Ann. Inst. Statist. Math.
Vol. 53, No. 2, 418–426 (2001)


# A NEW DESIGN CRITERION WHEN HETEROSCEDASTICITY IS IGNORED

GRACE MONTEPIEDRA[1] AND WENG KEE WONG[2]

[1]*Department of Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, OH 43403, U.S.A.*
[2]*Department of Biostatistics, UCLA, 10833 Le Conte Ave, Los Angeles, CA 90095-1772, U.S.A.*

(Received July 21, 1998; revised August 17, 1999)

**Abstract.** This paper examines the construction of optimal designs when one assumes a homoscedastic linear model, but the underlying model is heteroscedastic. A criterion that takes this type of misspecification into account is formulated and an equivalence theorem is given. We also provide explicit optimal designs for single-factor and multi-factor experiments under various heteroscedastic assumptions and discuss the relationship between the D-optimal design sought here and the conventional D-optimal design.

*Key words and phrases*: Heteroscedasticity, D-optimal, efficiency function, equivalence theorem, mean squared error, L-optimal, multi-factor experiment.

## 1. Introduction

Much of the work on design problems in the statistical literature assumes the model is known or the mean structure of the response is approximately known. The dispersion part of the model is usually taken to be known for simplicity although in practice this assumption is always questionable. In this work, we consider the design problem when there is concern about the underlying heteroscedastic structure of the model and there is belief that the mean structure is correctly specified. The optimality criterion of interest is based on the variance of the unweighted least squares estimator under the assumption that the true model is a member of a known class of heteroscedastic models. Additionally, we address the related issues of finding conditions under which the optimal design for the homoscedastic model is also optimal when there is heteroscedasticity, and conditions under which the D-optimal designs defined in this paper coincide with the conventional ones where the model is assumed to be completely known.

We suppose the model that best approximates the relationship between a response variable, y, and a $p \times 1$ vector of predictor variables, $\boldsymbol{x}$, is heteroscedastic, i.e.,

$$y = \boldsymbol{\beta}^T f(\boldsymbol{x}) + \frac{1}{\sqrt{\lambda(\boldsymbol{x}, \boldsymbol{\theta})}}\epsilon. \tag{1.1}$$

Here $f(\boldsymbol{x})$ is a $m \times 1$ vector of linearly independent functions of $\boldsymbol{x}$, $\beta \in \Re^m$ is a vector of unknown constants, the $\epsilon_i$'s are uncorrelated errors with variance $\sigma^2$ and $\lambda(\boldsymbol{x}, \boldsymbol{\theta})$ is an efficiency function. The form of the efficiency function is assumed to be known apart from the value of $\boldsymbol{\theta}$ and the values of $\boldsymbol{x}$ are confined to a given compact design space $\chi$. When $\lambda(\boldsymbol{x}, \boldsymbol{\theta})$ is constant on $\chi$, we have the usual homoscedastic model. Atkinson and

Cook (1995) provides a good reference to previous works on optimal design for models with heteroscedastic errors.

If interest were strictly in estimating $\beta$, then $\theta$ would be a nuisance parameter. Some analytical designs for this situation were obtained by Dette and Wong (1996) where they assumed $f(\boldsymbol{x})$ is a vector of monomials. Since the least squares estimate of $\beta$ under the heteroscedastic model requires knowledge of the nuisance parameter, which is usually unavailable, the researcher may prefer to use the simpler homoscedastic model:

$$y = \boldsymbol{\beta}^T f(\boldsymbol{x}) + \epsilon, \tag{1.2}$$

for estimating $\beta$ while taking into account the possibility that (1.1) is the more appropriate model. Hence with a decision to use the unweighted estimator $\hat{\beta}$ under (1.2), one needs to look at the variance of $\hat{\beta}$ under model (1.2) as the basis for the precision of this estimate. This approach is somewhat similar to Box and Draper (1959) who introduced this design criterion when there is model misspecification in the mean response only. In this paper, we find optimal designs for estimating $\beta$ under a corresponding criterion when the researcher ignores the heteroscedsticity in the model.

We suppose at the outset that a total of $N$ observations is allocated to the experiment and the response at $\boldsymbol{x}_i$ is $y_i$, $i = 1, 2, \ldots, N$. The least squares estimate for $\boldsymbol{\beta}$ in model (1.2) is

$$\hat{\beta} = \left[\sum_{i=1}^{N} f(\boldsymbol{x}_i) f^T(\boldsymbol{x}_i)\right]^{-1} \left[\sum_{i=1}^{N} f(\boldsymbol{x}_i) y_i\right].$$

Note that an attractive feature of this estimate is that it does not depend on the unknown nuisance parameter.

Let $\xi$ denote the design which takes $n_i$ observations at the support points $\boldsymbol{x}_i$, $i = 1, 2, \ldots, n$, so that $\sum_{i=1}^{n} n_i = N$. It is straightforward to show that the variance-covariance matrix of $\hat{\beta}$ under model (1.2) is proportional to

$$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{\theta}) = M_{11}^{-1}(\xi) \bar{M}(\boldsymbol{\theta}, \xi) M_{11}^{-1}(\xi) \equiv V(\boldsymbol{\theta}, \xi),$$

where

$$M_{11}(\xi) = \sum_{i=1}^{n} n_i f(\boldsymbol{x}_i) f^T(\boldsymbol{x}_i) \quad \text{and} \quad \bar{M}(\boldsymbol{\theta}, \xi) = \sum_{i=1}^{n} n_i \lambda^{-1}(\boldsymbol{x}_i, \boldsymbol{\theta}) f(\boldsymbol{x}_i) f^T(\boldsymbol{x}_i).$$

The matrix $V(\boldsymbol{\theta}, \xi)$ will play a crucial role in the rest of the paper. In what is to follow, we adopt Kiefer's approach and focus on continuous designs. This means we consider designs of the form $\{\boldsymbol{x}_i, p_i\}_{i=1}^{n}$ where we now permit each $p_i \in (0, 1)$ and $\sum_{i=1}^{n} p_i = 1$. The implemented design then takes approximately $Np_i$ observations at $x_i$, $i = 1, 2, \ldots, n$ subject to the constraint $\sum_{i=1}^{n} Np_i = N$. Consequently, an optimal design is determined by the number of support points ($n$), the location of each $\boldsymbol{x}_i$ in $\chi$ and its mass $p_i$ at $\boldsymbol{x}_i$, $i = 1, 2, \ldots, n$.

Section 2 discusses the determinant and the linear optimality criteria and presents equivalence conditions for checking if a design is optimal under our setup. In Section 3, we consider various heteroscedastic structures when the mean response is linear, and optimal designs under the mean-squared error criterion are found. Section 4 contains a discussion of our results and Section 5 extends some of the results in Section 2 to multi-factor experiments.

## 2. Optimality criteria and equivalence conditions

To further simplify the optimization problem, the criterion is chosen to be a convex function to facilitate the search of the globally optimal designs. This implies that if our criterion is $\Phi[V(\theta,\xi)]$, we assume $\Phi[V(\theta,\xi)]$ is a convex function of the design $\xi$. An optimal design is one which minimizes this function over the set $\Xi$ consisting of all designs (probability measures) on $\chi$.

Let $\xi_1$ and $\xi_2$ be any two designs on $\chi$ and let $\xi = (1-\alpha)\xi_1 + \alpha\xi_2$, where $0 \le \alpha \le 1$. The following inequality (Fedorov (1972), p. 20):

$$\begin{aligned} M_{11}(\xi)\bar{M}^{-1}(\boldsymbol{\theta},\xi)M_{11}(\xi) \le\ & (1-\alpha)M_{11}(\xi_1)\bar{M}^{-1}(\boldsymbol{\theta},\xi_1)M_{11}(\xi_1) \\ & +\alpha M_{11}(\xi_2)\bar{M}^{-1}(\boldsymbol{\theta},\xi_2)M_{11}(\xi_2) \end{aligned}$$

ensures that $\Phi$ as a function of $V^{-1}$ is also convex. (Note that we have used the same notation for information matrices of continuous designs.) It follows that equivalence theorems for the optimal designs sought here can be given.

Let us first consider the determinant optimality criterion given by

$$\begin{aligned} \Phi[V(\boldsymbol{\theta},\xi)] &= \log|V(\boldsymbol{\theta},\xi)| \\ &= \log|\bar{M}(\boldsymbol{\theta},\xi)| - 2\log|M_{11}(\xi)|. \end{aligned}$$

We call a design which minimizes this criterion over the set of all designs on $\Xi$ a D-optimal design. To distinguish this D-optimal design from the conventional one, we will refer to the latter as the standard D-optimal design. (The standard D-optimal design is found assuming that the model (1.1) is given and $\theta$ is known.) The verification of a standard D-optimal design is straightforward and well known, see Fedorov ((1972), p. 71) for example.

To find the D-optimal design, we first let

$$d_\lambda(\boldsymbol{x},\xi) = \lambda^{-1}(\boldsymbol{x},\boldsymbol{\theta})f^T(\boldsymbol{x})\bar{M}^{-1}(\boldsymbol{\theta},\xi)f(\boldsymbol{x}),$$

and let $d_1(\boldsymbol{x},\xi) = f^T(\boldsymbol{x})M_{11}^{-1}(\xi)f(\boldsymbol{x})$. Using directional derivatives considerations as described in several design monographs, Fedorov (1972) and Pukelsheim (1993) for example, we have the following conditions are equivalent:

(i) $\xi^*$ minimizes $\log|V(\boldsymbol{\theta},\xi)|$ over all $\xi$ in $\Xi$;

(ii) $\min_\Xi \max_\chi\{2d_1(\boldsymbol{x},\xi) - d_\lambda(\boldsymbol{x},\xi)\} = \max_\chi\{2d_1(\boldsymbol{x},\xi^*) - d_\lambda(\boldsymbol{x},\xi^*)\}$;

(iii) $\max_\chi\{2d_1(\boldsymbol{x},\xi^*) - d_\lambda(\boldsymbol{x},\xi^*)\} = m$, where the maxima occur at the support points of $\xi^*$.

We will apply part (iii) of the equivalence conditions to find the optimal designs in Section 3. When $\chi$ is an interval, condition (iii) above can be easily verified by means of a graph.

*Note* 1. It is clear that multiplying $\lambda(\boldsymbol{x},\theta)$ by any constant factor does not change expressions (ii) and (iii), thus preserving these equalities which are necessary and sufficient for optimality. Consequently, the optimal designs are invariant to change of scale in the ignored true efficiency function.

*Note* 2. The optimal design is also invariant to nondegenerate linear transformations of $f(\boldsymbol{x})$. Suppose $\phi(\boldsymbol{x}) = L^{-1}f(\boldsymbol{x})$, where $|L| \neq 0$. The regression function $\eta(\boldsymbol{x},\theta)$

can also be written in a reparametrized form of model (1.1) as $\eta(\boldsymbol{x}, \theta) = \rho^T \phi(\boldsymbol{x}) = \rho^T L^{-1} f(\boldsymbol{x})$ where $\rho = L^T \beta$. The variance of the estimate for $\rho$, say $\hat{\rho}$, is

$$\operatorname{Var}(\hat{\rho} \mid \theta) = \operatorname{Var}(L^T \hat{\beta} \mid \theta) = L^T \operatorname{Var}(\hat{\beta} \mid \theta) L, \quad \text{and}$$
$$|\operatorname{Var}(\hat{\rho} \mid \theta)| = |L|^2 |\operatorname{Var}(\hat{\beta} \mid \theta)| = |L|^2 |V(\theta, \xi)|.$$

Consequently, the design $\xi^*$ that minimizes $|V(\theta, \xi)|$ is the same design that minimizes $|\operatorname{Var}(\hat{\rho} \mid \theta)|$.

*Note* 3. Condition (iii) is useful for checking whether the D-optimal design for the homoscedastic model is also optimal under certain heteroscedastic assumptions. Recall that $\xi_D$ is a standard D-optimal design for the homoscedastic model if and only if $\max_{\boldsymbol{x} \in \mathcal{X}} d_1(\boldsymbol{x}, \xi_D) = m$, see Fedorov ((1972), p. 71). Hence, in order to check whether the standard $\xi_D$ is also D-optimal when the true efficiency function $\lambda(\boldsymbol{x}, \theta)$ is non-constant, one only needs to verify whether

$$\max_{\boldsymbol{x} \in \mathcal{X}} [d_1(\boldsymbol{x}, \xi_D) - d_\lambda(\boldsymbol{x}, \xi_D)] = 0. \tag{2.1}$$

It should be noted that condition (2.1) is only sufficient for optimality of $\xi_D$.

Another popular design criterion is L-optimality; see for example Cook and Nachtsheim (1982), where they applied this criterion to design a calibration experiment. This criterion seeks to find a design $\xi^*$ such that $\xi^* = \operatorname{Arg\,min}_\Xi \operatorname{tr}[AV(\theta, \xi)]$, where $A$ is a user-specified matrix. For instance, if $A$ is the identity matrix, we seek an optimal design to minimize the average variance of the estimated parameters. On the other hand, if $A = f(z) f(z)^T$, we seek a design to minimize the variance of the estimated response at the point $z$. Since trace $V(\boldsymbol{\theta}, \xi)$ is a convex function with respect to $\xi$, corresponding equivalence conditions can be similarly formulated. To do this, let $g(\boldsymbol{x}, \xi, \lambda) = f^T(\boldsymbol{x}) M_{11}^{-1}(\xi) AV(\boldsymbol{\theta}, \xi) f(\boldsymbol{x})$ and $h(\boldsymbol{x}, \xi) = f^T(\boldsymbol{x}) M_{11}^{-1}(\xi) A M_{11}^{-1}(\xi) f(\boldsymbol{x})$. The following statements are equivalent:

(i) the optimal design $\xi^*$ minimizes $\operatorname{tr}[AV(\boldsymbol{\theta}, \xi)]$ over all $\xi$ in $\Xi$;

(ii) $\min_\Xi \max_\mathcal{X} [2g(\boldsymbol{x}, \xi, \lambda) - \lambda^{-1}(\boldsymbol{x}, \theta) h(\boldsymbol{x}, \xi) - \operatorname{tr} AV(\boldsymbol{\theta}, \xi)] =$

$$\max_{\mathcal{X}} [2g(\boldsymbol{x}, \xi^*, \lambda) - \lambda^{-1}(\boldsymbol{x}, \theta) h(\boldsymbol{x}, \xi) - \operatorname{tr} AV(\boldsymbol{\theta}, \xi^*)];$$

(iii) $\max_\mathcal{X} [2g(\boldsymbol{x}, \xi^*, \lambda) - \lambda^{-1}(\boldsymbol{x}, \theta) h(\boldsymbol{x}, \xi)] = \operatorname{tr} AV(\boldsymbol{\theta}, \xi^*)$, with the maxima attained at the support points of $\xi^*$.

## 3. Examples

In this section we present examples when the assumed model is the homoscedastic simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad x \in [-1, 1] \tag{3.1}$$

but the true model is heteroscedastic. In each example, we provide the optimal design when the known efficiency function is ignored. Thus, these optimal designs can be viewed as robust designs under a preselected class of efficiency functions. Example 1 gives a simple set of conditions whereby our optimal designs remain optimal whether there is heteroscedasticity or not and also shows that a design can be D and L-optimal simultaneously. Example 2 presents a case involving a monotonic efficiency function where our D-optimal design is independent of the value of the nuisance parameter $\theta$. Example 3 shows our designs can have support points other than at the extreme ends of the design space.

*Example* 1. Suppose that the efficiency function $\lambda(x,\theta)$ is symmetric and nondecreasing in $|x|$. The standard D-optimal design, $\xi_D$, for model (3.1) is equally supported at $\pm 1$ and it is easy to show that the equivalence condition (iii) in Section 2 is the same as requiring for every $x \in [-1,1]$,

$$\lambda(x,\theta)[2x^2] \leq \lambda(1,\theta)[1+x^2]. \tag{3.2}$$

This inequality holds since $2x^2 \leq 1+x^2$ and by assumption, $\lambda(x,\theta) \leq \lambda(1,\theta)$ for every $x \in [-1,1]$. Hence, $\xi_D$ is also optimal for the D-criterion defined in this paper. Similarly, it can be shown that condition (2.1) from Note 3 is satisfied by $\xi_D$ for this case, which leads us to the same conclusion.

Consider now the linear criterion with $A = I$ and set $\xi^* = \xi_D$ in the equivalence condition (iii) for the L-criterion. After some algebra, we verify that we have

$$2\lambda^{-1}(1,\theta)[1+x^2] - \lambda^{-1}(x,\theta)[1+x^2] \leq 2\lambda^{-1}(1,\theta)$$

for every $x \in [-1,1]$ if $\lambda(x,\theta)$ is symmetric and nondecreasing in $|x|$. The above inequality also reduces to (3.2), which implies the same design is L-optimal for this problem as well.

In this example, we verified that the standard D-optimal design is optimal under both the log determinant and linear criteria defined in this paper whenever $\lambda(x,\theta)$ is symmetric and nondecreasing in $|x|$. More generally, it is easy to see that if we have an efficiency function with its maximum occuring at the endpoints of the design region, the standard D-optimal design is also optimal under the determinant and linear criteria.

*Example* 2. Suppose that the efficiency function is now $\lambda(x,\theta) = 1/(x+\theta)$ and $\theta > 1$.

If we use the same design $\xi_D$ as in Example 1, direct algebra shows that condition (iii) for the D-criterion reduces to

$$2(1+x^2) - \frac{\theta+x}{\theta^2-1}(\theta x^2 - 2x + \theta) \leq 2 \tag{3.3}$$

for every $x \in \chi = [-1,1]$. Since $\theta > 1$, (3.3) can be equivalently written as $(\theta - x)(x^2 - 1) \leq 0$, which is clearly true whenever $|x| \leq 1$. So, $\xi_D$ is D-optimal. Note that condition (2.1) from Note 3, being only a sufficient optimality condition, is not satisfied in this case.

*Example* 3. Suppose the efficiency function is given by $\lambda(x,\theta) = e^{-\theta x^2}$. By symmetry, we search for the optimal design among all symmetric two-point designs of the form

$$\xi = \left\{ \begin{matrix} -s & s \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}.$$

For these designs, condition (iii) for the D-criterion becomes

$$h(x;\theta) \le 2 \tag{3.4}$$

where $h(x;\theta) = 2(1+\frac{x^2}{s^2}) - e^{\theta(x^2-s^2)}(1+\frac{x^2}{s^2})$.

Note that for any given $s$, equality in (3.4) is attained whenever $x^2 = s^2$. Since we want the extrema of $h(x;\theta)$ to occur at the design point $s$, we find $s \in (-1,1)$ such that the first derivative of $h(x;\theta)$ is 0 at the point $x = s$. This gives

$$\frac{\partial h(x;\theta)}{\partial x} = -2\theta x e^{\theta(x^2-s^2)}\left(1+\frac{x^2}{s^2}\right) + 2\frac{x}{s^2}(2 - e^{\theta(x^2-s^2)}) = 0, \tag{3.5}$$

if $s^2 = \frac{1}{2\theta}$. To ensure that the maximum is attained at $x^2 = s^2 = \frac{1}{2\theta}$ and $s^2 \le 1$, a direct calculation shows that $\xi$ is D-optimal if $s = \frac{1}{\sqrt{2\theta}}$ whenever $\theta > \frac{1}{2}$.

Next, we observe that when $s = 1$, we have $h(x;\theta) = 2 + 2x^2 - e^{\theta(x^2-1)}(1+x^2)$ in (3.4) and $h(\pm 1;\theta) = 2$ for any $\theta$. However for the inequality in (3.4) to be true for all $x \in [-1,1]$, $h(x;\theta)$ has to be decreasing at $x = -1$ (i.e., $\frac{\partial h(x;\theta)}{\partial x}\,|_{x=-1} < 0$) and it has to be increasing at $x = 1$ (i.e., $\frac{\partial h(x;\theta)}{\partial x}\,|_{x=1} > 0$). Omitting the calculus, it can be shown that whenever $\theta < \frac{1}{2}$, $h(x;\theta)$ does behave in this way. In summary, $\xi_D$ is D-optimal for all efficiency functions of the form $\lambda(x;\theta) = e^{-\theta x^2}$ provided $\theta \le \frac{1}{2}$. When $\theta \ge \frac{1}{2}$, the design with equal mass at the points $\{-\frac{1}{\sqrt{2\theta}}, \frac{1}{\sqrt{2\theta}}\}$ is D-optimal.

We point out that the above D-optimal design is also the standard D-optimal design given in Wong (1995). The same argument can be used to show the coincidence of these two types of optimal designs for the following efficiency functions: $x+\theta$, $e^{-\theta x}$, $\theta - x^2$, $\theta - |x|$, where $\theta > 1$.

## 4. Discussion

In the previous examples, the optimal designs sought here for the determinant criterion always coincide with the standard D-optimal design. This is not always the case as the following shows.

Consider the simple linear regression problem, where $f(x) = [1,x]^T$ is defined on the design region $\chi = [-1,1]$ and the efficiency function is $\lambda(x,\theta) = 2 + \cos(\theta x)$, where $\theta = 4$. It can be verified that the design, $\xi_V^*$, which minimizes $\log|V(\theta,\xi)|$ is equally supported at $\pm 1$ with $|V(\theta,\xi_V^*)| = 0.5517$. However, the standard D-optimal design (for minimizing $\log|M(\theta,\xi)|^{-1}$, where $M(\theta,\xi) = \int_\chi \lambda(\boldsymbol{x},\theta) f(\boldsymbol{x}) f^T(\boldsymbol{x}) \xi(d\boldsymbol{x})$) is:

$$\xi_D^* = \left\{ \begin{matrix} -1 & 0 & 1 \\ 0.4535 & 0.0929 & 0.4535 \end{matrix} \right\},$$

with $|M(\theta,\xi_D^*)|^{-1} = 0.5459$ and no two-point D-optimal design exists. Note that for this example, $\min_\xi |M(\theta,\xi)|^{-1} < \min_\xi |V(\theta,\xi)|$ for $\theta = 4$. This implies that using ordinary least squares estimator with the $MSE$-based optimal design $\xi_V^*$ is not as efficient as using the appropriate weighted least squares estimator and the standard D-optimal design $\xi_D^*$ for this type of heteroscedasticity. However, when $\theta = 2$ the optimal design under both criteria is the same; it is a two-point design with the same value for the two criteria, i.e $\min_\xi |M(\theta,\xi)|^{-1} = \min_\xi |V(\theta,\xi)|$, assuring the same efficiency under both criteria. To quantify the robustness properties of the ordinary least squares estimators and standard designs under the homoscedastic model to the presence of heteroscedasticity, the concept of model validity introduced in Fedorov and Hackl (1997) and Fedorov *et al.* (1998) may be helpful.

The following result gives a condition under which the D-optimal design defined in this paper coincides with the standard D-optimal design for a heteroscedastic model. In essence, it says that the two types of D-optimal designs are the same, provided that for each criterion there exists an optimal design supported on the smallest number of points for which the information matrix is non-singular.

THEOREM 1. *Let* $f(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})]^T$ *and assume that the components of* $f(\boldsymbol{x})$ *are linearly independent. Suppose that the efficiency function is* $\lambda(\boldsymbol{x}, \theta)$ *and there exists a m-point design* $\xi_V^*$ *which minimizes* $\log |V(\theta, \xi)|$. *If there also exists a m-point design* $\xi_D^*$ *which minimizes* $\log |M(\theta, \xi)|^{-1}$, *then* $\xi_V^* = \xi_D^* \equiv \xi^*$ *and* $|V(\theta, \xi^*)| = |M(\theta, \xi^*)|^{-1}$ *for all* $\theta$.

PROOF. Let $\xi$ be a design with support points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ and corresponding weights $\{p_1, p_2, \ldots, p_m\}$. Let $\lambda_i = \lambda(\boldsymbol{x}_i, \theta)$. Then by Corollary 1 (2.3.3) of Fedorov ((1972), p. 84), it is clear that

$$|M(\theta, \xi)| = \prod_{i=1}^{m} p_i \prod_{i=1}^{m} \lambda_i |F|^2,$$
$$|\bar{M}(\theta, \xi)| = \prod_{i=1}^{m} p_i \prod_{i=1}^{m} \lambda_i^{-1} |F|^2 \quad \text{and} \quad |M_{11}(\xi)| = \prod_{i=1}^{m} p_i |F|^2,$$

where $F = [f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_m)]$ . Hence, for all $\theta$,

$$|V(\theta, \xi)| = |\bar{M}(\theta, \xi)| / |M_{11}(\xi)|^2 = 1 \Big/ \left[ \prod_{i=1}^{m} p_i \prod \lambda_i |F|^2 \right] = |M(\theta, \xi)|^{-1}$$

and the desired result follows.

Conditions under which a minimally-supported optimal design exists can be given using a similar argument in Fedorov ((1972), pp. 85–87). We state one set of such conditions here without proof:

THEOREM 2. *Suppose the mean response is a polynomial of degree* $m$, $\chi = [a, b]$ *and the efficiency function is* $\lambda(x, \theta)$. *For fixed* $\theta$, *the design which minimizes* $\log |V(\theta, \xi)|$, *is concentrated at* $m$ *points if the system of functions*

$$\{1, x, x^2, \ldots, x^{2m-2}, \lambda^{-1}(x, \theta), \lambda^{-1}(x, \theta)x, \ldots, \lambda^{-1}(x, \theta)x^{2m-2}\}$$
$$\equiv \{\phi_1(x), \phi_2(x), \ldots, \phi_{2(2m-1)}(x)\}$$

*is such that any linear combination* $\sum_{i=1}^{2(2m-1)} c_i \phi_i(x)$ *has no more than* $2m - 1$ *distinct roots in the interval* $\chi = [a, b]$.

## 5. Extensions to multi-factor experiments

In this section, we find D-optimal designs for multi-factor experiments. This is a difficult design problem in general, especially when heteroscedastic errors are involved. Optimal designs for the multi-factor experiments can be found from optimal designs for each of the single factor experiment only under stringent conditions, see Lau (1988), Rafajlowicz and Myszka (1988, 1992) and Wong (1994). Frequently, optimal designs for multi-factor experiments are found for the Kronecker product models. These models are built from smaller models by exploiting the nice properties of the Kronecker product (see Eaton (1983), p. 36, for example) and are the easiest to study analytically.

It is instructive to first consider the case when two factors are involved. Suppose $\xi_i$ is optimal for the problem $(\chi_i, f_i(x_i), \lambda_i(x_i, \theta_i))$, $i = 1, 2$ and the dimensions of $f_1(x_1)$ and $f_2(x_2)$ are $m_1$ and $m_2$ respectively. If $f_1(x_1) \times f_2(x_2)$ denotes the Kronecker product of $f_1$ and $f_2$, it is straightforward to show that the product design $\xi_1 \times \xi_2$ is D-optimal if $2d_1(x_1, \xi_1)d_2(x_2, \xi_2) - d_{\lambda_1}(x_1, \xi_1)d_{\lambda_2}(x_2, \xi_2) \leq m_1 m_2$ for all $x_1$, $x_2 \in \chi_1 \times \chi_2$. In particular, we have

$$
\begin{aligned}
(5.1) \qquad & 2d_1(x_1, \xi_1)d_2(x_2, \xi_2) - d_{\lambda_1}(x_1, \xi_1)d_{\lambda_2}(x_2, \xi_2) \\
& \leq (2d_1(x_1, \xi_1) - d_{\lambda_1}(x_1, \xi_1)) \times (2d_2(x_2, \xi_2) - d_{\lambda_2}(x_2, \xi_2)) \\
& \leq m_1 m_2
\end{aligned}
$$

provided that either (i) $d_i(x_i, \xi_i) \geq d_{\lambda_i}(x_i, \xi_i)$ for all $x_i \in \chi_i$, $i = 1, 2$ or (ii) $d_i(x_i, \xi_i) \leq d_{\lambda_i}(x_i, \xi_i)$ for all $x_i \in \chi_i, i = 1, 2$. Consequently, if one of these conditions holds, the product design $\xi_1 \times \xi_2$ is the D-optimal design for the multi-factor experiment.

Here are some examples of D-optimal designs for two-factor experiments. In each case, we assume $\chi_i = [-1, 1], f_i(x_i) = (1, x_i)^T$, $i = 1, 2$ so that $f_1(x_1) \times f_2(x_2) = (1, x_1, x_2, x_1 x_2)^T$ on $\chi_1 \times \chi_2 = [-1, 1]^2$. The optimality of each design can be verified directly using inequality (5.1).

*Example* 4. Suppose the efficiency function is $\lambda(\theta_1, \theta_2, x_1, x_2) = e^{-\theta_1 x_1^2 - \theta_2 x_2^2}$. The proposed D-optimal design is $\xi_1 \times \xi_2$ where

$$
(5.2) \qquad \xi_i = \left\{ \begin{array}{cc} -s_i & s_i \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}
$$

and $s_i = 1$ if $\theta_i \leq \frac{1}{2}$ and, $s_i = \frac{1}{\sqrt{2\theta_i}}$ if $\theta_i \geq \frac{1}{2}$, $i = 1, 2$.

*Example* 5. Consider now when the efficiency function is $\lambda(\theta_1, \theta_2, x_1, x_2) = (\theta_1 - x_1^2)(\theta_2 - x_2^2)$, where $\theta_1 > 1$ and $\theta_2 > 1$. The D-optimal design is $\xi_1 \times \xi_2$ where $\xi_i$ is defined by (5.2) with $s_i = 1$ whenever $\theta_i \geq 3$ and $s_i = \sqrt{\frac{\theta_i}{3}}$ when $1 < \theta_i < 3$.

*Example* 6. Suppose the efficiency function is $\lambda(\theta_1, \theta_2, x_1, x_2) = (\theta_2 - x_2^2)e^{-\theta_1 x_1^2}$, $\theta_2 > 1$ and $0 \leq \theta_1 \leq \frac{1}{2}$. The D-optimal design is $\xi_1 \times \xi_2$, where $\xi_i$ is defined by (5.2) with $s_1 = 1$ and, $s_2 = 1$ if $\theta_2 \geq 3$ and $s_2 = \sqrt{\frac{\theta_2}{3}}$ if $1 < \theta_2 < 3$.

When there are more than two factors in an experiment, an argument similar to the two-factor experiments can be used to construct the D-optimal design. We omit details and state the following result. Let $\chi = \chi_1 \times \chi_2 \times \cdots \times \chi_k$ and let $\xi_i$ be D-optimal for the design problem $(\chi_i, f_i(x_i), \lambda_i(\theta_i, x_i))$, $i = 1, 2, \ldots, k$. Then the D-optimal design for the problem $(\chi, f_1(x_1) \times f_2(x_2) \times \cdots \times f_k(x_k), \prod_{i=1}^k \lambda_i(\theta_i, x_i))$ is $\xi_1 \times \cdots \times \xi_k$ if and only if $2\prod_{i=1}^k d_i(x_i, \xi_i) - \prod_{i=1}^k d_{\lambda_i}(x_i, \xi_i) \leq \prod_{i=1}^k m_i$.

## Acknowledgements

The authors wish to thank the referees for their helpful comments. The research of Wong is partially supported by a NIH research grant R29 AR44177-01A1.

## References

Atkinson, A. C. and Cook, R. D. (1995). D-optimum designs for heteroscedastic linear models, *J. Amer. Statist. Assoc.*, **90**, 204–212.

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design, *J. Amer. Statist. Assoc.*, **54**, 622–654.

Cook, R. D and Nachtsheim, C. J. (1982). Model robust, linear-optimal designs, *Technometrics*, **24**, 49–54.

Dette, H. and Wong, W. K. (1996). Bayesian optimal designs for models with partially specified heteroscedastic structure, *Ann. Statist.*, **24**, 2108–2127.

Eaton, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*, Wiley, New York.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.

Fedorov, V. V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*, Lecture Notes in Statist., No. 125, Springer, New York.

Fedorov, V. V., Montepiedra, G. and Nachtsheim, C. (1998). Optimal design and the model validity range, *J. Statist. Plann. Inference*, **72**, 215–227.

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

Pukelsheim, F. (1993). *Optimal Design of Experiments*, Wiley, New York.

Rafajlowicz, E. and Myszka, W. (1988). Optimum experimental design for a regression on a hypercube-generalization of Hoel's result, *Ann. Inst. Statist. Math.*, **40**, 821–828.

Rafajlowicz, E. and Myszka, W. (1992). When product type experimental design is optimal? Brief survey and new results, *Metrika*, **39**, 321–333.

Wong, W. K. (1994). G-optimal designs for multi-factor experiments with heteroscedastic errors, *J. Statist. Plann. 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.