Optimal Design with Finite Model Validity Range

By

Grace Montepiedra, Dept. of Applied Statistics
School of Statistics, University of Minnesota

Technical Report #586

December 3, 1992

Abstract

December 3, 1992

The construction of optimal designs using a "simplified" response surface model (i.e., we ignore some terms in what is suspected to be the actual response surface model) has been dealt with as far back as Box and Draper(1959). In this paper, we introduce a class of optimality criteria which can be considered as a general version of the "standard" class of optimal design criteria extending to cases when such model inadequacies occur. We define and focus on designs called D_R-optimal designs, which are counterparts of standard D-optimal designs. An equivalence theorem is presented along with examples which illuminate its importance. Validity of the use of the "simplified" model is examined by comparing the results in this paper with "standard" experimental design theory.

Optimal Design with Finite Model Validity Range

Grace Montepiedra

December 3, 1992

1 Introduction

Suppose an experimenter is interested in the relationship between a response variable y and the $k \times 1$ vector of variables, $\mathbf{x}$, which can be controlled. The actual response surface can be written as

$$y_i = \theta^T f_1(\mathbf{x}_i) + \delta^T f_2(\mathbf{x}_i) + \epsilon_i \tag{1}$$

where $\theta \in \mathbf{R}^m, \delta \in \mathbf{R}^p, \mathbf{x}_i = (x_{i1}, \ldots, x_{ik})$ and the $\{\epsilon_i\}$ $(i = 1, \ldots, N)$ are uncorrelated random errors with zero means and variances σ^2. Here let n be the number of distinct support points. For some reason, the investigator decides to use a more parsimonious model (which is typical in practice, since the true model is rarely known)

$$\hat{y}_i(\mathbf{x}_i) = \hat{\theta}^T f_1(\mathbf{x}_i) \tag{2}$$

to describe the relationship. For example, a simplified version of (1) can be used when the number of support points is not sufficient to handle the exact model. Taking into account the possible presence of what is called the contamination function $\delta^T f_2(\mathbf{x}_i)$, this paper attempts to find a design

$$\xi = \{\mathbf{x}_i, p_i\}_{i=1}^n, p_i = \frac{r_i}{N}, \sum_{i=1}^n r_i = N$$

in a specified design region χ which is optimal in the sense of a criterion defined in section 2.

In their pioneering paper, Box and Draper(1959) recognized the danger of the above model misspecification and introduced a bias term in their design optimality criterion. Since then, numerous studies on optimal design for this problem have been published. The approaches basically differ in two aspects: (a) the form of estimation used, and (b) the main estimation goal on which the design criterion is based (e.g., interest might be in the estimation of response surfaces instead of the coefficients).

Much of the work has centered on design criteria for response surfaces. Box and Draper(1959), Draper and Lawrence(1965), Draper and Guttman (1986) and Huber(1975) (the latter using a more general formulation of the response function) used the integrated mean squared error or functions of it as their criteria. Karson, Manson and Hader(1969) introduced minimum bias estimation (MBE) and proposed a design which minimizes the variance of the estimated response. Karson(1970), Khuri and Cornell(1977), Cote, Manson and Hader(1973) and more recently Draper and Sanders(1988) also examined designs to minimize the variance of the estimated response.

Karson and Spruill(1975) were interested in determining the factor levels associated with the maximum response as derived from the assumed model. Atkinson(1970) and Lui and Karson(1980) were concerned with estimating the slope of the response surface. Steinberg(1985) used a Bayesian approach.

The approach to optimality criteria in this paper differs from those previously mentioned: main interest centers on finding "best" estimates for the coefficients $\{\theta_j\}$. Previous work has used design criteria based on the mean squared error of $\hat{\theta}$ (which we abbreviate by MSE($\hat{\theta}$)). The earliest work was by Myers and Lahoda(1975). Their design criterion was the **A**-optimal counterpart for MSE($\hat{\theta}$). Marcus and Sacks(1976), Li(1984), Pesotchinsky(1982) and Notz(1989) assumed more general forms than in model (1) for their contamination function. The first two papers dealt with variations of the **A**-optimality criterion applied to MSE($\hat{\theta}$) and the latter two contained proposed families of MSE($\hat{\theta}$)-based criteria, of which **D**, **G**, and **A**-optimality analogs were special cases. Instead of using least squares, Evans and Manson(1978) examined minimum bias estimates of θ and proposed designs which minimize the variance of such estimates. They used **D**, **G** and **A**-optimal counterparts. The "resistance" or "robustness" of **D**-optimal designs for the one-dimension polynomial regression to the impact of relatively small remainders was considered by Fedorov and Malyutov (1971). **A**, **D** and **G**-optimality criteria

are defined in most optimal design books (e.g., Fedorov, 1972).

In this paper, we estimate θ by least squares under the assumed simplified model and define a general criterion in section 2 based on the mean squared error of the estimate for θ under the true model (1). We present an equivalence theorem for the **D**-optimal counterpart in section 3 and provide numerical examples in section 4. Computational aspects are described in section 5.

2 The Optimality Criterion

Under the assumed model, the mean function $\eta(\mathbf{x}, \theta)$ is

$$\eta(\mathbf{x}, \theta) = \theta^T f_1(\mathbf{x}) \tag{3}$$

A reasonable choice of estimator for θ is

$$\hat{\theta} = \text{Arg}\min_{\theta} \sum_{i=1}^{n} (y_i - \theta^T f_1(\mathbf{x}_i))^2$$

which is the least squares estimator. The usual criteria for "goodness" of designs are based on the variance-covariance matrix of $\hat{\theta}$

$$V = E\{[\hat{\theta} - E(\hat{\theta})][\hat{\theta} - E(\hat{\theta})]^T\}$$

However, since (1) is the suspected correct model, it is better to consider the matrix

$$R = E[(\hat{\theta} - \theta_t)(\hat{\theta} - \theta_t)]$$

(the mean squared error of $\hat{\theta}$; the subscript "t" on θ_t stands for true value) as a "measure" of "goodness" of the estimate $\hat{\theta}$.

It is straightforward to show that

$$R = \sigma^2 \mathbf{M}_{11}^{-1} + \mathbf{M}_{11}^{-1} \mathbf{M}_{12} \delta_t \delta_t^T \mathbf{M}_{21} \mathbf{M}_{11}^{-1} \tag{4}$$

where

$$\mathbf{M}_{\alpha\beta} = \sum_{i=1}^{n} p_i f_\alpha(\mathbf{x}_i) f_\beta(\mathbf{x}_i), \alpha, \beta = 1, 2$$

and δ_t is the true value of δ. We can express (3) as

$$R = V + B$$

where

$$B = [E(\hat{\theta}) - \theta_t][E(\hat{\theta}) - \theta_t]^T$$

is the corresponding bias term. Note that if model (3) is true, R simply reduces to V. In practice an investigator may suspect the existence of the remainder $\delta^T f_2(\mathbf{x})$ and thus nonzero bias matrix B, but will neglect it (perhaps with some hesitation).

Consider the set $\mathbf{R}$ of all such $m \times m$ matrices R and let Ψ be a suitably chosen functional. The optimal (*continuous*) design is defined as

$$\xi^* = \text{Arg} \min_{\xi} \Psi[\mathbf{R}] \tag{5}$$

where ξ is a probability measure whose supporting set belongs to χ. The simplest structure for ξ is of the form:

$$\xi^* = \begin{pmatrix} p_i \\ \mathbf{x}_i \end{pmatrix}, i = 1, \ldots, n$$

where p_i is the proportion of the total number of observations that are allocated to the supporting point $\mathbf{x}_i$. After normalization we have:

$$R(\xi) = \mathrm{M}_{11}^{-1}(\xi) + \mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)B_t\mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi) \tag{6}$$

and

$$N\mathrm{M}_{\alpha\beta}(\xi) = \int_{\chi} f_\alpha(\mathbf{x}) f_\beta^T(\mathbf{x}) \xi(d\mathbf{x}) \tag{7}$$

where N is the total number of observations and $B_t = \sigma^{-2} N \delta_t \delta_t^T = \gamma_t \gamma_t^T$.

Generally, any reasonable *convex* function, $\Psi[R(\xi)]$, can be considered as the optimality criterion. We start with the **D**-optimality criterion:

$$\Psi[R(\xi)] = \ln |R(\xi)| \tag{8}$$

which is a popular choice in applications and also convenient in calculus.

Definition 2.1 *The* D_R*-optimal design* ξ^* *is defined as*

$$\xi^* = \text{Arg} \min_{\xi \in \Xi} \ln |R(\xi)| \tag{9}$$

where Ξ *is the set of all probability measures with supporting points belonging to* χ.

The rest of the paper will focus on this newly defined optimality criterion, and compare it to "standard" experimental design theory.

3 An Equivalence Theorem

We now consider the optimization problem of finding a D_R-optimal design. Let

$$M(\xi) = \int_\chi f(x) f^T(x) \xi(dx) \tag{10}$$

where

$$f(x) = [f_1(x), f_2(x)]^T \tag{11}$$

A general version of the equivalence theorem (Fedorov, 1992) states that, under the following assumptions:

(a) χ is compact and $f(x)$ is continuous on χ,
(b) $\Psi[M(\xi)]$ is a convex function,
(c) $\{\xi : \Psi[M(\xi)] \leq q \leq \infty, \xi \in \Xi\} = \Psi_q \neq \emptyset$,
(d) for any $\bar{\xi} \in \Psi_q$

$$\begin{aligned} \Psi[(1-\alpha)M(\xi^*) + \alpha M(\xi^*)] \;=\; & \Psi[M(\xi^*)] + \\ & \alpha \int_\chi \psi(x, \xi^*) \bar{\xi}(dx) + o(\alpha), \end{aligned}$$

a necessary and sufficient condition for ξ^* to be optimal is fulfillment of the inequality

$$\min_{x \in \chi} \psi(x, \xi^*) \geq 0$$

When $\Psi[M(\xi)] = \ln |R(\xi)|$, then (a)-(c) are satisfied when χ is compact and $f(x)$ is continuous and linearly independent on this compact set.

We now use the above result to prove the following theorem:

Theorem 3.1 *Let χ be a compact set and $f(\mathbf{x})$ is a vector of continuous and linearly independent functions of $\mathbf{x}$. Assume also that δ_t is fixed. Then the following statements are equivalent:*

(i) ξ^* minimizes

$$\Psi[R(\xi)] = \ln|R(\xi)|$$

(ii) ξ^* satisfies the inequality

$$\max_{\mathbf{x}\in\mathbf{X}}[d_1(\mathbf{x},\xi^*) + d_2(\mathbf{x},\xi^*)] \leq m - K(\xi^*)/[K(\xi^*)+1] \tag{12}$$

where

$$d_1(\mathbf{x},\xi^*) = f_1^T(\mathbf{x})\mathrm{M}_{11}^{-1}(\xi^*)f_1(\mathbf{x})$$

and

$$\begin{aligned} d_2(\mathbf{x},\xi^*) \;=\; & \{f_1^T(\mathbf{x})\mathrm{M}_{11}^{-1}(\xi^*)\mathrm{M}_{12}(\xi^*)B_t[\mathrm{M}_{21}(\xi^*)\mathrm{M}_{11}^{-1}(\xi^*)f_1(\mathbf{x}) - \\ & 2f_2(\mathbf{x})]\}/[K(\xi^*)+1] \end{aligned}$$

where $K(\xi^*) = \mathrm{tr}\{B_t\mathrm{M}_{21}(\xi^*)\mathrm{M}_{11}^{-1}(\xi^*)\mathrm{M}_{12}(\xi^*)\}$.

Proof. Conditions (a) and (c) can be proven without much difficulty if one puts

$$\mathrm{M}(\xi) = \begin{pmatrix} \mathrm{M}_{11}(\xi) & \mathrm{M}_{12}(\xi) \\ \mathrm{M}_{21}(\xi) & \mathrm{M}_{22}(\xi) \end{pmatrix}$$

To prove the other two conditions (i.e., (b) and (d)) it is convenient to write $\Psi[R(\xi)]$ in a special form as suggested by the following result from matrix theory:

Let A be a regular $r \times r$ matrix and let F be an $r \times s$ matrix; then

$$|A + FF^T| = |A||I_s + F^TA^{-1}F|$$

Hence if we let $A = \mathrm{M}_{11}^{-1}(\xi)$ and $F = \mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)\gamma_t$ we can write $\Psi[R(\xi)]$ as

$$\Psi[R(\xi)] = S_1(\xi) + S_2(\xi) \tag{13}$$

where

$$S_1(\xi) = \ln|\mathrm{M}_{11}^{-1}(\xi)|$$

and

$$S_2(\xi) = \ln[1 + \gamma_t^T\mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)\gamma_t]$$

Let us now prove condition (b). It is well-known that $S_1(\xi)$, which is the standard **D**-optimal criterion, is a convex function.

Letting

$$\xi = (1-\alpha)\xi^* + \alpha\bar{\xi}$$

then

$$\mathrm{M}(\xi) = (1-\alpha)\mathrm{M}(\xi^*) + \alpha\mathrm{M}(\bar{\xi})$$

and

$$\mathrm{M}_{\alpha\beta}(\xi) = (1-\alpha)\mathrm{M}_{\alpha\beta}(\xi^*) + \alpha\mathrm{M}_{\alpha\beta}(\bar{\xi})$$

Recall the following basic inequality (see, e.g., Fedorov, 1972):

If the matrix A_j has dimension $r \times s$ and B_j is an $s \times s$ positive-definite matrix (j = 1,2) then

$$[(1-\alpha)A_1 + \alpha A_2][(1-\alpha)B_1 + \alpha B_2]^{-1}[(1-\alpha)A_1^T + \alpha A_2^T] \leq (1-\alpha)A_1B_1^{-1}A_1^T + \alpha A_2B_2^{-1}A_2^T$$

So if we let $A_1 = \gamma_t^T\mathrm{M}_{21}(\xi^*), A_2 = \gamma_t^T\mathrm{M}_{21}(\bar{\xi}), B_1 = \mathrm{M}_{11}(\xi^*)$ and $B_2 = \mathrm{M}_{11}(\bar{\xi})$ then, using the above inequality, we have:

$$\begin{aligned}\gamma_t^T\mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)\gamma_t \;\leq\; & (1-\alpha)\gamma_t^T\mathrm{M}_{21}(\xi^*)\mathrm{M}_{11}^{-1}(\xi_*)\mathrm{M}_{12}(\xi^*)\gamma_t \\ & +\alpha\gamma_t^T\mathrm{M}_{21}(\bar{\xi})\mathrm{M}_{11}^{-1}(\bar{\xi})\mathrm{M}_{12}(\bar{\xi})\gamma_t\end{aligned}$$

This means that $S_2(\xi)$ is also a convex function. Note that the sum of convex functions is also convex, so we obtain the desired result that $\ln|R(\xi)|$ is convex.

We now prove condition (d) by finding $\psi(\mathbf{x}, \xi^*)$ directly. We first find the derivative of $S_2(\xi)$ with respect to α,

$$\begin{aligned}\frac{d}{d\alpha}S_2(\xi) &= \{\frac{d}{d\alpha}[1 + \gamma_t^T\mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)\gamma_t]\}/[K(\xi)+1] \\ &= \{\gamma_t^T[\mathrm{M}_{21}(\bar{\xi}) - \mathrm{M}_{21}(\xi^*)]\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)\gamma_t - \\ &\quad \gamma_t^T\mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi)[\mathrm{M}_{11}(\bar{\xi}) - \mathrm{M}_{11}(\xi^*)]\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)\gamma_t + \\ &\quad \gamma_t^T\mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi)[\mathrm{M}_{12}(\bar{\xi}) - \mathrm{M}_{12}(\xi^*)]\gamma_t\}/[K(\xi)+1],\end{aligned}$$

where $K(\xi) = \gamma_t^T\mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)\gamma_t$.

Evaluating the last expression as $\alpha \to 0$ and simplifying further we obtain:

$$\frac{d}{d\alpha}S_2(\xi)|_{\alpha\to 0} = \{\gamma_t^T[2\mathrm{M}_{21}(\bar{\xi}) - \mathrm{M}_{21}(\xi^*) - \mathrm{M}_{21}(\xi^*) \mathrm{M}_{11}^{-1}(\xi^*)\mathrm{M}_{11}(\bar{\xi})]\mathrm{M}_{11}^{-1}(\xi^*)\mathrm{M}_{12}(\xi^*)\gamma_t\}/[K(\xi^*)+1]. \quad (14)$$

By using the definition of $\mathrm{M}_{\alpha\beta}(\xi)$ given in (6), we can rewrite (14) as:

$$\frac{d}{d\alpha}S_2(\xi)|_{\alpha\to 0} = \int_{\chi} \psi_2(\mathbf{x},\xi^*)\bar{\xi}(d\mathbf{x}), \quad (15)$$

where

$$-\psi_2(\mathbf{x},\xi^*) = \{f_1^T(\mathbf{x})\mathrm{M}_{11}^{-1}(\xi^*)\mathrm{M}_{12}(\xi^*)B_t[\mathrm{M}_{21}(\xi^*)\mathrm{M}_{11}^{-1}(\xi^*)f_1(\mathbf{x}) - 2f_2(\mathbf{x})] + K(\xi^*)\}/[K(\xi^*)+1]. \quad (16)$$

Also, from standard theory

$$\frac{d}{d\alpha}S_1(\xi)|_{\alpha\to 0} = \int_{\chi} \psi_1(\mathbf{x},\xi^*)\bar{\xi}(d\mathbf{x}), \quad (17)$$

where

$$-\psi_1(\mathbf{x},\xi^*) = f_1^T(\mathbf{x})\mathrm{M}_{11}^{-1}(\xi^*)f_1(\mathbf{x}) - m. \quad (18)$$

So from equations (13), (15) and (17), we get

$$\frac{d}{d\alpha}\Psi[R(\xi)]|_{\alpha\to 0} = \int_{\chi} \psi(\mathbf{x},\xi^*)\bar{\xi}(d\mathbf{x})$$

where

$$\psi(\mathbf{x},\xi^*) = \psi_1(\mathbf{x},\xi^*) + \psi_2(\mathbf{x},\xi^*). \quad (19)$$

From the general equivalence theorem, ξ^* also fulfills the inequality

$$\min_{\mathbf{x}\in\chi} \psi(\mathbf{x},\xi^*) \geq 0$$

or

$$\max_{\mathbf{x}\in\chi} -\psi(\mathbf{x},\xi^*) \leq 0$$

which gives us (ii). This proves the theorem.

Note 1. The optimization problem (8) can actually be considered as a special case of optimal design for moving regression studied by Fedorov and

Nachtsheim(1992). The emphasis in their paper, however, is very different and not directly applicable to our problem. In addition, some results which were not discussed there are thoroughly proven in this paper (for instance, the convexity of $\ln|R(\xi)|$).

Note 2. At first glance, the above result is useful only when δ_t is known, which is not a realistic scenario. But in a number of practical cases the use of Theorem 3.1 helps to find some features in optimal designs which are independent of γ_t.

All results can be generalized for the Bayesian case, replacing B_t everywhere by $B = \int \gamma\gamma^T p_0(\gamma)d\gamma$, where $p_0(\gamma)$ is a prior distribution for γ.

Note 3. Recall that if we use the standard least squares estimators for the simplified model then

$$\ln|R(\xi)| = \ln|\mathrm{M}_{11}^{-1}(\xi)| + \ln[1 + \mathrm{tr}\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)B_t\mathrm{M}_{21}(\xi)] \qquad (20)$$

where $B_t = \sigma^{-2}N\delta_t\delta_t^T = \gamma_t\gamma_t^T$.

On the other hand, if we consider the standard least squares estimators for the complete model (i.e., the contamination function is included in the analysis) then

$$\ln|R(\xi)| = \ln|\mathrm{M}_{11}^{-1}(\xi)| + \ln[1 + \mathrm{tr}\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)H^{-1}(\xi)\mathrm{M}_{21}(\xi)] \qquad (21)$$

where $\mathrm{H}(\xi) = \mathrm{M}_{22}(\xi) - \mathrm{M}_{21}(\xi)\mathrm{M}_{11}^{-1}(\xi)\mathrm{M}_{12}(\xi)$.

Comparing (20) and (21) or matrices B and $\mathrm{H}(\xi)$, a practitioner can evaluate boundaries for N and δ_t for which a "simplified" model can guarantee better estimates than an "exact" model.

For instance, when $p = 1$, one must compare $\gamma^2 = N^2\delta^2/\sigma^2$, and $\rho^2 = \mathrm{H}(\xi)$ which is independent of of N and σ^2. Comparison of γ^{-2} and ρ^2 shows that the "simplified" model can provide better estimators when $\sigma^2/N^2\delta^2 \geq \rho^2$. One may call $\gamma^{-1} = \sqrt{\sigma^2/N^2\delta^2}$ the "model validity range" (see also example 1 for additional comments).

4 Examples

In many relatively simple regression problems, the equivalence theorem allows the construction of optimal designs using semi-intuitive considerations, for example, such as using symmetry or examining the form of the response

surface. In the following examples, the number of observations and weights are predetermined and the exact values of the design points are determined to ensure that $\max_{\mathbf{x}\in\chi}[d_1(\mathbf{x},\xi) + d_2(\mathbf{x},\xi)] + K(\xi)/[K(\xi)+1] \leq m$, hence giving us an optimal design under the defined criterion.

EXAMPLE 1. Suppose we have only one predictor variable x. Let the assumed model be

$$\eta(x,\theta) = \theta^T f_1(x) = \theta_1 + \theta_2 x$$

with the contamination function being $\delta^T f_2(x) = \delta x^2$. Design points are chosen in the region χ which coincides with the interval $[-1,1]$.

From symmetry considerations it is natural to assume that one of the optimal designs belongs to the set of two-point designs of the form:

$$\xi = \left\{ \begin{array}{cc} \bar{x} & -\bar{x} \\ 1/2 & 1/2 \end{array} \right\}$$

The information matrix for the complete model will have the submatrices:

$$\mathrm{M}_{11}(\xi) = \left[\begin{array}{cc} 1 & 0 \\ 0 & \bar{x}^2 \end{array} \right]$$

and

$$\mathrm{M}_{12}(\xi) = \left[\begin{array}{c} \bar{x}^2 \\ 0 \end{array} \right]$$

Hence, with $f_1^T(x) = (1,x), f_2^T(x) = x^2, \theta^T = (\theta_1,\theta_2)$ and scalar δ then it can be shown that

$$d_1(x,\xi) + d_2(x,\xi) + K(\xi)/[K(\xi)+1] = 1 + \frac{2\bar{x}^4\gamma^2}{1+\bar{x}^4\gamma^2} + x^2[\bar{x}^{-2} - \frac{2\bar{x}^2\gamma^2}{1+\bar{x}^4\gamma^2}] \quad (22)$$

where $\gamma^2 = N^2\delta^2/\sigma^2$. Using the equivalence theorem, in order for ξ to be an optimal design, the inequality

$$\max_{\mathbf{x}\in\chi}[d_1(x,\xi) + d_2(x,\xi)] + K(\xi)/[K(\xi)+1] \leq m = 2$$

must be satisfied (which is just inequality (12)).

If we choose $\bar{x} = \min\{1, 1/\sqrt{|\gamma|}\}$ then the above inequality is true. Hence, ξ for this choice of $\bar{x}$ is an optimal design. A more general result is given by

Notz (1989), which considers polynomial contamination functions of higher order.

The considered example emphasizes that model validity is relative and depends upon variance of the error, the number of observations, the value of δ, and the behavior of the neglected basis function $f_2(\mathbf{x})$ (x^2 in our case).

Instead of using a two-point optimal design for the above regression problem, consider designs which belong to the set of three-point designs of the form:

$$\xi = \begin{Bmatrix} -\bar{x} & 0 & \bar{x} \\ p & 1-2p & p \end{Bmatrix}$$

It can be verified that

$$\mathrm{M}_{11}(\xi) = \begin{bmatrix} 1 & 0 \\ 0 & 2p\bar{x}^2 \end{bmatrix}$$

and

$$\mathrm{M}_{12}(\xi) = \begin{bmatrix} 2p\bar{x}^2 \\ 0 \end{bmatrix}$$

so that

$$d_1(x,\xi) + d_2(x,\xi) = 1 + \frac{8p^2\bar{x}^4\gamma^2}{1+4p^2\bar{x}^4\gamma^2} + x^2\left[\frac{1}{2p\bar{x}^2} - \frac{4p\bar{x}^2\gamma^2}{1+4p^2\bar{x}^4\gamma^2}\right]$$

which can also be written as

$$d_1(x,\xi) + d_2(x,\xi) = 1 + \frac{2\bar{z}^4\gamma^2}{1+\bar{z}^4\gamma^2} + x^2\left[\bar{z}^{-2} - \frac{2\bar{z}^2\gamma^2}{1+\bar{z}^4\gamma^2}\right] \qquad (23)$$

where $\bar{z}^2 = 2p\bar{x}^2$.

Note that (23) is of the same form as (22) with $\bar{x}^2$ replaced by $\bar{z}^2$. This implies that the value of $\bar{x}$ which satisfies (12) is $\sqrt{2p}\bar{x} = \bar{z} = \min\{1, 1/\sqrt{|\gamma|}\}$.

Under the constraint that $0 < p < 1/2$, a three-point design is a candidate for an optimal design if $|\gamma| > 1$. If we choose the design points to be $\{-1, 0, 1\}$, then p should be $1/\sqrt{|\gamma|}$. Recall that a standard **D**-optimal design for quadratic regression also has $\{-1, 0, 1\}$ for design points but with equal weights on each point. Thus the structure of the optimal design is determined by the parameter γ^{-1} and therefore the introduction of the "model validity range" in the previous section looks reasonable.

EXAMPLE 2. Let us consider the case when we have two predictors x_1 and x_2. The assumed model is the linear case

$$\eta(\mathbf{x}, \theta) = \theta_1 + \theta_2 x_1 + \theta_3 x_2$$

so that $\theta^T = (\theta_1, \theta_2, \theta_3)$ and $f_1^T(\mathbf{x}) = (1, x_1, x_2)$ and the contamination function contains the interaction term:

$$\delta^T f_2(\mathbf{x}) = \delta x_1 x_2$$

so that δ is scalar and $f_2(\mathbf{x}) = (x_1 x_2)$.

The design space χ will be the square:

$$\{(x_1, x_2) : -1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1\}$$

Again, by symmetry considerations, an optimal design is expected to be a four-point design of the following form:

$$\xi = \left\{ \begin{array}{cccc} (\bar{x}_1, \bar{x}_2) & (\bar{x}_1, -\bar{x}_2) & (-\bar{x}_1, \bar{x}_2) & (-\bar{x}_1, -\bar{x}_2) \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array} \right\}$$

According to this assumed design, we have

$$\mathrm{M}_{11}(\xi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \bar{x}_1^2 & 0 \\ 0 & 0 & \bar{x}_2^2 \end{bmatrix}$$

and

$$\mathrm{M}_{12}(\xi) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

In other words, $x_1 x_2$ is orthogonal to x_1 and x_2 (and this is a commonly used fact) on the set of supporting points. Direct calculation gives:

$$d_1(\mathbf{x}, \xi) + d_2(\mathbf{x}, \xi) = 1 + x_1^2 \bar{x}_1^{-2} + x_2^2 \bar{x}_2^{-2}$$

Hence, for inequality (12) to be satisfied, we let $\bar{x}_1 = 1$ and $\bar{x}_2 = 1$ so that ξ is an optimal design for such a choice of design points. Note that $d_1(\mathbf{x}, \xi) + d_2(\mathbf{x}, \xi)$ for this example does not depend on γ_t, giving an optimal design which is independent of the value of γ_t. Also, note that our criterion reduces to the "standard" **D**-optimal criterion so that the optimal design here is simply the "standard" **D**-optimal design for a linear regression with two predictor variables.

5 Computational Method for Constructing D_R-Optimal Designs

For more complicated problems, analytic solutions for finding optimal designs may not be available. A numerical procedure for obtaining such designs would be useful.

Fedorov(1986) suggested a first-order iterative algorithm which can be used for a general set of optimality criteria. Convergence is assured, provided certain basic assumptions are met. The iterative method presented here for finding the D_R-optimal design is a straightforward application of this algorithm.

On the $(s+1)$st iteration, we construct

$$\xi_{s+1} = (1-\alpha_s)\xi_s + \alpha_s \xi(\mathbf{x}_s)$$

where $\xi(\mathbf{x}_s)$ is the design degenerate at point $\mathbf{x}_s$.

If the iteration deletes the "least informative" point from the current design, a backward procedure is performed. If the iteration includes a new, "most informative" point from the current design, then a forward procedure is performed. The selection of $\mathbf{x}_s$ and α_s in each iteration depends on whether it is a forward or a backward procedure.

(1) For the forward procedure, choose:

$$\mathbf{x}_s = \mathbf{x}_s^+ = \text{Arg}\min_{\mathbf{x}\in X} \phi(\mathbf{x}, \xi_s)$$

$$\alpha_s = v_s$$

where

$$\begin{aligned} -\phi(\mathbf{x}, \xi_s) &= d_1(\mathbf{x}, \xi_s) + \{f_1^T(\mathbf{x})\mathrm{M}_{11}^{-1}(\xi_s)\mathrm{M}_{12}(\xi_s)B_t[\mathrm{M}_{21}(\xi_s)\mathrm{M}_{11}^{-1}(\xi_s)f_1(\mathbf{x}) \\ &\quad -2f_2(\mathbf{x})]\}/[K(\xi_s)+1] \end{aligned}$$

(2) For the backward procedure, choose:

$$\mathbf{x}_s = \mathbf{x}_s^- = \text{Arg}\max_{\mathbf{x}\in X_s} \phi(\mathbf{x}, \xi_s)$$

$$\alpha_s = \begin{cases} -v_s, & p_j^s \geq v_s \\ -p_s^s/(1-p_s^s), & p_s^s < v_s \end{cases}$$

where χ_s is the set of design points in the current design and $p_s^s = p(\mathbf{x}_s^-)$ is a weight for a point $\mathbf{x}_s^-$.

Three possible choices for $\{v_s\}$ are:

(a) $v_s = \frac{1}{n_0+s}, s = 1, 2, \ldots; n_0$ is a number of supporting points in an initial design

(b) v_s is determined by the steepest descent method, giving the largest decrease in the objective function in the direction of $\xi(\mathbf{x}_s)$.

(c) $v_s \equiv C_0$ where C_0 is a user-specified small constant.

Note 1. The number of steps for the forward and backward procedures are defined by the user.

Note 2. For a stopping rule, computations are terminated after a given number of iterations is achieved or if, in the forward procedure,

$$\frac{|\phi(\mathbf{x}_s^+)|}{m} < \epsilon,$$

where ϵ is user-specified.

Note 3. Amendments in the algorithm can be made which allows for the merging of neighboring supporting points and deletion of supporting points with small weights (see Fedorov, 1986 for further details).

Acknowledgement

I am deeply grateful to Professor V. V. Fedorov for his guidance and support throughout this project. I also wish to thank Professors Kinley Larntz and Kathryn Chaloner for their helpful comments and suggestions.

References

Atkinson, A.C. (1970). The Design of Experiments to Estimate the Slope of a Response Surface. *Biometrika 57*, 319-328.

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.

Cote, R., Manson, A.R., and Hader, R.J. (1973). Minimum Bias Approximation of a General Regression Model. *J. Amer. Statist. Assoc. 68*, 633-638.

Draper, N.R. and Guttman, I. (1986). Response Surface Designs in Flexible Regions. *J. Amer. Statist. Assoc. 81*, 1089-1094.

Draper, N.R. and Lawrence, W.E. (1965). Designs which Minimize Model Inadequacies: Cuboidal Regions of Interest. *Biometrika 52*, 111-118.

Draper, N.R. and Sanders, (1988). Designs for Minimum Bias Estimation. *Technometrics 30*, 319-325.

Evans, J.W. and Manson, A.R. (1978). Optimal Experimental Designs in Two Dimensions Using Minimum Bias Estimation. *J. Amer. Statist. Assoc. 73*, 171-176.

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

Fedorov, V.V. (1986) Optimal Design of Experiments: Numerical Methods. WP-86-55. Laxenburg, Austria, International Institute for Applied Systems Analysis.

Fedorov, V.V. (1992). Optimal Design Construction with Constraints I, Technical Report No. 573, University of Minnesota.

Fedorov, V.V. and Malyutov, M.B. (1971) Possible Model Inadequacy in Experimental Design, Moscow State University, Laboratory of Statistical Methods, *18* pp. 1-14.

Fedorov, V.V. and Nachtsheim, C.J. (1992). Optimal Experimental Design for Moving Regression Problems, Working Paper 92-6, University of Minnesota.

Huber, P. (1975). *Robustness and Designs. A Survey of Statistical Design and Linear Models.* (ed. Srivastava) 287-303. North Holland, Amsterdam.

Karson, M.J. (1970). Design Criterion for Minimum Bias Estimation of Response Surfaces. *J. Amer. Statist. Assoc. 65*, 1565-1572.

Karson, M.J., Manson, A.R., and Hader, R.J. (1969). Minimum Bias Estimation and Experimental Design for Response Surfaces. *Technometrics 11*, 461-475.

Karson, M.J. and Spruill, M.L. (1975). Design Criteria and Minimum Bias Estimation. *Commun. Statist.-Theor. Meth. 4*, 339-355.

Khuri, A.I. and Cornell, J.A. (1977). Secondary Design Considerations for Minimum Bias Estimation. *Commun. Statist.-Theor. Meth. 6*, 631-646.

Li, K.-C. (1984). Robust Regression Designs When the Design Space Consists of Finitely Many Points. *Annals of Statistics 12*, 511-526.

Liu, W.C. and Karson, M.J. (1980). Minimum Bias Estimation and Designs for the Slope of a Response Surface. *Proceedings of the Business and Economic Statistics Section*, ASA, 415-420.

Marcus, M.B. and Sacks, J. (1976). Robust Designs for Regression Problems. *Statistical Decision Theory and Related Topics II.* 245-268.

Myers, R.H. and Lahoda, S.J. (1975). A Generalization of the Response Surface Mean Square Error Criterion with a Specific Application to the Slope. *Technometrics 17*, 481-486.

Notz, W.I. (1989). Optimal Designs for Regression Models with Possible Bias. *JSPI 22*, 43-54.

Pesotchinsky, L. (1982). Optimal Robust Designs: Linear Regression in R^k. *Annals of Statistics 10*, 511-526.

Silvey, S.D. (1980). Optimal Designs. Chapman and Hall, London.

Steinberg, D.M. (1985). Model Robust Response Surface Designs: Scaling Two-level Factorials. *Biometrika 72*, 513-526.