Elfving's Theorem and Optimal Designs for Quadratic Loss

by

W.J. Studden*

Department of Statistics

Division of Mathematicsl Science

Mimeograph Series #227

June 23, 1970

* This research was supported by the National Science Foundation Grants GP 8986 and GP 20306. Reproduction is permitted in whole or in part for any purposes of the United States Government.

Abstract - Elfving's Theorem and Optimal

Designs for Quadratic Loss

by W.J. Studden

The theorem due to Elfving mentioned in the title is concerned with the optimal allocation of experiments in estimating linear functions of regression parameters. The purpose of the present paper is to give a matrix analog of this theorem and to give some simple applications.

Elfving's Theorem and Optimal

Designs for Quadratic Loss

by

W.J. Studden*

Purdue University

§1. Introduction. The purpose of this paper is to give a matrix analog of a geometric result of Elfving in the theory of optimal design of experiments. The connection with quadratic loss is indicated below.

Let $f=(f_1,\ldots,f_m)$ denote m linearly independent continuous functions on a compact set X. For each $x\varepsilon X$ an experiment can be performed. The outcome is a random variable y(x) with mean value $\theta\, f'(x)=\sum_i \theta_i f_i(x)$ and a variance σ^2 independent of x. (Primes will denote transposes.) The functions $f_1,\ldots,f_m$, called the regression functions, are assumed known while $\theta=(\theta_1,\ldots,\theta_m)$ and σ^2 are unknown. An experimental design is a probability measure μ on X. In practice, the experimenter is allowed N uncorrelated observations and the number of observations that he takes at each $x\varepsilon X$ is "proportional" to the measure μ. For a given design μ let $m_{ij}=m_{ij}(\mu)=\int f_i f_j\, d\mu$ and $M(\mu)=||m_{ij}||_{i,j=1}^{m}$. The matrix $M(\mu)$ is called the information matrix of the design .

Suppose μ concentrates mass μ_i at the points x_i, $i=1,\ldots,r$ and $N\mu_i=n_i$ are integers. If N uncorrelated observations are made, taking n_i observations of x_i, then the variance of the best linear unbiased estimate of

*This research was supported by the National Science Foundation Grants GP 8986 and GP 20306. Reproduction is permitted in whole or in part for any purposes of the United States Government.

$a\theta' = \sum_i a_i\theta_i$ is given by $\sigma^2 N^{-1} aM^{-1}(\mu)a'$. The inverse $M^{-1}(\mu)$ must be suitably defined if $M(\mu)$ is nonsingular. A design μ is called a-optimal if μ minimizes $V(a,\mu) = aM^{-1}(\mu)a'$. The following geometric result was given by Elfving (1952); see also Karlin and Studden (1966).

Theorem (Elfving). Let R denote the smallest convex set in Euclidean m-space which is symmetric with respect to the origin and contains all of the vectors $f(x)=(f_1(x),\ldots,f_m(x))$, $x\varepsilon X$. A design μ_o is a-optimal if and only if there exists a scalar valued function $\phi(x)$ satisfy $|\phi(x)|\equiv 1$ such that (i) $\int\phi(x)\ f(x)d\mu_o(x)= \beta a$ for some β and (ii) βa is a boundary point of R. Moreover βa lies on the boundary of R if and only if $\beta^2=v^{-1}$ where $v= \min_\mu V(a,\mu)$.

The quantity, analogous to $V(a,\mu)$, that we wish to consider is

$$(1.1)\qquad V(A,\mu) = \text{tr } A'M^{-1}(\mu)A = \text{tr } M^{-1}(\mu)\ AA'$$

where A is an mxk matrix and tr denotes the trace. We thus wish to minimize the sum of quantities $V(a,\mu)$ where the a's are given by the columns of A.

The expression $V(A,\mu)$ can be also be seen to be proportional to $E(\hat{\theta}-\theta)\ AA'(\hat{\theta}-\theta)'$ where $\hat{\theta}$ denotes the least squares estimate of θ. This is the reason for part of the title of the paper.

In the following sense the expression $V(A,\mu)$ provides some "generality". Let $L(B)$ denote a linear function on the set of mxm matrices which is positive in the sense that $L(B) \geq 0$ for B positive semidefinite. Then $L(B)=\text{tr } BC$ for some positive definite C. Thus $V(A,\mu)$ is the most general positive linear function in $M^{-1}(\mu)$.

A design μ is called A-optimal if it minimizes $V(A,\mu)$. In order to state the matrix analog of Elfving's theorem we let $\phi=(\phi_1,\ldots,\phi_k)$ and define R as the smallest convex set of mxk matrices which contains all the matrices $f'(x)\phi$ where $x\varepsilon X$ and $\Sigma\phi_i^2=|\phi|^2 \leq 1$. (The symbol $|\cdot|$ will denote the usual Euclidean norm.) We then have the following result.

Theorem 1.1. A design μ_o is A_o-optimal if and only if there exists a function $\phi(x)$ satisfying $|\phi(x)| \equiv 1$ such that (i) $\int f'(x)\ \phi(x)d\mu_o(x)= \beta A_o$ for some scalar β and (ii) βA_o is contained in the boundary of R. Moreover βA_o lies on the boundary of R if and only if $\beta^{-2} = v_o= \min_{\mu} V(A_o,\mu)$.

A more complete discussion of the function $V(A,\mu)$ is given in §2 while the proof of Theorem 1.1 and some preliminary lemmas are given in §3. A more useful from of the theorem is given in Theorem 3.1. Various simple applications are given in §4 and in §5 we discuss briefly the choice of a basis in regression theory.

The application of Theorem 1.1 is, at present, somewhat limited (as are most results on the optimal choice of design) in that it appears difficult in any given situation to determine the points where the observations are to be taken. Some iterative computational procedures are available both for the minimization of tr $M^{-1}(\mu)$ AA' and for maximizing the determinant of $M(\mu)$. See for example Fedorov (1968) and Fedorov and Dubova (1968).

We wish to thank Professor J. Yackel for a helpful discussion concerning Lemma 3.1.

§2. The Function $V(A,\mu)$. Whenever $M(\mu)$ is nonsingular the quantity $V(A,\mu)=\text{tr } AA'\ M^{-1}(\mu)$ is well defined. With the aid of Schwartz's inequality it can be shown that for any mxk matrix E

$$(2.1) \qquad \mathrm{tr}^2\, E'A \leq \mathrm{tr}\, E'M(\mu)\, E\ \mathrm{tr}\, A'M^{-1}A.$$

and equality occurs if and only if A is proportional to $M(\mu)E$. Therefore

$$(2.2) \qquad V(A,\mu) = \sup_E \frac{\mathrm{tr}^2\, E'A}{\mathrm{tr}\, E'M(\mu)E}$$

When $M=M(\mu)$ is singular we take $V(A,\mu)$ as defined by (2.2) where the sup is over those E such the both numerator and denominator do not vanish simultaneously. Thus, in order that $V(A,\mu)$ be finite we must have each column of A orthogonal to every vector e such that $Me = 0$. That is, the columns of A must be in the range of $M(\mu)$. We can therefore restrict the columns of E to also be in the range of M. Let $\lambda_1,\ldots,\lambda_s$ be the nonzero eigenvalues of M with associated orthonormal eigenvectors $v_1,\ldots,v_s$. Then $M=\Sigma\lambda_i v_i' v_i$ and if we define

$$(2.3) \qquad M^{\varepsilon} = \Sigma\lambda_i^{\varepsilon}\, v_i'\, v_i \quad \text{for } \varepsilon = \pm 1 \text{ or } \pm 1/2$$

then $M^{1/2}\, M^{-1/2} = \Sigma v_i'\, v_i$. If the columns of A are in the range of M it follows that $(\Sigma v_i'\, v_i)A = A$. Then by Schwartz's inequality

$$(2.4) \qquad \mathrm{tr}^2\, E'A = \mathrm{tr}^2\, E'M^{1/2}\, M^{-1/2}\, A$$
$$\leq \mathrm{tr}\, E'ME\ \mathrm{tr}\, A'M^{-1}\, A$$

and equality occurs if and only if A is proportional to ME. We shall usually take the proportionality constant so that

$$\beta A = ME \quad \text{or} \quad \beta M^{-1}A = E \tag{2.5}$$

where $\beta^{-2} = \text{tr } A'M^{-1}A$.

We have now shown that when the columns of A are in the range of $M(\mu)$ then $V(A,\mu)=\text{tr } A'M^{-1}(\mu)A$ where the inverse is given by (2.3). Otherwise $V(A,\mu)= \infty$.

§3. Preliminary Lemmas and Proof of Theorem 1.1.

Lemma 3.1. Let R denote the smallest convex set containing the mxk matrices $f'(x)\phi$, $x\varepsilon X$ and $|\phi|^2= \Sigma\phi_i^2 \leq 1$. Then

$$R = \{A|\text{tr}^2\ E'A \leq \sup_x f(x)EE'f'(x) \quad \forall\ E\}$$

where E is an mxk matrix.

Proof. Let R_o denote the convex set defined in parenthesis above. Then for $A = f'(x)\phi$,

$$\begin{aligned} \text{tr}^2\ E'f'\phi &\leq \text{tr } f(x)\ EE'f'(x)\ \text{tr } \phi'\ \phi \\ &\leq f(x)\ EE'\ f'(x). \end{aligned} \tag{3.1}$$

Therefore $R \subset R_o$. Now suppose $A_o \notin R$. Since R is easily seen to be closed and bounded there exists a hyperplane strictly separating A_o and R. Thus there exists E_o and a_o such that

$$\text{tr } E_o'\ A \leq a_o < \text{tr } E_o'\ A_o \quad \text{for all } A\varepsilon R \tag{3.2}$$

Without loss of generality we take $a_o= 1$. In (3.2) we take $A=f'(x)\phi$ where

$\phi = f(x)\, E_o / \,|f(x)E_o|$. Then

$$f(x)\, E_o E_o'\, f'(x) \le 1 < \mathrm{tr}^2\, E_o'\, A_o \tag{3.3}$$

for all x and hence $A_o \notin R_o$.

Corollary 3.1. (i) Every matrix $A \varepsilon R$ has a representation $A = \sum_\nu f'(x_\nu)\phi(\nu)\, p_\nu$ where $|\phi(\nu)| \le 1$ and $\sum_\nu p_\nu = 1$ and the x_ν are not necessarily distinct.

(ii) Every matrix A in the boundary of R has a representation

$$A = \sum_\nu f'(x_\nu)\, \phi(x_\nu)\, p_\nu \tag{3.4}$$

where $|\phi(x_\nu)| = 1$, $\Sigma p_\nu = 1$ and the x_ν are all distinct. Both of the sums in the above representations are finite.

Arguments similar to those used in Lemma 3.1 may be used to prove the following lemma.

Lemma 3.2. A matrix A of the form (3.4) is a boundary point of R if and only if there exists a "supporting plane" E such that

$$f(x)\, EE'\, f'(x) \le 1 \quad \text{for all } x \varepsilon X \tag{3.5}$$

and equality holds for each x_ν (if $p_\nu > o$). Moreover $\phi(x_\nu) = f(x_\nu)E/|f(x_\nu)E|$ and tr E' A = 1.

Proof of Theorem 1.1. First suppose that μ_o and ϕ are such that $\int f'(x)\, \phi(x)\, d\mu_o(x) = \beta A_o$ and that βA_o is on the boundary of R. Then by

Lemma 3.2 there exists an E_o such that

(3.6) $$\beta \text{ tr } E_o' A_o = 1 \quad \text{and} \quad f(x)\, E_o E_o'\, f'(x) \leq 1 \quad \text{for all } x$$

with equality holding for x in the spectrum of μ_o. Therefore,

(3.7) $$\sup_x f(x)\, E_o E_o'\, f'(x) = 1.$$

For any design μ we have

$$\begin{aligned} \text{tr } E' M(\mu)\, E &= \text{tr } EE' \int f' f \, d\mu \\ &\leq \sup_x \text{tr } EE'\, f'(x)\, f(x) \\ &= \sup_x f(x)\, EE'\, f'(x) \end{aligned}$$

Then

$$\begin{aligned} V(A_o, \mu) &\geq \frac{\text{tr}^2\, E_o' A_o}{\text{tr } E_o'\, M(\mu) E_o} \\ &\geq \frac{\text{tr}^2\, E_o'\, A_o}{\sup_x f(x) E_o E_o'\, f'(x)} \end{aligned}$$

This inequality together with (3.6) and (3.7) imply that

(3.8) $$V(A_o, \mu) \geq \frac{1}{\beta^2}\ .$$

Now for the measure μ_o and any E we apply Schwartz's inequality twice to give

$$\begin{aligned}\mathrm{tr}^2\, E'A_o &= \beta^{-2}(\mathrm{tr}\, E' \int f'(x)\ \phi(x)\ d\mu_o(x))^2 \\ &\leq \beta^{-2} \int [\mathrm{tr}\, E'f'(x)\ \phi(x)]^2\ d\mu_o(x) \\ &\leq \beta^{-2} \int \mathrm{tr}(E'f'(x)\ f(x)\ E)\ d\mu_o(x) \\ &= \beta^{-2}\ \mathrm{tr}\, E'M(\mu_o)\ E\end{aligned}$$

Therefore

$$V(A_o,\mu_o) = \sup \frac{\mathrm{tr}^2\, E'A_o}{\mathrm{tr}\, E'M(\mu_o')E} \leq \frac{1}{\beta^2}\ .$$

This inequality combined with (3.8) shows that μ_o is A_o-optimal.

Note that there always exists a design μ satisfying (i) and (ii) so that the above analysis proves the last sentence of the theorem, namely that $v_o = \beta^{-2}$ for βA on the boundary of R.

We now let μ_o be any A_o-optimal design and wish to show that (i) and (ii) are satisfied for some ϕ. We take $\beta^{-2} = v_o$ so the βA_o lies on the boundary of R. Then there exists E_o so that

$$f(x)\ E_oE_o'\ f'(x) \leq 1 = \beta^2\ \mathrm{tr}^2\ E_o'A_o. \tag{3.9}$$

Integrating the left side with respect to μ_o we obtain

$$\mathrm{tr}\ E_o'\ M(\mu_o)\ E_o \leq 1 \tag{3.10}$$

However since μ_o is A_o-optimal we have

$$\frac{\mathrm{tr}^2\, E_o' A_o}{\mathrm{tr}\, E_o' M(\mu) E_o} \leq V(A_o, \mu_o) = \frac{1}{\beta^2}$$

so that $\mathrm{tr}\, E_o' M(\mu_o) E_o \geq \beta^2\, \mathrm{tr}^2\, E_o' A_o = 1$. Therefore $\mathrm{tr}\, E_o' M(\mu_o)\, E_o = \beta^2\, \mathrm{tr}^2 E_o' A_o$ and by the sentence containing (2.4) we must have A_o proportional to $M(\mu_o) E_o$. The latter part of (3.9) shows that

(3.11) $$\beta A_o = \varepsilon M(\mu_o)\, E_o \qquad \text{where } \varepsilon = \pm 1$$

In this case

$$\beta A_o = \varepsilon \int f'(x)\, f(x)\, E_o\, d\mu_o(x)$$

$$= \int f'(x)\, \phi(x)\, d\mu_o(x)$$

where $\phi(x) = \varepsilon f(x) E_o$ for x in the spectrum of μ_o. The vector ϕ has length one since equality must occur in (3.9) for x in the spectrum of μ_o.

For a given matrix A it is usually difficult to determine the spectrum of any A-optimal design μ. Theorem 3.1 below is sometimes useful in determining those A which have an optimal design supported on a given set of points.

In many cases the "boundary representation"

(3.12) $$\beta A = \sum_{\nu} f'(x_\nu)\, \phi(x_\nu) p_\nu$$

will reduce to a finite sum with at most m terms. If the number of terms is less than m we add arbitrary points with corresponding $p_\nu = 0$. We shall assume in this case that the determinant F with columns $f(x_\nu)$ is nonsingular.

Let $\ell'(x) = Tf'(x)$ denote the vector of Lagrange functions for the points $x_1,\ldots,x_m$, i.e. $\ell_i(x_j) = \delta_{ij}$. Inserting the values $x_1,\ldots,x_m$ in $\ell'(x) = Tf'(x)$ gives $I = TF$ so that $T = F^{-1}$. If we multiply (3.12) by T and let $TA = B$ then

$$\beta B = \sum_\nu \ell'(x_\nu) \quad \phi(x_\nu)p_\nu \ .$$

In this case $\beta b_\nu = \phi(x_\nu)p_\nu$ where b_ν denotes the νth row of B. Then

(3.13) $$\beta = (\Sigma|b_j|)^{-1},\ p_\nu = \beta|b_\nu| \quad \text{and} \quad \phi(x_\nu) = b_\nu \ |b_\nu|^{-1}$$

In case $|b_\nu| = 0$ we have $p_\nu = 0$ and $\phi(x_\nu)$ need not be defined.

For any matrix B we take each nonzero row and replace it by $b_\nu|b_\nu|^{-1}$. The resulting matrix is denoted by B_o. Thus if B_d^{-1} denotes the diagonal matrix with diagonal elements $|b_\nu|^{-1}$ for $|b_\nu| \neq 0$ and zero if $|b_\nu| = 0$ then

(3.14) $$B_o = B_d^{-1} B.$$

The following theorem characterizes those A with an optimal design supported on a given set $x_1,\ldots,x_m$.

<u>Theorem 3.1</u>. If F is nonsingular then an A-optimal design is supported on $x_1,\ldots,x_m$ if and only if there exists a matrix B such that

(i) $\ell(x)\ B_o\ B_o'\ \ell'(x) \leq 1 \qquad \forall\ x.$

(ii) $A = FB$

The optimal weights are then proportional to the lengths of the rows of B.

<u>Proof</u>. Suppose first that a matrix B exists satisfying (i) and (ii).

An A-optimal design then concentrates mass p_ν on x_ν where p_ν is proportional to the νth row of B. To see this we observe that with p_ν and $\phi(x_\nu)$ as in (3.13) we have (3.12) holding. Moreover (i) implies that

$$f(x)\ T'\ B_o B_o'\ T\ f'(x) \leq 1 \qquad \text{for all } x$$

and

$$\operatorname{tr} B_o'\ T(\beta A) = \beta \operatorname{tr} B'\ B_d^{-1}\ B$$
$$= \beta \operatorname{tr} B_d^{-1}\ BB' = 1$$

Therefore $\beta A \varepsilon \text{Bdry } R$ and the result follows by Theorem 1.1.

Now suppose that an optimal design μ_o is supported on $x_1,\ldots,x_m$. The optimal weights p_ν must be as in (3.13) and $\beta A = \Sigma p_\nu\ f'(x_\nu)\ \phi(x_\nu)$ with $\beta A \varepsilon \text{Bdry } R$. The hyperplane supporting R at βA then gives

(3.15) $$f'(x)\ E_o E_o'\ f(x) \leq 1 = \operatorname{tr} E_o'\ A$$

so that (i) holds with $B_o = F'E_o$. From (2.5) we know that $\beta A = M_o E_o$ so that $\beta A = \beta F C_d\ F'\ E_o$ where $C = TA$. In this case (iii) holds with $B = C_d B_o$.

§4. Applications. Polynomial extrapolation: Theorem 3.1 with $k=1$, $X=[-1,1]$, $f(x)=(1,x,\ldots,x^n)$ reduces fairly readily to the extrapolation result of Hoel and Levine (1964); see also Studden (1968). If k=1 the matrix A has one column. We take x_ν $\nu=0,\ldots,n$ to be the extrema of the Tchebycheff polynomial T_n of the first kind, i.e. $x_\nu = -\cos\frac{\nu\pi}{n}$, $\nu=0,1,\ldots,n$ and $T_n^2(x) \leq 1$ with equality holding at $x=x_\nu$. If we take the elements of the column vector B to have alternating sign then $\ell(x)\ B_o B_o'\ \ell'(x) \leq 1$ since $\ell(x)\ B_o = \pm T_n(x)$. Clearly $A=FB$ for some such B if $A=f'(x_o)$, $|x_o| > 1$. Thus the optimal design for extrapolating to x_0 concentrates on the x_ν defined above.

Linear Regression. In this case we take $f(x)=(1,x)$ and $X=[a,b]$ and apply Theorem 3.1. It is readily seen that (i) holds with $x_1=a$ and $x_2=b$ for any matrix B due to the linearity of the regression functions. That is, if $\ell(x)\ B_o=(P_1(x),\ P_2(x))$ then $P_1^2(a)+P_2^2(a)\leq 1$ and $P_1^2(b)+P_2^2(b)\leq 1$, (usually equality will hold). Then $x=\alpha a+(1-\alpha)b$ where $\alpha=(b-x)/(b-a)$ so that $P_i(x)=\alpha P_i(a)+(1-\alpha)\ P_i(b)$ and $P_1^2(x)+P_2^2(x)\leq 1$. For any matrix A we let a_1 and a_2 denote its rows. Since the weights of the A-optimal design are then proportional to the rows of B, we find that the weights on a and b are proportional to the square roots of $b^2|a_1|^2+|a_2|^2-ba_1a_2'$ and $a^2|a_1|^2+|a_2|^2-aa_1a_2'$. Note that in the case $a=-b$ the weights will be equal if and only if $a_1a_2'=0$, i.e. the two rows of A are orthogonal. This is the situation when, for example, (i) A is diagonal or (ii) A has rows (1,1) and (1,-1), i.e. we estimate the sum and difference of the regression coefficients.

Linear Spline Regression. Here we take $X=[a,b]$ and let $f(x)$ consist of the functions $1,(x-\xi_0)_+,(x-\xi_1)_+,(x-\xi_2)_+,\ldots,(x-\xi_h)_+$ where $\xi_0=a<\xi_1<\ldots<\xi_h<\xi_{h+1}=b$ and $z_+=z$ for $z>0$ and 0 for $z\leq 0$. The regression function is a polygonal line segment. The argument used for the ordinary linear case shows that (i) again holds for $x_1,\ldots,x_m$ equal $\xi_0,\xi_1,\ldots,\xi_{h+1}$ and any matrix B. The matrix $T=F^{-1}$ has three nonzero entries starting at the diagonal (except for the last two rows). The first row has $1,\ -(\xi_1-\xi_0)^{-1},\ (\xi_1-\xi_0)^{-1}$ while the ith row, for $i=2,\ldots,h+2$, has entries

$$\frac{1}{\xi_i-\xi_{i-1}}\ ,\ \frac{-(\xi_{i+1}-\xi_{i-1})}{(\xi_{i+1}-\xi_i)(\xi_i-\xi_{i-1})}\ ,\ \frac{1}{\xi_{i+1}-\xi_i}$$

If we take h=1, and A to have zero entries except in the lower right corner we then wish to estimate the coefficient of $(x-\xi_1)_+$. The optimal design has weights

$$\frac{\xi_2-\xi_1}{2(\xi_2-\xi_0)}, \quad \frac{1}{2} \quad , \frac{\xi_1-\xi_0}{2(\xi_2-\xi_0)}$$

on the points $a = \xi_0, \xi_1$ and $b = \xi_2$.

For general h we take $A = (a_{ij})$ again to be diagonal with $a_{11} = a_{22} = 0$ and $a_{ii} = \gamma$ for $i=2,\ldots,h+2$. If the ξ_i are equally spaced on (ξ_o, ξ_{h+1}) the optimal design has weights on $\xi_0, \xi_1, \ldots, \xi_{h+1}$ proportional to $1, \sqrt{5}, \sqrt{6}, \sqrt{6}, \ldots, \sqrt{6}, \sqrt{5}, 1$.

Quadratic Regression. For simplicity we take $X = [-1,1]$ and $f(x)=(1,x,x^2)$ and consider those designs supported on the three points -1,0,1. Since $\ell(x)\ B_oB_o'\ \ell'(x)$ is a quadratic form and a polynomial of degree four, it can be checked that it is at most one on [-1,1] if and only if its derivative vanishes at x = 0. This can be seen to be the case if and only if the second row of B is orthogonal to the first minus the second. For example we can take B of the form

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & 0 & b_{23} \\ b_{11} & -b_{12} & b_{13} \end{pmatrix}$$

The corresponding matrix A = FB is of the form

$$A = \begin{pmatrix} \alpha & 0 & \varepsilon \\ 0 & \beta & 0 \\ \delta & 0 & \gamma \end{pmatrix}$$

Again the weights are proportional to the square roots of the diagonal of $BB' = TA\,A'T'$. If a_1, a_2 and a_3 denote the rows of A then the diagonal elements of $TA\,A'T'$ are

$$|b_1|^2 = |b_3|^2 = \frac{1}{4}\left(|a_2|^2 + |a_3|^2\right), \quad |b_2|^2 = |a_1|^2 + |a_3|^2 - 2a_1 a_3'$$

As special cases we take $\delta = \varepsilon = 0$ then

$$|b_1|^2 = |b_3|^2 = (\beta^2 + \gamma^2)/4$$

$$|b_2|^2 = \alpha^2 + \gamma^2$$

If $A = I$ = the identity, then $\alpha = \beta = \gamma = 1$ and the weights on $-1, 0, 1$ are proportional to 1,2,1. This design can also be shown to minimize $\int f(x)\, M^{-1}(\mu)\, f'(x)dx = \operatorname{tr} M^{-1}(\mu)\, C$ where

$$C = \begin{pmatrix} c_0 & 0 & c_2 \\ 0 & c_2 & 0 \\ c_2 & 0 & c_4 \end{pmatrix}$$

and $c_i = \int_{-1}^{1} x^i\, dx$.

<u>Cubic Regression</u>. For simplicity we take $A = I$, $X = [-1,1]$ and $f(x) = (1, x, x^2, x^3)$. One can show that there exists an A-optimal symmetric design on four points $-1, -s, s, 1$. The quantities A and F are thus determined and $B = TA$. We can argue that $\ell(x)\, B_0 B_0'\, \ell'(x) \leq 1$ for all x if and only if the derivative of the left side is zero at $x=s$. A rather tedious calculation

shows that $s = (\sqrt{7}-2)/3$ and that the weights on $-1,-s,s,1$ are proportional to the square roots of $1+s^4$, $(1+s^2)s^{-2}$, $(1+s^2)s^{-2}$, $1+s^4$. These values are approximately $s = .215$ and the weights are .087, .413, .413 and .087.

§5. Choice of Basis. In this section we indicate a connection between the quadratic loss designs discussed above and the design which maximizes the determinant of $M(\mu)$ (see Kiefer (1960)). The result is of a simple nature and follows fairly readily from the known result that if G is a positive semidefinite matrix and $|G|$ denotes the determinant then

$$n|G|^{1/n} = \min_{|H|=1} \operatorname{tr} GH \tag{5.1}$$

where H is also positive semidefinite.

If we consider a change of basis $g' = Pf'$, then $M_g(\mu) = \int g'g\, d\mu = PM_f(\mu)P'$ and $\operatorname{tr} M_g^{-1}(\mu) = \operatorname{tr} M_f^{-1}(\mu)\, AA'$ where $A = P^{-1}$. As a measure of how good the basis is we consider

$$L(P) = \min_{\mu} \operatorname{tr} M_g^{-1}(\mu). \tag{5.2}$$

Some normalization of P must be used and we consider those P with $|P| = 1$. Using (5.1) we then have

Theorem 5.1. If $L(P)$ is defined as in (5.2) then

$$\min_{|P|=1} L(P) = m|M_f^{-1}(\mu_o)|^{1/m}$$

where μ_o is the design maximizing $|M_f(\mu)|$.

As an example we consider $f(x) = (1,x,\ldots,x^n)$ on $X = [-1,1]$ for n=1,2. It is well known that the design maximizing $|M_f(\mu)|$ concentrates equal mass on -1 and 1 for n=1 and on -1,0 and 1 for n=2. (The general case has equal mass on the n+1 zeros of $(1-x^2)\ P_n'(x) = 0$ where P_n is the Legendre polynomial). We consider four different basis; namely

1. $f(x) = (1,x,\ldots,x^n)$
2. T-basis: $g=k(T_0,\ldots,T_n)$ where T_n is the nth Tchebycheff polynomial.
3. B-basis: $g' = k(B_0,B_1,\ldots,B_n)$ where B_i denotes the Bernoulli polynomial $B_i(x) = \binom{n}{i}(1-x)^i\ (1+x)^{n-i}$.
4. L-basis: where $L_i(x)$ denotes the ith Lagrange polynomial corresponding to n+1 points $x_0,x_1,\ldots,x_n$, i.e. $L_i(x_j) = \delta_{ij}$.

In each case the proportionality constant k is used so that $P = 1$.

The case n=1 shows no distinction between the four basis. In each case $L(P) = 2$ as a direct calculation will verify. For n=2 however we get 1. $L(P) = 8$; 2. $L(P) = 5.90$; 3. $L(P) = 8.03$; 4. $L(P) = 5.67$. It is not clear that the ordering will be the same for higher values of n. The result for n=2 is in accord with results in approximation theory which indicate that the Tchebycheff basis is "good". By the above definition the Lagrange polynomials on -1,0,1 are better.

References

Elfving, G. (1952). Optimum allocation in linear regression theory. Ann. Math. Statist. 23 255-262.

Fedorov, V.V. (1968). Plans of experiments in the case of simultaneous measurements of several response surfaces. Publishing House of Moscow Univ. Preprint #4 (in Russian).

Fedorov, V.V. and Dubova, I.S. (1968). The methods of the construction of the optimal plans for regression experiments. Publishing House of Moscow Univ. Preprint #3 (in Russian).

Hoel, P.G. and Levine, A. (1964). Optimal spacing and weighing in polynomial regression. Ann. Math. Statist. 35 1553-1560.

Karlin, S. and Studden, W.J. (1966). Optimal experimental designs. Ann. Math. Statist. 37 783-815.

Kiefer, J. (1960). Optimum experimental designs V. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1 381-405. Univ. of California Press.

Studden, W.J. (1968). Optimal designs on Tchebycheff points. Ann. Math. Statist. 39 1435-1447.