Design Workshop Lecture Notes
ISI, Kolkata, November, 25-29, 2002, pp. 23-32

Optimal Regression Designs in Symmetric Domains

Nripes K Mandal
Calcutta University, Kolkata, India

1. Introduction

In this talk, we propose to discuss optimality study in the set-up of polynomial regression designs, following Pukelsheim (1993) and Liski et al (2002). Two classic texts in this area of research [containing lucid descriptions of early work] are : Fedorov (1972) and Silvey (1980).

We postulate a polynomial fit model of degree $k \geq 1$.

$$Y_{ij} = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k + e_{ij} \tag{1.1}$$

where

$$E(e_{ij}) = 0 \text{ and } V(e_{ij}) = \sigma^2$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, N_i$. The responses Y_{ij} are uncorrelated and the experimental conditions $x_1, x_2, \ldots, x_n$ are assumed to lie in $[-1, 1]$. Note that now the experimental domain $\mathcal{T} = [-1, 1]$ is an interval symmetric with respect to origin. The corresponding regression range $\chi = \{(1, x, \ldots, x^k)' : x \in \mathcal{T}\}$ is a one-dimensional curve embedded in $\mathbf{R}^{k+1}$. We remind ourselves that any collection $d_n = \{x_1, x_2, \ldots, x_n; p_1, p_2, \ldots, p_n\}$ of $n \geq 1$ distinct points $x_i \in \mathcal{T}$ and positive numbers p_i, $i = 1, 2, \ldots, n$ such that $\sum_{i=1}^{n} p_i = 1$, induces a continuous design d on the regression range χ (cf. Pukelsheim (1993, p. 32)). In what follows we will denote by $\mathcal{D}$ the set of all such designs. The *exact* design above yields $p_i = N_i/N$ where $N = \sum N_i$ is the total number of observations.

We stated the well-known de la Garza (1954) phenomenon. Let $d_n = \{x_1, x_2, \ldots, x_n; p_1, p_2, \ldots, p_n\}$ with $n > k + 1$ be an n-point design for the LSE of $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ in the polynomial fit model (1.1) of degree k. Then there exists a $(k + 1)$-point design $d^*_{k+1} = \{x^*_1, x^*_2, \ldots, x^*_{k+1};\ p^*_1, p^*_2, \ldots, p^*_{k+1}\}$ for the LSE of β in (1.1) such that $\mathbf{I}(d^*_{k+1}) = \mathbf{I}(d_n)$, where $\mathbf{I}(d_n)$ denotes the information matrix of the design d_n.

1.1 Symmetric Polynomial Designs

First we consider the *reflection* operation. Let $d \in \mathcal{D}$ be a design for the LSE of $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ on $\mathcal{T} = [-1, 1]$ in the polynomial fit model (1.1). The reflected design d^R is given by $d^R = \{-x_1, -x_2, \ldots, -x_n; p_1, p_2, \ldots, p_n\}$. The designs d and d^R have the same even moments, while the odd moments of d^R have a reversed sign.

If $\mathbf{I}(d^R)$ denotes the $(k + 1) \times (k + 1)$ information matrix of d^R, then

$$\mathbf{I}(d^R) = \mathbf{Q}\mathbf{I}(d)\mathbf{Q}, \tag{1.2}$$

where $\mathbf{Q} = Diag(1, -1, 1, -1, \ldots, \pm 1)$ is a diagonal matrix with diagonal elements $1, -1, 1, -1, \ldots, \pm 1$

The symmetrized design

$$\bar{d} = \frac{1}{2}(d + d^R) \equiv \{\pm x_i;\ \frac{p_i}{2}, \frac{p_i}{2} | 1 \leq i \leq n\}$$

assigns the weights $\frac{p_i}{2}$ to x_i and $-x_i$ for each i. The information matrix of $\bar{d}$ is

$$\mathbf{I}(\bar{d}) = \frac{1}{2}[\mathbf{I}(d) + \mathbf{Q}\mathbf{I}(d)\mathbf{Q}],$$

where all odd moments are zero and the even moments are equal to the corresponding moments of the original design d. Hence the averaging operation simplifies information matrices by letting all odd moments vanish. Since $\mathbf{I}(d^R)$ is obtained from $\mathbf{I}(d)$ by the similarity transformation (1.2), $\mathbf{I}(d^R)$ and $\mathbf{I}(d)$ have the same eigenvalues.

From the above, it follows that any optimality criterion which is a function of the eigenvalues of the information matrices will be invariant with respect to the reflection operation. It follows that superadditivity and invariance of an optimality functional ϕ (with respect to the reflection) imply

$$\begin{aligned} \phi[\mathbf{I}(\bar{d})] &= \phi\left\{\frac{1}{2}[\mathbf{I}(d) + \mathbf{I}(d^R]\right\} \\ &\geq \frac{1}{2}\{\phi[\mathbf{I}(d)] + \phi[\mathbf{I}(d^R)]\} \\ &= \phi[\mathbf{I}(d)]. \end{aligned}$$

Thus symmetrization improves the value of the criterion ϕ, or at least maintains the same value, provided that ϕ is superadditive and invariant with respect to the reflection. Therefore, for such criteria, we may confine ourselves to the class of symmetric designs.

1.2 Symmetric Designs for Quadratic Regression

Let d_n denote a symmetric n-point design on $\mathcal{T} = [-1,\ 1]$ for the LSE of $\beta\ =\ (\beta_0, \beta_1, \beta_2\}'$ in the quadratic regression model

$$Y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_{ij} \tag{1.3}$$

with the assumptions similar to the general polynomial fit model (1.1). It follows that

$$\mathbf{I}(d_n) \ = \ \begin{pmatrix} 1 & 0 & \mu_2 \\ 0 & \mu_2 & 0 \\ \mu_2 & 0 & \mu_4 \end{pmatrix}.$$

Whenever $n > 3$, we may obtain the *same* information matrix by using $d_3 = \{-a, 0, a; w/2, 1-w, w/2\}$ by choosing $a = \sqrt{\mu_4/\mu_2}$ and $w\ = \mu_2^2/\mu_4$. This is the spirit of de la Garza (DLG) Phenomenon. Henceforth, we will confine only to 3-point symmetric designs.

Now we prove for the model (1.3) the following result.

Theorem. Let d_3 be any 3-point symmetric design for the LSE of $\beta = (\beta_0, \beta_1, \beta_2)'$ in (1.3) with its support points in the interior of $\mathcal{T}$. Then there exists a symmetric 3-point design $d_3^* = \{-1, 0, 1;\ \frac{p}{2}, 1-p, \frac{p}{2}\}$ with $p < 1$ such that $d_3^* \succ d_3$ in the sense that $\mathbf{I}(d_3^*) \geq \mathbf{I}(d_3)$.

We skip the proof and instead refer to Liski et al (2002).

Now the A-optimal design, D-optimal design, E-optimal design and MV-optimal design can be determined in a straightforward manner. Note that in terms of matrix means ϕ_t, $-\infty \le t \le 1$, ϕ_{-1}, ϕ_0 and $\phi_{-\infty}$ are the A-optimality criterion, D-optimality criterion, and E-optimality criterion respectively. The characteristic function of $\mathbf{I}(d_p)$ corresponding to a 3-point design d_p is

$$f(\lambda) = (p-\lambda)[(1-\lambda)(p-\lambda) - p],$$

which yields the eigenvalues of $\mathbf{I}(d_p)$:

$$\begin{aligned} \lambda_1 &= \frac{1}{2}(p+1) + \frac{1}{2}\sqrt{5p^2 - 2p + 1}, \\ \lambda_2 &= p, \\ \lambda_3 &= \frac{1}{2}(p+1) - \frac{1}{2}\sqrt{5p^2 - 2p + 1}. \end{aligned}$$

The optimal designs are given in Liski et al (2002).

In the above, MV-optimal design is based on the criterion of minimization of the larger of the two variances of the estimators involving the linear and the quadratic terms.

It is clear from the nature of the above designs that, within the class of 3-point designs, further Loewner order domination is *not* possible. So there exists no Loewner optimal design. In spite of this limitation, it would be natural to examine the nature of a complete class of designs for a general polynomial regression model. Pukelsheim [(1993); Claim 10.7] states that d is admissible in (1.1) if and only if d has at most $k-1$ support points in the open interval $(-1,1)$, besides including the extreme points $+1/-1$. Thus the $(k+1)$-point designs $d_{k+1} = \{-1, t_2, \ldots, t_s, 1; p_1, p_2, \ldots, p_s, p_{k+1}\}$ with t_2, t_3, $\ldots$, $t_k \in (-1,1)$ and $\sum_{i=1}^{s+1} p_i = 1$ are admissible. This gives a starting point to look for spesific optimal designs. Pukelsheim (1993) has listed A-optimal designs, E-optimal designs and D-optimal designs for polynomial regression from the 1st to the 10th degree over $[-1,1]$. The theoretical developments leading to such computations are all explained in Pukelsheim (1993).

2. Multi-Factor First-Degree Polynomial Fit Models

Let us first look at an m-factor first-degree polynomial fit model

$$Y_{ij} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im} + e_{ij} \tag{2.1}$$

with m regressor variables, n experimental conditions $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})'$, $i = 1, 2, \ldots, n$; $j = 1,2,\ldots, N_i$ where the model has no constant term. In polynomial fit models of the previous section the experimental domain is $\mathcal{T} = [-1,1]$.

For the above m-way polynomial fit model (2.1) the experimental domain $\mathcal{T}$ is a subset of the m-dimensional Euclidean space $\mathbf{R}^m$. In this section we consider two extensions of the one-dimensional domain $\mathcal{T} = [-1,1]$: A Euclidean ball of radius $\sqrt{m}$ and a symmetric m-dimensional hypercube $\mathcal{T} = [-1,1]^m$ with half of sidelength 1. Later in this section we also study an m-way first-degree polynomial fit model with a constant term.

2.1 Designs in a Euclidean Ball

We assume now that the experimental domain for the model (2.1) is an m-dimensional Euclidean ball of radius $\sqrt{m}$, that is $\mathcal{T}_{\sqrt{m}} = \{\mathbf{x} \in \mathbf{R}^m : \ ||\mathbf{x}|| \leq \sqrt{m}\}$, where $\| \cdot \|$ denotes the Euclidean norm.

Denote $\mu_{jk} = \sum_{i=1}^{n} p_i x_{ij} x_{ik}$ for j, $k = 1, 2, \ldots, m$. Then the information matrix of an n-point design

$$d = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; \ p_1, p_2, \ldots, p_n\} \tag{2.2}$$

is of the form

$$\mathbf{I}(d) = \sum_{i=1}^{n} p_i \mathbf{x}_i \mathbf{x}_i' = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1} & \mu_{m2} & \cdots & \mu_{mm} \end{pmatrix}. \tag{2.3}$$

Consider an n-point design (2.2), $n \geq m$. Let

$$\lambda_1 \mathbf{w}_1 \mathbf{w}_1' + \lambda_2 \mathbf{w}_2 \mathbf{w}_2' + \cdots + \lambda_m \mathbf{w}_m \mathbf{w}_m'$$

be the spectral decomposition of $\mathbf{I}(d)$, where $\mathbf{w}_i$ and λ_i (> 0) are orthonormal eigenvectors and the eigenvalues of $\mathbf{I}(d)$, respectively. Note that

$$\begin{aligned} tr[\mathbf{I}(d)] &= \sum_{i=1}^{m+1} \lambda_i \\ &= \sum_{i=1}^{n} p_i \mathbf{x}_i' \mathbf{x}_i \ \leq m, \end{aligned}$$

since by assumption $\mathbf{x}_i \in \mathcal{T}_{\sqrt{m}}$ for all $i = 1, 2, \ldots, n$.

Denote $\widetilde{\mathbf{w}}_i = \sqrt{\sum_{i=1}^{n} \lambda_i} \mathbf{w}_i$ and $r_i = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}$, $i = 1, 2, \ldots, m$ and consider the m-point design

$$\widetilde{d} = \{\widetilde{\mathbf{w}}_1, \widetilde{\mathbf{w}}_2, \ldots, \widetilde{\mathbf{w}}_m; \ r_1, r_2, \ldots, r_m\}.$$

Clearly, $\widetilde{\mathbf{w}}_i \in \mathcal{T}_{\sqrt{m}}$, $i = 1, 2, \ldots, m$, and the designs $\widetilde{d}$ and d have the same information matrix, i.e. $\widetilde{d}$ and d are information equivalent designs. Thus for any n-point design d for the LSE of β in (2.1) there exists an information equivalent m-point design $\widetilde{d}$ from the regression range $\mathcal{T}_{\sqrt{m}}$ such that the support vectors are orthogonal. We say that $\widetilde{d}$ is an *orthogonal design*. This incidentally demonstrates validity of the DLG phenomenon in the present set-up as well.

Now we will prove that any design d for the LSE of $\beta = (\beta_1, \beta_2, \ldots, \beta_m)'$ in (2.1) can be *dominated* by a suitably defined orthogonal design as well. Hence an optimal design, if it exists, is necessarily an orthogonal design.

Theorem 1. Let d be an n-point design for the LSE of $\beta = (\beta_1, \beta_2, \ldots, \beta_m)'$ in (2.1), $n \geq m$. Then there exists an m-point orthogonal design $\widehat{d}$ that dominates d in the Loewner Order Domination sense.

Proof. Let any n-point design d with $n \geq m$ be given. Then there exists, as shown above, an information equivalent m-point design $\tilde{d} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m; r_1, r_2, \ldots, r_m\}$ with orthogonal support vectors $\mathbf{v}_i \in \mathcal{T}_{\sqrt{m}}$, $i = 1, 2, \ldots, m$. Then we can always define an

m-point design $\widehat{d} = \{\widehat{\mathbf{v}}_1, \widehat{\mathbf{v}}_2, \dots \widehat{\mathbf{v}}_m;\ r_1, r_2, \dots r_m\}$ with $\widehat{\mathbf{v}}_i = \sqrt{m}\mathbf{v}_i/\|\mathbf{v}_i\|$. Note that $\widehat{d}$ is an orthogonal design *on the ball* $\mathcal{T}_{\sqrt{m}}$.

Next, we prove that $\widehat{d}$ dominates d in the Loewner sense. Since $\|\mathbf{v}_i\|/\sqrt{m} \le 1$ for $i = 1, 2, \dots, m$, we have

$$\mathbf{I}(\widehat{d}) - \mathbf{I}(\tilde{d}) = \sum_{i=1}^{m} r_i \left(1 - \frac{\|\mathbf{v}_i\|^2}{m}\right) \widehat{\mathbf{v}}_i \widehat{\mathbf{v}}_i' \ge 0, \tag{2.4}$$

and consequently $\widehat{d} \succ \tilde{d}$. Since $\mathbf{I}(d) = \mathbf{I}(\tilde{d})$, we have $\mathbf{I}(\widehat{d}) \ge \mathbf{I}(d)$. This justifies the claim.

We consider now the m-factor first-degree polynomial fit model

$$Y_{ij} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + e_{ij}, \quad i = 1, 2, \dots, n;\ j = 1, 2, \dots, N_i \tag{2.5}$$

with a constant term β_0. The regression range of an n-point design d for LSE of β in (2.5) is of the form

$$\chi = \left\{ \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} \mid \mathbf{x} \in \mathcal{T}_{\sqrt{m}} \right\} \subset \mathcal{T}_{\sqrt{m+1}}. \tag{2.6}$$

The information matrix of an n-point design d in (2.2) is of the form

$$\mathbf{I}(d) = \sum_{i=1}^{n} p_i \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} (1, \mathbf{x}_i') = \begin{pmatrix} 1 & \mu_{01} & \dots & \mu_{0m} \\ \mu_{10} & \mu_{11} & \dots & \mu_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m0} & \mu_{m1} & \dots & \mu_{mm} \end{pmatrix}, \tag{2.7}$$

where, additionally, $\mu_{0k} = \sum_{i=1}^{n} p_i x_{ik}$, $k = 1, 2, \dots, m$.

Consider an n-point design d, $n \ge m+1$. Let

$$\lambda_1 \mathbf{w}_1 \mathbf{w}_1' + \lambda_2 \mathbf{w}_2 \mathbf{w}_2' + \cdots + \lambda_{m+1} \mathbf{w}_{m+1} \mathbf{w}_{m+1}' = \mathbf{I}(d)$$

be the spectral decomposition of $\mathbf{I}(d)$, where $\mathbf{w}_i$ and $\lambda_i > 0$ are orthonormal eigenvectors and the eigenvalues of $\mathbf{I}(d)$ respectively. Note that

$$\begin{aligned} tr[\mathbf{I}(d)] &= \sum_{i=1}^{m+1} \lambda_i \\ &= \sum_{i=1}^{n} p_i (1 + \mathbf{x}_i' \mathbf{x}_i) \le m+1, \end{aligned}$$

since by assumption $\mathbf{x}_i \in \mathcal{T}_{\sqrt{m}}$ for all $i = 1, 2, \dots, n$.

Denote $\tilde{\mathbf{w}}_i = \sqrt{\sum_{i=1}^{m+1} \lambda_i} \mathbf{w}_i$ and $r_i = \frac{\lambda_i}{\sum_{i=1}^{m+1} \lambda_i}$, $i = 1, 2, \dots, m+1$, and consider the $(m+1)$-point design

$$\tilde{d} = \{\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \dots, \tilde{\mathbf{w}}_{m+1};\ r_1, r_2, \dots, r_{m+1}\}.$$

Clearly, the support vectors $\tilde{\mathbf{w}} \in \mathcal{T}_{\sqrt{m+1}}$, $i = 1, 2, \dots, m+1$ are orthognal, and the designs $\tilde{d}$ and d have the same information matrix, i.e. $\tilde{d}$ and d are information equivalent

designs. Thus, for any n-point design d from the regression range (2.6) for the LSE of β in (2.5) there exists an information equivalent $(m+1)$-point design $\tilde{d}$ from the regression range $\mathcal{T}_{\sqrt{m+1}}$ such that the support vectors are orthogonal, i.e. $\tilde{d}$ is an *orthogonal design.*

If the vectors $\mathbf{x}_i \in \mathcal{T}_{\sqrt{m}}$, $i = 1, 2, \ldots, m+1$ fulfill the conditions

$$1 + \mathbf{x}_i'\mathbf{x}_i = 1 + m, \quad 1 + \mathbf{x}_i'\mathbf{x}_j = 0 \tag{2.8}$$

for all $1 \leq i \neq j \leq m+1$, then the vectors span a convex body in $\mathbf{R}^m$ called a *regular simplex* (cf. Pukelsheim (1993), p. 391). A design $d = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1};\ p_1, p_2, \ldots, p_{m+1}\}$ which places weights p_i, $i = 1, 2, \ldots, m+1$, on the vertices of a regular simplex in $\mathbf{R}^m$ is called a *simplex design.* A design with equal weights $p_1 = p_2 = \ldots = p_{m+1} = \frac{1}{m+1}$ is called a *uniform simplex design.* If the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1}$ satisfy the conditions (2.8), then the vectors $(1, \mathbf{x}_1')$, $(1, \mathbf{x}_2')$, $\ldots$, $(1, \mathbf{x}_{m+1}')$ are orthogonal, they belong to the boundary of $\mathcal{T}_{m+1}$ and are of the form (2.6). Given an orthogonal design on a Euclidean ball (the support vectors belong to the boundary), then any other orthogonal design can be obtained from it by orthogonal rotation of support vectors.

It is obvious that we can always find $m+1$ orthogonal vectors $(1, \mathbf{x}_i')'$, $i = 1, 2, \ldots, m+1$ such that every $\mathbf{x}_i$ belongs to the boundary of $\mathcal{T}_{\sqrt{m}}$. For $m = 1$ the support points are $x_1 = 1$ and $x_2 = -1$ so that $(1, 1)'$ and $(1, -1)'$ satisfy the conditions (2.8). The support points $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ satisfying the conditions (2.8) belong to the boundary of $\mathcal{T}_{\sqrt{2}}$ and they span an *equilateral triangle*equilateral triangle on the sphere $\mathcal{T}_{\sqrt{2}}$. For example the support points $(1, 1)'$, $-\frac{1}{2}(1+\sqrt{3}, 1-\sqrt{3})'$, $-\frac{1}{2}(1-\sqrt{3}, 1+\sqrt{3})'$, and every rotation of them span an equilateral triangle.

We consider now design optimality criteria ϕ which are isotonic with respect to the Loewner ordering. We prove that any design d for the LSE of β in (2.5) can be dominated by an $(m+1)$-point simplex design. Hence an optimal design with respect to ϕoptimality criterion, if it exists, can be found among the $(m+1)$-point simplex designs.

Theorem 2. Let d be an n-point design for the LSE of β in (2.5) over the ball $\mathcal{T}_{\sqrt{m+1}}$, $n \geq m+1$ and let ϕ be any optimality criterion that is (1) isotonic with respect to the Loewner ordering and (2) depends on the information matrix only through its eigenvalues. Then there exists an $(m+1)$-point simplex design d^* that dominates d with respect to ϕ, i.e. $\phi(\mathbf{I}(d^*)) \geq \phi(\mathbf{I}(d))$.

Proof. Let any n-point design d with $n \geq m+1$ be given. Then there exists as shown above, an information equivalent $(m+1)$-point design $\tilde{d} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{m+1};\ r_1, r_2, \ldots, r_{m+1}\}$ with orthogonal support vectors $\mathbf{w}_i \in \mathcal{T}_{\sqrt{m+1}}$, $i = 1, 2, \ldots, m+1$. Then we can always define such an $(m+1)$-point design $d_u = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m+1};\ r_1, r_2, \ldots, r_{m+1}\}$ that $\mathbf{u}_i = \sqrt{m+1}\mathbf{w}_i/\|\mathbf{w}_i\|$. Now d_u is an orthogonal design on the ball $\mathcal{T}_{\sqrt{m+1}}$. We note that the range of d_u is *not* of the form (2.6).

We have noted earlier that we can find $m+1$ orthogonal vectors $(1, \mathbf{x}_i')'$, $i = 1, 2, \ldots, m+1$ such that every $\mathbf{x}_i$ belongs to the boundary of $\mathcal{T}_{\sqrt{m}}$. Let $\widehat{\mathbf{u}}_i = (1, \mathbf{x}_i')'$ for every i. It then follows that

$$\widehat{\mathbf{u}}_i'\widehat{\mathbf{u}}_i = 1 + \widehat{\mathbf{x}}_i'\widehat{\mathbf{x}}_i = 1 + m, \quad \widehat{\mathbf{u}}_i'\widehat{\mathbf{u}}_j = 1 + \widehat{\mathbf{x}}_i'\widehat{\mathbf{x}}_j = 0$$

Thus $d^* = \{\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2, \ldots, \widehat{\mathbf{u}}_{m+1};\ r_1, r_2, \ldots, r_{m+1}\}$ is an orthogonal design on the ball $\mathcal{T}_{\sqrt{m+1}}$. Since any other orthogonal design on the ball can be obtained by an orthogonal rotation of d^*, we can find an $(m+1) \times (m+1)$ orthogonal matrix $\mathbf{P}$ such that $\widehat{\mathbf{u}}_i = \mathbf{P}\mathbf{u}_i$

for each i. Clearly, the range of d^* is of the form given by (2.6). We note that d^* is a regular simplex design.

Next we prove that d^* dominates d with respect to criterion ϕ. The argument used to prove (2.4) shows that $\mathbf{I}(d_u) \geq \mathbf{I}(\widetilde{d}) = \mathbf{I}(d)$. Further, $\mathbf{I}(d_u) = \mathbf{P}'\mathbf{I}(d^*)\mathbf{P}$ and hence, $\mathbf{I}(d_u)$ and $\mathbf{I}(d^*)$ have the same eigenvalues. Thus by the assumption (2), $\phi(\mathbf{I}(d^*)) = \phi(\mathbf{I}(d_u))$. It is now clear that d^* dominates d with respect to ϕ. We note that the range of d^* is of the form (2.6) and hence d^* is a valid design for the model given by (2.5). Hence the proof is complete.

Let $\mathbf{X}$ denote a model matrix whose rows are the support vectors $(1 \;\; \mathbf{x}_1')'$, $(1 \;\; \mathbf{x}_2')'$, $\ldots$, $(1 \;\; \mathbf{x}_{m+1}')'$ of a simplex design $d^{(m+1)}$. For such a design

$$\mathbf{I}(d^{(m+1)}) = \mathbf{X}'\mathbf{D}\mathbf{X}, \quad \mathbf{D} = Diag(p_1, p_2, \ldots, p_{m+1})$$

and the model matrix $\mathbf{X}$ is square. The non-zero eigenvalues of $\mathbf{X}'\mathbf{D}\mathbf{X}$ and $\mathbf{D}\mathbf{X}\mathbf{X}'$ are the same and

$$\mathbf{D}\mathbf{X}\mathbf{X}' = (m+1)\mathbf{D} = (m+1)\; Diag(p_1, p_2, \ldots, p_{m+1}).$$

Thus optimum designs with respect to matrix mean criteria are easy to determine. A design with equal weights $p_1 = p_2 = \cdots = p_{m+1} = 1/(m+1)$, called a *uniform simplex design*, is A-, D- and E-optimal design.

3. Designs in a Unit Hypercube

A symmetric m-dimensional unit-cube $[-1,1]^m$ is a natural extension of $[-1,1]$. Note that $[-1,1]^m$ is the convex hull of its extreme points, the 2^m vertices of $[-1,1]^m$. It is known that in order to find optimal support points, we need to search the extreme points of the regression range χ only. If the support of a design contains other than extreme points, then it can be Loewner dominated by a design with extreme support points only. This result was basically presented by Elfving (1952, 1959). A unified general theory is given by Pukelsheim ((1993); Chapter 8).

As an example consider a 2-factor first degree model (2.1) that has *no* constant term. The experimental domain $\mathcal{T}$ is the square $[-1,1]^2$. The extreme points (vertices) of $[-1,1]^2$ are

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

Suppose now that the support of a design d consists of the extreme points only. Then the information matrix of d takes finally the form

$$\mathbf{I}(d) = p \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + (1-p) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2p-1 \\ 2p-1 & 1 \end{pmatrix} \tag{3.1}$$

with $0 < p < 1$, since $\begin{pmatrix} 1 \\ 1 \end{pmatrix} (1 \;\; 1) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} (-1 \;\; -1)$, $\begin{pmatrix} 1 \\ -1 \end{pmatrix} (1 \;\; -1) = \begin{pmatrix} -1 \\ 1 \end{pmatrix} (-1 \;\; 1)$. The eigenvalues of $\mathbf{I}(d)$ are $\lambda_{1,2} = 1 \pm (2p-1)$. Thus the optimum values of various matrix means criteria ϕ_t are easy to determine. The 2-point design

$$d^{(2)}_{\frac{1}{2}} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}; \; \frac{1}{2} \right\} \tag{3.2}$$

is A-, D- and E-optimal.

Note that any extreme point design with 2, 3 or 4 support points such that the total weight at the points $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}-1\\-1\end{pmatrix}$ is equal to p yields the same information matrix (2.9), i.e. they constitute a class of information equivalent designs. The information equivalent 2-point designs with the support $\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\right\}$, $\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix}\right\}$, $\left\{\begin{pmatrix}-1\\-1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\right\}$ and $\left\{\begin{pmatrix}-1\\-1\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix}\right\}$ have the minimal support size. For example, a 3-point design $\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix};\ \frac{1}{2},p_2,p_3\right\}$ with $p_2+p_3=\frac{1}{2}$ is information equivalent to (2.10), and hence also it is A-, D- and E-optimal.

Now consider the model (2.5) with $m=2$. The information matrix of d defined below for the LSE of β in (2.5) is

$$\mathbf{I}(d)=\begin{pmatrix}1 & p_1+p_2-p_3-p_4 & p_1-p_2-p_3+p_4\\ p_1+p_2-p_3-p_4 & 1 & p_1+p_2-p_3-p_4\\ p_1-p_2-p_3+p_4 & p_1+p_2-p_3-p_4 & 1\end{pmatrix} \tag{3.3}$$

when the support of d is $\left\{\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}1\\1\\-1\end{pmatrix},\begin{pmatrix}1\\-1\\-1\end{pmatrix},\begin{pmatrix}1\\-1\\1\end{pmatrix}\right\}$ and $p_i>0$, $i=1,\ 2,\ 3,\ 4$ with $p_1+p_2+p_3+p_4=1$ are the corresponding weights. It may be noted that it is enough to concentrate on the above extreme points (in the suport of d) even for a model *with* the intercept term.

Let $\lambda_1,\ \lambda_2,\ \lambda_3$ denote the eigenvalues of the information matrix (2.11). Then by Hadamard's inequality (Horn and Johnson (1985), p. 477)

$$|\mathbf{I}(d)|=\lambda_1\lambda_1\lambda_2\le 1. \tag{3.4}$$

Equality holds in (2.12) if and only if $\mathbf{I}(d)$ is equal to $\mathbf{I}_3$, and $\mathbf{I}(d)=\mathbf{I}_3$ exactly when $p_1=p_2=p_3=p_4=\frac{1}{4}$. Consequently, the design that assigns uniform weights $\frac{1}{4}$ to each of these four extreme points of the regression range is the D-optimal design. Since $\lambda_1+\lambda_2+\lambda_3=1$, the design is also A- and E- optimal. There is no 3-point design $d^{(3)}=\{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3;\ p_1,p_2,p_3\}$ such that $\mathbf{I}(d^{(3)})=\mathbf{I}_3$. This is easy to show if the support is chosen to be any 3-point subset of the vertices of the regression range χ. Therefore a 3-point design cannot be D-, A- and E- optimal at the same time !!

We now turn to the general case for arbitrary but fixed $m(>2)$. Note that the *complete factorial design which assigns equal weight* 2^{-m} *to each of the* 2^m *extreme points of the form* $\{\pm1,\ \pm1,\ldots,\pm1\}$ *provides an information matrix exactly equal to* $\mathbf{I}_{m+1}$.

Liski et al (2002) have illustrated optimality of such designs.

Remark. Though the complete factorial design assigning uniform weight $1/2^m$ *to each of the* $n=2^m$ *vertices of the* m *dimensional cube* $[-1,1]^m$ *is optimal, its support size* 2^m *grows very quickly when the dimsension* m *increases.*

For certain values of m there exists a design $d^{(m+1)} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m+1};\ p_1, p_2, \ldots p_{m+1}\}$ with the minimum support size $m+1$ for which $\mathbf{I}(d^{m+1}) = \mathbf{I}_{m+1}$. An $(m+1) \times (m+1)$ matrix $\mathbf{X}$ with entries 1 and -1 is called a *Hadamard matrix* (denoted by $\mathbf{H}_{m+1}$) if $\mathbf{X}'\mathbf{X} = (m+1)\mathbf{I}_{m+1}$. Then the model matrix $\mathbf{X}$ is square and hence $|\mathbf{X}'\mathbf{D}\mathbf{X}| = |\mathbf{D}\mathbf{X}\mathbf{X}'|$. Now by Hadamard's inequality

$$\left|\mathbf{I}(d^{(m+1)})\right| = \lambda_1\lambda_2\cdots\lambda_{m+1} \leq \prod_{i=1}^{m+1} p_i(1+\mathbf{x}_i'\mathbf{x}_i) \leq 1. \tag{3.5}$$

Equality holds if and only if the matrix $\mathbf{X}\mathbf{X}'$ is diagonal, that is if $1+\mathbf{x}_i'\mathbf{x}_j = 0$ for all $i \neq j \leq m+1$, and $p_1 = p_2 = \cdots = p_{m+1} = \frac{1}{m+1}$.

If the model matrix $\mathbf{X}$ of an $(m+1)$-point design $d^{(m+1)}$ is an $(m+1) \times (m+1)$ Hadamard matrix $\mathbf{H}_{m+1}$, then $\mathbf{I}(d^{(m+1)}) = \mathbf{I}_{m+1}$. Note that the design $d^{(m+1)}$ with $\mathbf{I}(d^{(m+1)}) = \mathbf{I}_{m+1}$ is a uniform simplex design which assigns weight $1/(m+1)$ to each of the vertices $\mathbf{x}_1$, $\mathbf{x}_2$, $\ldots$, $\mathbf{x}_{m+1}$ of a regular simplex. The support points $\mathbf{x}_i \in \{\pm 1\}^m$, $i = 1, 2, \ldots, m+1$ are also vertices of the m dimensional cube $[-1,1]^m \subset \mathcal{T}_{\sqrt{m}}$. It is known that if a Hadamard matrix of order k exists then $k = 2$ or $k = 4q$ for some positive integer q. Although it has not yet been shown that Hadamard matrices of order $4q$ exist for all $q \geq 1$, many infinite families of Hadamard matrices have been constructed. These include all values of q which are of practical interest. A useful reference is Hedayat and Wallis (1978).

In conclusion, we note that optimal designs with both the support sizes $m+1$ and 2^m are available whenever $\mathbf{H}_{m+1}$ exists. It would be interesting to examine what other support sizes also provide optimal desgns.

In this context, Liski et al (220@0 have discussed an important result due to Caratheodory and examined its performance vis-a-vis Hadamard and factorial bounds.

References

de la Garza, A. (1954). Spacing of information in polynomial regression. *Annals of Mathematical Statistics* **25**, 123–130.

Elfving, G. (1952). Optimum allocation in linear regression theory. *Annals of Mathematical Statistics* **23**, 255–262.

Elfving, G. (1959). Design of linear experiments. *Probability and statistics. The Harald Cramér volume* (ed. by Ulf Grenander), 58–74. Wiley, New York.

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

Hedayat, A. S. and Wallis, W. D. (1978). Hadamard matrices and their applications. *Annals of Statistics* **6**, 1184–1238.

Horn, R. A. and Johnson, C. R. (1985). *Matrix analysis.* Cambridge: Cambridge University Press.

Kiefer, J. C. (1959). Optimum experimental designs (with discussion). *Journal of the Royal Statistical Socciety* Series B **21**, 272–319.

Liski, E. P. and Zaigraev, A. (2001). A stochastic cracterization of Loewner optimality design criterion in linear models. Appears in *Metrika.*

Liski, E. P. Mandal, N. K. Shah, K. R. and Sinha, B. K. (2002). Topics in Optimal Design. Springer-Verlag Lecture Notes in Statistics, Vol. 163.

Pukelsheim, F. (1993). *Optimal design of experiments.* Wiley, New York.

Shah, K. R. and Sinha, Bikas K. (1989). Theory of optimal designs. *Lecture Notes in Statistics* **54**, Springer, New York.

Silvey, S. D. (1980). *Optimum design.* Chapman & Hall, London.