



Mandal, S., Torsney, B. and Chowdhury, M. (2017) Optimal designs for minimising covariances among parameter estimators in a linear model. *Australian and New Zealand Journal of Statistics*, 59(3), pp. 255-273.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

This is the peer reviewed version of the following article: Mandal, S., Torsney, B. and Chowdhury, M. (2017) Optimal designs for minimising covariances among parameter estimators in a linear model. *Australian and New Zealand Journal of Statistics*, 59(3), pp. 255-273, which has been published in final form at <http://dx.doi.org/10.1111/anzs.12195>. This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

<http://eprints.gla.ac.uk/149214/>

Deposited on: 13 November 2017

Optimal designs for minimising covariances among parameter estimators in a linear model

S. Mandal^{1*}, B. Torsney² and M. Chowdhury³

University of Manitoba, University of Glasgow and University of Calgary

Summary

We construct approximate optimal designs for minimising absolute covariances between least-squares estimators of the parameters (or linear functions of the parameters) of a linear model, thereby rendering relevant parameter estimators approximately uncorrelated with each other. In particular, we consider first the case of the covariance between two linear combinations. We also consider the case of two such covariances. For this we first set up a compound optimisation problem which we transform to one of maximising two functions of the design weights simultaneously. The approaches are formulated for a general regression model and are explored through some examples including one practical problem arising in Chemistry.

Key words: Lagrangian optimality conditions; multiplicative algorithms; near uncorrelated parameter estimators; optimal design theory; vertex directional derivatives.

1. Introduction

In regression designs it may be desirable to render certain parameter estimators uncorrelated with others. Interest in this objective is motivated by a number of practical problems. One example of such a problem arises in Chemistry, and was considered by Torsney & Alahmadi (1995). They studied a three parameter regression model which describes the relationship between the viscosity and the concentration (x) of a chemical solution, the latter being the design variable. It was needed to construct a design such that the parameter estimators of interest are as uncorrelated as possible. The design interval for x was extremely short: $(0, 0.2]$. Parameter estimators under any design would be highly correlated. It was essential to limit this. Further details are provided in Section 5. Another such problem is a study of lipoprotein levels in relation to albumin concentration in childhood nephrotic syndrome considered by Noto et al. (1999). They considered relationships among several variables including lipoproteins, plasma albumin and proteinuria, and studied a group of 84

* Author to whom correspondence should be addressed.

¹ Department of Statistics, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada
Email: saumen.mandal@umanitoba.ca

² School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QW, United Kingdom

³ Department of Community Health Sciences, University of Calgary, Calgary, AB, T2N 4Z6, Canada

children in different clinical stages of the disease for five years. They performed a series of multiple regression analyses so that the resulting parameter estimators were uncorrelated with each other. A third such problem gave rise to a factorial experimental design, constructed so as to identify the most significant factors in the formulation of ascorbic acid tablets with hydrophilic polymers, used as stabilizers against oxidative degradation. This was considered by Odeniyi & Jaiyeoba (2009). They structured the design in such a way that the estimators of the parameters corresponding to polymer type, polymer concentration and relative density were uncorrelated with each other. A 2^3 full factorial design was used for this study.

The foregoing practical problems motivate us to construct a design such that certain parameter estimators in a regression model are as uncorrelated with each other as possible. Ideally we would aim at *independence* of parameter estimators, but such a goal is much more demanding. Recall that a lack of covariance or correlation implies independence only if the corresponding variables have a joint Gaussian distribution. In this work we aim to minimise the magnitude of covariances among relevant parameter estimators. This problem is an example of optimal regression design in which we need to calculate an optimising probability distribution. Some previous work in this direction includes Mandal, Torsney & Carriere (2005), Torsney & Mandal (2001), and Torsney & Alahmadi (1995). Mandal, Torsney & Carriere (2005) and Torsney & Mandal (2001) constructed optimal designs by maximising a standard criterion subject to the constraint of zero covariance of the estimators of parameters of interest. They solved this constrained optimisation problem by a Lagrangian method. Torsney & Alahmadi (1995) considered the optimisation problem by maximising covariance and correlation criteria, and formulated the criterion function as the square of a linear combination of the reciprocals of the design weights. They constructed optimal designs by considering the cases of common or opposite signs of the coefficients of the linear combination.

The contribution of the present work, and the differences between it and the references cited above, are as follows. In contrast with the work of Mandal, Torsney & Carriere (2005) and Torsney & Mandal (2001), the present work is based directly on optimising covariances without considering a standard optimality criterion such as D or A -optimality. In contrast with the work of Torsney & Alahmadi (1995), we formulate the criterion as a general function of the weights and do not need to consider the signs of the coefficients.

Another important contribution of this paper is that we address not only the problem of obtaining two parameter estimators, which in the examples considered are uncorrelated with each other, but also the problem of obtaining more than two parameter estimators with more than one zero correlation among them. We achieve this goal by considering a compound optimisation problem and solving the problem by means of simultaneous optimisation techniques. More specifically, we transform the problem to an optimisation problem in which

we maximise two functions of the design weights simultaneously. In the examples considered the functions have a common maximum of zero which is simultaneously attained at the optimal design weights.

The remainder of this paper is organised as follows. In Section 2, we first provide a brief introduction to optimal design theory. We determine the first and second order optimality conditions firstly in Lagrangian terms, and then interpret them in terms of optimal design concepts. In Section 3, we formulate the problem of minimising covariances and consider the cases of two or more linear combinations of the parameters. We also propose two compound criteria in the case of more than two linear combinations of the parameters. In Section 4, we describe a detailed algorithmic approach to our optimisation problems. In Section 5, we consider some examples and report results. Finally, we make some concluding remarks in Section 6.

2. Optimality conditions

Before we determine the optimality conditions of our optimisation problem, it is expedient that we review optimal design theory. In an optimal design setup, we first assume a probability model $y \sim \pi(y|\mathbf{x}, \boldsymbol{\theta}, \sigma)$, where y is the response variable; \mathbf{x} is a vector of design variables, $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^m$, \mathcal{X} is the design space; $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^\top$ is a vector of unknown parameters; and σ is a nuisance parameter. In linear models we assume $E(y|\mathbf{v}) = \mathbf{v}^\top \boldsymbol{\theta}$, where $\mathbf{v} \in \mathcal{V}$, $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^k: \mathbf{v} = \boldsymbol{\eta}(\mathbf{x})\}$ with $\boldsymbol{\eta}(\mathbf{x}) = (\eta_1(\mathbf{x}), \eta_2(\mathbf{x}), \dots, \eta_k(\mathbf{x}))^\top$.

An approximate design is characterised by a probability measure, say p , defined on \mathcal{X} and hence on \mathcal{V} . In practice we discretise these spaces. Suppose we discretise the design space \mathcal{X} into J distinct points, say, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J$. Then $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_J\}$, where $\mathbf{v}_j = \boldsymbol{\eta}(\mathbf{x}_j)$, $j = 1, 2, \dots, J$. We then specify p by a set of weights p_j satisfying $p_j \geq 0$, $j = 1, 2, \dots, J$, $\sum_{j=1}^J p_j = 1$. The weight p_j is assigned to \mathbf{v}_j . We seek to choose the vector $\mathbf{p} = (p_1, p_2, \dots, p_J)^\top$ optimally. Let $\hat{\boldsymbol{\theta}}$ be the least squares estimator of $\boldsymbol{\theta}$. Then the corresponding covariance matrix is $\text{cov}(\hat{\boldsymbol{\theta}}) \propto \mathbf{M}^{-1}(\mathbf{p})$, where $\mathbf{M}(\mathbf{p})$ is the per observation information matrix, given by $\mathbf{M}(\mathbf{p}) = \sum_{j=1}^J p_j \mathbf{v}_j \mathbf{v}_j^\top = \mathbf{V} \mathbf{P} \mathbf{V}^\top$ where $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_J]$ and $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_J)$. Generally we choose the proportion p_j of observations taken at \mathbf{x}_j , for good estimation of $\boldsymbol{\theta}$, by optimising some criterion, say $\phi(\cdot)$. For further details, see Fedorov (1972), Silvey (1980), Pukelsheim (2006), Atkinson, Donev & Tobias (2007), and Berger & Wong (2009).

The criterion $\phi(\cdot)$ should make $\text{cov}(\hat{\boldsymbol{\theta}})$ ‘small’ and hence $\mathbf{M}(\mathbf{p})$ large. There are many standard criteria. Here we opt for a non-standard one, namely the negative of the squared covariance between two parameter estimators. As our criterion function is a non-standard

one, and is not concave, we first establish first and second order optimality conditions in the following sub-sections.

2.1. Problem of interest

Our problem is an example of the following.

Problem (1)

$$\text{Maximise a criterion } \phi(\mathbf{p}) \text{ over } \mathcal{P} = \left\{ \mathbf{p} = (p_1, p_2, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1 \right\}. \quad (1)$$

We first appeal to Lagrangian Theory to identify sufficient conditions for a local maximum. We then transform these to equivalent optimal design theory terminology as this provides both clear physical interpretations of the sufficient conditions, and also a basis for some versions of the algorithm we will use later. Of note is that our feasible region is the convex set \mathcal{P} which has the canonical unit vectors \mathbf{e}_j , i.e. the columns of the relevant identity matrix, as extreme vertices.

2.2. Lagrangian conditions

Our problem is a quite simple example of Problem (1), as defined by Avriel (2003, p. 28). In our notation this defines optimisation (minimisation in Avriel, maximisation in our case) of a criterion $\phi(\mathbf{p})$, subject to one set of functions $\mathbf{g}(\mathbf{p})$ being non-negative, and also to a second set of functions $\mathbf{h}(\mathbf{p})$ being zero. In our problem constraints are $\mathbf{p} \geq \mathbf{0}$, $\sum_{i=1}^J p_i = 1$. So $\mathbf{g}(\mathbf{p}) = \mathbf{p}$ (identity functions), while $\mathbf{h}(\mathbf{p})$ is the scalar function $h(\mathbf{p}) = \sum_{i=1}^J p_i - 1$. Note that these are linear and therefore have constant first derivatives (values of 0 or 1 in fact) and zero second derivatives / Hessian matrices.

Following Avriel (2003, p. 32), the Lagrangian for our problem is

$$L(\mathbf{p}) = \lambda_0 \phi(\mathbf{p}) - \boldsymbol{\lambda}^\top \mathbf{p} - \mu \left(\sum_{i=1}^J p_i - 1 \right). \quad (2)$$

Subsequently λ_0 is set to a value of 1.

Sufficient first and second order conditions are set out in Theorem 3.11 of Avriel (2003).

First order conditions:

These conditions are derived from Avriel (2003, Conditions 3.108 – 3.110, p. 48).

They depend linearly on first derivatives of all functions. In our case, for a maximum, with

$\mathbf{d} = \partial\phi/\partial\mathbf{p}$, these are

$$\mathbf{d} - \lambda - \mu\mathbf{1} = \mathbf{0}, \quad (3)$$

$$\lambda_i p_i = 0, \quad (4)$$

$$\lambda \leq \mathbf{0}. \quad (5)$$

We will shed light on these later in Section 2.3. Of note is that, given (4), $\lambda_i = 0$ if $p_i > 0$. Otherwise $p_i = 0$ and either $\lambda_i = 0$ or $\lambda_i < 0$. Conditions (7) and (8) below distinguish between these two cases.

Second order conditions:

These derive from Avriel (2003, Definition 3.107, p. 48) (which in turn relies on Definition 3.5, p. 28) and Condition 3.111 of Avriel (2003, p. 49). The latter depends on quadratic forms of the Hessians of all functions. So given the linearity of $\mathbf{g}(\mathbf{p})$ and $\mathbf{h}(\mathbf{p})$, with zero Hessians, these reduce to

$$\mathbf{z}^\top \mathbf{H}(\mathbf{p}) \mathbf{z} < 0, \quad (6)$$

where \mathbf{H} is the Hessian of $\phi(\mathbf{p})$ and the vector \mathbf{z} must satisfy three conditions

$$z_i = 0 \text{ if } p_i = 0 \text{ and } \lambda_i < 0, \quad (7)$$

$$z_i \geq 0 \text{ if } p_i = 0 \text{ and } \lambda_i = 0, \quad (8)$$

$$\sum_{i=1}^J z_i = 0. \quad (9)$$

In the following we check for (6) over the wider class of \mathbf{z} satisfying (7) and (9), while we refer to (8) in Section 2.5. Given (7), the quadratic form in (6) depends only on the terms of the submatrix of $\mathbf{H}(\mathbf{p})$ and the subvector of \mathbf{z} which remain after terms corresponding to condition (7) are removed. Denote the corresponding sub-terms as \mathbf{p}_+^* , \mathbf{z}^* , and $\mathbf{H}(\mathbf{p}_+^*)$, and their orders as $J^* \times 1$, $J^* \times 1$ and $J^* \times J^*$ respectively. We require

$$\mathbf{z}^{*\top} \mathbf{H}(\mathbf{p}_+^*) \mathbf{z}^* < 0, \quad (10)$$

with \mathbf{z}^* satisfying (8) and (9). Clearly (10) is true for any \mathbf{z}^* if $\mathbf{H}(\mathbf{p}_+^*)$ is negative definite; a maximally sufficient condition.

However (9) implies that $\mathbf{z}^* = \mathbf{B}\mathbf{u}^*$, where $\mathbf{B} = (\mathbf{I} : -\mathbf{1})^\top$, with \mathbf{I} being the identity matrix of order $(J^* - 1)$ and $\mathbf{1}$ is a constant vector of length $(J^* - 1)$, all components equalling 1. So \mathbf{B} is of order $J^* \times (J^* - 1)$. This corresponds to choosing free values for the first $(J^* - 1)$ of the components of \mathbf{z}^* and setting the final one equal to minus

their sum. Further this in turn corresponds to removing the summation to one constraint on \mathbf{p}_+^* by substituting for its final component in terms of the others; i.e. assuming (briefly) that $\mathbf{p}_+^* = (p_1^*, \dots, p_{j^*}^*)$, $p_{j^*}^* = 1 - (p_1^* + \dots + p_{j^*-1}^*)$. We opt for this choice in our examples. The elements of the square matrix $\mathbf{B}^\top \mathbf{H}(\mathbf{p}_+^*) \mathbf{B}$ of order $(J^* - 1)$ are of the form: $d_{rs}^* - d_{rj^*}^* - d_{j^*s}^* + d_{j^*j^*}^*$, where $d_{rs}^* = \partial^2 \phi / \partial p_r^* \partial p_s^*$.

Going further we could have $\mathbf{z}^* = \mathbf{C} \mathbf{B} \mathbf{u}^*$, where the columns of \mathbf{C} are a permutation of the identity matrix of order J^* . (Note that $\mathbf{C} \mathbf{C}^\top = \mathbf{C}^\top \mathbf{C} = \mathbf{I}$.) Each choice of $\mathbf{C} \mathbf{B}$ corresponds to choosing free values for some set of $(J^* - 1)$ of the components of \mathbf{z}^* and setting the remaining one equal to minus their sum. This in turn corresponds to removing the summation to one constraint on \mathbf{p}_+^* by substituting for one of its components in terms of the others as above. The choice should not matter. To illustrate this, suppose that $\mathbf{H}(\mathbf{p}_+^*)$ has orthogonal decomposition $\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top$ and let $\mathbf{Q}^* = \mathbf{C}^\top \mathbf{Q}$. Then $\mathbf{G} = \mathbf{C}^\top \mathbf{H}(\mathbf{p}_+^*) \mathbf{C} = \mathbf{C}^\top \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top \mathbf{C} = \mathbf{Q}^* \mathbf{\Lambda} \mathbf{Q}^{*\top}$.

Now the columns of \mathbf{Q}^* are permutations of the columns of \mathbf{Q} , and hence are normalised and orthogonal. Thus $\mathbf{Q}^* \mathbf{\Lambda} \mathbf{Q}^{*\top}$ is an orthogonal decomposition of \mathbf{G} . Further, with $\mathbf{R} = \mathbf{Q}^{*\top} \mathbf{B}$, $\mathbf{S} = \mathbf{B}^\top \mathbf{C}^\top \mathbf{H}(\mathbf{p}_+^*) \mathbf{C} \mathbf{B} = \mathbf{B}^\top \mathbf{G} \mathbf{B} = \mathbf{B}^\top \mathbf{Q}^* \mathbf{\Lambda} \mathbf{Q}^{*\top} \mathbf{B} = \mathbf{R} \mathbf{\Lambda} \mathbf{R}^\top$. This is a non-standard decomposition of \mathbf{S} .

Now suppose that $\mathbf{H}(\mathbf{p}_+^*)$ has a single positive eigenvalue corresponding to a constant vector \mathbf{v} of length J^* ; i.e. $\mathbf{v} = k\mathbf{1}$. Then $\mathbf{C}^\top \mathbf{v} = \mathbf{v} = k\mathbf{1}$, and, crucially, $\mathbf{B}^\top \mathbf{v} = \mathbf{B}^\top \mathbf{C}^\top \mathbf{v} = \mathbf{0}$.

Hence $\mathbf{S} = \mathbf{B}^\top \mathbf{Q}^* \mathbf{\Lambda} \mathbf{Q}^{*\top} \mathbf{B} = \mathbf{B}^\top \mathbf{Q}^{**} \mathbf{\Lambda}^{**} \mathbf{Q}^{**\top} \mathbf{B}$, where \mathbf{Q}^{**} is \mathbf{Q}^* reduced by the exclusion of \mathbf{v} and $\mathbf{\Lambda}^{**}$ is $\mathbf{\Lambda}$ with the exclusion of the positive eigenvalue. Further $\mathbf{S} = \mathbf{R}^{**} \mathbf{\Lambda}^{**} \mathbf{R}^{**\top}$. Note that \mathbf{R}^{**} is a square matrix and $\mathbf{\Lambda}^{**}$ a diagonal matrix, each of order $(J^* - 1) \times (J^* - 1)$. This is a non-orthogonal decomposition of \mathbf{S} . However clearly $\mathbf{u}^{*\top} \mathbf{H}(\mathbf{p}_+^*) \mathbf{u}^* < 0$ if $\mathbf{\Lambda}^{**} < \mathbf{0}$. Further, if so, this would be true for all \mathbf{C} .

In general $\mathbf{B}^\top \mathbf{C}^\top \mathbf{H}(\mathbf{p}_+^*) \mathbf{C} \mathbf{B}$ is the Hessian of the criterion, in respect of the components of \mathbf{p}_+^* , which remain after the chosen substitution. Given $\mathbf{z}^* = \mathbf{B} \mathbf{u}^*$ condition (10) transforms to

$$\mathbf{u}^{*\top} \mathbf{B}^\top \mathbf{C}^\top \mathbf{H}(\mathbf{p}_+^*) \mathbf{C} \mathbf{B} \mathbf{u}^* < 0. \quad (11)$$

Clearly (11) is true for any \mathbf{u}^* if $\mathbf{C}^\top \mathbf{H}(\mathbf{p}_+^*) \mathbf{C}$ is negative definite; a minimally sufficient condition.

Note it is only necessary to consider the subvector \mathbf{p}_+^* when checking for second order conditions. Otherwise the full vector \mathbf{p} should be assumed to be under consideration.

2.3. Optimal design/directional derivative illumination of optimality conditions

Lagrangian theory can seem a bit mysterious. Indeed, as Whittle (1971) says on page 25. ‘... it is not surprising that people, who have known no other, should find the Lagrangian multiplier a mysterious notion to the end of their days.’!

However we can reveal a clear practical interpretation of first and second order conditions, and we need to do this to motivate one version of the algorithm we will use later. Consider the first order conditions (3) – (5). One implication of (3) is $\lambda = \mathbf{d} - \mu \mathbf{1}$. So $\lambda_i = d_i - \mu$, while (4) implies $\lambda_i = 0$ if $p_i > 0$. Let λ^+ , \mathbf{d}^+ and $\mathbf{1}^+$ correspond to the positive components of \mathbf{p} . Then $\mathbf{d}^+ = \mu \mathbf{1}^+$. That is, the components of \mathbf{d}^+ all share a common value μ . It is convenient to uncover a formula for μ for arbitrary \mathbf{d}^+ and hence \mathbf{d} , which will yield the common value under local optimality. There are many. One would be the average of the components of a current \mathbf{d}^+ . A more convenient one exploits the constraint $\sum_{i=1}^J p_i = 1$ and accommodates zeroes in \mathbf{p} ; namely $\mu = \sum_{i=1}^J p_i d_i$. (Another would be $\mu = \sum_{i=1}^J w_i d_i$ with $\sum_{i=1}^J w_i = 1$, subject to $w_j = 0$ if $p_j = 0$.) Opting for $\mu = \sum_{i=1}^J p_i d_i$, let

$$F_j = d_j - \sum_{i=1}^J p_i d_i.$$

(Note: If $\phi(\mathbf{p})$ is homogeneous of degree t , as most criteria of interest are, then $\sum_{i=1}^J p_i d_i = t\phi(\mathbf{p})$; so $F_j = d_j - t\phi(\mathbf{p})$. Often constraints on \mathbf{p} , such as those of Problem (1), are needed to ensure a finite optimiser \mathbf{p}^* .)

Then first order optimality conditions for a local maximum are

$$\begin{aligned} F_j &= 0 & \text{if } p_j > 0, \\ F_j &\leq 0 & \text{if } p_j = 0. \end{aligned} \tag{12}$$

Now $F_j = F_\phi(\mathbf{p}, \mathbf{e}_j)$, where \mathbf{e}_j is the j th (canonical) unit vertex, and $F_\phi(\mathbf{p}, \mathbf{q})$ is a (directional) derivative of $\phi(\cdot)$ at \mathbf{p} in the direction of \mathbf{q} , which is defined below.

Definition: Directional Derivative $F_\phi(\mathbf{p}, \mathbf{q})$

$$F_\phi(\mathbf{p}, \mathbf{q}) = \lim_{\varepsilon \downarrow 0} \frac{\phi\{(1 - \varepsilon)\mathbf{p} + \varepsilon\mathbf{q}\} - \phi(\mathbf{p})}{\varepsilon}.$$

If $\phi(\cdot)$ is differentiable at \mathbf{p} , $F_\phi(\mathbf{p}, \mathbf{q}) = (\mathbf{q} - \mathbf{p})^\top \mathbf{d}$. Given this,

$$F_\phi(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^J q_j (F_\phi(\mathbf{p}, \mathbf{e}_j)) = \sum_{j=1}^J q_j F_j.$$

In $F_\phi(\mathbf{p}, \mathbf{q})$ we have a rate of change in ϕ at \mathbf{p} in the direction of \mathbf{q} . It is an apt formula if the constraint region is convex. Then all points between \mathbf{p} and \mathbf{q} belong to that region if \mathbf{p} and \mathbf{q} do so. $F_\phi(\mathbf{p}, \mathbf{q})$ could be normalised as it depends on the distance between \mathbf{p} and \mathbf{q} . However, it is useful as is, if we are only focussing on its sign. So $F_j = F_\phi(\mathbf{p}, \mathbf{e}_j)$ is the directional derivative of $\phi(\cdot)$ at \mathbf{p} in the direction of the extreme vertex \mathbf{e}_j of the constraint region of our problem, namely $\{\mathbf{p} : p_j \geq 0, \sum_{j=1}^J p_j = 1\}$.

First order conditions (12), for a local maximum, are that F_j 's should be zero or negative. Note that, given $F_\phi(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^J q_j F_j$, it follows that $F_\phi(\mathbf{p}, \mathbf{q}) \leq 0$ for all \mathbf{q} . So, for all \mathbf{q} , $\phi(\cdot)$ is either decreasing at \mathbf{p} in the direction of \mathbf{q} , or it is stationary in that direction, as should be the case.

Further these conditions will be sufficient if $\phi(\cdot)$ is concave. This statement constitutes the general equivalence theorem of optimal design theory; see Whittle (1973), Kiefer (1974). It states that we need only to check derivatives at \mathbf{p} in the direction of the extreme points (\mathbf{e}_j) of our convex constraint region.

If concavity is lacking, we must check for second order conditions as outlined above. We can shed further light on these too. To our knowledge this has not been explored before. We propose defining a second order directional derivative of $\phi(\cdot)$ at \mathbf{p} in the direction of \mathbf{q} ; namely the directional derivative at \mathbf{p} in the direction of \mathbf{q} of $\psi(\mathbf{p}) = F_\phi(\mathbf{p}, \mathbf{q})$ itself. Denote this by $F_\phi^{(2)}(\mathbf{p}, \mathbf{q})$. Assuming differentiability, we can claim that $F_\phi(\mathbf{p}, \mathbf{q}) = (\mathbf{q} - \mathbf{p})^\top \mathbf{d}$, and in turn applying this to $\psi(\mathbf{p})$, we obtain that

$$F_\phi^{(2)}(\mathbf{p}, \mathbf{q}) = -F_\phi(\mathbf{p}, \mathbf{q}) + \mathbf{z}^\top \mathbf{H}(\mathbf{p})\mathbf{z}, \text{ with } \mathbf{z} = (\mathbf{q} - \mathbf{p}).$$

We wish that this be non-positive when $F_\phi(\mathbf{p}, \mathbf{q}) = 0$, in which case

$$F_\phi^{(2)}(\mathbf{p}, \mathbf{q}) = \mathbf{z}^\top \mathbf{H}(\mathbf{p})\mathbf{z}.$$

The restriction $F_\phi(\mathbf{p}, \mathbf{q}) = 0$ imposes restrictions on \mathbf{q} and hence on \mathbf{z} . These are as in Section 2.2. In particular, $p_j = 0$ implies $z_j = q_j - p_j = q_j = 0$, while we have $\sum_{i=1}^J z_i = 0$, given that $\sum_{i=1}^J p_i = \sum_{i=1}^J q_i = 1$. So \mathbf{p} , \mathbf{z} , $\mathbf{H}(\mathbf{p})$ are reduced to \mathbf{p}_+^* , \mathbf{z}^* , $\mathbf{H}(\mathbf{p}_+^*)$ as in Section 2.2 and for a local maximum, we require

$$\mathbf{z}^{*\top} \mathbf{H}(\mathbf{p}_+^*)\mathbf{z}^* < 0. \quad (13)$$

Thus we have uncovered conditions (6), (8) and (9) of Section 2.2.

Further we have given (expected) physical expression to the differential calculus conditions which must hold. These are that, if first and second order conditions are satisfied respectively at \mathbf{p}^* and \mathbf{p}_+^* , then we can claim that, for each \mathbf{q} in the convex feasible region,

$\phi(\cdot)$ is either decreasing at \mathbf{p}^* towards \mathbf{q} , or it is stationary in the direction of \mathbf{q} , but its first derivative at \mathbf{p}^* towards \mathbf{q} is decreasing (from zero). Thus at \mathbf{p}^* , $\phi(\cdot)$ is non-increasing at \mathbf{p}^* towards all feasible \mathbf{q} ; i.e. we have a local maximum at \mathbf{p}^* .

(Note: There would be parallels to the results of Sections 2.2 and 2.3 if the summation to one constraint were changed to general linear constraints $\mathbf{A}\mathbf{p} = \mathbf{b}$.)

2.4. Simple examples on first order conditions

In these examples $\mathbf{p} = (p_1, p_2, p_3, p_4)$, $(p_1 + p_2 + p_3 + p_4) = 1$. It is convenient to also let $\boldsymbol{\rho} = (p_1, p_2, p_3)$. We still have $d_j = \partial\phi/\partial p_j$ and $F_j = F_\phi(\mathbf{p}, \mathbf{e}_j)$, $j = 1, 2, 3, 4$.

Example 2.4.1

Let $\phi(\mathbf{p}) = -[\psi(\boldsymbol{\rho}) + p_4]$, $\psi(\boldsymbol{\rho}) = [\gamma(\boldsymbol{\rho})]^2$, $\gamma(\boldsymbol{\rho}) = [a_1/p_1 - (a_2/p_2 + a_3/p_3)]$. Thus $\phi(\mathbf{p}) = -\{[\gamma(\boldsymbol{\rho})]^2 + p_4\}$. This is a simple example of our main criterion of interest.

Suppose we wish to maximise $\phi(\mathbf{p})$ subject to $p_i \geq 0$ and $(p_1 + p_2 + p_3 + p_4) = 1$. Under these constraints $\phi(\mathbf{p})$ can obtain a global maximum of zero. There are many choices of \mathbf{p} attaining this, each with $\gamma(\boldsymbol{\rho}) = 0 = p_4$. Hence we have a global maximum of $[\gamma(\boldsymbol{\rho})]^2$ and of $[-p_4]$ too.

One solution is $(p_1, p_2, p_3) = (a_1, 2a_2, 2a_3)/D$, $D = (a_1 + 2a_2 + 2a_3)$.

Given $d_j = \gamma(\boldsymbol{\rho})[a_j/p_j^2]$, $j = 1, 2, 3$, while $d_4 = -1$, it follows that, at all these optima, $d_j = 0$, $j = 1, 2, 3$, while $d_4 = -1$. Hence $F_j = 0$, $j = 1, 2, 3$, while $F_4 = -1 < 0$, confirming first order conditions.

We could view the maximisation process in two steps. First we maximise $(-p_4)$ subject to $0 \leq p_4 \leq 1$, yielding $p_4 = 0$. Then we maximise $[\gamma(\boldsymbol{\rho})]^2$ subject to $p_i \geq 0$ and $(p_1 + p_2 + p_3) = 1$, yielding a second example of Problem (1). We note the following two points.

(i) Since we have found a global maximum of $[\gamma(\boldsymbol{\rho})]^2$, we must have $d_j = 0$, $j = 1, 2, 3$.

(ii) We are maximising a criterion of three variables over the triangle whose vertices are the unit vertices on the positive axes of three dimensional space. For any \mathbf{p} , the F_j , $j = 1, 2, 3$, are the directional derivatives towards these vertices. The class of optima form a quadratic ridge over the triangle defined by

$$a_1 p_2 p_3 = a_2 p_1 p_3 + a_3 p_1 p_2.$$

At each one of the points on this ridge $F_j = 0$, $j = 1, 2, 3$.

Example 2.4.2

A similar example has $\phi(\mathbf{p}) = -[\gamma(\boldsymbol{\rho}) + p_4]$, $\gamma(\boldsymbol{\rho}) = [c_1^2/p_1 + c_2^2/p_2 + c_3^2/p_3]$, $c_j > 0$, $j = 1, 2, 3$. The function $\gamma(\boldsymbol{\rho})$ is a simple example of the c -optimal criterion (Atkinson,

Donev & Tobias (2007). It is a variance, which we wish to minimise. It is convex. There is a unique globally optimal explicit solution to this problem; namely, $p_j = c_j/D$, $D = (c_1 + c_2 + c_3)$. At this $d_j = -[c_j/p_j]^2$, $j = 1, 2, 3$ have a common value of $-D^2$. Thus $F_j = 0$, $j = 1, 2, 3$. Further with $p_4 = 0$, $d_4 = -1$ and, for $\phi(\mathbf{p})$, $d_j = [c_j/p_j]^2$, we have $F_j = 0$, $j = 1, 2, 3$, $F_4 < 0$.

Examples 2.4.1 and 2.4.2

The criteria $\phi(\mathbf{p})$, in both of these, is of the form, $\phi(\mathbf{p}) = -\{[\gamma(\boldsymbol{\rho})]^k + p_4\}$, where $\gamma(\boldsymbol{\rho})$ is linear in the reciprocals of p_1, p_2, p_3 , and $k > 0$. Hence $\{[\gamma(\boldsymbol{\rho})]^k\}$ is homogeneous of degree $(-k)$. We could approach both problems by maximising $\{-[\gamma(\boldsymbol{\rho})]^k\}$ subject to $(p_1 + p_2 + p_3) = (1 - p_4)$ (denoting the maximising \mathbf{p} , by \mathbf{p}^{**} , and the maximum by $C > 0$), and then maximising the resultant optimum over $0 \leq p_4 \leq 1$.

At the first stage, given the homogeneity of $[\gamma(\boldsymbol{\rho})]^k$, $p_j^{**} = (1 - p_4^*)p_j^*$, where p_j^* is the optimiser of $\{-[\gamma(\boldsymbol{\rho})]^k\}$ subject to $(p_1 + p_2 + p_3) = 1$.

Then $\phi(\mathbf{p})$ is of the form $\phi(\mathbf{p}) = -\{p_4 + [C/(1 - p_4)]^k\}$. Given $k > 0$, this decreases over $0 \leq p_4 \leq 1$ and is therefore maximised at $p_4 = 0$.

We now explore examples of Problem (1) with more structure yielding further geometrical characterisations of first order conditions.

2.5. Optimisation over bounded convex sets

One class of examples of Problem (1) derives from the following problem.

Maximise $\psi(\mathbf{x})$ subject to \mathbf{x} being a convex combination of extreme points \mathbf{v}_j , $j = 1, 2, \dots, J$; i.e.

$$\mathbf{x} = \mathbf{x}(\mathbf{p}) = \sum_{i=1}^J p_i \mathbf{v}_i, p_i \geq 0, \sum_{i=1}^J p_i = 1.$$

So solve Problem (1) with $\phi(\mathbf{p}) = \psi\{\mathbf{x}(\mathbf{p})\}$. Now directional derivatives of $\phi(\cdot)$ are expressed in terms of those of $\psi(\cdot)$. In particular, $F_\phi(\mathbf{p}, \mathbf{q}) = F_\psi(\mathbf{x}(\mathbf{p}), \mathbf{x}(\mathbf{q}))$; $F_j = F_\phi(\mathbf{p}, \mathbf{e}_j) = F_\psi(\mathbf{x}(\mathbf{p}), \mathbf{v}_j)$. Still necessary conditions are $F_j = 0$ if $p_j > 0$, $F_j \leq 0$ if $p_j = 0$.

One example of this problem arises when $\mathbf{x}, \mathbf{v}_j \in \mathbb{R}^2$ and the feasible region is a quadrilateral \mathcal{Q} with vertices $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) = (A, B, C, D)$, edges AB, BC, CD, DA and diagonals AC, BD . Suppose $\mathbf{x}^* = \mathbf{x}^*(\mathbf{p}^*)$ is a potential local maximiser. For first order conditions we must check the four vertex directional derivatives at \mathbf{x}^* towards A, B, C, D ; i.e. $F_j = F_\psi(\mathbf{x}^*, \mathbf{v}_j)$, $j = 1, 2, 3, 4$. There are three main possibilities, namely

- (i) \mathbf{x}^* lies at a vertex, say A , in which case we must have $F_j \leq 0$, $j = 2, 3, 4$;

- (ii) \mathbf{x}^* lies on an edge, say AB ; so it is a convex combination of only \mathbf{v}_1 and \mathbf{v}_2 , in which case we must have $F_j = 0$, $j = 1, 2$, $F_j \leq 0$, $j = 3, 4$;
- (iii) \mathbf{x}^* is an internal point of \mathcal{Q} , in which case it can be expressed as a convex combination of all four vertices, all convex weights being positive. So we must have $F_j = 0$, $j = 1, 2, 3, 4$.

However this latter case sub-divides into three sub-possibilities, namely

- (i) \mathbf{x}^* lies within two triangles, each formed by three of the vertices. So it can be expressed as a convex combination of just three vertices (with two choices for these), with the remaining vertex zero-weighted;
- (ii) \mathbf{x}^* lies within one such triangle and on one diagonal; the latter implies that it can be expressed as a convex combination of just two vertices, with the remaining two zero-weighted;
- (iii) \mathbf{x}^* is the ‘centroid’ of \mathcal{Q} ; i.e. it lies on both ‘diagonals’. So it can be expressed as a convex combination of just two vertices (with two choices for these), with the remaining two vertices zero-weighted.

This is in keeping with Caratheodory’s Theorem which, since \mathbf{x}^* lies in two-dimensional space, asserts that \mathbf{x}^* can be expressed as a convex combination of a set of at most three vertices. Here we have two choices for these sets. With the inclusion of zero weights, denote these two sets of weights by \mathbf{r} and \mathbf{s} . So $\mathbf{x}^* = \mathbf{x}(\mathbf{r}) = \mathbf{x}(\mathbf{s})$. It follows then that $\mathbf{x}^* = \mathbf{x}(\alpha\mathbf{r} + (1 - \alpha)\mathbf{s})$ for $0 \leq \alpha \leq 1$. Now for $0 < \alpha < 1$, $\alpha\mathbf{r} + (1 - \alpha)\mathbf{s}$ has four positive components. This reveals that for a given \mathbf{x}^* , there can be many solutions \mathbf{p}^* to the equation $\mathbf{x}^* = \mathbf{x}^*(\mathbf{p}^*)$, with some \mathbf{p}^* having zero components. However, at all of these we must have zero vertex directional derivatives, whether \mathbf{p}^* contains zero components or not. This yields examples of the scenario envisaged under condition (8) of Section 2.2.

2.6. Optimal design summary

We now return to our primary area of interest; namely optimal design problems. These are extensions of the examples of Section 2.5 under which \mathbf{x} takes the form of a symmetric matrix \mathbf{M} , constrained to be a convex combination of rank one matrices of the form $\mathbf{v}\mathbf{v}^\top$. So

$$\phi(\mathbf{p}) = \psi\{\mathbf{M}(\mathbf{p})\}, \mathbf{M}(\mathbf{p}) = \sum_{j=1}^J p_j \mathbf{v}_j \mathbf{v}_j^\top,$$

while

$$F_\phi(\mathbf{p}, \mathbf{q}) = F_\psi(\mathbf{M}(\mathbf{p}), \mathbf{M}(\mathbf{q})); F_j = F_\phi(\mathbf{p}, \mathbf{e}_j) = F_\psi(\mathbf{M}(\mathbf{p}), \mathbf{v}_j \mathbf{v}_j^\top).$$

Note a summary of the approach we will use to solve Problem (1) is as follows:

1. Apply an iterative numerical algorithm (see equation 21 in Section 4), the successive approximate solutions of which have at least the potential of converging to a candidate local maximum \mathbf{p}^* .
2. Check that the limit point \mathbf{p}^* is a reasonable candidate in that it is in a certain sense a critical point, i.e. that it satisfies the first order conditions (12).
3. Given that the limit point satisfies the first order conditions check that it satisfies a second order condition for constituting a local maximum. This condition is expressed in terms of an associated Hessian being negative definite (13). So all of its eigenvalues must be negative. Zero values would yield inconclusive results.

3. Designing to minimise covariances

We consider the problem of constructing designs which ensure that the estimators of two or more components or linear combinations of components of $\boldsymbol{\theta}$ are uncorrelated or as uncorrelated as possible. To find such designs we minimise the squared covariances between the estimators of parameters of interest, or equivalently we maximise negative of the squared covariances.

3.1. Case of two linear combinations of $\boldsymbol{\theta}$

Denote the two linear combinations of interest by $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$, where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$. We wish to find a design such that the covariance of the estimators of $\mathbf{a}^\top \boldsymbol{\theta}$ and $\mathbf{b}^\top \boldsymbol{\theta}$ is as close to zero as possible.

As an example of Problem (1) we wish to maximise $\phi(\mathbf{p}) = -[\text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}})]^2 = -[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2$ over $\mathcal{P} = \{\mathbf{p} = (p_1, p_2, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1\}$. If the maximum value of zero is attained, the estimators of the two linear combinations of the parameters are uncorrelated. There may not exist any such design. Nevertheless maximising the foregoing criterion will yield a design with minimum numerical covariance, in which case the estimators of the two linear combinations are as near to being uncorrelated as possible.

Let us denote the covariance between $\mathbf{a}^\top \hat{\boldsymbol{\theta}}$ and $\mathbf{b}^\top \hat{\boldsymbol{\theta}}$ by $h(\mathbf{p})$. That is, $h(\mathbf{p}) = \mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}$. Note that $h(\mathbf{p})$ has a similar structure to the c -optimal design criterion. This criterion function is $\phi_c(\mathbf{p}) = \mathbf{c}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{c}$, where \mathbf{c} is a known vector of constants

(Atkinson, Donev & Tobias 2007). Minimisation of $\phi_c(\mathbf{p})$ leads to a c -optimal design. This criterion was used in Example 2.4.2 of Section 2.

The c -optimal criterion is a special case of linear optimality according to which the criterion function $\phi_L(\mathbf{p}) = \text{tr}\{\mathbf{M}^{-1}(\mathbf{p})\mathbf{L}\}$ is minimised, where \mathbf{L} is a $k \times k$ matrix of coefficients. ‘‘Linear optimality’’ criteria are linear in the entries of the covariance matrix $\mathbf{M}^{-1}(\mathbf{p})$. In the context of c -optimality, our interest is in choosing a design so as to estimate the linear function of $\boldsymbol{\theta}$, $\mathbf{c}^\top \boldsymbol{\theta}$, with minimum variance. It is known that the criterion $\phi_c(\mathbf{p})$ is a convex function on \mathbb{M} , where \mathbb{M} is the set of all positive definite symmetric matrices (Fedorov 1972). However, the covariance function $h(\mathbf{p})$ is not a convex function on \mathbb{M} . On the other hand, similar to these criteria, the function $h(\mathbf{p})$ is linear in the entries of the covariance matrix $\mathbf{M}^{-1}(\mathbf{p})$.

Since $\phi(\mathbf{p}) = -h^2(\mathbf{p})$, the partial derivatives of $\phi(\mathbf{p})$ with respect to the p_j are $d_j^\phi = -2h(\mathbf{p})d_j^h$, where the d_j^h are the corresponding partial derivatives of $h(\mathbf{p})$. Accordingly the second order partial derivatives of $\phi(\mathbf{p})$ with respect to p_i and p_j are $d_{ij}^\phi = -2d_i^h d_j^h - 2h(\mathbf{p})d_{ij}^h$. In order to run the algorithms so as to find the optimal design and to check conditions for optimality (as established in Section 2) we need to find the first and second order partial derivatives d_j^h and d_{ij}^h respectively. Derivatives of $h(\mathbf{p})$ are provided by the following theorem.

Theorem 3.1. For $h(\mathbf{p}) = \text{cov}(\mathbf{a}^\top \hat{\boldsymbol{\theta}}, \mathbf{b}^\top \hat{\boldsymbol{\theta}})$, $d_j^h, d_{ij}^h, i, j = 1, 2, \dots, J$, are given by

$$\begin{aligned} d_j^h &= -\mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j, \\ d_{ij}^h &= \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i + \mathbf{v}_i^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{a} \mathbf{b}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_j \mathbf{v}_j^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{v}_i, \end{aligned}$$

where $\mathbf{v}_j = \eta(\mathbf{x}_j)$ are the points in the induced design space as defined in Section 2.

This theorem is easily proved by noting that $\mathbf{M}(\mathbf{p})\mathbf{M}^{-1}(\mathbf{p}) = \mathbf{I}$ (where \mathbf{I} is the $k \times k$ identity matrix), differentiating the product on the left hand side appropriately and solving the resulting equation for the derivative of $\mathbf{M}^{-1}(\mathbf{p})$.

As we mentioned in Section 2, our criterion function is not concave and (given that first order conditions are met) we will check for second order conditions by dealing with our summation constraint via the substitution $p_J = 1 - (p_1 + p_2 + \dots + p_{J-1})$, where $J = J^*$ as defined in Section 2. Let the criterion function $\phi(\mathbf{p})$, in terms of p_1, p_2, \dots, p_{J-1} be $\tilde{\phi}(\mathbf{p})$. We will check that standard conditions on the Hessian matrix of $\tilde{\phi}(\mathbf{p})$ are satisfied. The first and second order derivatives of $\tilde{\phi}(\mathbf{p})$ are given by

$$d_j^{\tilde{\phi}} = d_j^\phi - d_{Jj}^\phi, \quad j = 1, 2, \dots, J-1, \quad (14)$$

$$d_{ij}^{\tilde{\phi}} = d_{ij}^\phi - d_{iJ}^\phi - d_{Jj}^\phi + d_{JJ}^\phi, \quad i, j = 1, 2, \dots, J-1. \quad (15)$$

The first partial derivatives at a critical point of $\tilde{\phi}(\mathbf{p})$ are zero, and hence so are the vertex directional derivatives of $\phi(\mathbf{p})$ as discussed in Section 2. The Hessian matrix of $\tilde{\phi}(\mathbf{p})$ is given by $\mathbf{H}(\mathbf{p}) = ((d_{ij}^{\tilde{\phi}}))$, $i, j = 1, 2, \dots, J - 1$.

First, suppose $J - 1 = 2$. Then, \mathbf{p}^* is a local maximum of $\tilde{\phi}(\mathbf{p})$ if $\det(\mathbf{H}(\mathbf{p}_+^*)) > 0$ and $d_{ii}^{\tilde{\phi}}(\mathbf{p}_+^*) < 0$, $i = 1, 2$. This condition applies to Examples 1 and 2 of Section 5.

Now suppose that $\tilde{\phi}(\mathbf{p})$ is a function of more than two p_i 's (i.e., $J - 1 > 2$). Then, \mathbf{p}^* is a local maximum of $\tilde{\phi}(\mathbf{p})$ if the Hessian matrix at \mathbf{p}_+^* is negative definite (as discussed in Section 2), that is, if the eigenvalues of $\mathbf{H}(\mathbf{p}_+^*)$ are negative. This condition applies to Example 3 of Section 5.

Note that, as we mentioned in Section 2, we apply the Hessian test to the positive candidate points \mathbf{p}_+^* found by using algorithm (21). We describe the detailed algorithmic approach in Section 4.

A simple example: We now illustrate the zero covariance case, by a simple example in the context of quadratic regression model

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

The information matrix is

$$\mathbf{M}(\mathbf{p}) = \mathbf{V}\mathbf{P}\mathbf{V}^\top = \sum_{j=1}^J p_j (1 \ x_j \ x_j^2)^\top (1 \ x_j \ x_j^2), \quad (16)$$

where $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_J)$ and $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_J]$.

Suppose we are interested in estimating the parameters θ_2 and θ_3 so that the corresponding parameter estimators are uncorrelated. Now $\text{cov}(\hat{\theta}_2, \hat{\theta}_3) = -\det(\mathbf{M}_{23})/\det(\mathbf{M}(\mathbf{p}))$ where \mathbf{M}_{23} is the 2×2 matrix

$$\mathbf{M}_{23} = \sum_{j=1}^J p_j (1 \ x_j^2)^\top (1 \ x_j). \quad (17)$$

The matrix \mathbf{M}_{23} is the submatrix of the information matrix $\mathbf{M}(\mathbf{p})$ obtained by deleting row 2 and column 3. The covariance $\text{cov}(\hat{\theta}_2, \hat{\theta}_3)$ would be zero if $\det(\mathbf{M}_{23})$ were zero. From (17), $\det(\mathbf{M}_{23})$ is a polynomial of degree 2 in the design weights \mathbf{p} .

Let us consider the case $J = 3$ and suppose the support points of our probability measure are $\{-1, 0, 2\}$. Then the matrix \mathbf{M}_{23} will be

$$\mathbf{M}_{23} = \begin{bmatrix} 1 & -p_1 + 2p_3 \\ p_1 + 4p_3 & -p_1 + 8p_3 \end{bmatrix}.$$

Let us denote $\det(\mathbf{M}_{23})$ by $\tilde{h}(\mathbf{p})$. Thus $\text{cov}(\hat{\theta}_2, \hat{\theta}_3)$ will be zero if

$$\tilde{h}(\mathbf{p}) = p_1^2 - 8p_3^2 - p_1 + 8p_3 + 2p_1p_3 = (p_1 + p_3 - 1/2)^2 - 9(p_3 - 1/2)^2 + 2 = 0.$$

If $p_3 = 1/2$, then $\tilde{h}(\mathbf{p}) = p_1^2 + 2 = 0$, which is impossible. If $p_1 + p_3 = 1/2$, then $\tilde{h}(\mathbf{p}) = -9(p_3 - 1/2)^2 + 2$, which is equal to zero if $p_3 = 1/2 \pm \sqrt{2}/3$, that is, $p_3 = 0.971405$ or 0.028595 . But if $p_1 + p_3 = 1/2$ the only possible value of p_3 is 0.028595 . Hence one solution, in which the estimators of the parameters θ_2 and θ_3 are uncorrelated, is $(0.471, 0.5, 0.029)$.

For different sets of support points it may be possible to obtain a solution by following a similar procedure to that given above. However in general this sort of approach is not a practical one.

Instead of trying to make the covariance between the estimators of two linear combinations of parameters equal (or as close as possible) to zero, we might wish to make this covariance equal to some pre-specified constant, say c_0 . This objective is readily achieved by adjusting the objective function that we wish to optimise to be

$$\phi(\mathbf{p}) = -[h(\mathbf{p}) - c_0]^2.$$

This idea can obviously be extended to settings in which we wish to deal with more than two linear combinations, as discussed in Section 3.2.

3.2. Case of more than two linear combinations of θ

Suppose that we are interested in making a parameter estimator uncorrelated with another *two* parameter estimators in a regression model, or more generally that we wish to make two pairs of parameter estimators uncorrelated (or as near to being uncorrelated as possible) with each other. So consider four linear combinations $\mathbf{a}^\top \theta$, $\mathbf{b}^\top \theta$, $\mathbf{c}^\top \theta$ and $\mathbf{d}^\top \theta$, where $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^k$. Our objective is to minimise the magnitude of the covariance between the estimators of $\mathbf{a}^\top \theta$ and $\mathbf{b}^\top \theta$ and between the estimators of $\mathbf{c}^\top \theta$ and $\mathbf{d}^\top \theta$.

The two covariances whose magnitude we wish to minimise are thus

$$h_1(\mathbf{p}) = \text{cov}(\mathbf{a}^\top \hat{\theta}, \mathbf{b}^\top \hat{\theta}) \text{ and } h_2(\mathbf{p}) = \text{cov}(\mathbf{c}^\top \hat{\theta}, \mathbf{d}^\top \hat{\theta}). \quad (18)$$

In order to solve the problem of simultaneously minimising the magnitude of these two covariances we set $\phi_i(\mathbf{p}) = -h_i(\mathbf{p})^2$, $i = 1, 2$, and create a compound criterion involving both $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. The compound criterion must be chosen in such a way that if there exists a \mathbf{p}^* which simultaneously maximises the two functions $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$ then this \mathbf{p}^* should (uniquely) maximise the compound criterion. Ideally we want the \mathbf{p}^* to produce

a common value of zero for $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. The use of a compound criterion transforms the problem of maximising $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$ simultaneously into a problem of the form of (1). The optimum of the compound criterion must of course satisfy constraints $p_j \geq 0$ and $\sum_{j=1}^J p_j = 1$.

There are a number of compound criteria that can be used; two such are

$$\phi(\mathbf{p}) = \phi_1(\mathbf{p}) + \phi_2(\mathbf{p}), \quad (19)$$

$$\phi(\mathbf{p}) = \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\}. \quad (20)$$

Note that in maximising any of these compound criteria we must consider two sets of partial derivatives d_j corresponding respectively to the two functions $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. Given the vectors \mathbf{a} , \mathbf{b} , \mathbf{c} and \mathbf{d} , these derivatives can be found using Theorem 3.1.

As mentioned before, ideally we want a \mathbf{p}^* to produce a common value of zero for $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$. Assuming this to be possible, each of the criterion (19) and (20) will have a maximum of zero.

Numerical techniques are needed to determine the optimising \mathbf{p}^* . In Section 4 we consider an appropriate algorithm to solve the foregoing optimisation problems.

4. Algorithms

It is usually not possible to calculate an explicit solution to Problem (1). We will use the following algorithm to search for an optimising \mathbf{p}^* . If we find a \mathbf{p} for which $\phi(\mathbf{p}) = 0$ we have found one. Otherwise, if the algorithm has converged to \mathbf{p} , then, since our criterion is not concave, we must check for first and second order conditions outlined in Section 2, to confirm that at least a local maximum has been found. An iterative procedure which has the potential to converge to a solution is given by

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(x_j^{(r)}, \delta)}{\sum_{j=1}^J p_j^{(r)} f(x_j^{(r)}, \delta)}, \quad (21)$$

where the $x_j^{(r)}$ may be the partial derivatives $d_j^{(r)}$ of $\phi(\cdot)$ with respect to p_j , or the corresponding directional derivatives $F_j^{(r)}$, evaluated at $\mathbf{p}^{(r)}$.

The function $f(\cdot, \cdot)$ must be positive and must be a strictly increasing function of its first argument x while $f(x, 0) = \text{constant}$. The second argument δ is a positive free ‘‘tuning’’ parameter. Torsney (1977) first proposed this type of algorithm, taking $f(x, \delta) = x^\delta$ which requires $x \geq 0$. For related discussion of this algorithm, see Torsney (1983, 1988), Silvey, Titterington & Torsney (1978), Titterington (1976), Mandal & Torsney (2000), Torsney &

Mandal (2001), Torsney & Alahmadi (1995), and Mandal, Torsney & Carriere (2005). For further developments based on clustering, see Mandal & Torsney (2006).

One clear property of algorithm (21) is that an iterate $\mathbf{p}^{(r)}$ is always feasible. This can be easily seen by noting the following. We always start with an initial design (usually with equal weights \mathbf{p}). The function $f(\cdot, \cdot)$ is always positive. Moreover, the summation of the right hand side of (21) for $j = 1, 2, \dots, J$ is always unity. That is, an iterate $\mathbf{p}^{(r)}$ is always feasible. A second property is that the criterion is increasing at $\mathbf{p}^{(r)}$ in the direction of $\mathbf{p}^{(r+1)}$ for positive δ ; i.e., $F_\phi(\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)}) \geq 0$, where $F_\phi(\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)})$ is the directional derivative of ϕ at the current iteration $\mathbf{p}^{(r)}$ in the direction of the next iteration $\mathbf{p}^{(r+1)}$. This result can be proved by noting that, after some simplification, the directional derivative can be written as: $F_\phi(\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)}) = \text{cov}(F, f(F, \delta))/E(f(F, \delta))$, where F is a random variable taking values in the set of vertex directional derivative values with distribution given by $\text{Pr}(F = F_j^{(r)}) = p_j^{(r)}$. Finally if, at $\mathbf{p}^{(r)} = \mathbf{p}^*$ (for some \mathbf{p}^*), $F_j^{(r)} = 0$, then $\mathbf{p}^{(r+1)} = \mathbf{p}^{(r)} = \mathbf{p}^*$, so that $F_\phi(\mathbf{p}^{(r)}, \mathbf{p}^{(r+1)}) = 0$ and \mathbf{p}^* is a fixed point of the algorithm. Also, given that $\mathbf{p}^{(r+1)} = \mathbf{p}^{(r)}$ if $\delta = 0$, iterations should be monotonic for small positive δ . Roughly speaking the algorithm increases those p_j which should be increased and decreases the others; equivalently it increases those p_j for which $F_j > 0$, and decreases those for which $F_j < 0$. So F_j is a good choice and can have a beneficial effect on convergence.

We explore use of algorithm (21), with a suitable choice of $f(\cdot, \cdot)$ and its arguments to solve our optimisation problem. These choices play an important role in the convergence of the algorithm; see Torsney (1977, 1988). Some possible choices of $f(x, \delta)$ include x^δ ($x \geq 0$) and $\exp(\delta x)$. Consider the case $x_j = F_j$ and recall that $\sum_{j=1}^J p_j F_j = 0$ from Section 2. This condition suggests that a suitable function should be centred on zero (in respect of its first argument “ x ”) and should change quickly about $x = 0$. Two appropriate choices of $f(x, \delta)$ are $f(x, \delta) = \Phi(\delta x)$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution, $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, i.e. the logistic c.d.f. evaluated at δx . We run the algorithm until the first order conditions (12) are satisfied, and then we need to check for the second order conditions (13) as discussed in Section 2.

Some further points are given in the following.

In the case of minimising two covariances, the choice of x depends on which of the compound criteria given in (19), (20) is used. One possible set of choices is as follows

$$x_j = d_j^{\phi_1} + d_j^{\phi_2} \text{ if we maximise } \phi_1(\mathbf{p}) + \phi_2(\mathbf{p}) \text{ (i.e. criterion (19)),} \quad (22)$$

$$x_j = d_j^{\min\{\phi_1, \phi_2\}} \text{ if we maximise } \min\{\phi_1(\mathbf{p}), \phi_2(\mathbf{p})\} \text{ (i.e. criterion (20)),} \quad (23)$$

where $d_j^{\phi_i}$ denotes the partial derivative of ϕ_i with respect to p_j , $i = 1, 2$, $j = 1, 2, \dots, J$. Here also we improve the rate of convergence of the algorithm by replacing $d_j^{\phi_i}$ by the corresponding directional derivatives.

In addition to depending on the choice of $f(\cdot, \cdot)$ and of its first argument, convergence rates also depend on the parameter δ and of course on properties of the criterion function being optimised. We explore various choices of the function, its first argument and the parameter δ , and compare the resulting convergence rates in the following section.

5. Examples

Example 1. Quadratic regression revisited

As a first example we revisit the quadratic regression model of Section 3.1. In this example we seek to estimate the parameters θ_2 and θ_3 so that these parameter estimators are uncorrelated with each other. To do so we choose \mathbf{p} by maximising $\phi(\mathbf{p}) = -[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2$, where $\mathbf{a} = (0, 1, 0)^\top$ and $\mathbf{b} = (0, 0, 1)^\top$.

We start off by taking $f(x, \delta) = \Phi(\delta x)$ and choose the x_j to be the partial derivatives d_j of $\phi(\cdot)$ with respect to the p_j . We take the initial design to be $p_j^{(0)} = 1/3$, $j = 1, 2, 3$, set $\delta = 0.001$ and apply algorithm (21). After 856 iterations, we obtain

$$\begin{aligned} \phi(\mathbf{p}) &= -5.86939 \times 10^{-17}, \\ -0.0000005 &\leq F_i^\phi \leq 0.00000001, \end{aligned}$$

for $i = 1, 2, 3$.

In a second experiment we use the same settings as above but take $\delta = 0.01$. In this case, after 98 iterations, we obtain

$$\begin{aligned} \phi(\mathbf{p}) &= -1.883629 \times 10^{-30}, \\ -9.32 \times 10^{-14} &\leq F_i^\phi \leq 2.76 \times 10^{-15}, \end{aligned}$$

for $i = 1, 2, 3$.

As a third experiment we take $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, $x_j = F_j$ and $\delta = 0.001$. Starting with equal initial weights, after 1225 iterations, we obtain

$$\begin{aligned} \phi(\mathbf{p}) &= -1.300011 \times 10^{-12}, \\ -0.000077 &\leq F_i^\phi \leq 0.000002, \end{aligned}$$

for $i = 1, 2, 3$.

As a fourth experiment we use the same choices as in the third experiment but increase δ to 0.05. We then obtain, after 56 iterations

$$\begin{aligned}\phi(\mathbf{p}) &= -3.83605 \times 10^{-15}, \\ -0.000004 &\leq F_i^\phi \leq 0.0000001,\end{aligned}$$

for $i = 1, 2, 3$.

We see that different choices of $f(\cdot, \cdot)$ and of δ have a very substantial effect on the rate of convergence of the algorithm. As we proposed in Section 4, using the directional derivatives for the first argument of $f(\cdot, \cdot)$ (third and fourth experiments), the results are very satisfactory in terms of speed of convergence. In all four experiments $\phi(\mathbf{p}) = 0$ (up to at least 11 decimal places) and the design to which the algorithm converged was $\mathbf{p}^* = (0.4729515, 0.4984532, 0.0285953)$. The corresponding support points are $(-1, 0, 2)$. As the global maximum is achieved, we do not need to provide the eigenvalues of the Hessian matrix. We have found a \mathbf{p} for which the criterion function attains the maximum possible value of zero. Thus the covariance is zero indicating that the estimators of the parameters θ_2 and θ_3 are uncorrelated with each other.

Note that the above optimal design is approximately the same as the solution $(0.471, 0.5, 0.029)$ which was obtained by solving the equation $\tilde{h}(\mathbf{p}) = (p_1 + p_3 - 1/2)^2 - 9(p_3 - 1/2)^2 + 2 = 0$ in Section 3.1.

Example 2. A practical problem in Chemistry

The following example was discussed briefly in Section 1. It involves a practical problem arising in Chemistry (Torsney & Alahmadi 1995). In this example we cannot obtain zero covariance. However we can still construct a design such that the parameter estimators of interest are as uncorrelated as possible. The regression model which is involved in this example describes the relationship between the viscosity y and the concentration x of a chemical solution. Viscosity is the response. The model is given by

$$E(y|x) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2.$$

The design interval is restricted to $(0, 0.2]$. It is desired to minimise the magnitude of the covariance between $\hat{\theta}_3$ and each of $\hat{\theta}_1$ and $\hat{\theta}_2$. That is we seek to minimise each of $|\text{cov}(\hat{\theta}_1, \hat{\theta}_3)|$ and $|\text{cov}(\hat{\theta}_2, \hat{\theta}_3)|$. As we cannot obtain a design to produce a common value of zero covariance, we seek two separate designs, one to minimise $|\text{cov}(\hat{\theta}_1, \hat{\theta}_3)|$ and one to minimise $|\text{cov}(\hat{\theta}_2, \hat{\theta}_3)|$, rather than to minimise the two simultaneously.

In dealing with the first covariance case we take $\mathbf{a} = (1, 0, 0)^\top$ and $\mathbf{b} = (0, 0, 1)^\top$. We discretise the design space at intervals of 0.01 between 0.02 and 0.20. We start with equal

initial weights. Note that, for this particular model, both the partial and directional derivatives are numerically very large in early iterations of the algorithm (21). The convergence of the algorithm is consequently slow in terms of the number of iterations, and sometimes the functional values of $f(\cdot, \cdot)$ become unbounded. To alleviate this problem, we replace the partial derivatives d_i by standardised versions $\tilde{d}_i = d_i / \sum_{j=1}^J p_j d_j$. Note that $\sum_{i=1}^J p_i d_i = -2\phi(\mathbf{p})$, which is positive when $\phi(\mathbf{p})$ is nonzero. However, we do not need to standardize the directional derivatives as these values (based on the standardized partial derivatives) become bounded. We then run algorithm (21) with $f(x, \delta) = \exp(\delta x) / (1 + \exp(\delta x))$, $x_j = \tilde{d}_j$ and $\delta = 1.5$. After 1909 iterations, we obtain

$$\begin{aligned} \phi(\mathbf{p}) &= -1487305833, \\ -1.025987 &\leq F_i^\phi \leq 5.316815 \times 10^{-10}, \end{aligned}$$

for $1 \leq i \leq J = 19$. The algorithm converges to a solution having three support points, namely 0.02, 0.12 and 0.20 with corresponding weights (0.4233560, 0.4049047, 0.1717393). So the solution is on a boundary of \mathcal{P} . The directional derivatives corresponding to the above three support points are zero and are negative towards all zero weighted remaining design points. At this solution the value of the covariance whose magnitude is being minimised is -38565.6. Here \mathbf{p}_+^* (the vector of the positive components of \mathbf{p}) is of length 3, so that $\mathbf{H}(\mathbf{p}_+^*)$ is of order 2×2 . The Hessian $\mathbf{H}(\mathbf{p}_+^*)$ is calculated using (15). Based on the above solution, we find that $\det(\mathbf{H}(\mathbf{p}_+^*)) = 1.202243 \times 10^{21}$, $d_{11}^{\tilde{\phi}}(\mathbf{p}_+^*) = -48693553129$ and $d_{22}^{\tilde{\phi}}(\mathbf{p}_+^*) = -49333921945$. The design satisfies both the first and second order conditions. We have found a \mathbf{p} for which the criterion function attains a local maximum.

With the same settings, but replacing $x_j = \tilde{d}_j$ by $x_j = F_j$ the algorithm converges to a solution with the same three support points (0.02, 0.12, 0.20) and the same weights (0.4233560, 0.4049047, 0.1717393) after only 527 iterations.

In dealing with the second covariance case we take $\mathbf{a} = (0, 1, 0)^\top$, $\mathbf{b} = (0, 0, 1)^\top$ and again start with equal initial weights. Again the values of both the partial and directional derivatives are very large in the initial stages of the algorithm so we employ the same technique of standardising the partial derivatives. The logistic c.d.f. seems to be a good choice for $f(\cdot, \cdot)$ so we use this function with $x_j = \tilde{d}_j$ and $\delta = 1.3$. After 892 iterations, we obtain

$$\begin{aligned} \phi(\mathbf{p}) &= -47739048, \\ -1.184627 &\leq F_i^\phi \leq 2.198301 \times 10^{-8}, \end{aligned}$$

for $1 \leq i \leq J = 19$. The algorithm again converges to a solution with three support points, namely 0.02, 0.12 and 0.20 with corresponding weights (0.5089060, 0.3468093, 0.1442847).

Again the solution is on a boundary of \mathcal{P} . The directional derivatives corresponding to the above three support points are zero and are negative towards all zero weighted remaining design points. The value of the covariance at this solution is 6909.345. Let \mathbf{p}_+^* be as above. Again $\mathbf{H}(\mathbf{p}_+^*)$ is of order 2×2 and is calculated using (15). Based on the above solution, we find that $\det(\mathbf{H}(\mathbf{p}_+^*)) = 1.431919 \times 10^{18}$, $d_{11}^{\tilde{\phi}}(\mathbf{p}_+^*) = -1698694792$ and $d_{22}^{\tilde{\phi}}(\mathbf{p}_+^*) = -1874075146$. The design satisfies both the first and second order conditions. We have found a \mathbf{p} for which the criterion function attains a local maximum.

We again endeavour to improve the rate of convergence by taking the x_j to be the directional derivatives. With this choice of the x_j , and with $\delta = 1.9$, the algorithm converges to a solution with the same three support points and same corresponding probability weights after only 223 iterations.

As we mentioned, the values of the covariance are -38565.6 and 6909.345 at the optimum for the above two cases respectively. We cannot achieve beyond these minimum values. The corresponding correlation values are -0.94 and 0.81 approximately. Given the very short design interval (0,0.2], all three parameter estimators will be highly correlated, positively or negatively, under any design. (The parameters will be inestimable if all observations are taken at one value of x — a one point design.) It was precisely because of this that the chemists, who supplied the problem, wanted a design under which some correlations were somewhat reduced, in the knowledge that they would still be high.

Note that all the above results are given only for the best choices of the parameter δ . Values of δ both lower and higher than those presented were also investigated. We also investigated discretising the design space at intervals of 0.001 and found that the algorithm converges to the same design. However, in this case it takes more than 100,000 iterations to converge.

Example 3. Simultaneous optimisation

The following example, briefly discussed in Section 1, is one of five which were studied by Silvey, Titterton & Torsney (1978) and Wu (1978) in the context of constructing D -optimal designs on a finite design space. The model is $E(y) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 = \mathbf{v}^\top \boldsymbol{\theta}$, $\mathbf{v} \in \mathcal{V}$, where

$$\mathcal{V} = \{ \mathbf{v} = (1, x_1, x_2)^\top : (x_1, x_2) \in \mathcal{Q} \},$$

and where \mathcal{Q} is a quadrilateral with vertices (-1, -1), (-1, 1), (1, -1), (2, 3). We take the design points to consist of the vertices of the quadrilateral, and wish to find the optimal weights for these four vertices according to the following covariance criteria.

Our goal is to obtain an estimator of the parameter θ_1 which is uncorrelated with both of the estimators of the other two parameters θ_2 and θ_3 . Thus we take $\phi_1(\mathbf{p}) = -[\mathbf{a}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{b}]^2$, and $\phi_2(\mathbf{p}) = -[\mathbf{c}^\top \mathbf{M}^{-1}(\mathbf{p}) \mathbf{d}]^2$, where $\mathbf{a} = (1, 0, 0)^\top$, $\mathbf{b} = (0, 1, 0)^\top$, $\mathbf{c} = \mathbf{a}$ and $\mathbf{d} = (0, 0, 1)^\top$. We seek to maximise the compound criterion as given in (19). We use $f(x, \delta) = \exp(\delta x)/(1 + \exp(\delta x))$, $\delta = 0.05$ and x_j as given in (22) (corresponding to the criterion (19)) of Section 3.2.

With equal initial weights ($\mathbf{p}^{(0)} = (0.25, 0.25, 0.25, 0.25)$) we obtain at the initial iteration the following values

$$\begin{aligned} (\phi_1, \phi_2) &= (-0.001371742, -0.02777778), \\ -0.1033379 &\leq F_i^{\phi_1} \leq 0.08179647, \\ -0.2695473 &\leq F_i^{\phi_2} \leq 0.3518519. \end{aligned}$$

After 882 iterations we obtain

$$\begin{aligned} (\phi_1, \phi_2) &= (-4.599769 \times 10^{-17}, -1.236737 \times 10^{-16}), \\ -0.00000001 &\leq F_i^{\phi_1} \leq 0.00000001, \\ -0.00000003 &\leq F_i^{\phi_2} \leq 0.00000002, \end{aligned}$$

for $1 \leq i \leq 4$.

We also consider maximising the minimum of $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$, i.e. using compound criterion (20) rather than maximising their sum. We use the same $f(\cdot, \cdot)$ and $\delta = 0.12$, and the same starting values but replace the partial derivatives by the corresponding directional derivatives. After 662 iterations we obtain

$$\begin{aligned} (\phi_1, \phi_2) &= (-3.366706 \times 10^{-16}, -4.121759 \times 10^{-16}), \\ -0.00000004 &\leq F_i^{\phi_1} \leq 0.00000004, \\ -0.00000005 &\leq F_i^{\phi_2} \leq 0.00000003, \end{aligned}$$

for $1 \leq i \leq 4$. The solution obtained by both procedures is given by $\mathbf{p}^* = (0.3498955, 0.2200836, 0.2900627, 0.1399582)$. Note that, in our setup, $\phi_1(\mathbf{p}) \leq 0$ and $\phi_2(\mathbf{p}) \leq 0$. We obtained the criterion values $-3.366706 \times 10^{-16}$ and $-4.121759 \times 10^{-16}$ respectively. So the criterion values differ from zero only by numerical noise. As the global maximum is achieved, we do not need to provide the eigenvalues of the Hessian matrix. We have found a \mathbf{p} for which both of the criterion functions attain the maximum possible value of zero. So both covariances are zero indicating that the estimator of the parameter θ_1 is uncorrelated with both of the estimators of the other two parameters θ_2 and θ_3 .

Although we investigated the best values of δ in running algorithm (21), there is no clear rule by which a user could determine an efficacious value of δ a priori without resorting to experimentation. In general, there is no explicit formula for δ in terms of $\mathbf{p}^{(r)}$ and $\mathbf{d}^{(r)}$ in algorithm (21), but one can suggest an approximate range of values of it according to the problems and regression models considered in this paper.

6. Concluding remarks

In the present work we have tried to address an important problem of estimation and optimal regression design. The problem is that of estimating regression model, parameters or linear combinations of these in such a manner that their estimators are uncorrelated with each other. To solve this problem we seek to choose a design which renders the covariances among relevant parameter estimators as close to zero as possible.

We developed a class of algorithms for solving the resulting optimisation problem and also investigate techniques for improving convergence rates. The algorithm is based on a differential calculus approach which extends to a Lagrangian approach.

If one seeks to obtain a parameter estimator which is uncorrelated with two other parameter estimators, a compound optimisation problem arises. To solve this problem we transformed it into one of maximising two functions of the design weights simultaneously. In the present work we restricted attention to problems involving only one or two covariances. However it is clear that the simultaneous optimisation approach which we have developed can be generalised to optimising more than two covariances.

References

- ATKINSON, A.C., DONEV, A.N. & TOBIAS, R.D. (2007). *Optimum experimental designs, with SAS*. Oxford: Oxford University Press.
- AVRIEL, M. (2003). *Nonlinear programming: analysis and methods*. New York: Dover Publications, Inc.
- BERGER, M.P.F. & WONG, W.K. (2009). *An introduction to optimal designs for social and biomedical research*. West Sussex: John Wiley & Sons.
- FEDOROV, V.V. (1972). *Theory of optimal experiments*. New York: Academic Press.
- KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics* **2**, 849–879.
- MANDAL, S. & TORSNEY, B. (2000). Algorithms for the construction of optimizing distributions. *Communications in Statistics – Theory and Methods* **29**, 1219–1231.
- MANDAL, S. & TORSNEY, B. (2006). Construction of optimal designs using a clustering approach. *Journal of Statistical Planning and Inference* **136**, 1120–1134.
- MANDAL, S., TORSNEY, B. & CARRIERE, K.C. (2005). Constructing optimal designs with constraints. *Journal of Statistical Planning and Inference* **128**, 609–621.
- NOTO, D., BARBAGALLO, C.M., CASCIO, A.L., CEFALÚ, A.B., CAVERA, G., CALDARELLA, R., MARINO, G., TRAVALI, S., CUTAIA, I., MARINGHINI, S., NOTARBARTOLO, A. & AVERNA, M.

- (1999). Lipoprotein(a) levels in relation to albumin concentration in childhood nephrotic syndrome. *Kidney International* **55**, 2433–2439.
- ODENIYI, M.A. & JAIYEGBA, K.T. (2009). Optimization of ascorbic acid tablet formulations containing hydrophilic polymers. *Farmacia* **57**, 157–166.
- PUKELSHEIM, F. (2006). *Optimal design of experiments*. Philadelphia: Society for Industrial and Applied Mathematics.
- SILVEY, S.D. (1980). *Optimal design: an introduction to the theory for parameter estimation*. London: Chapman and Hall.
- SILVEY, S.D., TITTERINGTON, D.M. & TORSNEY, B. (1978). An algorithm for optimal designs on a finite design space. *Communications in Statistics – Theory and Methods* **7**, 1379–1389.
- TITTERINGTON, D.M. (1976). Algorithms for computing D -optimal designs on a finite design space. In *Proceedings of the 1976 Conference on Information Sciences and Systems*. pp. 213–216. Baltimore: Department of Electronic Engineering, John Hopkins University.
- TORSNEY, B. (1977). Contribution to discussion of ‘maximum likelihood from incomplete data via the EM algorithm’ by A.P. Dempster, N.M. Laird & D.B. Rubin. *Journal of the Royal Statistical Society: Series B* **39**, 26–27.
- TORSNEY, B. (1983). A moment inequality and monotonicity of an algorithm. In *Proceedings of the International Symposium on Semi-Infinite Programming and Applications, Lecture Notes in Economics and Mathematical Systems*, eds. K.O. Kortanek & A.V. Fiacco, vol. 215. pp. 249–260.
- TORSNEY, B. (1988). Computing optimizing distributions with applications in design, estimation and image processing. In *Optimal design and analysis of experiments*, eds. Y. Dodge, V.V. Fedorov & H.P. Wynn. pp. 361–370. North Holland: Elsevier Science Publishers B.V.
- TORSNEY, B. & ALAHMADI, A.M. (1995). Designing for minimally dependent observations. *Statistica Sinica* **5**, 499–514.
- TORSNEY, B. & MANDAL, S. (2001). Construction of constrained optimal designs. In *Optimum design 2000*, eds. A. Atkinson, B. Bogacka & A. Zhigljavsky. pp. 141–152. London: Kluwer Academic Publishers.
- WHITTLE, P. (1971). *Optimisation under constraints: theory and applications of nonlinear programming*. London: Wiley-Interscience.
- WHITTLE, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society: Series B* **35**, 123–130.
- WU, C.F.J. (1978). Some iterative procedures for generating nonsingular optimal designs. *Communications in Statistics – Theory and Methods* **7**, 1399–1412.