ELSEVIER

Journal of Statistical Planning and Inference 85 (2000) 103–114

journal of statistical planning and inference

www.elsevier.com/locate/jspi

An algorithmic construction of optimal minimax designs for heteroscedastic linear models

Lawrence D. Brown[a], Weng Kee Wong[b,*]

[a]Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[b]Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA 90095, USA

Abstract

We construct optimal designs to minimize the maximum variance of the fitted response over an arbitrary compact region. An algorithm is proposed for finding such optimal minimax designs for the simple linear regression model with heteroscedastic errors. This algorithm always finds the optimal design in a few simple steps. For more complex models where there is a symmetric error variance structure, we suggest a strategy to help find some hitherto elusive optimal minimax designs.

Keywords: Approximate designs; Efficiency function; Extrapolation; Information matrix

1. Introduction

We consider the problem of finding an optimal design to estimate the response surface of an experiment over an arbitrary region Y when there is heteroscedasticity in the model. The optimality criterion is to minimize the maximum variance of the estimated response surface over Y. Motivation for these problems can be found in Ehrenfeld (1955), Elfving (1958), Kiefer and Wolfowitz (1964a, b, 1965), Gaylor and Sweeny (1965), Wong and Cook (1993), and Wong (1993). In addition, these types of designs are useful for two other reasons: (i) they can be motivated from the cost perspectives, Pazman (1986, p. 36) and (ii) these designs are related to multiple-objectives optimal design problems, Cook and Fedorov (1995). The latter set of researchers showed that solving a multiple-objective design problem invariably requires solving a sub-minimax problem similar to those discussed here.

The verification of these optimal designs and other types of optimal minimax designs generally involves finding a maximin measure μ^*, see Atkinson and Fedorov (1975a, b), Fedorov and Khabarov (1986), and Cook and Fedorov (1995). However, despite

* Corresponding author. Fax: 1-310-267-2113.

E-mail address: wkwong@norfolk.sunlab.ph.ucla.edu (W.K. Wong)

0378-3758/00/$ - see front matter
PII: S0378-3758(99)00073-7

its ubiquity, very little is known about this measure μ^*, except that an effective way of determining μ^* will help considerably in finding the optimal minimax design.

In this paper, we provide an efficient algorithm for finding optimal minimax designs for the simple linear model, and elicit a property of μ^* useful for finding optimal designs in more complex models. Algorithms are useful because they can be implemented in computers and generate the optimal designs for practitioners. The assumptions in this work are that both the design space X and the region Y are known and compact and the model is linear. The variance of the response at the point x is described by a positive function $\lambda(x)$ which is frequently called the efficiency function. Following previous work, (Lim and Studden, 1988; Lau, 1988; Dette, 1990, 1992; DasGupta et al., 1992; Wong, 1993, 1994), we shall assume $\lambda(x)$ is known up to a multiplicative constant. This is not an unrealistic assumption especially if one relates $\lambda(x)$ to the cost incurred in taking an observation at the point x (Pazman, 1986, p. 36). All observations are uncorrelated.

Designs are treated as probability measures on the design space X and so they are approximate designs in the sense of Kiefer and Wolfowitz (1964a, b). Thus, a probability measure with mass m_i at the point $x_i \in X$ is a design which takes Nm_i observations at x_i, $i = 1, 2, \ldots, k$, and N is the number of observations for the experiment.

Following convention (Kiefer and Wolfowitz, 1964a, b), the usefulness of a design ξ on X is measured by its information matrix,

$$M(\xi) = \int_X \lambda(x) f(x) f^\top(x) \xi(\mathrm{d}x),$$

where the coordinates of $f(x)$ are a basis for the set of regression functions. Our design criterion is to find a design ξ^* among all designs on X to minimize, for some given set Y,

$$\bar{d}(\xi) = \max_{x \in Y} d(x, \xi),$$

where $d(x,\xi) = f^\top(x) M^{-1}(\xi) f(x)$ is the variance (apart from a constant) of the fitted response at the point x using design ξ. Gaylor and Sweeny (1965) found optimal designs for the simple linear model when $X = Y$ and $\lambda(x)$ is constant, i.e. errors are homoscedastic. As we will show in the next two sections, optimal designs for heteroscedastic models need not include the extreme points of the design space as their support points and consequently the design problem is more complicated.

When $X = Y$, these optimal designs are called heteroscedastic G-optimal designs and have been studied in Wong and Cook (1993) and Wong (1993). For convenience, we will abbreviate them as simply G-optimal designs. It was shown there that one may verify if a design ξ^* is G-optimal by first letting $A(\xi) = \{a \in X \mid d(a, \xi) = \bar{d}(\xi)\}$ and then checking that there exists a probability measure μ^* defined on $A(\xi^*)$ such that

$$g_\lambda(x, \mu^*, \xi^*) = \int_{A(\xi^*)} \lambda(x) \{f^\top(x) M^{-1}(\xi^*) f(a)\}^2 \mu^*(\mathrm{d}a) - \bar{d}(\xi^*) \leqslant 0 \quad \text{for all } x \in X \tag{1.1}$$

with equality at the support points of ξ^*. This approach is generally labor intensive largely because there is no efficient way of finding μ^*.

In Section 2, we propose an algorithm for generating these optimal designs for the simple linear model. Justifications for this algorithm are given in Section 3, and in Section 4, as illustrations, we provide optimal designs found from our algorithm. Section 5 studies properties of μ^* in (1.1) and suggests a strategy for finding G-optimal designs for polynomials of higher degrees.

2. An algorithm for optimal minimax designs in the simple linear model

We provide here an algorithm for the simple linear model, which avoids the problem of having to find μ^* in (1.1). Here $f^\top(x) = (1, x)$. We temporarily assume $X = Y = [-1, 1]$ and defer the more general case to the end of this section. Since our criterion is strictly convex on the space of information matrices, we can confine our search for the optimal design ξ^* to 2- or 3-point designs by the second part of Caratheodory's Theorem (Pazman, 1986, p. 57). In addition, for simple linear regression, the variance function $d(x, \xi^*)$ is a quadratic in x and so it is maximized at one of 3 possible sets in X: (i) $A(\xi^*) = \{1\}$, (ii) $A(\xi^*) = \{-1\}$ and (iii) $A(\xi^*) = \{-1, 1\}$. These correspond to off-diagonal elements of $M(\xi^*)$ (i.e. $\{M(\xi^*)\}_{12}$) as being negative, positive or zero, respectively.

A salient feature of our algorithm is that it always finds the optimal design in 2 or 3 steps. At each stage of the algorithm, a set of support points is hypothesized and the mass at each support point of the optimal design is found (Theorems 3.1 and 3.2). Afterwards, the maximum of the variance function is expressed as a function of the support points and minimized by solving the derivative of $\bar{d}(\xi^*)$ for the points in X. Alternatively, a numerical search routine such as the Newton–Raphson method could be used for this last step.

We introduce a bit more notation for the rest of the paper. The value of $\lambda(x)$ at the point $x_i \in X$ is denoted by λ_i and the design with mass ξ_i at x_i, $i = 1, 2, \ldots, k$, is denoted by ξ. When $x_1, \ldots, x_k$ are known, it is sometimes convenient to view ξ as a $k \times 1$ vector with elements $\xi_1, \ldots, \xi_k$. In all cases, $k = 2$ or 3. In addition, let $v_i^\top = (-x_i, 1)$, $i = 1, 2$, and let $\{t\}_i$ denote the ith element in the vector t. The algorithm proceeds as follows:

Step A1: Hypothesize a two-point support set $S_2 = \{x_1, x_2\}$ with $-1 \leqslant x_1 < x_2 \leqslant 1$. Use the algebraic formulae in Theorem 3.1 to calculate the optimal design ξ^* and the corresponding value $\bar{d}(\xi^*)$.

Step A2: Minimize $\bar{d}(\xi^*)$ over all two point sets S_2.

Step B1: Hypothesize a three-point support set $S_3 = \{x_1, x_2, x_3\}$ with $-1 \leqslant x_1 < x_2 < x_3 \leqslant 1$. If $x_2 \leqslant 0$, set $z_i = x_i$ and $\tilde{\lambda}(z_i) = \lambda(z_i)$, $i = 1, 2, 3$; otherwise, set $z_i = -x_{4-i}$ and $\tilde{\lambda}(z_i) = \lambda(z_{4-i})$, $i = 1, 2, 3$.

$$\text{Define } n_1 = \frac{\tilde{\lambda}_1 \tilde{\lambda}_3 (z_3 - z_1) - \tilde{\lambda}_1 \tilde{\lambda}_2 (z_2 - z_1) - \tilde{\lambda}_2 \tilde{\lambda}_3 (z_3 - z_2)}{z_3 \tilde{\lambda}_3 - z_2 \tilde{\lambda}_2},$$

$$n_2 = \frac{\tilde{\lambda}_1\tilde{\lambda}_3(z_3-z_1)(-z_1z_3)+\tilde{\lambda}_1\tilde{\lambda}_2(z_2-z_1)(z_1z_2)+\tilde{\lambda}_2\tilde{\lambda}_3(z_3-z_2)(z_2z_3)}{z_3\tilde{\lambda}_3-z_2\tilde{\lambda}_2},$$

$$\beta = \left\{\frac{\tilde{\lambda}_2\tilde{\lambda}_3(z_3-z_2)}{z_3\tilde{\lambda}_3-z_2\tilde{\lambda}_2}(\sqrt{|n_2|}+z_2z_3\sqrt{|n_1|})\right\}/\{n_2\sqrt{|n_1|}-n_1\sqrt{|n_2|}\}$$

and

$$\xi_1=\beta, \quad \xi_2=\frac{z_3\tilde{\lambda}_3-\beta(z_3\tilde{\lambda}_3-z_1\tilde{\lambda}_1)}{z_3\tilde{\lambda}_3-z_2\tilde{\lambda}_2} \quad \text{and} \quad \xi_3=\frac{-z_2\tilde{\lambda}_2+\beta(z_2\tilde{\lambda}_2-z_1\tilde{\lambda}_1)}{z_3\tilde{\lambda}_3-z_2\tilde{\lambda}_2}.$$

Step B2: If $n_1n_2<0$ and $0<\beta<\beta_{\max}=z_3\tilde{\lambda}_3/\{z_3\tilde{\lambda}_3-z_1\tilde{\lambda}_1\}$, set $\xi_i^*=\xi_i$, $i=1,2,3$, if $x_2\leqslant 0$; otherwise set $\xi_i^*=\xi_{4-i}$, $i=1,2,3$.

Step B3: If $n_1n_2\geqslant 0$ or $\beta\leqslant 0$ or $\beta\geqslant z_3\tilde{\lambda}_3/\{z_3\tilde{\lambda}_3-z_1\tilde{\lambda}_1\}$, stop. (The optimal design supported on X will actually be a two point design and hence will have already been included in Steps A1 and A2.)

Step B4: Minimize $\bar{d}(\xi^*)$ over all three point sets satisfying the conditions in Step B2.

Step C: Steps A2 and B4 will produce the optimal 2- and 3-point designs, respectively. The optimal design is the one with the smaller value of $\bar{d}(\xi^*)$.

The above results can be generalized to the case when $X\neq Y$. It appears that only Kiefer and Wolfowitz (1964a, b) and, Gaylor and Sweeny (1965) had investigated this problem when the model is homoscedastic; no work is available for the heteroscedastic case. The first set of researchers focused on the case when $X=[-1,1]$ and $Y=[-a,a]$ and a is either very large or very small, while the latter set of researchers obtained optimal designs for the simple linear model when X is fixed and Y is an arbitrary compact set. It is not hard to see our argument in Section 3 can be readily applied to any compact sets X and Y. For the simple linear heteroscedastic model, it can be shown that if $X=[z_1,z_2]$ and $Y=[y_1,y_2]$, and the variance function of the two-point design ξ on Y is maximized only at y_2, the mass of the optimal two-point design at the larger support point x_2 is

$$\xi_2=\frac{\sqrt{\lambda_1}(y_2-z_1)}{\sqrt{\lambda_1}(y_2-z_1)+\sqrt{\lambda_2}|y_2-z_2|}. \tag{2.1}$$

When $\lambda(x)$ is a constant and $X=[0,1]$, we have $A(\xi)=\{y_2\}$ if $y_1>1$ for any design ξ supported on $\{0,1\}$. The optimal design is supported at 0 and 1 with mass at 1 equals to $y_2/(y_2+|y_2-1|)=y_2/(2y_2-1)$. This is a more compact expression than the one given in (15) of Gaylor and Sweeny (1965). Other formulas can be similarly deduced, but are omitted for space consideration.

3. Justication for the algorithm

In this section, we provide justification for the algorithm proposed in Section 2 when $X=Y=[-1,1]$. We begin with Theorem 3.1 which gives the mass distribution of a two-point optimal design if the support points are known.

Theorem 3.1. *Let ξ be the optimal design among those supported on two given points x_1 and x_2 with $-1 \leqslant x_1 < x_2 \leqslant 1$. Then*

(i) $A(\xi) = \{1\}$ *if and only if* $$\xi_2 = \frac{\sqrt{\lambda_1}(1 - x_1)}{\sqrt{\lambda_1}(1 - x_1) + \sqrt{\lambda_2}(1 - x_2)}, \tag{3.1}$$

(ii) $A(\xi) = \{-1\}$ *if and only if* $$\xi_2 = \frac{\sqrt{\lambda_1}(1 + x_1)}{\sqrt{\lambda_1}(1 + x_1) + \sqrt{\lambda_2}(1 + x_2)},$$

(iii) $A(\xi) = \{-1, 1\}$ *if and only if (i) and (ii) fail and*

$$\xi_2 = \frac{-\lambda_1 x_1}{\lambda_2 x_2 - \lambda_1 x_1} \quad \textit{with} \quad x_1 < 0 < x_2. \tag{3.2}$$

Proof. (i) Let ξ be a design with mass ξ_i and x_i, $i = 1, 2$. In order that $A(\xi) = \{1\}$ holds, ξ_2 must be the unique solution to

$$\frac{\partial}{\partial \rho_2} d(1, \rho)|_{\rho_2 = \xi_2} = 0 \tag{3.3}$$

for any two-point design ρ with mass ρ_i on x_i, $i = 1, 2$. Since $d(1, \rho)$ is convex, differentiable for $\rho_2 \in (0, 1)$, (3.3) has at most one solution. Furthermore, $\max\{d(1, \rho), d(-1, \rho)\}$ tends to infinity as ρ_2 tends to 0 or 1; thus any G-optimal design ξ^* with $A(\xi^*) = \{1\}$ must satisfy (3.3). Let m_{ij} denote the (i, j)th element of $M(\rho)$. Straightforward algebra shows the left-hand side of Eq. (3.3) is proportional to

$$f^\top(1) L \{\lambda_2 f(x_2) f^\top(x_2) - \lambda_1 f(x_1) f^\top(x_1)\} L f(1)$$

and

$$L = \begin{pmatrix} m_{22} & -m_{12} \\ -m_{12} & m_{11} \end{pmatrix}.$$

Thus (3.3) is equivalent to

$$\lambda_2 f^\top(1) L f(x_2) f^\top(x_2) L f(1) = \lambda_1 f^\top(1) L f(x_1) f^\top(x_1) L f(1)$$

and has a solution which must satisfy either

$$\sqrt{\lambda_2} f^\top(1) L f(x_2) = \sqrt{\lambda_1} f^\top(1) L f(x_1) \quad \text{or} \quad \sqrt{\lambda_2} f^\top(1) L f(x_2) = -\sqrt{\lambda_1} f^\top(1) L f(x_1). \tag{3.4}$$

But $L = \sum_{i=1}^2 \rho_i \lambda_i v_i v_i^\top$ with $v_i^\top f(x_j) \neq 0$ for $i \neq j$ and $v_1^\top f(x_2) = -v_2^\top f(x_1)$. Thus, (3.4) simplifies to $\sqrt{\lambda_1}(1 - \rho_2)(1 - x_1) = \pm\sqrt{\lambda_2} \rho_2 (x_2 - 1)$. Since $1 - x_i \geqslant 0$, $i = 1, 2$ and $1 - x_1 > 0$, only the right-hand side of Eq. (3.4) can have a solution and this is given by (3.1).

(ii) The proof for this case is entirely similar to (i), and is omitted.

(iii) Clearly, $A(\xi) = \{-1, 1\}$ if and only if the off-diagonal elements of $M(\xi)$ are 0, which is true if and only if ρ_2 is given in (3.2). The desired conclusion follows. $\square$

The next result decribes a relationship between $A(\xi^*)$ and the number of support points of the optimal design ξ^*. It implies that if a three-point optimal design ξ^* exists, then $A(\xi^*)=\{-1,1\}$ and $M(\xi^*)$ is diagonal.

Lemma 3.1. *Let ξ^* be an optimal design supported on at most 3 points which satisfies $A(\xi^*)=\{1\}$ ($A(\xi^*)=\{-1\}$). Then either ξ^* is actually supported on two points or there exists a design ξ^{**} supported on two points such that $d(1,\xi^*)=d(1,\xi^{**})$ ($d(-1,\xi^*)=d(-1,\xi^{**})$) and $A(\xi^{**})=\{1\}$ ($A(\xi^{**})=\{-1\}$).*

Proof. Let $-1\leqslant x_1<x_2<x_3\leqslant 1$ be 3 potential support points of a design ρ with mass ρ_i at x_i, $i=1,2,3$. Since $M(\rho)$ is continuous, convex and continuously differentiable on Δ, any three-point optimal design ξ^* satisfying $A(\xi^*)=\{1\}$ must satisfy

$$d(1,\xi^*)=\inf_{(\rho_1,\rho_2,\rho_3)\in\Delta} d(1,\rho). \tag{3.5}$$

Here Δ is the simplex $\{(\rho_1,\rho_2,\rho_3)\in R^3 \mid \rho_1+\rho_2+\rho_3=1 \text{ and } \rho_i>0,\ i=1,2,3\}$.

Note that $d(1,\rho)=f^\top(1)[\sum_{i=1}^3 \rho_i\lambda_i f(x_i)f^\top(x_i)]^{-1}f(1)$ and its gradient $\nabla d(1,\rho)$ has components given by

$$\frac{\partial}{\partial\rho_i}d(1,\rho)=-\lambda_i f^\top(1)M^{-1}(\rho)f(x_i)f^\top(x_i)M^{-1}(\rho)f(1)$$

$$=-\lambda_i(f^\top(x_i)M^{-1}(\rho)f(1))^2.$$

Since $|M(\rho)|>0$, we have $M^{-1}(\rho)=\sum_{i=1}^3 \rho_i\lambda_i v_i v_i^\top/|M(\rho)|$ and $|M(\rho)|f^\top(x_i)M^{-1}(\rho)$ $f(1)=\{K\rho\}_i$, where

$$K=\begin{pmatrix} 0 & \lambda_2(1-x_2)(x_1-x_2) & \lambda_3(1-x_3)(x_1-x_3)\\ \lambda_1(1-x_1)(x_2-x_1) & 0 & \lambda_3(1-x_3)(x_2-x_3)\\ \lambda_1(1-x_1)(x_3-x_1) & \lambda_2(1-x_2)(x_3-x_2) & 0 \end{pmatrix}.$$

Now, let w be any direction within the simplex Δ, i.e. $w^\top 1=0$. Then

$$\frac{\mathrm{d}}{\mathrm{d}z}d(1,\rho+zw)|_{z=0}=w^\top\nabla d(1,\rho)=\sum_{i=1}^{3}\lambda_i\{K\rho\}_i^2 w_i/|M(\rho)|^2. \tag{3.6}$$

Clearly, (3.5) yields that $w^\top 1=0$ implies that $w^\top\nabla d(1,\xi^*)=0$ since the optimal design ξ^* is in the relative interior of Δ so that $w^\top 1=0$ implies ξ^*+zw is in Δ for sufficiently small z. This can equivalently be expressed as $\lambda_i\{K\xi^*\}_i^2=h$ for some non-negative h, $i=1,2,3$. This can then be expressed in matrix terminology as

$$D^{1/2}K\xi^*\propto 1, \tag{3.7}$$

where $D=\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3)$. Note that $D^{1/2}K\xi^*\neq 0$ since all elements in the first row of K are non-positive and $\xi_i^*>0$, $i=1,2,3$. Hence (3.7) implies that the column space of $D^{1/2}K$ contains the vector $(1,1,1)^\top$. Let $w^\top=(x_3-x_2,x_1-x_3,x_2-x_1)$ and note that $(w^\top D^{-1/2})$ $(D^{1/2}K)=w^\top K=0$. Hence $D^{1/2}K$ is of rank at most 2. Inspection shows that K has rank at least 2 and so $\mathrm{rank}(D^{1/2}K)=2$.

It follows that there must be a vector γ, say, linearly independent of ξ^* such that $D^{1/2}K\gamma = 0$. If $\gamma^\top 1 \neq 0$, set $\gamma^* = \xi^* - \gamma/(\gamma^\top 1)$; otherwise set $\gamma^* = \gamma$. Then $1^\top \gamma^* = 0$ and $D^{1/2}K\gamma^* \propto 1$. It follows that $\xi^\alpha = \xi^* + \alpha\gamma^*$ satisfies

$$1^\top \xi^\alpha = 1 \quad \text{and} \quad D^{1/2}K\xi^\alpha \propto 1. \tag{3.8}$$

Let $\xi^\alpha = \xi^* + \alpha u$, let $\alpha_1 = \inf\{\alpha \mid \xi^\alpha \in \Delta\}$ and let $\alpha_2 = \sup\{\alpha \mid \xi^\alpha \in \Delta\}$. Then $\alpha_1 < 0 < \alpha_2$. Further, it follows from (3.6) and (3.8) that for any $\alpha_1 \leqslant \alpha \leqslant \alpha_2$,

$$d(1, \xi^\alpha) = d(1, \xi^*) = \inf_{\rho \in \Delta} d(1, \rho)$$

since

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} d(1, \xi^\alpha) = \sum_{i=1}^{3} \lambda_i \{K\xi^\alpha\} i^2 \gamma_i^* \propto 1^\top \gamma^* = 0.$$

In addition,

$$M(\xi^*) = \frac{\alpha_2 M(\xi^{\alpha_1})}{|\alpha_1| + \alpha_2} + \frac{|\alpha_1| M(\xi^{\alpha_2})}{|\alpha_1| + \alpha_2}.$$

Hence either $\{M(\xi^{\alpha_1})\}_{12} < 0$ or $\{M(\xi^{\alpha_2})\}_{12} < 0$ since $\{M(\xi^*)\}_{12} < 0$ because of the assumption $A(\xi^*) = \{1\}$. Hence either ξ^{α_1} or ξ^{α_2} (or both) also satisfy $A(\xi^{\alpha_i}) = \{1\}$ and so is (are) the required design supported on just two points. The proof when $A(\xi^*) = \{-1\}$ is entirely similar and hence is omitted. □

Theorem 3.2. *Let ξ be a design supported on three points x_1, x_2 and x_3 with $-1 \leqslant x_1 < x_2 < x_3 \leqslant 1$ and the mass at x_i is ξ_i, $i = 1, 2, 3$. Define*

$$n_1 = \frac{\lambda_1\lambda_3(x_3 - x_1) - \lambda_1\lambda_2(x_2 - x_1) - \lambda_2\lambda_3(x_3 - x_2)}{x_3\lambda_3 - x_2\lambda_2},$$

$$n_2 = \frac{\lambda_1\lambda_3(x_3 - x_1)(-x_1x_3) + \lambda_1\lambda_2(x_2 - x_1)(x_1x_2) + \lambda_2\lambda_3(x_3 - x_2)(x_2x_3)}{x_3\lambda_3 - x_2\lambda_2}, \tag{3.9}$$

$$\beta^* = \left\{ \frac{\lambda_2\lambda_3(x_3 - x_2)}{x_3\lambda_3 - x_2\lambda_2} (\sqrt{|n_2|} + x_2x_3\sqrt{|n_1|}) \right\} \Big/ \{n_2\sqrt{|n_1|} - n_1\sqrt{|n_2|}\}.$$

(i) Suppose $x_2 \leqslant 0$. Among all designs ξ satisfying $A(\xi) = \{-1, 1\}$, there is either an optimal one given by (3.2), or if $n_1n_2 < 0$ and β^* satisfies

$$0 < \beta^* < \beta_{\max} = x_3\lambda_3/(x_3\lambda_3 - x_1\lambda_1), \tag{3.10}$$

the optimal design ξ on $\{x_1, x_2, x_3\}$ has mass given by1

$$\xi_1 = \beta^*, \quad \xi_2 = \frac{x_3\lambda_3 - \beta^*(x_3\lambda_3 - x_1\lambda_1)}{x_3\lambda_3 - x_2\lambda_2} \quad \text{and} \quad \xi_3 = \frac{-x_2\lambda_2 + \beta^*(x_2\lambda_2 - x_1\lambda_1)}{x_3\lambda_3 - x_2\lambda_2}. \tag{3.11}$$

(ii) Suppose $x_2 > 0$. Define $z_3 = -x_1$, $z_2 = -x_2$, $z_1 = -x_3$, $\tilde{\lambda}_3 = \lambda_1$, $\tilde{\lambda}_2 = \lambda_2$ and $\tilde{\lambda}_1 = \lambda_3$. The optimal design has mass given by

$$\tilde{\xi}_1 = \beta^*, \quad \tilde{\xi}_2 = \frac{z_3\tilde{\lambda}_3 - \beta^*(z_3\tilde{\lambda}_3 - z_1\tilde{\lambda}_1)}{z_3\tilde{\lambda}_3 - z_2\tilde{\lambda}_2} \quad \text{and} \quad \tilde{\xi}_3 = \frac{-z_2\tilde{\lambda}_2 + \beta^*(z_2\tilde{\lambda}_2 - z_1\tilde{\lambda}_1)}{z_3\tilde{\lambda}_3 - z_2\tilde{\lambda}_2}.$$

Proof. We prove only the case when $x_2 \leqslant 0$ since the case for $0 \leqslant x_2$ is easily proved by a similar argument. We show among all designs satisfying $A(\xi)=\{-1,1\}$, there is either an optimal one supported on only 2 points (and hence given by (3.10)), or if $n_1 n_2 < 0$ and β^* satisfies (3.10), the optimal design is given by (3.11).

Since $A(\xi)=\{-1,1\}$, it must be that $\{M(\xi)\}_{12}=0$. Thus we have $\xi_1+\xi_2+\xi_3=1$ and $\lambda_1 x_1 \xi_1 + \lambda_2 x_2 \xi_2 + \lambda_3 x_3 \xi_3 = 0$. Since $\lambda_3 x_3 - \lambda_2 x_2 > 0$, these two equations have the general solution given in (3.11). Let

$$c_1 = \lambda_2\lambda_3(x_3-x_2)/\{x_3\lambda_3 - x_2\lambda_2\} \quad \text{and} \quad c_2 = -x_2 x_3 c_1.$$

Then a direct calculation yields $\{M(\xi)\}_{ii} = c_i + \beta n_i$, $i=1,2$, where n_1 and n_2 are given by (3.9). If $(M(\xi))_{ii}$, $i=1,2$ are non-zero, let $s(\beta)=d(1,\xi)=1/(c_1+\beta n_1)+1(c_2+\beta n_2)$. If $n_1, n_2 \geqslant 0$ and both non-zero, it is clear that $s(\beta)$ is strictly monotone on the region Δ and so $s(\beta)$ does not assume its maximum on Δ. If $n_1=n_2=0$, $s(\beta)$ is constant in β and either the choice $\beta=0$ or

$$\beta = \beta_{\max} = x_3\lambda_3/(x_3\lambda_3 - x_1\lambda_1) > 0$$

yields an optimal design among two-point designs, i.e. $\{x_2, x_3\}$ if $\beta=0$ or $\{x_1,x_3\}$ if $\beta=\beta_{\max}$. (In that case the computational scheme will have already found an optimal design.)

Thus assume $n_1 n_2 < 0$. Differentiating $s(\beta)$, we have $s'(\beta) = -n_1/(c_1+\beta n_1)^2 - n_2/(c_2+\beta n_2)^2$ and the minimum of $s(\beta)$ on Δ, if it exists, satisfies $s'(\beta)=0$. Since the design is supported on three points and $c_i+\beta n_i > 0$, $i=1,2$, this is equivalent to $\sqrt{|n_1|}(c_2+\beta n_2) = \sqrt{|n_2|}(c_1+\beta n_1)$. The solution β^* to $s'(\beta)=0$ is given in (3.10). □

4. Examples

We apply the algorithm to construct several optimal minimax designs for the model $f^\top(x)=(1,x)$ on $X=[-1,1]$ with various efficiency functions. These examples are for illustrative purposes; in actual applications other functional forms for $\lambda(x)$ could occur depending on the application.

Example 4.1. Suppose $X=Y$, $\lambda(x)=4+x-x^2$ and a G-optimal design is sought.

If we suppose a two-point optimal design ξ with support at x_1, x_2 exists, the assumption that $x_1 x_2 < 0$ implies the mass at x_2 is given by (3.2). A straightforward computation shows the diagonal elements of $M(\xi)$ are

$$m_{11} = \frac{(4+x_1-x_1^2)(x_2^2-x_2-4)}{-4-x_1+x_1^2-x_2+x_1x_2+x_2^2}, \qquad m_{22} = \frac{-x_1x_2(4+x_1-x_1^2)(x_2^2-x_2-4)}{-4-x_1+x_1^2-x_2+x_1x_2+x_2^2}$$

and

$$\bar{d}(x,\xi) = \frac{(1-x_1x_2)(-4-x_1+x_1^2-x_2+x_1x_2+x_2^2)}{-x_1x_2(4+x_1-x_1^2)(x_2^2-x_2-4)}, \quad x^\top = (x_1,x_2).$$

It is now easy to verify that on X, $\bar{d}(\xi)$ is minimized when $x_1^* = -0.868517$ and $x_2^*=1$. This is consistent with the result in Wong (1996), where a class of G-optimal designs was found for the same problem but with $\lambda(x) = 4 + x - cx^2$, $c > 0$. For each c inside a certain interval, μ^* in (1.1) is identified by a complicated expression, and so is the optimal design. The algorithm here avoids the rather tedious algebra involved.

Example 4.2. Suppose $X = Y$, $\lambda(x) = 2 + \text{Cos}(3x)$ and a G-optimal design is sought.
A direct calculation yields $n_1 = -1.14048$, $n_2 = 0.528383$, $\beta_{\max} = 0.5$ and $\beta = 0.252782$. Since $n_1 n_2 < 0$, $0 < \beta \leqslant \beta_{\max} = 0.5$, Step 2 of the algorithm is satisfied and the optimal design ξ^* is supported on $x_1 = -1$, $x_3 = 1$ and $x_2 = -0.471961$ with mass $\xi_2^* = 0.246396$ at x_2 and $\xi_3^* = 0.500822$ at 1, respectively.

The optimal design in Example 4.2 is not unique. It can be checked that the design which puts mass 0.500822 at $-1, 0.246396$ at 0.471961 and 0.252782 at 1 is also optimal. Likewise, a symmetrized four-point design on ± 1 and ± 0.461961 with mass at 1 equals to 0.376802 is also optimal. Examples where an optimal design is supported on two points with the same sign are given in Wong (1993). Our last example assumes $X \neq Y$.

Example 4.3. Suppose $X = [-1, 1]$, $\lambda(x) = 2 + x^2$ and an optimal minimax design over the interval $Y = [2, 4]$ is sought. Set $y_1 = 2$, $y_2 = 4$, $z_1 = -1$, $z_2 = 1$ and assume the optimal design ξ^* is supported on x_1 and x_2 ($> x_1$). We have by (2.1),

$$\xi_2^* = \frac{\sqrt{\lambda_1}(4 - (-1))}{\sqrt{\lambda_1}(4 - (-1)) + \sqrt{\lambda_2}|4 - (1)|} = \frac{5\sqrt{2 + x_1^2}}{5\sqrt{2 + x_1^2} + 3\sqrt{2 + x_2^2}}, \qquad \xi_1^* = 1 - \xi_2^*.$$

It can be shown $\bar{d}(\xi^*) = d(4, \xi^*)$ for all x_1 and x_2 in X and is equal to

$$\frac{2048 - 384x_1 + 832x_1^2 - 192x_1^3 + 24x_1^4 - 640x_2 - 200x_1^2 x_2 + 320x_2^2 + 25x_1^2 x_2 - 120x_2^3 + 15x_2^4}{15(2 + x_1^2)(x_2 - x_1)^2(2 + x_2^2)}.$$

It is straightforward to check that on X, this function is minimized at $x_1^* = -1$ and $x_2^* = 1$ so that $\xi_2^* = \frac{5}{8}$ from above and $\bar{d}(\xi^*) = \frac{16}{3}$. The optimal design is thus supported on 1 with mass $\frac{5}{8}$ and on -1 with mass $\frac{3}{8}$.

5. Extensions to polynomial models of higher degrees

The problem of finding numerically G-optimal designs for more complex models has persisted because of lack of an efficient algorithm. These algorithms are notoriously difficult to obtain because the criterion used here is 'non-differentiable' in Kiefer's terminology and requires use of sub-gradients, such as μ^* in (1.1). To our knowledge, the above algorithm is the first one available that guarantees to find an optimal design under a non-differentiable criterion in a few steps. Extensions of the algorithm for use in more complex models are not available at this time but we provide here some suggestions for finding optimal designs for more complex models when the efficiency function is symmetric on $[-1, 1]$. This assumption, along with the convexity of the function $\bar{d}(\xi)$, implies a symmetric optimal design always exists; see Example 4.2 and

Table 1
Numerical G-optimal designs for the quadratic model with $\lambda(x) = \exp(-cx^2)$; support points are $\pm s$ and 0, with p as its mass at 1 and q is the mass of μ^* at 1

c	p	q	s
0.5	0.383652	0.383652	1
1.5	0.449816	0.449816	1
1.75	0.459748	0.461531	0.970303
2	0.467463	0.473896	0.946385
3.69868	0.450000	0.5	0.76573
4	0.44526	0.5	0.739223
8	0.418331	0.5	0.539124
16	0.404723	0.5	0.389176

the comments following it. Consequently, the support points of μ^* are symmetrical and are at most of the form

$$\pm 1 \quad \text{if } p = 1,$$

$$0, \pm 1 \quad \text{if } p = 2,$$

$$\{\pm a_1, \ldots, \pm a_{(p-1)/2}, \ldots, \pm 1\} \quad \text{if } p = 3, 5, \ldots,$$

$$\{0, \pm a_1, \ldots, \pm a_{(p-2)/2}, \ldots, \pm 1\} \quad \text{if } p = 4, 6, \ldots.$$

Here a_1 could take the value 0 when $p = 3, 5, \ldots$ and ξ^* can be chosen to be a symmetric measure supported on, at most $2p+2$ points of the form $\{\pm x_1, \ldots, \pm x_{p+1}\}$ with x_1 possibly equals to 0. This observation can help us find the optimal design by reducing the number of variables needed to consider in (1.1); see the numerical examples to follow.

Our strategy for finding optimal minimax designs for symmetric efficiency functions relies partly on the above observation and partly on heuristics. Empirical work suggests that when there is mild heteroscedasticity in the model, the D- and G-optimal designs frequently have the same set of support points; see Table 4.1 of Wong (1995). Thus the first part of our strategy is to apply any known algorithm for generating a weighted D-optimal designs (Pazman, 1986, Ch. 5) and afterwards, assume the G-optimal designs are supported on the same set of points.

As an illustration, consider a bell-shaped efficiency function $\lambda(x) = \exp\{-cx^2\}$, $c \geqslant 0$, and the quadratic model $f^\top(x) = (1, x, x^2)$ defined on $[-1, 1]$. If $0 \leqslant c \leqslant 1.5$ (mild heteroscedasticity for this problem), it is readily verified that the G-optimal design is symmetric and supported at $0, \pm 1$ with mass at 0 equals to $1/(1 + 2\exp(c))$. The corresponding μ^* is the same as the optimal design, implying $A(\xi^*) = \{\text{Support of } \xi^*\}$. If $c > 1.5$, the non-zero support points are no longer ± 1 but tends toward zero as c increases (Table 1). Furthermore, if $c \geqslant t$ ($t \approx 3.7$), the variance function of the optimal design is maximized only at ± 1. Some numerical optimal designs are shown in Table 1. Their optimality can be verified by plotting $g_\lambda(x, \mu^*, \xi^*)$. The quantities in (1.1) are first calculated by letting $\xi^* = p\delta_s + (1-2p)\delta_0 + p\delta_{-s} +$ and $\mu^* = q\delta_{-1} + (1-2q)\delta_0 + q\delta_1$, where δ_a is the point mass at a and, $0<p,\ q<1$. Then if $c \geqslant 1.5$, the optimal design is found by solving the set of equations:

$$g_\lambda(s,\mu^*,\xi^*) = g_\lambda(0,\mu^*,\xi^*) = 0$$

and

$$\left.\frac{\mathrm{d}}{\mathrm{d}x} g_\lambda(x,\mu^*,\xi^*)\right|_{x=0} = \left.\frac{\mathrm{d}}{\mathrm{d}x} g_\lambda(x,\mu^*,\xi^*)\right|_{x=s} = 0;$$

if $0 \leqslant c \leqslant 1.5$, the optimal design is given above.

Applications to polynomial models of higher degrees are possible. For example, if we have a cubic model and $\lambda(x) = 2 - x^2$, this strategy produces a symmetric optimal design with positive support points at $1, 0.411431$ with mass 0.323367 and 0.176633, respectively. Likewise, if $\lambda(x) = 3 - x^2$, the symmetric optimal design has positive support points at $1, 0.423258$, with mass 0.292568 and 0.207432, respectively.

Acknowledgements

The research of Wong is partially supported by a NIH research grant R29 AR44177-01A1.

References

Atkinson, A.C., Fedorov, V.V., 1975a. The design of experiments for discriminating between two rival models. Biometrika 62, 57–70.

Atkinson, A.C., Fedorov, V.V., 1975b. Optimal design: experiments for discriminating between several models. Biometrika 62, 289–303.

Cook, R.D., Fedorov, V.V., 1995. Constrained optimization of experimental design. (Discussion paper) Statistics. 26, 129–178.

DasGupta, A., Mukhopadhyay, S., Studden, W.J., 1992. Compromise designs in heteroscedastic linear models. J. Statist. Plan. Inference 32, 363–384.

Dette, H., 1990. A generalization of D- and D_1-optimal design in polynomial regression. Ann. Statist. 18, 1784–1805.

Dette, H., 1992. Experimental designs for a class of weighted polynomial regression models. Computat. Statist. Data Anal. 14, 359–373.

Elfving, G., 1958. In: Ulf Grenander (Ed.), Design of Linear Experiments. Probability and Statistics, Wiley, New York.

Ehrenfeld, S., 1955. On the efficiency of experimental designs. Ann. Math. Statist. 26, 247–255.

Fedorov, V.V., Khabarov, V., 1986. Duality of optimal designs for model discrimination and parameter estimation. Biometrika 73, 183–190.

Gaylor, D.W., Sweeny, H.C., 1965. Design for optimal prediction in simple linear regression. J. Amer. Statist. Assoc. 60, 205–216.

Kiefer, J., Wolfowitz, J., 1964a. Optimum extrapolation and interpolation designs, I. Ann. Inst. Statist. Math. 16, 79–108.

Kiefer, J., Wolfowitz, J., 1964b. Optimum extrapolation and interpolation designs, II. Ann. Inst. Statist. Math. 16, 295–303.

Kiefer, J., Wolfowitz, J., 1965. On a theorem of Hoel and Levine on extrapolation designs. Ann. Inst. Statist. Math. 36, 1627–1655.

Lau, T.S., 1988. D-optimal designs on the unit q-ball. J. Statist. Plan. Inference 19, 299–315.

Lim, Y.B., Studden, W.J., 1988. Efficient D_s-optimal designs for multivariate polynomial regression on the q-cube. Ann. Statist. 16, 1225–1240.

Pazman, A., 1986. Foundations of Optimum Experimental Design. Reidel, Dordrecht.

Wong, W.K., 1993. Minimal number of support points for mini-max optimal designs. Statist. Probab. Lett. 17, 405–409.

Wong, W.K., 1994. Multifactor G-optimal designs with heteroscedastic errors. J. Statist. Planning Inference 40, 127–133.

Wong, W.K., 1995. On the equivalence of D and G-optimal designs in heteroscedastic models. Statist. Probab. Lett. 25, 317–321.

Wong, W.K., 1996. How robust are G-optimal designs to heteroscedasticity? In: Ermakov, S.M., Melas, V.B. (Eds.), Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design. Publishing House of St. Petersburg University, pp. 269–273.

Wong, W.K., Cook, R.D., 1993. Heteroscedastic G-optimal designs. J. Roy. Statist. Soc. Ser. B 55, 871–880.