j} = \underline{v}_j^t M^-(p) A \left[A^t M^-(p) A \right]^{-1} A^t M^-(p) \underline{v}_j \quad , p_j > 0. \text{(see Appendix 1).}$$

These derivatives are invariant for any generalised inverse $M^-(p)$ of $M(p)$ if $\underline{v}_j$, $j = 1, 2, \ldots, J$ and A are in the column space of $M(p)$ (see Graybill(1983), Theorem(6.6.9) and corollaries(6.6.9.1) ,(6.6.9.2)).

Note that if $A = \left[I_s : 0 \right]$ and we partition the matrix $M(p)$ as follows:

$$M(p) = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^t & M_{22} \end{bmatrix} \begin{matrix} s \\ k-s \end{matrix}$$

$$s \qquad k-s$$

then the matrix $\left(A^t M^-(p) A \right)$ can be expressed of the form:

$\left(M_{11} - M_{12} M_{22}^- M_{12}^t \right)^{-1}$ see Rhode(1965) ,Torsney(1981).

So maximizing ϕ_5 in this particular case is equivalent to maximizing

$$\phi_6(p) = \log \det \left(M_{11} - M_{12} M_{22}^- M_{12}^t \right)$$

which is known as the D_S-optimality criterion (see Karlin and Studden (1966), Atwood(1969), Silvey and Titterington(1973), Silvey(1980)).

(II) The Linear Optimality Criterion :

The linear optimality criterion is defined by the criterion function :

$$\phi_7(p) = -Trace\left[A'M^-(p)A\right] \tag{1.3.8}$$

Properties of $\phi_7(p) = \varphi_7\{M(p)\}$:

(a) φ_7 is an increasing function of the set of positive definite symmetric matrices.

(b) φ_7 is a concave on M and strictly concave on M_+.

(c) φ_7 has unique partial derivatives corresponding to positive weights, namely

$$\frac{\partial \phi_7}{\partial p_j} = \underline{v}_j'M^-(p)AA'M^-(p)\underline{v}_j,\ p_j > 0 .$$

Note that ϕ_2 is the particular case of φ_7 corresponding to $A = I$ and also the case $A = \underline{c}$, where $\underline{c}$ is a $k \times 1$ vector, corresponds to another standard criterion known in the literature as the c-optimality criterion. This is of the form

$$\phi_8(p) = -\underline{c}'M^-(p)\underline{c}. \tag{1.3.9}$$

(III) The E_A-optimality criterion :

The criterion of E_A-optimality is defined by the criterion function:

$$\phi_9(p) = -\lambda_{max}\left[A'M^-(p)A\right] \tag{1.3.10}$$

where λ_{max} denotes the largest eigenvalue of the matrix $A'M^{-}(p)A$; (see Pazman(1986).

Properties of $\phi_9(p) = \varphi_9\{M(p)\}$:

(a) φ_9 is an increasing function over the set of positive definite symmetric matrices.

(b) φ_9 is concave on M and strictly concave on M_+; (see Ford(1976).

(c) The differentiability properties of this criterion are similar to those of E-optimality; see section (1.3.1).

CHAPTER TWO

Further Development of Algorithms for Constructing Optimizing Distributions

2.1 Introduction

2.2 Optimality Conditions

2.3 A class of Algorithm

2.4 Properties of Proposed Algorithm

2.5 Construction of Optimal Designs:Empirical Results on Convergence

2.6 Tables

CHAPTER TWO

Further Development of Algorithms for Constructing Optimizing Distributions

2.1 Introduction

In Chapter (1) we introduced the design problem which seeks to maximise one of the criteria $\phi_i(p)$ subject to the constraint $\sum_{i=1}^{J} p_i = 1 ,\ 0 \le p_i \le 1$.

For some design problems it may be possible to obtain an explicit solution for the optimal weights. But in general this will not be possible and iterative numerical methods of solving design problems are necessary.

In this chapter we are going to study a class of algorithms, indexed by a function which depends on derivatives and a free parameter (say δ) for a constrained maximisation problem which requires the calculation of an optimizing probability distribution.

The performance of the algorithm is investigated in constructing D-optimal designs under optimal choices of the parameter, δ , and in constructing c-optimal designs starting from difficult initial designs ,using an optimal and fixed value of the parameter δ.

First we must establish conditions of optimality .It helps to consider the following general problems ,of which the design problem is an example .

Problem (P1):

" Maximise a criterion $\phi(p)$ subject to the constraints $\sum_{i=1}^{J} p_i = 1; p_i \geq 0.$"

Problem (P2):

" Maximise $\psi(X)$ over the polygon whose vertices are the points $G(v_1), G(v_2),..., G(v_J)$, where $G(\bullet)$ is a given one to one function and $V = \{v_1, v_2,..., v_J\}$ is a known set of vector (or matrix) vertices of fixed dimension. That is solve (P1) for:

$$\phi(p) = \psi\{E_p[G(v)]\},\ X = E_p[G(v)] = \sum_{j=1}^{J} p_j\, G(v_j)\text{"}$$

2.2 Optimality Conditions:

We concentrate on Problem (P2) and define optimality conditions in terms of point to point directional derivatives.

2.2.1 Directional Derivatives :

Let $$f(X,Y,\xi) = \Psi\{(1-\xi)X + \xi Y\}$$

$$F_{\Psi}(X,Y) = \left.\frac{\partial f(X,Y,\xi)}{\partial \xi}\right|_{\xi=0^+} = \lim_{\xi \to 0} \frac{1}{\xi}\left[\psi\{(1-\xi)X + \xi Y\} - \psi(X)\right]$$

$$F_{\Psi}^{(2)}(X,Y) = \left.\frac{\partial^2 f(X,Y,\xi)}{\partial \xi^2}\right|_{\xi=0^+}$$

$F_{\Psi}(X,Y)$ is known as the directional derivative of $\Psi(\bullet)$ at X in the direction of Y; see Whittle (1973). This derivative may well exist in the absence of differentiability of $\Psi(\bullet)$ but we will in general wish to assume such differentiability which implies that

$$F_{\Psi}(X,Y) = (Y-X)'\frac{\partial \Psi}{\partial X} = Trace\left[(Y-X)'\frac{\partial \Psi}{\partial X}\right] \quad (2.2.1)$$

We call F_j a vertex directional derivative of $\Psi(\bullet)$ at X where $F_j = F_{\Psi}\{X, G(v_j)\}$. If $\Psi(\bullet)$ is differentiable, then so is the function $\phi(p)$ = $\Psi\{E_p[G(v)]\}$ and

$$F_j = \frac{\partial \phi}{\partial p_j} - \sum_{i=1}^{J} p_i \frac{\partial \phi}{\partial p_i} \quad (2.2.2)$$

2.2.2 Condition for Local Optimality :

If $\Psi(\bullet)$ is differentiable at $X^* = E_{p^*}\{G(v)\}$, then $\Psi(X^*)$ is a local maximum of $\Psi(\bullet)$ in the feasible region of problem (P2) if

$$
\begin{aligned}
F_j^* &= F_\Psi\{X^*, G(v_j)\}\begin{cases} = 0 \text{ if } p_j^* > 0 \\ \\ \leq 0 \text{ if } p_j^* = 0 \end{cases} && (i) \\
F_j^{*(2)} &= F_\Psi^{(2)}\{X^*, G(v_j)\} \leq 0 \text{ if } p_j^* > 0 && (ii)
\end{aligned}
\qquad (2.2.3)
$$

See Whittle (1973) for a proof. If $\Psi(\bullet)$ is concave on its feasible region then the first order stationarity condition (2.2.3) (i) is both necessary and sufficient for a solution to problem (P2). This of course recovers the General Equivalence Theorem.

2.3 A Class of Algorithm:

Problems (P1) and (P2) have a well defined set of constraints, which are that the variables $p_1, p_2, ..., p_J$ must be positive and sum to 1. An iteration which preserves these and has respectable properties is

$$p_j^{(r+1)} = p_j^{(r)} f\left(d_j, \delta\right) / \sum_{i=1}^{J} p_i^{(r)} f\left(d_i, \delta\right) \tag{2.3.1}$$

where now $d_j = \left.\dfrac{\partial \phi}{\partial p_j}\right|_{p=p^{(r)}}$, while $f(d, \delta)$ satisfies the following conditions:

(a) $f(d, \delta) > 0$.

(b) $f(d, 0) =$ constant$\neq 0$.

(c) $f(d, \delta)$ is strictly increasing in d for some set of δ-values, say $\delta > 0$.

(d) The variable δ is a free parameter.

This type of algorithm was first proposed by Torsney (1977), taking $f(d, \delta) = d^{\delta}$, with δ>0. Subsequent empirical studies include Silvey et al (1978), which is a study of the choice of δ when $f(d, \delta) = d^{\delta}$, and Torsney (1988), which mainly considers $f(d, \delta) = e^{\delta d}$ in a variety of applications, including estimation and image processing problems. We continue these investigations exploring other choices of $f(d, \delta)$ for which an approximate optimal finite δ can be determined.

Of course other iterations for problems like (P2) have been proposed. Vertex direction algorithms which perturb one p_j and change the others proportionately were first proposed by Fedorov (1972) and Wynn (1972). These are useful when many of the p_j are zero at the optimum as happens in design problems . At the other extreme, when all p_j are positive at the optimum or when it has been established which are positive, constrained steepest ascent or Newton type iterations may be appropriate. See Wu (1978) and Atwood (1976,1980) on these respectively. It is in a context intermediate to these, when only a few optimal weights might be zero that iteration (2.3.1) is to be recommended in its raw form. See Torsney (1983) for further discussion of this.

2.4 Properties Of Proposed Algorithm:

2.4.1 General Properties:

Under the conditions impose on $f(d,\delta)$, iterations under (2.3.1) possess the following properties:

(a) $p^{(r)}$ is always feasible, since $p_i^{(r)}$ are normalized.

(b) $F_\phi\left(p^{(r)}, p^{(r+1)}\right) \geq 0$ with equality when d_j corresponding to nonzero p_j are equal, in which case

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(d_j,\delta)}{\sum_{i=1}^{J} p_i^{(r)} f(d_i,\delta)} = \frac{p_j^{(r)} f(d_j,\delta)}{f(d_j,\delta)\sum_{i=1}^{J} p_i^{(r)}} = p_j^{(r)}$$

then $F_\phi\left(p^{(r)}, p^{(r+1)}\right) = F_\phi\left(p^{(r)}, p^{(r)}\right) = 0$.

The property of this inequality can be seen by letting a positive random variable Z take the value $\partial\phi / \partial p_j$ with probability p_j $\left(p_j = p_j^{(r)}\right)$. Then

$$F_{\phi}\left(p^{(r)},\, p^{(r+1)}\right) = \left(\underline{p}^{(r+1)} - \underline{p}^{(r)}\right)^{t}\, \underline{d} \tag{2.4.1}$$

$$= \sum_{j=1}^{J}\left(p_j^{(r+1)} - p_j^{(r)}\right) d_j$$

$$= \sum_{j=1}^{J} p_j^{(r+1)}\, d_j - \sum_{j=1}^{J} p_j^{(r)}\, d_j$$

$$= \frac{\sum_{j=1}^{J} p_j f\left(d_j, \delta\right) d_j}{\sum_{j=1}^{J} p_j f\left(d_j, \delta\right)} - \sum_{j=1}^{J} p_j\, d_j \tag{2.4.2}$$

$$= \frac{\left(\sum_{j=1}^{J} p_j f\left(d_j, \delta\right) d_j\right) - \left(\sum_{j=1}^{J} p_j\, d_j\right)\left(\sum_{j=1}^{J} p_j f\left(d_j, \delta\right)\right)}{\sum_{j=1}^{J} p_j f\left(d_j, \delta\right)} \tag{2.4.3}$$

$$\therefore \quad F_{\phi}\left(p^{(r)},\, p^{(r+1)}\right) = Cov\left[Z, f(Z, \delta)\right] / E\left[f(Z, \delta)\right].$$

If $f\left(Z, \delta\right)$ is increasing in Z it must have nonnegative covariance with Z. This result implies that an increase in the criterion can be obtained by stepping from $p^{(r)}$ to $p^{(r+1)}$ though it does not guarantee that $\phi\left(p^{(r+1)}\right) \geq \phi\left(p^{(r)}\right)$.

(c) Under the above iteration $Supp\left(p^{(r+1)}\right) \subseteq Supp\left(p^{(r)}\right)$.

(d) An iterate $p^{(r)}$ is a fixed point of the iteration (2.3.1) if the derivatives $\partial\phi / \partial p_j^{(r)}$ corresponding to nonzero $p_j^{(r)}$ share a common value. This is a necessary but not a sufficient condition for $p^{(r)}$ to solve (P1) or (P2). Thus in view of the conditions for (local) optimality, a solution to (P2) is a fixed point of the iteration but so also are the solutions to (P2) for any subset of V, see Torsney(1988).

(e) Let $g(\delta) = F\left(p^{(r)}, p^{(r+1)}\right)$. Then from (2.4.2)

$$g(\delta) = \frac{\sum_{j=1}^{J} p_j f(d_j, \delta)\, d_j}{\sum_{j=1}^{J} p_j f(d_j, \delta)} - \sum_{j=1}^{J} p_j\, d_j$$

and then

$$g'(\delta) = \frac{(\sum_{j=1}^{J} p_j f_j)(\sum_{j=1}^{J} p_j f_j' d_j) - (\sum_{j=1}^{J} p_j f_j\, d_j)(\sum_{j=1}^{J} p_j f_j')}{(\sum_{j=1}^{J} p_j f_j)^2}$$

where $f_j = f_j(\delta, d)$ and $f_j' = \partial f_j(\delta, d)/\partial\delta$

$$= \frac{(\sum_{j=1}^{J} p_j f_j)(\sum_{j=1}^{J} p_j f_j\, d_j \frac{f_j'}{f_j}) - (\sum_{j=1}^{J} p_j f_j\, d_j)(\sum_{j=1}^{J} p_j f_j \frac{f_j'}{f_j})}{(\sum_{j=1}^{J} p_j f_j)^2}$$

$$= \frac{(\sum_{j=1}^{J} p_j f_j\, d_j \frac{f_j'}{f_j})}{(\sum_{j=1}^{J} p_j f_j)} - \frac{(\sum_{j=1}^{J} p_j f_j\, d_j)(\sum_{j=1}^{J} p_j f_j \frac{f_j'}{f_j})}{(\sum_{j=1}^{J} p_j f_j)^2}$$

$$= (\sum_{j=1}^{J} q_j\, d_j \frac{f_j'}{f_j}) - (\sum_{j=1}^{J} q_j\, d_j)(\sum_{j=1}^{J} q_j \frac{f_j'}{f_j})$$

$$\therefore\ g'(\delta) = Cov(D, G) \tag{2.4.4}$$

where

$$G = \left(\frac{\partial f(D,\delta)}{\partial \delta}\right) / f(D,\delta) = \frac{\partial\, ln\left[f(D,\delta)\right]}{\partial \delta}$$

and D is a random variable taking the value d_j with probability q_j,

$$q_j = \frac{p_j f(d_j, \delta)}{\sum_{i=1}^{J} p_i f(d_i, \delta)}.$$

2.4.2 Properties Of Specific Cases:

2.4.2.1

To begin with we consider the two choices of the functions $f(d,\delta) = d^\delta$ and $f(d,\delta) = e^{d\,\delta}$ together. These share two properties, namely :

(a) If there is a unique maximum derivative at $p^{(r)}$, say $d_s = \left.\frac{\partial \phi}{\partial p_s}\right|_{p=p^{(r)}}$ then in the case of $f(d,\delta) = e^{d\,\delta}$:

$$\lim_{\delta \to \infty} p^{(r+1)} = \lim_{\delta \to \infty} \frac{p_j\, e^{\delta d_j}}{\sum_{i=1}^{J} p_i\, e^{\delta d_i}}$$

$$= \lim_{\delta \to \infty} \frac{p_j e^{\delta d_j}}{e^{\delta d_s} \sum_{i=1}^{J} p_i \left(\frac{e^{\delta d_i}}{e^{\delta d_s}} \right)} = \lim_{\delta \to \infty} \frac{p_j \left(\frac{e^{\delta d_j}}{e^{\delta d_s}} \right)}{\sum_{i=1}^{J} p_i \left(\frac{e^{\delta d_i}}{e^{\delta d_s}} \right)}$$

$$= \lim_{\delta \to \infty} \frac{p_j \left(\frac{e^{d_j}}{e^{d_s}} \right)^{\delta}}{\sum_{i=1}^{J} p_i \left(\frac{e^{d_i}}{e^{d_s}} \right)^{\delta}} = \begin{cases} 1 & \text{if } s = j \\ 0 & \text{if } s \neq j \end{cases}.$$

So $p^{(r+1)} \to e_s \;\; as \;\; \delta \to \infty$, where e_s is the s^{th} unit vector.

And similarly for the function $f(d, \delta) = d^{\delta}$.

(b) $g(\delta) = F\left(p^{(r)}, p^{(r+1)}\right)$ is nondecreasing in δ. The first property is trivial. In respect of the second we note that the function $G(D)$ of section 2.4.1 (e) is given by $G(D, \delta) = ln(D)$ and $G(D, \delta) = D$ in the two cases respectively. Both are increasing functions and therefore $g'(\delta) = Cov[D, G(D, \delta)] \geq 0$.

Note care must be taken in interpreting the latter . In the optimal design context the vector e_s corresponds to a single point design. For a number of optimal design criteria $\phi(e_s) = -\infty$. The implication is that for such criteria iteration (2.3.1) is unlikely to be monotonic and possibly not convergent if δ is large. In fact non-convergence occurs under the following combinations:

$$\phi_i(p) = \prod_{j=1}^{J} p_j \;\; , \; f(d, \delta) = d^{\delta} \; , \; \delta = 2 \; ;$$

$$\phi_{ii}(p) = \sum_{j=1}^{J} p^{-t} \quad , \; f(d,\delta) = d^{\delta} \; , \; \delta = 2/(t+1) \; ;$$

$$\phi_{iii}(p) = \sum_{j=1}^{J} p_j \ln(p_j) \quad , \; f(d,\delta) = e^{\delta d} \; , \; \delta = 2 \quad .$$

In each case iterations oscillate between two values unless the initial value is the optimizing p^*, which is $p^* = \frac{1}{j}$ for each $\phi\ (p)$.

In contrast this optimum is attained in one step from any initial $p^{(0)}$ if $\delta = 1 \; , \; 1/(t+1) \; , \; 1$ respectively in the three examples. An implication would seem to be that iteration (2.3.1) would be convergent if not monotonic at least for $\delta \le 1 \; , \; \delta \le 1/(t+1) \; , \; \delta \le 1$ in the three examples respectively. For large δ we recall that property (b) in section 2.4.1 only guarantees an increase in the criterion if we take a small enough step from $p^{(r)}$ to what we have defined to be $p^{(r+1)}$. This would mean a different formula from (2.3.1) for the next iterate. If we adopt such a method, property (b) suggests taking $\delta = \infty$. The revised iterative rule would then be a vertex direction one but not a steepest ascent method since $F_{\Psi}(X,Y)$ depends on the distance between X and Y. Constrained steepest ascent techniques choose directions which maximise normalised directional derivatives .

2.4.2.2 :

We again consider two cases of $f(d,\delta)$; namely $f(d,\delta) = \ln(e+\delta d)$ and $f(d,\delta) = F(\delta d)$ where $F(x)$ is increasing in x and bounded above so that it must have asymptote as $x \to \infty$. Examples include cumulative distribution functions. In these examples the following is true:

(a) $p^{(r+1)} \rightarrow p^{(r)}$ as $\delta \rightarrow \infty$;

(b) $g(\delta)$ is maximised by some finite δ, say δ^*.

The first is again trivial. It implies that $g(\infty) = g(0) = 0$ since $F(p,p) = 0$. Given that $g(\delta) \geq 0$ from 2.4.1 (b), property (b) follows.

It is a possibility then that convergence, if not monotonicity are obtained for any δ. An optimal choice might be the δ^* of (b). In general there is no explicit formula for δ^* in terms of $p^{(r)}$ and $d = \left.\frac{\partial \phi}{\partial p}\right|_{p=p^{(r)}}$, (terms on which it must depend), but we can suggest an approximation to it in the case of $f(d,\delta) = F(\delta d)$. Recall that $g'(\delta)$ is a covariance between a random variable D and $G(D,\delta)$ where $G(D,\delta) = \partial \ln[f(D,\delta)] / \partial\delta$. Thus $g'(\delta)$ is likely to be zero if δ is such that $G(D,\delta)$ has a turning point in the range of $d_1, d_2, \ldots, d_J$.

Now

$$\frac{\partial G(d,\delta)}{\partial d} = \frac{\partial\left[\{\partial f(d,\delta)/\partial\delta\}/f(d,\delta)\right]}{\partial d} = \frac{\partial^2 \ln[f(d,\delta]}{\partial d\ \partial\delta} \qquad (2.4.5),$$

and for $f(d,\delta) = F(\delta d)$

$$\frac{\partial f}{\partial\delta} = d\,F'(\delta d) \text{ and } \frac{\partial f}{\partial\delta}\Big/ f = dF'(\delta d)/f \qquad (2.4.6)$$

Then from (2.4.6) the derivatives (2.4.5) will be of the form

$$\frac{\partial G(d,\delta)}{\partial d} = \frac{F(\delta d)\{d\,\delta F''(d\,\delta) + F'(\delta d)\} - \{\delta d\,[F'(\delta d)]^2\}}{[F(\delta d)]^2}$$

$$= \frac{F'(\delta d)}{F(\delta d)} + \frac{\delta d\, F''(\delta d)}{F(\delta d)} - \frac{\delta d\, [F'(\delta d)]^2}{[F(\delta d)]^2} \tag{2.4.7}$$

Now

let $\quad x = \delta d \quad$ and $\quad H(x) = \dfrac{\partial G(d, \delta)}{\partial d}$

then (2.4.7) will be of the form

$$H(x) = \frac{F'(x)}{F(x)} + \frac{x\, F''(x)}{F(x)} - \frac{x\, [F'(x)]^2}{[F(x)]^2} \tag{2.4.8}$$

Let $H(x^*) = 0$. A possibly simplistic suggestion is to approximate δ^* by $\delta^* = x^* / \left(\sum_{i=1}^{J} p_i\, d_i \right)$ or by corresponding terms based on other moments of the d_i's .

We focus attention on this choice of δ in the next section .

(2.5) Construction of Optimal Designs: Empirical Results on Convergence:

(2.5.1) Construction of Optimal Designs:

We report the performance of iteration (2.3.1) in calculating D-optimal designs when $f(d,\delta)$ satisfies the conditions of section (2.4.2.2) and

$\delta = \delta^* = x^* / \left(\sum_{i=1}^{J} p_i d_i \right)$

Optimal regression design problems are examples of (P2) in which

(1) $V \subset \Re^k$ and is called the (induced) design space.

(2) $G(v) = \underline{v}\underline{v}'$.

(3) X is a symmetric $k \times k$ matrix.

(4) A variety of criteria $\Psi(\bullet)$ have been considered including $\Psi(X) = ln\left[det(X)\right]$ which is the D-optimality criterion.

We calculate D-optimal designs for five examples considered by Silvey et al (1978) and Wu (1978). The examples are defined by their design space:

<u>Example (1):</u> $V = V_1 = \left\{ (1,-1,-1)', (1,-1,1)', (1,1,-1)', (1,2,2)' \right\}$

<u>Example (2):</u> $V = V_2 = \left\{ (1,-1,-1)', (1,-1,1)', (1,1,-1)', (1,2,3)' \right\}$

<u>Example (3):</u> $V = V_3 = \left\{ (1,-1,-2)', (1,-1,1)', (1,1,-1)', (1,2,2)' \right\}$

Example (4): $V = V_4 = \{(1,1,-1,-1)', (1,-1,1-1)', (1,-1,-1,-1)', (1,2,2,-1)'$

$(1,2,2,-1)', (1,1,-1,1)', (1,-1.5,1,1)', (1,-1,-1,2)' \}$

Example (5): $V = V_5 = V_4 \cup \{(1,1,1.5,1)'\}$.

In Tables (2.1) -(2.3) we report the number of iterations needed to achieve $max\ F_j \le 10^{-n}, n = 1,2,3,4$ under three choices of $f(d,\delta)$ namely, $f(d,\delta) = \ln(e+\delta d)$; $f(d,\delta) = e^{\delta d} / (1+e^{\delta d})$ (Logistic Distribution Function) and $f(d,\delta) = c - e^{-\delta d}$, $c > 1$. For c close to 1 the last choice of $f(d,\delta)$ is close to an exponential cumulative distribution function. It is clear that on the whole convergence is slow in terms of number of iterations. However arguably it is fast to begin with. It must be remembered too that at each iteration only first derivatives are required. One marginally positive result is that convergence is faster under the case $f(d,\delta) = c - e^{-\delta d}$ with $c = 1.0001$.

Convergence was slower for large values of c. Interestingly if $c \approx 1$ and δ is small then $f(d,\delta) \approx \delta d$. Iterations are then approximately those under $f(d,\delta) = d$. This suggest that $f(d,\delta) = d$ is an efficient choice for D-optimality criterion. Certainly it is known to be monotonic for this criterion.

(2.5.2) The convergence of The Algorithm

We have not addressed the topic of convergence of iteration (2.3.1). So far only isolated results have been established in the literature, and mainly on monotonicity. Titterington (1976) describes a proof of monotonicity of $f(d,\delta) = d$ in the case of D-optimality, while Torsney (1983) establishes a sufficient condition for monotonicity of $f(d,\delta) = d^{\delta}$, $\delta = \delta_t = 1/(t+1)$

when the criterion $\phi(p)$ is a homogeneous function of degree $-t$, $t > 0$ with positive derivatives. He further shows this condition is satisfied by linear design criteria such as the c-optimal and A-optimal criteria. For these $t = 1$ so that $\delta_t = 1/2$. Also the case $f(d,\delta) = d$ sometimes proves to yield EM iterations which are therefore monotonic and convergent. See Dempster et al (1977). The EM algorithm is known to have notoriously slow convergence. This also seems to be the case with iteration (2.3.1). Silverman et al (1990) proposed a smoothed version of the EM algorithm to improve convergence in stereology and emission tomography problems, but convergence per se has not been proved. This too is the case with iteration (2.3.1). The extent of the difficulty is emphasised by the fact that Gaffke and Mather (1990) prove convergence of a wide class of algorithms for design problems but they cannot fit iteration (2.3.1) into their class. Of course convergence results depend on properties of the criterion $\phi(p)$, on the function $f(d,\delta)$ and on c.

We believe that if c is sufficiently small convergence and probably monotonicity will be assured in a wide range of problems. Certainly this happened in many examples .

In the absence of analytic progress we report some empirical results obtained when $f(d,\delta) = d^{1/2}$ for constructing c-optimal designs under fairly testing conditions. The form of this criterion is $-\underline{c}'X^{-}\underline{c}$ for a given vector $\underline{c}$. Pukelsheim and Torsney (1990) report that there always exists a c-optimal design with a linearly independent support and given the support points there is an explicit solution for the optimal weights i.e.

$$p_j^* = \frac{|\eta_j|}{\sum_{i=1}^{J} |\eta_i|} \quad \text{and optimal value for the criterion} \quad \underline{c}' X^- \underline{c} = \left(\sum_{i=1}^{J} |\eta_i| \right)^2 \qquad (5.2.1)$$

where $\underline{\eta} = (XX')^{-1} X \ \underline{c}$.

This combines results of Fellman (1974) and Kitsos et al (1988). Moreover iteration (2.3.1) with $f(d, \delta) = d^{1/2}$ will find this optimum in one step, starting from a design which assigns weights only to the optimal support points. More generally if an initial design $\underline{p}^{(0)}$ has a linearly independent support, this particular case of iteration (2.3.1) will identify the c-optimal design on this support in one step.

Consider $\underline{c} = (1, 2, 3)'$ in examples 1,2 and 3. In each case V contains four points, say v_1, v_2, v_3, v_4. If v_1, v_2, v_3, v_4 represent the four design points of example (1), then the support of the c-optimal design is $\{v_2, v_3, v_4\}$ with optimal weights $p^* = \{0.072,\ 0.214,\ 0.714\}$. In example (2) the fourth point is the only optimal support point. Finally if v_1, v_2, v_3, v_4 represent the four design points of example (3), the optimal support is $\{v_1, v_2, v_4\}$ with optimal weights $p^* = \{0.2,\ 0.2,\ 0.6\}$.

We started iteration (2.3.1) with $f(d, \delta) = d^{1/2}$ from various initial designs $\underline{p}^{(0)}$, which put small weights on at least one of these support points. These included permutations $\underline{p}^{(0)} = (\alpha, \alpha, \alpha, \beta)$, $\underline{p}^{(0)} = (\alpha, \alpha, \beta, \beta)$ and $\underline{p}^{(0)} = (\alpha, \beta, \beta, \beta)$ with $\beta \leq 10^{-12}$. At the first iteration the algorithm irresistibly moves immediately towards the optimal design on the subset of points receiving weights α. However the algorithm slowly moves away from this and converges to the optimum. The numbers of iterations needed to achieve

$max\ F_j \leq 10^{-n}, n = 1, 2, 3, 4$ are recorded in Tables (2.4)-(2.8) for these three examples. Similar results were found in the other two examples.

(2.6) Tables :

In the following three tables we report results when using three choices of $f(d,\delta)$ with $\delta = \delta^* = x^* / \left(\sum_{i=1}^{J} p_i d_i \right)$ (see section (2.4)) in examples 1-5 to calculate D-optimal designs. In particular we record the number of iterations needed to achieve $max\ F_j \leq 10^{-n}, n = 1, 2, 3, 4$ for all $j = 1, 2, \ldots, J$ where F_j are the vertex directional derivatives. We note that $\delta = \delta^* = x^* / \left(\sum_{i=1}^{J} p_i d_i \right) = x^* / k$,(see Appendix 1), for D-optimality criterion when $V \subset \Re^k$.

TABLE (2.1)

$$f(d,\delta) = \ln(e + \delta d)$$

Example	n=1	n=2	n=3	n=4
1	6	25	50	75
2	6	41	89	141
3	6	24	45	66
4	18	121	339	714
5	13	190	488	880

TABLE (2.2)

$$f(d,\delta) = e^{\delta d} / \left(1 + e^{\delta d}\right)$$

Example	n=1	n=2	n=3	n=4
1	7	29	57	86
2	7	48	101	161
3	6	28	52	76
4	20	139	388	815
5	15	217	557	1004

TABLE (2.3)

$$f(d,\delta) = c - e^{-\delta d} \quad , \quad c = 1.0001$$

Example	n=1	n=2	n=3	n=4
1	1	7	14	22
2	3	13	27	43
3	2	7	13	19
4	6	39	109	229
5	5	61	157	283

In the following five tables we recorded the number of iterations needed to achieve max $F_j \leq 10^{-n}, n = 1,2,3,4,5,6$ for different initial designs ,$p^{(0)}$,in the case of $f(\delta, d) = d^{\delta}$, $\delta = 1/2$.

TABLE (2.4)

$$P^{(0)} = (\alpha\ ,\ \alpha\ ,\ \beta\ ,\ \beta)\ \ ,\ \beta = 10^{-12}$$

	n=1	n=2	n=3	n=4	n=5	n=6
Example(1)	16	19	23	26	29	33
Example(2)	4	6	7	8	9	11
Example(3)	4	14	23	32	41	51

TABLE (2.5)

$$P^{(0)} = (\alpha\ ,\ \beta\ ,\ \beta\ ,\ \alpha)\ \ ,\ \beta = 10^{-12}$$

	n=1	n=2	n=3	n=4	n=5	n=6
Example(1)	8	11	15	18	21	25
Example(2)	2	2	2	3	3	3
Example(3)	11	21	30	39	49	58

TABLE (2.6)

$$P^{(0)} = (\beta\ ,\ \alpha\ ,\ \alpha\ ,\ \beta)\ \ ,\ \beta = 10^{-12}$$

	n=1	n=2	n=3	n=4	n=5	n=6
Example(1)	30	33	36	40	43	46
Example(2)	4	5	6	8	9	10
Example(3)	3	6	13	24	33	43

TABLE (2.7)

$$P^{(0)} = (\beta\ ,\ \beta\ ,\ \alpha\ ,\ \alpha)\ \ ,\ \beta = 10^{-12}$$

	n=1	n=2	n=3	n=4	n=5	n=6
Example(1)	3	4	5	9	12	15
Example(2)	2	2	2	2	2	2
Example(3)	26	36	45	54	63	72

TABLE (2.8)

$$P^{(0)} = (\alpha\ ,\ \beta\ ,\ \alpha\ ,\ \beta)\ \ ,\ \beta = 10^{-12}$$

	n=1	n=2	n=3	n=4	n=5	n=6
Example(1)	13	17	20	23	27	30
Example(2)	4	5	7	8	9	10
Example(3)	5	15	24	33	43	52

CHAPTER THREE

Covariance and Correlation Criteria

3.1 Introduction

3.2 Simplifying the Covariance Criterion in the Case of *kxk* Design Matrix

3.3 The Invariance of the Covariance and Correlation Criteria Under Specific Choices of $\underline{a}$ and $\underline{b}$

3.4 Algorithm

3.5 Examples and Discussions

CHAPTER THREE

Covariance and Correlation Criteria

3.1 Introduction:

In chapter one we introduced the main criteria in an optimal design problem i.e. $\phi_i(p)$, $i = 1,2,...,9$. These criteria have the following common properties:

(a) ϕ is a concave function on the set of positive definite symmetric matrices.

(b) ϕ is an increasing function over the set of positive definite symmetric matrices.

The choice of the criterion depends on the aim of the experimenter. If the experimenter is interested in all the unknown parameters $\underline{\theta} \subseteq \Theta \in R^k$, he or she can choose one of the criteria $\phi_i(p), i = 1,...,4$, but if the experimenter is interested in s (linear combinations) of the unknown parameters , then criteria $\phi_i(p), i = 5,..,9$ serve this purpose.

In this chapter we are going to study two new criteria (first discussed by Torsney (1988)) with the aim of estimating one or more of the unknown parameters as independently of the others as possible. Thus we wish to make numerical covariances or correlations between the relevant parameter estimates as small as possible e.g.

$$\text{maximize } \phi_c(p) = -\left[\underline{a}'M^-(p)\underline{b}\right]^2 \text{ subject to } 0 \le p_j \le 1, \ \sum_{i=1}^{J} p_i = 1 \quad (3.1.1)$$

where $\underline{a}$ and $\underline{b}$ are given vectors in $\Re^k$, and

$M(p) = VPV'$, V $k \times J$ design matrix, $P = \text{diag}(p_1, p_2, \ldots, p_J)$, and M^- is any generalised inverse of M. For estimation of $\underline{a}'\underline{\theta}$, $\underline{b}'\underline{\theta}$ (which seems desirable in this context) we require that the range space (column space) of the vectors $\underline{a}$ and $\underline{b}$ are in the range space of M. This guarantees invariance of $\underline{a}'M^-(p)\underline{b}$ to the choice of M^-.See Graybill (1983) corollary (6.6.9.2) ;

or maximize

$$\phi_p(p) = -\frac{\left[\underline{a}'M^-(p)\underline{b}\right]^2}{\left[\underline{a}'M^-(p)\underline{a}\right]\left[\underline{b}'M^-(p)\underline{b}\right]} \text{ subject to } 0 \le p_j \le 1, \sum_{i=1}^{J} p_i = 1. \quad (3.1.2)$$

We call $\phi_c(p)$ and $\phi_p(p)$ the covariance and the correlation criteria respectively. These seem to be new criteria except if $\underline{a} \propto \underline{b}$, when the covariance criterion is equivalent to a c-optimal criterion and the correlation criterion is constant. The covariance and c-optimal criterion are likely to have similar properties. One difference is that its derivatives can be negative. These are

$$\frac{\partial \phi_c}{\partial p_j} = 2\left(\underline{a}'M^-\underline{b}\right)\left(\underline{a}'M^-\underline{v}_j\right)\left(\underline{v}_j'M^-\underline{b}\right) \quad (3.1.3)$$

and also in the case of the correlation criterion the derivatives will be of the form

$$\frac{\partial \phi_p}{\partial p_j} = \phi_p(p)\left\{\frac{2(\underline{a}'M^- \underline{v}_j)(\underline{v}_j' M^- \underline{b})}{(\underline{a}'M^- \underline{b})} - \frac{(\underline{b}'M^- \underline{v}_j)^2}{(\underline{b}'M^- \underline{b})} - \frac{(\underline{a}'M^- \underline{v}_j)^2}{(\underline{a}'M^- \underline{a})}\right\} \qquad (3.1.4)$$

3.2 Simplifying the Covariance Criterion in the Case of $k\times k$ Design Matrix:

Suppose that $M(p)$ is non-singular. Then the covariance criterion will be of the form:

$$\phi_c(p) = -[\underline{a}'M^{-1}(p)\underline{b}]^2$$

$$= -[\underline{a}'(V\ PV')^{-1}\underline{b}]^2 \quad \text{since } M = V\ PV'$$

$$= -[\underline{a}'V'^{-1}P^{-1}V^{-1}\underline{b}]^2 = -[(V^{-1}\underline{a})'P^{-1}(V^{-1}\underline{b})]^2$$

$$= -[\underline{c}'P^{-1}\underline{d}]^2$$

where $\underline{c} = V^{-1}\underline{a}$ and $\underline{d} = V^{-1}\underline{b}$.

Then $$\phi_c(p) = -\left[\sum_{i=1}^{k}\frac{c_i d_i}{p_i}\right]^2 = -[\phi^+(p)]^2 \qquad (3.2.1)$$

where $\phi^+(p) = \sum_{i=1}^{k}\frac{c_i d_i}{p_i}$.

Clearly from (3.2.1) when the number of design points is equal to the number of unknown parameters we can write the covariance criterion as a square of a linear combination of the reciprocals of the weights. Also the value of the criterion depends on the sign of $c_i d_i$, $i = 1,2,...,k$. There are two different cases to distinguish :

(a) When all $c_i d_i$, $i = 1,2,...,k$ have the same sign an explicit solution can be found for the optimal weights p_i, $i = 1,2,...,k$ as we shall see in the next section.

(b) When $c_i d_i$, $i = 1,2,...,k$ have differing signs the criterion $\phi_c(p)$ can be set to zero. We will discus this case in detail later on in Chapter four.

3.2.1 Explicit Solution

Suppose $c_i d_i > o$, $i = 1,2,...,k$,then in this case we can get an explicit solution for the optimal weights (p_i, $i = 1,2,...,k$) as follows :

From (3.2.1) $\phi_c(p) = -\left[\sum_{i=1}^{k} \frac{c_i d_i}{p_i}\right]^2$,

then by taking the first derivatives for $\phi_c(p)$ with respect to the weights p_i, $i = 1,2,...,k$ we find:

$$\frac{\partial \phi_c}{\partial p_j} = -2\left(\sum_{i=1}^{k} \frac{c_i d_i}{p_i}\right)\frac{c_j d_j}{p_j^2} \qquad (3.2.2)$$

But from the optimality conditions we know that at the optimum:

$$\frac{\partial \phi_c}{\partial p_i} = \sum_{j=1}^{k} p_j \frac{\partial \phi_c}{\partial p_j} \qquad (3.2.3)$$

See Whittle (1973).

Then from (3.2.2) and (3.2.3) we get

$$\sum_{j=1}^{k} p_j \frac{\partial \phi_c}{\partial p_j} = \sum_{j=1}^{k} \left[p_j \left(\frac{-2c_j d_j}{p_j^2} \sum_{i=1}^{k} \frac{c_i d_i}{p_i} \right) \right]$$

$$= -2\left(\sum_{j=1}^{k} \frac{c_j d_j}{p_j} \right)\left(\sum_{i=1}^{k} \frac{c_i d_i}{p_i} \right) \quad \text{and then}$$

$$-2\left(\sum_{i=1}^{k} \frac{c_i d_i}{p_i} \right)\frac{c_j d_j}{p_j^2} = -2\left(\sum_{j=1}^{k} \frac{c_j d_j}{p_j} \right)\left(\sum_{i=1}^{k} \frac{c_i d_i}{p_i} \right) \quad \Rightarrow \frac{c_j d_j}{p_j^2} = \left(\sum_{i=1}^{k} \frac{c_i d_i}{p_i} \right)$$

Thus $$p_i^2 = c_i d_i \,/\, \sum_{j=1}^{k} \left(\frac{c_j d_j}{p_j} \right), \quad i = 1,2,\ldots,k \qquad (3.2.4)$$

and by taking the square-root of (3.2.4) we get

$$p_i = \pm \frac{\sqrt{|c_i\, d_i|}}{\sqrt{\sum_{j=1}^{k} \left(\frac{|c_j d_j|}{p_j} \right)}} \quad , \; i = 1,2,\ldots,k$$

But since $p_i > o$ for $i = 1,2,\ldots,k$ and $\sum_{i=1}^{k} p_i = 1$, then

$$p_i^* = \left[\sqrt{|c_i\, d_i|} \,/\, \sqrt{\sum_{j=1}^{k} \left(\frac{|c_j d_j|}{p_j} \right)} \right] / \sum_{j=1}^{k} \left[\sqrt{|c_j\, d_j|} \,/\, \sqrt{\sum_{j=1}^{k} \left(\frac{|c_j d_j|}{p_j} \right)} \right] \quad , \; i = 1,2,\ldots,k$$

$$= \frac{\sqrt{|c_i\, d_i|}}{\sum_{j=1}^{k} \sqrt{|c_j\, d_j|}} \quad , \; i = 1,2,\dots,k \tag{3.2.5}$$

So p^* are the optimal weights for the criterion $\phi_c(p)$ and by substituting from (3.2.5) in (3.2.1) we get

$$\phi_c^*(p^*) = -\left\{ \sum_{i=1}^{k} \frac{c_i d_i}{p_i^*} \right\}^2 = -\left\{ \sum_{i=1}^{k} \frac{c_i d_i}{\sqrt{|c_i d_i|}} \times \sum_{j=1}^{k} \sqrt{|c_j d_j|} \right\}^2$$

$$= -\left\{ \sum_{i=1}^{k} \sqrt{|c_i d_i|} \right\}^4 \tag{3.2.6}$$

which is the maximum value for $\phi_c(p)$.

A similar explicit result would be possible if the support consists of s linearly independent points, $s \leq k$, as happens for the c-optimality criterion. See Pukelsheim and Torsney (1991) .

Note that when all $c_i d_i < o$, $i = 1,2,\dots,k$ the optimal weights p^* and the maximum value for $\phi_c(p)$ are similar to that in (3.2.4) and (3.2.5) respectively.

As an example for this case, consider the quadratic regression model

$$E(y) = \theta_1 + \theta_2 x + \theta_3 x^2 \quad , \quad 1 \leq x \leq 2 \quad .$$

and suppose we are interested to estimate the unknown parameter θ_1 as independently of θ_3 as possible. Thus we want to maximize the criterion $\phi_c(p) = -\left[\underline{a}' M^-(p) \underline{b}\right]^2$ with $\underline{a} = (1,0,0)', \underline{b} = (0,0,1)'$.

Suppose the optimal support points are $\{1, x_0, 2\}$,$1 < x_0 < 2$, then the design matrix will be of the form

$$V = \begin{bmatrix} 1 & 1 & 1 \\ 1 & x_0 & 2 \\ 1 & x_0^2 & 4 \end{bmatrix} \Rightarrow V^{-1} = \left(\frac{1}{\det(V)}\right) \begin{bmatrix} 2x_0(2-x_0) & x_0^2-4 & 2-x_0 \\ -2 & 3 & -1 \\ x_0(x_0-1) & 1-x_0^2 & x_0-1 \end{bmatrix}$$

where $\det(V) = 3x_0 - x_0^2 - 2$.

then $\underline{c} = \frac{1}{\det(V)}\left[2x_0(2-x_0), -2, x_0(x_0-1)\right]$, $\underline{d} = \frac{1}{\det(V)}\left[(2-x_0), -1, (x_0-1)\right]$

and then

$$\underline{cd} = \left(\frac{1}{\det(V)}\right)^2 \left[2x_0\,(2-x_0)^2\ ,\ 2\ ,\ x_0\,(x_0-1)^2\right]^t \qquad (3.2.7)$$

Clearly from (3.2.7) all the $c_i d_i$, $i = 1,2,\ldots,k$ are greater than zero, then by substituting from (3.2.7) in (3.2.5) we find the optimal weights and their support point which are recorded in the following table:

Support points	1	x_0	2
Optimal Weights	$\frac{\sqrt{2x_0}(2-x_0)}{w}$	$\frac{\sqrt{2}}{w}$	$\frac{(x_0-1)\sqrt{x_0}}{w}$

where $w = (2-x_0)\sqrt{x_0} + \sqrt{2} + (x_0-1)\sqrt{x_0}$.

In order to find the third support point x_0^*,by applying Newton-Raphson method (see Nielsen 1964), we have to find the first derivative of the criterion

$$\phi_c(p^*) = -\left\{ \frac{1}{\det(V)} \left[2\sqrt{x_0}\,(2-x_0) + \sqrt{2} + \sqrt{x_0}\,(x_0-1) \right] \right\}^{-1} \qquad (3.2.8)$$

such that $\dfrac{\partial \phi_c(p^*)}{\partial x_0} = 0$; the optimal value is found to be $x_0^* = 1.5$.

Similarly , for the same model we can find another example for which all the $c_i d_i$, $i = 1,2,\ldots,k$ are less than zero by taking $\underline{a} = (0,1,0)'$ and $\underline{b} = (0,0,1)'$. This means we are interested in making the numerical covariance between θ_2 and θ_3 as small as possible. Using the same procedure for the above example we find

Support points	1	x_0	2
Optimal Weights	$\dfrac{(2-x_0)\sqrt{(2+x_0)}}{w}$	$\dfrac{\sqrt{3}}{w}$	$\dfrac{(x_0-1)\sqrt{(x_0+1)}}{w}$

where $w = (2-x_0)\sqrt{2+x_0} + \sqrt{3} + (x_0+1)\sqrt{x_0+1}$ and the third optimal support point is $x_0^* = 1.5$.

In fact we can deduce what the optimal designs are for design intervals of the form $[c,2c], c > 0$ in view of the following result .

3.3 The invariance of the Covariance and Correlation Criteria Under Specific Choices of $\underline{a}$ and $\underline{b}$:

Suppose each of the vectors $\underline{a}$ and $\underline{b}$ have exactly one non-zero component, say, a_i and b_j respectively; $i,j=1,2,...,k$, and suppose the design space $V=\{v_1 , v_2 ,...,v_J \}$ is transformed to $\omega=\{w_1 , w_2 ,...,w_J \}$ under the transformation $W=D\ V$,where $D=diag(d_1,d_2,...,d_k)$. Then the design assigning weights p_i to w_i has the design matrix

$$M_w(p)=W PW'=DVPV'D$$

and $M_w^{-1}(p)=(DVPV'D)^{-1}=D^{-1}(VPV')^{-1}D^{-1}=D^{-1}M^{-1}(p)D^{-1}$, where V,W are respectively $k\times J$ matrices whose jth column is v_j, w_j. Then

$$\phi_c(p_w)=-\left[\underline{a}'\,M_w^{-1}(p)\underline{b}\right]^2=-\left[\underline{a}'D^{-1}M_w^{-1}(p)D^{-1}\underline{b}\right]^2$$

$$=-\left[\left(d_i^{-1}d_j^{-1}\right)\underline{a}'\,M^{-1}(p)\underline{b}\right]^2 \text{ since } \underline{a}'D^{-1}=d_i^{-1}\underline{a}',\underline{b}'D^{-1}=d_j^{-1}\underline{b}',$$

$$=-\left(d_i^{-1}d_j^{-1}\right)^2\times\left[\underline{a}'\,M^{-1}(p)\underline{b}\right]^2$$

$$=\phi_c(p)\times\text{constant}.$$

Similarly, for the correlation criterion we find

$$\phi_\rho(p_w)=\frac{-\left[\underline{a}'\,M_w^{-1}(p)\underline{b}\right]^2}{\left(\underline{a}'\,M_w^{-1}(p)\underline{a}\right)\left(\underline{b}'\,M_w^{-1}(p)\underline{b}\right)}$$

$$=-\frac{\left(d_i^{-1}d_j^{-1}\right)^2}{(d_i^{-1})^2(d_j^{-1})^2}\times\frac{\left[\underline{a}'\,M^{-1}(p)\underline{b}\right]^2}{\left(\underline{a}'\,M^{-1}(p)\underline{a}\right)\left(\underline{b}'\,M^{-1}(p)\underline{b}\right)}=\phi_\rho(p)$$

3.4 Algorithm:

To find optimal designs we will use a version of algorithm (2.3.1) i.e.

$$p_i^{(r+1)} = \frac{p_i^{(r)} f(d_i,\delta)}{\sum_{j=1}^{J} p_j^{(r)} f(d_j,\delta)} \quad , \quad i = 1,2,\ldots,J \qquad (3.4.1)$$

Since the covariance and the correlation criteria can have negative derivatives we need a choice of $f(d,\delta)$ which is defined for negative *d*. Indeed this was the main reason why Torsney (1988) considered choices of $f(d,\delta)$ such as $e^{\delta d}$. In its conception algorithm (2.3.1) was evolved for standard optimal design criteria which have positive derivatives and $f(d,\delta) = d^{\delta}$ proved to be a natural choice for particular values of δ; in particular $\delta=1$ for D-optimality and $\delta=1/2$ for c-optimality yield monotonic iterations .See Torsney (1983).

Since the covariance and correlation criteria are design criteria we have explored the use of

$$f(\delta,d_i) = \begin{cases} (1+d_i)^{\delta} & if \;\; d_i \geq o \\ (1-d_i)^{-\delta} & if \;\; d_i < o \end{cases} \qquad \text{where} \quad d_i = \left.\frac{\partial\phi}{\partial p_i}\right|_{\underline{p}=\underline{p}^{(r)}} \qquad (3.4.2)$$

in finding designs which optimise them.

3.5 Examples and Discussion

In this section we consider some examples with the aim of maximising the covariance or the correlation criteria subject to $\sum_{i=1}^{J} p_i = 1$ and $0 \le p_i \le 1$ by using the algorithm (3.4.1) under the choice of $f(d,\delta)$ defined in (3.4.2).

Example (1):

This an example for the covariance criterion in the case of the quadratic regression model:

$$E(y) = \theta_1 + \theta_2 x + \theta_3 x^2 \quad , \quad x \in [1,2].$$

we are interested to make three numerical covariances between the parameters $\underline{\theta} = (\theta_1,\ \theta_2,\ \theta_3)'$ as small as possible i.e. by using different choices of the vector $\underline{a}$, namely, $\underline{a} = (1,0,0)'$, $\underline{a} = (0,1,0)'$ and $\underline{a} = (-1,1,0)'$ while $\underline{b} = (0,0,1)'$always.

In all these cases we find the optimal support points, namely $supp(p^*) = \{1\ ,\ 1.5\ ,\ 2\}$ and corresponding optimal weights which are recorded in the following Table:

Table(3.1)	**optimal weights** (p^*)			$\underline{a}' M^{-1}(p^*)\underline{b}$	$\phi_p(p)$
$\underline{a} = (1,0,0)'$	0.2994	0.4889	0.2117	133.875	-0.9392
$\underline{a} = (0,1,0)'$	0.2705	0.5009	0.2286	-191.326	-0.986
$\underline{a} = (-1,1,0)'$	0.2826	0.4957	0.2217	-325.56	-0.970

In the following tables (3.2 to 3.4) we record for different choices of $\underline{a}$ and $\underline{b}$ the value of delta (δ) which attained $\max F_i \leq 10^{-4}$ in the smallest number of iterations when $p^{(0)}$ assigns equal weight to J equally spaced points in $[1,2]$ for $J = 21$, 11 , 3 respectively. Also recorded the number of iterations needed to achieve $\max F_i \leq 10^{-n}, n = 1,2,3,4$.

Table(3.2) $\underline{a} = (1,0,0)^t$ *and* $\underline{b} = (0,0,1)^t$ $\underline{a}^t M^{-1}(p^*)\underline{b} = 133.875$

δ	$p^{(0)}$	*n=1*	*n=2*	*n=3*	*n=4*
.95	*1/21*	265	374	384	444
.9	*1/11*	77	92	107	122
.5	*1/3*	2	2	2	2

Table(3.3) $\underline{a} = (0,1,0)^t$ *and* $\underline{b} = (0,0,1)^t$ $\underline{a}^t M^{-1}(p^*)\underline{b} = -191.328$

δ	$p^{(0)}$	*n=1*	*n=2*	*n=3*	*n=4*
.95	*1/21*	285	345	405	465
.9	*1/11*	82	98	113	128
.5	*1/3*	2	2	2	2

Table(3.4) $\underline{a} = (-1,1,0)^t$ *and* $\underline{b} = (0,0,1)^t$ $\underline{a}^t M^{-1}(p^*)\underline{b} = -325.56$

δ	$p^{(0)}$	*n=1*	*n=2*	*n=3*	*n=4*
.95	*1/21*	312	371	431	491
.9	*1/11*	89	104	120	135
.5	*1/3*	2	2	2	2

See figures (3.1 to 3.9) for plots of number of iterations needed to achieve $\max F_i \leq 10^{-4}$ against δ , $0 < \delta < 1$.

Example(2)

This is another example for the covariance criterion in the case of the model $E(y) = \theta_0 x + \theta_1 x^{1/2} + \theta_2 x^2$ which describes the relationship between the viscosity y and the concentration x, $0 < x \leq .2$, of a chemical solution. This stems from a practical problem reported by Torsney (1981). Interest was in the same choices of $\underline{a}$ and $\underline{b}$ as in the last example.

We find that the optimal support points in all these cases are the same for each choice of $\underline{a}$ and $\underline{b}$, namely, $supp(p^*) = \{.02\,,\ .12\,,\ .2\}$ and the corresponding optimal weights are as follows:

Table(3.5)	**optimal weights** (p^*)			$\underline{a}'M^{-1}(p^*)\underline{b}$	$\phi_p(p)$
$\underline{a} = (1,0,0)'$	0.4233	0.4050	0.1717	-38565.6	-0.889
$\underline{a} = (0,1,0)'$	0.5089	0.3468	0.1443	6909.34	-0.650
$\underline{a} = (-1,1,0)'$	0.4365	0.3959	0.1675	45649.5	-0.857

In the following Tables (3.6 to 3.8) we report the same information as in tables (3.2 - 3.4) for J equally spaced points in $[0.02, 0.2]$ for $J = 19, 10, 3$.

Table(3.6) $\underline{a} = (1,0,0)'$ *and* $\underline{b} = (0,0,1)'$ $\underline{a}'M^{-1}(p^*)\underline{b} = -38565.6$

δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
.95	*1/19*	637	747	821	889
.9	*1/10*	131	145	158	170
.5	*1/3*	2	2	2	2

Table(3.7) $\underline{a} = (0,1,0)^t$ *and* $\underline{b} = (0,0,1)^t$ $\underline{a}^t M^{-1}(p^*)\underline{b} = 6909.34$

δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
.95	1/19	313	353	393	433
.9	1/10	88	98	110	120
.5	1/3	2	2	2	2

Table(3.8) $\underline{a} = (-1,1,0)^t$ *and* $\underline{b} = (0,0,1)^t$ $\underline{a}^t M^{-1}(p^*)\underline{b} = 45649.51$

δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
.95	1/19	603	668	734	809
.9	1/10	128	141	154	168
.5	1/3	2	2	2	2

See figures (3.10 to 3.18) for plots of number of iterations needed to achieve max $F_i \leq 10^{-4}$ against δ , $0 < \delta < 1$.

Example(3):

Finally, we consider the cubic regression model $E(y) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3$, $x \in [1,2]$ as an example for the covariance criterion under different choices of the vector $\underline{a}$, namely, $\underline{a} = (1,0,0,0)^t$, $\underline{a} = (0,1,0,0)^t$ $\underline{a} = (-1,1,0,0)^t$ while $\underline{b} = (0,0,0,1)^t$ always.

In this example the algorithm converges to the same four support points for each choice of $\underline{a}$ and $\underline{b}$, namely , $supp(p^*) = \{1 , 1.2 , 1.8 , 2\}$ and the corresponding optimal weights are as follows :

Table(3.9)	**optimal weights** (p^*)				$\underline{a}^t M^{-1}(p^*)\underline{b}$	$\phi_p(p)$
$\underline{a} = (1,0,0,0)^t$	0.2237	0.3403	0.2778	0.1582	-3373.012	-0.9320
$\underline{a} = (0,1,0,0)^t$	0.2099	0.3332	0.2899	0.1670	7231.389	-0.9689
$\underline{a} = (-1,1,0,0)^t$	0.2144	0.3354	0.2860	0.1642	10609.01	-0.9586

In the following Tables (3.10 - 3.11) we again report the same information as in Tables (3.2 -3.4) for two choices of J .

Table(3.10) $\underline{a} = (1,0,0,0)'$ *and* $\underline{b} = (0,0,0,1)'$ $\underline{a}'M^{-1}(p^*)\underline{b} = -3373.011$

δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
.95	*1/11*	927	1.40	1152	1258
.5	*1/3*	2	2	2	2

Table(3.11) $\underline{a} = (0,1,0,0)'$ *and* $\underline{b} = (0,0,0,1)'$ $\underline{a}'M^{-1}(p^*)\underline{b} = 7231.38$

δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
.95	*1/11*	959	2195	2409	2634
.5	*1/3*	2	2	2	2

Table(3.12) $\underline{a} = (-1,1,0,0)'$ *and* $\underline{b} = (0,0,0,1)'$ $\underline{a}'M^{-1}(p^*)\underline{b} = 10609.01$

δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
.95	*1/11*	1608	1787	1960	2228
.5	*1/3*	2	2	2	5

See figures (3.19 to 3.24) for plots of number of iterations needed to achieve max $F_i \leq 10^{-4}$ against δ , $0 < \delta < 1$ . .

Example (4)

This is an example for the correlation criterion for the same model and same choices of vectors $\underline{a}$ and $\underline{b}$ mentioned in Example (2).

The optimal support points and corresponding optimal weights are recorded in the following table:

Table(3.13)	$\underline{a}=(1\ ,\ 0\ ,\ 0)'$			$\underline{a}=(0\ ,\ 1\ ,\ 0)'$			$\underline{a}=(-1\ ,\ 1\ ,\ 0)'$		
$supp(p*)$	0.02	0.12	0.2	0.02	0.14	0.2	0.02	0.12	0.2
p^*	.0133	.9845	.0022	.0158	.9810	.0032	.0132	.9849	.0019
$\phi_\rho(p^*)$	-0.8155			-0.5537			-0.7755		

Clearly from Table(3.13) algorithm (3.4.1) converges to three design points under the three choices of $\underline{a}$ and $\underline{b}$. These designs have the unusual feature of one large weight corresponding to the middle support point and small weights for the other end points. Also the convergence is slow in terms of the number of iterations. This would seem to be due to a combination of small weights and zero homogeneity of the correlation criterion. The latter implying zero partial derivatives at the optimum corresponding to the positive optimal weights. This is because the partial derivatives and the directional derivatives are equal under zero homogeneity.

In general $$F_j=\frac{\partial\phi_\rho}{\partial p_j}-\sum_{i=1}^{J}p_i\frac{\partial\phi_\rho}{\partial p_i}\ ,\ i=1,2,..,J\quad .$$

But since $\phi_\rho(p)$ is homogeneous of degree zero in the weights $\underline{p}$ then

$$\sum_{i=1}^{J}p_i\frac{\partial\phi_\rho}{\partial p_i}=0\times\phi_\rho(p)=0\ ,\ i=1,2,..,J\ ,$$

Thus when the algorithm approaches what appears to be the optimum in this case both derivatives and some weights will be small. Accordingly, proceeding from $p^{(r)}$ *to* $p^{(r+1)}$ will only slightly change criterion and weight values .In order to improve the convergence in such cases, we might use $f(d,\delta) = (1+s\,\alpha d)^{s\delta}$ instead of $f(d,\delta) = (1+s\,d)^{s\delta}$, for appropriate α .For example ,if the initial weights are $p_i^{(0)} = 1/3, i = 1,2,3$ which puts equal weights to the support points, then in this case convergence improves under the choice of $f(d,\delta) = (1+s\,\alpha d)^{s\delta}$ for $\alpha = 10^{2n}$ *if* $\max F_j < 10^{-n}, n = 1,2,3$ compared to the case of $f(d,\delta) = (1+s\,d)^{s\delta}$; see Table(3.14).

For this criterion a common value of $\delta \approx 1$ attained ,in all these examples, $\max F_j < 10^{-2}$ in the smallest number of iterations, but the value of δ which attained $\max F_j < 10^{-3}$ varied between examples. These are recorded in Table(3.14) and Table(3.15).

Table(3.14) This table illustrates the number of iterations needed to achieve $\max F_j < 10^{-n}$ in the case of $f(d,\delta) = (1+s\,\alpha d)^{s\delta}$ for $\alpha = 1$ and $\alpha = 10^{2n}, n = 1,2,3$.

	$\alpha = 1$				$\alpha = 10^{2n}, n = 1,2,3$			
$\underline{a} \backslash n$	δ	n=1	n=2	n=3	δ	n=1	n=2	n=3
$\underline{a} = (1,0,0)'$	.07	0	405	909	.008	0	36	51
$\underline{a} = (0,1,0)'$	.07	0	543	1113	.007	0	38	51
$\underline{a} = (-1,1,0)'$	.05	0	548	1222	.007	0	42	59

Table(3.15) This table illustrates the number of iterations needed to achieve $\max F_j < 10^{-n}, n = 1,2$ for different value of δ in the case of $f(d,\delta) = (1+s\,\alpha d)^{s\delta}$ for $\alpha = 1$.

	$\underline{a} = (1,0,0)'$		$\underline{a} = (0,1,0)'$		$\underline{a} = (-1,1,0)'$	
$\delta \backslash n$	$n=1$	$n=2$	$n=1$	$n=2$	$n=1$	$n=2$
0.1	0	284	0	272	0	274
0.2	0	142	0	136	0	138
0.3	0	95	0	91	0	92
0.4	0	72	0	68	0	69
0.5	0	57	0	55	0	56
0.6	0	48	0	46	0	46
0.7	0	41	0	39	0	40
0.8	0	36	0	34	0	**35**
0.9	0	32	0	**31**	0	>35
1.0	0	29	0	>31	0	>35
1.1	0	**27**	0	>31	0	>35

(3.6) The convergence of the algorithm :

From these results we note that for the covariance criterion algorithm (3.4.1) with $f(d,\delta) = (1+s\,d)^{s\delta}$, $s = sign(d)$ converges to three design points in the case of the quadratic model and the model mentioned in example(2) , while in the case of the cubic regression model it converges to four design points.

Also the number of iterations needed to achieve max $F_i \leq 10^{-n}, n = 1,2,3,4$ depends on the number of points in the initial design and on the value of δ. For instance, if the initial design consists only of the supporting points of the optimal design and if $\delta = .5$ then the optimal design may be obtained in two steps. In contrast,when the initial design consists of J points for $J=11$ or 21, higher values of δ, namely, $\delta = .9$ or $.95$ respectively, attain max $F_i \leq 10^{-4}$ in the smallest number of iterations. In terms of the number of iterations the convergence is slow especially in the case of the cubic regression model when $J \geq 11$, but it can be improved by setting weights to zero when $p_j < \epsilon_1$ and $F_j < -\epsilon_2$ for some small ϵ_1, ϵ_2 or just when p_j goes below a fixed value ϵ $(= 0.00001)$. Moreover $f(d,\delta) = |d|^{1/2}$ would also attain the optimum on one step in such circumstances in the case of our examples since derivatives corresponding to positive weights share a common sign (see section 3.2).

Similar results have been obtained by Fellman (1989) for the c-optimality criterion when $f(d,\delta) = d^{\delta}$. In particular as noted by Torsney (1983) $f(d,\delta) = d^{1/2}$ attains the optimum in one step for the c-optimality criterion when the support points form a linearly independent set of vectors. Clearly $f(d,\delta) = (1 + s\,d)^{s\,\delta}$, $s = sign(d)$ has similar effects .

We note finally that it would be unwise to take δ to be too large a value in view of the following the result that $p^{(r+1)} \rightarrow e_m$ *as* $\delta \rightarrow \infty$ where e_m is the m^{th} unit vector, assuming that $d_m = \left.\frac{\partial \phi_c}{\partial p_m}\right|_{p=p^{(r)}}$ is a unique maximum derivative at $p^{(r)}$.

Proof:

$$\lim_{\delta\to\infty} p_j^{(r+1)} = \lim_{\delta\to\infty} \frac{p_j(1+s\,d_j)^{s\,\delta}}{\sum_{i=1}^{J} p_i(1+s\,d_i)^{s\,\delta}} = \lim_{\delta\to\infty} \frac{p_j\left(\dfrac{1+s\,d_j}{1+s\,d_m}\right)^{s\,\delta}}{\sum_{i=1}^{J} p_i\left(\dfrac{1+s\,d_i}{1+s\,d_m}\right)^{s\,\delta}} \quad , \quad s=\text{sign}(d)$$

but $\left(\dfrac{1+d_j}{1+d_m}\right)^{\delta} \to 0$ as $\delta\to\infty, j\neq m$ since $0<\dfrac{1+d_j}{1+d_m}<1$ *if all* $d_i>0$

and

$\left(\dfrac{1-d_j}{1-d_m}\right)^{-\delta} \to 0$ as $\delta\to\infty, j\neq m$ since $0<\dfrac{1-d_m}{1-d_j}<1$ *if all* $d_i<0$,and

also

$$\frac{(1-d_j)^{-\delta}}{(1+d_m)^{\delta}} = \left(\frac{1}{(1+d_m)(1-d_j)}\right)^{\delta} \to 0 \text{ as } \delta\to\infty, j\neq m$$

since $0<\dfrac{1}{(1-d_j)(1+d_m)}<1$ if the derivatives have a mixture of signs.

Then

$$\lim_{\delta\to\infty} p_j^{(r+1)} = \lim_{\delta\to\infty} \frac{p_j\left(\dfrac{1+s\,d_j}{1+s\,d_m}\right)^{s\,\delta}}{p_j+\sum_{i,i\neq j}^{J} p_i\left(\dfrac{1+s\,d_i}{1+s\,d_m}\right)^{s\,\delta}} = \begin{cases} 1 & \text{if} \quad j=m \\ 0 & \text{if} \quad j\neq m \end{cases} .$$

The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve max $F_i \leq 10^{-4}$ against delta $(0 < \delta \leq 1)$ in the case of the quadratic regression model when $\underline{a} = (1,0,0)^t$ and $\underline{b} = (0,0,1)^t$.

Figure (3.1) $p^{(o)} = 1/21$

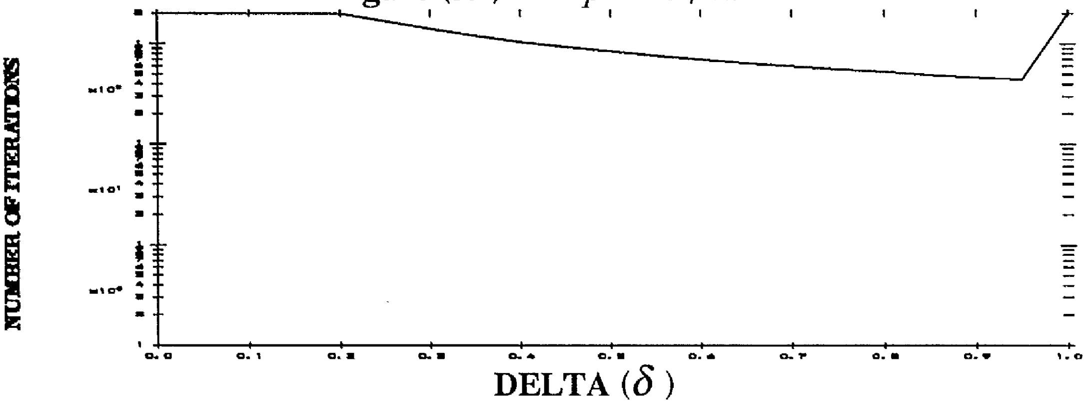


Figure (3.2) $p^{(o)} = 1/11$

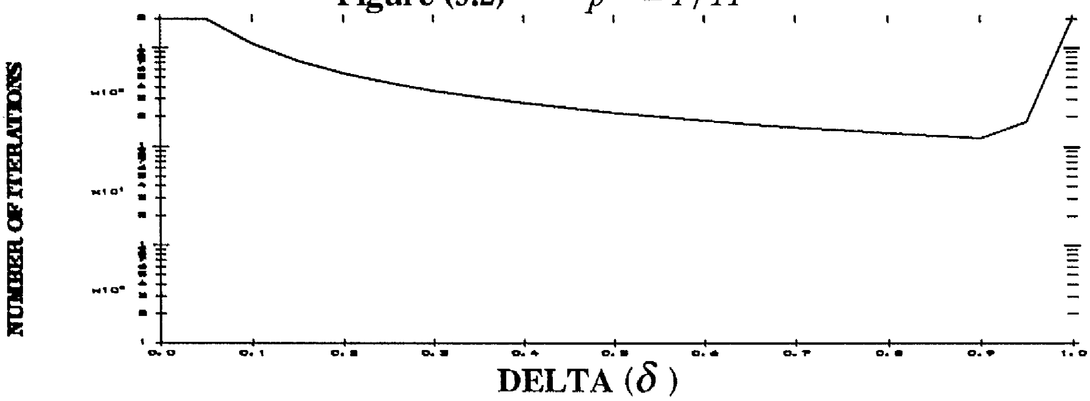
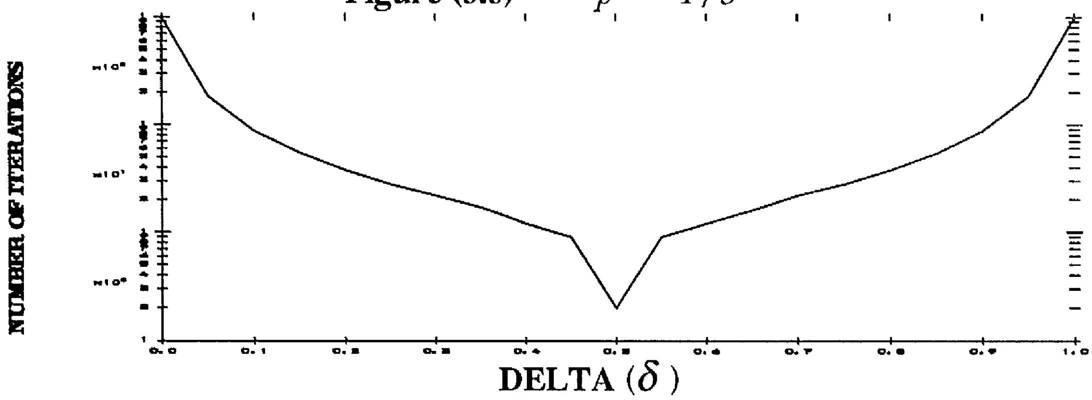


Figure (3.3) $p^{(o)} = 1/3$



The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve max $F_i \leq 10^{-4}$ against delta $(0 < \delta \leq 1)$ in the case of the quadratic regression model when $\underline{a} = (0,1,0)'$ and $\underline{b} = (0,0,1)'$.

Figure (3.4) $p^{(o)} = 1/21$

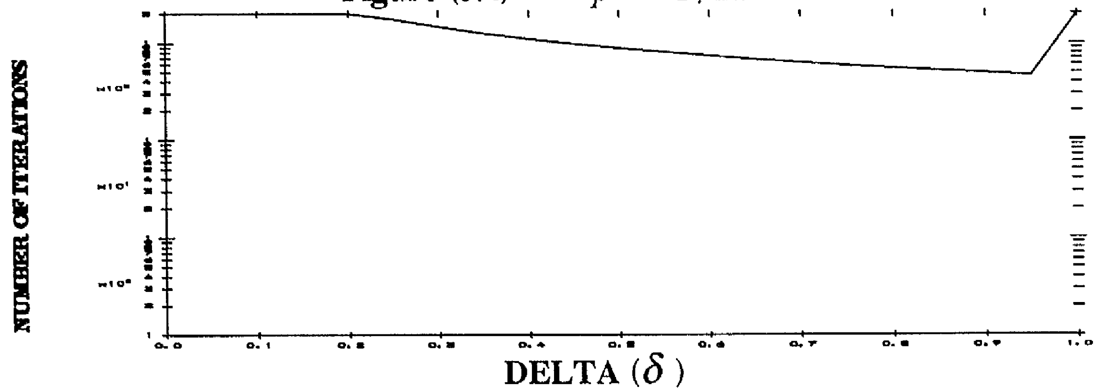


Figure (3.5) $p^{(o)} = 1/11$

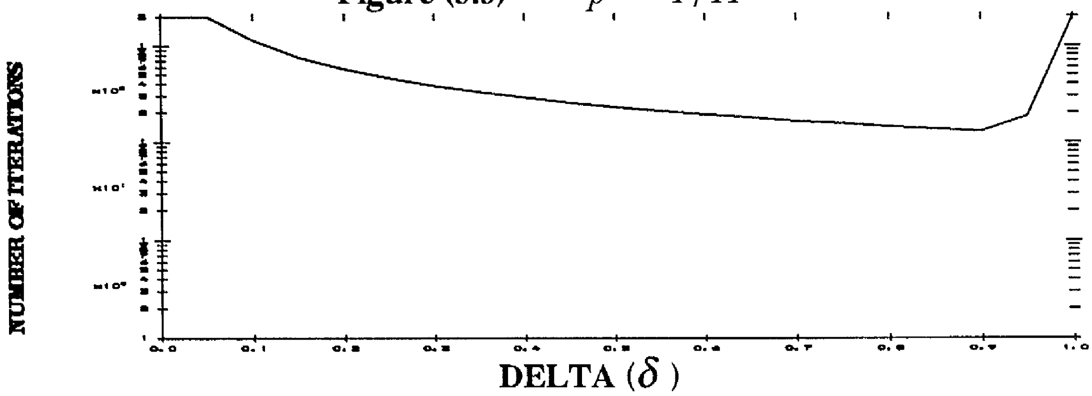
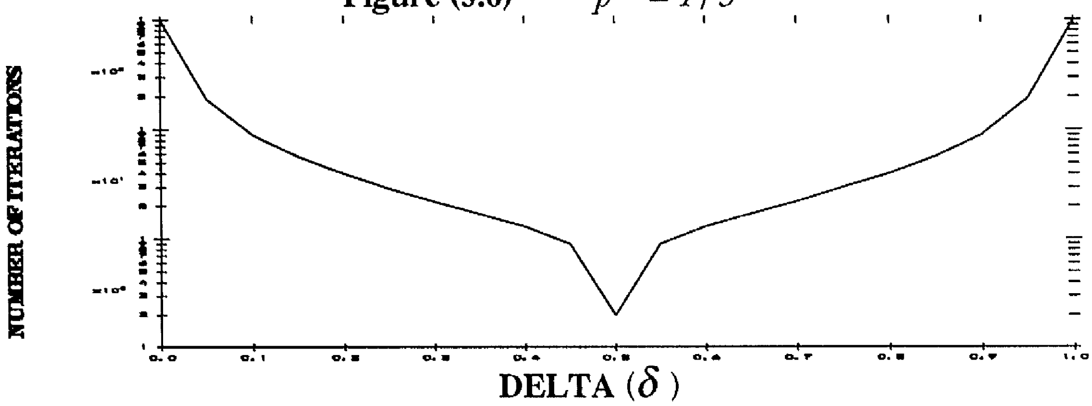


Figure (3.6) $p^{(o)} = 1/3$



The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve max $F_i \le 10^{-4}$ against delta $(0 < \delta \le 1)$ in case of the quadratic regression model when $\underline{a} = (-1,1,0)^t$ and $\underline{b} = (0,0,1)^t$.

Figure (3.7) $p^{(o)} = 1/21$

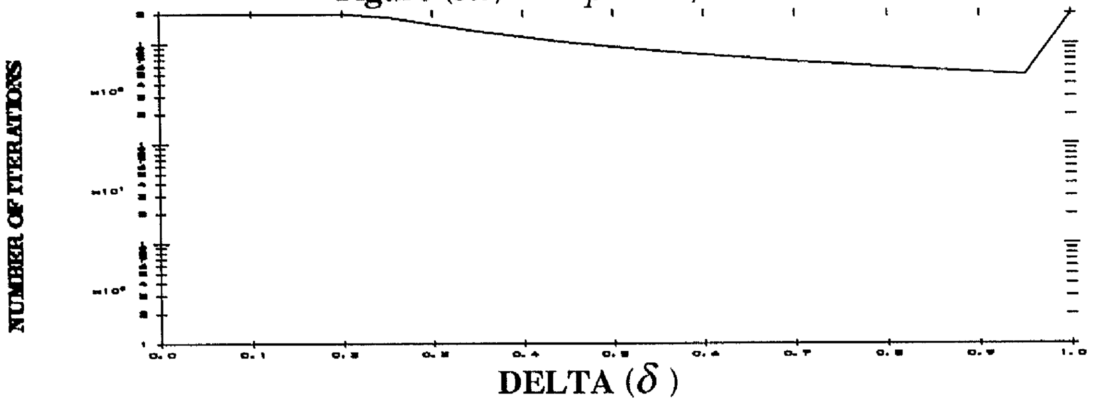


Figure (3.8) $p^{(o)} = 1/11$

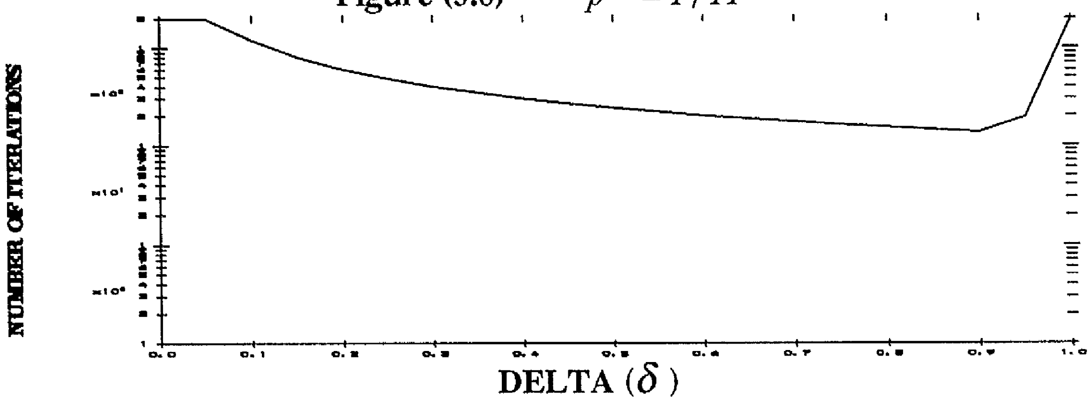
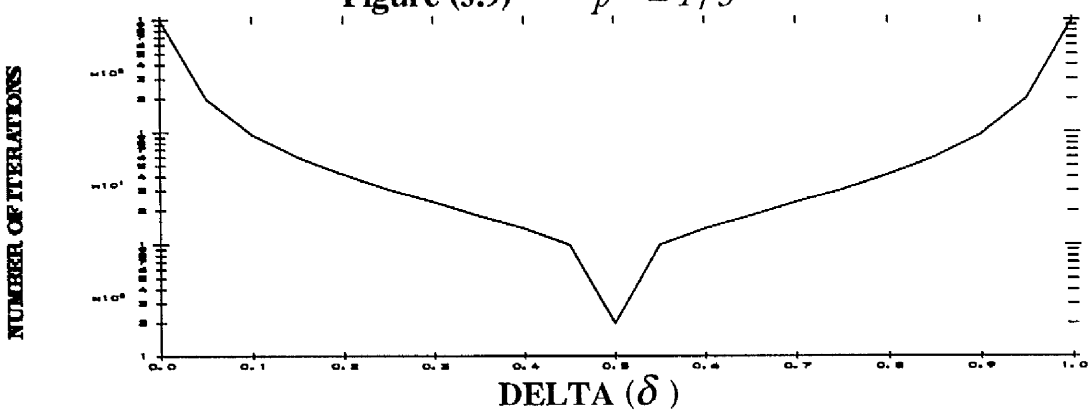
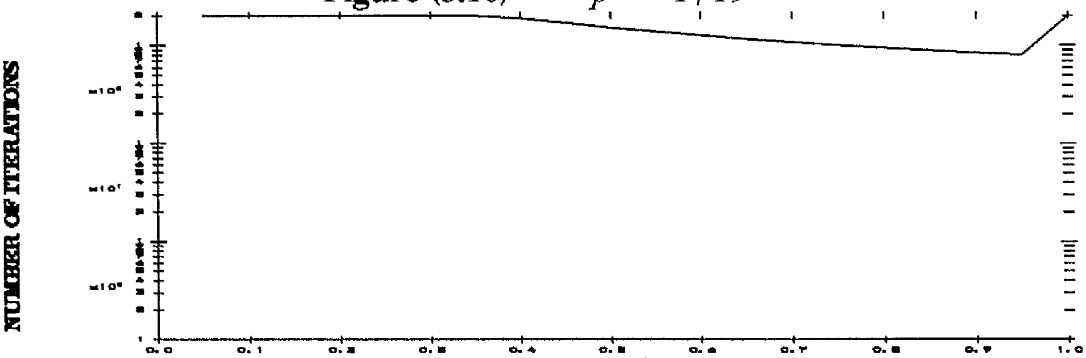


Figure (3.9) $p^{(o)} = 1/3$



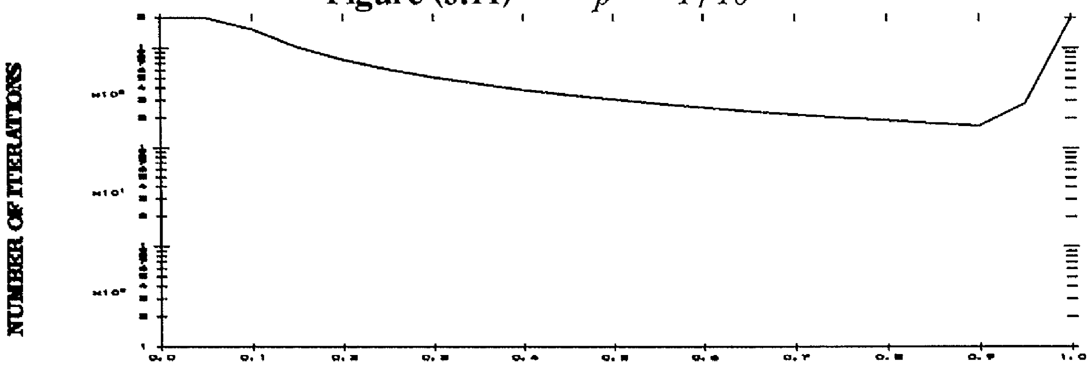
The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve max $F_i \leq 10^{-4}$ against delta ($0 < \delta \leq 1$) in case of the model $E(y) = \theta_0 x + \theta_1 x^{1/2} + \theta_2 x^2$ when $\underline{a} = (1,0,0)^t$ and $\underline{b} = (0,0,1)^t$.

Figure (3.10) $p^{(o)} = 1/19$



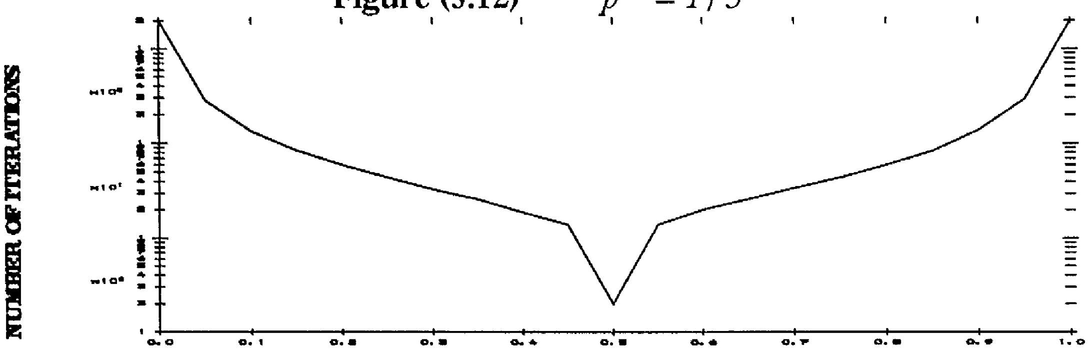
DELTA (δ)

Figure (3.11) $p^{(o)} = 1/10$



DELTA (δ)

Figure (3.12) $p^{(o)} = 1/3$



DELTA (δ)

The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve $\max F_i \le 10^{-4}$ against delta $(0 < \delta \le 1)$ in case of the model $E(y) = \theta_0 x + \theta_1 x^{1/2} + \theta_2 x^2$ when $\underline{a} = (0,1,0)'$ and $\underline{b} = (0,0,1)'$.

Figure (3.13) $p^{(o)} = 1/19$

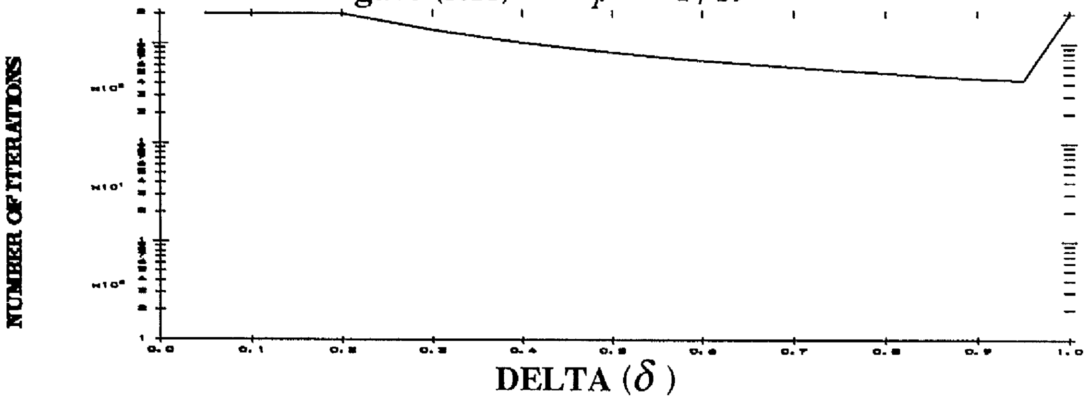


Figure (3.14) $p^{(o)} = 1/10$

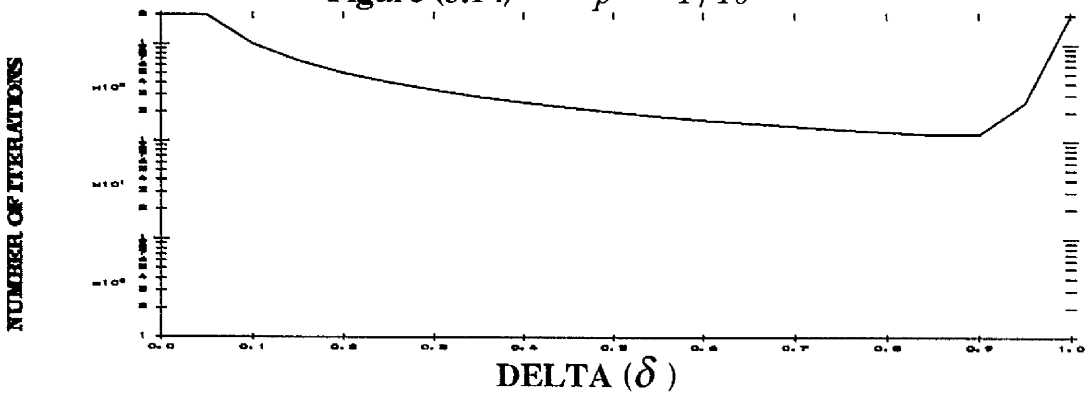
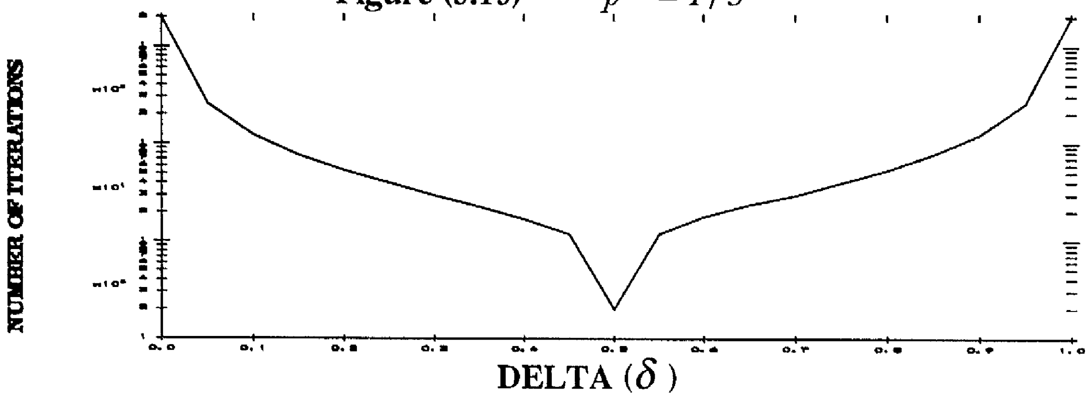


Figure (3.15) $p^{(o)} = 1/3$



The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve $\max F_i \leq 10^{-4}$ against delta $(0 < \delta \leq 1)$ in case of the model $E(y) = \theta_0 x + \theta_1 x^{1/2} + \theta_2 x^2$ when $\underline{a} = (-1,1,0)'$ and $\underline{b} = (0,0,1)'$.

Figure (3.16) $p^{(o)} = 1/19$

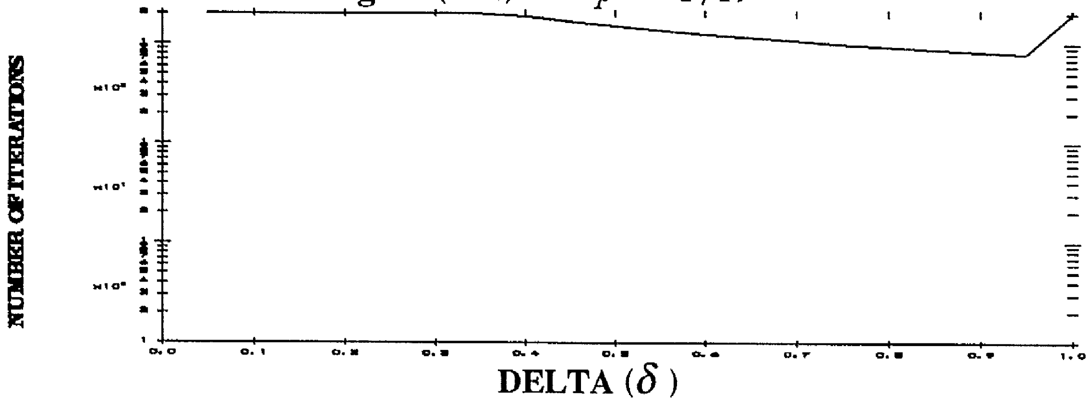


Figure (3.17) $p^{(o)} = 1/10$

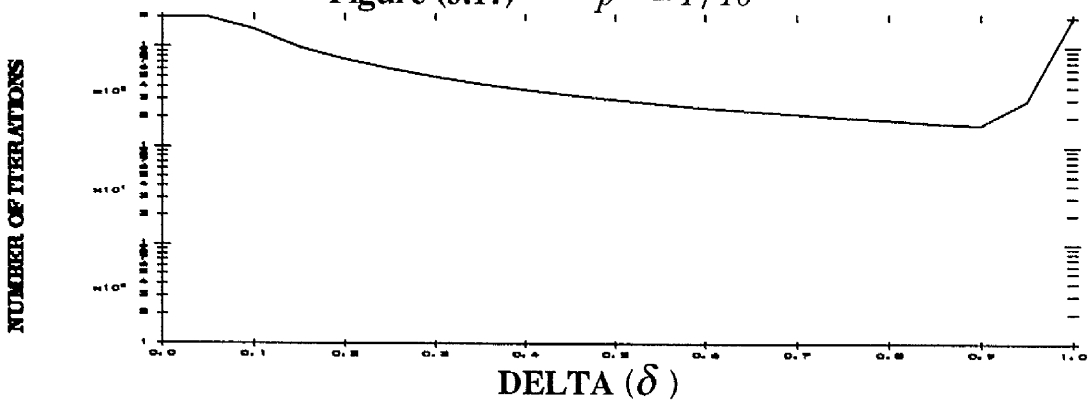
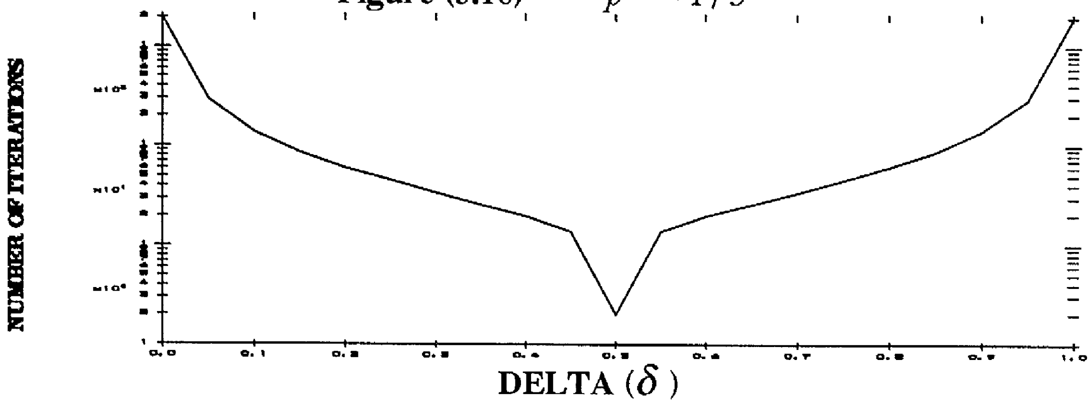


Figure (3.18) $p^{(o)} = 1/3$



The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve max $F_i \leq 10^{-4}$ against delta $(0 < \delta \leq 1)$ in case of the cubic regression model .

Figure (3.19) $p^{(o)} = 1/11$ $\underline{a} = (1,0,0,0)^t$ and $\underline{b} = (0,0,0,1)^t$

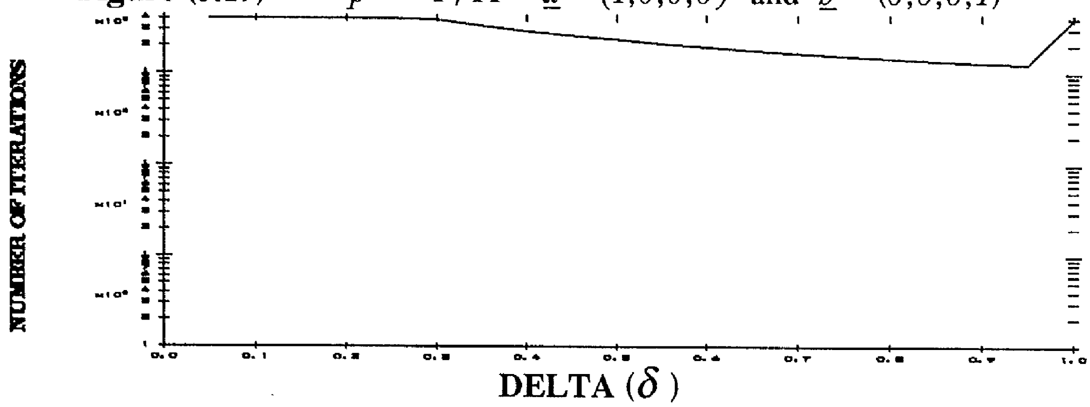


Figure (3.20) $p^{(o)} = 1/11$ $\underline{a} = (0,1,0,0)^t$ and $\underline{b} = (0,0,0,1)^t$

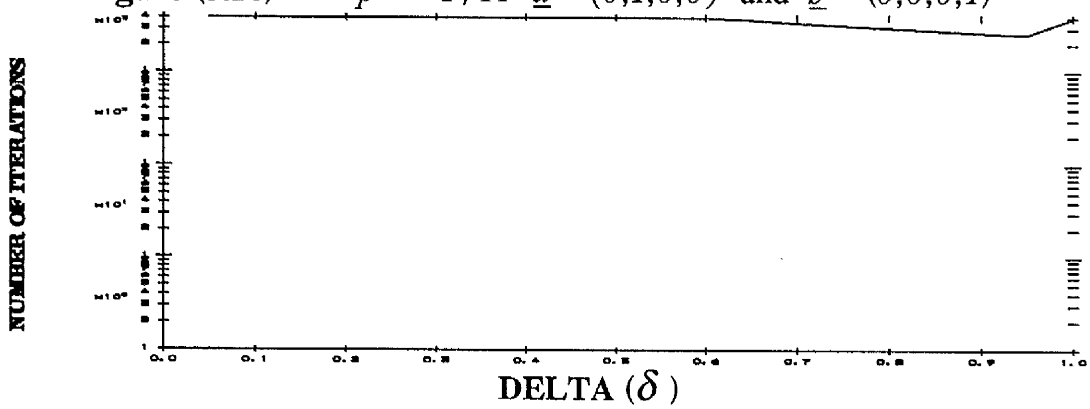
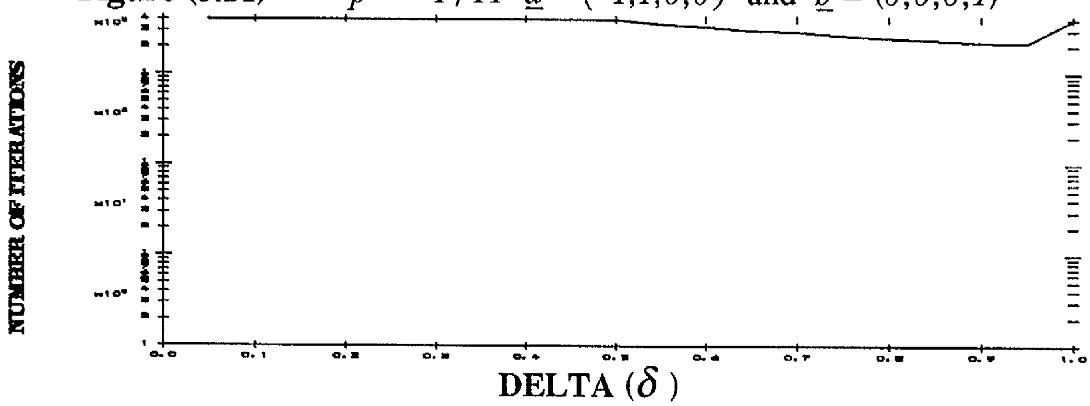


Figure (3.21) $p^{(o)} = 1/11$ $\underline{a} = (-1,1,0,0)^t$ and $\underline{b} = (0,0,0,1)^t$



The following figures illustrate a plot of the number of iterations (in log scale) needed to achieve max $F_i \le 10^{-4}$ against delta ($0 < \delta \le 1$) in case of the cubic regression model .

Figure (3.22) $p^{(o)} = 1/4$ $\underline{a} = (1,0,0,0)^t$ and $\underline{b} = (0,0,0,1)^t$

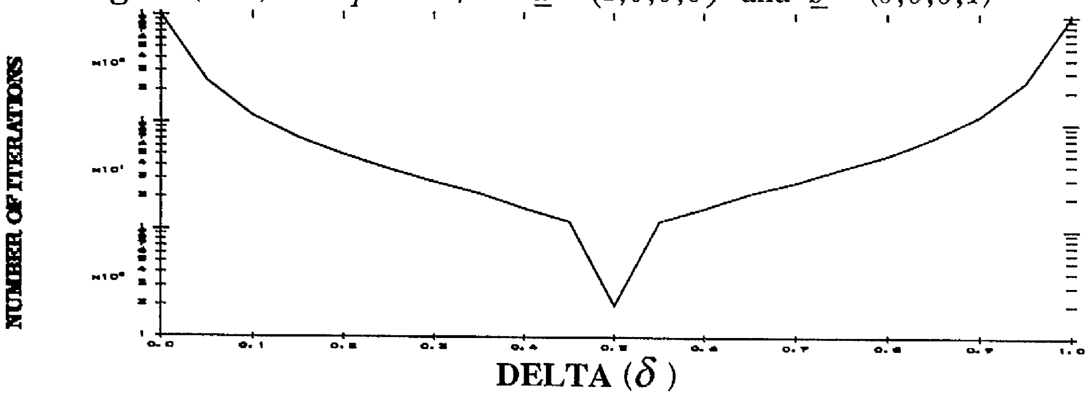


Figure (3.23) $p^{(o)} = 1/4$ $\underline{a} = (0,1,0,0)^t$ and $\underline{b} = (0,0,0,1)^t$

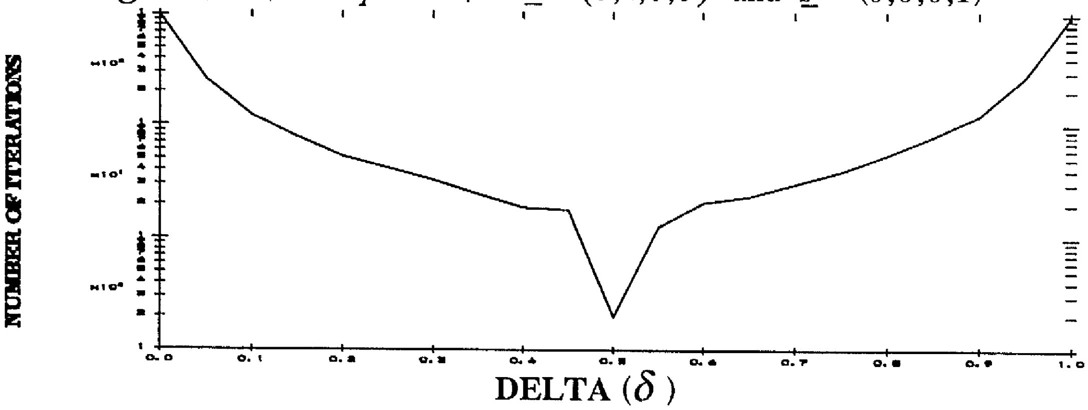
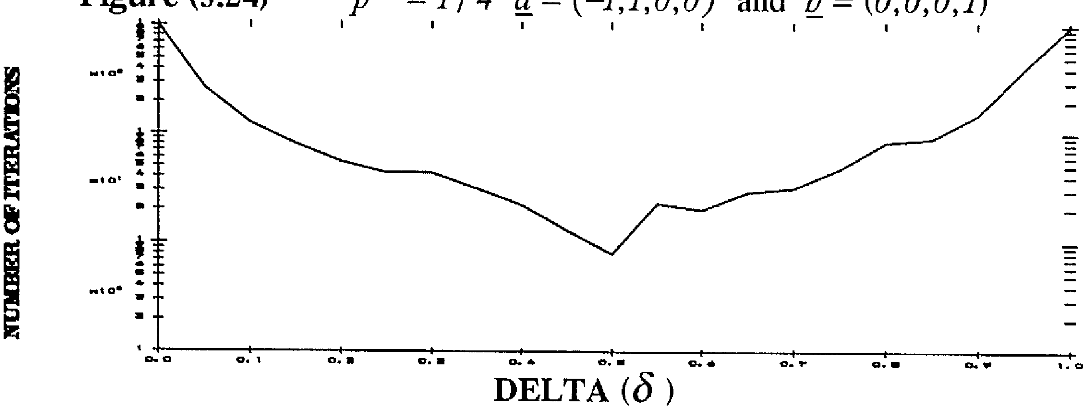


Figure (3.24) $p^{(o)} = 1/4$ $\underline{a} = (-1,1,0,0)^t$ and $\underline{b} = (0,0,0,1)^t$



CHAPTER FOUR

Maximise A Concave Criterion Subject To Zero Covariance

4.1 Introduction

4.2 Study Different Cases for Design Satisfying Zero Covariance

4.3 Algorithm

4.4 Examples and Discussions

4.5 Efficiencies of Constraint Optimal Design

CHAPTER FOUR

Maximize A Concave Criterion Subject To Zero Covariance

4.1 Introduction

As was seen in chapter (3) the covariance criterion depends on the signs of $c_i d_i \ \forall i \ , \ i = 1, 2, \dots, k$ where the covariance criterion is of the form:

$$\phi_c(p) = -\left(\underline{a}' M^{-1}(p) \underline{b}\right)^2 \tag{4.1.1}$$

$$= -\left(\sum_{i=1}^{k} \frac{c_i d_i}{p_i}\right)^2 \quad . \tag{4.1.2}$$

This cannot be zero if $c_i d_i > 0 \ \forall i$ or $c_i d_i < 0 \ \forall i \ , i = 1, 2, \dots, k$, in which case an explicit solution for the optimal weights p^* and the maximum value of the criterion can be found.

In this chapter we study the case in which zero covariance can be attained. So we concentrate on the maximization of a concave criterion (e.g. ϕ_5, ϕ_7 or ϕ_8) subject to zero covariance. We consider the case when the number of the design points equal the number of the unknown parameters.

4.2 Study Different Cases for Design Satisfying Zero Covariance

4.2.1 Case (1):

A simple case which attains zero covariance is when one of $c_i d_i$, say $c_1 d_1$, is greater than zero, one is less than zero, say $c_k d_k$, and the others are equal to zero. So equating (4.1.2) to zero implies:

$$\frac{c_1 d_1}{p_1} = \frac{c_k d_k}{p_k}$$

$$\Rightarrow \; p_k = \left(\frac{c_k d_k}{c_1 d_1}\right) \times p_1$$

$$\Rightarrow \; p_k = \alpha p_1 \quad \text{where } \alpha = {}^{c_k d_k}\!\big/_{c_1 d_1} \tag{4.2.1}$$

Then there are many designs guaranteeing zero covariance. So we try to find an optimal choice for $p_1, p_2, \ldots, p_k$ by maximizing another criterion (e.g. ϕ_5, ϕ_7 or ϕ_8) with the aim of a good estimation of $\underline{a}'\underline{\theta}$ and $\underline{b}'\underline{\theta}$.

One possibility is ϕ_7 with $A' = [\underline{a} \vdots \underline{b}]$. Our aim in this case is to maximize ϕ_7 subject to zero covariance i.e. when $p_k = \alpha\, p_1$ and $0 \le p_i \le 1$, $\sum_{i=1}^{k} p_i = 1$, $i = 1,2,\ldots,k$. But since ϕ_7 is of the form

$$\phi_7(p) = -Trace\left[AM^{-1}(p)A'\right] \quad , \quad A' = [a \;\vdots\; b]$$

$$= -Trace\left[A\left(VPV'\right)^{-1}A'\right] \quad , \quad \text{since } M(p) = VPV'$$

$$= -Trace\left[A\left(\left(V^{t}\right)^{-1}P^{-1}V^{-1}\right)A^{t}\right]$$

$$= -Trace\left[P^{-1}V^{-1}A^{t}A\left(V^{t}\right)^{-1}\right]$$

$$= -\sum_{i=1}^{k}\frac{\eta_i^2}{p_i} \qquad (4.2.2)$$

where η_i^2 is the diagonal elements of the matrix $V^{-1}A^{t}A\left(V^{t}\right)^{-1}$. Then by substituting from (4.2.1) in (4.2.2) we find:

$$\phi_7(p) = -\left(\sum_{i=1}^{k-1}\frac{\eta_i^2}{p_i} + \frac{\eta_k^2}{\alpha p_1}\right)$$

$$= -\left[\frac{\eta_1^2 + \left(\eta_k^2/\alpha\right)}{p_1} + \sum_{j=2}^{k-1}\frac{\eta_j^2}{p_j}\right]$$

But since $\sum_{i=1}^{k} p_i = 1$, $0 \le p_i \le 1$ then

$$\alpha p_1 + \sum_{i=1}^{k-1} p_i = 1 \quad \Rightarrow \quad (1+\alpha)p_1 + \sum_{i=2}^{k-1} p_i = 1$$

If we let $q_1 = (1+\alpha)p_1$ *and* $q_j = p_j$, $j = 2,\ldots,k-1$, then $\sum_{i=1}^{k-1} q_i = 1$ and

$$\phi_7(p) = -\left[\frac{(1+\alpha)\left(\eta_1^2 + [\eta_k^2/\alpha]\right)}{q_1} + \sum_{i=2}^{k-1}\frac{\eta_i^2}{q_i}\right]$$

$$= -\sum_{i=1}^{k-1} \frac{\eta_i^{*^2}}{q_i} \tag{4.2.3}$$

where $\eta_1^{*^2} = (1+\alpha)\left[\eta_1^2 + \left(\eta_k^2 / \alpha\right)\right]$, $\eta_j^{*^2} = \eta_j^2$, $j = 1, 2, \ldots, k-1$. Hence (4.2.3) leads to the explicit solution for the optimal weights :

$$q_j^* = |\eta_j^*| / \left(\sum_{i=1}^{k-1} |\ \eta_i^*\ | \right) \tag{4.2.4}$$

and the maximum value for $\phi_7(p)$ subject to zero covariance is:

$$\phi_7(q^*) = -\left\{ \sum_{i=1}^{k-1} |\eta_i^*| \right\}^2 \tag{4.2.5}$$

A second possibility might be to maximize ϕ_8 i.e.

maximize $\phi_8(p) = -\underline{c}M^{-1}(p)\underline{c}'$ with $\underline{c} = \underline{a} + \underline{b}$ subject to zero covariance and $\sum_{i=1}^{k} p_i = 1$, $0 \le p_i \le 1$. Following the same procedure as for the criterion ϕ_7 we again find that an explicit solution for the optimal weights exist and is of the form:

$$q_j^* = |\eta_j^*| / \left[\sum_{i=1}^{k-1} |\eta_i^*| \right] \quad , \tag{4.2.6}$$

and the maximum value for ϕ_8 is

$$\phi_8(q^*) = -\left\{ \sum_{i=1}^{k-1} |\eta_i^*| \right\}^2 , \tag{4.2.7}$$

where $\eta_1^{*2} = (1+\alpha)\left[\eta_1^2 + \left(\eta_k^2 / \alpha\right)\right]$, $\eta_j^{*2} = \eta_j^2$, $j = 1, 2, \ldots, k-1$ and $\eta_j^2 = \left(V^{-1}\underline{c}\right)^2, j = 1, 2, \ldots, k$.

4.2.2 Equivalence of ϕ_7 and ϕ_8:

In fact subject to zero covariance our two choices of ϕ_7 and ϕ_8 are equivalent. A direct proof of this derives from the fact that if $\text{cov}\left(\underline{a}'\hat{\underline{\theta}}, \underline{b}'\hat{\underline{\theta}}\right) = 0$ then $\left(V^{-1}\underline{c}\right) = diag\left(V^{-1}A'AV'\right)$ which leads to:

$$\phi_7(p) = -Trace\left[AM^{-1}(p)A'\right]$$

$$= \sum_{i=1}^{k} \frac{\eta_i^2}{p_i} = -\underline{c}'M^{-1}(p)\underline{c} = \phi_8(p) \qquad (4.2.8)$$

where $\underline{\eta}^2 = \left(V^{-1}\underline{c}\right)^2 = diag\left(V^{-1}A'AV'\right)$, and $\underline{c} = \underline{a} + \underline{b}$, $A' = \left[\underline{a} \vdots \underline{b}\right]$.

4.2.3 Case(2):

In this case we study the covariance criterion (4.1.1) when some of $c_i d_i$ are greater than zero and the others are less than zero. So without loss of generality suppose:

$c_i d_i > 0$ for $i = 1, 2, \ldots, n_1$; $c_i d_i < 0$ for $i = n_1 + 1, \ldots, k$ and let

$$\left.\begin{array}{l} e_i = c_i d_i \\ q_i = p_i \end{array}\right\} \text{ if } \ i = 1, 2, \ldots, n_1$$

$$, \left.\begin{array}{l} f_j = -c_{n_1+j} d_{n_1+j} \\ w_j = p_{n_1+j} \end{array}\right\} \text{ if } \ j = 1, 2, \ldots, n_2 \ , \ n_1 + n_2 = k$$

So we can rewrite (4.1.2) as follows:

$$\phi_c\left(\underline{q},\underline{w}\right)=-\left(\sum_{s=1}^{n_1}\frac{e_s}{q_s}-\sum_{t=1}^{n_2}\frac{f_t}{w_t}\right)^2 \qquad (4.2.9)$$

where $\quad q_i, w_j > 0 \;\; and \;\; \sum_{i=1}^{n1} q_i + \sum_{j=1}^{n_2} w_j = 1.$

Clearly (4.2.9) can be zero if

$$\sum_{i=1}^{n_1}\frac{e_i}{q_i}=\sum_{j=1}^{n_2}\frac{f_j}{w_j}. \qquad (4.2.10)$$

So there are many designs which will guarantee zero covariance i.e. $\phi_c\left(\underline{q},\underline{w}\right)=0$, and we try to choose $\underline{q}, \underline{w}$ optimally for good estimation of $(\underline{a}^t\underline{\theta}, \underline{b}^t\underline{\theta})$ by maximising another design criterion(e.g. ϕ_5, ϕ_7 or ϕ_8). The complete class of such designs is given by the following transformation:

$$\left(\underline{q},\underline{w}\right)\rightarrow\left(\underline{g},\underline{h}\right) \;:\; \begin{array}{ll} q_i=\dfrac{g_i\sum_{s=1}^{n_1}\frac{e_s}{g_s}}{z}, & i=1,2,\ldots,n_1 \\ w_j=\dfrac{h_j\sum_{t=1}^{n_2}\frac{f_t}{h_t}}{z}, & j=1,2,\ldots,n_2 \end{array} \qquad (4.2.11)$$

where $\quad z=\left[\left(\sum_{i=1}^{n_1}g_i\right)\times\left(\sum_{s=1}^{n_1}\frac{e_s}{g_s}\right)\right]+\left[\left(\sum_{j=1}^{n_2}h_j\right)\times\left(\sum_{t=1}^{n_2}\frac{f_t}{h_t}\right)\right] \quad , \quad g_i$ and $h_j > 0$.

4.2.4 Properties of the Transformation (4.2.11):

(1) That transformation satisfies zero covariance, is clear by substituting from (4.2.10) in (4.2.9) when we obtain:

$$\text{LHS} \quad \Rightarrow \quad \sum_{i=1}^{n_1} \frac{e_i}{q_i} = \left(\sum_{i=1}^{n_1} e_i \Big/ \left[\left(g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s} \right) / z \right] \right)$$

$$= z \left[\left(\sum_{i=1}^{n_1} \frac{e_i}{g_i} \right) \Big/ \left(\sum_{s=1}^{n_1} \frac{e_s}{g_s} \right) \right] = z \,. \qquad (4.2.12)$$

$$\text{Similarly the RHS} \quad \Rightarrow \quad \sum_{j=1}^{n_2} \frac{f_j}{w_j} = z \qquad (4.2.13)$$

So (4.2.12) and (4.2.13) show that $\varphi(\underline{q}, \underline{w}) = 0$.

(2) The transformation satisfies the constraint $\sum_{i=1}^{n_1} q_i + \sum_{j=1}^{n_2} w_j = 1$, because:

$$\sum_{i=1}^{n_1} q_i + \sum_{j=1}^{n_2} w_j = \sum_{i=1}^{n_1} \left[\left(g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s} \right) / z \right] + \sum_{j=1}^{n_2} \left[\left(h_j \sum_{t=1}^{n_2} \frac{f_t}{h_t} \right) / z \right] = 1$$

(3) If $\sum_{i=1}^{n_1} g_i = \sum_{j=1}^{n_2} h_j = 1$ the transformation is injective (one-one). To see this recall from (4.2.11):

$$q_i = \left(g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s} \right) / z \qquad \rightarrow g_i = (z \; q_i) \Big/ \left(\sum_{s=1}^{n_1} \frac{e_s}{g_s} \right)$$

But since $\sum_{i=1}^{n_1} g_i = 1$, then

$$g_i = \left\{ (z\ q_i) / \left(\sum_{s=1}^{n_1} \frac{e_s}{g_s} \right) \right\} / \left\{ \left(\sum_{j=1}^{n_1} z\ q_i \right) / \left(\sum_{s=1}^{n_1} \frac{e_s}{g_s} \right) \right\}$$

$$= \frac{z\ q_i}{\left(\sum_{s=1}^{n_1} \frac{e_s}{g_s} \right)} \times \frac{\sum_{s=1}^{n_1} \frac{e_s}{g_s}}{z \sum_{j=1}^{n_1} q_j} = \frac{q_i}{\sum_{j=1}^{n_1} q_j}, \tag{4.2.14}$$

and also similarly

$$h_j = \frac{w_j}{\sum_{i=1}^{n_2} w_i}. \tag{4.2.15}$$

So (4.2.14) and (4.2.15) show that this transformation is one-one.

(4) The criterion $\phi\left[\underline{q}(\underline{g}, \underline{h}), \underline{w}(\underline{g}, \underline{h})\right]$ under this transformation is homogeneous of degree zero, since each of q_i and w_j are homogeneous of degree zero in both $\underline{g}$ and $\underline{h}$ i.e.

$\phi\left[\underline{q}(\lambda \underline{g}, \lambda \underline{h}), \underline{w}(\lambda \underline{g}, \lambda \underline{h})\right] = \lambda^0 \phi\left[\underline{q}(\underline{g}, \underline{h}), \underline{w}(\underline{g}, \underline{h})\right] = \phi\left[\underline{q}(\underline{g}, \underline{h}), \underline{w}(\underline{g}, \underline{h})\right]$ for any $\lambda > 0$.

This is the reason for invoking the constraints $\sum_{i=1}^{n_1} g_i = \sum_{j=1}^{n_2} h_j = 1$.

4.2.5 Case(3):

A final case which attains zero covariance is when some of the $c_i\, d_i$ are greater than zero, some less than zero and the others are equal to zero. This case is similar to case(2) in this chapter but with extra weights r_m corresponding to zero $c_i\, d_i$. Without loss of generality suppose

$c_i\, d_i > 0$ for $i = 1,2,\ldots,n_1$; $c_{n_1+j}\, d_{n_1+j} < 0$ for $j = 1,2,\ldots,n_2$

and $c_{n_1+n_2+m}\, d_{n_1+n_2+m} = 0$ for $m = 1,2,\ldots,n_3$.

Then the covariance criterion in this case will be of the form

$$\phi_c\left[\underline{q},\underline{w},\underline{r}\right] = -\left(\sum_{i=1}^{n_1}\frac{e_i}{q_i} - \sum_{j=1}^{n_2}\frac{f_j}{w_j} + \sum_{m=1}^{n_3}\frac{o_m}{r_m}\right)^2 \qquad (4.2.16)$$

where $\left.\begin{matrix} e_i = c_i d_i \\ q_i = p_i \end{matrix}\right\}$ if $i = 1,2,\ldots,n_1$; $\left.\begin{matrix} f_j = -c_{n_1+j} d_{n_1+j} \\ w_j = p_{n_1+j} \end{matrix}\right\}$ if $j = 1,2,\ldots,n_2$

and $\left.\begin{matrix} o_m = 0 \\ r_m = p_{n_1+n_2+m} \end{matrix}\right\}$ if $m = 1,2,\ldots,n_3$, $\sum_{i=1}^{n_1} q_i + \sum_{j=1}^{n_2} w_j + \sum_{m=1}^{n_3} r_m = 1$, $n_1 + n_2 + n_3 = k$.

As in case(2) a complete class of designs which guarantee zero covariance is given by the following transformation:

$$(\underline{q},\underline{w},\underline{r}) \rightarrow (\underline{g},\underline{h},\underline{u}): \quad \begin{array}{ll} q_i = u_{n_3+1} \dfrac{g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s}}{z} & , \quad i = 1,2,\ldots,n_1 \\ w_j = u_{n_3+1} \dfrac{h_j \sum_{t=1}^{n_2} \frac{f_t}{h_t}}{z} & , \quad j = 1,2,\ldots,n_2 \\ r_m = u_m & , \quad m = 1,2,\ldots,n_3 \end{array} \qquad (4.2.17)$$

where $u_{n_3+1} = \sum_{i=1}^{n_1} q_i + \sum_{j=1}^{n_2} w_j$.

This transformation has similar properties to the transformation mentioned in case(2). In addition it is homogeneous of degree -1 for the criterion $\phi\left[\underline{q}(\underline{g},\underline{h},u_{n_3+1}),\underline{w}(\underline{g},\underline{h},u_{n_3+1}),\underline{r}(\underline{u})\right]$ with respect to u.

Now after using the transformations (4.2.11) and (4.2.17), the original problem changes to:

Either: maximize $\phi\left[\underline{q}(\underline{g},\underline{h}),\underline{w}(\underline{g},\underline{h})\right]$ subject to $\sum_{i=1}^{n_1} g_i = \sum_{j=n_1+1}^{n_2} h_j = 1$, or to

maximize $\phi\left[\underline{q}(\underline{g},\underline{h},u_{n_3+1}),\underline{w}(\underline{g},\underline{h},u_{n_3+1}),\underline{r}(\underline{u})\right]$ subject to $\sum_{i=1}^{n_1} g_i = \sum_{j=1}^{n_2} h_j = \sum_{m=1}^{n_3+1} u_m = 1$,

g_i, h_j and u_m are greater than zero for $i = 1,2,\ldots,n_1$, $j = 1,2,\ldots,n_2$ and $m = 1,2,\ldots,n_3+1$.

Where $\phi\left[\underline{q}(\underline{g},\underline{h}),\underline{w}(\underline{g},\underline{h})\right]$ and $\phi\left[\underline{q}(\underline{g},\underline{h},u_{n_3+1}),\underline{w}(\underline{g},\underline{h},u_{n_3+1}),\underline{r}(\underline{u})\right]$ are one of the criterions ϕ_5, ϕ_7 or ϕ_8 under these transformations.

Note that to distinguish between the optimality criteria ϕ_5, ϕ_7 or ϕ_8 and these criterion under the above transformations we will use ψ_5, ψ_7 or ψ_8 instead of ϕ_5, ϕ_7 or ϕ_8.

4.3 Algorithm:

In Chapter 3 we described and subsequently used algorithm (3.4.1) for maximising a criterion with respect to one set of weights $p_1, p_2, \ldots, p_J$. That algorithm was of the form

$$p_j^{(r+1)} \propto p_j^{(r)} f(d_j, \delta) \tag{4.3.1}$$

Where $f(d_j, \delta)$ satisfied certain properties. In this problem we have two sets of weights g_i, $i = 1,2,\ldots,n_1$ and h_j, $j = 1,\ldots,n_2$ and possibly a third set u_m, $m = 1,\ldots,n_3$. A natural extension of this algorithm is:

$$\left.\begin{array}{l} \left.\begin{array}{l} g_i^{(r+1)} = \dfrac{g_i^{(r)} f_1(d_{1i},\delta)}{\sum_{s=1}^{n_1} g_s^{(r)} f_1(d_{1s},\delta)}, i = 1,2,\ldots,n_1 \\ \\ h_j^{(r+1)} = \dfrac{h_j^{(r)} f_2(d_{2j},\delta)}{\sum_{t=1}^{n_2} h_t^{(r)} f_2(d_{2t},\delta)}, j = 1,2,\ldots,n_2 \end{array}\right\} , n_1 + n_2 = k \quad (4.3.2)(a) \\ \\ \text{and if necessary} \\ \\ u_m^{(r+1)} = \dfrac{u_m^{(r)} f_3(d_{3m},\delta)}{\sum_{v=1}^{n_3+1} u_v^{(r)} f_3(d_{3v},\delta)}, m = 1,2,\ldots,n_3+1 \end{array}\right\} (4.3.2)(b)$$

where $f_1(d_{1i}, \delta)$, $f_2(d_{2j}, \delta)$ and $f_3(d_{3m}, \delta)$ satisfy similar properties to the function $f(d_j, \delta)$, and $d_{1i} = \dfrac{\partial \psi}{\partial g_i}$, $d_{2j} = \dfrac{\partial \psi}{\partial h_j}$ and $d_{3m} = \dfrac{\partial \psi}{\partial u_m}$. These derivatives are of course given by:

$$\frac{\partial \psi}{\partial g_i} = \sum_{s=1}^{n_1}\left[\left(\frac{\partial \psi}{\partial q_s}\right)\left(\frac{\partial q_s}{\partial g_i}\right)\right] + \sum_{t=1}^{n_2}\left[\left(\frac{\partial \psi}{\partial w_t}\right)\left(\frac{\partial w_t}{\partial g_i}\right)\right]$$
$$\frac{\partial \psi}{\partial h_j} = \sum_{s=1}^{n_1}\left[\left(\frac{\partial \psi}{\partial q_s}\right)\left(\frac{\partial q_s}{\partial h_j}\right)\right] + \sum_{t=1}^{n_2}\left[\left(\frac{\partial \psi}{\partial w_t}\right)\left(\frac{\partial w_t}{\partial h_j}\right)\right] , \qquad (4.3.3)$$

where

$$\frac{\partial\psi}{\partial q_s} = \frac{\partial\phi}{\partial p_s}, s = 1,2,\ldots,n_1 ; \quad \frac{\partial\psi}{\partial w_t} = \frac{\partial\phi}{\partial p_{n_1+t}}, t = 1,2,\ldots,n_2;$$

$$\frac{\partial q_i}{\partial g_i} = \frac{z\left[g_i\left(-\frac{e_i}{g_i^2}\right) + \sum_{s=1}^{n_1}\frac{e_s}{g_s}\right] - \left(g_i\sum_{s=1}^{n_1}\frac{e_s}{g_s}\right)\frac{\partial z}{\partial g_i}}{z^2} \quad for\ \ i = 1,2,..,n_1$$
$$\frac{\partial q_i}{\partial g_j} = \frac{z\left[g_i\left(-\frac{e_j}{g_j^2}\right)\right] - \left(g_i\sum_{s=1}^{n_1}\frac{e_s}{g_s}\right)\frac{\partial z}{\partial g_j}}{z^2} \quad for\ \ j \neq i ,$$

and

$$\frac{\partial z}{\partial g_i} = \left(\sum_{s=1}^{n_1} g_s\right)\left(-\frac{e_i}{g_i^2}\right) + \sum_{s=1}^{n_1}\frac{e_s}{g_s},$$
$$\frac{\partial z}{\partial g_j} = \left(\sum_{s=1}^{n_1} g_s\right)\left(-\frac{e_j}{g_j^2}\right) + \sum_{s=1}^{n_1}\frac{e_s}{g_s}$$

and also

$$\frac{\partial w_i}{\partial h_j} = \frac{\left(-g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s}\right)\left(\frac{\partial z}{\partial h_j}\right)}{z^2}$$

where

$$\frac{\partial z}{\partial h_j} = \left(\sum_{t=1}^{n_2} h_t\right)\left(-\frac{f_j}{h_j^2}\right) + \sum_{t=1}^{n_2} \frac{f_t}{h_t}$$

and finally

$$\frac{\partial \psi}{\partial u_m} = \begin{cases} \dfrac{\partial \psi}{\partial r_m} & if \quad m = 1, 2, \ldots, n_3 \\ \sum_{s=1}^{n_1} \left(\dfrac{g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s}}{z} \right) \times \dfrac{\partial \psi}{\partial q_m} + \sum_{t=1}^{n_2} \left(\dfrac{h_i \sum_{s=1}^{n_2} \frac{e_t}{h_t}}{z} \right) \times \dfrac{\partial \psi}{\partial w_m} & if \quad m = n_3 + 1. \end{cases}$$

The partial derivatives $\frac{\partial w_i}{\partial h_i}$, $\frac{\partial w_i}{\partial h_j}$ *and* $\frac{\partial w_i}{\partial g_l}; l = 1, 2, \ldots, n_1;\ i, j = 1, 2, \ldots, n_2$ are similarly defined.

4.4 Examples and Discussion:

This section will be devoted to a set of examples and we derive the solution to some optimal design problems which can be calculated either explicitly for case(1) or by using algorithm (4.3.1) for case(2) and case(3).

Example (1)

An example of case(1) is EX(1) considered by Silvey et al (1978) and Wu (1978). The design space for this example is:

$$V = \left\{ (1,-1,-1)', (1,-1,1)', (1,1,-1)', (1,2,2)' \right\}$$

We consider $\underline{a} = (1, 0, 0)'$, $\underline{b} = (0, 0, 1)'$. Then a subset of the design space which satisfies case(1) is $V_1 = \left\{ (1,-1,1)', (1,1,-1)', (1,2,2)' \right\}$.

Values which emerge are:

$$\left.\begin{aligned} \underline{\underline{cd}} &= (0.0625, 0.0, -0.1875)' \\ \alpha &= \frac{0.1875}{0.0625} = 3 \\ \underline{\eta}^2 &= (0.2656, 0.3906, 0.0625)' \end{aligned}\right\} \qquad (4.4.1)$$

Then substituting from (4.4.1) into (4.2.4) and (4.2.5) we obtain a maximum value for ψ_7 of 2.2750 and for the optimal weights:

$$\underline{q}^* = (0.83425, 0.16575)', \underline{p}^* = (0.2086, 0.6257, 0.1657)'.$$

Note that we get the same result for the criterion ψ_8 in this example since ψ_7 and ψ_8 are equivalent (see section (4.2.2)).

Example (2)

This is an example of case(2) for the criterion ψ_7. We consider the quadratic regression model:

$$E(y) = \theta_0 + \theta_1 x + \theta_2 x^2 \ , \ x \in [-\beta, \beta] \ , \beta = 1, 2, 3, 4, 5$$

and we choose $A' = [\underline{a} \ , \ \underline{b}]$ where $\underline{a} = (1 \ , \ 0 \ , \ 0)'$ and $\underline{b} = (0 \ , \ 0 \ , \ 1)'$.

We determined ψ_7-optimal design on $(-\beta \ , \ \beta \ )$ from among those designs with three support points, two of them being the endpoints.

Weights and support points are recorded in TABLE (4.1). These designs are not unique as can be seen from FIGURE (4.1) and (4.2), which are a plot against x of the ψ_7-criterion value of the optimal design on $\{-\beta, x \ , \ \beta\}$. Alternative designs are got by exchanging the weights on $\pm\beta$ and changing the sign of the third support point, as the following confirms:

From (1.2.8) the design matrix is of the form

$$M_v(p) = V \ P \ V' \text{, where } \quad V = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ x_1^2 & x_2^2 & x_3^2 \end{bmatrix}$$

Changing the sign of x-values is equivalent to using the design matrix

$$M_w(p) = W \ P \ W' \ ; \ W = DV \ , \ D = diag(1, -1, 1)$$

Then $M_w(p) = (D \ V \ P \ V' \ D)^{-1} = D^{-1} M_v^{-1}(p) \ D^{-1}$. So in the case of criterion ϕ_7

$$Trace \ (A' \ M_w^{-1}(p) \ A \ ) = Trace \ (A' \ D^{-1} \ M_v^{-1}(p) \ D^{-1} \ A \ ) == Trace(A' \ M_v^{-1}(p) \ A \ ).$$

since $\quad A' \ D^{-1} = A', \ D^{-1} \ A = A$ when $a_{12} = a_{22} = 0,$ where $A = \{a_{ij}\}$

$;i = 1,2$ and $j = 1,2,3.$

Corresponding results for the ψ_5-criterion are reported in TABLE(4.2).

Clearly from TABLE(4.1) and TABLE(4.2) these designs have the unusual feature of one large weight and two small weights. In part this is due to constraints imposed on them by the zero-covariance requirement e.g. in the case $\beta = 1$ limits on these weights are: $0.001 \leq p_1 \leq 0.0199$, $0.0028 \leq p_2 \leq 0.7135$ and $0.2666 \leq p_3 \leq 0.9997$.

Also from these tables, when β increases the value for the ψ-optimality criterion (ψ_5, ψ_7 and ψ_8) is increased (see Figures (4.3) and (4.4)). A proof for this will be given later on in this section in the case of criterion ψ_5. So the optimal design whose support includes $\pm\beta$ is optimal among all possible symmetric designs on the interval $[-\beta, \beta]$. Moreover, these designs seems globally optimal since they have a higher criterion value than the best three point designs on the interval $[-\alpha, \gamma]$ subject to the endpoints being support points, where $[-\alpha, \gamma] \subseteq [-\beta, \beta]$. See TABLE (4.4) and TABLE(4.5) .

There is no particular relationship between the middle support point and the endpoints in the case of criteria ψ_7 but in the case of ψ_5 if x is the middle support point on $[-1, 1]$ then βx is the middle support point on $[-\beta, \beta]$. To see this recall from (1.2.8) that:

$$M_v(p) = V P V' = DV_1 P V_1' D = D M_{v_1}(p) D \quad ,$$

where $V = \begin{bmatrix} 1 & 1 & 1 \\ -\beta & x_1 & \beta \\ \beta^2 & x_1^2 & \beta^2 \end{bmatrix}$, $V_1 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & (x_1/\beta) & 1 \\ 1 & (x_1^2/\beta^2) & 1 \end{bmatrix}$,

$D = diag(1, \beta, \beta^2)$ and $P = diag(p_1, p_2, p_3)$.

Then $M_v^{-1}(p) = [D\, M_{v_1}(p)\, D]^{-1} = D^{-1}\, M_{v_1}^{-1}(p)\, D^{-1}.$

So $Det[A'\, M_v^{-1}(p)\, A] = Det[A'\, D^{-1}\, M_{v_1}^{-1}(p)\, D^{-1}\, A]$

$= Det[E\, A' M_{v_1}^{-1}(p)\, A\, E]$, where $E\, A' = A'\, D^{-1}$, $E = Diag\left[1, \left(1/\beta^2\right)\right]$

$$= Det(E^2) \times Det[A'\, M_{v_1}^{-1}(p)\, A] = \left(\frac{1}{\beta^4}\right) \times Det[A'\, M_{v_1}^{-1}(p)\, A]. \qquad (4.4.2)$$

Clearly there is an equivalence between ψ_5-optimal designs on $\{-\beta, x_1, \beta\}$ and $\{-1, (x_1/\beta), 1\}$ in that the points in these sets have respectively the same optimal weights.

Moreover, $x^* = \dfrac{x_1^*}{\beta} \Rightarrow x_1^* = x^*\beta$ where x^*, x_1^* are the middle support points on $[-1, 1]$ and $[-\beta, \beta]$ respectively.

We note finally that the optimal values of $\underline{g}$ *and* $\underline{h}$ were determined by algorithm (4.3.2)(a) with

$$f_1(d_{1i}, \delta) = \begin{cases} (1+d_{1i})^{\delta} & if \;\; d_{1i} \geq 0 \\ (1-d_{1i})^{-\delta} & if \;\; d_{1i} < 0 \end{cases} \quad \text{and} \quad f_2(d_{2j}, \delta) = \begin{cases} (1+d_{2j})^{\delta} & if \;\; d_{2j} \geq 0 \\ (1-d_{2j})^{-\delta} & if \;\; d_{2j} < 0 \end{cases}$$

and initial starting weights $g_i^{(0)} = \dfrac{1}{n_1}$, $h_j^{(0)} = \dfrac{1}{n_2}$ where $n_1 + n_2 = k$; for $i = 1, 2, \ldots, n_1$ and $j = n_1 + 1, \ldots, k$.

During the iterative process the value of δ was kept fixed. The program was homemade and written in FORTRAN. The program has the possibility of choosing the criteria ψ_5, ψ_7 or ψ_8.

The number of iterations needed to achieve $max\ F_j \leq 10^{-n}; n = 1, 2, 3, 4, 5$ and $j = 1,2,3$ are recorded in Table (4.3) when finding the optimal weights for the optimal support points in case of the criterion ψ_7 and ψ_8 for those values of δ which attain max $F_j \leq 10^{-5}$ in the smallest number of iterations. The number of iterations in the case of the criterion ψ_5 are 4, 5 , 6, 6, 8 (when δ=.32) which are the same for various β. See Figures (4.5),(4.6) and (4.11) for different values of δ.

Example (3):

This is another example of case (2) for the criteria ψ_5, ψ_7 and ψ_8. We consider the cubic regression model:

$$E(y) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \ ; \ x \in [-\beta, \beta], \ \beta = 1, 2, 3, 4 \text{ and } 5.$$

and we choose $A = [\underline{a} : \underline{b}]$ where $\underline{a} = (1, 0, 0, 0)'$ and $\underline{b} = (0, 0, 0, 1)'$.

As in example (2) by using the same algorithm (4.3.2)(a), we determined the ψ_7-optimal design on the intervals $[-\beta, \beta]$ from among those designs with four support points two of them being the end points.

Clearly from TABLE (4.7) the ψ_7-optimal design has symmetric support points which puts small weights to the end points and large weights to the middle points. Explanation for the unusual feature of these optimal weights were given in example(2). The optimal design with symmetric support points seems to be global since at the optimal the partial derivatives for the ψ_7 -optimal criterion are equal i.e.

$$\frac{\partial \phi_7}{\partial p_j} = \sum_{i=1}^{4} p_i \frac{\partial \phi_7}{\partial p_i} \quad ; j = 1, 2, 3, 4 \ .$$

The number of iterations needed to achieve $max\ F_j \le 10^{-n}; n = 1, 2, 3, 4, 5$ and $j = 1,2,3,4$ are recorded in Table (4.9) when finding the optimal weights of these particular optimal support points for those values of δ which attain max $F_j \le 10^{-5}$ in the smallest number of iterations. Figures (4.7) and (4.8) show the number of iterations for different values of δ.

In the case of ψ_5-optimality criterion we determined the optimal value for the criterion ψ_5 of -33.97197 when $\beta = 1$ with support points -1 , -0.35, 0.35, 1 and optimal weights 0.09806, 0.40194, 0.40194 and 0.09806. So from these result we can get the optimal support points on the intervals $[-\beta, \beta]$ for β =2,3,4,5 i.e. $-\beta$, $-0.35 \times \beta$, $0.35 \times \beta$, β and optimal value of $(-33.97197/\beta^6)$ The proof for this is similar to that of example (2). The number of iterations needed to achieve $max\ F_j \le 10^{-n}; n = 1, 2, 3, 4, 5$ and $j = 1,2,3,4$ are 3, 4, 4, 6, 6 (when δ =.63). See Figure (4.12) for different value of δ.

We note that these designs are symmetric. In fact this might have been anticipated. To see this consider the symmetric four support points $\{-\beta,\ -x,\ x,\ \beta\}$. Then the design matrix will be of the form:

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -\beta & -x & x & \beta \\ \beta^2 & x^2 & x^2 & \beta^2 \\ -\beta^3 & -x^3 & x^3 & \beta^3 \end{bmatrix} \quad , \; 0 < x < \beta \;,\; \beta = 1,2,3,4,5 \;.$$

Hence

$$\underline{c} = V^{-1}\underline{a} = \left(\frac{1}{Det(V)}\right)\left[-2x^3\beta\left(\beta^2 - x^2\right), 2x\beta^3\left(\beta^2 - x^2\right), 2x\beta^3\left(\beta^2 - x^2\right), -2x^3\beta\left(\beta^2 - x^2\right)\right]$$

and

$$\underline{d} = V^{-1}\underline{b} = \left(\frac{1}{Det(V)}\right)\left[-2x\left(\beta^2 - x^2\right), 2\beta\left(\beta^2 - x^2\right), -2\beta\left(\beta^2 - x^2\right), 2x\left(\beta^2 - x^2\right)\right]$$

and then

$$\begin{aligned} \underline{cd} = \left(\frac{1}{Det(v)^2}\right)&[4x^4\beta\left(\beta^2 - x^2\right), 4x\beta^4\left(\beta^2 - x^2\right), \\ &-4x\beta^4\left(\beta^2 - x^2\right), -4x^4\beta\left(\beta^2 - x^2\right)] \end{aligned} \qquad (4.4.3)$$

Clearly from (4.4.3) the number of positive $c_i d_i$ equals the number of the negative $c_i d_i$ i.e. $n_1 = n_2 = 2$. Moreover, $e_1 = f_2$, $e_2 = f_1$. For simplicity suppose we reassign labels such that $e_i = f_i$ for $i = 1,2$. It therefore follows that a symmetric design will gurantee zero covariance i.e. $q_1 = w_1$, $q_2 = w_2$. Hence the transformation to $\underline{g}$ and $\underline{h}$ is strictly unnecessary. Our optimisation problem reduces to one with the constraint $q_1 + q_2 = 1/2$ or to $g_1 + g_2 = 1$, $q_i = g_i/2$. In fact if we do transform to $\underline{g}$ and $\underline{h}$ but then argue that $g_i = h_i$ the transformation (4.2.11) reduces to $q_i = g_i/2$, $w_j = h_j/2$. It must follow that if

$\phi(\underline{q},\underline{w}) = \phi(\underline{q})$ is concave then so is $\psi\left[\underline{q}(\underline{g}),\underline{w}(\underline{h})\right] = \psi\left[\underline{q}(\underline{g})\right]$. We can be assured that the designs listed in Table(4.7) are optimal for their supports. Of course there may be asymmetric optimal designs. To discover these we would want to transform to $\underline{g}$ and $\underline{h}$.

Example (4):

In this example we consider again the cubic regression model as in example (3) in this chapter but with $\underline{a} = (1,0,0,0)'$ and $\underline{b} = (0,0,1,0)'$ for the criteria ψ_5 , ψ_7 and ψ_8 .

In Table (4.6) we report the optimal four point designs for the criteria ψ_7 and ψ_8 on $[-\beta,\beta]$ subject to the endpoints being support points for $\beta = 1,2,..,5$. In Table (4.8) we report the number of iterations needed to achieve max $F_i \le 10^{-n}$ for $n = 1,2,3,4,5$ for those values of δ which attain max $F_j \le 10^{-5}$ in the smallest numder of iterations, by algorithm (4.3.2)(a), when finding the optimal weights of these particular support points. Figures (4.9) and (4.10) show the number of iterations for different values of δ.

The two middle points were found by a search through the supports $\{-\beta,\ x_1,\ x_2,\ \beta\}$ where $x_1 < x_2$ and $x_1,\ x_2 < 0$ or $x_1,\ x_2 > 0$ (Note that we exclude the case when x_1 and x_2 have opposite signs because the $c_i d_i$ are either all negative or all positive. In this case no design satisfies zero covariance), algorithms (4.3.2)(a) or (4.3.2)(b) being used for each pair x_1 and x_2.

Hidden in these calculations are examples of case(3) which arise when $x_1 + x_2 = \pm\beta$. To see this consider the four point support $\{-1,\ \alpha,\ (1-\alpha),\ 1\}$. The design matrix will be of the form

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & \alpha & (1-\alpha) & 1 \\ 1 & \alpha^2 & (1-\alpha)^2 & 1 \\ -1 & \alpha^3 & (1-\alpha)^3 & 1 \end{bmatrix}, \qquad 0<\alpha<1.$$

Then if $\underline{a} = (1,0,0,0)'$ and $\underline{b} = (0,0,1,0)'$ we get

$$\underline{d} = V^{-1}\underline{a} = \left(\frac{1}{Det(V)}\right)[\alpha^2(\alpha-1)^2(2\alpha-1), -2\alpha(\alpha^2-1), 2\alpha(1-\alpha)(\alpha-2), \\ -\alpha(1-\alpha)(2\alpha^3-3\alpha^2-3\alpha+2)]'$$

and

$$\underline{c} = V^{-1}\underline{b} = \left(\frac{1}{Det(V)}\right)\left[2\alpha(1-\alpha)(2\alpha-1), 2\alpha(1-\alpha^2), 2\alpha(2-\alpha)(1-\alpha), 0\right]'$$

and then

$$\underline{cd} = \left(\frac{1}{Det(V)^2}\right)\left[-2\alpha^3(\alpha-1)^3(2\alpha-1)^2, -4\alpha^2(1-\alpha^2)^2, -4\alpha(1-\alpha)^2(\alpha-2)^2, 0\right]'$$

So in this case we have one positive $c_i d_i$, two negatives and one zero thus satisfying case(3).

4.5 Efficiencies Of Constrained Optimal Designs:

In the previous examples we calculated the optimal design on intervals $[-\beta,\beta]$ for the criteria ψ_5,ψ_7 and ψ_8 subject to zero covariance under three choices of $\underline{a}$ and $\underline{b}$, namely, $\underline{a}=(1,0,0)^t, \underline{b}=(0,0,1)^t$ for the quadratic regression model and $\underline{a}=(1,0,0,0)^t$, $\underline{b}=(0,0,0,1)^t$; $\underline{a}=(1,0,0,0)^t$, $\underline{b}=(0,0,1,0)^t$ for the cubic regression model.

We wish to calculate efficiencies of these designs. For this purpose we calculated the unrestricted optimal design on the intervals $[-\beta,\beta]$ for the criteria ϕ_5,ϕ_7 and ϕ_8 under the above choices of $\underline{a}$ and $\underline{b}$ by using the algorithm (2.3.1) when $f(\delta,d)=d^\delta$ taking δ to be $\delta= 1, (1/2), (1/2)$ respectively, thereby guaranteeing monotonicity; see Torsney (1983). Results are recorded in Tables (4.10) -(4.15).

We now define efficiencies. Suppose ϕ_i^* denotes the optimal value for the criterion ϕ_i, where ϕ_i is one of the criteria D-optimality, A-optimality or c-optimality for $i = 5, 7, 8$ respectively. And suppose ψ_i^* denotes the optimal value for these criteria under the zero covariance constraint.

The efficiencies for ϕ_i^* relative to ψ_i^* are defined as follows:

(a) In the case of the D-optimality criterion :

$$Eff(\phi_5^*,\ \psi_5^*) = \left(\frac{\phi_5^*}{\psi_5^*}\right)^{1/s} \tag{4.5.1}$$

where s is the number of linear combinations of the parameters of interest. The scaling with the *sth* root makes the criterion homogeneous of degree -1. See Atkinson and Donev (1992) and Pukelsheim and Rosenberger (1993).

(b) In the case of the A-optimality and c-optimality criteria:

$$Eff(\phi_i^*,\ \psi_i^*) = \frac{\phi_i^*}{\psi_i^*} \quad \text{for} \quad i = 7, 8. \tag{4.5.2}$$

In Tables (4.16) to (4.18) we list the efficiencies for ϕ_i^* relative to ψ_i^* on the intervals $[-\beta, \beta]$ for $\beta = 1, 2, 3, 4$ in the case of the quadratic and cubic regression models listed above. Clearly in the case of D-optimality efficiencies are equal for any β under each choice of $\underline{a}$ and $\underline{b}$. This is because the criterion is invariant under these particular choices of $\underline{a}$ and $\underline{b}$ given the models considered. However in the case of the other two criteria the efficiencies are increased when β increases. Moreover, under the choice of $\underline{a} = (1,0,0,0)^t$ and $\underline{b} = (0,0,0,1)^t$ for the cubic regression model efficiencies are unity for the case of the A and D-optimality criteria. For these criteria the designs which maximize ϕ_i for $i = 5, 7$ are identical with the designs which maximize ψ_i subject to zero covariance. In the case of c-optimality for the same model and same choice of $\underline{a}$ and $\underline{b}$ the design achieves efficiencies ranging from $.85$ to $.95$ for $\beta = 1, 2, 3, 4$.

In general the constrained optimal designs only have good efficiencies for large β in the case of the linear criteria.

TABLE (4.1)

This table shows the optimal support points in the intervals $(-\beta, \beta)$ for the criteria ψ_7 *or* ψ_8 where β takes the values 1, 2, 3, 4 and 5. The value of δ was kept fixed during the iterative process with value equal 0.05.

	β =1	β =2	β =3	β =4	β =5
The No. of points in the design space	199	399	599	799	999
The maximum value for ψ_7 *or* ψ_8	-13.5004	-2.4268	-1.52333	-1.2743	-1.1699
Weights $\underline{g}^*$ and $\underline{h}^*$	1 0.7280 0.2719	1 0.8521 0.1478	1 0.9192 0.0807	1 0.9518 0.0481	1 0.9682 0.0317
Weights $\underline{q}^*$ and $\underline{w}^*$	0.0199 0.7135 0.2665	.0166 0.8379 0.1454	0.0112 0.9089 0.0798	0.0071 0.9450 0.0477	0.0049 0.9634 0.0316
Corresponding design points	-1 0.26 1	-2 0.26 2	-3 0.21 3	- 4 0.16 4	-5 0.15 5

TABLE (4.2)

This table shows the optimal support points in the intervals $(-\beta\,,\,\beta\,)$ for the criterion ψ_5 where β takes the values 1 ,2 ,3 ,4 and 5. The value of δ was kept fixed during the iterative process with value equal 0.05.

	β =1	β =2	β =3	β =4	β =5
The No. of points in the design space	199	399	599	799	999
The maximum value for ψ_5	-21.8661	-1.3666	-0.2699	-0.08546	-0.03498
Weights $\underline{g}^*$ and $\underline{h}^*$	1 0.82003 0.17996	1 0.82003 0.17996	1 0.82003 0.17996	1 0.82003 0.17996	1 0.82003 0.17996
Weights $\underline{q}^*$ and $\underline{w}^*$	0.0182 0.8051 0.1767	0.0182 0.8051 0.1767	0.0182 0.8051 0.1767	0.0182 0.8051 0.1767	0.0182 0.8051 0.1767
Corresponding design points	-1 0.16 1	-2 0.32 2	-3 0.48 3	-4 0.64 4	-5 0.80 5

TABLE (4.3)

This table shows the number of iterations needed to achieve $max\ F_j \leq 10^{-n}$; $n = 1, 2, 3, 4, 5$ and $j = 1, 2, 3$ in the intervals $(-\beta, \beta)$ for the criterion ψ_7 *or* ψ_8 where β takes the values 1,2,3,4 and 5.

δ	β \n	n=1	n=2	n=3	n=4	n= 5
δ=0.046	β =1	8	8	9	11	13
δ=0.24	β =2	5	6	6	8	10
δ= 0.37	β =3	5	7	7	7	9
δ= 0.56	β =4	4	6	6	6	8
δ= 0.63	β =5	4	6	6	8	10

TABLE(4.4)

This table shows the best three points design on $[-\alpha, \gamma]$ subject to the endpoints being support points in the case of ψ_7 -optimality criterion.

$[-\alpha, \gamma]$	The Maximum Value For ψ_7	Support Points
[-1.9 , 1.1]	-3.06868	-1.9 , -0.35 , 1.1
[-1.8 , 1.2]	-3.27932	-1.8 , -0.33 , 1.2
[-1.7 , 1.3]	-3.53473	-1.7 , -0.31 , 1.3
[-1.6 , 1.4]	-3.84751	-1.6 , -0.30 , 1.4
[-1.5 , 1.5]	-4.23605	-1.5 , -0.28 , 1.5
[-2 , 2]	**-2.42680**	**-2 , -0.26 , 2**
[-2 , 1.9]	-2.45866	-2 , -0.27 , 1.9
[-2 , 1.7]	-2.52945	-2 , -0.28 , 1.8
[-2 , 1.5]	-2.61108	-2 , -0.30 , 1.5
[-2 , 1.3]	-2.70754	-2 , -0.33 , 1.3
[-2 , 1.1]	-2.82455	-2 , -0.36 , 1.1
[-2 , 0.9]	-2.97193	-2 , -0.39 , 0.9
[-2 , 0.7]	-3.16889	-2 , -0.45 , 0.7
[-2 , 0.5]	-3.46195	-2 , -0.53 , 0.5
[-2 , 0.3]	-4.00542	-2 , -0.66 , 0.3

TABLE(4.5)

This table shows the best three points design on $[-\alpha,\gamma]$ subject to the endpoints being support points in the case of ψ_7 -optimality criterion.

$[-\alpha,\gamma]$	The Maximum Value For ψ_7	Support Points
[-0.5 , 0.01]	-2130.96	-0.5 , -0.31 , 0.01
[-0.7 , 0.01]	-703.750	-0.7 , -0.45 , 0.01
[-0.9 , 0.2]	-46.5821	-0.9 , -0.34 , 0.2
[-1.1 , 0.4]	-17.0152	-1.1 , -0.36 ,0.4
[-1.3 , 0.6]	-8.69952	-1.3 , -0.37 ,0.6
[-1.5 , 0.8]	-5.42187	-1.5 , -0.37 ,0.8
[-1.7 , 1.0]	-3.85344	-1.7 , -0.35 , 1.0
[-1.9 , 1.1]	-3.06868	-1.9 , -0.35 ,1.1
[-2.1 , 1.3]	-2.52181	-2.1 , 0.33 , 1.3
[-2.3 , 1.5]	-2.17053	-2.3 , -0.31 , 1.5
[-2.5 , 1.7]	-1.93106	-2.5 , -0.29 , 1.7
[-2.7 , 1.9]	-1.76012	-2.7 , -0.27 , 1.9
[-2.9 , 2.0]	-1.64381	-2.9 , -0.26 , 2.0
[-3.1 , 2.2]	-1.54465	-3.1 , -0.25 , 2.2
[-3.3 , 2.4]	-1.46743	-3.3 , -0.23 , 2.4

TABLE(4.6)

This table shows the optimal support points when $\underline{a} = (1,0,0,0)'$ and $\underline{b} = (0,0,1,0)'$ on the intervals $(-\beta\ ,\ \beta\ )$ for the criterion ψ_7 *or* ψ_8 where β takes the values 1 ,2 ,3 ,4 and 5. The value of δ was kept fixed during the iterative process with value equal 0.1 when β =1,2 and $\delta = .5$ when β =3,4,5 .

	β =1	β =2	β =3	β =4	β =5
The maximum value for ψ_7 *or* ψ_8	-9.45062	-2.29000	-1.53441	-1.29294	-1.17614
Weights $\underline{g}^*$ and $\underline{h}^*$	0.16440 0.14852 0.85148 0.83560	0.21991 0.07006 0.92994 0.78009	0.24656 0.04216 0.95784 0.75344	0.25422 0.02631 0.97369 0.74578	0.25034 0.01529 0.98471 0.74966
Weights $\underline{q}^*$ and $\underline{w}^*$	0.03166 0.11992 0.68751 0.16091	0.02300 0.06273 0.83269 0.08158	0.01422 0.03973 0.90261 0.04344	0.00900 0.02538 0.93923 0.02639	0.00600 0.01492 0.96110 0.01798
Corresponding design points	-1 -.97 -0.01 1	-2 -1.93 -0.01 2	-3 -2.74 -0.02 3	- 4 -3.65 -0.02 4	-5 -4.64 -0.01 5

TABLE (4.7)

This table shows the optimal support points when $\underline{a} = (1, 0, 0, 0)'$ and $\underline{b} = (0, 0, 0, 1)'$ on the intervals $(-\beta\ ,\ \beta\ )$ for the criterion $\psi_7\ or\ \psi_8$ where β takes the values 1 ,2 ,3 ,4 and 5.

	β =1	β =2	β =3	β =4	β =5
The maximum value for $\psi_7\ or\ \psi_8$	-18.7627	-1.8000	-1.20506	-1.08341	-1.04215
Weights $\underline{g}^*$ and $\underline{h}^*$	0.29932 0.70069 0.70069 0.29932	0.11112 0.88888 0.88888 0.11112	0.03940 0.96060 0.96060 0.03940	0.01685 0.98315 0.98315 0.01685	0.00910 0.99090 0.99090 0.00910
Weights $\underline{q}^*$ and $\underline{w}^*$	0.14965 0.35035 0.35035 0.14965	0.05556 0.44444 0.44444 0.05556	0.01970 0.48030 0.48030 0.01970	0.00842 0.49158 0.49158 0.00842	0.00455 0.49545 0.49545 0.00455
Corresponding design points	-1 -0.46 0.46 1	-2 -0.5 0.5 2	-3 -0.43 0.43 3	- 4 -0.35 0.35 4	-5 -0.34 0.34 5

TABLE (4.8) $\underline{a} = (1, 0, 0, 0)', \underline{b} = (0, 0, 1, 0)'$

δ	β \n	n=1	n=2	n=3	n=4	n= 5
δ =.072	β =1	3	4	4	6	6
δ =0.33	β =2	5	6	8	11	13
δ =0.6	β =3	4	8	12	18	24
δ =0.7	β =4	3	10	20	29	39
δ =0.8	β =5	3	13	25	37	49

TABLE (4.9) $\underline{a} = (1, 0, 0, 0)', \underline{b} = (0, 0, 0, 1)'$

δ	β \n	n=1	n=2	n=3	n=4	n= 5
δ =.06	β =1	7	7	8	9	10
δ = 0.6	β =2	4	5	7	7	9
δ = 0.9	β =3	3	6	7	9	9
δ =1	β =4	3	7	9	9	9
δ = 1.3	β =5	2	6	7	9	11

These tables show the number of iterations needed to achieve $max\ F_j \leq 10^{-n}$; $n = 1, 2, 3, 4, 5$ and $j = 1, 2, 3$ on the intervals $(-\beta, \beta)$ for the criterion ψ_7 or ψ_8.

TABLE(4.10) D-Optimality criterion $\underline{a} = (1,0,0)', \underline{b} = (0,0,1)'$

	$\beta=1$	$\beta=2$	$\beta=3$	$\beta=4$
Support points	-1 0 1	-2 0 2	-3 0 3	-4 0 4
p^*	.25 .5 .25	.25 .5 .25	.25 .5 .25	.25 .5 .25
$\phi_5{}^*$	-4	-.25	-0.04938	-0.01562

TABLE(4.11) c-Optimality criterion $\underline{a} = (1,0,0)', \underline{b} = (0,0,1)'$

	$\beta=1$	$\beta=2$	$\beta=3$	$\beta=4$
Support points	-1 1	-2 0 2	-3 0 3	-4 0 4
p^*	.5 .5	.125 .75 .125	.056 .889 .056	.031 .938 .031
$\phi_7{}^*$	-1	-1	-1	-1

TABLE(4.12) A-Optimality criterion $\underline{a} = (1,0,0)', \underline{b} = (0,0,1)'$

	$\beta=1$	$\beta=2$	$\beta=3$	$\beta=4$
Support points	-1 0 1	-2 0 2	-3 0 3	-4 0 4
p^*	.207 .596 .207	.098 .805 .098	.05 .9 .05	.029 .942 .029
$\phi_8{}^*$	-5.8284477	-1.640388	-1.24828	-1.133056

These tables illustrate the optimal value for the criteria ϕ_5, ϕ_7 and ϕ_8 and the support points with theire optimal weights for the quadratic regression model.

TABLE(4.13) D-Optimality criterion $\underline{a} = (1,0,0,0)^t, \underline{b} = (0,0,0,1)^t$

	$\beta=1$	$\beta=2$	$\beta=3$	$\beta=4$
Supp(p^*)	-1 -.35 .35 1	-2 -.7 .7 2	-3 -1.05 1.05 3	-4 -1.4 1.4 4
p^*	.098 .402 .402 .098	.098 .402 .402 .098	.098 .402 .402 .098	.098 .402 .402 .098
ϕ_5^*	-33.97197	-.5123248	-.0466007	-.0082939

TABLE(4.14) c-Optimality criterion $\underline{a} = (1,0,0,0)^t, \underline{b} = (0,0,0,1)^t$

	$\beta=1$	$\beta=2$	$\beta=3$	$\beta=4$
Supp(p^*)	-1 .5 .5 1	-2 -.3 -.2 2	-3 -.2 -.1 3	-4 -.1 0.0 4
p^*	.208 .5 .167 .125	.063 .447 .442 .048	.019 .107 .857 .171	.008 .615 .369 .007
ϕ_7^*	-16	-1.3082	-1.07866	-1.0323

TABLE(4.15) A-Optimality criterion $\underline{a} = (1,0,0,0)^t, \underline{b} = (0,0,0,1)^t$

	$\beta=1$	$\beta=2$	$\beta=3$	$\beta=4$
Supp(p^*)	-1 -.46 .46 1	-2 -.5 .5 2	-3 -.43 .43 3	-4 -.35 .35 4
p^*	.15 .35 .35 .15	.06 .44 .44 .06	.02 .48 .48 .02	.009 .491 .491 .009
ϕ_8^*	-18.7627	-1.8000	-1.20506	-1.08341

These tables illustrate the optimal value for the criteria ϕ_5, ϕ_7 and ϕ_8 and the support points with theire optimal weights for the cubic regression model.

TABLE(4.16) $\underline{a} = (1,0,0)', \underline{b} = (0,0,1)'$ Quadratic regression Model

$i \backslash [-\beta, \beta]$	β=1	β=2	β=3	β=4
i=8	0.43	.68	0.82	0.89
i=7	0.07	0.41	0.66	0.78
i=5	0.42	0.42	0.42	0.42

TABLE(4.17) $\underline{a} = (1,0,0,0)', \underline{b} = (0,0,0,1)'$ Cubic Regression Model

$i \backslash [-\beta, \beta]$	β=1	β=2	β=3	β=4
i=8	1	1	1	1
i=7	0.85	0.73	0.89	0.95
i=5	1	1	1	1

TABLE(4.18) $\underline{a} = (1,0,0,0)', \underline{b} = (0,0,1,0)'$ Cubic Regression Model

$i \backslash [-\beta, \beta]$	β=1	β=2	β=3	β=4
i=8	0.62	0.72	0.81	0.92
i=7	0.11	0.44	0.65	0.77
i=5	0.48	0.48	0.48	0.48

These tables illustrate the efficiencies for ϕ_i^* relative to ψ_i^* for $i = 5$, $7,8$, and for various vectors $\underline{a}$, $\underline{b}$.

FIGURE (4.1)

This figure illustrates the optimal support points on the interval $(-1,\ 1)$ where the y-axes represent the maximum value for the criteria ψ_7 *or* ψ_8 and the x-axes represent the design points $x,\ x \in (-1,1)$.

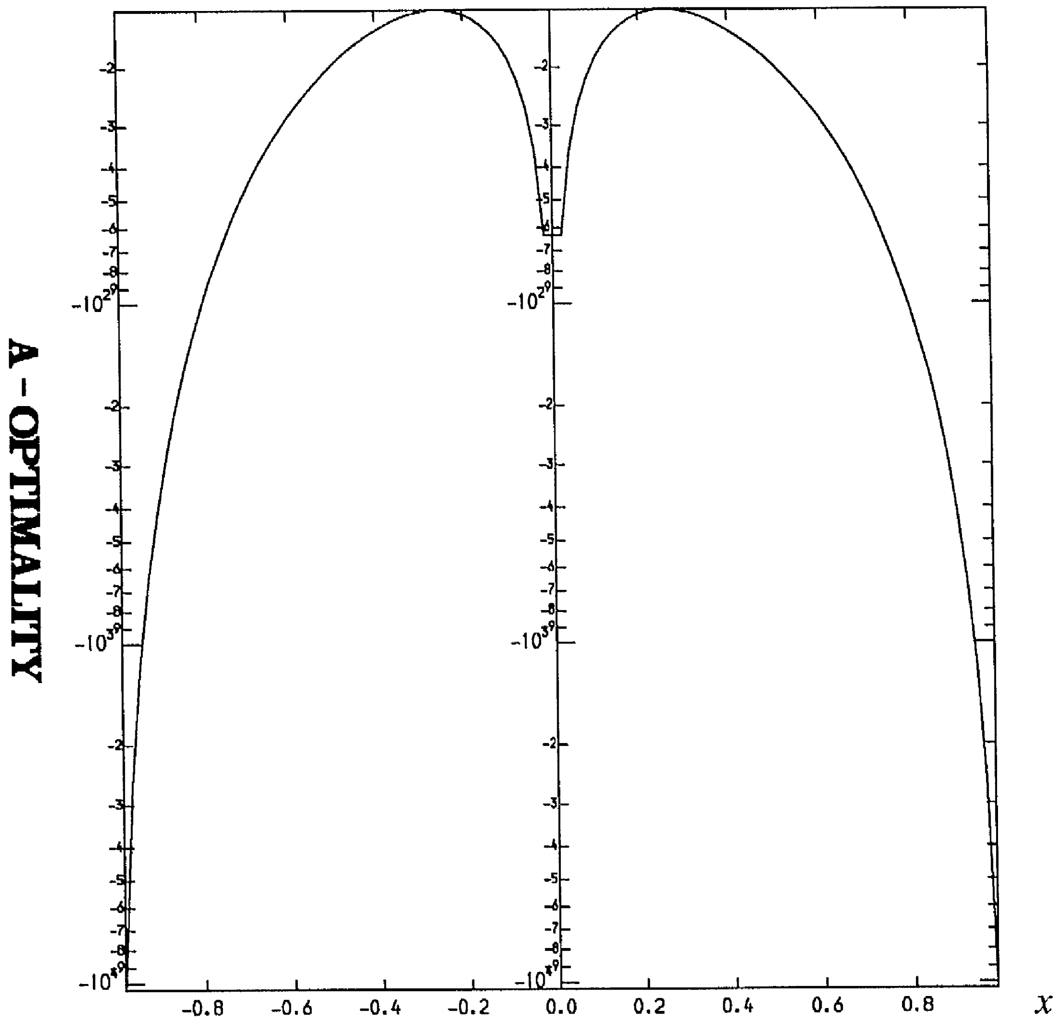


FIGURE (4.2)

This figure illustrates the optimal support points on the interval $(-2,\ 2)$ where the y-axes represent the maximum value for the criteria ψ_7 *or* ψ_8 and the x-axes represent the design points x, $x \in (-2,2)$.

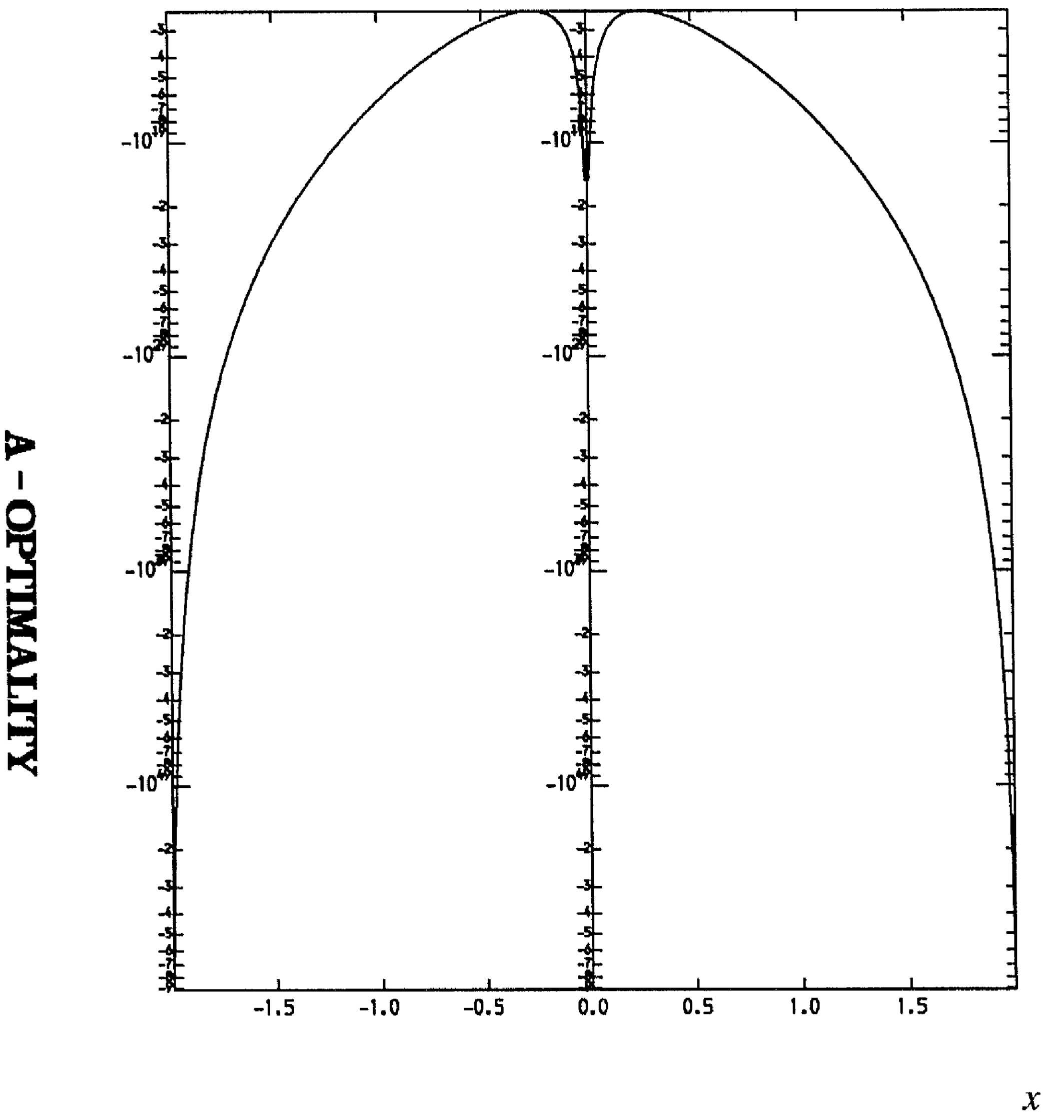


FIGURE (4.3)

This figure illustrates the value for the ψ_7 -optimality criterion against the weight h_1 on the intervals $[-\beta, \beta]$ for β=1,2,3,4 and 5 .

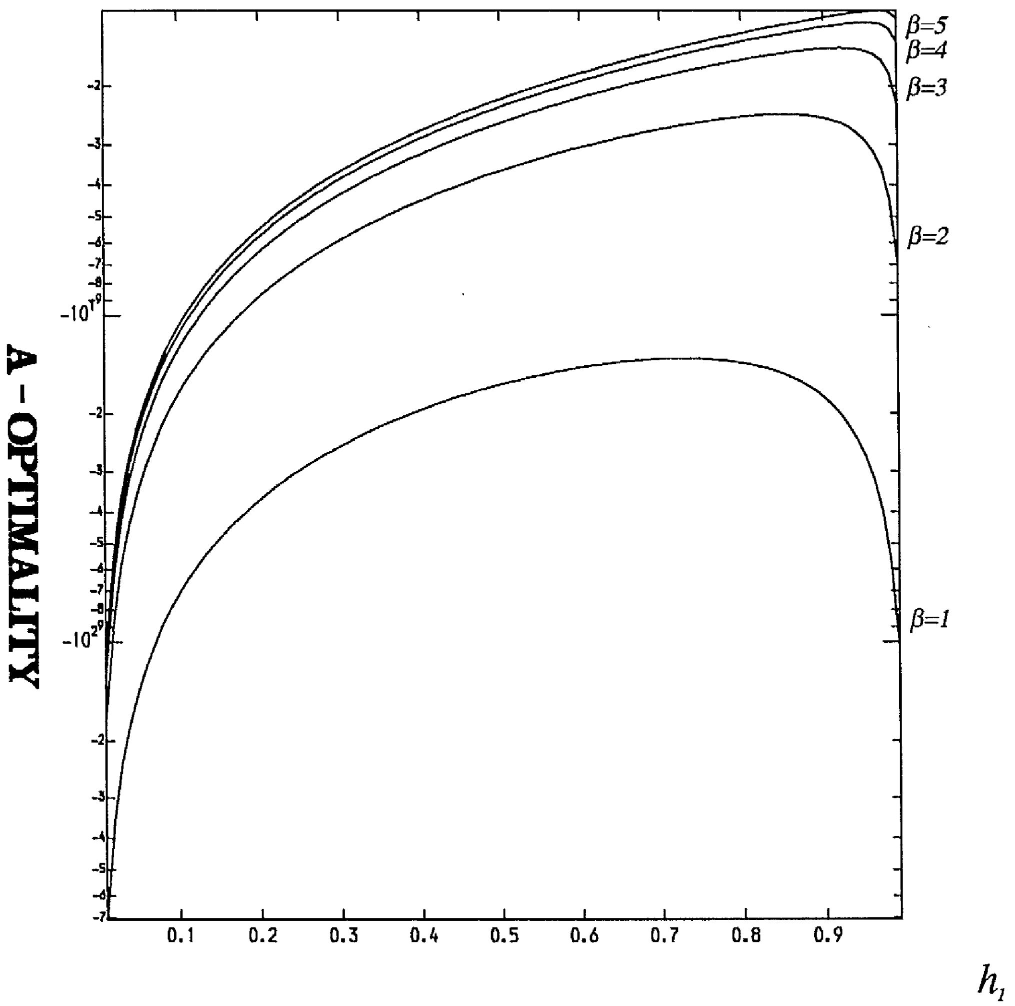
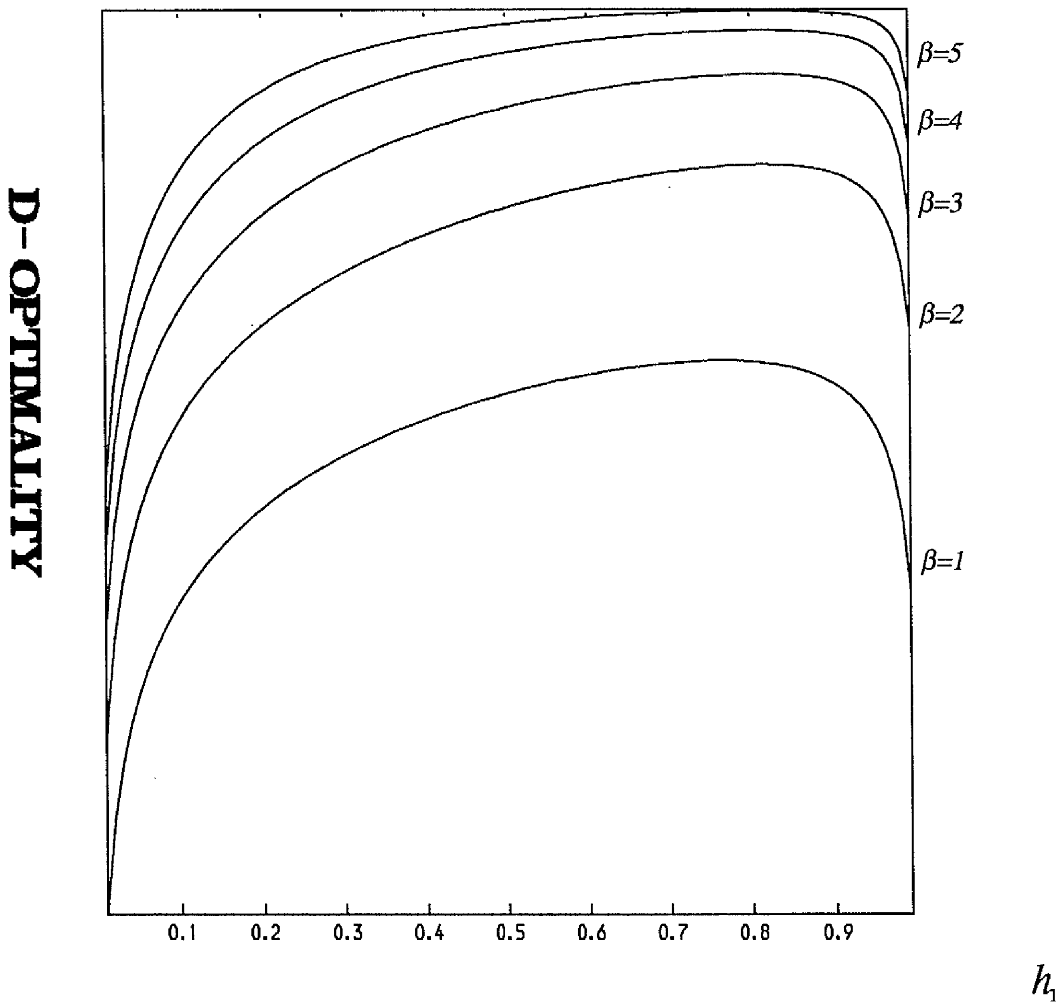


FIGURE (4.4)

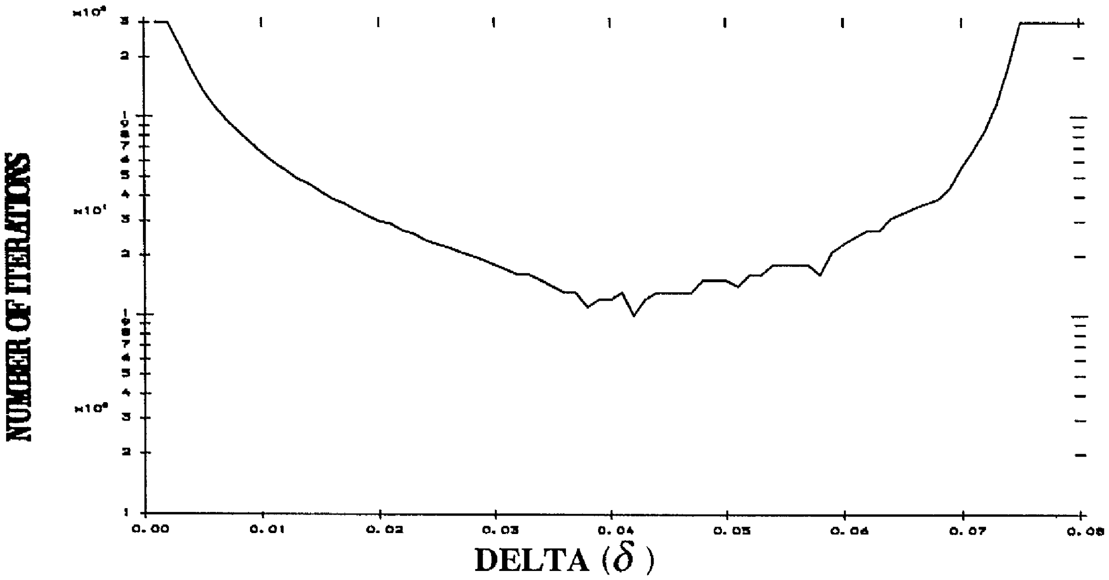
This figure illustrates the value for the ψ_5 -optimality criterion against the weight h_1 on the intervals $[-\beta, \beta]$ for β=1,2,3,4 and 5 .



These figures illustrate a plots of the number of iterations needed to achieve $\max F_i = 10^{-5}$ for different value of delta in the case of the criterion ψ_7.

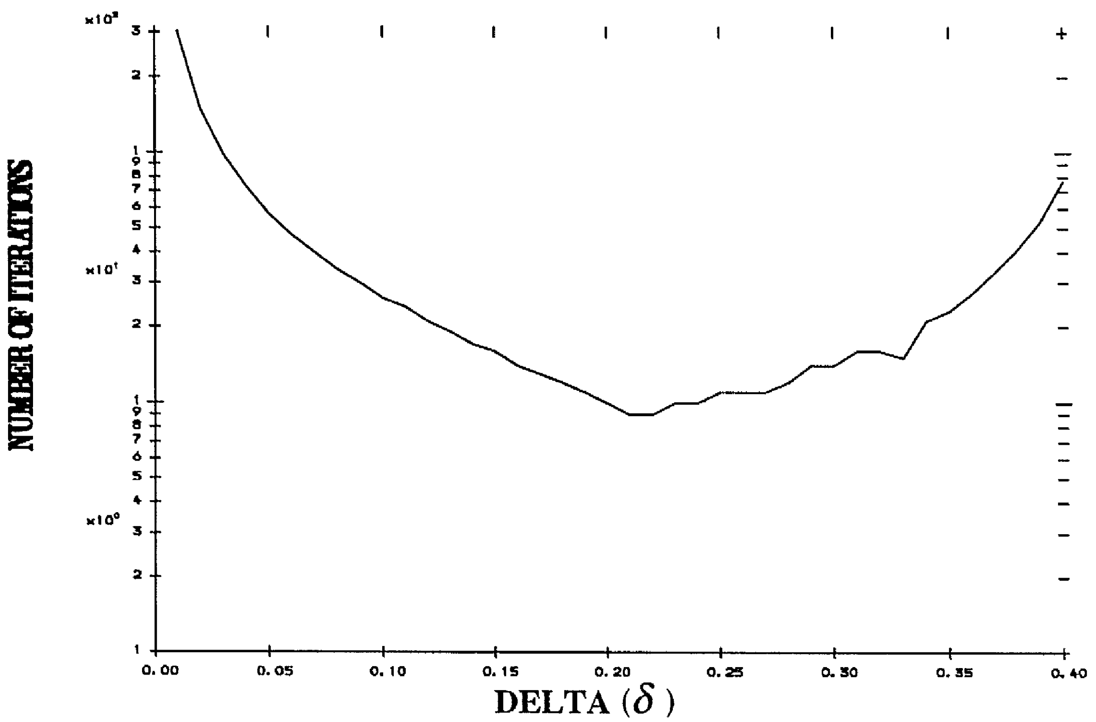
FIGURE(4.5)

Quadratic regression model $x \in [-1,1]$ and $\underline{a} = (1,0,0)'$, $\underline{b} = (0,0,1)'$



FIGURE(4.6)

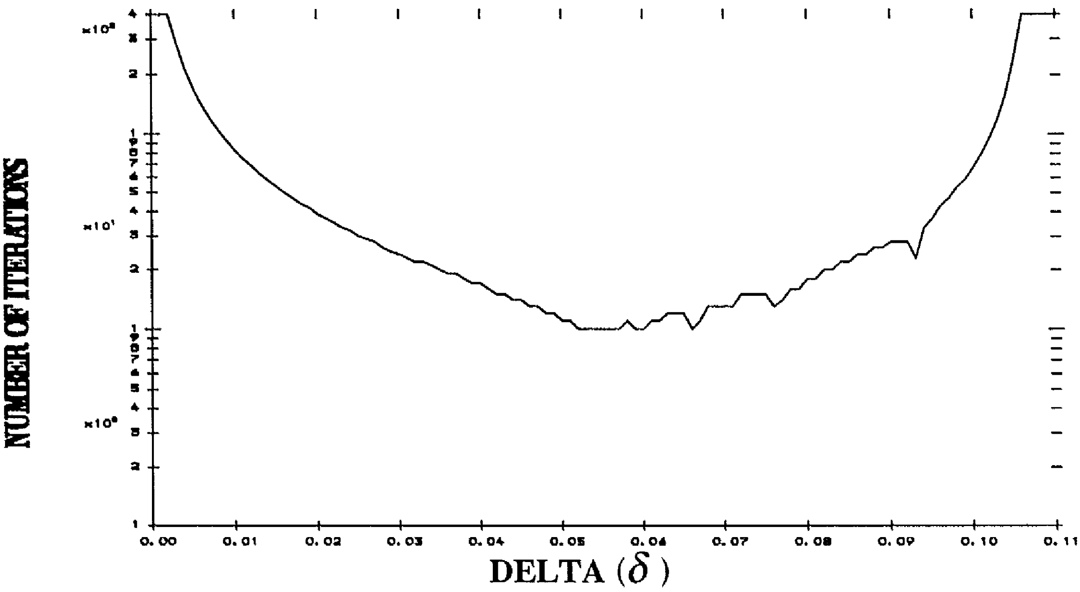
Quadratic regression model $x \in [-2,2]$ and $\underline{a} = (1,0,0)'$, $\underline{b} = (0,0,1)'$



These figures illustrate a plots of the number of iterations needed to achieve $\max F_i = 10^{-5}$ for different value of delta in the case of the criterion ψ_7.

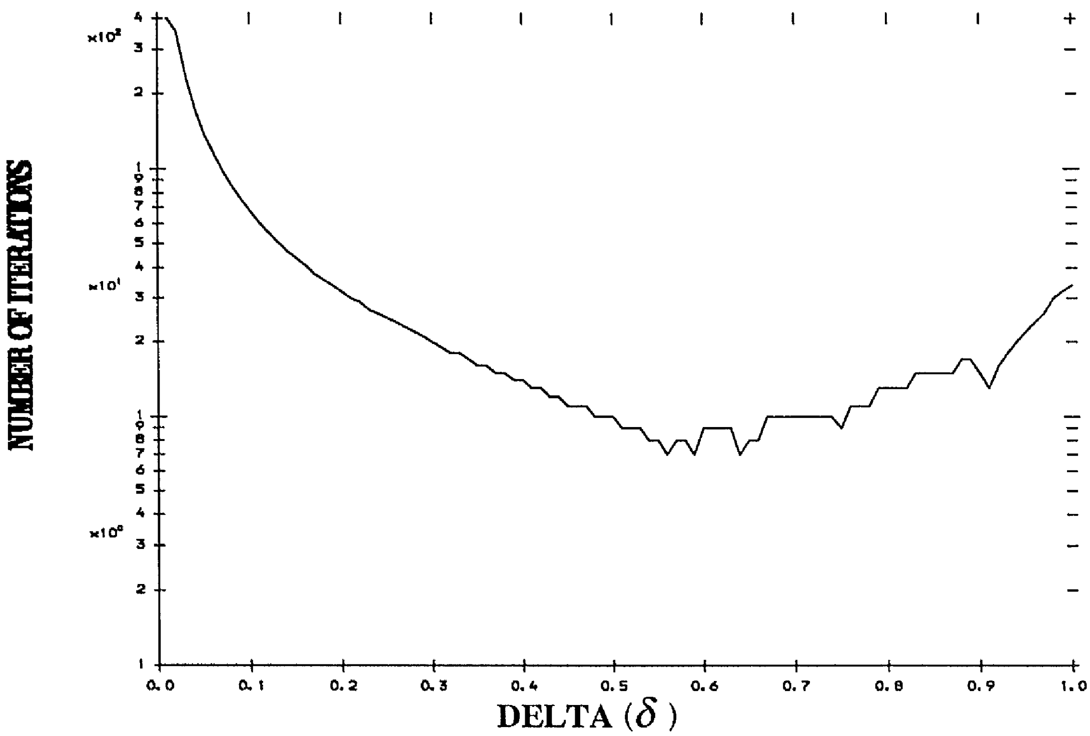
FIGURE(4.7)

Cubic regression model $x \in [-1,1]$ and $\underline{a} = (1,0,0,0)'$, $\underline{b} = (0,0,0,1)'$



FIGURE(4.8)

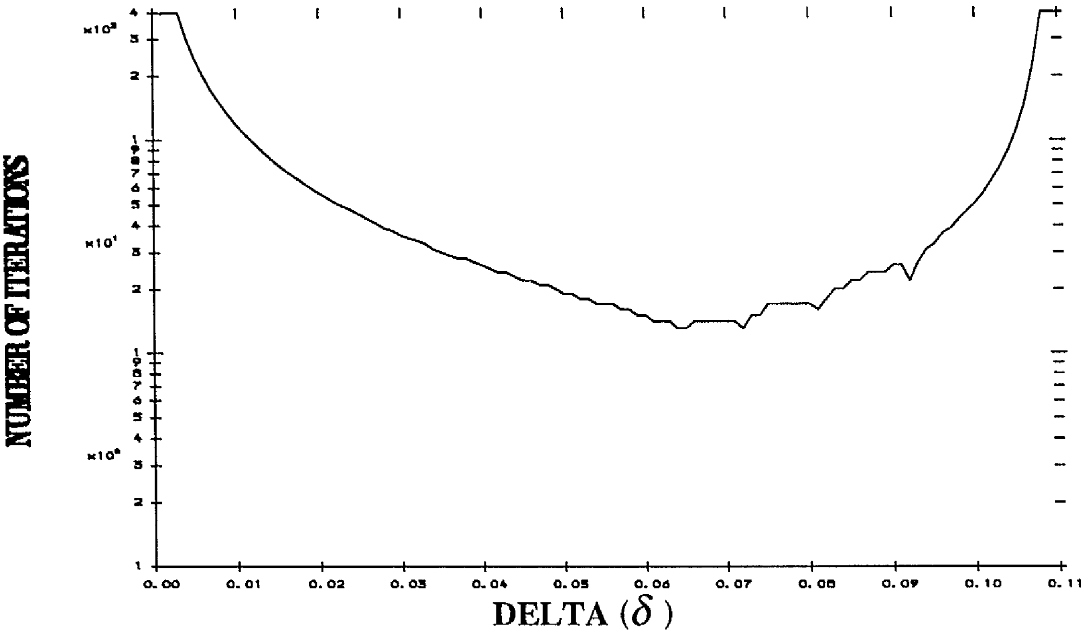
Cubic regression model $x \in [-2,2]$ and $\underline{a} = (1,0,0,0)'$, $\underline{b} = (0,0,0,1)'$



These figures illustrate a plots of the number of iterations needed to achieve $\max F_i = 10^{-5}$ for different value of delta in the case of the criterion ψ_7.

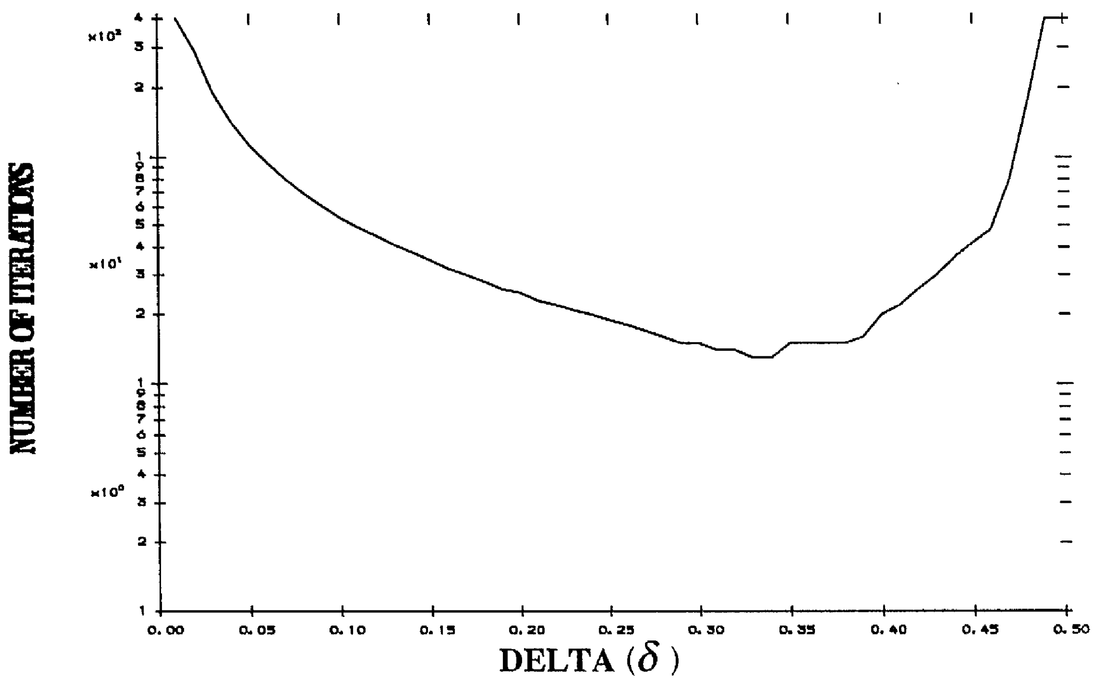
FIGURE(4.9)

Cubic regression model $x \in [-1,1]$ and $\underline{a} = (1,0,0,0)'$, $\underline{b} = (0,0,1,0)'$



FIGURE(4.10)

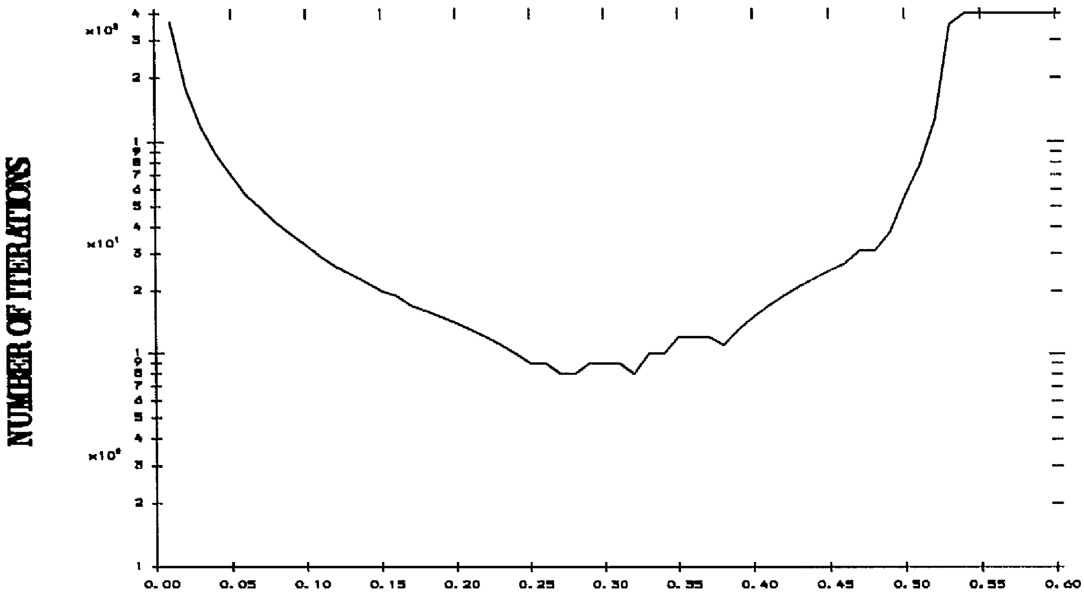
Cubic regression model $x \in [-2,2]$ and $\underline{a} = (1,0,0,0)'$, $\underline{b} = (0,0,1,0)'$



These figures illustrate a plots of the number of iterations needed to achieve $\max F_i = 10^{-5}$ for different value of delta in the case of the criterion ψ_5.

FIGURE(4.11)

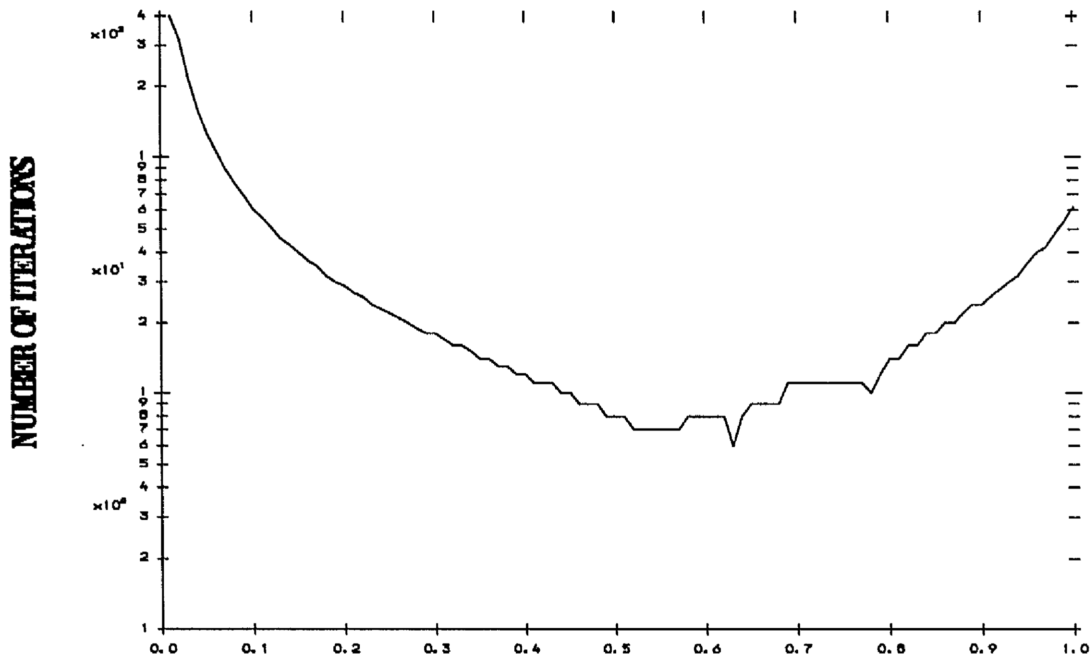
Quadratic regression model $x \in [-1,1]$ and $\underline{a} = (1,0,0)^t$, $\underline{b} = (0,0,1)^t$



DELTA (δ)

FIGURE(4.12)

Cubic regression model $x \in [-1,1]$ and $\underline{a} = (1,0,0,0)^t$, $\underline{b} = (0,0,0,1)^t$



DELTA (δ)

CHAPTER FIVE

Optimal Design Subject To Zero Covariance

(General Case)

5.1 Introduction

5.2 Quadratic Regression Model

5.3 General Problem (GP)

5.4 Proposed Algorithm

5.5 Examples and Discussions

CHAPTER FIVE

Optimal Design Subject To Zero Covariance

(General Case)

5.1 Introduction

In the previous Chapter we maximised a concave criterion subject to zero covariance by using a transformation of the weights which guaranteed zero covariance. This has been done just for the case when the number of design points equals the number of the unknown parameters $\underline{\theta}$.

In this Chapter we attempt to generalise the above problem by considering the case when the number of design points greater than the number of the unknown parameters $\underline{\theta}$. This can be done by applying a Lagrangian approach,see Bertsekas(1982), as we shall see in section (5.3). In the following section we consider a special case for the quadratic regression model.

5.2 Quadratic Regression Model

As a simple case we consider the quadratic regression model

$$E(y) = \theta_0 + \theta_1 x + \theta_2 x^2 \quad , \quad x \in [-1, 1]$$

Suppose $M(p) = V\ PV' = \sum_{i=1}^{J} p_i \begin{pmatrix} 1 \\ x_i \\ x_i^2 \end{pmatrix} \begin{pmatrix} 1 & x_i & x_i^2 \end{pmatrix}$, (5.2.1) where $P = diag(p_1, p_2, \dots, p_J)$ and V is the design matrix .

Suppose we are interested in estimating the parameter θ_0 independently from θ_2 i.e. by making the covariance between these two parameters equal to zero. Now $Cov(\theta_0, \theta_2) = Det(M_{21}) / Det(M(p))$ where M_{21} is the 2×2 matrix

$M_{21} = \sum_{i=1}^{J} p_i \begin{pmatrix} x_i \\ x_i^2 \end{pmatrix} (1 \quad x_i)$, a matrix element of a partition of $M(p)$ such that

$$M(p) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

Thus the covariance is equal to zero if $Det(M_{21})=0$.

Now $$M_{21} = \sum_{i=1}^{J} x_i \, p_i \begin{pmatrix} 1 \\ x_i \end{pmatrix} (1 \quad x_i)$$

$$= \sum_{i=1}^{J} q_i \begin{pmatrix} 1 \\ x_i \end{pmatrix} (1 \quad x_i) \ , \text{ where } q_i = x_i p_i \qquad (5.2.2)$$

So from (5.2.2) $Det(M_{21})$ is a polynomial of degree 2 in the weights $\underline{p}$, namely

$$Det(M_{21}) = \sum_{1 \le u < v \le J} q_u \, q_v \left(Det \begin{bmatrix} 1 & 1 \\ x_u & x_v \end{bmatrix} \right)^2 = \sum_{1 \le u < v \le J} q_u \, q_v \left(x_u - x_v \right)^2$$

$$= \sum_{1 \le u < v \le J} x_u x_v \, p_u \, p_v \, (x_u - x_v)^2 = \sum_{1 \le u < v \le J} p_u \, p_v \, c_{uv} , \qquad (5.2.3)$$

where $c_{uv} = x_u x_v (x_v - x_u)^2$.

Clearly from (5.2.3) the covariance can be zero if the values of the terms $c_{uv} = x_u x_v (x_v - x_u)^2$ have a mixture of signs i.e. when some of the x_i's are positive and the others are negative.

As an example of this, consider the case when $J = 4$ and suppose the support points are $\{x_1, x_2, x_3, x_4\}$ with $x_1 = -1 < x_2 < 0 < x_3 < x_4 = 1$. Then $c_{12}, c_{34} > 0$ and the others are negative. Letting $d_{ij} = -c_{ij}$ when $c_{ij} < 0$, the determinant of the matrix M_{21} can be zero if

$$c_{12}p_1p_2 + c_{34}p_3p_4 = d_{13}p_1p_3 + d_{14}p_1p_4 + d_{23}p_2p_3 + d_{24}p_2p_4 \ . \tag{5.2.4}$$

Since $$\sum_{i=1}^{4} p_i = 1 \Rightarrow p_3 = 1 - (p_4 + p_1 + p_2), \tag{5.2.5}$$

then by substituting for p_3 in (5.2.4) we get

$$\alpha p_4^2 + \beta p_4 + \gamma = 0, \tag{5.2.6}$$

where $\alpha = -c_{34}$, $\beta = (c_{34}(1 - p_1 - p_2) + d_{13}p_1 - d_{14}p_1 + d_{23}p_2 - d_{24}p_2)$ and

$\gamma = (c_{12}p_1p_2 - d_{13}p_1(1 - p_1 - p_2) - d_{23}p_2(1 - p_1 - p_2))$.

Also in the case when the support points are $\{x_1, x_2, x_3, x_4\}$ with $x_1 = -1 < x_2 < x_3 < 0 < x_4 = 1$ $Det(M_{21}) = 0$ implies (5.2.6) with

$\alpha = -d_{34}$, $\beta = (-p_1c_{13} - c_{23}p_2 - d_{14}p_1 - d_{24}p_2 - d_{34}(1 - p_1 - p_2))$ and

$\gamma = (c_{12}p_1p_2 - c_{13}p_1(1 - p_1 - p_2) - c_{23}p_2(1 - p_1 - p_2))$.

where in this case c_{12}, c_{13} and $c_{23} > 0$ and $d_{ij} = -c_{ij}$ when $c_{ij} < 0$.

So in these two cases we find

$$p_4 = (-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}) / 2\alpha \tag{5.2.6}$$

In (5.2.5) and (5.2.7) we have equations which enable us to substitute for p_3 and p_4 in terms of p_1 and p_2 in a standard criterion (e.g. D-optimality , A-optimality or c-optimality) to yield a function of p_1 and p_2 say $\phi_5(p_1,p_2),\phi_7(p_1,p_2)$ or $\phi_8(p_1,p_2)$ which we wish to maximize.

We explored this approach maximizing $\phi_5(p_1,p_2),\phi_7(p_1,p_2)$ or $\phi_8(p_1,p_2)$ by a search method (which ensured that all constraints on the weights were respected i.e $0 \le p_i \le 1, i = 1,2,3,4$ and $\sum_{i=1}^{4} p_j = 1$) in Example(2) of Chapter 4 in the case of the quadratic regression model for the above three criteria. We obtained the same solution as in that example.

5.3 General Problem (GP)

The general problem (GP) which we will consider is

Maximize $\phi_i(p)$ subject to the constraints $g(p) = 0$, $p_i \ge 0$, $\sum_{j=1}^{J} p_j = 1$, where $\phi_i(p)$ is one of the criteria A , c or D-optimality (i.e. for $i = 5,7,8$ respectively) and $g(p) = \underline{a}' M^{-1} \underline{b}$ for appropriate $\underline{a}$ and $\underline{b}$.

So by applying Lagrangian Theory(see Winston(1987)) we get

$$L(\phi,\underline{p},\lambda) = \phi(p) + \lambda g(p) \tag{5.3.1}$$

Then

$$d_i^L = \frac{\partial L}{\partial p_i} = \frac{\partial \phi}{\partial p_i} + \lambda \frac{\partial g}{\partial p_i} = d_i^\phi + \lambda d_i^g \tag{5.3.2}$$

and also the vertex directional derivatives are

$$F_i^L = d_i^L - \sum_{j=1}^{J} p_j d_j^L \tag{5.3.3}$$

$$= d_i^\phi - \sum_{j=1}^{J} p_j d_j^\phi + \lambda \left(d_i^g - \sum_{j=1}^{J} p_j d_j^g \right)$$

$$= F_i^\phi + \lambda F_i^g. \tag{5.3.4}$$

Clearly from (5.3.4) the vertex derivatives can be zero if

$$F_i^g \lambda = -F_i^\phi. \tag{5.3.5}$$

Now suppose $A = \underline{F}^g$, $\underline{b} = -\underline{F}^\phi$ *and* $\underline{\lambda} = \lambda$,then we can rewrite (5.3.5) in the form

$$A \underline{\lambda} = \underline{b}. \tag{5.3.6}$$

According to Graybill (1969) the set of solutions to the system of equations (5.3.6) ,if solutions exist , is given by:

$$\underline{\lambda} = A^- \underline{b} + (I - A^- A)\underline{z} \qquad \text{for any } \underline{z}, \tag{5.3.7}$$

which is unique if and only if

$$A A^- \underline{b} = \underline{b}. \tag{5.3.8}$$

Now let $A^- = (A'A)^{-1} A' = \left[(\underline{F}^g)' \underline{F}^g \right]^{-1} (\underline{F}^g)'$, then

$$A^- \underline{b} = \frac{-(\underline{F}^g)' \underline{F}^\phi}{\left[(\underline{F}^g)' \underline{F}^\phi \right]} = \hat{\lambda} \qquad \text{and} \qquad A A^- \underline{b} = \hat{\lambda} A = \hat{\lambda}\, \underline{F}^g.$$

So we need

$$\hat{\lambda}\,\underline{F}^{g} = -\underline{F}^{\phi} \quad \text{i.e} \quad \hat{\lambda}\,\underline{F}^{g} - \underline{F}^{\phi} = \underline{0} \tag{5.3.8}$$

and then substituting by $\hat{\lambda}$ in (5.3.8) we get

$$-\frac{(\underline{F}^{g})^{t}\underline{F}^{\phi}}{(\underline{F}^{g})^{t}\underline{F}^{g}}\,\underline{F}^{g} + \underline{F}^{\phi} = \underline{0} \tag{5.3.9}$$

or $$\underline{h} = \left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]\underline{F}^{\phi} - \left[(\underline{F}^{g})^{t}\underline{F}^{\phi}\right]\underline{F}^{g} = \underline{0}, \tag{5.3.10}$$

i.e $$\underline{h}^{t}\underline{h} = 0 \tag{5.3.11}$$

where

$$h_i = \left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]F_i^{\phi} - \left[(\underline{F}^{g})^{t}\underline{F}^{\phi}\right]F_i^{g},$$

$$= \left[\sum_{j=1}^{J}(F_j^{g})^{2}\right]F_i^{\phi} - \left[\sum_{j=1}^{J}(F_j^{g}F_j^{\phi})\right]F_i^{g}.$$

Now we try to simplify $\underline{h}^{t}\underline{h}$ as follows:

$$\underline{h}^{t}\underline{h} = \left\{\left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]\underline{F}^{\phi} - \left[(\underline{F}^{g})^{t}\underline{F}^{\phi}\right]\underline{F}^{g}\right\}^{t}\left\{\left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]\underline{F}^{\phi} - \left[(\underline{F}^{g})^{t}\underline{F}^{\phi}\right]\underline{F}^{g}\right\}$$

$$= \left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]^{2}\left[(\underline{F}^{\phi})^{t}\underline{F}^{\phi}\right] + \left[(\underline{F}^{g})^{t}\underline{F}^{\phi}\right]^{2}\left[(\underline{F}^{g})^{t}\underline{F}^{g}\right] -$$

$$\left[(\underline{F}^{g})^{t}\underline{F}^{\phi}\right]^{2}\left[(\underline{F}^{g})^{t}\underline{F}^{g}\right] - \left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]\left[(\underline{F}^{\phi})^{t}\underline{F}^{g}\right]^{2}$$

$$= \left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]^{2}\left[(\underline{F}^{\phi})^{t}\underline{F}^{\phi}\right] - \left[(\underline{F}^{g})^{t}\underline{F}^{\phi}\right]^{2}\left[(\underline{F}^{g})^{t}\underline{F}^{g}\right]$$

$$= \left[(\underline{F}^g)' \underline{F}^g \right]^2 \left[(\underline{F}^\phi)' \underline{F}^\phi \right] \left\{ 1 - \frac{\left[(\underline{F}^g)' \underline{F}^\phi \right]^2}{\left[(\underline{F}^\phi)' \underline{F}^\phi \right] \left[(\underline{F}^g)' \underline{F}^g \right]} \right\}$$

$$= \left[(\underline{F}^g)' \underline{F}^g \right]^2 \left[(\underline{F}^\phi)' \underline{F}^\phi \right] \{ 1 - R \} \qquad (5.3.12)$$

$$\text{where} \quad R = \frac{\left[(\underline{F}^g)' \underline{F}^\phi \right]^2}{\left[(\underline{F}^\phi)' \underline{F}^\phi \right] \left[(\underline{F}^g)' \underline{F}^g \right]} = \frac{\xi_1^2}{\xi_2 \, \xi_3}, \qquad (5.3.13).$$

$\xi_1 = \left[(\underline{F}^g)' \underline{F}^\phi \right]$, $\xi_2 = (\underline{F}^\phi)' \underline{F}^\phi$ and $\xi_3 = (\underline{F}^g)' \underline{F}^g$.

Clearly from (5.3.13) $0 \leq R \leq 1$. Thus from (5.3.12) $\underline{h}' \underline{h}$ is minimised when R is maximised i.e $\underline{h}' \underline{h} = 0 \iff R = 1$.

We need to be clear about what we have done here . We have replaced the first order conditions of a solution to problem (GP) by an optimisation problem , namely, maximisation of R (or any increasing function $Q(R)$ of R) by a p which must satisfy the constraints $0 \leq p_i \leq 1, i = 1,2,\ldots,J \text{ and } \sum_{i=1}^{J} p_i = 1$ and there is still $g(p) = 0$ to insure. We note that the latter is equivalent to maximising $G(p) = -g^2(p) = -(\underline{a}' M^{-1}(p) \underline{b})^2$ (assuming there exist p such that $g(p) = 0$).

If there is a solution to problem (GP), then the implication is that $G(p)$ and $Q(R)$ are simultaneously maximised by the same p^* subject to $0 \leq p_i \leq 1, i = 1,2,\ldots,J \text{ and } \sum_{i=1}^{J} p_i = 1$.

Further consider the case $Q_0(R) = -(1-R)^2$. Then $G_0(p)$ and $Q_0(R)$ have a common maximum of zero and hence so does $G_0(p)$ + $Q_0(R)$, at p^*. Of course determination of p^* will require numerical techniques.We discuss approaches in the following section .

5.4 Proposed Algorithm

In the previous section we transformed the solution of problem (GP) to maximisation of two functions simultaneously. So for this type of problem we suggest the following algorithm:

$$p_j^{(r+1)} = \frac{p_j^{(r)} \left[f(\delta, d_j^Q, d_j^G) \right]}{\sum_{i=1}^{J} p_i^{(r)} \left[f(\delta, d_i^Q, d_i^G) \right]} \tag{5.4.1}$$

where $$f(d_i^Q, d_i^G, \delta) = \begin{cases} \left[1 + (d_i^Q + d_i^G)\right]^{\delta} & \text{if } d_i^Q + d_i^G \geq 0 \\ \left[1 - (d_i^Q + d_i^G)\right]^{-\delta} & \text{if } d_i^Q + d_i^G < 0 \end{cases} \tag{5.4.2}$$

and d_i^Q , d_i^G are the first partial derivatives for the two functions which we want to maximise. These of course are given by

$$d_i^G = \frac{\partial G}{\partial p_i} = 2\left(\underline{a}' M^{-1}(p)\underline{b}\right)\left(\underline{a}' M^{-1}(p)\underline{v}_i\right)\left(\underline{v}_i' M^{-1}(p)\underline{b}\right) \quad \text{and}$$

$$d_i^Q = \frac{\partial Q}{\partial p_i} = 2R(1-R)\left[\frac{2(\partial \xi_1 / \partial p_i)}{\xi_1} - \frac{(\partial \xi_2 / \partial p_i)}{\xi_2} - \frac{(\partial \xi_3 / \partial p_i)}{\xi_3}\right],$$

$$\frac{\partial \xi_1}{\partial p_i} = \left[\sum_{j=1}^{J} \left(F_j^g \frac{\partial F_j^\phi}{\partial p_i} + F_j^\phi \frac{\partial F_j^g}{\partial p_i} \right) \right];$$

$$\frac{\partial \xi_2}{\partial p_i} = 2\left[\sum_{j=1}^{J} F_j^g \frac{\partial F_j^g}{\partial p_i} \right] \text{ and } \frac{\partial \xi_3}{\partial p_i} = 2\left[\sum_{j=1}^{J} F_j^\phi \frac{\partial F_j^\phi}{\partial p_i} \right].$$

and finally

$$\frac{\partial F_j^\phi}{\partial p_i} = \frac{\partial^2 \phi}{\partial p_j p_i} - \left[\frac{\partial \phi}{\partial p_i} + \sum_{s=1}^{J} p_s \frac{\partial^2 \phi}{\partial p_s p_i} \right] = \frac{\partial^2 \phi}{\partial p_j p_i} - \left[\frac{\partial \phi}{\partial p_i} + (h^\phi - 1)\frac{\partial \phi}{\partial p_i} \right] = \frac{\partial^2 \phi}{\partial p_j p_i} - h^\phi \frac{\partial \phi}{\partial p_i}$$

if ϕ is homogeneous of degree h^ϕ.

Similarly

$$\frac{\partial F_j^g}{\partial p_i} = \frac{\partial^2 g}{\partial p_j p_i} - \left[\frac{\partial g}{\partial p_i} + \sum_{s=1}^{J} p_s \frac{\partial^2 g}{\partial p_s p_i} \right] = \frac{\partial^2 g}{\partial p_j p_i} - h^g \frac{\partial g}{\partial p_i},$$ if g is homogeneous of degree h^g.

where $\frac{\partial g}{\partial p_i}, \frac{\partial \phi}{\partial p_i}, \frac{\partial^2 g}{\partial p_i p_j}$ and $\frac{\partial^2 \phi}{\partial p_i p_j}$ are the first partial derivatives and the second derivatives for the covariance function and the design criterion which we want to maximise.

The motivation for algorithm (5.4.1) is that it is algorithm (3.4.1) for maximisation of $\phi(p) = G_0(p) + Q_0(p)$. However since there are two sets of derivatives each of which must satisfy first order conditions (the d_j^Q must share a common value and the d_j^G must share a common value), the variations of (5.4.1) obtained by replacing $f(d_i^Q, d_i^G, \delta)$ by $f_1(d^Q, \delta) + f_2(d^G, \delta)$ or $f_1(d^Q, \delta) \times f_2(d^G, \delta)$ were considered, where

$$f_j(d_i,\delta) = \begin{cases} \left[1+d_i\right]^{\delta_j} & if \;\; d_i \geq 0 \\ \left[1-d_i\right]^{-\delta_j} & if \;\; d_i < 0 \end{cases} \quad ,j = 1,2.$$

5.5 Examples and Discussion

In this section we will consider three examples which were considered by Silvey et al (1978), with the aim of maximising the functions $Q(R) = -(1-R)^2$ and $G(p) = -\left(\underline{a}'M^{-1}(p)\underline{b}\right)^2$ simultaneously, under the choice of $\underline{a} = (1,0,0)'$ and $\underline{b} = (0,0,1)'$. The examples are defined by their design space as follows:

Example(1) $V = V_1 = \{\, (1,-1,-1)', (1,-1,1)', (1,1,-1)', (1,2,2)' \,\}$

Example(2) $V = V_2 = \{\, (1,-1,-1)', (1,-1,1)', (1,1,-1)', (1,2,3)' \,\}$

Example(3) $V = V_3 = \{\, (1,-1,-2)', (1,-1,1)', (1,1,-1)', (1,2,2)' \,\}$

In Table (5.1) we report the optimal weights and the optimal support points for all these examples in the case of the A-optimality criterion .

In Table (5.2) to Table (5.4) we report respectively for the algorithm(5.4.1) (modified in one instance)and its two variations the number of iterations needed to achieve $F_{\max}^{(Q+G)} = \max\ (F_i^Q + F_i^G) \;\; \leq 10^{-n} \;\; for\ n = 1,2,\ldots,5$ when $\delta = \delta_1 = \delta_2$ is taken to be the value which achieves $F_{\max}^{(Q+G)} \leq 10^{-5}$ in the smallest number of iterations. Initial weights were $p^{(0)} = 1/J$, where J is the number of design points.

Clearly from these tables there are widely varying values of δ which attain $F_{\max}^{(Q+G)} \leq 10^{-5}$ in the smallest number of iterations. This may be due to the nonhomogeneity of the function $Q(R) = -(1-R)^2$. Also from the same table, in terms of the number of iterations convergence is slow. One possible reason for this is that at the optimum p_j^*, partial derivatives corresponding to positive weights must be zero for both functions. Hence if $p^{(r)}$ is close to p^* proceeding from $p^{(r)}$ to $p^{(r+1)}$ will only slightly change the functions and weight values. We encountered this situation in chapter 3 with the correlation criterion. As in that case we can gain improvement by using the function $f(d,\delta) = (1+s\,\alpha\, d)^{s\delta}$, $s=sign(d)$, for some suitably chosen value of α instead of the function $f(d,\delta) = (1+s\;d)^{s\delta}$. It was necessary to apply this modification to the basic algorithm(5.4.1) when considering Example(2). The result reported in Table(5.2) for this example were obtained when $\alpha = 15$ (with $\delta = .08$ attaining faster convergence for this value of α).

Table(5.1) This table illustrate the optimal weights and corresponding support points for the above three examples .

	Example (1)				Example (2)				Example (3)			
p^*	.237	.270	.330	.163	.259	.230	.359	.152	.255	.355	.215	.175
supp(p^*)	v_{11}	v_{12}	v_{13}	v_{14}	v_{21}	v_{22}	v_{23}	v_{24}	v_{31}	v_{32}	v_{33}	v_{34}
$\phi(p^*)$	-1.7130				-1.5251				-1.49607			
λ	0.2406				0.3494				-0.0841			

Table(5.2) This table shows the number of iterations needed to achieve $F_{\max}^{Q+G} \le 10^{-n}$ for $n=1,2,...,5$ in the case of the A-optimality criterion under the choice of the function $f(d_i^Q, d_i^G, \delta)$, (with $\alpha = 15$ for Example(2)).

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	δ
Example(1)	12	28	34	69	233	0.3
Example(2)	30	38	40	62	163	0.08
Example(3)	12	273	288	313	426	0.2

Table(5.3) This table shows the number of iterations needed to achieve $F_{\max}^{Q+G} \leq 10^{-n}$ for $n=1,2,...,5$ in the case of the A-optimality criterion under the choice of the function $f_1(d^Q,\delta)+f_2(d^G,\delta)$.

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	δ
Example(1)	20	23	27	35	82	0.8
Example(2)	25	35	48	84	247	1.0
Example(3)	24	55	59	65	107	0.8

Table(5.4) This table shows the number of iterations needed to achieve $F_{\max}^{Q+G} \leq 10^{-n}$ for $n=1,2,...,5$ in the case of the A-optimality criterion under the choice of the function $f_1(d^Q,\delta)\times f_2(d^G,\delta)$.

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	δ
Example(1)	84	91	98	148	394	0.2
Example(2)	23	32	43	73	208	0.6
Example(3)	24	252	274	305	486	0.1

CHAPTER SIX

Conclusion and Further Work

6.1 Conclusion

6.2 Further Work

CHAPTER SIX

Conclusion and Further Work

6.1 Conclusion

The aim of the work which has been described in the previous chapters was to explore algorithms for solving the constrained optimisation problems mentioned early on in this thesis.

In chapter 2 for solving **problem(a)** we studied a class of multiplicative algorithms of the form $p_i^{(r+1)} = p_i^{(r)} f(d_i, \delta) / \sum_{j=1}^{J} p_j^{(r)} f(d_j, \delta)$, indexed by a function $f(d, \delta)$ which depends on the derivatives of the criterion $\phi(p)$ and a free parameter δ.

In the case of the D-optimality criterion, the algorithm has been investigated under three choices of the function $f(d, \delta)$, namely $f(d, \delta) = \ln(e + \delta d)$, $e^{\delta d} / (1 + e^{\delta d})$ *and* $c - e^{-\delta d}$ with optimal choices of $\delta = \delta^* = x^* / \sum_{i=1}^{J} p_i d_i$ (see section 2.4). Under the choice of $f(d, \delta) = c - e^{-\delta d}$ and for $c = 1.001$ the convergence of the algorithm is faster than that of the algorithm under the other two choices of $f(d, \delta)$. Moreover, if $c \approx 1$ and δ small then $f(d, \delta) \cong \delta d$. Iterations are then approximately those under

$f(d,\delta) = d$. This suggests that $f(d,\delta) = d$ is an efficient choice for the D-optimality criterion. Certainly it is known to be monotonic for this criterion (see Titterington (1976)).

In the case of the c-optimality criterion the algorithm also has been investigated under the choice of $f(d,\delta) = d^{1/2}$ using various initial designs, $p^{(0)}$, which put small weights on at least one of the support points. These included permutations $p^{(0)} = (\alpha,\alpha,\alpha,\beta)$, $p^{(0)} = (\alpha,\alpha,\beta,\beta)$ and $p^{(0)} = (\alpha,\beta,\beta,\beta)$ with $\beta \leq 10^{-12}$. At the first iteration the algorithm irresistibly moves immediately towards the optimal design on the subset of points receiving weights α. However the algorithm slowly moves away from this point and converges to the optimum.

In chapter 3, we first study the covariance and correlation criteria of **problem(b)** with the intention of estimating one or more of the unknown parameters as independently of the others as possible. Thus we wish to make the numerical covariance or correlation between the relevant parameter estimates as small as possible. In the case, when the number of design points exceeds the number of parameters, we explore a new version of the above algorithm by taking $f(d,\delta) = (1+s\,d)^{s\delta}$, $s = \text{sign}(d)$ which is defined for negative $\underline{d}$, where $\underline{d}$ is the vector of first partial derivatives of the covariance criterion $\phi_c(p)$. From the empirical results described in this chapter we found that the convergence of this algorithm is similar to that obtained by Fellman(1989) for the c-optimality criterion when $f(d,\delta) = d^{\delta}$. In particular as noted by Torsney(1983) $f(d,\delta) = d^{1/2}$ attains the optimum in one step for the c-optimality criterion when the support points form a linearly independent set of vectors.

However, in the case when the number of design points equals the number of parameters , the covariance criterion can be simplified by the formula :

$$\phi_c(p) = -\left[\underline{a}^t M^{-1}(p)\underline{b}\right]^2 = -\left[\sum_{i=1}^{k} \frac{c_i d_i}{p_i}\right]^2, \tag{6.1.1}$$

where $M(p) = V P V^t$, $\underline{c} = V^{-1}\underline{a}$ and $\underline{d} = V^{-1}\underline{b}$.

Clearly from (6.1.1) the value of the covariance criterion depends on the sign of $c_i d_i, i = 1,2,\ldots,k$. There are two cases to distinguish :

case(1): $\phi_c(p)$ cannot be zero if all $c_i d_i, i = 1,\ldots,k$ have the same sign.

case(2): $\phi_c(p)$ can be zero if the $c_i d_i, i = 1,2,\ldots,k$ have differing signs.

For **case(1)** the optimal weights has been found explicitly to be:

$$p_i^* = \sqrt{|c_i d_i|} \,/\, \sum_{j=1}^{k} \sqrt{|c_j d_j|} \quad , \; i = 1,2,\ldots,k$$

and the maximum value for the covariance criterion is $\phi_c(p^*) = \left\{\sum_{i=1}^{k} \sqrt{|c_i d_i|}\right\}^4$.

Chapters 4 and 5 are concerned with the solution of **problem(c)**.

In chapter 4 , we concentrated on **case(2)**. For this case we identify a class of designs which guarantee zero covariance . This class of design has been found by using a transformation for the weights $\underline{p}$ to two or three sets of variables each of which forms a probability vector. Under this transformation, **problem(c)** changes to a problem of maximising a criterion with respect to two probability vectors which yields an extension of **problem(a)** and in finding the optimal weights p^* and the optimal value for the criterion $\phi(p)$ under the zero covariance condition we use a natural extension of the algorithm used for that problem.

The performance of the algorithm in the case of the quadratic and cubic regression models has been investigated under three choices of the design criteria $\phi(p)$ namely, D_A-, c- and the Linear optimality criterion for $\underline{c} = \underline{a} + \underline{b}$ and $A' = [\underline{a} : \underline{b}]$, when the design space $\chi \in [-\beta, \beta]$, $\beta = 1,2,\ldots,5$. We observed that, the optimal value tends to increase as β increases for all these criteria. This has been proved in the case of the D_A-optimality criterion. The equivalence of the c- and the Linear optimality criteria also has been proven under the zero covariance condition.

Finally in this chapter, the calculation of the efficiencies for the restricted optimal design under the zero covariance constraint relative to the unrestricted optimal design suggested that these designs have good efficiencies for large β in the case of the c- and the Linear optimality criteria.

In chapter 5, we generalised **case(2)** by considering the case when the number of design points exceeds the number of parameters. Using a Lagrangian approach, the problem is transformed to one of simultaneous maximisation of two functions of the same probability vector each of which is maximised at the same value of this vector and have a common maximum of zero. For the purpose of finding the optimal value for these two functions we suggested the algorithm $p_i^{(r+1)} = p_i^{(r)} f(d_i^Q, d_i^G, \delta) / \sum_{j=1}^{J} p_j^{(r)} f(d_j^Q, d_j^G, \delta)$, where d_i^Q, d_i^G are the first partial derivatives for the two functions we want to maximise.

This algorithm is investigated under three choices of the function $f(d_i^Q, d_i^G, \delta)$, namely $f(d_i^Q, d_i^G, \delta) = f_1(d_i^{Q+G}, \delta)$, $f_2(d_i^Q, \delta) + f_3(d_i^G, \delta)$ and $f_2(d_i^Q, \delta) \times f_3(d_i^G, \delta)$ where $d_i^{Q+G} = d_i^Q + d_i^G$ and $f_j(d, \delta) = (1 + s\, d)^{s\,\delta}, j = 1,2,3$,$s = \text{sign}(d)$.Various examples were considered and we noted for all these examples, under the three choices of the function $f(d_i^Q, d_i^G, \delta)$, that the above algorithm often converges to the optimal solution but in terms of the number of iterations this convergence was slow . One possible reason for this, is that at the optimum p_j^* , partial derivatives corresponding to positive weights must be zero for both functions. Hence if $p^{(r)}$ is close to p^* proceeding from $p^{(r)}$ to $p^{(r+1)}$ will only slightly change the functions and weight values. But we suggested an improvement for this algorithm by considering the function $f_j(d,\delta) = (1+\alpha s\, d)^{s\delta}, j = 1,2,3$ for some suitably chosen value of α instead of $f_j(d,\delta) = (1+s\, d)^{s\delta}$.

6.2 Further Work

The following list sets out several ways in which the work of this thesis may be extended:

(1) Continue our investigation of the algorithms(2.3.1) in constructing optimal experimental designs with respect to other criteria namely, c- , A-optimality and other optimality criteria. Note that in these cases the optimal value of $\delta = \delta^* = x^* / \sum_{i=1}^{J} p_i d_i$ will have non-fixed values at each iteration. For example consider the case of c-optimality which is defined by the function $\phi_8(p) = -\underline{c}M^{-1}(p)\underline{c}$. This criterion is homogenous of degree -1 and then the value of $\sum_{i=1}^{k} p_i d_i = -\phi_8(p)$ (see Appendix 1) which yields a non-fixed value of δ^* during the iterative processing.

(2) Extend **problem(b)** to one of maximising a convex combination of two or more covariances or correlations e.g.

"Maximise $\psi_c(p) = -\left\{\alpha\left(\underline{a}M^{-1}(p)\underline{b}\right)^2 + (1-\alpha)\left(\underline{c}M^{-1}(p)\underline{b}\right)^2\right\}$ **or**

$$\psi_p(p) = -\left\{\alpha \frac{\left(\underline{a}M^{-1}(p)\underline{b}\right)^2}{\left(\underline{a}M^{-1}(p)\underline{a}\right)\left(\underline{b}M^{-1}(p)\underline{b}\right)} + (1-\alpha)\frac{\left(\underline{c}M^{-1}(p)\underline{b}\right)^2}{\left(\underline{c}M^{-1}(p)\underline{c}\right)\left(\underline{b}M^{-1}(p)\underline{b}\right)}\right\}$$

for $0 \leq \alpha \leq 1$ subject to the constraint $\sum_{i=1}^{k} p_i = 1;\ 0 \leq p_i \leq 1$".

(3) In chapter 3, we suggested for improving the convergence of the algorithm(3.4.1) the choice of the function $f(d,\delta) = (1+\alpha s\, d)^{s\,\delta}$ instead of $f(d,\delta) = (1+s\, d)^{s\,\delta}$ and we recommended $\alpha = 10^{2n}$ if $\max F_j < 10^{-n}$, $n=1,2,...,5$, where F_j , $j = 1,...,J$ are the vertex directional derivatives of the function we want to maximise. Presumably we could gain faster convergence of this algorithm by choosing the value of α optimally. One possible way of doing this would be to choose α to maximise the directional derivative of the criterion at $p^{(r)}$ in the direction of $p^{(r+1)}$ i.e. by maximising the function $F\left(p^{(r)}, p^{(r+1)}\right)$ with respect to α.

Appendixes

Appendix 1

Differentiation Of Matrices:

(1) Suppose A, B are $n \times n$ matrices whose elements are functions of a scalar variable z and $\underline{a} \in R^n$ is a constant vector, then :

(a) $\frac{\partial}{\partial z}\left[\underline{a}' A\right] = \underline{a}' \frac{\partial A}{\partial z}$.

(b) $\frac{\partial}{\partial z}\left[A + B\right] = \frac{\partial}{\partial z}(A) + \frac{\partial}{\partial z}(B)$.

(c) $\frac{\partial}{\partial z}\left[A\, B\right] = \frac{\partial}{\partial z}(A)\, B + A\, \frac{\partial}{\partial z}(B)$.

(d) $\frac{\partial}{\partial z}\left[Trace(A)\right] = Trace\left[\frac{\partial}{\partial z}(A)\right]$.

(2) Suppose A is non-singular and C is $k \times n$ constant matrix, then :

(a) $\frac{\partial}{\partial z}\left[A^{-1}\right] = -A^{-1}\left[\frac{\partial}{\partial z}(A)\right]A^{-1}$.

(b) $\frac{\partial}{\partial z}\left[C\, A^{-1}\, C'\right] = -C\, A^{-1}\left[\frac{\partial}{\partial z}(A)\right]A^{-1}\, C'$.

(c) $\frac{\partial}{\partial z}\left[\ln(Det\{\, A^{-1}\, \}\right] = Trace\left[A\, \frac{\partial}{\partial z}(\, A^{-1}\,)\right]$.

(d) $\frac{\partial}{\partial z}\left[\ln(Det\{C\, A^{-1}\, C'\,\}\right] = Trace\left[\,(C\, A^{-1}\, C'\,)^{-1} \frac{\partial}{\partial z}(C\, A^{-1}\, C'\,)\right]$.

So in the case of the D_A-optimality criterion defined by the function $\phi_5(p) = -\log \det\left[A'M^{-1}A\right]$, the first partial derivatives with respect to the weights p_i will be of the form :

$$\frac{\partial\phi_5}{\partial p_i} = \frac{\partial}{\partial p_i}\left[-\log \det\left\{A'M^{-1}(p)A\right\}\right],$$

$$= -Trace\left[\left(A'M^{-1}(p)A\right)^{-1}\frac{\partial}{\partial p_i}\left(A'M^{-1}(p)A\right)\right] \quad \text{by using (2)-(d),}$$

$$= -Trace\left[\left(A'M^{-1}(p)A\right)^{-1}\left\{-A'M^{-1}(p)\frac{\partial}{\partial p_i}[M(p)]M^{-1}(p)A\right\}\right], \text{by (2)-(b)}$$

but since the $J \times J$ matrix $M(p) = VPV'$,where V is a $J \times k$ design matrix which does not depend on the weights p and $\underline{P} = diag(p_1, p_2, p_3, \ldots, p_k)$, then

$$\frac{\partial\phi_5}{\partial p_i} = -Trace\left[\left(A'M^{-1}(p)A\right)^{-1}\{-A'M^{-1}(p)\underline{v}_i\underline{v}_i'M^{-1}(p)A\}\right],$$

$$= Trace\left[\underline{v}_i'M^{-1}(p)A\left(A'M^{-1}(p)A\right)^{-1}A'M^{-1}(p)\underline{v}_i\right], \text{since } Trace(AB) = Trace(BA),$$

$$= \underline{v}_i'M^{-1}(p)A\left(A'M^{-1}(p)A\right)^{-1}A'M^{-1}(p)\underline{v}_i \ .$$

Note that ,the first partial derivatives for the D-optimality criterion can be derived from the above derivatives by taking $A=I_k$ where I_k is the kxk identity matrix. Moreover for this criterion, the value of $\sum_{i=1}^{K} p_i\frac{\partial\phi_1}{\partial p_i} = k$,where k is the number of the unknown parameters .This can be shown as follows:

$$\sum_{i=1}^{K} p_i\frac{\partial\phi_1}{\partial p_i} = \sum_{i=1}^{K} p_i\left[\underline{v}_i' M^{-1}(p)\ \underline{v}_i\right], \text{ since } \frac{\partial\phi_1}{\partial p_i} = \underline{v}_i' M^{-1}(p)\ \underline{v}_i,$$

$$=Trace\left\{\sum_{i=1}^{K} p_i\left[\underline{v}_i^t M^{-1}(p)\ \underline{v}_i\right]\right\}=Trace\left\{M^{-1}(p)\sum_{i=1}^{K} p_i\underline{v}_i^t\ \underline{v}_i\right\},$$

$$=Trace\left\{M^{-1}(p)\,M(p)\right\}=Trace\left\{I_k\right\}=k.$$

Similarly ,we can derive the first partial derivatives for the other optimality criteria. We note that $\sum_{i=1}^{K} p_i\frac{\partial\phi}{\partial p_i}=h^{\phi}\phi\ (p)$ if the criterion is a homogeneous function of degree h^{ϕ}. For example, in the case of the c-optimality criterion which is defined by the function $\phi_8(p)=-\underline{c}^t M^{-1}(p)\underline{c}$,this criterion is homogeneous of degree -1 and then

$$\sum_{i=1}^{k} p_i\frac{\partial\phi_8}{\partial p_i}=\sum_{i=1}^{k} p_i\left[\underline{c}^t M^{-1}(p)\underline{v}_i\right]^2,\ \text{since}\ \frac{\partial\phi_8}{\partial p_i}=\left[\underline{c}^t M^{-1}(p)\underline{v}_i\right]^2$$

$$=\underline{c}^t M^{-1}(p)\left\{\sum_{i=1}^{k} p_i\underline{v}_i\underline{v}_i^t\right\}M^{-1}(p)\ \underline{c}$$

$$=\underline{c}^t M^{-1}(p)\ M(p)M^{-1}(p)\ \underline{c}=\underline{c}^t M^{-1}(p)\ \underline{c}=-\phi_8(p).$$

REFERENCES

REFERENCES

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs.* Oxford Statistical Science Series-8. Oxford University Press.

Atwood, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist.*, **Vol. 40**, 1570-1602.

Atwood, C. L. (1976). Convergent design sequences for sufficiently regular optimality criteria . *Annals of Statistics*, **Vol. 4**,1124-1138.

Atwood, C. L. (1980). Convergent design sequences for sufficiently regular optimality criteria, II: singular case. *Annals of Statistics*. **Vol. 8**, 894-912.

Bertsekas, D. P. (1982). *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, Inc. , New York and London.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.*, **Vol. 24**, 586-602.

Dempster, A. P. , Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *JRSSB*, **Vol. 39**, 1-38.

Elfving, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.*, **Vol. 23**, 255-262.

Fedorov, V. V. (1972). *Theory of optimal experiments*. Academic Press: New

York and London.

Fellman, J. (1974). On the allocation of linear observations. *Commentationes Physico-Mathematicae (Finska Vetenskaps Societeten)* ,**Vol. 44**, 27-77.

Fellman, J. (1989). An empirical study of a class of iterative searches for optimal designs. *Journal of Statistical Planning and Inference*, North-Holland, **Vol. 21**, 85-92.

Ford, I. (1976). *Optimal static and sequential design*: A critical review. Ph.D. Thesis. University of Glasgow.

Gaffke, N. and Mather, R. (1990). On a class of algorithms from experimental design theory. *Report No. 196. Institut für Mathematik.* Universität Augsburg, Germany.

Graybill, F. A. (1969). *Introduction to matrices with applications in statistics*. Wadsworth Pub. Co. Inc., Belmont, calif.

Karlin, S. and Studden, W.J. (1966). Optimal experimental Designs. *Ann. Math. Statist.*, **Vol. 37**, 783-815.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory. *Ann. Statist.*, **Vol 2**, 849-879.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian J, Math.* , **Vol. 12**, 363-366.

Kitsos, C. P., Titterington, D. M. and Torsney, B. (1988). An optimal design problem in rhythmometry. *Biometrics*, **Vol. 44**, 657-671.

Nielsen, K. L. (1964). *Methods in Numerical Analysis*. The Macmillan Company, New York and Coller-Macmillan Limited, London.

Pazman, A. (1986). *Foundations of optimum experimental design*. Reidel, Dordrecht.

Pukelsheim, F. and Rosenberger, J. L. (1993). Experimental designs for model discrimination, *Journal of the American Association*, **Vol. 88**, No. 422, 642-649.

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Annals of statistics*, **Vol. 19**, 1614-1625.

Rhode, C. A. (1965). Generalized inverses of partitioned matrices. *J. Soc. Indust., Appl. Math.*,**Vol. 13**,1033-1053.

Seber, G. A. F. (1977). *Linear regression analysis*. Wiley, New York.

Shah, K. R. and Sinha, B. K. (1989). Theory of optimal designs. *Lecture Notes in Statistics*, **Vol. 54**. Springer-Verlag.

Sibson, R. (1974). D_A-optimality and duality. Progress in Statistics. *Colloq. Math. Soc. Janos Bolyai*, **Vol 9**, 677-692.

Silverman, B. W., Jones, M. C., Wilson, J. D. and Nychka, D. W. (1990). A smoothed EM approach to indirect estimation problems with particular reference to stereology and emission tomography (with discussion). *JRSSB*, **Vol. 52**, 271-324.

Silvey, S. D. (1980). *Optimum design*. Chapman & Hall. London.

Silvey, S. D., Titterington, D. M. (1973). A geometric approach to optimal design theory. *Biometrika*, **Vol. 60**,21-32.

Silvey, S. D., Titterington, D. M. and Torsney, B. (1978). An algorithm for optimal designs on a finite design space. *Communications in Statistics*, **A7 Vol. 14**, 1379-1389.

Titterington, D. M. (1976). Algorithms for computing *D*-optimal designs on a finite design space. Proceedings, *Conference on Information Sciences and Systems, Dept. Elect. Eng.* , Jones Hopkins Univ., Baltimore **MD**, 213-216.

Torsney, B. (1977). Contribution to discussion of "Maximum likelihood estimation via the EM algorithm" by Dempster et al., *JRSSB*, **Vol. 39**, 26-27.

Torsney, B. (1981). *Algorithms for a constrained optimization problem with applications in statistics and optimum design*. Ph.D. Thesis, University of Glasgow.

Torsney, B. (1983). A moment inequality and monotonicity of an algorithm. Proceedings, International Symposium on Semi-Infinite Programming and Applications (Kortanek, K. O. and Fiacco, A. V. Eds.) at Univ. of Texas, Austin, *Lecture Notes in Economics and Mathematical Systems*, **Vol. 215**, 249-260.

Torsney, B. (1986). Moment Inequalities via optimal design theory. *Lin. Alg. &*

Its Applns., **Vol. 82**, 237-253.

Torsney, B. (1988). Computing optimizing distributions with applications in design, estimation and image processing. *Optimal Design and Analysis of Experiments* (Y. Dodge, V. V. Fedorov and H. P. Wynn -Eds) North Holland, 361-370.

Torsney, B. and Alahmadi, Abdulhadi, M. (1992). Further development of algorithms for constructing optimizing distributions. *Model Oriented Data Analysis*. Proceedings of 2nd IIASA Workshop in St. Kyrik, Bulgaria (V. V. Fedorov, W. G. Müller, I. N. Vuchkov -Eds), Physica-Verlag, 121-129.

Whittle, P. (1973). Some general points in the theory of optimal experimental designs. *JRSSB*, **Vol. 35**, 123-130.

Winston, W. L.(1987). *Operations Research: Applications and Algorithms*. Duxbury Press, Boston.

Wu, C. F. (1978). Some iterative procedures for generating nonsingular optimal designs. *Communications in Statistics* A7, **Vol. 14**,1399-1412.

Wynn, H. P. (1972). Results in the theory and construction of *D*-optimum experimental designs (with Discussion). *JRSSB*, **Vol. 34**, 133-147,170-186.

83, 117