\[
u - \left[ \Sigma c_{i_1, \ldots, i_{n-1}, q} a_{\tau(i_1) \cdots a_{\tau(i_{n-1})}} u \right.
\quad + \left. \Sigma c_{i_1, \ldots, i_{n-1}, q+1} a_{\tau(i_1) \cdots a_{\tau(i_{n-1})}} v \right] = K(\tau).
\]

Summing over all \( \tau \) in the set \( C \) (say) which leave \( q \) and \( q+1 \) fixed, we get

\[
(q-1)!u - \left[ \Sigma c_{i_1, \ldots, i_{n-1}, q} \Sigma_{\tau \in C} a_{\tau(i_1) \cdots a_{\tau(i_{n-1})}} u \right.
\quad + \left. \Sigma c_{i_1, \ldots, i_{n-1}, q+1} \Sigma_{\tau \in C} a_{\tau(i_1) \cdots a_{\tau(i_{n-1})}} v \right] = \Sigma_{\tau \in C} K(\tau).
\]

Now in the same way as we went from (3.15) to (3.16) we see that for any \( i_1, \ldots, i_{n-1} \) each product \( a_{j_1} \cdots a_{j_{n-1}} \) with \( j_1 < \ldots < j_{n-1} \), occurs in the sum over \( C \) exactly \((q-1)!/(q-1)!) = (q-n)!(n-1)!\) times. Thus we get

\[
(q-1)!u - \left[ \Sigma c_{i_1, \ldots, i_{n-1}, q} \Sigma_{\tau \in C} (q-n)!(n-1)! a_{j_1} \cdots a_{j_{n-1}} u \right.
\quad + \left. \Sigma c_{i_1, \ldots, i_{n-1}, q+1} \Sigma_{\tau \in C} (q-n)!(n-1)! a_{j_1} \cdots a_{j_{n-1}} v \right] = \Sigma_{\tau \in C} K(\tau).
\]

With the obvious definition of \( A \) and \( K \), this is

\[
(3.17) \quad u - A[ \Sigma c_{i_1, \ldots, i_{n-1}, q} u + \Sigma c_{i_1, \ldots, i_{n-1}, q+1} v ] = K.
\]

Now

\[
\Sigma c_{i_1, \ldots, i_{n-1}, q+1} = \Sigma c_{i_1, \ldots, i_{n-1}, q+1} - \Sigma c_{i_1, \ldots, i_{n-2}, q, q+1}.
\]
There are \((q-1)!\) permutations \(\tau\) in \(B_{h,t}\). Any product \(a_{\tau(i_1)} \cdots a_{\tau(i_{n-2})}\) can be rewritten \(a_{j_1} \cdots a_{j_{n-2}}\) with \(j_1 < \cdots < j_{n-2}\). Since the \(a\)'s must be in the set \(\{a_1, \ldots, a_{q-1}\}\), there are \(\binom{q-1}{n-2}\) possible such products \(a_{j_1} \cdots a_{j_{n-2}}\) with \(j_1 < \cdots < j_{n-2}\). By symmetry each product occurs the same number of times in the sum over \(B_{h,t}\). Thus each occurs \((q-1)!/(q-2)! = (q+1-n)!/(n-2)!\) times.

Therefore (3.15) becomes

\[
(3.16) \quad \sum_{i_1 < \cdots < i_{n-2}, h,t} c_{i_1, \ldots, i_{n-2}, h,t} = 0,
\]

for any \(h \neq t, 1 \leq h < q+1, 1 \leq t \leq q+1\). We would like to show that the sum of the \(a\)'s is nonzero. Since each \(a_{j_i} \geq 0\) (the \(a_i\) are coordinates in a simplex), we need only show that some \(n-2\) of the \(a_{j_i}\) are positive; then the sum will be positive. If no \(n-2\) of the \(a\)'s were positive, (3.13) would reduce to \(x_q\). Choosing \(\tau\) with \(\tau(q) = q\) would make \(x_q = u\), a variable. But (3.13) is constant for any \(\tau\). Therefore some \(n-2\) of the \(a\)'s are positive, and (3.16) reduces to

\[
\sum_{i_1 < \cdots < i_{n-2}, h,t} c_{i_1, \ldots, i_{n-2}, h,t} = 0, \quad \text{all } h \neq t.
\]

We fix \(t\) and sum over \(h \neq t\) to get

\[
(n-1) \sum_{i_1 < \cdots < i_{n-1}, \ell} c_{i_1, \ldots, i_{n-1}, \ell} = 0,
\]

which reduces to (3.14), the desired equation. We will use (3.14) shortly.

Now let us return to (3.13). We have examined the quadratic terms; we now examine the linear terms.

It turns out that we only need to consider those \(\tau\) with \(\tau(q) = q\), \(\tau(q+1) = q+1\). That is, \(h = q, t = q+1\). Since (3.13) is constant, the linear terms of (3.13) must total a constant \(K(\tau)\). Thus
\[
x_q - \left[ \sum_{i_1 < \ldots < i_n} c_{i_1, \ldots, i_n} a_{\tau(i_1)} \cdots a_{\tau(i_n)} \right.
\]
\[
+ \sum_{i_1 < \ldots < i_{n-1}} c_{i_1, \ldots, i_{n-1}, h} a_{\tau(i_1)} \cdots a_{\tau(i_{n-1})} u
\]
\[
+ \sum_{i_1 < \ldots < i_{n-2}} c_{i_1, \ldots, i_{n-2}, h, t} a_{\tau(i_1)} \cdots a_{\tau(i_{n-2})} v
\]
(3.13)

where \( x_q = a_{\tau(q)}, u, \) or \( v, \) depending on whether \( \tau(q) < q, = q, \) or \( = q+1. \)

Suppose for the moment that \( n > 2. \) If \( n = 2 \) much of what follows simplifies, as discussed at the end of the proof.

Our first goal is to prove

\[
\sum_{i_1 < \ldots < i_{n-1}} c_{i_1, \ldots, i_{n-1}, t} = 0, \quad \text{all } t, 1 \leq t \leq q+1.
\]
(3.14)

Since (3.12) is constant on \( L_\tau, \) the expression (3.13) is a constant, (the value of which may depend on \( \tau \)). Therefore the coefficients of the quadratic term \( uv \) must total zero:

\[
\sum_{i_1 < \ldots < i_{n-2}} c_{i_1, \ldots, i_{n-2}, h, t} a_{\tau(i_1)} \cdots a_{\tau(i_{n-2})} = 0, \quad h = \tau^{-1}(q), \quad t = \tau^{-1}(q+1).
\]

Let \( B_{h, t} = \{ \tau | h = \tau^{-1}(q), \quad t = \tau^{-1}(q+1) \}. \) Summing over all \( \tau \) in \( B_{h, t}, \) we get

\[
\sum_{i_1 < \ldots < i_{n-2}} c_{i_1, \ldots, i_{n-2}, h, t} \sum_{\tau \in B_{h, t}} a_{\tau(i_1)} \cdots a_{\tau(i_{n-2})} = 0.
\]
(3.15)
PROOF. Suppose \( d_s(x, \xi^*, D) \) is bounded on such a line, say the line \( L \) determined by setting \( x_1 = a_1, i < q \). We will show that a contradiction follows.

In general if \( z \) is an \( s \)-vector and \( B \) is a positive definite symmetric \( s \times s \) matrix with smallest eigenvalue \( \lambda_0 \), then \( |z' B z| \geq \lambda_0 |z|^2 \), where \( | | \) denotes the absolute value of a number and the Euclidean norm of a vector. Thus if \( d_s(x, \xi^*, D) \) is bounded on a line, \( |f(1)(x) - D'f(2)(x)|^2 \) must be bounded on that line. Hence each of the \( s \) components of \( f(1)(x) - D'f(2)(x) \) is bounded on that line. Since each component of \( f(1)(x) - D'f(2)(x) \) is a polynomial, each component must be constant on that line. Let us write the \( q \)th component as

\[
(3.12) \quad x_q = \sum_{i_1 < \ldots < i_n} c_{i_1, \ldots, i_n} x_{i_1} \cdots x_{i_n}.
\]

Here we have used the assumption that \( n = m + 1 \), so \( f^{(2)} \) consists only of terms of degree \( n \). Observe that \( n = m + 1 \geq 2 \). To avoid notational problems later, we define \( c_{j_1, \ldots, j_n} = c_{i_1, \ldots, i_n} \) whenever \( (j_1, \ldots, j_n) \) is a rearrangement of \( (i_1, \ldots, i_n) \).

If \( \tau \) is any permutation of \( q + 1 \) elements, let \( L_\tau \) be the image of \( L \) under that permutation of the coordinates. \( L_\tau \) is defined by setting \( x_i = a_\tau(i) \) for those \( i \) with \( \tau(i) < q \). We will write \( h = \tau^{-1}(q) \) and \( \xi = \tau^{-1}(q + 1) \), and write \( x_n = u \) and \( x_\xi = v \). This is simply a notational convenience. Thus \( u \) and \( v \) are variables satisfying \( u + v + \sum_{i=1}^{q-1} a_i = 1 \).

We assumed \( d_s(x, \xi^*, D) \) to be bounded on \( L \). By the invariance of \( d_s(x, \xi^*, D) \) it is therefore bounded on \( L_\tau \) for every \( \tau \). Therefore on every \( L_\tau \), (3.12) is constant.

For any \( \tau \), on the line \( L_\tau \), (3.12) becomes
where in each matrix we have used the fact that \( f^{(2)} = 0 \) on the support of \( \xi_s \). \( O(\varepsilon) \) is in each case a matrix which approaches 0 as \( \varepsilon \to 0 \). Therefore

\[
d_s|k(x, \xi_s) = (1 - \varepsilon)^{-1} h^{(1)}(x) h^{(1)}(x) + h'(x) O(\varepsilon) h(x).
\]

Hence

\[
s \leq \max_x d_s|k(x, \xi_s) \leq (1 - \varepsilon)^{-1} \max_x h^{(1)}(x) h^{(1)}(x) + \max_x h'(x) O(\varepsilon) h(x) - \max_x h^{(1)}(x) h^{(1)}(x) \leq s,
\]

the last inequality following from (3.11). Therefore \( \xi_s \) is optimal for \( s \) out of \( k \) parameters, proving the theorem.

We now return to the example of special polynomial regression on the simplex. Let \( X \) be the \( q \)-simplex. Let \( f(x) \) be the \( k \)-vector whose components are the special monomials of degree \( \leq n \), where \( n \leq q + 1 \). Let \( f^{(1)}(x) \) be the \( s \)-vector consisting of those components of degree \( \leq m \), where \( m \leq n \). And let \( G \) be the group of permutations of the coordinates of the simplex. It is easily seen that there is a corresponding \( \overline{G} \) so that \( G \) and \( \overline{G} \) satisfy the conditions of Theorem 3.3. Let \( d_s(x, \xi, D) \) be as in Theorem 3.3.

**Lemma.** In the model described above, let \( n = m + 1 \), and let \( \xi^* \) be optimal for \( f^{(1)} \). Then \( d_s(x, \xi^*, D) \) is unbounded on any line determined by setting all but two of the \( x_i \) constant and holding \( \sum_{i=1}^{q+1} x_i = 1 \).

When \( s = k \) the analogous statement about \( d(x, \xi^*) \) is trivial, but when \( s < k \), \( f^{(1)}(x) - D' f^{(2)}(x) \) could conceivably be constant on such a line, making \( d_s(x, \xi^*, D) \) constant there.
To see that equality holds for some $x$, recall that (3.6) says that
$$\int d_s(x, \xi, X) \, d\xi(x) \geq s, \text{ for any } r \times s \text{ matrix } X.$$ Thus (i) implies (iii).

Moreover from (3.6) we see that $\max_x d_s(x, \xi, D) \geq s$ for any $\xi$, so (iii) implies (ii). But there is a $\xi$, namely an optimal $\xi$, for which $\max_x d_s(x, \xi, D) = s$, so (ii) implies (iii). Finally, (iii) implies (i) by the corollary to the theorem of Karlin and Studden given earlier. This completes the proof of the theorem.

The remaining results in this chapter concern special situations.

**THEOREM 3.4.** If $\xi_{s}$ is optimal for estimating the $s$ out of $s$ coefficients of $A\theta(1) + B\theta(2)$, where $A$ is $s \times s$ nonsingular and $B$ is $s \times r$, and if $\theta(2) = 0$ on the support of $\xi_{s}$, then $\xi_{s}$ is optimal for estimating $\theta(1)$ out of $\theta$.

**REMARK.** As a special case we may have $A = I$ and $B = 0$.

**PROOF.** Write $g(1) = A\theta(1) + B\theta(2)$. There is a nonsingular $s \times s$ matrix $L_{1}$ with $L_{1}[\int \, g(1) \, d\xi_{s}] \, L_{1}' = I$. Write $h(1) = L_{1}g(1) = L_{1}A\theta(1) + L_{1}B\theta(2)$. Since $\xi_{s}$ is optimal for the $s$ out of $s$ coefficients of $g(1)$,

$$s \geq g(1)'(x)[\int \, g(1) \, d\xi_{s}]^{-1} \, g(1)(x) = h(1)'(x) \, h(1)(x).$$

Now let

$$L = \begin{bmatrix} L_{1}A & L_{1}B \\ 0 & I \end{bmatrix} \text{ and } Lf = \begin{pmatrix} h(1) \\ f(2) \end{pmatrix} = h.$$ 

Let $\xi'$ be a design so that $M(\xi_{c})$ is nonsingular for $0 < c < 1$, where $\xi_{c} = (1 - c)\xi_{s} + c\xi'$. Then

$$d_s|_{k}(x, \xi_{c}) = f'(x)M_{-1}(\xi_{c})f(x) - f(2)'(x)M_{-1}(\xi_{c})f(2)(x)$$

$$= h'(x)[(1 - c)LM(\xi_{s})L' + c \, LM(\xi')L']^{-1}h(x)$$

$$- f(2)'(x) \, M_{-1}(\xi_{c})f(2)(x).$$
\begin{equation}
(3.10) \quad d_s(x, \xi^*, X) = f'(g^{-1}x)(I, -B'X'C^{-1}g)[M^*(\xi^*)]^{-1}(I, -B'X'C^{-1}g)f(g^{-1}x)
= d_s(g^{-1}x, \xi^*, C^{-1}XBg).
\end{equation}

In particular

\[ d_s(gx, \xi^*, D) = d_s(x, \xi^*, C^{-1}D_0g) = d_s(x, \xi^*, D). \]

We now must show that (i) implies (iii). Let \( \xi^* \) be optimal. Then by Theorem 3.2, we know that \( d_s(x, \xi^*, D_0) \leq s \) for all \( x \). Therefore, by (3.10),

\[ d_s(x, \xi^*, D_0) = d_s(gx, \xi^*, C^{-1}D_0g) \]
\[ = d_s(gx, \xi^*, D_0) \leq s \]

for all \( x \).

Observe that if \( P \) is any symmetric positive definite \( s \times s \) matrix and \( z_1 \) and \( z_2 \) are any two \( s \)-vectors, then

\[ ([1-\alpha]z_1 + \alpha z_2)'P([1-\alpha]z_1 + \alpha z_2) \leq (1-\alpha)z_1'Pz_1 + \alpha z_2'Pz_2 \]

for \( 0 \leq \alpha \leq 1 \). (For the expression on the left is convex in \( \alpha \).) For fixed \( x \) let \( z_g = (I, -D_g')f(x) \), and let \( h(z_g) = z_g'[M^*(\xi^*)]^{-1}z_g \). Then Jensen's inequality states that \( h(\int z_g\,d\mu(g)) \leq \int h(z_g)d\mu(g) \), so we have

\[ d_s(x, \xi^*, D) = f'(x)(I, -\int D'_g d\mu(g))'[M^*(\xi^*)]^{-1}(I, -\int D'_g d\mu(g))f(x) \]
\[ = \int [f'(x)(I, -D'_g)]d\mu(g)[M^*(\xi^*)]^{-1}f(x)]d\mu(g) \]
\[ \leq \int [f'(x)(I, -D'_g)[M^*(\xi^*)]^{-1}(I, -D'_g)f(x)]d\mu(g) \]
\[ = \int d_s(x, \xi^*, D_g)d\mu(g) \]
\[ \leq s. \]
because $A \cdot A_1 = A$, as mentioned in Chapter 1.

But this last expression equals

$$
\int C^{-1}_g D_0 B_g \, d\mu(g^{-1}_1) = \int C^{-1}_g D_0 B_g \, d\mu(g') = D
$$

by the invariance of Haar measure. Thus $D$ is invariant in the sense defined.

Now suppose $\xi^*$ is optimal, i.e., (i) holds. We must show that $d_g(x, \xi^*, D)$ is invariant.

For each fixed $g$, note first that $\theta f(x) = (A_g \theta)^t f(gx) = \theta A^t g f(gx)$ for all $\theta$ and $x$, and therefore $f(x) = A^t g f(gx)$.

For any $r \times s$ matrix $X$, write

$$
d_g(x, \xi^*, X) = f'(x)(I, -X)^t [M^*(\xi^*)]^{-1}(I, -X') f(x).
$$

Because $A_g$ is an element of a group, $A_g$ has an inverse. Therefore for any $g$ we have

$$
d_g(x, \xi^*, X) = f'(x) A_g^{-1}(I, -X')^t B_g^{-1} [M^*(\xi^*)]^{-1} B_g^{-t} [M^*(\xi^*)]^{-1} (I, -X') A_g^{-1} f(x)
$$

$$
= f'(g^{-1} x)(I, -B_g X C_g^{-1})^t B_g^{-1} [M^*(\xi^*)]^{-1} B_g^{-t} (I, -B_g X C_g^{-1}) f(g^{-1} x).
$$

Now let $\xi_0$ be any invariant optimal design. Then for each $g$,

$$
A_g^t M(\xi_0) A_g = \int A_g^t f(x) f'(x) A_g \, d\xi_0(x)
$$

$$
= \int f(g^{-1} x) f'(g^{-1} x) \, d\xi_0(x)
$$

$$
= \int f(g^{-1} x) f'(g^{-1} x) \, d\xi_0(g^{-1} x)
$$

$$
= M(\xi_0).
$$

Since $M^*(\xi)$ is the same for all optimal $\xi$, we conclude

$$
M^*(\xi^*) = M^*(\xi_0) = B_g^t M^*(\xi_0) B_g = B_g^t M^*(\xi^*) B_g.
$$

The middle equality follows from the remarks in connection with (3.9). Therefore
with \( L_1 \) \( s \times s \), and if \( LML' = N \), then \( N^* = L_1M^*L_1' \). If \( Lf = g \) and \( M(\xi) \) is nonsingular, then \( d_{s}(x, \xi) \) is the same whether written in terms of \( f \) and \( M \) or \( g \) and \( N \).

Let us now consider the question of invariance in connection with Theorem 3.2. When estimating all \( k \) parameters we sometimes had the case that for \( \xi^* \) optimal, \( d(x, \xi^*) \) was invariant under some group of transformations. The next theorem gives an analogous result.

**Theorem 3.3.** Let \( G, \overline{G} \) and the matrices

\[
A_g = \begin{bmatrix} B_g & 0 \\ 0 & Q \end{bmatrix}
\]

be as described in the section on invariance in Chapter 1. There exists an \( r \times s \) matrix \( D \) such that for any \( A_g \) in \( G \) we have \( C^{-1}DB_g = D \), and

if \( \xi^* \) is optimal then \( d_s(x, \xi^*, D) = d_s(gx, \xi^*, D) \) for any \( g \) in \( G \) and \( x \) in \( X \). Moreover the sets of designs \( \xi^* \) satisfying (i), (ii) or (iii) coincide.

1. \( \xi^* \) maximizes \( \det M^*(\xi) \). (D-optimality)
2. \( \xi^* \) minimizes \( \max_x d_s(x, \xi^*, D) \).
3. \( \max_x d_s(x, \xi^*, D) = s \).

**Proof.** The order of proof is as follows. We define \( D \), show that for any \( A_g \) in \( G \) we have \( C^{-1}DB_g = D \), and then show that \( d_s(x, \xi^*, D) \) is invariant under \( G \), for \( \xi^* \) optimal. We then show that (i), (ii) and (iii) are equivalent.

First we must define \( D \). Let \( D_0 \) be as in Theorem 3.2. For \( g \) in \( G \) and corresponding matrix \( A_g \), write \( C^{-1}D_0 B_g = D_g \). Let \( \mu \) denote Haar measure on \( G \) with \( \mu(G) = 1 \). We define \( D = \int D_g d\mu(g) \).

For any \( g_1 \) in \( G \) we have

\[
C^{-1}D_0 B_g = \int C^{-1}C^{-1}D_0 B_g B_g d\mu(g) = \int C^{-1}D_0 B_g d\mu(g)
\]
- 23 -

- $\infty < c < \infty$, then $d_s(x, \xi^*, X) = (x - c(1 - x))^2$. This has maximum value max $1, c^2$, which equals 1 $s$ if $|c| \leq 1$, and equals $c^2 > s$ if $|c| > 1$. Thus in the theorem as originally stated (i) may hold while (iii) does not.

In this example $D_0$ can be computed. To do this we consider the game described earlier, which in this case simplifies to

$$ (3.8) \quad (q, e)M(\xi)(q, e)' $$

where $e \leq N$ and $q \geq \sup_\xi \det M^*(\xi) = M^*(\xi^*) = 1$. Now (3.8) equals

$$ \int (q, e)f(x)f'(x)(q, e)' \, d\xi(x) = \int (q x + e(1 - x))^2 \, d\xi(x). $$

The expression $(q x + e(1 - x))^2$ is convex in $x$. Therefore (3.8) is maximized if $\xi$ is concentrated at 0 or 1, according as $e$ or $q$ is larger. The value attained is $q^2$ or $e^2$, and thus the minimax strategy $(q_N, e_N)$ is $(1, 0)$. Thus $D_0 = - \lim e_N q_N^{-1} = 0$. (By the way, the maximin strategy $\xi_N$ concentrates at 1, and therefore $\xi_N \to \xi^*$.)

Thus in this example, for any design $\xi$,

$$ \max_x d_s(x, \xi, D_0) = \max_x x [M^*(\xi)]^{-1} x = [M^*(\xi)]^{-1}, $$

and a design minimizes $\max_x d_s(x, \xi, D_0)$ if and only if it maximizes $\det M^*(\xi)$.

Finally we remark that if $X = -1$, then $d_s(x, \xi^*, X) = 1$, identically in $x$. The fact that $d_s(x, \xi^*, X)$ can conceivably be constant on certain intervals will cause considerable difficulty in a later example. (See the lemma preceding Theorem 3.5;)

Before going further we point out a fact which will be used from time to time. If $L$ is a nonsingular $k \times k$ matrix of the form

$$ (3.9) \quad L = \begin{bmatrix} L_1 & L_2 \\ 0 & L_3 \end{bmatrix} $$
Both the Karlin-Studden Corollary and Theorem 3.1 give sufficient conditions for optimality. It seems as easy in principle to guess an $X$ which works in the corollary as to guess a $\xi'$ (hence $\xi_{\varepsilon}$) which works in Theorem 3.1.

However after proving Theorem 3.7 by using a $\xi_{\varepsilon}$ we will point out that it appears hard to find an $X$ with $\max_x d_s(x, \xi^o, X) = s$. In particular $X(\xi_{\varepsilon})$ is uniquely determined and equal to $M^{-1}_3(\xi_{\varepsilon})M'_2(\xi_{\varepsilon}) = 0$, but we cannot take $X = \lim_{\varepsilon} X(\xi_{\varepsilon})$, since for that $X$, $\max_x d_s(x, \xi^o, X) > s$. On the other hand there are problems in which the $X$ method is very quick. Also we know that a suitable $X$ always exists, $D_0$. We have not proved that a $\xi_{\varepsilon}$ satisfying (3.1) and (3.2) always exists.

Let us conclude this discussion with the promised example.

EXAMPLE 3.1. Let $X$ be the unit interval, $0 \leq x \leq 1$, let $s = 1$ and $k = 2$, let $f_1(x) = x$, $f_2(x) = 1 - x$, and let $\xi^*$ be concentrated at 1.

If $\xi'$ is concentrated at $c/(c+1)$, for $c > 0$, we have

$$M(\xi_{\varepsilon}) = \frac{\varepsilon}{(c+1)^2} \begin{vmatrix} (c+1)^2(1-\varepsilon) + c^2 & c \\ c & 1 \end{vmatrix}.$$ 

Then $d_s|k(x, \xi_{\varepsilon}) = (1 - \varepsilon)^{-1}(x - c(1 - x))^2$. This is maximized at an end point of the interval, and $\max_x d_s|k(x, \xi_{\varepsilon}) = (1 - \varepsilon)^{-1}\max (1, c^2)$. If $c \leq 1$,

$$\lim_{\varepsilon} \max_x d_s|k(x, \xi_{\varepsilon}) = 1 = s,$$

proving that $\xi^*$ is optimal, by Theorem 3.1. If $c > 1$

$$\lim_{\varepsilon} \max_x d_s|k(x, \xi_{\varepsilon}) = c^2 > s,$$

demonstrating that (3.3) does not imply (3.2).

Let us now use the Karlin-Studden method. We have

$$M(\xi^*) = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix},$$

so any $1 \times 1$ matrix $X$ satisfies $M_3(\xi^*)X = M'_2(\xi^*)$. If we let $X = c$, 

\[
\min_\xi \max_\eta \int d_s(x, \xi, X(\xi)) \, d\eta(x) = \min_\xi \max_\eta \phi([M^*(\xi)]^{-1}, X(\xi), \eta) \\
\leq \max_\eta \phi([M^*(\xi^*)]^{-1}, D_0, \eta).
\]

It is not clear without additional argument that the inequality is true unless we choose \(X(\xi^*) = D_0\).

We now see how to define \(X(\xi)\) to make the theorem true. Find \(D_0\) and choose \(X(\xi) = D_0\) whenever \(M_3(\xi)D_0 = M_2^*(\xi)\). Otherwise choose any \(X(\xi)\) satisfying (3.5). Then the proof holds as it is, and the modified theorem is correct.

Alternatively we could replace \(X(\xi)\) by \(D_0\) for all \(\xi\). We obtain this modification of the theorem if we show that \(\max_x d_s(x, \xi, D_0) \geq s\) for all \(\xi\), with equality if and only if \(\xi\) is optimal. But this is quick to show. By (3.6) we have \(\max_x d_s(x, \xi, D_0) \geq s\), for all \(\xi\). If equality holds then by the corollary given above, \(\xi\) is optimal. On the other hand if \(\xi^*\) is optimal then (3.7) holds, so choosing \(X(\xi)\) as in the last paragraph we have \(X(\xi^*) = D_0\) and

\[
\max_x d_s(x, \xi^*, D_0) = \max_x d_s(x, \xi^*, X(\xi^*)) = s.
\]

We write out the theorem just obtained.

**THEOREM 3.2.** (Karlin and Studden) *There is an \(r \times s\) matrix \(D_0\) so that the sets of designs \(\xi^*\) satisfying (i), (ii), or (iii) coincide.*

(i) \(\xi^*\) maximizes \(\det M^*(\xi)\). (D-optimality)

(ii) \(\xi^*\) minimizes \(\max_x d_s(x, \xi, D_0)\).

(iii) \(\max_x d_s(x, \xi^*, D_0) = s\).

We will use this theorem later in this chapter. However since \(D_0\) is in general so hard to determine, it seems that the corollary given earlier will usually be more applicable in examples than Theorem 3.2.
equality holds in (3.6) and \( M_3(\xi)X = M_2'(\xi) \). The correct portion of the theorem then gives the result.

We now consider the proof of the theorem in detail.

The authors consider the sequence of games

\[ \Psi_N(Q, E, \xi) = \text{tr} \| q', E' \| M(\xi) \| q', E' \| ' \]

where \( E \) is an \( r \times s \) real matrix whose entries are bounded by \( N \), and \( Q \) is a positive definite \( s \times s \) matrix, normalized so that \( \det^{\frac{2}{2}} Q \geq \sup_{\xi} \det M^*(\xi) > 0 \), and the smallest eigenvalue of \( Q \) is \( \geq N^{-1} \). The minimizing player has as strategy space the product of the space of such \( Q \) and the space of such \( E \), while the maximizing player has as strategy space the space of all \( \xi \). Denote the minimax strategy by \( (Q_N, E_N) \). Letting \( N \to \infty \) (perhaps via a subsequence) we obtain limit matrices \( Q_0 \) and \( E_0 \), where \( Q_0 \) is nonsingular. Define \( D_0 = -E_0Q_0^{-1} \). Finally define

\[ \phi(P, D, \xi) = \int \text{tr} \ P(f^{(1)}(x) - D'f^{(2)}(x))(f^{(1)}(x) - D'f^{(2)}(x))' \, d\xi(x), \]

where \( P \) is \( s \times s \) and \( D \) is \( r \times s \).

In proving that (i) implies (iii) it is shown that if \( \xi^* \) satisfies (i) then

\[ (3.7) \quad M_3(\xi^*)D_0 = M_2'(\xi^*). \]

Hence it is claimed that

\[ \phi([M^*(\xi^*)]^{-1}, D_0, \xi) = \int d_s(x, \xi^*, X(\xi^*)) \, d\xi(x). \]

This is true if we choose \( X(\xi^*) = D_0 \), but it is not true in general, as Example 3.1 will show.

There is certainly some \( \xi^* \) satisfying (i). For this \( \xi^* \), (3.7) holds, so \( D_0 \) is an allowed choice for \( X(\xi^*) \). In proving that (ii) implies (iii) it is claimed that
Such an $X(\xi)$ is shown always to exist. If $M(\xi)$ is nonsingular,
$X(\xi) = M^{-1}_3(\xi)M'_2(\xi)$, and $d_s(x, \xi; X(\xi))$ is equal to $d_s|_k(x, \xi)$ as defined by (1.2). If $M_3(\xi)$ is singular $X(\xi)$ is not unique, but in any case the authors write $d_s(x, \xi, X(\xi)) = d_s(x, \xi)$, which is supposed to be well defined.
Example 3.1 will show that neither $d_s(x, \xi)$ nor $\max_x d_s(x, \xi)$ is independent of the particular $X(\xi)$ chosen. The following theorem is then asserted to be true.

THEOREM. (Karlin and Studden) If for any $\xi$, $X(\xi)$ satisfies (3.5), then the sets of designs $\xi^*$ satisfying (i), (ii) or (iii) coincide.

(i) $\xi^*$ maximizes $M(\xi)$. (D-optimality)

(ii) $\xi^*$ minimizes $\max_x d_s(x, \xi, X(\xi))$.

(iii) $\max_x d_s(x, \xi, X(\xi)) = s$.

The proofs that (iii) implies (i) and (iii) implies (ii) are correct no matter what $X(\xi)$ is chosen. It is not clear that (ii) implies (iii) unless $X(\xi)$ is chosen in a certain way. Finally (i) implies (iii) only if $X(\xi)$ is chosen in a special way. Example 3.1 will show that (i) may hold while (iii) does not.

Before discussing the proof in detail we point out that the theorem as stated does give a correct sufficient condition for optimality, which we state as a corollary.

COROLLARY. (Karlin and Studden) If for some $r \times s$ matrix $X$

$$\max_x d_s(x, \xi, X) = s$$

then $\xi$ is D-optimal.

PROOF. It is not hard to show (see equation (6.9) of Karlin and Studden [1966a]) that $d_s(x, \xi, X)$ satisfies

$$\int d_s(x, \xi, X) d\xi(x) \geq s,$$

with equality holding if and only if $M_3(\xi)X = M'_2(\xi)$. Thus if $\max_x d_s(x, \xi, X) = s$,
Taking the limit as $\varepsilon \to 0$,

$$\min_\xi \det M^{(1)}(\xi) \geq \det M^{(1)}(\xi^*),$$

that is, $\xi^*$ is optimal. This completes the proof of the theorem.

We will show in Example 3.1 that (3.3) does not imply (3.2). However since that example illustrates several other things as well, we postpone it until we can give a full discussion.

Let us compare Theorem 3.1 with previously known results. Kiefer [1961] defines functions $\overline{D}(x, \xi^*), D(\xi, \xi^*)$ and $D(x, \xi^*)$, and shows that for estimating $\theta^{(1)}$

$$\max_x \overline{D}(x, \xi^*) = s \implies \max_\xi D(\xi, \xi^*) = s$$

$$\iff \xi^* \text{ optimal}$$

$$\implies \max_x D(x, \xi^*) = s.$$

Later in this chapter we will consider the example of special $(q+1)$-tic regression on the $q$-simplex, where $\theta^{(1)}$ consists of the multilinear monomials of degree $\leq m$. In that example Kiefer's results are inconclusive, because for the design $\xi^0$ under consideration, $\max_x \overline{D}(x, \xi^0) > s$, $\max_x D(x, \xi^0) = s$, and $D(\xi, \xi^0)$ is extremely hard to compute. However Theorem 3.1 can be applied to show that $\xi^0$ is optimal.

Karlin and Studden, [1966a, Sec. 6] and [1966b, Ch. 10], take the following approach. For any $r \times s$ matrix $X$ define

$$d_s(x, \xi, X) = (f^{(1)}(x) - X'f^{(2)}(x))' [M^*(\xi)]^{-1} (f^{(1)}(x) - X'f^{(2)}(x)).$$

(Their definition of $M^*(\xi)$ is of a different form from ours, but it gives the same matrix.) For any $\xi$ let $X(\xi)$ be an $r \times s$ matrix satisfying

$$M^*_2(\xi)X(\xi) = M^*_2(\xi).$$
the pseudo-inverse \( X^{-1} = \lim_{\lambda \to 0^+} (X + \lambda Y)^{-1} \). Chernoff proves that \( X^{-1} \) is independent of the particular \( Y \) used. \( X^{-1} \) may have infinite entries, but we will only be interested in \( M^{(1)}(\xi^*) \), which will always be finite if \( \theta^{(1)} \) is estimable under \( \xi^* \).

In our case this pseudo-inverse \( M^{-1}(\xi^*) = \lim_{\xi \to 0^+} M^{-1}(\xi \xi^*) \), and \( M^{(1)}(\xi^*) = \lim_{\xi \to 0^+} M^{(1)}(\xi \xi^*) \) (As so defined \( \sigma^2 N^{-1} M^{(1)}(\xi^*) \) really is the covariance matrix of the best linear estimator of \( \theta^{(1)} \) under \( \xi^* \)).

Kiefer [1961, p. 305 ff.] shows that for all measures \( \xi \) supported on one point \( x \), and \( \xi_1 \) with \( M(\xi_1) \) nonsingular, the following derivative exists and satisfies the equation:

\[
- \frac{\partial}{\partial \alpha} \log \det M^{(1)}([1 - \alpha] \xi_1 + \alpha \xi) \bigg|_{\alpha = 0^+} = d_s |k(x, \xi_1)| - s.
\]

He also shows there that the left side is linear in \( \xi \) for all designs \( \xi \). Thus for any number \( a \)

\[
- \frac{\partial}{\partial \alpha} \log \det M^{(1)}([1 - \alpha] \xi_1 + \alpha \xi) \bigg|_{\alpha = 0^+} \leq a, \text{ all } \xi
\]

\[
\iff d_s |k(x, \xi_1)| \leq s + a, \text{ all } x.
\]

By assumption, \( d_s |k(x, \xi \xi^*)| \leq s + c(\xi) \) for all \( x \), where \( c(\xi) \to 0 \) as \( \xi \to 0 \). So

\[
c(\xi) \geq - \frac{\partial}{\partial \alpha} \log \det M^{(1)}([1 - \alpha] \xi \xi^* + \alpha \xi) \bigg|_{\alpha = 0^+}, \text{ all } \xi.
\]

Since (see Kiefer [1961, p. 306]) for \( 0 \leq \alpha \leq 1 \) and any \( \xi \),

\[- \log \det M^{(1)}([1 - \alpha] \xi \xi^* + \alpha \xi) \text{ is concave in } \alpha, \text{ we have}
\]

\[
c(\xi) \geq - \log \det M^{(1)}(\xi) + \log \det M^{(1)}(\xi \xi^*), \text{ all } \xi.
\]

Hence

\[
\min_{\xi} \det M^{(1)}(\xi) \geq e^{-c(\xi)} \det M^{(1)}(\xi \xi^*). \text{ ...}
\]
We assume that \( \lim_{\varepsilon} d_s k(x, \xi) \) exists for all \( x \) and is \( \leq \varepsilon \). Just as before we obtain that for \( \varepsilon < \) some \( \varepsilon_0 \) and for \( |x - x'| \) sufficiently small, (3.4) holds. Now choose a sequence \( \varepsilon_n \) so that

\[
\max_x d_s k(x, \xi) \to \limsup_{\varepsilon} \max_x d_s k(x, \xi).
\]

The value \( \max_x d_s k(x, \xi) \) is attained at \( x \). A subsequence (again written \( x_{\varepsilon_n} \)) approaches some \( x_0 \). So choose \( \varepsilon_1 < \varepsilon_0 \) so that if \( 0 < \varepsilon_n < \varepsilon_1 \) then

\[
|\limsup_{\varepsilon} \max_x d_s k(x, \xi) - d_s k(x_{\varepsilon_n}, \xi_{\varepsilon_n})| < \delta
\]

and

\[
|d_s k(x_{\varepsilon_n}, \xi_{\varepsilon_n}) - d_s k(x_0, \xi_{\varepsilon_n})| < \delta
\]

and

\[
|d_s k(x_0, \xi_{\varepsilon_n}) - \lim_{\varepsilon} d_s k(x_0, \xi)| < \delta.
\]

Then, since \( \delta \) is arbitrary,

\[
\limsup_{\varepsilon} \max_x d_s k(x, \xi) = \lim_{\varepsilon} d_s k(x_0, \xi) \leq s.
\]

But for any \( \xi \) with \( M(\xi) \) nonsingular, \( \max_x d_s k(x, \xi) \geq s \), so

\[
\liminf_{\varepsilon} \max_x d_s k(x, \xi) \geq s,
\]

proving that \( \lim_{\varepsilon} \max_x d(x, \xi) \) exists and equals \( s \).

Finally we prove that either (3.1) or (3.2) implies (3.3). Since (3.1) and (3.2) are equivalent we will assume both and obtain (3.3).

If \( M(\xi^*) \) is nonsingular, we assume (3.1). Then

\[
s = \max_x \lim_{\varepsilon} d_s k(x, \xi) = \max_x d_s k(x, \xi^*),
\]

so \( \xi^* \) is optimal by Theorem 1.2.

In general we assume (3.2) and proceed as follows.

Chernoff [1953] gives the following definition of a pseudo-inverse. For \( X \) a symmetric nonnegative definite matrix, let \( Y \) be any symmetric matrix such that \( X + \lambda Y \) is positive definite for \( \lambda > 0 \) sufficiently small. Define
with entries $F_i(x_j)$, is nonsingular. Then (letting $| |$ denote the Euclidean norm of a vector and the corresponding operator norm of a matrix)

$$|b(e)| = |b(e)'F^{-1}| \leq |b(e)'F||F^{-1}|$$

which is bounded as $e \to 0$ because $b(e)'F(x) = d_\|k(x, \xi_e)\|$ is bounded.

Therefore for $0 < e < e_0$ and $|x - x'|$ sufficiently small

$$|d_\|k(x, \xi_e) - d_\|k(x', \xi_e)| \leq N \max_{0 < e \leq e_0} \sup_i |b_i(e)| \max_i |F_i(x) - F_i(x')| < \delta$$

proving (3.4). The last inequality uses the continuity of the $F_i$'s and the compactness of $x$.

Let $x_e$ be a point at which $\max_x d(x, \xi_e)$ is attained. Since $X$ is compact there is a point $x_0$ and a sequence $x_n$ approaching $x_0$ as $e_n \to 0$.

Pick $e_1 \leq e_0$ so that if $0 < e_n < e_1$ then

$$|s - d_\|k(x_n, \xi_e)| < \delta$$

and $|x_n - x_0|$ is small enough so that

$$|d_\|k(x_n, \xi_e) - d_\|k(x_0, \xi_e)| < \delta$$

and

$$|d_\|k(x_0, \xi_e) - \lim_{e_n \to 0} d_\|k(x_0, \xi_e)| < \delta.$$  

(Recall that we have shown that $\lim_{e \to 0} d_k(x, \xi_e)$ exists for each $x$.) Therefore

$$|s - \lim_{e \to 0} d_k(x_0, \xi_e)| < 3\delta, \text{ all } \delta > 0.$$  

So $\max_x \lim_{e \to 0} d_k(x, \xi_e) = s$. But also

$$\max_x \lim_{e \to 0} d_k(x, \xi_e) \leq \lim_{e \to 0} \max_x d_k(x, \xi_e) = s.$$  

Therefore $\max_x \lim_{e \to 0} d_k(x, \xi_e) = s$.

We next prove that (3.1) implies (3.2).
3. Optimal Designs for Estimating $\theta(1)$

In this chapter we first prove some general results for estimating $s$ out of $k$ parameters and then apply them in the case of special $n$-tic regression on the $q$-simplex.

As always we assume below that the regression functions $f_i$ are continuous and that $X$ is compact.

**THEOREM 3.1.** Let $\xi^*$ be any design and $\xi'$ any design such that $M(\xi_c)$ is nonsingular for $0 < c < 1$, where $\xi_c = (1-c)\xi^* + c\xi'$. Then the statements (3.1) and (3.2) are equivalent, and either implies (3.3).

(3.1) $\lim_{c \to 0^+} \max_{x \in X} d_s |k(x, \xi_c)|$ exists, and $\max_{x \in X} \lim_{c \to 0^+} d_s |k(x, \xi_c)| = s$.

(3.2) $\lim_{c \to 0^+} \max_{x \in X} d_s |k(x, \xi_c)|$ exists and $= s$.

(3.3) $\xi^*$ is D-optimal for estimating $\theta(1)$.

**PROOF.** We prove first that (3.2) implies (3.1).

If $\lim_{c \to 0^+} \max_{x \in X} d_s |k(x, \xi_c)| = s$ there is some $c_0 > 0$ with $\max_{x \in X} d_s |k(x, \xi_c)| \leq s + 1$ for $c \leq c_0$. Therefore $0 \leq d_s |k(x, \xi_c)| \leq s + 1$ for $x \in X, 0 < c \leq c_0$. For each $x$, $d_s |k(x, \xi_c)|$ is a rational function in $c$. It is bounded as $c \to 0$, and therefore $\lim_{c \to 0} d_s |k(x, \xi_c)|$ exists.

We now want to show that for $\delta > 0$ and $c < c_0$, if $|x - x'|$ is sufficiently small then

(3.4) $|d_s |k(x, \xi_c)| - d_s |k(x', \xi_c)| < \delta$.

Write

$$d_s |k(x, \xi_c)| = \sum_{i,j=1}^{N} a_{ij}(c)f_i(x)f_j(x) = \sum_{i=1}^{N} b_i(c)F_i(x),$$

where the $F_i$'s are linear combinations of the terms $f_i f_j$, chosen to be linearly independent. Then there is a set $\{x_1, \ldots, x_N\}$ so that $F$, the matrix
An optimal design when \( n < q + 1 \) is not known. As has been shown in the second special case of Theorem 2.2 any optimal design must be supported on the barycenters. The fact that \( d(z, \xi_0) > k \) when \( z \) is a barycenter of depth \( n + 1 \) suggests that an optimal design would assign positive measure to these points, but it is not known whether this is correct, or what weights should be used.
Because \( n < q + 1 \) there is a barycenter of depth \( p + m = n + 1 \).

Evaluated there the expression in brackets in (2.1) equals

\[
\sum_{s=0}^{m-1} (-1)^s s^{-1} (p + s)(s)(p + m)^{-s} = 0 - (-1)^m m^{-1}(p + m)^{-m-1} \quad \text{using (2.2)}
\]

\[
= (-p)(n-p)(n+1)^{-n-p}.
\]

Therefore at such a barycenter \( z \)

\[
k^{-1}d(z, \xi^0) = \sum_{i=1}^{k} g_i^+(z) = \sum_{p=1}^{n} (n+1)(\frac{n}{n+1})^{2n}.
\]

As a lower bound on this we consider the upper two terms,

\[
(n + 1)(\frac{n}{n+1})^{2n} + \frac{n(n + 1)}{2}(\frac{n-1}{n+1})^{2n}
\]

which we write as \( A + B \). When \( n = \frac{1}{4} \) a direct computation yields

\[
A + B = \frac{78658}{78125} > 1.
\]

For \( n \geq 4 \) we show as follows that \( A + B \) is monotone increasing (in fact it increases without limit).

Treating \( n \) as a continuous parameter,

\[
\frac{d}{dn} \log A = \frac{3}{(n+1)} - 2 \log (1 + 1/n) > \frac{3}{(n+1)} - \frac{2}{n}.
\]

Therefore \( A \) is monotone increasing for \( n \geq 2 \). In fact

\[
A = (n + 1)(1 + 1/n)^{-2n} \sim (n + 1)e^{-2}
\]

so \( A \to \infty \).

Likewise,

\[
\frac{d}{dn} \log B = \frac{6n^2 - n - 1}{n(n-1)(n+1)} - 2 \log (1 + 2/(n-1)) > \frac{6n^2 - n - 1}{n(n-1)(n+1)} - \frac{1}{n-1}
\]

\[
= \frac{2n^2 - 5n - 1}{n(n-1)(n+1)}.
\]

The numerator of the last expression has zeros \( (5 + 33^{1/2})/4 \), so \( B \) is monotone increasing for \( n \geq 3 \). Therefore \( d(z, \xi^0) > k \) for \( n \geq 4 \), completing the proof.
(1/n, ..., 1/n, 0, ..., 0) and their images under permutations of the coordinates. These will sometimes be referred to as the barycenters of depth \( \leq n \). The number of such points is \( \sum_{p=1}^{n} (q+1)^p = k \), the number of regression functions. 

Trivially \( \xi^0 \) is optimal for any \( q \) when \( n = 1 \). It was shown by Kiefer [1961, p. 320] that \( \xi^0 \) is optimal for all \( q \) when \( n = 2 \), and by Uranisi [1964] that \( \xi^0 \) is optimal for all \( q \) when \( n = 3 \). The natural conjecture for greater \( n \) is shown here to be only partly true.

**THEOREM 2.4.** For special \( n \)-tic regression on the \( q \)-simplex, \( 4 \leq n \leq q+1 \), the design \( \xi^0 \) is optimal when \( n = q+1 \) and is not optimal when \( n < q+1 \).

**PROOF.** Optimality when \( n = q + 1 \) is just the second special case of Theorem 2.2. We now prove that \( \xi^0 \) is not optimal when \( 4 \leq n < q + 1 \).

First observe that if \( L \) is a nonsingular matrix, \( g(x) = Lf(x) \), and \( N(\xi) = \int g(x)g'(x)d\xi(x) \), then \( \xi^* \) maximizes \( \det N(\xi) \) if and only if it maximizes \( \det M(\xi) \) and \( g'(x)N^{-1}(\xi)g(x) = f'(x)M^{-1}(\xi)f(x) \) identically in \( x \) and \( \xi \). The following \( k \) functions \( g_i(x) \) are so obtained.

\[
(2.1) \quad p^p x_1 \ldots x_p \left[ a_0 + a_1, a_2, \ldots, a_{n-p}, \Sigma x_j x_{j_1} \ldots x_{j_{n-p}} \right].
\]

Here \( 1 \leq p \leq n \), \( a_s, p = (-1)^s p^{-1}(p+s) \), \( 1 \leq i_1 < \ldots < i_p \leq q+1 \), and in each summation \( 1 \leq j_1 < j_2 < j_3 < \ldots < q+1 \) and none of the \( j \)'s are in \( \{ i_1, \ldots, i_p \} \).

It is immediate that such a polynomial equals 1 at the barycenter which has all coordinates zero except for \( i_1, \ldots, i_p \), and that the polynomial equals 0 at all other barycenters of depth \( \leq p \). At a barycenter of depth \( p + m \leq n \), \( m > 0 \), the expression in brackets is equal to:

\[
(2.2) \quad \sum_{s=0}^{m} (-1)^s p^{-1}(p+s)(p+m)^{-s} = (1 - \frac{p}{p+m})^{-m} \frac{d}{dx} \bigg|_{x=-p} \sum_{s=0}^{m} (s^m x^{p+m})^{-s}
\]

which equals 0. Therefore the polynomials are orthogonal with respect to \( \xi^0 \) and the sum of their squares equals \( k^{-1}d(x, \xi^0) \).
xy, x^2y, x^2y^2z, etc.) Of course n ≤ q. The number of such functions is
\[ \sum_{i=0}^{n} \binom{q}{i} x^i. \]

**THEOREM 2.3.** Let the regression functions be as described above, and let X be any space in \( \mathbb{R}^q \) which for some coordinate \( x_i \) is symmetric under interchange of \( x_i \) and \(-x_i\). Let I be a line segment in X on which all coordinates but \( x_i \) are constant and which is symmetric under interchange of \( x_i \) and \(-x_i\). Then an optimal design \( \xi \) can have points of support on I only at the midpoint and at the end points.

**PROOF.** As before \( d(x, \xi) \to \infty \) as \(|x| \to \infty \) and \( d(x, \xi) \) is at most quartic in \( x_i \). So \( d(x, \xi) \) can have at most one interior maximum in I. If \( \pi \) is the map which just interchanges \( x_i \) and \(-x_i\), then as in the last theorem \( d(x, \xi) \) is symmetric under \( \pi \), because \( \xi \) is optimal. Hence \( d(x, \xi) \) can attain its maximum value on I only at the midpoint and the end points.

**SPECIAL CASES OF X.**

1. X any convex body which for each \( x_i \) is symmetric under interchange of \( x_i \) and \(-x_i\) (e.g. q-ball).

Any optimal design is supported on the boundary and the origin. This result does not overlap the first special case of Theorem 2.2 because the functions are different in the two cases.

2. X the q-cube.

Any optimal design is supported on the lattice of points with coordinates only 0 or ±1. (There are \( 3^q \) such points.) If \( n = q \) the unique optimal design is uniform on this lattice.

Again we note that 2 is a special case of 1.

We conclude this chapter by considering the example of special n-tic regression on the q-simplex.

The design \( \xi^0 \), which was introduced by Scheffé [1958], assigns equal measure to the points on the simplex \((1,0,\ldots,0), (1/2,1/2,0,\ldots,0), \ldots,\)
Now if \( \pi \) is the permutation which just interchanges \( x_1 \) and \( x_2 \), there is a linear mapping \( L \) so that \( g(x) = Lf(x) = f(\pi x) \). And since \( \xi \) is optimal, it has the same information matrix as an invariant optimal design \( \xi' \), by Theorems 1.3 and 1.1. Therefore

\[
\begin{align*}
d(x, \xi) &= f'(x)L'L^{-1}([\int f(y)f'(y)d\xi'(y)]L^{-1}L f(x) \\
&= g'(x)[\int g(y)g'(y)d\xi'(y)]^{-1}g(x) \\
&= f'(\pi x)[\int f(\pi y)f'(\pi y)d\xi'(\pi y)]^{-1}f(\pi x) \\
&= d(\pi x, \xi).
\end{align*}
\]

That is, \( d(x, \xi) \) is symmetric, and thus can have maxima in \( I \) only at the midpoint and end points.

**SPECIAL CASES OF** \( X \).

1. \( X \) any convex body symmetric under all permutations of the coordinates (e.g., q-ball), constrained by \( \Sigma x_1 = \alpha \).

Any optimal design is supported on the boundary of \( X \) and the center point of \( X \).

2. \( X \) the q-simplex.

Any optimal design is supported on the barycenters. If \( n = q + 1 \) the unique optimal design is uniform on the barycenters (because the number of functions equals the number of points of support).

3. \( X \) determined by \( 0 \leq x_i \leq 1, \Sigma x_1 = \alpha, 1 < \alpha < q+1 \).

Any optimal design is supported on points of the form \( (0, \ldots, 0, 1, \ldots, 1, \beta, \ldots, \beta) \) with \( \beta = (\alpha - N_1)/(q+1 - N_0 - N_1) \), where \( N_0 \) is the number of 0's and \( N_1 \) the number of 1's. This space is considered in a different problem by Keilson [1966].

We remark that 2 and 3 are special cases of 1.

For the last theorem of this type we assume that the regression functions are the monomials in \( q \) independent variables which involve at most \( n \) variables, and which in each variable are at most quadratic (e.g., \( 1, x, x^2 \),
\[ |x| \to \infty \text{ so does } |f(x)| \text{ and therefore so does } d(x, \xi). \] Since \( f(x) \) is multilinear \( d(x, \xi) \) is at most quadratic in each variable. Now hold all the variables but one constant. The restriction of \( d(x, \xi) \) to this line is a non-negative unbounded quadratic function. Therefore it is strictly convex and cannot have a maximum in the interior of any segment of the line.

For \( f(x) \) as above we list two corollaries.

**SPECIAL CASES OF \( X \).**

1. \( X \) an arbitrary set in \( \mathbb{R}^q \).
   
   Any optimal design is supported on the boundary of \( X \).

2. \( X \) a set whose convex hull is the \( q \)-cube.
   
   Any optimal design is supported on the vertices of the cube. If \( n = q \) the unique optimal design is uniform on the vertices (because the number of regression functions equals the number of points of support.) If \( n < q \), the design uniform on the vertices is the unique optimal design which is invariant under the group of symmetries of the cube.

**THEOREM 2.2.** Suppose the regression is special \( n \)-tic on a set \( X \) in \( \mathbb{R}^{q+1} \) satisfying the linear constraint \( \sum x_i = \alpha \) and symmetric under the interchange of some two coordinates. Let \( I \) be any line segment in \( X \) which is invariant under the interchange of those same two coordinates, and on which all coordinates but those two are constant. Then an optimal design \( \xi \) can have points of support on \( I \) only at the midpoint and at the end points.

**PROOF.** Let \( \xi \) be optimal. As before \( d(x, \xi) \to \infty \) as \( |x| \to \infty \) and \( d(x, \xi) \) is at most quadratic in each variable. Without loss of generality \( X \) and \( I \) are symmetric under interchange of \( x_1 \) and \( x_2 \). Since \( \sum x_i = \alpha \) and since \( x_3, \ldots, x_{q+1} \) are constant on the line containing \( I \), the restriction of \( d(x, \xi) \) to this line when expressed in terms of one variable, say \( x_1 \), is at most quartic. Since \( d(x, \xi) \) is non-negative and not constant on the line, it can have at most one interior maximum in \( I \).
2. Optimal Designs for Estimating $\theta$

When the $k$ regression functions are the monomials of degree $\leq k - 1$ on an interval the optimal design is supported on $k$ points, as described by Guest [1958] and Hoel [1958]. If more than one variable is present, natural generalizations of the interval might be the cube, the ball and the simplex. Of these, the simplex seems to give the most analogous and simplest results. Farrell, Kiefer and Walbran [1965] compare the spaces in some detail. If the regression functions are all the monomials of degree $\leq n$ the problem is, in general, quite difficult. However if the regression functions belong to a more restricted class we can sometimes obtain extensive results.

We begin with three theorems in which symmetry of $X$ and restrictions on the regression functions enable us to make assertions about the support of the optimal designs. For the difficulty in proving analogues for estimation of $\theta^{(1)}$, see the lemma preceding Theorem 3.5.

Scheffe [1958, p. 352] introduced the special n-tic polynomials, defined as the multilinear polynomials of total degree $\leq n$. Under special n-tic regression we may take the regression functions $f_i(x)$ to be of the form $x_{i_1} \ldots x_{i_p}$, with $i_1 < \ldots < i_p$ and $p \leq n$. If there are $q$ independent variables then $p \geq 0$, of course $n \leq q$, and the number of regression functions is $k = \sum_{p=q}^{n} \binom{n}{p}$. If there are $q + 1$ variables constrained by $\sum x_i = \alpha$ then $p \geq 1$, $n \leq q + 1$, and the number of functions $k = \sum_{p=1}^{n} \binom{n}{q+1}$. 

THEOREM 2.1. Suppose the regression is special n-tic on a set $X$ in euclidean q-space $\mathbb{R}^q$. Then an optimal design $\xi$ can have no points of support in the interior of any line segment in $X$ on which all the variables but one are constant.

PROOF. Suppose $\xi$ is optimal. Then $M^{-1}(\xi)$ exists and is positive definite and $d(x, \xi) = f'(x)M^{-1}(\xi)f(x) \geq \lambda |f(x)|^2$, where $\lambda > 0$ is the smallest eigenvalue of $M^{-1}(\xi)$ and $|f(x)|$ is the euclidean norm of $f(x)$. As
optimally) may be any number from \( s \) to \( k \). It is not hard to show that if \( g^{(1)} \) is estimable using an \( s \)-point design then \( f^{(2)} \) is zero on the support of such a design, and the best such design is uniform on \( s \) points.

For let \( \xi_s \) be any \( s \)-point design. There is a nonsingular matrix

\[
L = \begin{bmatrix}
I & L_3 \\
0 & L_2
\end{bmatrix}
\]

with

\[
LM(\xi_s)L' = \begin{bmatrix}
M^*(\xi_s) & 0 & 0 \\
0 & N_3 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

where \( N_3 \) is a square nonsingular matrix of size \( \geq 0 \). But \( \text{rank } LM(\xi_s)L' = \text{rank } M(\xi_s) \leq s \). If \( \xi_s \) estimates \( b^{(1)} \), rank \( M^*(\xi_s) = s \) and therefore \( N_3 \) must have size \( 0 \). So \( f^{(2)} \) is 0 on the support of \( \xi_s \) and \( M^*(\xi) = M_1(\xi) \) for any \( \xi \) supported on these points. The best such \( \xi \) is therefore uniform.
The following theorem holds (Kiefer [1959, p. 296], [1960, p. 387], and [1961, p. 302]).

**THEOREM 1.3.** Under the above assumptions there is a $G$-invariant optimum design.

**NUMBER OF POINTS NEEDED FOR OPTIMALITY.** For a given $f$ and $X$ we consider the space of all possible $M(\xi)$, which we denote $\Gamma$. (See Kiefer [1960, p. 389], and Farrell, Kiefer, and Walbran [1965, p. 114 ff.].) There are at most $k(k+1)/2$ distinct functions $f_i f_j$, so $\Gamma$ has dimension at most $k(k+1)/2$. Since $M([1-t]f_1 + tf_2) = [1-t]M(f_1) + tM(f_2)$ for $0 \leq t \leq 1$, $\Gamma$ is convex and the extreme points are contained in the set of $M(\xi)$ with $\xi$ supported on a single point. Since $f$ is continuous and $X$ compact $\Gamma$ is compact, and $\max M(\xi)$ attains its maximum. It is easily seen that this maximum can only be attained at a boundary point of $\Gamma$. Therefore for estimating $\theta$ there is an optimum design supported on at most $k(k+1)/2$ points.

An example is given in this thesis for which all $k(k+1)/2$ points are needed.

The lower bound on the number of points needed for optimality is clearly $k$ when estimating $\theta$, since $M(\xi)$ must be of full rank. We remark that if a design supported on $k$ points is optimal, the design must be uniform on those points.

When estimating $\theta^{(1)}$ we might look at the space of all $M^*(\xi) = M^{(1)}(\xi)^{-1}$. This gives $s$ as a lower bound on the number of points needed, but we cannot obtain an upper bound as before because $M^*(\xi)$ is not linear in $\xi$. However Chernoff [1953, p. 590 ff.] gives an argument which can be applied to this problem, as does Stone [1959, p. 68]. The upper bound obtained is $s(s+1)/2 + rs$. See also Kiefer [1961, p. 303]. An example is given in Chapter 4 for which this bound is attained.

The minimum number of points needed to estimate $\theta^{(1)}$ (not necessarily...
As before it is immediate that if \( d(s|k(x, \xi^*) \) is defined for an optimal \( \xi^* \) (i.e., \( M(\xi^*) \) is nonsingular), then \( \xi^* \) is supported at most on those points satisfying \( d(s|k(x, \xi^*) = s \).

As with Theorem 1.1, it is shown that for \( M(\xi^*) \) nonsingular the three conditions of this theorem are equivalent to

\[
(iv) \frac{\partial}{\partial \alpha} \log \det M^*\left([1 - \alpha] \xi^* + \alpha \xi\right) \bigg|_{\alpha=0^+} < 0 \text{ for all } \xi.
\]

If \( M(\xi^*) \) is singular there is no simple known theorem analogous to the first part of the above. Karlin and Studden [ibid.] state one but it requires modification as discussed in Chapter 3 of this paper. Their sufficient condition for optimality seems to be a useful algorithm, but the condition equivalent to optimality will probably be of theoretical rather than computational use. Kiefer [ibid.] gives a condition equivalent to optimality, a necessary condition, and a sufficient condition. These are of use in some but not all problems. One of the results of this thesis is a new sufficient condition for optimality.

**INVARIANCE:** Suppose \( G \) is a compact group of transformations on \( X \). Let \( \tilde{G} \) be a group of linear transformations on the space of \( \theta \) of the form

\[
A = \begin{bmatrix}
B & 0 \\
0 & C
\end{bmatrix}
\]

where \( B \) is an \( s \times s \) matrix of determinant 1. (When estimating all of \( \theta \), i.e., \( s = k, B \) is the entire \( k \times k \) matrix \( A \).) Assume there is a homomorphism from \( G \) to \( \tilde{G} \). (Note then that \( A g_1 g_2 = A A_1 A_2 \).) And suppose \( \theta^t f(x) = (A \theta)^t f(gx) \).

When estimating all of \( \theta \) we may make the following alternative assumptions (as in Kiefer [1960]). Assume \( G \) is a compact group on \( X \), \( \tilde{G} \) is a group of transformations on the space of \( \xi \), there is a homomorphism from \( G \) to \( \tilde{G} \), and (with obvious notation) \( d(x, \xi) = d(gx, \tilde{G} \xi) \).

A design \( \xi \) is called \( G \)-invariant if \( \xi(gB) = \xi(B) \) for all \( g \) in \( G \) and Borel sets \( B \).
where $\theta(1)$ and $f(1)$ are $s$-vectors and $\theta(2)$ and $f(2)$ are $r$-vectors, with $r = k - s$. As before $EY(x) = \theta'f(x)$, but this time we are interested in estimating only $\theta(1)$ rather than all of $\theta$. If $M(\xi)$ is nonsingular, so that all of $\theta$ is estimable under $\xi$, the inverse below can be written directly. If $\theta(2)$ is not estimable, i.e., $M_3$ is singular, we understand that $M^{-1}$ is a pseudo-inverse. (This is discussed in Chernoff [1953] and in Chapter 3 of this thesis.) We write

$$M(\xi) = \begin{bmatrix} M_1(\xi) & M_2(\xi) \\ M_2(\xi) & M_3(\xi) \end{bmatrix}$$

$$M^{-1}(\xi) = \begin{bmatrix} M_1(\xi) & M_2(\xi) \\ M_2(\xi) & M_3(\xi) \end{bmatrix}^{-1}$$

where $M_1(\xi)$ and $M_1(\xi)$ are $s \times s$ matrices. The covariance matrix of the best linear estimator of $\theta(1)$ is then $\sigma^2N^{-1}M_1(\xi)$. A design $\xi^*$ is called D-optimal for estimating $s$ out of $k$ parameters if it minimizes $\det M_1(\xi)$. We will sometimes write $M^*(\xi) = [M_1(\xi)]^{-1} = M_1(\xi) - M_2(\xi)M_3^{-1}(\xi)M_2(\xi)$. Although there is no natural optimality criterion analogous to $G$-optimality, if $M(\xi)$ is nonsingular we still define

$$d_s|k(x, \xi) = f'(x)M^{-1}(\xi)f(x) - f(2)'(x)M_3^{-1}(\xi)f(2)(x)$$

$$= (f(1)(x) - M_2M_3^{-1}f(2)(x))'M_1(\xi)(f(1)(x) - M_2M_3^{-1}f(2)(x))$$

where in the second expression we have suppressed the $\xi$ for greater legibility. The two expressions are those used by Kiefer [1961] and Karlin and Studden [1966a] and [1966b] respectively. We now state the equivalence theorem (Kiefer [1961, pp. 305, 310]) analogous to the theorem above.

**Theorem 1.2.** If $M(\xi^*)$ is nonsingular, conditions (i), (ii), and (iii) are equivalent.

(i) $\det M_1(\xi^*) = \min_\xi \det M_1(\xi)$ (D-optimality)

(ii) $\max_x d_s|k(x, \xi^*) = \min_\xi \max_x d_s|k(x, \xi)$

(iii) $\max_x d_s|k(x, \xi^*) = s$.

In any case the set of D-optimal $\xi$ is convex, and $M_1(\xi)$ is the same for all optimal $\xi$. 

is the covariance matrix of the best linear estimator of \( \theta \). A design \( \xi^* \) is called D-optimal if it minimizes \( \det M^{-1}(\xi) \).

Define the function

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

If \( M(\xi) \) is nonsingular, the variance of the best linear estimator of \( E Y(x) = \theta'f(x) \) is \( \sigma^2N^{-1}d(x, \xi) \). A design \( \xi^* \) is called G-optimal if it minimizes \( \max_{x \in X} d(x, \xi) \). (We will simply write \( \max_{x} \).

These optimality criteria were shown to be equivalent in the following theorem of Kiefer and Wolfowitz [1960].

**THEOREM 1.1.** The conditions (i), (ii), and (iii) are equivalent.

(i) \( \det M^{-1}(\xi^*) = \min_{\xi} \det M^{-1}(\xi) \) (D-optimality)

(ii) \( \max_{x} d(x, \xi^*) = \min_{\xi} \max_{x} d(x, \xi) \) (G-optimality)

(iii) \( \max_{x} d(x, \xi^*) = k \)

The set \( B \) of all \( \xi \) satisfying these conditions is convex and \( M(\xi) \) is the same for all \( \xi \) in \( B \).

Since \( \int d(x, \xi)d\xi = \int M^{-1}(\xi)f(x)f'(x)d\xi = k \), it follows immediately that an optimal design \( \xi^* \) is supported (at most) on the points \( x \) satisfying \( d(x, \xi^*) = k \).

The theorem is proved by showing that the conditions are equivalent to a fourth condition

(iv) \( \frac{\partial}{\partial \alpha} \log \det M([1 - \alpha]\xi^* + \alpha\xi') |_{\alpha=0^+} \leq 0 \) for all \( \xi \).

This condition says that \( \xi^* \) gives a local maximum of \( \det M(\xi) \). A calculation in the proof which we shall use is that

\[ \frac{\partial}{\partial \alpha} \log \det M([1 - \alpha]\xi + \alpha\xi') |_{\alpha=0^+} = \text{tr} M^{-1}(\xi)M(\xi') - k. \]

The model discussed so far can be generalized as follows (see Kiefer [1961].)

We write

\[ \theta = \begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \end{pmatrix}, \quad f(x) = \begin{pmatrix} f^{(1)}(x) \\ f^{(2)}(x) \end{pmatrix} \]
1. Preliminaries

BASIC MODEL. Let \( f_1, \ldots, f_k \) be \( k \) continuous real-valued linearly independent functions, called the regression functions, on a compact space \( X \). Let \( \theta_1, \ldots, \theta_k \) be \( k \) unknown parameters. We shall write these as column vectors \( \mathbf{f} \) and \( \mathbf{\theta} \). For any \( x \) in \( X \) we may observe a random variable \( Y(x) \) with mean \( \theta'f(x) = \sum \theta_i f_i(x) \). (Throughout this thesis primes will denote transposes.) The variance \( \sigma^2 \) of \( Y(x) \) may be known or unknown but it is fixed independent of \( x \). The observed \( Y(x_i) \) and \( Y(x_j) \) are uncorrelated (whether \( x_i \) and \( x_j \) are the same point or not). We assign various values to \( x \) and make \( N \) observations in all.

A design \( \xi \) is a probability measure on \( X \). If \( \xi \) assigns to points probabilities which are all multiples of \( 1/N \), \( \xi(x) \) will be the proportion of observations of \( Y \) at \( x \). Such a \( \xi \) is called an exact design. A choice of the best exact design often presents a difficult combinatorial problem, perhaps depending on the number-theoretic properties of \( N \). If we do not require \( \xi \) to be exact (or even discrete) there are some simple and very useful theorems for obtaining an optimal design. Moreover any design \( \xi \) can be approximated by an exact design \( \xi_N \). If \( \xi \) is optimal \( \xi_N \) will be optimal to within order \( 1/N \). We will henceforth only consider this approximate or continuous theory, in which there are no restrictions on \( \xi \).

OPTIMALITY CRITERIA. A number of criteria for optimality of a design have been suggested. See Kiefer [1960, p. 383 ff.] for a discussion of some of these. We consider only two.

The information matrix of a design \( \xi \), denoted \( M(\xi) \), has components defined by

\[
m_{ij}(\xi) = \int f_i(x)f_j(x)d\xi(x).
\]

We will sometimes write \( M(\xi) = \int f(x)f'(x)d\xi(x) \). If \( M(\xi) \) is nonsingular, so that all the components of \( \mathbf{\theta} \) are estimable under \( \xi \), then \( \sigma^2_{N^{-1}N^{-1}(\xi)} \)
TABLE OF CONTENTS

Chapter 1. Preliminaries .................. 1

Chapter 2. Optimal Designs for
    Estimating $\theta$ ..................... 7

Chapter 3. Optimal Designs for
    Estimating $\theta^{(1)}$ ................ 14

Chapter 4. Efficiency of Various Designs ... 40

References .................................. 75
ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to Professor Jack Kiefer for his interest and helpful guidance while this thesis was being written.

I wish to thank also Mr. George Duncan for his check of the computation in Example 4.2 when \( r = s = 2 \), and Mr. Gerald DuChaine for his careful typing of the manuscript.

The research was supported in part by the Office of Naval Research under contract number Nonr-401(50).
April, 1968

OPTIMAL AND EFFICIENT DESIGNS
OF EXPERIMENTS

by

Corwin L. Atwood
Technical Report No. 106

University of Minnesota
Minneapolis, Minnesota

At last we use equation (3.14) here, letting \( t = q + 1 \), to get

\[
\sum_{i_1 < \ldots < i_{n-1}, i_{n-1}, q+1 = \sum_{i_1 < \ldots < i_{n-2}, i_{n-2}, q+1 :}
\sum_{i_{n-2} < \ldots < i_{n-1}, q+1 :}
\sum_{i_j < q}
\sum_{i_j < q}
\sum_{i_j < q}
\sum_{i_j < q}
\text{There is a similar identity if we reverse the roles of } q \text{ and } q + 1. \text{ Therefore (3.17) becomes}
\]

\[
u - A \left[ - \sum_{i_1 < \ldots < i_{n-2}, i_{n-2}, q+1 :}
\sum_{i_j < q}
\sum_{i_j < q}
\sum_{i_j < q}
\sum_{i_j < q}
\left( u + v \right) \right] = K.
\]

But \( u + v = 1 - \sum a_i \), a constant. Therefore \( u \) is a constant, a contradiction.

The proof is much simpler if \( n = 2 \). In that case the sole quadratic coefficient of (3.13) is \( c_{h,t} \), which must be zero because (3.13) is constant. This holds for any \( h \neq t \). Therefore (3.13) immediately reduces to \( x_q \). This is constant, but if \( \tau(q) = q \), \( x_q = u \), a variable. This is the desired contradiction.

Therefore \( d_s(x, \xi^*, D) \) cannot be bounded on \( L \).

It is not clear whether the conclusion of the lemma holds for other values of \( m \) and \( n \). If \( m = 1 \), \( n = 3 \), and \( q + 1 = 3 \), it is not hard (at least compared to the proof just completed) to show that the conclusion holds. However setting \( q + 1 = 3 \) is such a great simplification that this result probably gives no clue about the general case.

We obtain a result from the lemma which is given as a theorem.

**THEOREM 3.5.** In the model described above, special polynomial regression on the simplex with \( n = m + 1 \), the support of any optimal design for \( \phi^{(1)} \) is a subset of the barycenters.
PROOF. Let $G$ be the group of permutations of the coordinates of the simplex. Define $D$ as in Theorem 3.3. Let $\xi^*$ be optimal. On any line $L$ determined by setting all but two $x_i$ constant and holding $\Sigma x_i = 1$, $d_s(x, \xi^*, D)$ is unbounded, by the lemma just proved. Now $d_s(x, \xi^*, D)$ is at most quadratic in each variable, because the components of $f^{(1)}(x)-D'f^{(2)}(x)$ are multilinear in the $x_i$. Therefore the restriction of $d_s(x, \xi^*, D)$ to $L$, when expressed in terms of a single variable, is nonnegative, unbounded and at most quartic in terms of that variable. Thus if $I$ is the line segment defined as the intersection of $L$ and the simplex, $d_s(x, \xi^*, D)$ can attain a maximum on $I$ at most at the end points and at one interior point. But $d_s(x, \xi^*, D)$ is symmetric, by Theorem 3.3, and so can attain a maximum on $I$ at most at the end points and the midpoint. Thus $d_s(x, \xi^*, D)$ can attain its maximum in the simplex only at the barycenters, where it equals $s$.

Since $\int d_s(x, \xi^*, D)d\xi^* = s$, $\xi^*$ must be supported on a subset of the barycenters.

The result just proved is much weaker than the desired statement which would involve no restriction on $n$ and $m$. However as has been noted the difficulty lies in obtaining the conclusion of the lemma without assuming such a restriction.

For the rest of this chapter we will investigate the optimality of $\xi^0$, introduced in Chapter 2, when estimating $\theta^{(1)}$ in the case of special polynomial regression on the simplex. Let $X$, $f$ and $f^{(1)}$ be as described just before the lemma. Let $z_1, \ldots, z_s$ denote the barycenters of depth $\leq m$, and $z_{s+1}, \ldots, z_k$ those of depth from $m+1$ to $n$. Let $L$, of the form (3.9), be the nonsingular $k \times k$ matrix defined explicitly by letting the $i^{th}$ component of $g(x) = Lf(x)$ be (2.1). Then $g_i(z_j) = \delta_{ij}, 1 \leq i, j \leq k$. Let $\xi^0$ be the design uniform on $\{z_1, \ldots, z_s\}$. 
THEOREM 3.6. If $m \leq 3$, $\xi^o$ is optimal for $\theta^{(1)}$.

PROOF. This is immediate from Theorem 3.4 (setting $A=I$ and $B=0$ there) and the facts stated just before Theorem 2.4.

THEOREM 3.7. If $n = q + 1$, $\xi^o$ is optimal for $\theta^{(1)}$.

PROOF. Let $\xi'$ be uniform on $\{z_{s+1}, \ldots, z_k\}$, and let $\xi = (1-\varepsilon)\xi^o + \varepsilon\xi'$.

Then

$$d_s|k(x, \xi) = g'(x)[\int g g'd\xi]\cdot g(x) - g^{(2)'}(x)[\int g^{(2)}g^{(2)'}d\xi]\cdot 1g^{(2)}(x)$$

$$= g'(x) \begin{bmatrix} (1-\varepsilon)s^{-1} & 0 \\ 0 & \varepsilon r^{-1} \end{bmatrix}^{-1} g(x) - g^{(2)'}(x)[\varepsilon r^{-1}]^{-1} g^{(2)}(x),$$

where $(1-\varepsilon)s^{-1}I$ is $s \times s$ and $\varepsilon r^{-1}I$ is $r \times r$. Thus

$$d_s|k(x, \xi) = s(1-\varepsilon)^{-1}\Sigma_{i=1}^s g^2_i(x).$$

This is symmetric, nonnegative and at most quadratic in each variable $x_i$. Therefore it attains its maximum at a barycenter. (If the expression is constant on some line, it may also attain its maximum elsewhere.) But since $n = q + 1$, the only barycenters are $z_1, \ldots, z_k$, and there $\Sigma g^2_i(x) = 0$ or 1. Therefore $\lim_{\varepsilon} \max_x d_s|k(x, \xi) = s$, and $\xi^o$ is optimal by Theorem 3.1.

If we were to try to prove Theorem 3.7 using the corollary to the Karlin-Studden Theorem, we would have to find an $r \times s$ matrix $X$ such that

$$d_s(x, \xi^o, X) = f'(x)(I, -X')'[M^*(\xi^o)]^{-1}(I, -X')f(x)$$

is bounded by $s$. Let

$$A = \begin{bmatrix} L_1 & 0 \\ 0 & L_3 \end{bmatrix}$$

where $L_1$ and $L_3$ are the diagonal blocks of the matrix $L$ defined just before Theorem 3.6. One need not use the transformation $A$, but it seems to reduce the problem to the simplest form. Write $h = Af$. Then $h_i(z_j) = \delta_{ij}$,
\[ d_s(x, \xi^0, X) = f'(x)A' A^{-1}(I, -X')' L_1^{-1} [M^*(\xi^0)]^{-1} L_1^{-1}(I, -X') A^{-1} A f(x) \]

\[ = h'(x)(I, -L_1 X' L_2^{-1})'[s^{-1} l^{-1}(I, -L_1 X' L_2^{-1}) h(x) \]

\[ = s(h(1)(x) - Y' h(2)(x))'(h(1)(x) - Y' h(2)(x)), \]

defining \( Y' = L_1 X' L_2^{-1} \). It is not immediately clear how to choose \( Y \) so that the expression is bounded by \( s \) for all \( x \) in the simplex. In particular \( s Y = 0 \) does not work, because \( \sum h_i^2(x) > 1 \) at barycenters of depth \( s + 1 \), as was shown in the proof of Theorem 2.4.

The last theorem about \( \xi^0 \) gives a case in which it is not optimal.

**Theorem 3.8.** For \( q + 1 - n \) fixed \( > 0 \) and \( n - m \) fixed \( > 0 \) and \( n \) sufficiently large, \( \xi^0 \) is not optimal for \( \theta(1) \).

If we only wished to prove the theorem we would merely have to find some design better than \( \xi^0 \), and the proof below could be simplified. Instead we consider a more complicated design, which in the special case \( m + 1 = n = q \) is the most general symmetric competitor of \( \xi^0 \). Thus we not only prove the theorem, but also determine in this special case how large \( n \) must be before \( \xi^0 \) is not optimal.

**Proof.** \( \xi^0 \) is uniform on \([z_1, \ldots, z_s]\). Let \( \xi^+ \) be any symmetric design on \([z_1, \ldots, z_s]\), \( \xi' \) uniform on \([z_{s+1}, \ldots, z_k]\), and \( \xi_0 \) concentrated on \( z_{k+1}' \), a barycenter of depth \( n + 1 \). Define \( \xi = \xi(b, c) = a_\xi^+ + b_\xi' + c_\xi_0 \), where \( a = \xi - b - c \) and \( a, b, c \geq 0 \). (If \( m + 1 = n = q \), \( \xi(b, c) \) will be the most general symmetric competitor of \( \xi^0 \).) Without loss of generality we may consider the regression functions to be the \( g_i \) defined above. Then

\[
M([1 - \alpha] \xi^0 + \alpha \xi) = (1 - \alpha)s^{-1} \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \alpha a \begin{bmatrix} M_1(\xi^+) & 0 \\ 0 & 0 \end{bmatrix} + \alpha \beta r^{-1} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} + \alpha c g(z_{k+1}) g'(z_{k+1}).
\]
There exist orthogonal matrices $R_1$ and $R_2$ with $(R_1g^{(1)}(z_{k+1}))' = (t_1, 0, \ldots, 0)$ and $(R_2g^{(2)}(z_{k+1}))' = (t_2, 0, \ldots, 0)$ where

$$
(3.18) \quad \ell_1^2 = \sum_{i=1}^s g_i^2(z_{k+1}) = \sum_{p=1}^m (n+1)(\frac{p}{n+1})^2n
$$

and

$$
(3.19) \quad \ell_2^2 = \sum_{i=s+1}^k g_i^2(z_{k+1}) = \sum_{p=m+1}^n (n+1)(\frac{p}{n+1})^2n.
$$

(The evaluation of the sums is as in the proof of Theorem 2.4.) Write

$$
R = \begin{pmatrix}
R_1 & 0 \\
0 & R_2
\end{pmatrix}.
$$

Suppressing the argument of $M([1 - \alpha]x_0 + \alpha x_0)$, we have $\det M^* = \det (RMR')^*$. After straightforward manipulation, $(RMR')^*$ reduces to

$$
(1 - \alpha)s^{-1} + \alpha[a R_1 M_1(\ell_1^+) R_1' + \|d\|] = \frac{d}{0} = 0
$$

where $d = bc\ell_2^2/(b + rct_2^2)$. Let $P$ be generic for a polynomial. We have

$$
\det M^* = (1 - \alpha)s^{-1} - s + (1 - \alpha)s^{-1} - s + \alpha[a \tr R_1 M_1(\ell_1^+) R_1' + d] + \alpha^2 P(\alpha).
$$

The trace of $R_1 M_1(\ell_1^+) R_1'$ equals 1, because $R_1$ is orthogonal and $g_i(z_j) = \delta_{ij}$ for $1 \leq i, j \leq s$. Hence

$$
\det M^* = (1 - \alpha)s^{-1} - s[1 - \alpha + \alpha a + \alpha d + \alpha^2 P(\alpha)].
$$

Thus

$$
\frac{d}{\alpha} \log \det M^* \bigg|_{\alpha^+ = 0} = s(-1 + a + d) = s(bc\ell_2^2/(b + rct_2^2) - b - c).
$$

Now using (iv) of Theorem 1.2, $\xi^0$ is at least as good as $\xi(b, c)$ for all $b, c \geq 0$ with $b + c < 1$ if and only if this derivative is $\leq 0$. This is so if and only if

$$
0 \leq b^2 - (t_1^2 - 1 - rt_2^2)bc + rt_2^2c^2, \quad \text{all } b, c \geq 0 \text{ with } b + c < 1,
$$

which is true if and only if
We now show that for \( q + 1 - n \) fixed \( > 0 \) and \( n - m \) fixed \( > 0 \),
\[
(3.20) \quad t^2_1 - (1 + r^{1/2} t^2_2)^2 \leq 0.
\]

We now show that for \( q + 1 - n \) fixed \( > 0 \) and \( n - m \) fixed \( > 0 \),
\[
(3.20) \quad t^2_1 - (1 + r^{1/2} t^2_2)^2 \to \infty \text{ as } n \to \infty,
\]
and therefore for \( n \) sufficiently large \( t^0 \) is not optimal. Write \( q - n = t \geq 0 \), \( n - m = u \geq 1 \).

Write \( \binom{n+1}{p}(\binom{n}{p})^2 = G_p \). Then \( t^2_2 = \sum_{p=n-u+1}^n G_p \). To bound \( t^2_2 \) note that for \( n \) large, \( G_n/G_{n+1} \) is asymptotic to \( n e^{-2/(j+1)} \). So
\[
t^2_2 \leq u G_{n-u+1} \text{ for } n \text{ large.}
\]

Write \( r = \sum_{p=m+1}^n \binom{q+1}{p} \leq u \binom{q+1}{m+1} \) for \( n \) large.

Also, \( t^2_2 = \sum_{p=1}^{n-u} G_p \). Let \( G_{n-v} \) be a typical term. Then
\[
\frac{t^2_1}{rt^2_2} \geq G_{n-v}/u^{(n+1)} \quad u G_{n-u+1} = \binom{n-u+1}{n-u+1} \binom{t+u}{v+1} \binom{n-v}{v+1} u^2
\]
which we break up into three parts as shown. If we fix \( v \) sufficiently large (\( v \geq 2u + t \) is enough), and let \( n \to \infty \), application of Stirling's formula shows that the first ratio tends to infinity. The second is constant and the third approaches a fixed power of \( e \). Therefore \( t^2_2/t^2_1 \to \infty \). Also \( t^2_2 \to \infty \), so
\[
t^2_1 - (1 + r^{1/2} t^2_2)^2 \to \infty, \text{ which proves the assertion.}
\]

We have proved that in certain cases \( t^0 \) is not optimal by finding a better design \( t(x, c) \). In the case \( m + 1 = n = q \), a \( t \) of this form is the most general symmetric competitor of \( t^0 \). There is certainly some symmetric optimal design, so in this case \( t^0 \) is optimal if and only if there is no better \( t(x, c) \). A computer calculation of the expression in (3.20) gives the result that \( t^0 \) is optimal for \( m + 1 = n = q \leq 50 \), and not optimal for \( m + 1 = n = q = 51 \) through 55. Further calculations were not performed.

If \( m > 3 \) and if \( n - m > 1 \) or \( q > n \) it is not known whether \( t^0 \) is ever optimal, although it seems likely to be optimal for \( n - m \) and \( q - n \) and \( n \) all small. It is also not known how large \( n \) must be before a design
of the form $\xi(b, c)$ is better than $\xi^0$. However $\xi^0$ is better than $\xi(b, c)$ if and only if (3.20) holds, so $n$ must certainly be greater than 50. For (see (3.18) and (3.19))

\[(3.21) \quad t_1^2 - (1 + r^{1/2} t_2)^2 = \sum_{p=1}^{m} g_p - [1 + r^{1/2}(\sum_{p=m+1}^{n} g_p)^{1/2}]^2.\]

But

\[r = \sum_{p=m+1}^{n} p^{q+1} \geq \sum_{p=m+1}^{n} (n+1) \geq n + 1,\]

so the expression on the right in (3.21) is no greater than

\[\sum_{p=1}^{n-1} g_p - [1 + (n + 1)^{1/2} g_n^{1/2}]^2.\]

This last expression is $t_1^2 - (1 + r^{1/2} t_2)^2$ for the case $m + 1 = n = q$, and this is negative for $n \leq 50.$
4. Efficiency of Various Designs

In Chapter 1 upper bounds were given on the number of points needed for estimating \( \theta \) or \( \theta^{(1)} \). In this chapter we give examples in which the bounds are attained. In each model we also show that there are always designs supported on a small number of points (\( k \) or fewer) with D-efficiency bounded below by quantities which will be given. Unfortunately these bounds are sharp only in a few cases. However we do give a sharp lower bound on the G-efficiency of the best \( k \)-point design for estimating \( \theta'f(x) \). If \( \xi \) is a design for estimating \( \theta \), we obtain bounds in both directions relating the D-efficiency and G-efficiency of \( \xi \), including the result that any \( \xi \) has D-efficiency no less than its G-efficiency. These last relations are slight improvements over known results; however we also obtain corresponding relations if \( \xi \) is a design for estimating \( \theta^{(1)} \). Finally there is a discussion of the D-efficiency and G-efficiency of a single design used in several models, and an application of this to product designs considered by Hoel.

We begin by stating that for \( k \geq 1 \) there is an example in which \( k(k+1)/2 \) points are needed for an optimal design for \( \theta \), and for \( k > s \geq 1 \) there is an example in which \( s(s+1)/2 + rs \) points are needed for an optimal design for \( \theta^{(1)} \). However the actual verification of the second example is so tedious that we defer both examples to the very end of the chapter, and consider immediately the efficiency of designs with fewer points.

When estimating \( \theta \), we define the D-efficiency of a design \( \xi \) as

\[
\left[ \frac{\det M(\xi) / \max_{\xi'} \det M(\xi')} \right]^{1/k}.
\]

This definition is justified as follows, to within the usual approximations of the approximate theory. If we take \( m \) observations using \( \xi \) and \( n \) observations using an optimal design \( \xi^* \), the best linear unbiased estimators thus obtained have covariance matrices

\[
\sigma_m^{-1} M^{-1}(\xi) \quad \text{and} \quad \sigma_n^{-1} M^{-1}(\xi^*)
\]

respectively. These matrices will have equal determinants if and only if the ratio of sample sizes \( n/m = \left[ \det M(\xi)/\det M(\xi^*) \right]^{1/k} \).
Likewise we define the G-efficiency of $\xi$ to be $k/\max_x d(x, \xi)$. And when estimating $\theta^{(1)}$ we define the D-efficiency of $\xi$ to be $[\det M^*(\xi)/\max_\xi, \det M^*(\xi')]^{1/s}$. These definitions are also justified as ratios of sample sizes.

The first theorem gives a lower bound on the D-efficiency of certain designs supported on $k$ or fewer points. In the case $s = k$ this result has been obtained by M. J. Box [1966] and Meeter [1967]. The author found it independently later. However the case $s < k$ given here is not considered by Box or Meeter.

**THEOREM 4.1.** For $s \leq k$, let $\eta$ be a D-optimal design for $\theta^{(1)}$, supported on $n$ points, where $n \leq n_0 = s(s+1)/2 + rs$. Let $M(\eta)$ have rank $m, s \leq m \leq k$. There is a design $\zeta$ uniform on $m$ points of the support of $\eta$, satisfying

$$\frac{\det M^*(\zeta)}{\det M^*(\eta)} \geq \frac{1}{\binom{n}{s}} \frac{n^s}{m^s} \geq \frac{1}{\binom{n_0}{m_0}} \frac{n_0^s}{m_0^s}.$$

If $s > 1$ this in turn is bounded below by $n_0^{k-1}[(\binom{n_0}{k})^{k^s}]$. If $s = 1$ it is bounded below by $n_0/[(\binom{n_0}{m_0})m_0^2]$, where $m_0$ is the smallest positive integer satisfying $m_0^2 \geq (k - m_0)(m_0 + 1)$.

We shall comment extensively on the sharpness of these bounds, and on the asymptotic values of the bounds as the various parameters get large. One such result is given here as a corollary.

**COROLLARY.** When $s > 1$, the best design for $\theta^{(1)}$ on $k$ or fewer points had D-efficiency at least $n_0^{k-1}[(\binom{n_0}{k})^{k^s}]^{-1/s}$. If $r$ is fixed and $s \to \infty$, this bound has limit $e^{-1}$.

**PROOF OF THEOREM.** First we remark that if we operate on any $M(\xi)$ which has nonsingular $M_3(\xi)$ by a matrix

$$L = \left| \begin{array}{ccc}
I & -M_2M_3^{-1} \\
M_2 & I \\
0 & I
\end{array} \right|$$
we get immediately $\det M(\xi) = \det M^*(\xi) \det M_3(\xi)$.

The problem will be easier if we reformulate it in terms of matrices of full rank. Because $m = \text{rank } M(\eta)$, the $r \times n$ matrix with columns $f^{(2)}(x)$, $x$ in the support of $\eta$, has rank $t$, with $t \geq m - s$. We can multiply $f$ by a matrix of the form (3.9), in fact with $L_1 = I$ and $L_2 = 0$, to get a vector $Lf$ with $(Lf)_i(x) = 0$ for $i > t + s$ and $x$ in the support of $\eta$. Let $g$ be the $(s+t)$-dimensional column vector with components $(Lf)_1, \ldots, (Lf)_{s+t}$, and let $g^{(1)}$ be the vector consisting of the first $s$ components, $g^{(2)}$ the remaining $t$ components. Note $[LM(\xi)L']^* = L_1 M^*(\xi) L_1 = M^*(\xi)$ for any $\xi$.

Now for $\xi$ supported on a subset of the support of $\eta$, $[LM(\xi)L']^* = \int gg'd\eta$. In particular if $\xi = \eta$, then by the remark at the beginning of the proof

$$\det \int gg'd\eta = \det M^*(\eta) \det \int g^{(2)}g^{(2)'}d\eta.$$

The two matrices on the right are of full rank, while that on the left is of full rank only if $s + t = m$. Therefore $s + t = m$.

Throughout the rest of the proof we will use the following notation. Let the support of $\eta$ be written $\{x_1, \ldots, x_n\}$. If $M$ is an $m$-element subset of $\{1, \ldots, n\}$, then $\eta_M$ and $\xi_M$ will denote the $m$-fold products $\prod \eta(x_i)$ and $\prod \xi(x_i)$ for $i \in M$, and $\|g(x_M)\|$ will denote the $m \times m$ matrix with entries $g_i(x_j)$, for $g_i$ a component of $g$ and $j \in M$. If $S$ and $T$ are subsets of $\{1, \ldots, n\}$ with $s$ and $t$ elements respectively, similar definitions will hold for $\eta_S$, $\xi_T$, $\eta_T$, $\xi_S$, $\|g^{(1)}(x_S)\|$, and $\|g^{(2)}(x_T)\|$.

As a final preliminary to the proof we cite the (Binet-Cauchy) Theorem of Corresponding Minors (see Householder [1964, p. 14]): If $X$ and $Y$ are $p \times n$ matrices, $p \leq n$, then $\det(XY')$ equals the sum of the $\binom{n}{p}$ products of pairs of $p$-th order determinants formed by selecting $p$ columns from $X$ and the same $p$ columns from $Y$.

Consider now the case $m = s$. In this case $g$ and $g^{(1)}$ coincide, and $M^*(\eta) = \int gg'd\eta$. By the Theorem of Corresponding Minors, using the notation
defined two paragraphs above, we have

$$\det M^*(\eta) = \Sigma_M \eta_M \det^2 \left\| g(x_M) \right\|.$$  

Let $M_0$ be that $M$ which maximizes $\det^2 \left\| g(x_M) \right\|$, and let $\xi$ be uniform on $\{ x_i | i \in M_0 \}$. Then

$$\det M^*(\xi) = m^{-m} \max_M \det^2 \left\| g(x_M) \right\|.$$  

Therefore

$$\frac{\det M^*(\xi)}{\det M^*(\eta)} = m^{-m} \max_M \det^2 \left\| g(x_M) \right\| / \Sigma_M \eta_M \det^2 \left\| g(x_M) \right\|$$

$$\geq m^{-m}/\Sigma_M \eta_M.$$  

It is not hard to show (see Keilson [1966]) that if $\phi(x_1, \ldots, x_n)$ is a symmetric multilinear function on the $(n-1)$-simplex, then $\phi$ attains its extrema among the barycenters. So, putting $\phi(\eta_1, \ldots, \eta_n) = \Sigma_M \eta_M$, we get

$$\frac{\det M^*(\xi)}{\det M^*(\eta)} \geq m^{-m}/\max_j (j^2)^{-m}, \text{ where } m \leq j \leq n.$$  

Now we have the ratio

$$\frac{(j^2)^{-m}}{(j^2)^{-1}(j-1)^{-m}} = \frac{(i - 1/j)^{m}/(i - m/j)} \geq 1,$$

the last inequality being well known. Therefore in (4.2) the maximum occurs when $j$ is as large as possible. This proves that when $m = s$, we have

$$\det M^*(\xi)/\det M^*(\eta) \geq \frac{n^m}{[(\frac{n}{m})^m]} \geq \frac{n^m}{[(\frac{n}{m})^m]}.$$  

Let us now consider the case $m = s + t, t > 0$. In this case we have

$$\det M^*(\eta) = \det \int g^2 d\eta / \det \int g^{(2)} g^{(2)}' d\eta$$

$$= \Sigma_M \eta_M \det^2 \left\| g(x_M) \right\| / \Sigma_M \eta_T \det^2 \left\| g^{(2)}(x_T) \right\|.$$  

In the numerator we will delete any zero terms. Then
\[ (4.3) \quad \det M^*(\eta) \leq \sum_M [\eta_M \det^2 \| g(x_M) \| / \sum_{T \subseteq M} \eta_T \det^2 \| g^{(2)}(x_T) \| ], \]

where from now on we index only over those $M$ for which the numerator of the summand is nonzero. Because each numerator is nonzero, consideration of the expansion of $\det \| g(x_M) \|$ shows that each denominator is nonzero. The right hand expression in (4.3) equals

\[ (4.4) \quad \sum_M \left[ \det^2 \| g(x_M) \| / \sum_{T \subseteq M} (\eta_S)^{-1} \det^2 \| g^{(2)}(x_T) \| \right] \]

\[ \leq \sum_M \left[ \det^2 \| g(x_M) \| / \min_{S \subseteq M} (\eta_S)^{-1} \sum_{T \subseteq M} \det^2 \| g^{(2)}(x_T) \| \right] \]

\[ = \sum_M \max_{S \subseteq M} \eta_S \left[ \det^2 \| g(x_M) \| / \sum_{T \subseteq M} \det^2 \| g^{(2)}(x_T) \| \right]. \]

Let $M_0$ be that $M$ which maximizes the expression in square brackets in the last line of (4.4). Let $\zeta$ be uniform on $\{ x_i | i \in M_0 \}$. Then

\[ \det M^*(\zeta) = \sum_M \zeta_M \det^2 \| g(x_M) \| / \sum_M \zeta_T \det^2 \| g^{(2)}(x_T) \| \]

\[ = m^{-S} \max_M \left[ \det^2 \| g(x_M) \| / \sum_{T \subseteq M} \det^2 \| g^{(2)}(x_T) \| \right]. \]

Therefore

\[ (4.5) \quad \det M^*(\zeta) / \det M^*(\eta) \geq m^{-S} / \sum_M \max_{S \subseteq M} \eta_S. \]

We have been summing over $M$ as restricted just below (4.3). We now again sum over all $m$-element sets $M$; this can only decrease the right hand side of (4.5). Since for any $M$

\[ (4.6) \quad \max_{S \subseteq M} \eta_S \leq \sum_{S \subseteq M} \eta_S, \]

we have

\[ (4.7) \quad \det M^*(\zeta) / \det M^*(\eta) \geq m^{-S} / \sum_M \sum_{S \subseteq M} \eta_S. \]
The denominator on the right is a constant times \( \sum S \), which was shown in the proof for the case \( m = s \) to be maximized when \( \eta \) is uniform on as large a set as possible. Therefore

\[
\det M^*(\zeta)/\det M^*(\eta) \geq \frac{n^s}{[(n/m)^m]m^s} \geq \frac{n^s}{[(n_0/m)^m]m^s}.
\]

(The right inequality is easy.) We now only need to prove that for \( s \leq m \leq k \),

\[
B(m) = \frac{n^s}{[(n/m)^m]m^s}
\]

is bounded below by the quantities given in the statement of the theorem.

Consider the ratio

\[
\frac{B(m)}{B(m-1)} = \frac{(m-s)(m-1)^s}{(n_0-m+1)m^s}
\]

for \( s < m \leq k \). This is clearly increasing in \( m \). And we have

\[
\frac{B(k)}{B(k-1)} = \frac{r(k-1)^s}{k^s}
\]

which is less than 1 if \( s \geq 2 \). Thus \( B(m)/B(m-1) \) is less than 1 if \( s \geq 2 \), and \( B(m) \) is minimized by \( m = k \).

If \( s = 1 \), \( B(m+1)/B(m) = m^2/(k-m)(m+1) \). This is increasing in \( m \) for \( m > 0 \), and first becomes greater than or equal to 1 when \( m = m_0 \), the smallest positive integer satisfying \( m_0^2 \geq (k-m_0)(m_0+1) \). This proves the theorem.

**Proof of Corollary.** We must show that \( [(B(s+r)]^{1/s} \to -1 \) if \( r \geq 0 \) is fixed and \( s \to \infty \). By Stirling's approximation of the factorial

\[
[B(s+r)]^{1/s} = \frac{n^{0-k}r!s!}{k} \left[ \frac{(2\pi)^{1/2}e^{-n_0}n_0^{k+1/2}e^{-n_0+k+1/2}e^{-s}}{n_0^{n_0+k+1/2}e^{-n_0+k+s+1/2}e^{-s}} \right]^{1/s}
\]

\[
\sim \frac{n_0}{k} \frac{(2\pi)^{1/2}e^{-r/2}r!s^{1/2}e^{-s}}{n_0^{n_0+k+1/2}e^{-n_0+k+s+1/2}e^{-s}}
\]

\[
= n_0^{k-1}(2\pi)^{1/2}e^{-r/2}r!s^{1/2}e^{-s}(1-k/n_0)^{n_0-k+1/2}e^{-k/s}
\]

\[
\sim n_0^{r/s}(k^{-1}s)^{1/2}(1-k/n_0)^{(n_0-k+1/2)/s}[(1-k/n_0)^{n_0/k}]^{k/s}.
\]
All but the last factor approach 1. The last factor approaches $e^{-1}$. This proves the corollary.

In particular if $s = k$, the corollary says that the lower bound on efficiency approaches $e^{-1}$ as $k \to \infty$.

We now show that if $s$ is fixed and $r \to \infty$ or if the ratio $r/s$ is fixed and $r \to \infty$, then the lower bound on efficiency approaches 0. If $s$ is fixed $> 1$ and $r \to \infty$, then it is immediate that $[B(s + r)]^{1/s} \to 0$. For

$$
[B(s+r)]^{1/s} = \frac{n_0}{k} \left[ \binom{n_0}{s} \frac{k}{r} \right]^{-1/s} = \frac{n_0}{k} \left[ \frac{(n_0)!}{(k)! (r)!} \right]^{-1/r} \left[ \binom{n_0}{k} \binom{k}{r} \right]^{1/r - 1/s}.
$$

By the corollary just proved (interchanging the roles of $s$ and $r$), the product of the first two terms approaches $e^{-1}$, while the last term has limit 0.

If $s = 1$ the lower bound on efficiency is $n_0/[(m_0)_0^2]$, where $n_0 = k$ and $k/m_0 \to 2$ as $k \to \infty$. Thus as $r \to \infty$ the bound is asymptotic to $4/[(m_0)_0^k]$, which approaches 0.

Finally we must show that if $r = cs$ for fixed $c$ and if $r \to \infty$, then $[B(s + r)]^{1/s} \to 0$. We have

$$
[B(s+r)]^{1/s} = \frac{n_0}{k} \left[ \frac{(n_0-k)! r! s!}{n_0!} \right]^{1/s}
$$

$$
\sim \frac{n_0}{k} \left[ \frac{2\pi(n_0-k)^{n_0-k+1/2}r^{1/2}s^{1/2}}{n_0^{n_0+1/2}} \right]^{1/s}
$$

$$
= (n_0/k)(4\pi(n_0-k)rs/n_0)^{1/2}(1-k/n_0)^{(n_0-k)/s}(r/n_0)^{r/s(s/n_0)}
$$

$$
\sim (s/k)e^{-k/s}(1-k/n_0)^{-k/s}(r/n_0)^{r/s}
$$

$$
= (c+1)^{-1}e^{-(c+1)(1-k/n_0)}(r/n_0)^{c}
$$

The first three terms are bounded, while $(r/n_0)^c \to 0$. Thus $[B(s+r)]^{1/s} \to 0$.

Consideration of the sharpness of the bounds of the theorem is now in order. If $m > s$ strict inequality holds in (4.6), because $\eta(x_i) > 0$, $i = 1, \ldots, n$. 
Therefore the bound \( \frac{n^s}{s^s} \) is not attainable, and hence the (smaller) bounds which do not depend on \( m \) and \( n \) are not attainable.

If \( m = s \), the bounds are attainable only if equality holds in (4.1). Equality holds there if and only if \( \det^2 \|g(x_M)\| \) is the same for all \( m \). But this can happen only if \( m = n - 1 \) (or in the trivial cases \( m = 0, 1 \) or \( n \)). To see this, renumber if necessary so that \( \|g(x_M)\| \) is nonsingular, where \( M_1 = \{1, \ldots, m\} \). Operations on the rows of the \( m \times n \) matrix with entries \( g_i(x_j) \), (which do not change the ratio \( \det^2 \|g(x_M)\| / \det^2 \|g(x_{M'})\| \) for any \( M \) or \( M' \)), transform \( \|g(x_M)\| \) into the \( m \times m \) identity matrix, with determinant 1. If \( \det^2 \|g(x_M)\| \) is now to equal 1 for every \( M \) involving \( m \) of the first \( m + 1 \) columns, \( g(x_{m+1}) \) must have every entry equal to \( 1 \). By the same reasoning so must \( g(x_{m+2}) \). But then any determinant involving columns \( m + 1, m + 2 \), and all but two of the first \( m \) columns is not equal to \( 1 \).

So when \( m = s \) the bound \( \frac{n^s}{s^s} \) is attainable only if \( m = n - 1 \). Since this bound is strictly decreasing in \( n \), the bound \( \frac{n^s}{s^s} \) is attainable only if \( m = n_0 - 1 \), i.e., \( s = rs + s(s+1)/2 - 1 \). This occurs only if \( s \leq k = 2 \). We consider the two possibilities:

If \( m = s = 1, k = 2 \), we have \( \frac{n^s}{s^s} = 1 \), which is certainly attainable. But since \( n_0 = 2 \), the lower bound \( \frac{n_0^s}{s^s} \) equals 1/2 < 1. Thus the lower bound \( n_0^s/[s^s(1/2)] \) is not attainable.

If \( m = s = k = 2 \), the lower bound \( n_0^s/[k^s(1/2)] \) is attained. As an example let \( X = \{x_1, x_2, x_3\} \), let \( f_i(x_j) = \delta_{ij} \) for \( 1 \leq i,j \leq 2 \), and let \( f_i(x_3) = 1, i = 1, 2 \). It is easy to check that \( \eta \) is uniform on \( X \) and the bound is attained.

We remark in passing that if \( m = s \) the proof has a geometrical interpretation: the expressions \( \det \|g(x_M)\| \) are the contents of projections of an \( m \)-dimensional parallelepiped into suitable \( m \)-frames in \( n \)-space.
In most situations we will not know \( n \), and if \( s < k \) we will generally not know \( m \) either. In these cases we can only use the lower bounds which do not depend on these values. Moreover if \( m \) is unknown we do not know the number of points of support of \( \zeta \), only that the number is between \( s \) and \( k \). We might therefore try to find a theorem similar to 4.1 but involving a design supported on \( s \) points, or on \( t \) points, where \( t \) is the number of points necessary to estimate \( \theta(1) \). Neither of these ideas works, because of the following facts.

1. \( \theta(1) \) may not be estimable using only \( s \) points.
2. The best \( t \)-point design for \( \theta(1) \) may be arbitrarily bad, where \( t \) is as just defined above.

Both statements are easily proved. There are well known examples to demonstrate (1). One such is given at the end of this chapter, Example 4.2 when \( s = 1 \).

To show (2), let \( X = \{ x_i | i = 1, \ldots, k \} \cup \{ y_i | i = 1, \ldots, s \} \). Let \( f_i(x_j) = \delta_{ij} \) for \( i \leq s \) and \( j \leq k \), let \( f_i(x_j) = 1 \) for \( i > s \) and \( j \leq s \), let \( f_i(x_j) = \delta_{ij} \) for \( i > s \) and \( j > s \), and let \( f_i(y_j) = \delta_{ij} a \) for \( i \leq k \) and \( j \leq s \), where \( |a| \) is small. Then \( \theta(1) \) is estimable using \( s \) points, \( y_1, \ldots, y_s \). As was shown at the end of Chapter 1, if \( \theta(1) \) is estimable using an \( s \)-point design then \( f(2) \) is zero on the support of such a design, and the best such design is uniform on those \( s \) points. Therefore any \( s \)-point design which estimates \( \theta(1) \) is supported on \( \{ y_1, \ldots, y_s \} \), and the best such is uniform there. Denoting this design by \( \xi_s \), an optimal design by \( \xi^* \), and the design uniform on \( \{ x_1, \ldots, x_k \} \) by \( \xi_k \), we have

\[
\frac{\det M^*(\xi_s)}{\det M^*(\xi_k)} \leq \frac{\det M^*(\xi^*_s)}{\det M^*(\xi^*_k)} = a^{2s} s^{-s/(k^{-s}(1 + rs)^{-1})}
\]

which can be arbitrarily small.
We now turn to G-efficiency and give a bound for certain k-point designs.

**THEOREM 4.2.** The best (G) k-point design $\xi_G$ and the best (D) k-point design $\xi_D$ satisfy

$$\max_x d(x, \xi_G) \leq \max_x d(x, \xi_D) \leq k^2.$$ 

We could rephrase this to say that the G-efficiency of $\xi_G$ is no less than that of $\xi_D$, which is no less than $1/k$.

**PROOF.** We must show the right hand inequality. Let \{x_1, ..., x_k\} be the support of $\xi_D$. Let $L$ be the linear transformation so that $g = Lf$ satisfies $g_i(x_j) = \delta_{ij}$. Then $|g_i(x)| \leq 1$ for all $i$ and all $x \in X$; for if $|g_i(y)| > 1$ then substitution of $y$ for $x_j$ would yield a k-point design $\xi'$ with $\det M(\xi') > \det M(\xi_D)$. Thus we conclude

$$\max_x d(x, \xi_G) \leq \max_x d(x, \xi_D) = k \max_x \Sigma g_i^2(x) \leq k^2.$$ 

Here is an example in which the bound is attained. Let $X = \{x_1, ..., x_k, y\}$, let $f_i(x_j) = \delta_{ij}$, and let $f_i(y) = 1$ for $i = 1, ..., k$. If $\xi$ is supported on $\{x_1, ..., x_k\}$, $d(x_i, \xi) = 1/\xi_i$ for any $i$, and $d(y, \xi) = \Sigma(1/\xi_i) > d(x_i, \xi)$ for any $i$. So $d(x, \xi)$ is maximized at $y$. This is minimized by making $\xi$ uniform. In this case $d(y, \xi) = k^2$. If on the other hand $\xi$ is supported on $y$ and all the $x_i$'s but one, say $x_t$, let $L$ be the $k \times k$ matrix with entries $a_{ii} = 1$ for all $i$, $a_{it} = -1$ for $i \neq t$, and $a_{ij} = 0$ otherwise. If $g = Lf$ then $g_i(x_j) = \delta_{ij}$ for $i, j \neq t$, $g_i(y) = \delta_{it}$, $g_t(x_t) = 1$ and $g_t(y) = -1$ for $i \neq t$. Then writing $\xi(y) = \xi_t$, we have $d(x, \xi) = \Sigma g_i^2(x)(1/\xi_i)$. Therefore $d(x_i, \xi) = 1/\xi_i$ if $i \neq t$, $d(y, \xi) = 1/\xi_t$, and $d(x_t, \xi) = \Sigma(1/\xi_i)$. As before, minimizing over $\xi$ supported on all points but $x_t$, we have

$$\min_{\xi} \max_x d(x, \xi) = k^2.$$ 

This example shows that it is possible to cut the G-efficiency by a factor of $k$ by deleting even a single point from an optimal design. It would there-
fore seem undesirable in general to simplify a problem by considering only
k-point designs, although this has been done in the past, for example by Box
and Lucas [1959, p. 80].

Since D-efficiency is usually impossible to compute when no optimal design
is known, it is of practical value to have relations between the D-efficiency
and G-efficiency of a design.

We now develop such relations. The first such theorem is a slight improve­
ment over results obtained by Kiefer [1960, p. 389], and the method is
essentially the same.

Write \( \bar{d}(\xi) = \max_X d(x, \xi) \).

**THEOREM 4.3.** Any design \( \xi \) has D-efficiency no less than its G-efficiency.

For any design \( \xi \), if \( \bar{d}(\xi) \leq c \), where \( c > 3/2 \), then

\[
\frac{\det M(\xi) / \max_{\xi}}{\det M(\xi')} \leq \exp\left[ -(\bar{d}(\xi) - k)^2 / 2D \right]
\]

where \( D = \max \{k - 1 + (c - 1)^2, 4(c - 1)^2((k - 1)/(2c - 3)^2 + 1/9), 2(c - 1)(c - k)\} \).

For example we might wish to let \( c = \bar{d}(\xi) \) if this value is known and
> 3/2, or bearing in mind Theorem 4.2 we might want \( c = k^2 \).

**PROOF.** If \( \xi \) does not estimate \( \theta \), the first assertion reduces to \( 0 \geq 0 \).
If \( \xi \) does estimate \( \theta \), let \( \eta \) be an optimal design. Let \( A \) be a nonsingular
matrix such that \( AM(\xi)A' = I \) and \( AM(\eta)A' \) is a diagonal matrix with diagonal
entries \( d_i \geq 0 \). Then \( \bar{d}(\xi) \geq \int d(x, \xi) \, d\eta = \text{tr } M^{-1}(\xi)M(\eta) = \text{tr } A'^{-1}M^{-1}(\xi)A^{-1}AM(\eta)A' = \Sigma d_i \). Therefore by the arithmetic-geometric mean inequality, \( [\det M(\xi)/\det M(\eta)]^{1/k} \)
\( = (\prod d_i)^{-1/k} \geq k/\Sigma d_i \geq k/\bar{d}(\xi) \). This proves the first assertion of the theorem.

To prove the second assertion, let \( \bar{d}(\xi) = k + \varepsilon \leq c \), with \( c > 3/2 \). Then
there is some design \( \xi \) with \( \int d(x, \xi) \, d\xi = k + \varepsilon \). Writing \( q(\alpha) \) for
\( \log \det M([(1-\alpha)\xi + \alpha \xi]) \) we have (see the remarks following Theorem 1.1)
\( q'(0) = \text{tr } M^{-1}(\xi)M(\xi) - k = \int d(x, \xi) \, d\xi - k = \varepsilon \). Let \( A \) now be such that
\( AM(\xi)A' \) is diagonal with diagonal entries \( d_i \), and \( AM(\xi)A' = I \). The last is
possible because, by assumption, $\theta$ is estimable under $\xi$. Then

$$q''(\alpha) = \frac{\partial^2}{\partial \alpha^2} \left[ \log \det AM([1-\alpha]x + \alpha z)A' - \log \det^2 A \right] = \frac{\partial^2}{\partial \alpha^2} \sum \log(1-\alpha+\omega d_i^2) - 0$$

$$= -\frac{\partial}{\partial \alpha} \left[ \sum (d_i - 1)^2/(1-\alpha+\omega d_i) \right] \text{ and } q^{(1)}(\alpha) \leq 0 \text{ for } 0 \leq \alpha < 1. \text{ Therefore on any interval } 0 \leq \alpha \leq \beta < 1, q''(\alpha) \text{ attains its minimum value at } 0 \text{ or } \beta. \text{ Thus}$$

$$(4.8) \quad q''(\alpha) \geq -\max\{\Sigma (d_i - 1)^2, \Sigma (d_i - 1)^2/(1-b+bd_i)^2\} \text{ for } 0 \leq \alpha \leq \beta < 1.$$

Now $\Sigma (d_i - 1)^2$ is convex in the $d_i$ on the set $B = \{(d_1, ..., d_k) | \text{all } d_i \geq 0, \Sigma d_i \leq c\}$, which contains the $(d_1, ..., d_k)$ corresponding to $\xi$. Therefore $\Sigma (d_i - 1)^2$ takes its maximum value if all but one $d_i$ are 0 and the one equals $c$. Thus $\Sigma (d_i - 1)^2 \leq k-1 + (c-1)^2$.

To bound the second term on the right in (4.8) let $h(u) = (u-1)^2/(1-b+bu)^2$. Then $h''(u) = 2(1+2b-2bu)(1-b+bu)^{-4}$, which is nonnegative if $0 \leq u \leq 1 + 1/2b$. Since $d_i \leq c$ we conclude that if $b = 1/2(c-1)$, (which is less than 1, as required, because $c > 3/2$), then the second expression on the right side of (4.8) is convex in the $d_i$ on the set $B$. Therefore the expression attains its maximum if one $d_i$ equals $c$ and all the rest are 0. Thus

$$\Sigma (d_i - 1)^2/(1-b+bd_i)^2 \leq (k-1)/(1-b)^2 + (c-1)^2/(1+b(c-1))^2$$

$$= 4(k-1)(c-1)^2/(2(c-3)^2) + 4(c-1)^2/9$$

where we have replaced $b$ by its value $1/2(c-1)$.

Therefore for $0 \leq \alpha \leq 1/2(c-1)$,

$$q''(\alpha) \geq \max\{k-1 + (c-1)^2, 4(k-1)(c-1)^2/(2(c-3)^2) + 4(c-1)^2/9\}$$

$$= -K$$

defining $K$. Therefore for $0 \leq \alpha \leq 1/2(c-1)$,

$$(4.9) \quad q(\alpha) \geq q(0) + c\alpha - K\alpha^2/2.$$
If \((c-k)/K \leq 1/2(c-1)\) then \(\varepsilon/K \leq 1/2(c-1)\), and the expression on the right of (4.9) is minimized at \(\alpha_0 = \varepsilon/K\). Let \(\xi' = [1-\alpha_0]\xi + \alpha_0\xi\). Then
\[
\log[\det M(\xi')/\det M(\xi)] = q(\alpha_0) - q(0) \geq \varepsilon^2/2K.
\]

On the other hand if \((c-k)/K > 1/2(c-1)\) then
\[
q(\alpha) \geq q(0) + \alpha - (c-1)(c-k)\alpha^2,
\]
and the right hand side is minimized at \(\alpha_1 = \varepsilon/2(c-1)(c-k)\). In this case let \(\xi' = [1-\alpha_1]\xi + \alpha_1\xi\). Then \(\log[\det M(\xi')/\det M(\xi)] \geq \varepsilon^2/4(c-1)(c-k)\). This completes the proof of the theorem.

The bound given by the first statement of the theorem is seen to be sharp if we consider the following example. Let \(X = \{x_1, ..., x_k\} \cup \{y_1, ..., y_k\}\), let \(f_i(x_j) = \delta_{ij}\), and let \(f_i(y_j) = a\delta_{ij}\), where \(|a| > 1\). Then for any design \(\xi\), \(M(\xi)\) is diagonal with diagonal entries \(\xi(x_i) + a^2\xi(y_i)\). Therefore the optimal design \(\eta\) is uniform on \(\{y_1, ..., y_k\}\). Let \(\xi_1\) be uniform on \(\{x_1, ..., x_k\}\). Then \(\det M(\xi_1) = k^{-k} \) and \(\det M(\eta) = a^{2k}k^{-k}\). Therefore the D-efficiency of \(\xi_1\) is \(a^{-2}\). But \(d(\xi_1, x)\) is maximized if \(x\) is some \(y_i\). So \(\overline{d}(\xi_1) = ka^2\), and the G-efficiency of \(\xi_1\) is also \(a^{-2}\).

The bound given by the second assertion of the theorem is sharp only in the trivial case when \(\overline{d}(\xi) = k\). For if \(\overline{d}(\xi) > k\) then \(\varepsilon > 0\) and hence \(\alpha_0\) and \(\alpha_1\) are nonzero.(by definition). If the bound of the theorem is to be sharp, (4.9) must be equality at \(\alpha_0\) (resp. \(\alpha_1\)). But this can happen only if \(q''(\alpha) = -K\) for all \(\alpha\) between \(0\) and \(\alpha_0\) (resp. \(\alpha_1\)), because \(-K\) is a lower bound on \(q''(\alpha)\). But then \(q''(\alpha) = -\Sigma (d_i-1)^2/(1-\alpha+\alpha d_i)^2\) is constant on the interval, and hence
\[
q^{(h)}(\alpha) = -6 \Sigma (d_i-1)^4/(1-\alpha+\alpha d_i)^4 = 0.
\]
This is possible only if \(d_i = 1\) for every \(i\), in which case \(q''(\alpha) = 0 \uparrow -K\). Therefore (4.9) cannot be equality, and the bound in question cannot be sharp when \(\overline{d}(\xi) > k\).
We remark that we can use Theorems 4.2 and 4.3 together to get a lower bound on the D-efficiency of the best k-point design for estimating \( \theta \). This bound is \( 1/k \), which is not as good as the bound of Theorem 4.1.

As would be hoped there is an analogue of this theorem when estimating at least when \( M(\xi) \) is nonsingular. Write \( \overline{d}_{s|k}(\xi) = \max_x d_{s|k}(x, \xi) \).

**THEOREM 4.4.** If \( M(\xi) \) is nonsingular then

\[
[\det M^*(\xi)/\max_\xi, \det M^*(\xi')]^{1/s} \geq s/\overline{d}_{s|k}(\xi).
\]

If \( \overline{d}(\xi) \leq c \), where \( c > 3/2 \), then

\[
\det M^*(\xi)/\max_\xi, \det M^*(\xi') \leq \exp\left[-(\overline{d}_{s|k}(\xi)-s)^2/2D\right]
\]

where \( D = \max \{ k-1 + (c-1)^2, 4(c-1)^2((k-1)/(2c-3)^2 + 1/9), 2(c-1)(c-s) \} \).

We remark that if \( k > 1 \) then \( c \geq k > 3/2 \) automatically.

**proof.** Let \( L \) be of the form (3.9) and such that \( LM(\xi)L' = I \). If we let the vector of regression functions be \( Lf \) instead of \( f \), then for any \( \xi' \), \( d_{s|k}(x, \xi') \) and \( d(x, \xi') \) are unchanged and \( \det M^*(\xi') \) is only multiplied by a constant, \( \det^2L_L \). Thus we may simply consider \( M(\xi) = I \) throughout the proof, and will do so.

To prove the first statement of the theorem, let \( \eta \) be D-optimal for \( \theta^{(1)} \). Then, using the fact that \( M(\xi) = I \),

\[
\overline{d}_{s|k}(\xi) \geq \int d_{s|k}(x, \xi) d\eta
= \int f(x)f'(x) d\eta - \int f^{(2)}(x) f^{(2)'}(x) d\eta
= \text{tr } M(\eta) - \text{tr } M_3(\eta)
= \text{tr } M_1(\eta)
\geq \text{tr } M^*(\eta).
\]
The last inequality is true because $M^*(\eta) = M_1(\eta) - M_2(\eta)M_3^{-1}(\eta)M_2'(\eta)$, (using a limiting value if $M(\eta)$ is singular). Let $M^*(\eta)$ have eigenvalues $\lambda_1, \ldots, \lambda_s$, all nonnegative but not necessarily distinct. Then $[\det M^*(\eta)]^{1/s} = (\prod \lambda_i)^{1/s} \leq \Sigma \lambda_i/s = \text{tr} M^*(\eta)/s$. Thus

$$[\det M^*(\xi)/\det M^*(\eta)]^{1/s} = [\det M^*(\eta)]^{-1/s} \geq s/\text{tr} M^*(\eta) \geq s/\overline{d}_s \kappa(\xi).$$

To prove the second assertion of the theorem, let $\overline{d}_s \kappa(\xi) = s + \varepsilon$. Then there is a design $\zeta$ with $\int d_s \kappa(x, \xi) d\zeta = s + \varepsilon$. As in the above paragraph therefore $s + \varepsilon = \int d_s \kappa(x, \xi) d\zeta = \text{tr} M_1(\xi)$.

Define $q(\alpha) = \log \det M^*([1-\alpha] \zeta + \alpha \zeta) = \log \det M([1-\alpha] \zeta + \alpha \zeta) - \log \det M_3([1-\alpha] \zeta + \alpha \zeta)$. Let $P$ be a $k \times k$ matrix so that $PP' = I$ and $PM(\zeta)P'$ is a diagonal matrix with diagonal elements $d_i$. Let $Q$ be an $r \times r$ matrix so that $QQ' = I$ and $QM_3(\zeta)Q'$ is a diagonal matrix with diagonal elements $e_i$. (Note, $\text{tr} M(\zeta) = \text{tr} M(\zeta)P'P = \text{tr} PM(\zeta)P' = \Sigma d_i$, and similarly $\text{tr} M_3(\zeta) = \Sigma e_i$.) Then, again recalling that $M(\zeta) = I$,

$$q(\alpha) = \sum_{i=1}^{k} \log(1-\alpha + \alpha d_i) + \log \det PP' - \sum_{i=1}^{r} \log(1-\alpha + \alpha e_i) - \log \det QQ'$$

and

$$q(0) = \log \det PP' - \log \det QQ'.$$

The first derivative is

$$q'(\alpha) = \sum_{i=1}^{k} (-1+\alpha d_i)/(1-\alpha + \alpha d_i) - \sum_{i=1}^{r} (-1+\alpha e_i)/(1-\alpha + \alpha e_i)$$

and

$$q'(0) = \sum_{i=1}^{k} (-(1+d_i) - \sum_{i=1}^{r} (-(1+e_i)

= \text{tr} M(\zeta) - k - \text{tr} M_3(\zeta) + r

= \text{tr} M_1(\zeta) - s

= \varepsilon.$$
Differentiating again gives

\[ q''(\alpha) = - \sum_{i=1}^{k} (1-d_i)^2/(1-\alpha+\omega d_i)^2 + \sum_{j=1}^{r} (1-e_j)^2/(1-\alpha+\omega e_j)^2 \]

\[ \geq - \sum_{i=1}^{k} (1-d_i)^2/(1-\alpha+\omega d_i)^2. \]

We have \( \sum d_i = \text{tr} M(\xi) = \int d(x, \xi) d\eta \leq c. \) Therefore by the proof of Theorem 4.4, \( q''(\alpha) \geq - K \) for \( 0 \leq \alpha \leq 1/2(c-1), \) where

\[ K = \max \{k-1 + (c-1)^2, 4(c-1)^2((k-1)/(2c-3)^2 + 1/9)\} \]

Therefore for \( 0 \leq \alpha \leq 1/2(c-1) \)

\[(4.10) \quad q(\alpha) \geq q(0) + \epsilon \alpha - K\alpha^2/2. \]

Now \( \epsilon = \bar{d}_s|k(\xi) - s \leq \bar{d}(\xi) - s \leq c - s. \) Thus if \( (c-s)/K \leq 1/2(c-1) \)

then certainly \( \epsilon/K \leq 1/2(c-1) \) and the expression on the right of (4.10) is minimized at \( \alpha_0 = \epsilon/K. \) As in the proof of Theorem 4.4, let \( \xi' = (1-\alpha_0)\xi + \alpha_0\xi. \)

Then \( \log[\det M(\xi')/\det M(\xi)] = q(\alpha_0) - q(0) \geq \epsilon^2/2K. \)

On the other hand if \( (c-s)/K > 1/2(c-1), \) then

\[ q(\alpha) \geq q(0) + \epsilon \alpha - (c-1)(c-s)\alpha^2, \]

and the right side is minimized at \( \alpha_1 = \epsilon/2(c-1)(c-s). \) In this case let \( \xi' = (1-\alpha_1)\xi + \alpha_1\xi. \) Then \( \log[\det M(\xi')/\det M(\xi)] \geq \epsilon^2/4(c-1)(c-s). \) This completes the proof.

To see that the first statement of the theorem gives a sharp bound,

consider the simplest example: Let \( X = \{x_1, \ldots, x_k\}, f_i(x_j) = \delta_{ij}. \) The optimal design \( \eta \) for \( \theta^{(1)} \) is uniform on \( \{x_1, \ldots, x_s\}. \) Let \( \xi_1 \) be uniform on \( X. \) Then \( M^*(\eta) = s^{-1}I, \quad M^*(\xi_1) = k^{-1}I, \) and \( d_s|k(x, \xi_1) = k \) for all \( x. \)

Therefore

\[ [\det M^*(\xi_1)/\det M^*(\eta)]^{1/s} = s/k = s/\bar{d}_s|k(\xi). \]
To see that the second statement gives a sharp bound only when \( \varepsilon = 0 \),
we note as before that if \( \varepsilon > 0 \), the bound can be sharp only if (4.10) is
equality at \( \alpha_0 \) (resp. \( \alpha_1 \)). This in turn can happen only if
\[
\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(1-\varepsilon + \sigma_i)^2} = K
\]
for all \( x \) in the interval from 0 to \( \alpha_0 \) (resp. \( \alpha_1 \)). But this can never occur, as was shown in the discussion following
Theorem 4.3.

Theorem 4.4 gives some justification to examination of \( \overline{d_s k}(x, \xi) \)
even when \( \overline{d_s k}(\xi) > s \). Although \( \overline{d_s k}(\xi) \) does not measure the efficiency
of \( \xi \) with respect to any intuitively meaningful optimality criterion, it does
give an indication of the D-efficiency of \( \xi \).

There is an easy extension of the first part of Theorem 4.4 to the case
in which \( M(\xi) \) is singular. We state this as a corollary.

COROLLARY. Given any designs \( \xi \) and \( \xi_1 \), and \( 0 < \varepsilon < 1 \), write
\[
\xi_\varepsilon = (1-\varepsilon)\xi + \varepsilon\xi_1.
\]
Suppose \( M(\xi_\varepsilon) \) is nonsingular, \( 0 < \varepsilon < 1 \), and
\[
\lim_{\varepsilon \to 0} \overline{d_s k}(\xi_\varepsilon)
\]
exists. Then
\[
[\det M^*(\xi)/\max_{\xi}, \det M^*(\xi')]^{1/s} \geq s/\lim_{\varepsilon \to 0} \overline{d_s k}(\xi_\varepsilon).
\]

PROOF. For any \( \varepsilon \) between 0 and 1
\[
[\det M^*(\xi_\varepsilon)/\max_{\xi}, \det M^*(\xi')]^{1/s} \geq s/\overline{d_s k}(\xi_\varepsilon).
\]
Taking the limit in \( \varepsilon \) on each side gives the desired result.

We remark that for a given \( \xi \) the value of \( \lim_{\varepsilon \to 0} \overline{d_s k}(\xi_\varepsilon) \) will depend
in general on what design \( \xi_1 \) is used.

The second assertion of Theorem 4.4 cannot be extended in this manner to
the case when \( M(\xi) \) is singular, as Example 3.1 shows. There we had \( \xi^* \)
optimal. If \( \xi_1 \) was suitably chosen and \( \xi_\varepsilon \) defined as \( (1-\varepsilon)\xi^* + \varepsilon\xi_1 \), we
had \( \lim_{\varepsilon \to 0} \overline{d_s k}(x, \xi_\varepsilon) = c_2 \), for arbitrarily large \( c \). On the other hand
\[
det M^*(\xi_\varepsilon)/det M(\xi^*) = 1-\varepsilon.
\]
Let us now leave the problem of relating D- and G-efficiency. A similar problem is to obtain bounds on the efficiency of a single design in various related models.

THEOREM 4.5. Suppose $\xi_0$ is optimal for $\theta$. Then the following are true.

(4.11) $\left[\frac{\det M_1(\xi_0)}{\max_{\xi} \det M_1(\xi)}\right]^{1/s} \geq s/k$.

(4.12) $\left[\frac{\det M^*(\xi_0)}{\max_{\xi} \det M^*(\xi)}\right]^{1/s} \geq s/k$.

(4.13) $d_{s|s}(x, \xi_0) \leq k$.

(4.14) $d_{s|k}(x, \xi_0) \leq k$.

Statements (4.11) and (4.12) say that $\xi_0$ has D-efficiency $\geq s/k$ for estimating $\theta^{(1)}$ out of $\theta^{(1)}$ or $\theta^{(1)}$ out of $\theta$. Line (4.13) says that $\xi_0$ has G-efficiency $\geq s/k$ when estimating $f'(x)\theta$. Line (4.14) does not have a natural efficiency interpretation except to the extent given by Theorem 4.5, but it is included here for the sake of completeness.

PROOF. Statements (4.11) and (4.12) can be proved directly, but they follow immediately from (4.13) and (4.14) respectively, by Theorems 4.3 and 4.4. So we need only show (4.13) and (4.14).

We will suppress the $\xi_0$, writing $M_1$ for $M_1(\xi_0)$, etc. Let

$$L = \begin{bmatrix} I & 0 \\ -M_2^{{\prime}M_1^{-1}} & I \end{bmatrix}.$$ 

Then

$$LML' = \begin{bmatrix} M_1 & 0 \\ 0 & M_3 - M_2^{{\prime}M_1^{-1}M_2} \end{bmatrix}.$$ 

Write $g = Lf$, and note that $g^{(1)} = f^{(1)}$. Then
\[ k \geq d_k \left| k \right| (x, \xi_0) = g'(x)(LML')^{-1}g(x) \]
\[ = g^{(1)'}(x) M_1^{-1}g^{(1)}(x) + g^{(2)'}(x)[M_3 - M_1 M_2^{-1} M_1]^{-1}g^{(2)}(x) \]
\[ \geq g^{(1)'}(x) M_1^{-1}g^{(1)}(x) \]
\[ = f^{(1)'}(x) M_1^{-1}f^{(1)}(x) \]
\[ = d_s | s \right| (x, \xi_0), \]

proving (4.13).

To prove (4.14), write
\[ d_s \mid k (x, \xi_0) = f'(x) Mf(x) - f^{(2)'}(x) M_3^{-1}f^{(2)}(x) \]
\[ \leq f'(x) Mf(x) \]
\[ = d_k \mid k (x, \xi_0) \leq k, \]

as was to be shown.

The bounds are sharp, as is shown by the following example in which each bound is attained. (This example was also mentioned following Theorem 4.4.)

Note that in this example the inequalities in the proof are equality at \( x_1, \ldots, x_s \), and the arithmetic-geometric mean inequality, used in the proofs of Theorems 4.3 and 4.4, is equality.

Let \( X = \{ x_1, \ldots, x_k \} \), \( f_i(x_j) = \delta_{ij} \). Then \( \xi_0 \) is uniform on \( X \). For a competing design in (4.11) and (4.12) let \( \xi \) be uniform on \( \{ x_1, \ldots, x_s \} \). Then \( M(\xi_0) = (1/k)I \) and \( M_1(\xi) = (1/s)I, M_2(\xi) = 0, M_3(\xi) = 0 \). It is immediate that all four bounds in the theorem are attained.

We now apply this result to the product situation considered by Hoel [1965]. Suppose the regression functions \( f_{ij}(y, z) \) on \( Y \times Z \) can be factored into \( f_{ij}(y, z) = g_i(y)h_j(z) \). The optimal design for the coefficients of the \( g_i \) alone is some measure \( \eta \) on \( Y \), and for the coefficients of the \( h_j \) a
measure \( \zeta \) on \( Z \). Hoel then shows that an optimal design for the coefficients of the \( f_{ij} \) is simply \( \eta \times \zeta \). This gives a corollary to Theorem 4.5.

**COROLLARY.** Let \( \eta \) and \( \zeta \) be optimal designs for the coefficients of some \( n \)-vectors of regression functions \( g \) and \( h \) on spaces \( Y \) and \( Z \) respectively. If \( f(y, z) \) consists of some \( s \) of the products \( g_j(y)h_j(z) \) then for estimating the coefficients of \( f \) the product design \( \eta \times \zeta \) has D-efficiency \( \geq s/n^2 \) and G-efficiency \( \geq s/n^2 \).

For example if \( Y = Z = [-1, 1] \) and \( g(y) \) and \( h(z) \) each consist of the \( n \) monomials of degree \( \leq n-1 \), then the optimal design in each one dimensional problem is well known. (See Guest [1958] and Hoel [1958].) If \( f(y, z) \) consists of those monomials in \( y \) and \( z \) of degree \( \leq n-1 \), then the product design \( \xi_0 \) has D-efficiency \( \geq (n+1)/2n \) and G-efficiency \( \geq (n+1)/2n \).

In this situation an optimal design is known only for small \( n \). (See Kiefer [1961, sec. 4] and Farrell, Kiefer and Walbran [1965, sec. 3].) Even though \( \xi_0 \) is not optimal it is easy to obtain and use, so a bound such as that given, showing that \( \xi_0 \) is fairly efficient, is nice to have.

In fact in this example of polynomial regression, \( \xi_0 \) seems to be considerably more efficient than the bound indicates. If \( f \) is quadratic regression on the square (i.e., \( n = 3 \)), Kiefer [1961, pp. 314-317] has computed an optimal design \( \xi^* \) for the six regression coefficients. We compare this with the product design \( \xi_0 \). (M here is a \( 6 \times 6 \) matrix.)

\[
\begin{align*}
[\det M(\xi_0)]^{1/6} &= .462 \\
[\det M(\xi^*)]^{1/6} &= .475 \\
\text{D-efficiency of } \xi_0 &= .97 \\
\text{Theoretical lower bound} &= 6/9 = .67
\end{align*}
\]
\[
\max_x d_{\theta} (x, \xi_0) = 7.25
\]
G-efficiency of \( \xi_0 = .83 \)

Theoretical lower bound = 6/9 = .67

Although in this example the efficiencies are greater than the lower bounds of Theorem 4.5, there are product design situations in which the bounds are attained. As an example suppose \( Y = \{y_1, \ldots, y_m\}, Z = \{z_1, \ldots, z_n\}, g(y) \) an \( m \)-vector with \( g_i(y_j) = \delta_{ij} \), and \( h(z) \) an \( n \)-vector with \( h_i(z_j) = \delta_{ij} \).

Let the regression functions \( f_\ell(y, z) \) be some \( s \)-element subset of the \( mn \) functions \( g_i(y)h_j(z) \). Then \( f(y, z) \) is just that given in the example immediately following Theorem 4.5, in which all the lower bounds on efficiency were attained.

We conclude with the two examples mentioned at the beginning of the chapter. First is an example in which \( N = k(k+1)/2 \) points are needed for an optimal design for \( \theta \). This example is similar to the example which is concerned with optimality for \( \theta(1) \), but it is much simpler and easier to understand. The second example does not reduce to the first when we set \( s = k \), and the verification of the properties of the second example uses the fact that \( s < k \).

Thus the examples are similar but distinct.

EXAMPLE 4.1. If \( k = 1 \) the question is trivial. For \( k > 1 \), let \( X = \{x_{ij} \mid i \leq j \leq k\} \), so \( X \) has \( N \) points. Let \( \alpha \) be chosen with \( 2(2k-4)/(2k-3) < \alpha < 2 \). The reason for this choice will become apparent. Let \( f_\ell(x_{jj}) = \alpha^{1/2} \delta_{ij} \), and \( f_\ell(x_{ij}) = \delta_{i\ell} + \delta_{ij} \) for \( \ell \neq j \). We now verify that the only optimal design is supported on all \( N \) points of \( X \).

Writing \( \xi_{ij} \) for \( \xi(x_{ij}) \) we have for any \( \xi \)

\[
M_{ii}(\xi) = \xi_{ii} + \sum_{i < j} \xi_{ij} + \sum_{i < j} \xi_{ij},
\]

\[
M_{ij}(\xi) = M_{ji}(\xi) = \xi_{ij} \quad \text{for } i < j.
\]
Since \( M(\xi) \) is the same for all optimal \( \xi \), it is immediate that the optimal \( \xi \) is unique. It is therefore invariant, by Theorem 1.3, and we need henceforth consider only designs which are invariant in the following sense.

For any permutation \( \pi \) on \( k \) elements define \( g\xi_{ij} = x_{\pi_i \pi_j} \) or \( x_{\pi_j \pi_i} \).

(Just one of these will be in \( X \), as \( \pi_i \leq \pi_j \) or \( \pi_i \geq \pi_j \).) Define \( \bar{\theta}_i = \theta_{\pi_i} \). Then \( (\bar{\theta})'f(g\xi) = \theta'f(x) \), so the problem is invariant under such \( g \).

The invariant designs are those of the form \( \xi_{ii} = a \) and \( \xi_{ij} = b \) for \( i < j \).

Note \( ka + k(k-1)b/2 = 1 \).

For such designs \( M(\xi) = (a\alpha + (k-2)b)I + bU \), where \( U \) will from now on denote a matrix with every entry 1, here a \( k \times k \) matrix. Then

\[
\det M(\xi) = [a\alpha + 2(k-1)b][a\alpha + (k-2)b]^{k-1}.
\]

After substituting for \( b \) by the relation \( ka + k(k-1)b/2 = 1 \), the derivative of this determinant with respect to \( a \) at \( a = 0 \) is positive if \( \alpha > 2(2k-4)/(2k-3) \), and the derivative with respect to \( a \) at \( b = 0 \) is negative if \( \alpha < 2 \). Thus for optimality we must have \( a \geq 0 \) and \( b \leq 0 \), and all \( N \) points are needed for an optimal design.

In the second example \( N = s(s+1)/2 + rs \) points are needed for an optimal design for \( \theta(1) \).

**Example 4.2.** Let \( X = \{x_{ii} \mid i \leq s \} \cup \{x_{ij} \mid i < j \leq s \} \cup \{y_{ij} \mid i \leq s < j \leq k \} \).

Assume \( s < k \). There are \( N = s(s+1)/2 + rs \) points in \( X \), Choose \( \gamma \) so that

\[
0 < \gamma \quad \text{if } s = 1
\]

\[
0 < \gamma < 1/(3r) \quad \text{if } s = 2
\]

\[
0 < \gamma \quad \text{and } \gamma \text{ small enough so that}
\]

\[
(1/(r\gamma) - 1)^2 > 2(s-1)^2/(s-2) \quad \text{if } s > 2.
\]

In particular in the last case \( r\gamma < 1 \). Now let
Consider first the case \( s = 1 \). When \( s = 1 \) it is known that for polynomial regression on \([-1, 1]\), the unique optimal design requires \( N = k \) points. (See Kiefer and Wolfowitz [1959, sec. 3].) However for the sake of unity we now also show that in our example with \( s = 1 \), \( N = k \) points are needed to estimate \( \theta^{(1)} \), hence \( k \) points are needed for optimality.

When \( s = 1 \), \( X \) simplifies to \( \{x_i|i = 1, \ldots, k\} \), with \( f_i(x_j) = 1 \) for all \( j \), \( f_i(x_1) = \gamma \) for all \( i \), and \( f_i(x_j) = \delta_{ij} \) otherwise. Clearly \( \delta(x_1) \) cannot be 0 if \( \theta^{(1)} \) is estimable. If \( \delta(x_1) = 0 \) for some \( x_i \), say \( x_2 \), let \( L \) be the matrix with entries \( a_{ij} = \delta_{ij} \) for \( j \neq 2 \), \( a_{12} = 0 \), \( a_{22} = 1 \), and \( a_{ij} = -1 \) for \( i > 2 \). Then using \( \delta_2 = 0 \) we compute that \( LML' = N \) has entries \( n_{11} = 1, n_{12} = n_{21} = \gamma \delta_1, n_{22} = \gamma^2 \delta_1, n_{1i} = n_{i1} = n_{ii} = \delta_i \) for \( i > 2 \), and \( n_{ij} = 0 \) otherwise. Then \( \det M^*(\delta) = \det N^*(\delta) = 0 \).

Now we consider the case \( s > 1 \). We first show that an optimal design which is invariant must have \( N \) points of support. Under the obvious group of symmetries the only invariant design is of the form \( \delta(x_{1i}) = a, \delta(x_{ij}) = b, \delta(y_{ij}) = c \). For this \( \delta \), again with \( U \) denoting a matrix with all entries 1,

\[
M(\delta) = \begin{bmatrix}
[a + (s-2)b + rc]I + bU & (\gamma a + c)U \\
(\gamma a + c)U & scI + s\gamma^2aU
\end{bmatrix}
\]

where the upper left block is \( s \times s \). Here we have used the assumption that \( s < k \). Then \( M^{-1}(\delta) = 1/(sc)[I - \gamma^2a/(r\gamma^2a + c)U] \) except when \( c = 0 \). If \( c = 0 \), \( M^* \) will be computed by taking limits. Then

\[
M^*(\delta) = [a + (s-2)b + rc]I + [b - r(\gamma a + c)^2/(rs\gamma^2a + sc)]U
\]
and
\[
\det M^*(\xi) = [a + (s-2)b + rc]^{s-1}[a + 2(s-1)b + rc - r(\gamma a + c)^2/(r\gamma^2 a + c)]
\]

where if \( c = 0 \), \( M^* \) is computed as the limit of the last expression.

Observe that \( \det M^* = 0 \) when \( a = b = 0 \) or when \( b = c = 0 \). We now show that unless \( a > 0 \), \( b > 0 \) and \( c > 0 \) there is a \( \xi' \) with
\[
\frac{\partial}{\partial \epsilon} \bigg|_{\epsilon=0} \det M^*([1-\epsilon]\xi + \epsilon \xi') > 0.
\]
Denoting \([1-\epsilon]\xi + \epsilon \xi'\) by \( \xi_\epsilon \), we write
\[
\xi_\epsilon(x_{ij}) = A = (1-\epsilon)a + \epsilon x, \quad \xi_\epsilon(y_{ij}) = B = (1-\epsilon)b + \epsilon y, \quad \text{and} \quad \xi_\epsilon(z_{ij}) = C = (1-\epsilon)c + \epsilon z.
\]

**Case 1.** Suppose \( c = 0 \). We need only consider \( b > 0 \) since otherwise \( \det M^* = 0 \). Also we need only consider \( a + (s-2)b > 0 \) (i.e., \( a > 0 \) when \( s = 2 \)) for the same reason.

Define \( \xi' \) by letting \( x = a \), \( y = 0 \), \( z = b(s-1)/2r \). Then the derivatives with respect to \( \epsilon \) are \( A' = 0 \), \( B' = -b \) and \( C' = z \). The derivative of \( M^*(\xi_\epsilon) \) is positive if and only if
\[
(4.17) \quad 0 < -(s-1)^2(s-3)b^2 + [a + (s-2)b][-2(s-1)b + rz(1/(r\gamma)-1)^2] \]

When \( s = 2 \) this is satisfied if \( rz(1/(r\gamma)-1)^2 > 2(s-1)b \), i.e., when \( (1/(r\gamma)-1)^2 > 4 \). This holds if \( \gamma < 1/3r \). When \( s > 2 \) we use the relation \( sa + s(s-1)b/2 = 1 \) and the fact that \( b > 0 \) to see that \((4.17)\) is equivalent to
\[
[s(s-3)b + 2/s]/[1/s + (s-3)b/2] < (1/(r\gamma)-1)^2/2.
\]
The left side is maximized by \( b = 2/[s(s-1)] \). Thus by \((4.15)\) the derivative is positive.

**Case 2.** Suppose \( a = 0 \). We have just shown we may consider \( c > 0 \). Let \( x = rc \), \( y = b \), \( z = 0 \). Then \( A' = rc \), \( B' = 0 \), \( C' = -c \). Computing the derivative of \( \det M^*(\xi_\epsilon) \) and using \( c > 0 \) we see that \( \xi \) is not optimal if \( (1-r\gamma)^2 > 0 \). So by \((4.15)\) \( \xi \) is not optimal.
Case 3. Suppose \( b = 0 \). Let \( x = z = 0, y = 2/(s^2-s) \). Then \( A' = -a, B' = y, C' = -c \). Routine computation shows that the derivative of \( \det M^*(\xi) \) is positive if and only if

\[
0 < s(1/s - r(ya+c)^2/(ry^2a+c)) + 4r(ya+c)^2/(ry^2a+c).
\]

But \( (1/s - r(ya+c)^2/(ry^2a+c)) = s^{s-1} \det M^*(\xi) \geq 0 \). So this \( \xi \) is not optimal, and any invariant optimal design requires \( N \) points of support.

We now complete the verification by showing that any optimal design is invariant. This is by far the hardest part of the verification.

**Case 1.** Suppose \( \xi \) is optimal and that there is some \( x_{jj} \), say \( x_{11} \), with \( \xi(x_{11}) + \sum \xi(x_{ii})/s \). That is, not all the \( \xi_{ii} \)'s are equal. If \( \pi \) is a permutation on \( \{1, \ldots, k\} \) which leaves the sets \( \{1\}, \{2, \ldots, s\} \) and \( \{s+1, \ldots, k\} \) unchanged, define \( \xi_\pi \) by \( \xi_\pi(x_{ii}) = \xi(x_{\pi_i \pi_i}), \xi_\pi(y_{ij}) = \xi(y_{\pi_i \pi_j}) \) and \( \xi_\pi(x_{ij}) = \xi(x_{\pi_i \pi_j}) \). In the last case if \( \pi_i > \pi_j \) we will write \( x_{\pi_i \pi_j} \) to mean \( x_{\pi_i \pi_j} \).

We have

\[
\int f_\xi(x)f_m(x)d\xi(x) = \sum_{i \leq j \leq s} f_\xi(x_{ij})f_m(x_{ij})\xi_\pi(x_{ij})
\]

\[
+ \sum_{i < j} f_\xi(y_{ij})f_m(y_{ij})\xi_\pi(y_{ij})
\]

\[
= \sum_{\pi} f_\pi(x_{\pi_i \pi_j})f_m(x_{\pi_i \pi_j})\xi(x_{\pi_i \pi_j})
\]

\[
+ \sum_{\pi} f_\pi(y_{\pi_i \pi_j})f_m(y_{\pi_i \pi_j})\xi(y_{\pi_i \pi_j}),
\]

where in the last step we use the fact that \( f_\xi(x_{ij}) \) and \( f_\xi(y_{ij}) \) depend only on whether the respective indices are equal or not and whether they are greater than \( s \) or not. Thus \( \int f_\xi(x)f_m(x)d\xi(x) = \int f_{\xi_\pi}(x)f_m(x)d\xi(x) \) and \( \det M^*(\xi_\pi) = \det M^*(\xi) \). By the concavity of \( \log \det M^*[1-\alpha]\xi_1 + \alpha\xi_2 \) we have

\[
\det M^*((1/n)\sum_\pi \xi_\pi) \geq \det M^*(\xi), \text{ where } n \text{ is the number of permutations } \pi.
\]

We will henceforth denote \( (1/n)\sum_\pi \xi_\pi \) by \( \xi_1 \).
We therefore have a \( \xi_1 \) which is assumed optimal and of the form

\[
\begin{align*}
\xi_1(x_{ij}) &= a \quad i > 1 \\
\xi_1(x_{ij}) &= b \quad i > 1 \\
\xi_1(y_{ij}) &= c \quad i > 1 \\
\xi_1(x_{11}) &= d \\
\xi_1(x_{1j}) &= e \\
\xi_1(y_{1j}) &= f
\end{align*}
\]

with \( d \neq a \). We will prove this impossible by showing \( d = a \). Note that if \( s = 2 \) the only \( x_{ij} \) in \( X \) is \( x_{12} \); in what follows any \( b \)'s which appear are cancelled out if \( s = 2 \).

Then

\[
M_1(\xi_1) = \begin{bmatrix}
d + (s-1)e + rf & eV' \\
eV & [a + e + (s-3)b + rc]I + bU
\end{bmatrix}
\]

where \( I \) and \( U \) are \((s-1) \times (s-1)\), and \( V \) will be used from now on to denote a vector consisting of all 1's, in this case an \((s-1)\)-vector.

\[
M_2(\xi_1) = ((\gamma d + f)V, (\gamma a + c)U),
\]

where \( V \) is \( r \times 1 \) and \( U \) is \( r \times s \), and

\[
M_3(\xi_1) = [f + (s-1)c]I + [\gamma^2 d + (s-1)\gamma^2 a]U,
\]

where \( I \) and \( U \) are \( r \times r \). Then

\[
M^*(\xi_1) = \begin{bmatrix}
d + (s-1)e + rf - rJ^{-1}(\gamma d+f)^2 & (e - rJ^{-1}(\gamma d+f)(\gamma a+c))V' \\
(e - rJ^{-1}(\gamma d+f)(\gamma a+c))V & [a + e + (s-3)b + rc]I + (b - rJ^{-1}(\gamma a+c)^2)U
\end{bmatrix}
\]

where \( J = (s-1)(\gamma^2 + c) + \gamma^2 a + f \). If \( M(\xi_1) \) is singular we compute

\[
[M(\xi_1) + \epsilon I]^*,
\]

and get the above expression for \( M^*(\xi_1) \) by taking the limit as \( \epsilon \to 0 \).

Now \( \xi_1 \) distinguishes the index 1 from the indices 2, \ldots, \( s \). If we interchange the roles of 1 and any \( j, 2 \leq j \leq s \), we obtain a design \( \xi_j \) with \( \det M^*(\xi_1) = \det M^*(\xi_j) \). The average of these designs \( \xi_0 = s^{-1} \Sigma \xi_i \) is optimal and invariant under the group of all permutations which leave \( \{1, \ldots, s\} \) and \( \{1, \ldots, k\} \) fixed. We have

\[
\begin{align*}
\xi_0(x_{ij}) &= A = d/s + a(s-1)/s. \\
\xi_0(x_{ij}) &= B = 2e/s + b(s-2)/s. \\
\xi_0(y_{ij}) &= C = f/s + c(s-1)/s.
\end{align*}
\]
But from (4.16)

\[
M^*(\xi_0) = [A + (s-2)B + rC]I + [B - r(\gamma A + C)^2/(rs\gamma^2A + sC)]U
\]

\[
= s^{-1}[d + (s-1)a + 2(s-2)e + (s-2)b + rf + r(s-1)c]I
\]

\[
+ s^{-1}[2e + (s-2)b - rs^{-1}J^{-1}(\gamma d + \gamma(s-1)a + f + (s-1)c)^2]U.
\]

Because both designs are optimal, \( M^*(\xi_1) = M^*(\xi_0) \). Setting the upper left entries equal we obtain

\[
(4.18) \quad 0 = (a-d) + (s-2)(b-e) + r(c-f) + rs^{-1}J^{-1}[(s+1)(\gamma d+f)^2
\]

\[
- 2(\gamma d+f)(\gamma a+c) - (s-1)(\gamma a+c)^2].
\]

Now setting the lower right entries of the matrices equal we get

\[
(4.19) \quad 0 = (a-d) + (s-2)(b-e) + r(c-f) + rs^{-1}J^{-1}[(\gamma d+f)^2
\]

\[
+ 2(s-1)(\gamma d+f)(\gamma a+c) + (1-2s)(\gamma a+c)^2].
\]

Comparison of (4.18) and (4.19) gives

\[
0 = s[(\gamma d+f) - (\gamma a+c)]^2.
\]

Thus

\[
(4.20) \quad \gamma d + f = \gamma a + c.
\]

We now set the upper right entries of the matrices equal, and get

\[
(4.21) \quad (s-2)e = (s-2)b.
\]

Finally, setting the upper left and lower right entries of \( M^*(\xi_1) \) equal (because this equality holds in \( M^*(\xi_0) \)), we get

\[
d + (s-1)e + rf - rJ^{-1}(\gamma d+f)^2 = a + e + (s-2)b + rc - rJ^{-1}(\gamma a+c)^2
\]

which by (4.20) and (4.21) becomes \( d + rf = a + rc \), which again by (4.20)
becomes \((ry-1)d = (ry-1)a\). But since \(ry \not| 1\), \(d = a\), contrary to the assumption.

**Case 2.** Suppose \(\xi\) is optimal with \(\xi(x_{i2}) = 2[s^2-s]\sum_{i<j} \xi(x_{ij})\).

Of course \(s > 2\) or the case is vacuous. As before we can average over the permutations which leave \([1, 2], [3, \ldots, s]\) and \([s+1, \ldots, k]\) fixed and obtain an optimal \(\xi_{12}\) of the form

\[
\begin{align*}
\xi_{12}(x_{i1}) &= a & i & > 2 \\
\xi_{12}(x_{ij}) &= b & i & > 2 \\
\xi_{12}(y_{ij}) &= c & i & > 2 \\
\xi_{12}(x_{i1}) &= d & i & \leq 2 \\
\xi_{12}(x_{ij}) &= e & i & \leq 2 \\
\xi_{12}(y_{ij}) &= g & i & \leq 2 \\
\xi_{12}(x_{ij}) &= f & j & > 2 \\
\xi_{12}(y_{ij}) &= h & j & > 2
\end{align*}
\]

where \(e \not| b\) or \(e \not| f\). If \(s = 3\) there are no \(x_{ij}\) with \(i > 2\) and the assumption is simply \(e \not| f\). When \(s = 3\) the \(b\)'s below all cancel. By the result of the previous case, \(d = a\). Then

\[
M(\xi_{12}) = \begin{pmatrix}
[a+(s-2)f + rg]I & fU & (\gamma a + g)U \\
+ eU & fU & [a+2f+(s-4)b + rc]I \\
& & (\gamma a + c)U \\
& & [2g+(s-2)c]I \\
& & + s\gamma^2 aU
\end{pmatrix}
\]

where the matrices along the diagonal are \(2 \times 2, (s-2) \times (s-2)\) and \(r \times r\) respectively. It follows that

\[
M^*(\xi_{12}) = \begin{pmatrix}
[a+(s-2)f + rg]I & [f-rJ^{-1}(\gamma a+c)(\gamma a+c)]U \\
+ [e-rJ^{-1}(\gamma a+g)^2]U & [f-rJ^{-1}(\gamma a+g)(\gamma a+c)]U \\
& [a+2f+(s-4)b + rc]I \\
& + [b-rJ^{-1}(\gamma a+c)^2]U
\end{pmatrix}
\]

where \(J = rs\gamma^2 a + 2g + (s-2)c\). Again if \(M(\xi_{12})\) is singular, we compute \(M^*(\xi_{12})\) by taking limits. As before we average \(\xi_{12}\) over permutations of
\((l, \ldots, s)\) to obtain \(\xi_0\).

\[
\xi_0(x_{ij}) = A = a, \\
\xi_0(x_{ij}) = B = [2e + 4(s-2)f + (s-2)(s-3)b] / [s^2-s], \\
\xi_0(y_{ij}) = C = 2g/s + c(s-2)/s.
\]

By (4.16) (replacing \(a, b\) and \(c\) there by \(A, B\) and \(C\))

\[
M^*(\xi_0) = (s^2-s)^{-1}[s(s-1)a + 2(s-2)e + 4(s-2)^2f + (s-2)^2(s-3)b \\
+ 2r(s-1)g + (s-2)(s-1)rc]I \\
+ (s^2-s)^{-1}[2e + 4(s-2)f + (s-2)(s-3)b \\
- (s-1)s^{-1}rS^{-1}(sy_0+2g) + (s-2)c^2]U.
\]

We set the upper left entries of \(M^*(\xi_{12})\) and \(M^*(\xi_0)\) equal and use the fact that \(s > 2\) to obtain

\[(4.22) \quad 0 = (s-3)b - (s-4)f - e + r(c-g) + rs^{-1}J^{-1}[(s+2)(ya+g)^2 - 4(ya+g)(ya+c) \\
- (s-2)(ya+c)^2].\]

Now setting the lower right entries of the two matrices equal we get

\[(4.23) \quad 0 = (s-3)b - (s-4)f - e + r(c-g) + rs^{-1}J^{-1}[2(ya+g)^2 + 2(s-2)(ya+g)(ya+c) \\
- 2(s-1)(ya+c)^2].\]

Comparison of (4.22) and (4.23) gives

\[0 = s(ya+g)^2 - 2s(ya+g)(ya+c) + s(ya+c)^2,\]

and therefore \(c = g\).

Now we set the \((1, 2)\) and \((1, s)\) entries of \(M^*(\xi_{12})\) equal, because this equality holds in \(M^*(\xi_0)\), and use the fact that \(c = g\) to obtain \(e = f\).

If \(s = 3\) we have contradicted the assumptions. If \(s > 3\) we set the \((1, 2)\) entries of \(M^*(\xi_{12})\) and \(M^*(\xi_0)\) equal, getting \((s-2)(s-3)f = (s-2)(s-3)b\),
i.e., \( f = b \). This contradicts the assumptions.

**Case 3A.** Suppose \( r = 1 \), i.e., \( k = s + 1 \), and there is an optimal \( \xi \) with \( \xi(y_{1k}) \neq (rs)^{-1} \sum \xi(y_{ik}) \). As before we can average over the permutations of \( \{1, \ldots, k\} \) which have \( \{1\}, \{2, \ldots, s\} \) and \( \{k\} \) fixed. By the results already proved the resulting \( \xi_{1k} \) is constant on the set of points \( \{x_{ii}\} \) and on the set \( \{x_{ij}\} \). Thus \( \xi_{1k} \) is of the form

\[
\begin{align*}
\xi_{1k}(x_{ii}) &= a & \xi_{1k}(y_{ik}) &= f \\
\xi_{1k}(x_{ij}) &= b & \xi_{1k}(y_{ik}) &= h & 1 < i \leq s
\end{align*}
\]

where \( f \neq h \).

Then

\[
M(\xi_{1k}) = \begin{pmatrix}
a + (s-1)b + f & bV' & \gamma a + f \\
bV & [a + (s-2)b + h]I + bU & (\gamma a + h)V \\
\gamma a + f & (\gamma a + h)V' & \gamma^2 a + f + (s-1)h
\end{pmatrix}
\]

where in the central block \( I \) and \( U \) are both \( (s-1) \times (s-1) \). Then

\[
M^*(\xi_{1k}) = \begin{pmatrix}
a + (s-1)b + f - J^{-1}(\gamma a + f)^2 & [b - J^{-1}(\gamma a + f)(\gamma a + h)]V' \\
[b - J^{-1}(\gamma a + f)(\gamma a + h)]V & [a + (s-2)b + h]I \\
& + [b - J^{-1}(\gamma a + h)^2]U
\end{pmatrix}
\]

where \( J = \gamma^2 a + f + (s-1)h \). Averaging over permutations of \( \{1, \ldots, s\} \) we obtain an optimal \( \xi_0 \) with

\[
M^*(\xi_0) = [a + (s-2)b + f/s + h(s-1)/s]I
+ [b + s^{-2}J^{-1}(s\gamma a + f + (s-1)h^2)^2]U.
\]

Setting the upper left entries of \( M^*(\xi_{1k}) \) and \( M^*(\xi_0) \) equal gives

\[
(4.24) \quad 0 = s(f-h) + J^{-1}[-(s+1)(\gamma a+f)^2 + 2(\gamma a+f)(\gamma a+h) + (s-1)^2(\gamma a+h)^2].
\]

Setting the lower right entries equal gives
\( 0 = s(f-h) + \frac{1}{J}[-(\gamma_a+f)^2 - 2(s-1)(\gamma_a+f)(\gamma_a+h) + (2s-1)(\gamma_a+h)^2]. \)

Therefore

\[
0 = s(\gamma_a+f)^2 - 2s(\gamma_a+f)(\gamma_a+h) + s(\gamma_a+h)^2
\]

\[
0 = s(f-h)^2
\]

and \( f = h \), contradicting the assumption.

Case 3B. Suppose that \( r > 1 \) and that there is an optimal \( \xi \) with

\[
\xi(y_{1k}) = (r-1)^{-1} \sum \xi(y_{ij}).
\]

As before we average over the permutations of \( \{1,\ldots,k\} \) which leave \( \{1\}, \{2,\ldots,s\}, \{s+1,\ldots,k-1\} \) and \( \{k\} \) fixed. The resulting \( \xi_{1k} \) is constant on \( \{x_{ii}\} \) and on \( \{x_{ij}\} \), as shown by the first two cases. Thus

\[
\begin{align*}
\xi_{1k}(x_{ii}) &= a \\
\xi_{1k}(x_{ij}) &= b \\
\xi_{1k}(y_{ij}) &= c \quad i > 1, j < k \\
\xi_{1k}(y_{1k}) &= f \\
\xi_{1k}(y_{1j}) &= g \quad j < k \\
\xi_{1k}(y_{ik}) &= h \quad i > 1
\end{align*}
\]

where \( c, f, g \) and \( h \) are not all equal. Then \( M(\xi_{1k}) \) is given by

\[
M_1(\xi_{1k}) = 
\begin{bmatrix}
(a + (s+1)b + f + (r-1)g) & bV' \\
& bV \\
& [a + (s-2)b + k + (r-1)c]I + bU
\end{bmatrix}
\]

with \( I \) and \( U \) both \( (s-1) \times (s-1) \),

\[
M_2(\xi_{1k}) = 
\begin{bmatrix}
[g + (s-1)c]I + \gamma^2saU & \gamma^2saV' \\
& \gamma^2saV \\
& \gamma^2sa + f + (s-1)h
\end{bmatrix}
\]

with \( I \) and \( U \) both \( (r-1) \times (r-1) \), and

\[
M_3(\xi_{1k}) = 
\begin{bmatrix}
(\gamma a+g)V' & \gamma a+f \\
& (\gamma a+c)U \\
& (\gamma a+h)V
\end{bmatrix}
\]
where \( U \) is \((s-1) \times (r-1)\) and \( V' \) and \( V \) are of length \( r-1 \) and \( s-1 \) respectively. Let \( S = \gamma^2a, R = f + (s-1)h \) and \( T = g + (s-1)c \). Write \( J = (r-1)SR + ST + RT \). Then (if \( M_3(\xi_{1k}) \) is nonsingular) one can check that

\[
M_3^{-1}(\xi_{1k}) = \begin{bmatrix}
T^{-1}[I - SRJ^{-1}U] & -SJ^{-1}V \\
-SJ^{-1}V' & [(r-1)S + T]J^{-1}
\end{bmatrix}
\]

with \( I \) and \( U \) both \((r-1) \times (r-1)\). Taking limits if \( M(\xi_{1k}) \) is singular, we find after some cancellation that the entries of \( M^*(\xi_{1k}) \) are as follows:

\[
M^*_{11}(\xi_{1k}) = a + (s-1)b + f + (r-1)g - J^{-1}[(r-1)((\gamma a + g)^2R + (g-f)^2S) + (\gamma a + f)^2T],
\]

and for \( i > 1 \)

\[
M^*_{ii}(\xi_{1k}) = b - J^{-1}[(r-1)((\gamma a + g)(\gamma a + c)R + (g-f)(c-h)S) + (\gamma a + f)(\gamma a + h)T],
\]

and the lower right \((s-1) \times (s-1)\) block of \( M^*(\xi_{1k}) \) is

\[
[a + (s-2)b + h + (r-1)c]I + (b - J^{-1}[(r-1)((\gamma a + c)^2R + (c-h)^2S) + (\gamma a + h)^2T])U.
\]

As before we average \( \xi_{1k} \) over all permutations of \( \{1, \ldots, s\} \) and \( \{s+1, \ldots, k\} \) to get an optimal \( \xi_0 \),

\[
\xi_0(x_{1i}) = a
\]

\[
\xi_0(x_{ij}) = b
\]

\[
\xi_0(y_{ij}) = (rs)^{-1}[f + (r-1)g + (s-1)h + (r-1)(s-1)c]I
\]

and

\[
M^*(\xi_0) = s^{-1}[sa + s(s-2)b + f + (r-1)g + (s-1)h + (r-1)(s-1)c]I + [b - s^{-2}(\gamma ars + f + (r-1)g + (s-1)h + (r-1)(s-1)c)/K]U,
\]
where
\[ K = r^2 y^2 a + f + (r-1)g + (s-1)h + (r-1)(s-1)c. \]

Setting the upper left entries of the two matrices equal, we have
\[
(4.26) \quad 0 = (s-1)(f-h) + (s-1)(r-1)(g-c) + s^{-1} K^{-1} [yars + R + (r-1)T]^2
- s J^{-1} [(r-1)((y+a+g)^2 R + (g-f)^2 S) + (y+a+f)^2 T].
\]

Likewise setting the lower right entries equal gives
\[
(4.27) \quad 0 = (s-1)(f-h) + (s-1)(r-1)(g-c) + (s-1)s^{-1} K^{-1} [yars + R + (r-1)T]^2
+ s(s-1) J^{-1} [(r-1)((y+a+g)^2 R + (c-h)^2 S) + (y+a+h)^2 T].
\]

We equate (4.26) and (4.27) to get
\[
(4.28) \quad K^{-1} [yars + R + (r-1)T]^2
= s J^{-1} [(s-1) [(r-1)((y+a+g)^2 R + (c-h)^2 S) + (y+a+h)^2 T]
+ [(r-1)((y+a+g)^2 R + (g-f)^2 S) + (y+a+f)^2 T]].
\]

This can be simplified by looking at the off diagonal elements. We set the upper right entries of the two matrices equal, getting
\[
(4.29) \quad J^{-1} [(r-1)((y+a+c)(y+a+g)R + (g-f)(c-h)S) + (y+a+f)(y+a+h)T]
= s^{-2} K^{-1} [yars + R + (r-1)T]^2
= s^{-1} J^{-1} [(s-1) [(r-1)((y+a+g)^2 R + (c-h)^2 S) + (y+a+h)^2 T]
+ [(r-1)((y+a+g)^2 R + (g-f)^2 S) + (y+a+f)^2 T]],
\]

using (4.28) to get the last line.

If \( s > 2 \) there are other off diagonal elements of \( M^*(s_{1k}) \), which are equal to each other. Therefore
\[
(4.30) \quad [(r-1)((y+a+c)^2 R + (c-h)^2 S) + (y+a+h)^2 T]
= [(r-1)((y+a+g)(y+a+c)R + (g-f)(c-h)S) + (y+a+f)(y+a+h)T].
\]
Combining this with (4.29) we get that all the expressions in square brackets in (4.29) and (4.30) are equal, and thus we have

\[(4.31)\quad 0 = [(r-1)((ya+c)^2R + (c-h)^2S) + (ya+h)^2T] + [(r-1)((ya+g)^2R + (g-f)^2S) + (ya+f)^2T] - 2[(r-1)((ya+c)(ya+g)R + (c-h)(g-f)S) + (ya+h)(ya+f)T].\]

That was if \(s > 2\). If \(s = 2\) equation (4.29) reduces immediately to (4.31). Therefore in any case

\[(4.32)\quad 0 = R(r-1)[(ya+c) - (ya+g)]^2 + S(r-1)[(c-h) - (g-f)]^2 + T[(ya+h) - (ya+f)]^2 = R(r-1)(c-g)^2 + S(r-1)[(c-h) - (g-f)]^2 + T(h-f)^2.\]

Recall the definitions \(S = \gamma^2sa\), \(R = f + (s-1)h\) and \(T = g + (s-1)c\).

Now \(a = \xi_0(x_{11}^i) > 0\), since \(\xi_0\) is invariant and optimal. Thus \(S > 0\) (and \(r > 1\)) so by (4.32),

\[(4.33)\quad c - g + f - h = 0.\]

Therefore if \(c = g\) then \(h = f\). But if \(c\) and \(g\) were not equal, at least one of the two would be positive, and thus \(T > 0\). Therefore by (4.32) \(h = f\), so by (4.33), \(c = g\). Thus \(c = g\) and \(h = f\).

Substituting \(c = g\) and \(h = f\) into (4.28) we get

\[K^{-1}[\gamma a r + h + (r-1)c]^2 = J^{-1}[(r-1)((ya+c)^2sh + (c-h)^2\gamma^2sa + (ya+h)^2sc)].\]

After a lot of routine computation this reduces to

\[0 = (1-\gamma r)^2(c-h)^2.\]

Since \(\gamma r \neq 1\), \(c = h = f = g\), contradicting the assumptions.
We have thus shown that for $s > 1$ the only optimal designs are invariant and that the only invariant optimal designs require $N$ points of support. This proves that the example is as asserted.
REFERENCES


