COMPLETE CLASS RESULTS
FOR LINEAR REGRESSION OVER THE MULTI-DIMENSIONAL CUBE

Friedrich PUKELSHEIM, Cornell University*

BU-942-M† September 25, 1987

ABSTRACT

Complete classes of designs and of moment matrices for linear regression over the multi-dimensional unit cube are presented. An essentially complete class of designs comprises the uniform distributions on the vertices with a fixed number of entries being equal to unity, and mixtures of neighboring such designs. The corresponding class of moment matrices is minimally complete. The derivation is built on information increasing orderings, that is, a superposition of the majorization ordering generated by the permutation group, and the Loewner ordering of symmetric matrices.

* On leave from the Institut für Mathematik der Universität Augsburg. The work of this author is partially supported by the Mathematical Sciences Institute, Cornell University.

† Paper BU-942-M in the Biometrics Unit, Cornell University.

AMS 1980 subject classification. 62K05, 62C07.

Key words and phrases. Group majorization, permutation group, completely symmetric matrices, Loewner matrix ordering, p-mean optimality.

1. Introduction

In a brilliant paper C.-S. Cheng (1986) recently determined optimal designs over the k-dimensional unit cube $[0,1]^k$ for the linear model

$$\mathrm{E}[Y] = x'\theta, \qquad \mathrm{V}[Y] = \sigma^2.$$

In this setting the experimenter chooses the regression vector x in the cube $[0,1]^k$ prior to running the experiment, and then observes the response Y. The response is assumed to have expected value and variance as given above, furthermore repeated responses are taken to be uncorrelated. As pointed out by Cheng this model has interesting applications in Hadamard transform optics.

The optimal designs of Cheng (1986) are the j-vertex designs ξ_j and mixtures of $j+1$- and j-vertex designs, defined as follows. A *j-vertex* of the unit cube $[0,1]^k$ is a vector x with j entries equal to unity and the remaining $k-j$ entries equal to zero, for $j = 0, \ldots, k$. There are $\binom{k}{j}$ many j-vertices. The *j-vertex design* ξ_j is the design that has the j-vertices for support, and assigns uniform mass $1/\binom{k}{j}$ to each of them. For mixtures of the form $\alpha\xi_{j+1} + (1-\alpha)\xi_j$ the following notation is convenient, in that it provides a continuous parametrization in the support defining parameter $s \in [0,k]$. Given $j = 0, \ldots, k-1$ define the design

$$\xi_s = (s-j)\xi_{j+1} + \big(1-(s-j)\big)\xi_j \qquad \text{for } s \in (j, j+1).$$

In terms of s we have that j is the integer part of s, $j = \operatorname{int} s$. In other words, the two integers $j+1$ and j closest to s specify the vertices supporting ξ_s, and the fractional part $s-j$ determines the weight for mixing ξ_{j+1} and ξ_j. For example, $\xi_{2.11} = 0.11\,\xi_3 + 0.89\,\xi_2$, and $\xi_{7.4} = 0.4\,\xi_8 + 0.6\,\xi_7$; see also Figure 1.

Under the p-mean criteria considered by Cheng (1986) the class of optimal designs is

$$\mathcal{C} = \{\, \xi_s : s \in [\operatorname{int}\frac{k+1}{2}, k] \,\},$$

starting from the 'median vertex design' $\xi_{\operatorname{int}(k+1)/2}$ and running through the j-vertex designs ξ_j and mixtures ξ_s up to the design ξ_k that assigns all mass to the vector with every entry equal to unity. It is notationally convenient to set

$$m = \operatorname{int}\frac{k+1}{2};$$

note that m is the largest median of the set of numbers $0, \ldots, k$. As usual the class of all designs on $[0,1]^k$ is denoted by Ξ.

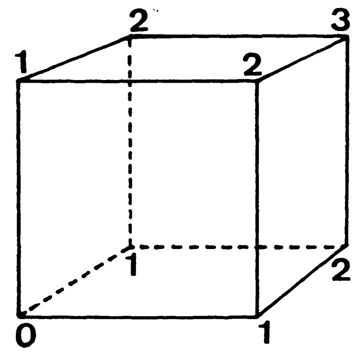

Figure 1. Corners of the unit cube with j entries 1 and the remaining entries 0 are called j-vertices. For the cube in dimension 3 the figure shows 0-, 1-, 2-, and 3-vertices.

In Section 2 we show that Cheng's class $\mathcal{C}$ is essentially complete, and that the corresponding class of moment matrices $M(\mathcal{C})$ is minimally complete, with respect to the information ordering generated by the permutation group Perm(k). As a consequence the class $\mathcal{C}$ contains an optimal design whenever the optimality criterion is given by an information function ϕ that is permutationally invariant. Cheng (1986) studied the subclass of p-means ϕ_p. In Section 3 we present some graphs showing how the optimal support parameter $s(p)$ and the optimal value $v(p)$ change with the order $p \in [-\infty, 1]$ of the mean ϕ_p, and with the dimensionality k.

2. Complete class results

The performance of a design ξ hinges on its moment matrix $M(\xi) = \int_{[0,1]^k} xx'\, d\xi$. These matrices are of order $k \times k$, and nonnegative definite. Our complete class results refer to the information increasing ordering generated by the group $\mathrm{Perm}(k)$ of $k \times k$ permutation matrices. The general theory is surveyed in Pukelsheim (1987), we here only recall such details as are necessary for the present discussion. A matrix B is said to be *more centered than* a moment matrix A whenever

$$B \in \mathrm{conv}\{\, QAQ' : Q \in \mathrm{Perm}(k)\,\},$$

that is, B lies in the convex hull of the orbit of A when the group $\mathrm{Perm}(k)$ acts through congruence. A moment matrix M is said to be *at least as informative as* another moment matrix A when in the Loewner ordering one has $M \geq B$ for some matrix B that is more centered than A. A moment matrix is said to be *more informative than* another moment matrix A when M is at least as informative as A, but does not lie in the orbit of A.

Theorem 1. *The class of designs $\mathcal{C}$ is essentially complete, that is, for all designs η in Ξ there exists a design ξ_s in $\mathcal{C}$ such that $M(\xi_s)$ is at least as informative as $M(\eta)$. The corresponding class of moment matrices $M(\mathcal{C})$ is minimally complete, that is, for all moment matrices A not in $M(\mathcal{C})$ there exists a moment matrix M in $M(\mathcal{C})$ such that M is more informative than A and there is no proper subclass of $M(\mathcal{C})$ with the same property.*

Proof. Let η be a design not in $\mathcal{C}$. First symmetrization leads to an invariant design $\overline{\eta}$; then a Loewner improvement produces a better design ξ, and another Loewner improvement yields a design ξ_s in the class $\mathcal{C}$.

I. Averaging η leads to a design $\overline{\eta}$ that is permutationally invariant. Its moment matrix matrix $\overline{A}$ is the average of the moment matrix A of η, $\overline{A} = \sum_{Q\in\mathrm{Perm}(k)} QAQ'$, and therefore more centered than A.

But it may happen that η has an invariant moment matrix A without η itself being invariant. In this case $\overline{A} = A$, so that the passage from A to $\overline{A}$ means no improvement whatsoever. The optimal balanced incomplete block designs of Corollary 3.5 in Cheng (1986) provide an instance of this.

II. Being invariant the design $\overline{\eta}$ must be a mixture of j-vertex designs for $j \geq 0$,

$$\overline{\eta} = \sum\nolimits_{j\geq 0} \beta_j \xi_j$$

with $\min \beta_j \geq 0$ and $\sum \beta_j = 1$. Let $\overline{J}$ be the $k \times k$ matrix with every entry equal to $1/k$, and set $K = I_k - \overline{J}$; this is an orthogonal pair of orthogonal projection matrices. The moment matrix of ξ_j is

$$M(\xi_j) = \Lambda_j \overline{J} + \lambda_j K, \qquad \text{where } \Lambda_j = \frac{j^2}{k}, \quad \lambda_j = \frac{j(k-j)}{k(k-1)}.$$

Therefore the moment matrix of $\overline{\eta}$ is

$$M(\overline{\eta}) = \sum\nolimits_{j\geq 0} \beta_j(\Lambda_j \overline{J} + \lambda_j K).$$

The eigenvalues Λ_j and λ_j increase as j runs over the initial section from 0 up to m. Hence we introduce new weights α_j that sweep the initial mass into the median m,

$$\begin{aligned} \alpha_j &= 0 \qquad && \text{for all } j < m, \\ \alpha_m &= \sum\nolimits_{j\leq m} \beta_j, && \\ \alpha_j &= \beta_j && \text{for all } j > m. \end{aligned}$$

This produces a design which is a mixture of j-vertex designs for $j \geq m$,

$$\xi = \sum\nolimits_{j\geq m} \alpha_j \xi_j,$$

with a Loewner improved moment matrix

$$M(\xi) \geq M(\overline{\eta}).$$

Furthermore the two moment matrices are distinct, unless the weights β_j vanish for $j < m$.

III. The moment matrix of ξ is $M(\xi) = \Lambda \overline{J} + \lambda K$, with

$$\Lambda = \sum\nolimits_{j\geq m} \alpha_j k \left(\frac{j}{k}\right)^2, \qquad \lambda = \sum\nolimits_{j\geq m} \alpha_j \frac{k}{k-1} \frac{j}{k}\left(1 - \frac{j}{k}\right).$$

Thus the eigenvalue pair (Λ, λ) varies over the convex set

$$\operatorname{conv}\{ (\Lambda_j, \lambda_j) : j = m, \ldots, k \} = \operatorname{conv}\left\{ \left(kz^2, \frac{k}{k-1} z(1-z)\right) : z = \frac{m}{k}, \ldots, 1 \right\}.$$

In other words, on the curve $x(z) = kz^2$ and $y(z) = \frac{k}{k-1} z(1-z)$ we pick the points (Λ_j, λ_j) corresponding to $z = j/k$ for $j \geq m$, and then form their convex hull. Figure 2 shows the limiting continuous curve $\big(x(z), y(z)\big)$ with $z \in [1/2, 1]$.

The geometry exhibits that for every eigenvalue pair (Λ, λ) there exist 'neighboring' points $(\Lambda_{j+1}, \lambda_{j+1})$ and (Λ_j, λ_j) on the curve such that with some $\alpha \in [0, 1]$ we obtain, with $s = j + \alpha$,

$$\Lambda \leq \alpha \Lambda_{j+1} + (1-\alpha)\Lambda_j = \Lambda_s, \qquad \lambda \leq \alpha \lambda_{j+1} + (1-\alpha)\lambda_j = \lambda_s.$$

Thus the design ξ_s has a Loewner improved moment matrix,

$$M(\xi_s) \geq M(\xi),$$

and lies in Cheng's class $\mathcal{C}$. Furthermore the two moment matrices are distinct, unless ξ itself lies in $\mathcal{C}$.

IV. As s varies over $[m, k]$ the eigenvalues Λ_s and λ_s strictly increase and decrease, respectively. Therefore a proper subclass of $M(\mathcal{C})$ cannot be complete. $\diamond$

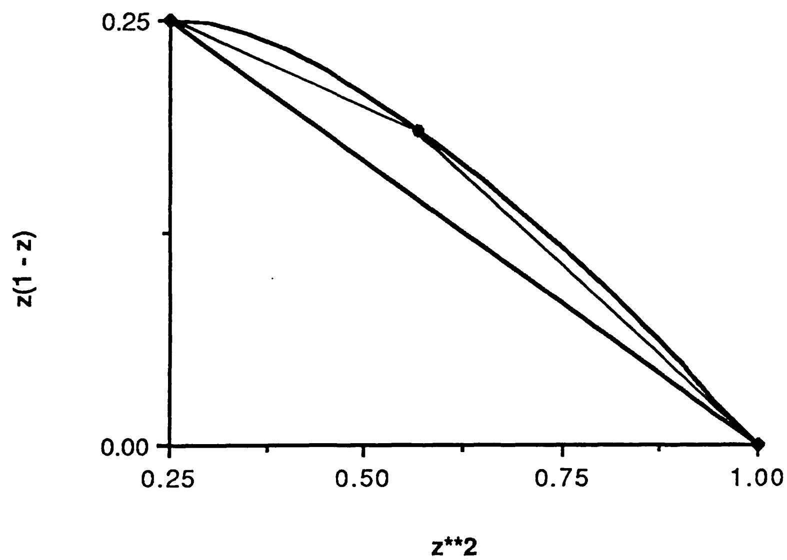

Figure 2. The figure illustrates the set of eigenvalue pairs (Λ, λ) of the designs ξ that are mixtures of j-vertex designs for $j \geq m$. The j-vertex designs have eigenvalue pairs (Λ_j, λ_j) that are the corners of the convex set shown in the picture. Dots indicate the pairs for $j = 2, 3, 4$ in dimension $k = 4$. See part III of the proof of Theorem 1.

The eigenvalue improvement in part III of the proof appears to be small, indicating that mixtures of j-vertex designs for $j \geq m$ may perform well even when they are not in the class $\mathcal{C}$.

Every optimality criterion ϕ that is isotonic relative to the Loewner ordering, concave, and permutaionally invariant is invariant also relative to the information increasing ordering: M is at least as informative as A if and only if

$$M \geq \sum \alpha_i Q_i A Q_i',$$

with $Q_i \in \text{Perm}(k)$, and $\min \alpha_i \geq 0$ and $\sum \alpha_i = 1$. The functional properties of ϕ then yield

$$\phi(M) \geq \phi(\sum \alpha_i Q_i A Q_i') \geq \sum \alpha_i \phi(Q_i A Q_i') = \phi(A).$$

The same reasoning also establishes that if there exists a design $\xi \in \Xi$ that is ϕ-optimal over Ξ then there actually exists a design $\xi_s \in \mathcal{C}$ with the same optimality property. An optimal design always exists provided the criterion ϕ is upper semicontinuous. The following corollary summarizes this behaviour.

Corollary 1.1. *Let ϕ be an optimality criterion that is Loewner-isotonic, concave, and permutationally invariant. If a moment matrix M is at least as informative as another moment matrix A then*

$$\phi(M) \geq \phi(A).$$

Moreover, there exists a design $\xi_s \in \mathcal{C}$ which is ϕ-optimal over Ξ,

$$\phi\big(M(\xi_s)\big) = \max_{\xi \in \Xi} \phi\big(M(\xi)\big),$$

provided ϕ is upper semicontinuous. ◇

A particular class of criteria to which this corollary applies are the p-means ϕ_p, for $p \in [-\infty, 1]$, studied by Cheng (1986).

3. Optimal designs for the p-mean criteria

As it happens the complete class of moment matrices $M(\mathcal{C}$ is in fact exhausted by the moment matrices $M(\xi_{s(p)})$ belonging to ϕ_p-optimal designs $\xi_{s(p)}$, as p varies over $[-\infty, 1]$. This follows from Theorem 3.1 in Cheng (1986); we now briefly recall this result. Cheng subdivides the interval $[-\infty, 1]$ using two interlacing sequences of numbers $f(j)$ and $g(j)$, $j = m, \ldots, k$, according to

$$-\infty = f(m) < g(m) < f(j) < g(j) < f(j+1) < g(j+1) < f(k) = g(k) = 1$$

for $j = m+1, \ldots, k-2$. His result can then be stated as follows.

Theorem 2. *For every order $p \in [-\infty, 1]$ there exists a support parameter $s(p) \in [m, k]$ such that the design $\xi_{s(p)}$ is ϕ_p-optimal over Ξ. As a function $s(p)$ is continuous, being equal to j on the closed intervals $[f(j), g(j)]$ and strictly increasing from j to $j+1$ on the open intervals $\big(g(j), f(j+1)\big)$.* ◊

Cheng (1986) actually provides explicit formulae for these quantities, namely

$$f(j) = 1 + \frac{\log\left(1 - \frac{k}{2j-1}\right)}{\log \frac{(k-1)j}{k-j}}, \qquad g(j) = 1 + \frac{\log\left(1 - \frac{k}{2j+1}\right)}{\log \frac{(k-1)j}{k-j}};$$

$$s(p) = \frac{j(j+1)\left(k - 1 + \left\{1 - \frac{k}{2j+1}\right\}^{\frac{1}{p-1}}\right)}{(2j+1)(k-1) + (2j+1-k)\left\{1 - \frac{k}{2j+1}\right\}^{\frac{1}{p-1}}} \qquad \text{for } p \in \big(g(j), f(j+1)\big).$$

Figures 3 and 4 illustrate how the standardized support parameter $s(p)/k$ and the optimal value $\phi(\xi_{s(p)})$ vary with p. Variation is small for $p < -1$ and is not shown, variation is relatively large for $p > 0$.

Writing $s_k(p)$ in place of $s(p)$ we show that for large dimensions k the support of the optimal designs tends to the the vertices with half of their entries unity,

$$\lim_{k\to\infty} \frac{s_k(p)}{k} = \frac{1}{2}.$$

Let j_k be the integer part of $s_k(p)$, so that $s_k(p) \in [j_k, j_k+1)$, and $p \in \big[f(j_k), f(j_k+1)\big)$. Hence it suffices to show that j_k/k tends to $1/2$. From $j_k > m$ we clearly get $\liminf j_k/k \geq 1/2$. We show that $\limsup j_k/k > 1/2$ is impossible. There is no loss in generality in assuming $\lim j_k/k = \alpha > 1/2$. Then we obtain

$$f(j_k) = 1 + \frac{\log\left(1 - \frac{1}{2j_k/k - 1/k}\right)}{\log(k-1)\frac{j_k/k}{1-j_k/k}} \quad \to \quad 1 + \frac{\log\left(1 - \frac{1}{2\alpha}\right)}{\lim_k \log(k-1)\frac{\alpha}{1-\alpha}} = 1.$$

Hence eventually p falls below $f(j_k)$, and this is impossible.

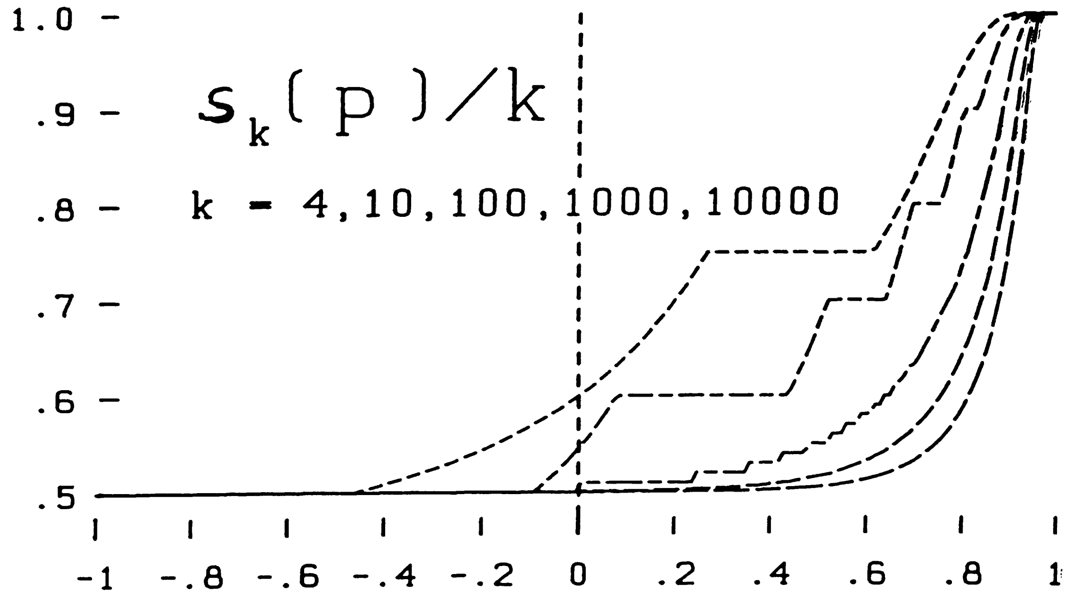

Figure 3. The graph shows the support parameter $s_k(p)/k$, standardized by the dimension k, of the ϕ_p-optimal design $\xi_{s_k(p)}$, as a function of the order p of the mean ϕ_p. Most of the variation takes place when p is positive. The limiting value for large dimensions k is 1/2.

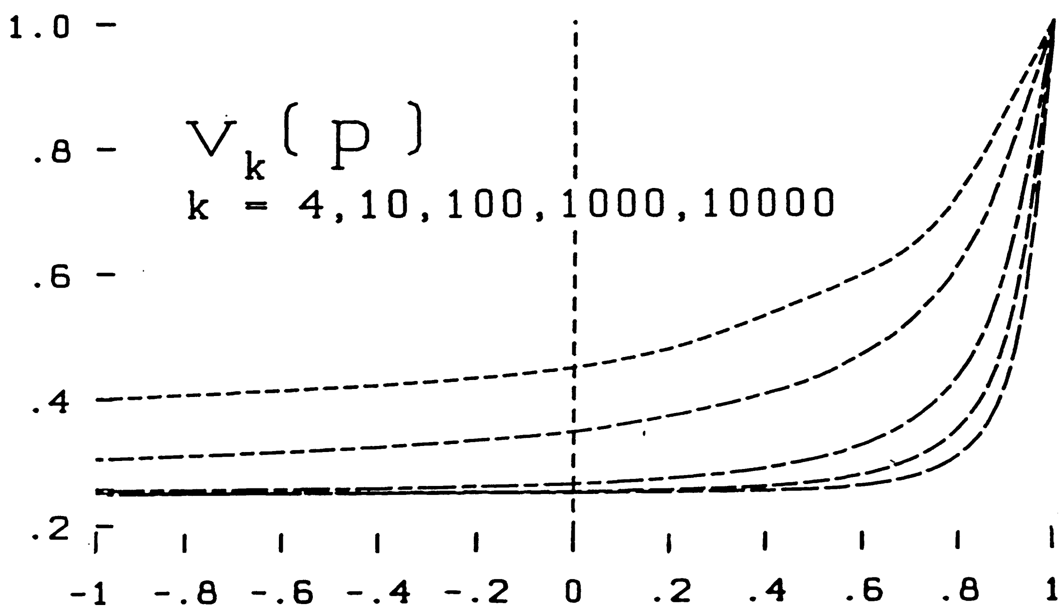

Figure 4. The graph shows the dependance of the optimal value $v(p) = \phi(\xi_{s(p)})$ on the order p of the mean ϕ_p, for varying dimensions k.

References

Cheng, Ching-Shui (1986). An application of the Kiefer-Wolfowitz equivalence theorem to a problem in Hadamard Transform Optics. Technical Report No. 63 May 1986 (revised January 1987), Department of Statistics, University of California, Berkeley, California. 22 pp.

Pukelsheim, Friedrich (1987). Information increasing orderings in experimental design theory. *International Statistical Review* **55** 203–219.