A Note on Approximate D-optimal Designs for $G \times 2^k$

by

Yong B. Lim $^1$, W. J. Studden $^2$ and H. P. Wynn $^3$
Purdue University and Imperial College

Technical Report #85-12

Department of Statistics
Purdue University

August 1985

$^1$Research supported by a David Ross Grant from Purdue University
$^2$Research supported by NSF Grant No. MCS-8200631
$^3$Research supported by NSF Grant No. MCS-8410183
ABSTRACT

Designs are considered for situations where the mean response consists of a general model together with any number of 2 level factors and suitable interactions. The D-optimal criterion is shown to be equivalent to a type of weighted model selection. Two examples are given.

Kew Words: G×2^k regression model, weighted model selection, D-optimality.

AMS 1980 subject classification: Primary 62K05, 62J05.
A Note on Approximate D-Optimal Designs for $G \times 2^k$

Yong B. Lim¹, W. J. Studden² and H. P. Wynn³

1. Introduction.

In experimental designs, we frequently have a response variable $y$ which depends on both qualitative and quantitative factors. The qualitative factors may be, for example, type of fertilizer, method of treatment, sex of patients, type of drug, curing condition, etc. The quantitative factors may be temperatures, or the amount of concentration, etc. In addition, the levels of some of the quantitative factors may be reduced to two levels, thus making them, in effect, qualitative factors. The reason for this may be for cost considerations, for ease of experimentation or to conduct preliminary studies. In the following all of the qualitative factors will have only two levels.

The purpose of this note is to show that one can leave part of the model more general and still analyze the situation from a D-optimal viewpoint; and that the analysis is equivalent to a type of weighted model selection which has a Kiefer-Wolfowitz equivalence theorem.

The general part of the model may consist of those factors of more interest to the experimenter.

2. Definitions and Formulation.

The basic criterion of design optimality which we shall use here is that of D-optimality developed largely by Kiefer (1959, 1961) and Kiefer and Wolfowitz (1959, 1960). It is assumed that for each point $z$ in some multidimension factor space a random variable or response $Y(z)$ can be observed. The variable $Y(z)$ has

¹Research supported by a David Ross Grant from Purdue University
²Research supported by National Science Foundation Grant No. MCS-8200081
³Research supported by National Science Foundation Grant No. MCS-8410183
expected value

\[ EY(z) = \sum \theta_i f_i = \theta' f(z) \]

and

\[ \text{var}Y(z) = \sigma^2 \]

where \( f(z) \) is a \( k \times 1 \) column vector of known functions \( f_i(z), i = 1, \ldots, k \), and \( \theta \) is a \( k \times 1 \) vector of unknown parameters. Uncorrelated observations are taken at \( z_1, z_2, \ldots, z_N \) (not necessarily distinct). If the least squares estimated of \( \theta \) are used, they have covariance matrix given by \( (\sigma^2/N) M^{-1}(\xi) \) where \( M(\xi) \) has elements

\[ m_{ij}(\xi) = \int f_i(z)f_j(z) d\xi(z) \quad (2.1) \]

and \( \xi(z) \) has mass \( 1/N \) at each point \( z_i \). We shall assume that our design measure \( \xi \) is an arbitrary probability measure. As usual, some approximation would be necessary in practice.

The D-optimal criterion is to choose \( \xi \) to maximize the determinant \( |M(\xi)| \). This is known, by the celebrated Kiefer-Wolfowitz theorem, to be equivalent to minimizing the supremum over \( z \) of

\[ d(z, \xi) = f'(z)M^{-1}(\xi)f(z). \quad (2.2) \]

In this general setting we let \( z = (x, y_1, \ldots, y_m) \). Here \( x \), which may be multidimensional or more general, consists of the more important factor(s) while \( y_1, \ldots, y_m \) are each set at two levels \( \pm 1 \). For the \( x \) factor we consider a general model, denoted here by \( G \), which involves \( g + 1 \) linearly independent functions
\( h_0(x), h_1(x), \ldots, h_g(x) \). The model G corresponds to or incorporates the variables of main interest. For example if the experiments are categorized by height, weight and sex, and height = x is the important variable, one might use \( h_i(x) = x^i \), \( i=0, \ldots, g \). Then take weight at two levels, light and heavy and sex at two levels, male and female. If height and weight are important then we may use \( x = (x_1, x_2) \) and \( h_i(x), i=0,1, \ldots, g \), could consist of a general polynomial model in two variables, leaving sex at two levels. The functions \( h_i \) should really be considered as function on \( z = (x, y_1, \ldots, y_k) \) as they will constitute part of our more general model involving \( f_1, \ldots, f_k \).

In addition to these \( h_i(x) \) we allow \( m \) linear variables \( y_1, \ldots, y_m \). To the functions \( h_i \) we will eventually add \( m + \binom{m}{2} \) terms \( y_i y_j, 1 \leq i < j \leq m \) where each \( y_i \in [-1,1] \). These will be eventually restricted, without loss of generality, to take on only the values \( \pm 1 \).

To describe the interaction between the two parts of the model it is convenient to introduce a number of submodels of G. This will be used later to describe the equivalence with the weighted model selection formulation in section 4. Thus let \( G_j, j = 1,2, \ldots, r \) denote distinct submodels of G, i.e. each \( G_j \) consists of a distinct subset of the functions \( h_0, h_1, \ldots, h_g \). The interactions will be of the form \( y_i \times G_j \), consisting of \( y_i \) times all the functions in \( G_j \). For later analysis it will be sufficient to note how many \( y_i \) occur with each specific \( G_j \). This we denote by \( s_j, j=1, \ldots, r \). That is, there are \( s_1 \) terms like \( y_1 G_1, s_2 \) of the type \( y_1 G_2, \) etc. For convenience we assume \( f_0 = 1 \) is contained in each \( G_i \). Then the totality of our functions \( f_i, i = 1, \ldots, k \) consists of

\[
G = y_1 G_{i_1}, y_2 G_{i_2}, \ldots, y_m G_{i_m}
\]

\[
y_{i_1} y_{i_2} \quad 1 \leq i < j \leq m
\]

To illustrate, (2.3) might consist of \( 1, x, x^2, x^3, y_1, y_1 x, y_2, y_2 x, y_3, y_3 x, y_3 x^2, y_4, y_4 x^2 \), and \( y_i y_j, 1 \leq i < j \leq 4 \). Here, \( G = \{1, x, x^2, x^3\} \), \( G_1 = \{1, x\} \), \( G_2 = \{1, x, x^2\} \), \( G_3 = \{1, x^2\} \).
3. **Symmetry Reduction**

In analysing the above situation, all of the functions listed in (2.3) are viewed as functions on \( z = (x, y_1, \ldots, y_k) \). An arbitrary design is a probability measure on \( z \). A symmetry argument will show the following.

**Lemma 3.1**

A solution \( \xi \) to the D-optimal design problem has each \( y_i \) on \( \pm 1 \) with probability \( 1/2 \) and a measure \( \mu \) on \( x \); the full \( \xi \) being a product of these measures. The D-optimal design problem reduces to maximizing

\[
\Delta = |M_G(\mu)| \prod_{i=1}^{r} |M_i(\mu)|^{\xi_i} \tag{3.1}
\]

where \( M_G \) denotes the information matrix corresponding to \( G \) and \( M_i \) to \( G_i \), \( i=1,2,\ldots,r \).

**Proof**

We consider the functions in (2.3) in blocks as indicated. Thus \( G \) is one block. Each \( y_j G_{ij} , j = 1,\ldots,m, \) is a block and all the terms \( y_i y_j , i < j \) are in a separate block. A symmetry argument on each \( y_i \) shows that the D-optimal design must have these blocks orthogonal. That is, integrals of products of functions from different blocks are zero. To show that \( y_i \) must be \( \pm 1 \) we observe that \( d(z,\xi) \) will split according to the blocks. The terms with \( y_1 G_{ii} \) separate out and \( y_1^2 \) factors out with a nonnegative coefficient. Terms from \( y_j y_j \) also will enter with \( y_1^2 \). Thus the supremum over \( y_1 \) must be at \( y_1 = \pm 1 \). Similarly for \( y_2 \), etc. The symmetry on \( y_i \) then gives weight \( 1/2 \) to \( y_i = \pm 1 \). We may then assume without loss of generality that \( \xi \) is of the form \( \xi = \mu \times \rho \) where \( \mu \) is on \( x \) and \( \rho \) has equal mass on the \( 2^m \) points \( (y_1, \ldots, y_m) \) with each \( y_i = \pm 1 \). (The measure \( \xi \) is only constrained by certain moment conditions and may differ from the one described).

If the design \( \xi \) is as described above then the determinant of the overall information matrix is seen to be equal to \( \Delta \) in (3.1).
4. **Weighted Model Selection**

The situation resulting in $\Delta$ given in (3.1) is exactly the same as certain analysis of Atkinson and Cox (1974) and Läuter (1974). In these situations a number of models (in our case $G$ and $G_i$, $i = 1, 2, \ldots, r$) are tentatively being considered. (Here $y_1, \ldots, y_m$ are not present.) We have assumed that the models $G_i$ are all distinct in our formulation. One of the $G_i$ may, however, be equal to $G$. In this case we could omit $G$, if desired, and replace the corresponding $s_i$ by $s_i + 1$. For convenience of notation we shall set $G = G_0$ and assume it is distinct from the other $G_i$. To design an experiment to distinguish between the models consideration is given directly to the weighted D-optimal criterion $\Delta$. At least two methods for choosing the weights seem to present themselves. The first is to make them proportional to some prior likelihood of the different models. The other is to make them proportional to the inverse of the dimension of the model.

For the criterion $\Delta$, a Kiefer-Wolfowitz equivalence theorem has been described in Läuter (1974) and Atkinson and Cox (1974). Thus the following are equivalent:

\[ \mu^* \text{maximizes } \Delta(\mu) \]

\[ \mu^* \text{minimizes sup} \sum_{x}^{m} s_i d_i(x, \mu) \]

\[ \Sigma s_i d_i(x, \mu^*) = \Sigma k_i \]

where $k_i$ is the dimension of $G_i$.

The analysis of $\Delta$ in previous sections gives another interpretation of the conditions in (4.1).

5. **Examples.**

**Example 1.**

Consider for $G$ the simple case $G = \{1, x, \ldots, x^m\}$ with $x \in [-1, 1]$ and let $G_i = \{1, x, \ldots, x^i\}$ $i = 0, 1, \ldots, m$. Thus in our model we have $x \in [-1, 1]$ and $m$ variables $y_1, \ldots, y_m$ on $\pm 1$, with 1st order interactions between the $y_i$. The subset $G_i$ appears as interactions with $s_i$ of the $y$ terms and
\[ \Delta(\mu) = |M_m(\mu)| \times \prod_{i=1}^{m} |M_i(\mu)|^{p_i} \] (5.1)

The minimization of (5.1) is given in Studden and Lau (1984) using canonical moments. We use the notation from that paper to describe the solution here. One can show that, except for some constants independent of \( \mu \),

\[ |M_k| = \prod_{i=1}^{k} (s_{2i-1}s_{2i})^{k+1-i} \] (5.2)

where \( s_i = q_{i-1} p_i, \ i \geq 1, \ q_0 = 1, \ 0 \leq p_i \leq 1 \) and \( q_i = 1 - p_i \). The quantities \( p_1, p_2, \ldots \) give a convenient parameterization to the design part \( \mu \). Substituting (5.2) into (5.1) the general model \( G \times 2^k \) can be solved, at least in terms of the \( p_i \). For \( m=2 \) the quantity \( \Delta \) reduces to

\[ \Delta = (p_1q_1p_2)^{s_1}[p_1q_1p_2]^2(q_2p_3q_3p_4)^{s_2+1} \]

Maximizing \( \Delta \) in terms of \( p_1, p_2, p_3, p_4 \) we find \( p_1 = p_3 = 1/2, \ p_4 = 1 \) and

\[ p_2 = \frac{2(s_2+1) + s_1}{3(s_2+1) + s_1} \] (5.3)

The resulting measure \( \mu^* \) can be shown to give weight \( p_2/2 \) to \( \pm 1 \) and \( 1-p_2 \) to zero.

For \( m=3 \) it can be shown that the measure \( \mu^* \) has weight \( \rho/2 \) on \( \pm 1 \) and \( (1-\rho)/2 \) on \( \pm t \) where \( t = (p_2q_4)^{1/2}, \ \rho = p_2p_4/(q_2+p_2p_4) \) and

\[ p_2 = \frac{3(s_3+1)+2s_2+s_1}{5(s_3+1)+3s_2+s_1} \quad p_4 = \frac{2(s_3+1)+s_2}{3(s_3+1)+s_2} \]
We remark that the usual D-optimal design for polynomial regression corresponds to $s_i = 0$, $i = 1, 2, \ldots, m$. This same solution results if these are zero only for $i$ up to $m-1$. In this case we have a "full product" model for some of factors.

**Example 2**

Consider for $G$ the quadratic polynomials on the $k$-cube,

$$G = \left\{ \begin{array}{c} x_1 & \cdots & x_k \\ 1, x_1^2 & \cdots & x_k^2 \\ x_1 x_2 & \cdots & x_{k-1} x_k \end{array} \right\}, \quad x_i \in [-1,1], i=1, \ldots, k, k \geq 2$$

and

$$G_1 = \{1, x_1, \ldots, x_k\}.$$ 

Thus we have the quadratic polynomials on the $k$-cube, $m$ variables $y_1, \ldots, y_m$, all the 1st order interactions between $x_i$ and $y_j$ and interactions between $y_j$ in our model. The subset $G_1$ interacts with each of the $y$ terms and

$$\Delta(\mu) = |M_G(\mu)| \cdot |M_1(\mu)|^m.$$ 

Note that the model is invariant under the group $H$ of sign changes and permutations of $x_i$'s, $i=1, \ldots, k$. By the invariance theorem ([3]) there exists a symmetric D-optimal design $\mu^*$ under $H$. By the same argument as one in the proof of Lemma 3.1 $d(z, \xi^*)$ is a quartic function of $x_i$ with the positive coefficient of $x_i^4$ and symmetric w.r.t. $x_i$. So $d(z, \xi^*)$ can be maximized at $x_i = \pm 1$ or 0. Thus we restrict $\mu$ to a symmetric design on $E$, $E = \{x: |x| = 0 \text{ or } 1\}$. Let

$$u = \int x_i^2 \mu(dx)$$
and

\[
v = \int \pi^2 x_1^2 x_2^2 \mu(dx).
\]

Noting that \( \mu \) is a symmetric design on \( E \), we get

\[
M_G(\mu) = \begin{bmatrix}
    uI_k & 0 & 0 \\
    0 & (u-v)I_{k-1} + vI & 0 \\
    0 & 0 & \frac{vI_{k-1}}{2}
\end{bmatrix}
\]

and

\[
M_I(\mu) = \begin{bmatrix}
    1 & 0 \\
    0 & uI_k
\end{bmatrix}
\]

where \( 1 \) is the \( k \times 1 \) vector of ones and \( I \) is the identity matrix. So

\[
\Delta(\mu) = \left[ u^k \cdot (u-v)^{k-1} \cdot (u+(k-1)v-ku^2) \right] \cdot v \cdot \frac{k(k-1)}{2} \cdot (u^k)^m
\]

\[
= u^{k(m+1)} \cdot v^{k(k-1)/2} \cdot (u-v)^{k-1} \cdot (u+(k-1)v-ku^2)
\]

(5.4)

Simple algebra shows that \( \Delta(\mu) \) is maximized at

\[
u^* = \frac{k+2m+3}{k^2+k(2m+3)+2} \cdot \frac{[(k-1)t^*+1]}{[(k-1)t^*+1]}
\]

and

\[
v^* = t^* \cdot u^*
\]

(5.5)

where

\[
t^* = \frac{(2k+2m+1)+\sqrt{4(k+m)^2+12(k+m)+17}}{4(k+m+2)}
\]
For $i=1,2,...,k$, let $E_i$ be the subset of $E$ consisting of those $\binom{k}{i} \cdot 2^i$ elements with the $(k-i)$ components of $x$ being equal to zero. Then a symmetric design $\mu^*$ on $E_{r_1} \cup E_{r_2} \cup E_{r_3}$ is D-optimal iff

$$0 \leq r_1 \leq (k-1)u^* \frac{1-t^*}{1-u^*} \leq r_2 \leq k-1, \ r_3 = k,$$

$$\mu^*(E_{r_1}) = \frac{k}{(k-r_1)(r_2-r_1)} \left[ r_2-(k+r_2-1)u^*+(k-1)v^* \right],$$

$$\mu^*(E_{r_2}) = \frac{k}{(k-r_2)(r_2-r_1)} \left[ -r_1+(k+r_1-1)u^*-(k-1)v^* \right] \tag{5.8}$$

and $\mu^*(E_k) = 1-\mu^*(E_{r_1})-\mu^*(E_{r_2})$.

A proof of this fact is given in an appendix.

The weights for a symmetric D-optimal design with $r_1 = 0$ are listed in Table 1 for some $k$ and $m$. That would be benificial if fewer points in design are desired.

We remark that a symmetric D-optimal design for quadratic regression on the $k$-cube corresponds to $m = 0$, and Farrel, Kiefer and Walbran ([2]) have shown that $E_{r_1} \cup E_{r_2} \cup E_k$ supports a symmetric D-optimal design iff

$$r_2 = k-1, \ r_1 < k-1 \quad \text{for } k \leq 5$$

$$r_2 = k-1 \text{ or } k-2, \ r_1 < k-2 \quad \text{for } k \geq 6.$$ 

But we have to choose $r_2$ to be $k-1$ for any $k$ and sufficiently large $m$ since $(k-1) \cdot u^* \frac{1-t^*}{1-u^*}$ converges to $k-1$ as $m$ goes to infinity.
Table 1. Weights for a symmetric D-optimal design

<table>
<thead>
<tr>
<th>m</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>m</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>k=2</td>
<td>$\mu^*(E_0)$</td>
<td>.583</td>
<td>.555</td>
<td>.706</td>
<td>.743</td>
<td>.772</td>
<td>$\mu^*(E_0)$</td>
<td>.510</td>
<td>.578</td>
<td>.629</td>
<td>.670</td>
</tr>
<tr>
<td></td>
<td>$\mu^*(E_1)$</td>
<td>.321</td>
<td>.284</td>
<td>.252</td>
<td>.226</td>
<td>.204</td>
<td>$\mu^*(E_0)$</td>
<td>.424</td>
<td>.378</td>
<td>.339</td>
<td>.306</td>
</tr>
<tr>
<td></td>
<td>$\mu^*(E_2)$</td>
<td>.096</td>
<td>.061</td>
<td>.042</td>
<td>.031</td>
<td>.024</td>
<td>$\mu^*(E_0)$</td>
<td>.088</td>
<td>.044</td>
<td>.032</td>
<td>.024</td>
</tr>
</tbody>
</table>

| k=4 | $\mu^*(E_4)$ | .451 | .516 | .567 | .609 | .643 | $\mu^*(E_4)$ | .402 | .465 | .516 | .558 | .593 |
|    | $\mu^*(E_0)$ | .502 | .451 | .408 | .371 | .341 | $\mu^*(E_4)$ | .552 | .509 | .464 | .426 | .393 |
|    | $\mu^*(E_2)$ | .047 | .034 | .025 | .020 | .016 | $\mu^*(E_0)$ | .036 | .026 | .020 | .016 | .013 |

| k=5 | $\mu^*(E_5)$ | .616 | .423 | .472 | .515 | .551 | $\mu^*(E_7)$ | .607 | .642 | .436 | .477 | .513 |
|    | $\mu^*(E_2)$ | .381 | .556 | .511 | .471 | .438 | $\mu^*(E_5)$ | .389 | .357 | .550 | .511 | .477 |
|    | $\mu^*(E_0)$ | .003 | .021 | .017 | .014 | .011 | $\mu^*(E_0)$ | .004 | .001 | .014 | .012 | .010 |
|    | $r_2$ | 4 | 5 | 5 | 5 | 5 | $r_2$ | 5 | 5 | 6 | 6 | 6 |

| k=8 | $\mu^*(E_8)$ | .599 | .632 | .404 | .445 | .481 | $\mu^*(E_9)$ | .592 | .623 | .549 | .417 | .452 |
|    | $\mu^*(E_{2g})$ | .397 | .366 | .584 | .545 | .510 | $\mu^*(E_2)$ | .404 | .375 | .350 | .574 | .540 |
|    | $\mu^*(E_0)$ | .004 | .002 | .012 | .010 | .000 | $\mu^*(E_0)$ | .004 | .002 | .001 | .000 | .008 |
|    | $r_2$ | 6 | 6 | 7 | 7 | 7 | $r_2$ | 7 | 7 | 7 | 8 | 8 |
Acknowledgment

The idea of this paper arose from conversations at the Workshop on Efficient Data Collection and Related Inference, held at Univ. of California at Berkely in summer 1984.

Appendix : Derivation of $\mu^*$ in Example 2.

Farrel, Kiefer and Walbran ([2]) have shown algebraically that there exists a symmetric D-optimal design for quadratic regression on the k-cube. Kono ([7]) has used geometric arguments for the moment space to get a symmetric D-optimal design. In the following we generalize the geometric ideas contained in [7].

The space of $(u,v)$ is the convex hull of $\{((i/k, i(i-1)(i-2)/k(k-1)), i=0,...,k\}$ since

$$u = \sum_{i=0}^{k} \frac{i}{k} \cdot \mu(E_i)$$

and

$$v = \sum_{i=2}^{k} \frac{\binom{i}{2}}{\binom{k}{2}} \mu(E_i) = \sum_{i=2}^{k} \frac{i(i-1)}{k(k-1)} \cdot \mu(E_i).$$

So we need to show that $(u^*, v^*)$ is in that convex hull for the existence of a symmetric D-optimal design. Now we try to find $r_1, r_2, r_3, \mu^*(E_{r_1})$ and $\mu^*(E_{r_2})$ such that

$$u^* = \sum_{i=1}^{3} \frac{r_i}{k} \cdot \mu^*(E_{r_i})$$

and

$$v^* = \sum_{i=1}^{3} \frac{r_i(r_i-1)}{k(k-1)} \cdot \mu^*(E_{r_i})$$

(A.1)

where $0 \leq r_1 < r_2 < r_3 \leq k$.

Since $t^* > \frac{k+m+1}{k+m+2}$ and $u^*$ is a linear increasing function of $t^*$. 
\[ u^* > \frac{k+2m+3}{k^2+k(2m+3)+2} \left[ (k-1) \cdot \frac{k+m+1}{k+m+2} + 1 \right] \]
\[ > \frac{k-1}{k} \cdot \]

So we have to choose \( r_3 \) to be \( k \). Let \( L_1 \) be the line which passes through \((1,1)\) and \( \left( \frac{r_1}{k}, \frac{r_1(r_1-1)}{k(k-1)} \right) \) and \( L_2 \) be the line which passes through \( \left( \frac{r_2}{k}, \frac{r_2(r_2-1)}{k(k-1)} \right) \) and \((u^*,v^*)\). Suppose \( u_1 \) is the abscissa of the intersection point of \( L_1 \) and \( L_2 \). Then \( u^* \leq u_1 \leq 1 \) if and only if there exists a symmetric D-optimal design \( \mu^* \) whose support is \( E_{r_1} \cup E_{r_2} \cup E_k \). It can be easily checked that

\[ u_1 = \frac{r_2(k \cdot u^* - r_1) + r_1[(r_1-1)u^*-(k-1)v^*]}{r_2(k \cdot u^* - r_1) - r_1(k-r_1)+k(k-1)(u^*-v^*)} \]

and

\[ u^* \leq u_1 \leq 1 \quad \text{iff} \quad r_1 \leq (k-1)u^* \frac{(1-t^*)}{(1-u^*)} \leq r_2 \quad \text{(A.2)} \]

By the direct substitution of (5.5) into \( u^* \frac{(1-t^*)}{(1-u^*)} \) we get

\[ u^* \frac{(1-t^*)}{(1-u^*)} < 1 \]

which assures the existence of a symmetric D-optimal design on \( E_{r_1} \cup E_{k-1} \cup E_k \). We get (5.6) from (A.1) and (A.2).
References


