The Annals of Statistics
1982, Vol 10, No. 2, 511–525

OPTIMAL ROBUST DESIGNS: LINEAR REGRESSION IN R^k [1]

BY L. PESOTCHINSKY

University of California, Santa Barbara

The model $E(y \mid x) = \theta_0 + \sum_{i=1}^k \theta_i x_i + \psi(\mathbf{x})$ is considered, where $\psi(\mathbf{x})$ is an unknown contamination with $|\psi(\mathbf{x})|$ bounded by given $\varphi(\mathbf{x})$. Optimal designs are studied in terms of least squares estimation and a family of minimax criteria. In particular, analogs of D-, A- and E-optimal designs are studied in the general case of an arbitrary k. Some commonly used integer designs are considered and their efficiencies with respect to optimal designs are determined. In particular, it is shown that star-point designs or regular replicas of 2^k factorials are very efficient under the appropriate choice of levels of factors.

1. Introduction. Consider the regression model given by

$$Y(\mathbf{x}_i) = \sum_{j=0}^k \theta_j f_j(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n,$$

where the errors, $\{\varepsilon_i\}$, are uncorrelated random variables with mean 0 and variance σ^2, $\mathbf{x}_i \in \mathscr{X} \subset R^P$ and functions f_j are linearly independent. The coefficients, $\{\theta_j\}$, are unknown and the regression problem is to make inference about $\{\theta_j\}$ in some "optimal" way. In particular, an optimal estimator of $\{\theta_j\}$ has to be chosen and in connection with this estimator the design problem is to choose the experimental points, $\mathbf{x}_i$'s in an optimal way.

If we choose, say, the least squares method of estimation, then a variety of optimality criteria could be considered in the associated design problem, and the most reasonable of them depend on functionals of $M(\xi)$, where $M(\xi)$ is the information matrix of the design ξ, defined in full below. An impressive list of papers by Kiefer and others refer to the problem of choice of the criterion and to the construction and study of various properties of optimal designs. The most relevant aspect of these studies for us is the robustness of the optimal designs under variations of criterion; see Box and Draper (1959) and Kiefer (1975). In particular, it was shown (Galil and Kiefer, 1977a, b; Pesotchinsky, 1978) that for the same regression equation the performance of optimal designs usually depends on choice of experimental region rather than on choice of criterion.

Unfortunately, as was noticed by Box and Draper (1959), the strict formulation of the regression function becomes dangerous in the situations when the "true" regression function $f(\mathbf{x})$ is only approximated by $\sum \theta_j f_j(\mathbf{x})$, thereby introducing a bias term which may be considerable. The corresponding model can be given now by

$$Y(\mathbf{x}_i) = \sum_{j=0}^k \theta_j f_j(\mathbf{x}_i) + \psi(\mathbf{x}_i) + \varepsilon_i, \tag{1.1}$$

where $\psi(\mathbf{x})$ is an unknown contamination function from some class $\mathscr{F}$ on $\mathscr{X}$. For more general model formulations see Kiefer (1973), Huber (1975) or Sacks and Ylvisaker (1978). Now, the least squares estimators, which disregard the presence of $\psi(x)$, are no longer optimal among linear estimators for $\{\theta_j\}$, and therefore the search for new estimators is of special interest. In connection with the latter problem we refer to the papers by Sacks and Ylvisaker (1978) and by Marcus and Sacks (1978). In the latter paper, one-dimensional regression

$$E(y \mid x) = ax + b + \psi(x)$$

Received October 1978; revised November 1981.

[1] Research supported by NSF Grant MCS 78-01881.

AMS 1970 *subject classification.* Primary 62K05, 62G35; secondary 62J05.

Key words and phrases. D-optimality; E-optimality; linear regression; optimal design; robust design.

is considered for a class of functions $\psi(x)$ bounded by given function $\varphi(x): |\psi(x)| \le \varphi(x)$, and for the problem of minimizing the weighted mean square error

$$\sup_\psi E\{(\hat{a} - a)^2 + \lambda^2(\hat{b} - b)^2\}.$$

If $\lambda = 1$ then the criterion above is an analogue of A-optimality, and for a convex $\varphi(x)$ the best linear estimator and the optimal design found in the paper do not differ too much from the least squares estimator and corresponding design. In other cases the results were close to those with least squares and from that point of view it is plausible to consider systematic use of least squares, especially because the best linear estimator depends on $\varphi(x)$ and hence has to be defined separately in each problem. In the related direction, Huber (1975) considers the design problem in which

$$\sup_f \left\{ \int E(f(x) - \hat{a} - \hat{b}x)^2 \, dx \right\}$$

is to be minimized. Here f belongs to a class of functions which can be "reasonably" approximated by a linear function $ax + b$, where $\hat{a}$, $\hat{b}$ are least squares estimates, and the optimal design from a class of symmetric designs is found.

In connection with these and some other papers, it is natural to look at an evident generalization of the problem. Firstly, it is interesting to consider a general regression function or at least a k-dimensional linear regression. Secondly, in the direction of Kiefer (1975) and others, we may introduce a family of optimality criteria and study both the robustness under the contamination of the model and variation of criterion.

In this paper we consider k-dimensional linear regression and for the construction of optimality criteria we use Φ_p-family introduced by Kiefer (1974, 1975).

We confine ourselves to the use of the standard least squares estimates both because these estimates do not depend on the type of deviation from the model (i.e., on the specification of a class of ψ's), and because in the case of small deviations, the performance of least squares is nearly the same as of the best linear estimators, as was shown by Marcus and Sacks (1978).

The formulation of the problem is given in Section 2 along with the required notation. We define a family of minimax criteria which is based on the Φ_p-family and on a natural analogue of the covariance matrix of least squares estimates.

In Section 3, D-, A- and E-optimal continuous minimax designs are found for a broad class of problems (depending on the values of $\varphi(1)$, σ^2 and n), and the latter two are shown to be unique. These designs are defined by uniform measures over spheres of radii determined by criteria; in particular, all the optimal minimax designs do not coincide if $\varphi > 0$. The case of $k = 1$ is the only exclusion; here we have uniqueness in all senses. The optimal design is the same one over points $-z$, z found by Marcus and Sacks (1978) for their criterion. (The choice of z depends on φ, σ^2 and n.)

Examples and integer designs are discussed in Section 4.

2. Preliminaries. We consider the setup given by (1.1), where $f_0(\mathbf{x}_i) \equiv 1$, $f_j(\mathbf{x}_i) = x_{ij}$, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ik}) \in \mathscr{X}$, and the experimental region $\mathscr{X}$ will be specified later. We assume also that $|\psi(\mathbf{x})| \le \varepsilon\varphi(\mathbf{x})$, where $\varepsilon > 0$ and the function φ is a convex function of $\|\mathbf{x}\|^2 = (x_1^2 + \cdots + x_k^2)$. If a design matrix is $X = \{x_{ij}\}_{i=1,n,j=0,k}$ where $x_{i0} \equiv 1$, then least squares estimates (L.S.E.) of $\theta' = (\theta_0, \theta_1, \cdots, \theta_k)$ for nonsingular information matrix $M = X'X$ are

$$\hat{\theta} = M^{-1}X'Y$$

and the covariance matrix of L.S.E. is $D = \sigma^2 M^{-1}$. If contamination $\psi(\mathbf{x})$ is present, then we define the mean squared error (M.S.E.) matrix.

$$D(\psi) = E(\hat{\theta} - \theta)(\hat{\theta} - \theta)' = \sigma^2 M^{-1} + M^{-1}\psi\psi' M^{-1}, \tag{2.1}$$

where $\psi' = (\sum_{i=1}^n \psi(\mathbf{x}_i), \sum_{i=1}^n \psi(\mathbf{x}_i)x_{i1}, \cdots, \sum_{i=1}^n \psi(\mathbf{x}_i)x_{ik})$ and $M^{-1}\psi\psi' M^{-1}$ is the bias term.

Introducing the concept of "continuous design" ξ, which is an arbitrary probability measure on $\mathscr{X}$ (Federov, 1972; Kiefer, 1974), we can consider the information matrix $M(\xi)$ "normalized to one observation," $M = nM(\xi)$; and, normalized in the same manner, the vector $\Psi(\xi)$ with the components

$$E_\xi\{\psi(\mathbf{x})\} = \frac{1}{n}\sum_{i=1}^n \psi(\mathbf{x}_i), \quad E_\xi\{\psi(\mathbf{x})x_j\} = \frac{1}{n}\sum_{i=1}^n \psi(\mathbf{x}_i)x_{ij}, \quad j = 1, 2, \cdots, k;$$

so that $\Psi' = n(E_\xi\{\psi(\mathbf{x})\}, E_\xi\{\psi(\mathbf{x})x_1\}, \cdots, E_\xi\{\psi(\mathbf{x})x_k\}) = n\psi'(\xi)$, where E_ξ denotes expectation over $\mathscr{X}$ with ξ. (In our original setting we could define ξ by $\xi(\mathbf{x}) = \nu(\mathbf{x})/n$, where $\nu(\mathbf{x})$ is the number of $\mathbf{x}_i$'s which equal $\mathbf{x}$. Such a ξ is usually called "integer design.") Then we can write instead of (2.1)

$$D(\xi, \psi) = \frac{\sigma^2}{n} M^{-1}(\xi) + M^{-1}(\xi)\Psi(\xi)\Psi'(\xi)M^{-1}(\xi) \tag{2.1$'$}$$

and the matrix $D(\xi, \psi)$ can serve as a natural analogue of the covariance matrix $(\sigma^2/n)\, M^{-1}(\xi)$ in the noncontaminated case.

Of course, we can assign weights to both "variance" and "bias" terms in $D(\xi, \psi)$ in accordance with our attitude towards the model and need of safeguarding against dangerous deviations, but it follows from the results below that this can be done simply by appropriate choice of ε in $|\psi(\mathbf{x})| \le \varepsilon\varphi(\mathbf{x})$.

The family of criteria considered is produced by optimality functionals $\Phi_p(\xi, \psi)$ derived from the M.S.E. matrix $D(\xi, \psi)$, which is nonsingular if $M(\xi)$ is nonsingular. If $\lambda_0(\xi, \psi) \le \lambda_1(\xi, \psi) \le \ldots \le \lambda_k(\xi, \psi)$ denote the eigenvalues of $D(\xi, \psi)$, then the functionals $\Phi_p(\xi, \psi)$ are defined as follows:

$$\Phi_p(\xi, \psi) = \left[\frac{1}{k+1}\operatorname{tr}\{D^p(\xi, \psi)\}\right]^{1/p} = \left\{\frac{1}{k+1}\sum_{j=0}^k \lambda_j^p(\xi, \psi)\right\}^{1/p}, \quad 0 < p < \infty,$$

$$\Phi_0(\xi, \psi) = \lim_{p\to 0+} \Phi_p(\xi, \psi) = \{\det D(\xi, \psi)\}^{1/(k+1)},$$

$$\Phi_\infty(\xi, \psi) = \lim_{p\to\infty} \Phi_p(\xi, \psi) = \max_{0\le j\le k}\{\lambda_j(\xi, \psi)\} = \lambda_{\max}\{D(\xi, \psi)\}.$$

Φ_0, Φ_1 and Φ_∞ are the familiar D-, A- and E-optimality criteria. Since the contamination $\psi(\mathbf{x})$ is unknown, it is natural to apply a minimax approach and to define a Φ_p-optimal design $\xi_p^*(\varphi)$ as one that minimizes

$$\max_{\psi:|\psi(\mathbf{x})|\le\varphi(\mathbf{x})\cdot\epsilon} \Phi_p(\xi, \psi). \tag{2.2}$$

The approach to optimality criteria in this work differs from that of, say, Box and Draper (1959) or Huber (1975): we are interested in the best (in some sense) estimates for the coefficients $\{\theta_j\}$ rather than in the best approximation of the response surface by functions of class $\{f'(\mathbf{x})\theta\}$. This may seem odd at first glance; indeed, if $E\{y\,|\,\mathbf{x}\} = f'(\mathbf{x})\theta + \psi(\mathbf{x})$, where $\psi(\mathbf{x})$ is a contamination, why do we prefer better estimates $\hat{\theta}$ of θ and not, say, such estimates $\tilde{\theta}$ which provide for a better fit to the response y? The following examples show that the estimation of the coefficients is often a more important problem than the estimation of the response function.

Consider the method of the steepest descent where we approximate a response function y in a neighborhood of a point $\mathbf{x}_0$ by a linear function $\theta_0 + \sum_{j=1}^k \theta_j \mathrm{x}_j$ in order to find the directions and rates which define the next experimental region. In this case, the setting given by (1.1) is evidently more appropriate than the one which does not include a contamination term. At the same time the estimation of the response function does not constitute the goal of the experiment.

A similar situation corresponds to the problem of estimating a response function and its slope in a fixed point x_0. We may consider in the neighborhood of x_0 a model

$$y(x) = y(x_0) + y'(x_0)(x - x_0) + \psi(x - x_0) + \varepsilon(x)$$

with the contamination term bounded by $y''(x_0)(x - x_0)^2/2$.

In the following we denote σ^2/n by ρ and $E_\xi\{x_i x_j\}$ by $m_{ij}(\xi) = m_{ij}$, with $m_{0i} = E_\xi\{x_i\}$, so that $M(\xi) = \{m_{ij}\}_{i,j=\overline{0,k}}$. We denote $E_\xi\{\psi(\mathbf{x})x_i\}$ by $\pi_i(\xi) = \pi_i$ and $\pi = \{\pi_i\}_{i=\overline{0,k}}$. The notion "symmetric design" is used for any first order symmetric design; the latter is defined as one for which $m_{ij} = 0$ if $i \neq j$, $0 \le i, j \le k$, and $m_{ii} = \text{const} = m$ for all $1 \le i \le k$. Obviously we can assume that $\varepsilon = 1$, because otherwise we would denote $\sigma^2/(\varepsilon^2 n)$ by ρ, without altering the optimal designs.

3. Optimal (continuous) designs. In the beginning we limit our consideration to the class $\Xi(m)$ of all symmetric designs ξ with fixed $E_\xi\{x_i^2\} = m$; it will be shown later that optimal minimax designs for our criteria are very often symmetric. In this case $M^{-1}(\xi)$ is diagonal with $m^{\overline{00}} = 1$ and $m^{\overline{ii}} = m^{-1}$, $i = 1, 2, \cdots, k$, and the structure of "bias" matrix $\underline{M^{-1}}(\xi)\Psi(\xi)\Psi'(\xi)M^{-1}(\xi)$ is simple. That is, if we consider the functions 1, $m^{-1/2}x_i$, $i = \overline{1,k}$, as an orthogonal normalized basis in linear space L_2 spanned by the vectors 1, x_1, $\cdots$, $x_k (L_2 \subset L_2(\xi))$, then

$$E_\xi\{\psi(\mathbf{x})\} + \textstyle\sum_{i=1}^k m^{-1/2}x_i\, E_\xi\{\psi(\mathbf{x})m^{-1/2}x_i\} = \pi_0 + \sum_{i=1}^k \pi_i m^{-1} x_i = \psi_L$$

is a projection of ψ into L_2 and hence

$$\|\psi_L\|^2 = \pi_0^2 + m^{-1}\textstyle\sum_{i=1}^k \pi_i^2. \tag{3.1}$$

At the same time the quantity

$$\lambda_\psi = \pi_0^2 + m^{-2}\textstyle\sum_{i=1}^k \pi_i^2 \tag{3.2}$$

is the only positive eigenvalue of $M^{-1}\Psi\Psi'M^{-1}$ and therefore minimization of the maximum "bias" is closely related to the minimization over designs from $\Xi(m)$ of the maximum of $\|\Psi_L\|$. The relations (3.1) and (3.2) provide for an estimate of λ_ψ with $m > 1$:

$$\begin{aligned}(E_\xi\{\psi\})^2 \le \lambda_\psi &= \|\psi_L\|^2 + m^{-2}\textstyle\sum_{i=1}^k \pi_i^2(1-m) \\ &\le \|\psi_L\|^2 \le E_\xi\{\psi^2(\mathbf{x})\} \le E_\xi\{\varphi^2(\mathbf{x})\}.\end{aligned} \tag{3.3}$$

It is easy to notice that (3.3) gives a sharp estimate for the $\max_\psi \lambda_\psi$ which corresponds to $\psi(\mathbf{x}) \equiv \varphi(\mathbf{x}) \equiv$ const. The latter condition means that $\|\mathbf{x}\| =$ const and since $E_\xi\{x_i^2\} = m$ we immediately have $\|\mathbf{x}\| = \sqrt{mk}$. This implies that a design which yields equalities in (3.3) is supported by the points of sphere S_R of radius $R = \sqrt{mk}$. However, for $m < 1$, (3.3) does not give a sharp upper bound on λ_ψ; the corresponding result is given in the second part of the following theorem.

THEOREM 3.1 *For any $\xi \in \Xi(m)$ and for φ convex in $\|\mathbf{x}\|^2$*

$$\operatorname{mix}_\xi \max_\psi \|\psi_L\|^2 = \varphi^2(\sqrt{km}), \tag{3.4}$$

and

$$\min_\xi \max_\psi \lambda_\psi = \begin{cases} a_k m^{-1}\varphi^2(\sqrt{km}), & m \le a_k, \\ \varphi^2(\sqrt{km}), & m > a_k, \end{cases} \tag{3.5}$$

where

$$a_k = \begin{cases} \dfrac{k[(k-2)!]^2}{2^{2k-6}(k-1)^2\left[\left(\dfrac{k-3}{2}\right)!\right]^4}, & k \text{ odd}, \\[3ex] \dfrac{2^{2k-2}k\left[\left(\dfrac{k-2}{2}\right)!\right]^4}{\pi^2[(k-1)!]^2}, & k \text{ even}. \end{cases}$$

The equalities are attained in (3.4) and in (3.5) for $m > a_k$ by $\psi(\mathbf{x}) \equiv \varphi(\mathbf{x})$ with such a symmetric $\xi \in \Xi(m)$ that $\xi\{S_{\sqrt{mk}}\} = 1$ and in (3.5) for $m \le a_k$ by $\psi(\mathbf{x}) \equiv \varphi(\mathbf{x})\,\mathrm{sgn}(x_i)$ with ξ being continuous uniform measure over $S_{\sqrt{mk}}$.

PROOF. First, (3.4) follows immediately from the inequalities

$$(E_\xi\{\psi(\mathbf{x})\})^2 \le \|\psi_L\|^2 \le E_\xi\{\psi^2(\mathbf{x})\} \le E_\xi\{\varphi^2(\mathbf{x})\}$$

because then $\max_\psi \|\psi_L\|^2 \ge (E_\xi\{\varphi(\mathbf{x})\})^2$ and

$$E_\xi\{\varphi(\mathbf{x})\} \ge \varphi(\sqrt{E_\xi\|\mathbf{x}\|^2}) = \varphi(\sqrt{km})$$

due to the fact that φ is convex in $\|\mathbf{x}\|^2$. With ξ symmetric over sphere $S_{\sqrt{mk}}$, both upper and lower bounds for $\max_\psi \|\psi_L\|^2$ coincide with $\varphi^2(\sqrt{mk}\,)$.

Result (3.5) requires more elaborate treatment. Firstly we will prove (3.5) for $k = 1, 2$ and then sketch an inductional proof for $k > 2$.

For $k = 1$ we can show, using the convexity of φ, that

$$E_\xi\{\varphi(x)\,|x|\} = E_\xi\{|x|\}E_\xi\left\{\varphi(x)\frac{|x|}{E_\xi\{|x|\}}\right\} \ge E_\xi\{|x|\}E_{\xi^*}\{\varphi(x)\},$$

where $\xi^*(x) = \xi(x)\dfrac{|x|}{E_\xi\{|x|\}}$. Hence

$$E_\xi\{\varphi(x)\,|x|\} \ge E_\xi\{|x|\}\varphi(E_{\xi^*}|x|) = E_\xi\{|x|\}\varphi\left(\frac{m}{E_\xi\{|x|\}}\right) \tag{3.6}$$

and the right-hand side in (3.6) decreases with $E_\xi\{|x|\}$ due to the fact that φ is convex and $\varphi(0) = 0$. Therefore the minimum corresponds to the maximal value of $E_\xi\{|x|\}$ which is $\sqrt{E_\xi\{x^2\}} = \sqrt{m}$. Thus with $\psi(x) = \varphi(x)\,\mathrm{sgn}(x)$

$$\min_\xi\max_\psi E_\xi\{\psi(x)x\} \ge \min_\xi E_\xi\{\varphi(x)\,|x|\} \ge \sqrt{m}\,\varphi(\sqrt{m}) \tag{3.7}$$

and the lower bound is attained by the design ξ_0 supported by points $\pm\sqrt{m}$ with equal weights. Also, with $\psi(x) = \varphi(x)$

$$\min_\xi\max_\psi E_\xi\{\psi(x)\} \ge \min_\xi E_\xi\{\varphi(x)\} \ge \min_\xi\varphi(\sqrt{E_\xi\{x^2\}}) = \varphi(\sqrt{m}). \tag{3.8}$$

At the same time

$$\begin{aligned}\min_\xi\max_\psi\lambda_\psi(\xi) &\le \max_\psi\lambda_\psi(\xi_0)\\ &= \max_\psi {\textstyle\frac14}\{[\psi(-\sqrt{m}) + \psi(\sqrt{m})]^2 + m^{-1}[\psi(\sqrt{m}) - \psi(-\sqrt{m})]^2\}\\ &\le \begin{cases}\varphi^2(\sqrt{m}), & m > 1,\quad \psi(-\sqrt{m}) = \psi(\sqrt{m}) = \varphi(\sqrt{m})\\ m^{-1}\varphi^2(\sqrt{m}), & m \le 1,\quad -\psi(-\sqrt{m}) = \psi(\sqrt{m}) = \varphi(\sqrt{m}).\end{cases}\end{aligned} \tag{3.9}$$

From (3.7) through (3.9) we conclude that (3.5) holds for $k = 1$ with $a_1 = 1$.

For $k = 2$, we use polar coordinates (R, α) and denote by $\mathscr{I}(\alpha)$ a marginal distribution of α associated with a design measure ξ. Without loss of generality, we assume that p.d.f. $f(\alpha)$ exists for $\mathscr{I}(\alpha)$ and for the beginning we assume also that ξ is supported only by the points of the circle S_R; that is $\xi(\mathbf{x} = (x_1, x_2) : \|\mathbf{x}\| = \sqrt{2m} = R) = 1$.

We define now $f^*(\alpha)$, a "symmetric version" of $f(\alpha)$, by

$$f^*(\alpha) = \frac18\left[\sum_{r=-2}^{1}\left\{f\left(\frac{r\pi}{2} + \alpha\right) + f\left(\frac{r\pi}{2} - \alpha\right)\right\}\right]$$

and we can notice that

$$E_{f^*}|\sin\alpha| = E_{f^*}|\cos\alpha| = {\textstyle\frac12}\{E_f|\sin\alpha| + E_f|\cos\alpha|\}. \tag{3.10}$$

For any θ, the functions 1, $\sqrt{2}\sin(\alpha - \theta)$, $\sqrt{2}\cos(\alpha - \theta)$ form a basis in L_2, and therefore for a function $\psi_\theta = \varphi(R)R\,\mathrm{sgn}(\sin(\alpha - \theta))$ we have

$$\|\psi_\theta\|_{L_2}^2 = 2\{E_{f^*}|\sin(\alpha - \theta)|\}^2\varphi^2(R)R^2.$$

Also, using basis 1, $\sqrt{2}\sin\alpha$, $\sqrt{2}\cos\alpha$, we may write

$$\|\psi_\theta\|_{L_2}^2 = 2[\{E_{f^*}(\sin\alpha)\psi_\theta\}^2 + \{E_{f^*}(\cos\alpha)\psi_\theta\}^2]$$

so that

$$\max_\psi[\{E_{f^*}(\sin\alpha)\psi\}^2 + \{E_{f^*}(\cos\alpha)\psi\}^2]$$
$$\geq \max_\theta[(E_{f^*}(\sin\alpha)\psi_\theta)^2 + (E_{f^*}(\cos\alpha)\psi_\theta)^2] \geq \max_\theta(E_{f^*}|\sin(\alpha-\theta)|)^2\varphi^2(R)R^2.$$

Below we establish a lower bound on

$$\max_\theta E_{f^*}|\sin(\alpha-\theta)|.$$

Due to the construction of f^*, we can write $f^*(\alpha)$ as a Fourier series

$$f^*(\alpha) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{p=1}^\infty d_{2p}\cos(2p\alpha), \tag{3.11}$$

where

$$d_{2p} = \frac{1}{\pi}\int_{-\pi}^{\pi}\cos(2p\alpha)f^*(\alpha)\,d\alpha.$$

We define function $G(\theta)$ by

$$G(\theta) = \int_\theta^{\pi+\theta}\sin(\alpha-\theta)f^*(\alpha)\,d\alpha = \frac{1}{2}\int_{-\pi}^{\pi}|\sin(\alpha-\theta)|f^*(\alpha)\,d\alpha. \tag{3.12}$$

Now we use (3.11) and (3.12) to derive that

$$\begin{aligned}
G(\theta) &= \frac{1}{\pi} - \frac{2}{\pi}\sum_{p=1}^\infty d_{2p}\frac{\cos(2p\,\theta)}{4p^2-1},\\
G'(\theta) &= \frac{2}{\pi}\sum_{p=1}^\infty d_{2p}\frac{2p\sin(2p\,\theta)}{4p^2-1},\\
G''(\theta) &= \frac{2}{\pi}\sum_{p=1}^\infty d_{2p}\frac{4p^2\cos(2p\,\theta)}{4p^2-1}\\
&= \frac{2}{\pi}\left|\sum_{p=1}^\infty d_{2p}\cos(2p\,\theta) + \sum_{p=1}^\infty d_{2p}\frac{\cos(2p\,\theta)}{4p^2-1}\right|\\
&= 2f^*(\theta) - \frac{1}{\pi} - \left\{G(\theta) - \frac{1}{\pi}\right\}.
\end{aligned} \tag{3.13}$$

It follows from (3.13) that

$$G''(\theta) + G(\theta) = 2f^*(\theta)$$

and therefore

$$\int_0^\pi G''(\theta)\,d\theta + \int_0^\pi G(\theta)\,d\theta = G'(\pi) - G'(0) + \int_0^\pi G(\theta)\,d\theta = 1$$

so that $\int_0^\pi G(\theta) = 1$. This implies $\max_{0\leq\theta\leq\pi}G(\theta) \geq \dfrac{1}{\pi}$ and

$$\max_\theta E_{f^*}\{|\sin(\alpha-\theta)|\} \geq \frac{2}{\pi}. \tag{3.14}$$

It is clear that the equality in (3.14) holds only for $f^*(\alpha)$ uniform on a circle. Thus for a uniform ξ_0 on a circle of radius $R = \sqrt{2m}$

$$\min_\xi \max_\psi \pi_i = \pi_i(\xi_0) = \frac{2}{\pi}\varphi(\sqrt{2m})\sqrt{2m}$$

with, say, the "least favorable function" $\psi(\mathbf{x}) = \varphi(R)\operatorname{sgn}(\sin\alpha)$. Since the projection of a ψ onto the linear space spanned by $\sin\alpha$ and $\cos\alpha$ can be written as

$$a\sin\alpha + b\cos\alpha = \sqrt{a^2+b^2}\sin(\alpha+\omega),$$

we can easily see that

$$\min_\xi \max_\psi (\pi_1^2 + \pi_2^2) = \frac{8}{\pi^2}\varphi^2(\sqrt{2m})m \tag{3.15}$$

with the equality attained by ξ uniform on S_R. The same also holds for a design supported not only by points of S_R. To show that we use the convexity of $\varphi(R)$ in R^2 and write that

$$E_{f^*}\{R|\sin\alpha|\varphi(R)\} = E\left[E\{|\sin\alpha| \;\middle|\; \|\mathbf{x}\| = R\}R\,\varphi(R)\right]$$
$$\geq \frac{2}{\pi}ER\,\varphi(R) \geq \frac{2}{\pi}\varphi(\sqrt{ER^2})\sqrt{ER^2} = \frac{2\sqrt{2}}{\pi}\varphi(\sqrt{2m})\sqrt{m}$$

which establishes the lower bound on $\pi_1^2+\pi_2^2$ for an arbitrary symmetric f. The equality above holds only for $R = \sqrt{2m}$ and ξ uniform on S_R.

To accomplish the proof of (3.5) for $k = 2$, we have to establish the bounds on $\lambda_\psi = \pi_0^2 + m^{-2}(\pi_1^2+\pi_2^2)$.
It is clear that

$$\max_\psi\lambda_\psi \geq \max\{\max_\psi\pi_0^2, \max\{m^{-2}(\pi_1^2+\pi_2^2)\}\} \geq \max\left\{\varphi^2(\sqrt{2m}), \frac{8\varphi^2(\sqrt{2m})}{\pi^2 m}\right\}. \tag{3.16}$$

We can also show that the equality in (3.16) is attained by a uniform on $S_{\sqrt{2m}}$ design. We omit the proofs of the following two lemmas. The first of them uses standard variation arguments, and the second is more or less a corollary of the first.

LEMMA 3.1. *Let ξ be a uniform measure on a circle of radius $\sqrt{2m}$ and let us fix $\pi_0 = E_\xi\psi = E\psi$ and denote $\pi(1-\pi_0/\varphi(\sqrt{2m}))$ by β. Then*

$$\max_{\psi:E\psi=\pi_0}\{|\pi_1|\} = \pi^{-1}\,2\sin\left(\frac{\beta}{2}\right)\varphi(\sqrt{2m})\sqrt{2m}.$$

The maximum in the above equality is yielded by a function

$$\psi(\alpha) = \begin{cases} -\operatorname{sgn}\beta & \textit{for } \dfrac{\pi-\beta}{2} \leq \alpha \leq \dfrac{\pi+\beta}{2}, \\ \operatorname{sgn}\beta & \textit{otherwise.} \end{cases}$$

LEMMA 3.2. *Under the conditions of Lemma* 3.1

$$\max_\psi(\pi_0^2 + m^{-2}\pi_1^2) = \max_\beta\varphi^2(\sqrt{2m})\left\{\left(1-\frac{\beta}{\pi}\right)^2 + \frac{8\sin^2(\beta/2)}{\pi^2 m}\right\}$$
$$= \begin{cases} \dfrac{8}{\pi^2 m}\varphi^2(\sqrt{2m}), & \textit{if } m \leq 8/\pi^2, \\ \varphi^2(\sqrt{2m}), & \textit{if } m > 8/\pi^2. \end{cases}$$

Thus the proof of (3.5) for $k = 2$ is accomplished and $a_2 = 8/\pi^2$.

For $k > 2$ the proof can be carried out by induction. In k-dimensional polar coordinates we can write that

$$x_k = R \sin \alpha_{k-1}, \quad x_{k-1} = R \cos \alpha_{k-1} \sin \alpha_{k-2}, \cdots, \quad x_1 = R \cos \alpha_{k-1} \cos \alpha_{k-2} \cdots \cos \alpha_1,$$

where $-\pi/2 \le \alpha_i \le \pi/2$ for $i = \overline{2, k-1}$ and $-\pi \le \alpha_1 \le \pi$; the uniform density on a sphere in R^k is proportional to

$$\cos^{k-2}\alpha_{k-1} \cos^{k-3}\alpha_{k-2} \cdots \cos \alpha_2.$$

Let us assume as before without loss of generality that a design ξ is supported only by the points of sphere S_R, $R = \sqrt{km}$, and that the density f_k of $\alpha = (\alpha_1, \cdots, \alpha_{k-1})$ is symmetric in the same sense as $f^*(\alpha)$ in (3.11). Then we can write

$$f_k(\alpha) = f^{(1)}(\alpha_1) \cdots f^{(k-1)}(\alpha_{k-1})$$

and from symmetry of a design $\xi \in \Xi(m)$

$$E_\xi(x_i^2) = E_\xi(x_k^2) = R^2 E_{f_k}(\sin^2 \alpha_{k-1}) = R^2 k^{-1}, \quad i = 1, 2, \cdots, k. \tag{3.17}$$

Since R is fixed and f is symmetric, we have that in $k - 1$ dimensions

$$\max_\psi E_\xi \psi(\mathbf{x}) x_{k-1} = \varphi(R) E_\xi(|x_{k-1}|) = \varphi(R) R E_{f_{k-1}}(|\sin \alpha_{k-2}|).$$

Due to the induction hypothesis for the dimension $k - 1$

$$E_{f_{k-1}}(|\sin \alpha_{k-2}|) \ge \int_{-\pi/2}^{\pi/2} |\sin \alpha_{k-2}| \cos^{k-3} \alpha_{k-2} \frac{d\alpha_{k-2}}{c(k-3)}, \tag{3.18}$$

where

$$c(k-3) = \int_{-\pi/2}^{\pi/2} \cos^{k-3} \alpha \, d\alpha.$$

We can use this now to find

$$\max_\psi E \psi(\mathbf{x}) x_i, \quad i = 1, 2, \cdots, k.$$

Indeed, since the design is symmetric, we may consider only one value for i. If we take $i = k - 1$, we have (using symmetry of f) that

$$\begin{aligned}\max_\psi E\psi(\mathbf{x}) x_{k-1} &= \varphi(R) R \max_{f_k} E_{f_k}(|\sin \alpha_{k-2} \cos \alpha_{k-1}|) \\ &= \varphi(R) R \max_{f, f_{k-1}} \{E_{f(\alpha_{k-1})}(|\cos \alpha_{k-1}|) E_{f_{k-1}}(|\sin \alpha_{k-2}|)\}.\end{aligned}$$

Taking into account (3.18), we need to find $\max_f E_{f(\alpha_{k-1})}(|\cos \alpha_{k-1}|)$. It remains to show that the lower bound for this maximum is yielded by $f(\alpha_{k-1}) = \cos^{k-2} \alpha_{k-1} / c(k-2)$ in order to have the proof almost completed. To do this, we reformulate the problem, rewriting (3.17),

to minimize $\int_{-\pi/2}^{\pi/2} \cos \alpha \, f(\alpha) \, d\alpha$ under the conditions

$$\int_{-\pi/2}^{\pi/2} \cos^2 \alpha \, f(\alpha) \, d\alpha = \frac{k-1}{k}, \; f(\alpha) \ge 0 \text{ and } \int_{-\pi/2}^{\pi/2} f(\alpha) \, d\alpha = 1. \tag{3.19}$$

Omitting the justification, we may consider only densities of the form

$$f(\alpha) = \left(\textstyle\sum_{j=0}^{\infty} e_j \cos^{j/2} \alpha\right)^2.$$

Using notation $\int_{-\pi/2}^{\pi/2} \cos^m \alpha \, d\alpha = c(m)$, we reformulate (3.19) as

$$\text{to minimize } \textstyle\sum_{j,m=0}^{\infty} e_j e_m c\left(\frac{j+m+2}{2}\right) \tag{3.20}$$
$$\text{given } \textstyle\sum_{j,m=0}^{\infty} e_j e_m c\left(\frac{j+m+4}{2}\right) = \frac{k-1}{k}, \qquad \sum_{j,m=0}^{\infty} e_j e_m c\left(\frac{j+m}{2}\right) = 1.$$

The latter problem can be solved with the help of a modification of the Kuhn-Tucker theorem (Pshenichnyi, 1971, Chapter II). The solution is

$$e_{k-2} = \{c(k-2)\}^{-1/2} \quad \text{and} \quad e_j = 0 \quad \text{for} \quad j \neq k-2.$$

Thus under the restrictions of (3.19)

$$\min_f E_f(\cos \alpha_{k-1}) \geq \frac{c(k-1)}{c(k-2)},$$

where the equality holds for $f(\alpha) = \{c(k-2)\}^{-1}\cos^{k-2}\alpha$, and

$$c(m) = \begin{cases} \dfrac{2^m\left\{\left(\frac{m-1}{2}\right)!\right\}^2}{m!}, & m \text{ odd}, \\[2ex] \dfrac{\pi m!}{2^m\left\{\left(\frac{m}{2}\right)!\right\}^2}, & m \text{ even}. \end{cases}$$

We find now that on a unit sphere $\min_{f_k} E_{f_k}(|x_i|) \geq 2(k-1)\{c(k-2)\}^{-1}$ and the equality holds only for a uniform on a sphere f_k. Next we find that

$$\max_\psi \pi_i^2 \geq \varphi^2(R)R^2\left\{\int_{-\pi/2}^{\pi/2} |\sin \alpha_{k-1}| (\cos \alpha_{k-1})^{k-2} \frac{d\,\alpha_{k-1}}{c(k-2)}\right\}^2$$
$$= 4k(k-1)^{-2}\{c(k-2)\}^{-2} m\, \varphi^2(\sqrt{km}),$$

so that $a_k = 4k(k-1)^{-2}\{c(k-2)\}^{-2}$, and

$$a_k = \begin{cases} \dfrac{k[(k-2)!]^2}{2^{2k-6}(k-1)^2\left[\left(\frac{k-3}{2}\right)!\right]^4}, & k \text{ odd}, \\[2ex] \dfrac{2^{2k-2}k\left[\left(\frac{k-2}{2}\right)!\right]^2}{\pi^2[(k-1)!]^2}, & k \text{ even}. \end{cases} \tag{3.21}$$

We can see from (3.21) that a_k decreases with k from $a_2 = 8/\pi^2 \cong .8106$ to $a_\infty = \lim_{k\to\infty} a_k = 2/\pi \cong .6366$.

The last part of the proof is similar to that for the case of $k = 2$. □

Theorem 3.1 enables us to find optimal symmetric minimax designs for any of the Φ_p-family criterion. We present below the results on A-, D- and E-optimality.

LEMMA 3.3. *Any symmetric design $\xi_0 \in \Xi(m)$ supported only by the points of sphere S_R of radius $R = \sqrt{mk}$ is D-optimal in $\Xi(m)$ in the sense of the minimax criterion* (2.2) *if $\varphi(\mathbf{x})$ is convex in $\|\mathbf{x}\|^2$.*

The result follows from (3.4) because

$$\det D(\xi, \psi) = \left(\frac{\rho}{m}\right)^k\left(\rho + \pi_0^2 + \frac{1}{m}\textstyle\sum_{i=1}^k \pi_i^2\right) = \left(\frac{\rho}{m}\right)^k(\rho + \|\psi_L\|^2).$$

Let m_0 denote a value m such that

$$(m_0^{-1}\rho)^k\{\rho + \varphi^2(\sqrt{m_0 k})\} = \min_{m>0}(m^{-1}\rho)^k\{\rho + \varphi^2(\sqrt{mk})\},$$

and let R_0 denote the value of $\sqrt{m_0 k}$ for $m_0 < +\infty$. Also let B_R denote a set of support of a first order symmetric design over sphere S_R. E.g., B_R is any set of points from S_R such that for some measure ξ over B_r, $E_\xi(x_i^2) = k^{-1}R^2$ and $E_\xi(x_i x_j) = 0$ for all $i \neq j$. It is clear now that if an experimental region $\mathscr{X}$ contains a set B_{R_0}, then the symmetric D-optimal design over $\mathscr{X}$ is one supported by B_{R_0}. The following theorem presents the result on the overall D-optimality of such a design.

THEOREM 3.2. *Consider the setup defined in Section 2 with $\varphi(x)$ convex in $\|x\|^2$. Then*
(i) *if $m_0 < +\infty$ and there exists such a set B_{R_0} which is contained in region $\mathscr{X}$, then the symmetric D-optimal design from $\Xi(m_0)$ supported by B_{R_0} is D-optimal in the class Ξ of all designs over $\mathscr{X}$;*
(ii) *if $\mathscr{X}$ is compact and symmetric (namely, invariant under the permutations and changes of signs of the coordinates) and $d_{\mathscr{X}} = \max_{x\in\mathscr{X}}\|x\| = R < R_0$, or if $m_0 = \infty$, then there is at least one symmetric D-optimal design in $\Xi(R^2/k)$ which is D-optimal over $\mathscr{X}$ in the class Ξ;*
(iii) *the reverse is also true. If either $\mathscr{X}$ is symmetric compact or $\mathscr{X} \supset B_{R_0}$, then the D-optimal designs over $\mathscr{X}$ are symmetric.*

PROOF. For an arbitrary design ξ with nonsingular information matrix $M(\xi)$, introduce an orthogonal matrix U such that $U^\top M(\xi)U = \lambda$, where λ is a diagonal matrix of eigenvalues of $M(\xi)$ and let us denote $U^\top\mathbf{x}$ by $\mathbf{z}$ and $E\{\psi(\mathbf{z})z_i\}$ by ν_i where the expectation is taken with respect to the design measure $\xi'(\mathbf{z}) = \xi(U^\top\mathbf{x})$. It is convenient here to use the notation $\mathbf{x} = (1, x_1, \cdots, x_k)$, so that the value of $\|\mathbf{x}\|^2$ is $1 + \sum_{i=1}^k x_i^2$ and the function $\varphi(\mathbf{x})$ becomes a function of $\|\mathbf{x}\|^2 - 1$. Then, since $\|\mathbf{z}\| = \|\mathbf{x}\|$, we have $\varphi(\mathbf{z}) = \varphi(\mathbf{x})$ and

$$\begin{aligned}\max_\psi \det D(\xi, \psi) \geq\ & \det D(\xi, \varphi)\\ &= \det(U[\rho\lambda^{-1} + \lambda^{-1}U^\top E(\mathbf{x}\varphi)\{E(\mathbf{x}\varphi)\}^\top U\lambda^{-1}]U^\top)\\ &= \det[\rho\lambda^{-1} + \lambda^{-1}E(\mathbf{z}\varphi)\{E(\mathbf{z}\varphi)\}^\top\lambda^{-1}]\\ &= \rho^k(\rho + \textstyle\sum_{i=0}^k \nu_i^2\lambda_i^{-1})/\prod_{i=0}^k \lambda_i.\end{aligned} \tag{3.22}$$

In the same way as in (3.1), we notice that $\sum_{i=0}^k \nu_i^2\lambda_i^{-1} = \|\varphi_{L_2(\mathbf{z})}\|^2$, where $L_2(\mathbf{z})$ is the linear space spanned by the orthonormal set of vectors $z_0\lambda_0^{-1/2}, \cdots, z_k\lambda_k^{-1/2}$ (the latter are orthonormal because $EZZ^\top = \lambda$), and since $L_2(\mathbf{z}) = L_2$, where L_2 is spanned by the vectors $1, x_1, \cdots, x_k$, we can see that

$$\|\varphi_{L_2(\mathbf{z})}\|^2 \geq \{E_\xi(\varphi)\}^2 \geq \varphi^2((\textstyle\sum_{i=1}^k m_{ii})^{1/2}),$$

where $m_{ii} = E_\xi(x_i^2)$. Also

$$\textstyle\prod_{i=0}^k \lambda_i = \det M(\xi) \leq \prod_{i=1}^k m_{ii} \leq (k^{-1}\sum_{i=1}^k m_{ii})^k. \tag{3.23}$$

Therefore, if we denote by m the value $k^{-1}\sum_{i=1}^k m_{ii}$, then

$$\max_\psi \det D(\xi, \psi) \geq \rho^k\{\rho + \varphi^2(\sqrt{mk})\}m^{-k} \tag{3.24}$$

and the equality in (3.24) holds if and only if $\lambda_i = m_{ii}$ for all $i = 1, 2, \cdots, k$, that is, if the design ξ is symmetric. The right hand side of (3.24) is exactly $\max_\psi \det D(\xi_0, \psi)$, where ξ_0 is D-optimal in $\Xi(m)$. Therefore, if ξ is nonsymmetric, it can be "improved" in the sense of the minimax D-optimality criterion (2.2), which implies statement (3) of the theorem. At the same time, the minimum of the function in the right-hand side of (3.24) is attained by $m = m_0$ and this function decreases with $m < m_0$ and increases with $m > m_0$ (if $m_0 < +\infty$), which implies statements (1) and (2). □

Unlike the D-optimal design, A- and E-optimal symmetric designs are unique and correspond to the uniform continuous measures on appropriate spheres. To prove that, we have to determine

$$\min_m[\max\{\rho + \varphi^2(\sqrt{km}), \frac{\rho}{m} + \frac{a_k}{m}\varphi^2(\sqrt{km})\}]$$

for the E-optimality and

$$\min_m[\rho + \frac{k\rho}{m} + \max\{\varphi^2(\sqrt{km}), \frac{a_k}{m}\varphi^2(\sqrt{km})\}]$$

for the A-optimality.

Let us fix the value of k and denote by m_1 that value of m which yields the minimum of the function

$$km^{-1}\rho + a_k m^{-1}\varphi^2(\sqrt{km}), \tag{3.25}$$

and by m_∞ that value of m which yields the minimum of

$$m^{-1}\rho + a_k m^{-1}\varphi^2(\sqrt{km}). \tag{3.26}$$

Let μ_∞ be the solution of the equation

$$\rho + \varphi^2(\sqrt{km}) = m^{-1}\rho + a_k m^{-1}\varphi^2(\sqrt{km}) \tag{3.27}$$

and let $\nu_\infty = \min(\mu_\infty, m_\infty)$ and $\nu_1 = \min(a_k, m_1)$. Clearly, $\nu_\infty \leq 1$ and $\nu_1 \leq a_k < 1$.

The following Lemma is similar to Lemmas 3.1 and 3.2, and its proof is omitted.

LEMMA 3.4. *Let φ be convex in $\|\mathbf{x}\|^2$. Then in a class of all symmetric designs, the E-optimal design is uniform on a sphere of radius $R_\infty = \sqrt{\nu_\infty k}$ and the A-optimal design is uniform on a sphere of radius $R_1 = \sqrt{\nu_1 k}$.*

Theorem 3.3 below does not give a general solution, but, as we will see later, it solves the problem in a wide range of important cases. Proof of the theorem is omitted because of similarity to the proof of Theorem 3.2.

THEOREM 3.3. *Let φ be convex in $\|\mathbf{x}\|^2$ and let χ be such an experimental region that the sphere of radius $R = \sqrt{m_1 k}$ (or $R = \sqrt{m_\infty k}$) is contained in χ. Then, if $m_1 \leq a_k$ ($m_\infty \leq \mu_\infty$), the A-optimal (E-optimal) design over χ is a symmetric one (uniform on S_R).*

The above theorem does not establish the overall optimality of symmetric designs. However, the conditions $m_1 \leq a_k$ or $m_\infty \leq \mu_\infty$ should be the most common ones. To show this, we consider functions $\varphi(\|\mathbf{x}\|) = \|\mathbf{x}\|^\alpha$, $\alpha \geq 2$, and the A-optimality criterion (for the E-optimality, the results are very similar). We find that

$$m_1 = k^{-1}\left\{\frac{\rho^k}{a_k(\alpha - 1)}\right\}^{1/\alpha}$$

and the inequality $m_1 \leq a_k$ is equivalent to

$$\rho \leq k^{\alpha-1}a_k^{\alpha+1}(\alpha - 1). \tag{3.28}$$

The right-hand side in (3.28) increases with k and α and it is not smaller than one for all k, α. Therefore, the symmetric designs may not be optimal in a general class only if ρ is sufficiently "large." But $\rho = \sigma^2/n\varepsilon^2$ and $|\psi(\mathbf{x})| \leq \varepsilon\varphi(\|\mathbf{x}\|)$ with $\varphi(1) = 1$, so "large" ρ's correspond to the contaminations of far smaller order than the random error. In this latter case the minimization of the variance part of the MSE may be of prime interest, and this is done by the standard least squares optimal design. In addition, ρ decreases with n so that for "large" n, the optimal designs are symmetric ones.

4. Integer designs and examples. The uniformity of the symmetric A- and E-optimal designs found above suggests that the integer optimal symmetric designs are also uniform on some spheres. This fact has been established for the dimension $k = 2$, in which case the optimal measures for an n-trial design are given by n uniformly distributed points on a circle. With the n fixed, all the results of Section 3 hold with a value $a_2(n)$ substituting for a_2, where

$$a_2(n) = \begin{cases} \dfrac{8}{n^2}\left[\dfrac{n-1}{2}\left\{1 - 2\cos\dfrac{\pi}{n} + \left(\cos\dfrac{3\pi}{2n}\right)\Big/\left(\cos\dfrac{\pi}{2n}\right)\right\}\right. \\ \quad \left. + \left(\cos\dfrac{\pi}{2n}\right)^2\Big/\left(\sin\dfrac{\pi}{n}\right)^2\right], \quad n \text{ odd}, \\ \dfrac{8}{n^2}\left(\sin\dfrac{\pi}{n}\right)^{-2}, \quad n \text{ even}. \end{cases}$$

In both cases $a_2(n) \to 8/\pi^2$ as $n \to \infty$. It is easy to verify that the sequences $a_2(2p + 1)$ and $a_2(2p)$ decrease with p and that

$$a_2(2p) > a_2(2p - 1) > a_2(2p + 2) > a_2(2p + 1).$$

Therefore

$$\max_n a_2(n) = a_2(4) = 1$$

and for any n, $.8106 \cong 8/\pi^2 < a_2(n) \le 1$. We also expect that similar results hold for any dimension k and that values $a_k(n)$ exhibit the same cyclic pattern depending on $n (\text{mod } k)$ as in the case of $k = 2$. However, from a practical point of view, an experiment over n uniform points on a k-dimensional sphere is hardly feasible and in the examples below we show that some standard designs preserve sufficient efficiency with respect to optimal designs.

EXAMPLE 1. $k = 1$. In this case for $m \le 1$ the symmetric design ξ_m supported by points $\pm\sqrt{m}\,(\xi_m(\sqrt{m}) = \xi_m(-\sqrt{m}) = ½)$ is Φ_p-optimal in the sense of criterion (2.2) for all $p \ge 0$ in the class of all designs with $E_\xi\{x_i^2\} = m$. If $\psi_1(x) = \varphi(x)\mathrm{sgn}\, x$, then (ξ_m, ψ_1) is a saddle point for all $p \ge 0$ in a game with loss function $\Phi_p(\xi, \psi)$. (For $p = 0$ we could, of course, take $\varphi(x)$ instead of $\psi_1(x)$.) This fact immediately follows from Theorem 3.1 because for any design ξ with $E_\xi\{x_i^2\} = m$ the minimal eigenvalue is not smaller than $\rho = \lambda_{\min}\{D(\xi_m, \psi_1)\}$ and the maximum over ψ of the maximal eigenvalue is not smaller than $m^{-1}\rho + m^{-1}\varphi^2(\sqrt{m}) = \lambda_{\max}\{D(\xi_m, \psi_1)\}$ and we can easily show that for any ψ and $p \ge 0$,

$$\lambda_0^p\{D(\xi_m, \psi)\} + \lambda_1^p\{D(\xi_m, \psi)\} \le \rho^p + (m^{-1}\rho + m^{-1}\varphi^2(\sqrt{m}))^p = \mathrm{tr}\{D(\xi_m, \psi_1)\}^p.$$

The minimum of $\mathrm{tr}\{D(\xi_m, \psi_1)\}^p$ is attained for all $p > 0$ as soon as $m^{-1}(\rho + \varphi^2(\sqrt{m}\,))$ is minimal (of course, this provides also for the minimum of $\det D(\xi_m, \psi_1)$) that is, with $m = m_1$. If $m_1 \le 1$, then the corresponding design is overall optimal for all $p \ge 0$.

As an example, we consider the case $\varphi(x) = x^\alpha$, $\alpha > 1$. Then the minimum of $\max_\psi \mathrm{tr}\{D(\xi_m, \psi)\}^p$ is yielded by $m_1 = \{\rho/(\alpha - 1)\}^{1/\alpha}$ with $\rho \le \alpha - 1$, so that $m_1 \le 1$.

In this case for all p the Φ_p-optimal design is supported by points $\pm\{\rho/(\alpha - 1)\}^{1/2\alpha}$ if $\rho < \alpha - 1$, which agrees with the result of Marcus and Sacks (1978) for their criterion.

EXAMPLE 2. We consider a linear model on a square $\chi = [-1, 1]^2$ with the contamination bounded by $\varphi(\mathbf{x}) = (x_1^2 + x_2^2) = \|\mathbf{x}\|^2$. Let us fix n and assume for simplicity that $\sigma = \varepsilon$ (that is, contamination is comparable with the experimental error) so that $\rho = 1/n$. We find then using notation of Section 3 that

$$m_\infty = (2\sqrt{na_2(n)})^{-1}, \quad m_1 = (\sqrt{2na_2(n)})^{-1}, \quad m_0 = \infty.$$

TABLE 1

Efficiencies of minimax optimal designs. Linear regression, $k = 2$, second order contamination

	n = 4			"large" n*		
p \ q	0	1	∞	0	1	∞
0	1.000	.492	.343	1.000	$\mathcal{O}(1/n\sqrt{n})$	$\mathcal{O}(1/n\sqrt{n})$
1	.883	1.000	.943	.893	1.000	.943
∞	.802	.947	1.000	.820	.943	1.000

* *Here $e_{q,p}$ is defined as* $\lim_{n\to\infty} e_{q,p}(n)$.

Since $8/\pi^2 < a_2(n) \le 1$, we can notice that $m_1 < a_2(n)$ for all n and that $m_\infty < \mu_\infty$. The latter value does not have to be found; we verify that $\rho + 4m_\infty^2 < (\rho + 4a_2(n)m_\infty^2)m_\infty^{-1}$. Therefore A- and E-optimal designs are uniform distributions of n points over spheres of, respectively, radii $R_1 = (2/na_2(n))^{1/4}$ and $R_\infty = (na_2(n))^{-1/4}$ and both spheres are contained in χ. It follows now from Theorems 3.2 and 3.3 that these designs are A- and E-optimal over χ and that the D-optimal design for $n = 4t$, $t = 1, 2, \cdots$, is also unique and supported by the vertices of the square. However, for $n \ne 4t$, the structure of the D-optimal design is not clear.

Below we compare the results for $n = 4$, when $a_2(4) = 1$, $\rho = 1/4$, $m_1 = 1/(2\sqrt{2})$ $m_\infty = 1/4$ and (for the D-optimal design) $\tilde{m}_0 = \sqrt{2}$, with the results for a "large n," so that $\rho = 1/n$, $a_2(n) \cong 8/\pi^2$, $m_1 = \pi/4\sqrt{n}$, $m_\infty = \pi/4\sqrt{2n}$ and $\tilde{m}_0 = \sqrt{2}$.

We determine the efficiency $e_{q,p}$ of a Φ_p-optimal design with respect to Φ_q-optimal one as the ratio

$$e_{q,p} = \frac{\max_\psi \Phi_q(\xi_q, \psi)}{\max_\psi \Phi_q(\xi_p, \psi)}$$

and present the results in Table 1.

From Table 1 we can easily see that A- and E-optimal designs are nearly robust with respect to changes of criterion, unlike D-optimal design with the efficiencies rapidly diminishing. The same kind of statement would also hold if $\sigma \ne \varepsilon$, but changes in φ would qualitatively change the situation, e.g., if φ is of a third order, then all m_i's, $i = 0, 1, \infty$, are of the same order $n^{1/3}$ and the limits of $e_{q,0}(q = 1, \infty)$ are bounded away from zero.

EXAMPLE 3. We compare below the performance of some common types of integer designs with the continuous optimal ones for $k > 2$.

We consider symmetric star-point designs over $2k$ star points (with $k - 1$ zero coordinates) on a sphere and regular replicas of 2^k factorial designs. From the condition $Ex_i^2 = m$ we obtain that in the first case the only nonzero coordinate is $\pm\sqrt{mk}$, and in the second case the levels of factors are $\pm\sqrt{m}$. It is easy to show that in both cases

$$\max_\psi \lambda_\psi = \begin{cases} m^{-1}\varphi^2(\sqrt{km}), & m < 1 \\ \varphi^2(\sqrt{km}), & m \ge 1, \end{cases}$$

so that at least for small values of ρ, for which the optimal designs are symmetric, A-optimal design corresponds to $\tilde{m}_1$ which minimizes

$$\rho + \frac{\rho k}{m} + \frac{\varphi^2(\sqrt{km})}{m},$$

and E-optimal design corresponds to $\tilde{m}_\infty$ which minimizes

$$\rho/m + \varphi^2(\sqrt{km})/m.$$

The respective values of m_i, $i = 1, \infty$, and m for continuous designs can be obtained from (3.24), (3.25).

Suppose that contamination is of order t. Then, with $\varphi(\mathbf{x}) = \|\mathbf{x}\|^t$, we obtain for the integer and E-optimal continuous designs, repsectively,

$$\tilde{m}_\infty = k^{-1}\{\rho/(t-1)\}^{1/t}, \quad \text{and} \quad m_\infty = k^{-1}\{\rho/a_k(t-1)\}^{1/t}. \tag{4.1}$$

This enables us to derive that efficiency of integer designs with respect to the optimal continuous one is equal to $(a_k)^{1/t} \geq (\min_k a_k)^{1/2} \geq (2/\pi)^{1/2} \cong .7979$. The result shows that star-point designs and regular replicas of factorial designs are very efficient with respect to E-optimal designs, provided the levels of the factors correspond to the optimal value $\tilde{m}_\infty$. The same is true also for A-optimal designs.

EXAMPLE 4. We are interested now in construction of integer designs for the experimental regions which are either non-spherical or do not satisfy the condition of Theorem 3.3. E. g., suppose that χ is a cube, $\chi = [-b, b]^k$ and we search for a design with high E-efficiency. Assume also that $\tilde{m}_\infty > b^2$, with $\tilde{m}_\infty$ as in Example 3, so that the cube is contained in a sphere $\tilde{S}_R$ of radius $\tilde{R} = \sqrt{\tilde{m}_\infty k}$. Consider now a sphere S_R of radius $b\sqrt{k}$. It is easy to see that if $b^2 \leq \mu_\infty$ then the E-optimal for this sphere design is uniform one over S_R and maximal eigenvalue is $b^{-2}\rho + a_k b^{-2}\varphi^2(b\sqrt{k})$. This value is to be compared with $b^{-2}\rho + b^{-2}\varphi^2(b\sqrt{k})$, which corresponds to a regular replica of 2^k factorial with levels $\pm b$. The efficiency ratio e is now

$$\frac{\rho + a_k\varphi^2(b\sqrt{k})}{\rho + \varphi^2(b\sqrt{k})} = 1 - \frac{\varphi^2(b\sqrt{k})(1-a_k)}{\rho + \varphi^2(b\sqrt{k})},$$

and for $\varphi(\mathbf{x}) = \|\mathbf{x}\|^t$ we obtain from (4.1)

$$(b\sqrt{k})^t < (\sqrt{\tilde{m}_\infty k})^t = \rho/(t-1),$$

so that

$$e \geq 1 - \frac{\rho(t-1)^{-1}(1-a_k)}{\rho(1+(t-1)^{-1})} = 1 - \frac{1-a_k}{t} > 1 - \frac{1-2/\pi}{t}.$$

For $t \geq 2$ we have

$$e \geq 1 - \frac{1-2/\pi}{2} \cong .8183$$

which demonstrates the lower bound on the efficiency of our integer design with respect to the E-optimal one. The same result is also true for the A-efficiency, and the case of D-criterion is covered by Theorem 3.1.

Acknowledgement. The author is gratuful to Ching-Shui Chen, A.S. Hedayat, and Jack Kiefer for valuable comments and remarks and to Ky Fan and Roy Leipnik for suggesting the ideas helpful in proving Theorem 3.1.

REFERENCES

BOX, G. E. P., and DRAPER, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc* **54** 622–654.

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

GALIL, Z. and KIEFER, J. (1977a). Comparison of rotatable designs for regression on balls, I (quadratic). *J. Statist. Plann. Inference* **1** 27–40.

GALIL, Z. and KIEFER, J. (1977b). Comparison of designs for quadratic regression on cubes. *J. Statist. Plann. Inference* **1** 121–132.

HUBER, P. (1975). *Robustness and Designs. A Survey of Statistical Design and Linear Models.* (ed. Srivastava) 287–303. North Holland, Amsterdam.

KIEFER, J. (1973). Optimal designs for fitting biased multiresponse surfaces. *Symposium on Multivariate Analysis III.* (ed. Krishnaiah) 287–297. Academic, New York.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879.

KIEFER, J. (1975). Optimal design: Variation in structure and performance under change of criterion. *Biometrica* **62** 277–288.

MARCUS, M. B. and SACKS, J. (1978). Robust designs for regression problems. *Statistical Decision Theory and Related Topics, II.* 245–268. Academic, New York.

PESOTCHINSKY, L. (1978). Φ_p-optimal second order designs for symmetric regions. *J. Statist. Plann. Inference* **2** 173–188.

PSHENICHNYI, B. N. (1971). *Necessary conditions for an extremum.* Dekker, New York.

SACKS, J. and YLVISACKER, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6** 1122–1137.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CA 93106