The Annals of Statistics
1993, Vol. 21, No. 1, 402–415

E-OPTIMAL DESIGNS FOR POLYNOMIAL REGRESSION

BY FRIEDRICH PUKELSHEIM[1] AND WILLIAM J. STUDDEN[2]

Pennsylvania State University and Purdue University

E-optimal designs for the full mean parameter vector, and for many subsets in univariate polynomial regression models are determined. The derivation is based on the interplay between E-optimality and scalar optimality. The scalar parameter systems are obtained as transformations of the coefficient vector c of the Chebyshev polynomial.

1. Introduction. In a linear model for polynomial regression of degree d on the interval $[-1, 1]$, we show that the E-optimal design for all $d + 1$ parameters and for many parameter subsets are supported by the Chebyshev points $s_i = \cos((d - i)\pi/d)$, for $i = 0, 1, \ldots, d$. This result was conjectured by Preitschopf [(1989), page 148], on the grounds of an extensive numerical study. In addition, we present an explicit formula for the weights w_i of s_i.

Whereas generally results on E-optimality appear to be less explicit than those for D- and A-optimality, the present paper provides an instance testifying to the contrary. The D-optimal polynomial regression design has the extreme points of the Legendre polynomial for its support, with constant weight $1/(d + 1)$, as established by Hoel (1958). A similar characterization for the support of the A-optimal design is not known, but knowledge of the support points permits an easy way to compute the weights, see Pukelsheim and Torsney (1991).

In Section 2 we start out by investigating the relationship between E- and c-optimality. There is a certain duality, but the kernel which relates the two problems to each other is convex in *both* variables. Hence duality gaps cannot be ruled out. Theorem 2.2 essentially states that duality holds true provided there exists a saddle-point.

In Section 3, we therefore first turn to c-optimality. The work of Studden (1968) suggests that the Chebyshev polynomial $T_d(x) = \sum_{i=0}^{d} c_i x^i$ plays a central role. Theorem 3.1 introduces the designs ξ_I that are optimal for $c'KK'\theta$, where c is the Chebyshev coefficient vector $c = (c_0, c_1, \ldots, c_d)'$ and the matrix $D = KK'$ is diagonal, with d_{ii} being 1 or 0 according as i is in the index set I or not. The proof relies on changing the polynomial basis from the power basis to the one provided by the Lagrange interpolating polynomials, with nodes given by the Chebyshev points.

Received July 1991; revised January 1992.

[1]On leave from the Universität Augsburg, supported by the Volkswagen-Stiftung.

[2]Partially supported by NSF Grant DMS-88-02535.

AMS 1991 *subject classification*. 62K05.

Key words and phrases. Approximate design theory, Chebyshev polynomial, c-optimality, E-optimality, parameter subset optimality, polynomial regression.

Section 4 takes up E-optimality. Our main result, Theorem 4.1, establishes E-optimality for the subsystem $\theta_I = K'\theta$ of the design ξ_I that is optimal for the scalar system $c'KK'\theta$. To do so we make the assumption that the index set I contains at least one index for which the Chebyshev coefficient is nonzero. The proof concentrates on showing that the information matrix C of the design ξ_I has $1/\|K'c\|^2$ for its smallest eigenvalue. For degree $d > 1$ this eigenvalue is shown to be simple.

Section 5 concludes the paper with a discussion of the results. There are interesting interrelations with the theory of polynomials, strengthening a result that flows from Erdös (1947), and rederiving a theorem on the extremum properties of Chebyshev polynomials that is due to Markoff (1916). We comment on the asymptotic behaviour of ξ_I as the degree d tends to infinity, and on the transition to an arbitrary interval $[a, b]$.

Independently Heiligers (1991) has obtained results of a similar nature. He convexifies the kernel $(d, M) \mapsto d'K'M^{-1}Kd$ on which we comment after Corollary 2.3, and he decomposes the E-optimal moment matrix $M(\xi_I)$ into two blocks of which each is investigated separately. In our exposition this corresponds to the polynomials P and Q which we introduce in part I of the proof of our Theorem 4.1. Our analysis is geometric in nature and connections with extremal properties of polynomials are made in Section 5. Besides these technical differences, Heiligers has a different outlook in that he exhibits the dependence of the results on the underlying interval $[a, b]$ which we here simply take to be $[-1, 1]$.

2. E-Optimality and c-optimality. We consider the usual linear model in which the experimenter chooses experimental conditions x from a compact experimental domain $\mathscr{X}$ and then observes a real-valued response Y with expectation $f(x)'\theta$, where the continuous regression function $f\colon \mathscr{X} \mapsto \mathbb{R}^k$ is known while $\theta \in \mathbb{R}^k$ is an unknown mean parameter vector. Responses under different experimental conditions, or replicated responses under identical experimental conditions are taken to be uncorrelated, with constant variance σ^2. This set-up for the design of experiments is standard, compare Kiefer (1974).

An experimental design ξ on $\mathscr{X}$ is a probability measure with finite support $x_1, \dots, x_l$ and corresponding weights $w_1, \dots, w_l$, directing the experimenter to realize a proportion w_i of all observations under experimental conditions x_i. The performance of a design ξ is evaluated through its moment matrix $M(\xi) = \sum_{i=1}^{l} w_i f(x_i) f(x_i)'$.

Let the parameter system of interest be $K'\theta$ where the $k \times s$ matrix K has full column rank s. With a design ξ we associate the information matrix $C_K(M(\xi))$ for $K'\theta$, given by

$$C_K(M(\xi)) = \min_{L \in \mathbb{R}^{s \times k}\colon LK = I_s} LM(\xi)L',$$

see Gaffke (1987). A design is called E-optimal for $K'\theta$ when it maximizes the smallest eigenvalue of the information matrices $C_K(M(\xi))$ among all possible designs.

Given a coefficient vector $c \in \mathbb{R}^k$, a design is called optimal for $c'\theta$ when it minimizes the standardized variance $c'M(\xi)^- c$ of the least squares estimator among all designs ξ under which the scalar parameter function $c'\theta$ is estimable.

The following two theorems point towards some intricate relations between E-optimality for $K'\theta$ and optimality for $z'K'\theta$. The first theorem considers the case when the smallest eigenvalue of the E-optimal information matrix is simple.

THEOREM 2.1. *Let ξ be a design that has a positive definite information matrix C for $K'\theta$, and let $\pm z \in \mathbb{R}^s$ be an eigenvector corresponding to the smallest eigenvalue of C. If the smallest eigenvalue of C has multiplicity one, then ξ is E-optimal for $K'\theta$ if and only if ξ is optimal for $z'K'\theta$.*

PROOF. By Theorem 8 of Pukelsheim (1980), if the smallest eigenvalue of C is simple, then ξ is E-optimal for $K'\theta$ if and only if there exists a generalized inverse G of $M(\xi)$ such that

$$(z'K'Gf(x))^2 \le \frac{\|z\|^2}{\lambda_{\min}(C)}, \quad \text{for all } x \in \mathscr{X}.$$

By the same theorem, the condition is necessary and sufficient for optimality for $z'K'\theta$. □

In general, E-optimality may obtain *without* any scalar optimality property. For $K = I_k$ and $f(x) = (\sin x, \cos x)'$, with $x \in (0, 2\pi]$, the only E-optimal design for θ has moment matrix $I_2/2$. But for every vector $0 \neq c \in \mathbb{R}^k$ the unique optimal design for $c'\theta$ has moment matrix $cc'/\|c\|^2$. See Example 5 of Pukelsheim (1981).

Following Elfving (1952) an important tool to discuss optimality for $c'\theta$ is the convex and compact set $\mathscr{R} \subseteq \mathbb{R}^k$ which is the convex hull of the vectors $\pm f(x)$, with $x \in \mathscr{X}$. We introduce

$$\rho(c) = \inf\{\mu \ge 0 : c \in \mu\mathscr{R}\}, \qquad r = \min\{\|c\| \ge 0 : c \in \mathbb{R}^k, \rho(c) = 1\}.$$

We assume that there are k linearly independent vectors $f(x_1), \dots, f(x_k)$; then $\mathscr{R}$ has nonempty interior, and ρ is a norm on $\mathbb{R}^k$. The quantity r is the in-ball radius of $\mathscr{R}$, that is, the radius of the largest Euclidean ball inscribed in $\mathscr{R}$.

The central role of the Elfving set $\mathscr{R}$ is generally well known. For instance, for every design ξ there exists a design η whose support points lie in the boundary of $\mathscr{R}$ such that η is at least as good as ξ, that is, $M(\eta) \geqq M(\xi)$. This is Proposition III.7 of Pázman (1986), see also Fellman [(1974), Theorem 2.1.2] and Elfving [(1959), Theorem 4.3]. In the present paper we take a different approach in starting out, not from the boundary of $\mathscr{R}$, but from the in-ball radius r.

In Example 5 of Pukelsheim (1981) the E-optimal value for θ, $v = 1/2$, is strictly smaller than the squared in-ball radius, $r^2 = 1$. In other instances, the two are equal. The following theorem implies the general inequality $v \le r^2$.

THEOREM 2.2. *Every design ξ with information matrix C for $K'\theta$ and every vector $0 \neq z \in \mathbb{R}^s$ fulfill*

$$(1) \qquad \lambda_{\min}(C) \le \left(\frac{\|z\|}{\rho(Kz)}\right)^2.$$

If a design ξ and a vector $z \neq 0$ satisfy (1) *with equality, then ξ is E-optimal for $K'\theta$ and every E-optimal design $\tilde{\xi}$ for $K'\theta$ is also optimal for $z'K'\theta$.*

PROOF. Maximization of the smallest eigenvalue of C is the same as minimization of the largest eigenvalue of C^{-1}. We have the trivial inequalities

$$\begin{aligned}
\lambda_{\max}(C^{-1}) &\ge \min_{\eta:\, M(\eta)>0}\ \max_{\|z\|=1} z'C_K(M(\eta))^{-1}z \\
(2) \qquad &\ge \max_{\|z\|=1}\ \inf_{\eta:\, M(\eta)>0} z'K'M(\eta)^{-1}Kz \\
&= \max_{\|z\|=1} (\rho(Kz))^2.
\end{aligned}$$

The last equality follows from the Elfving (1952) result that generally the optimal value for $c'\theta$ is $(\rho(c))^2$, compare Studden [(1968), Theorem 2.1] or Pukelsheim [(1981), Theorem 1]. The final expression in (2) may be written in various ways:

$$\max_{\|z\|=1} (\rho(Kz))^2 = \max_{z\neq 0}\left(\frac{\rho(Kz)}{\|z\|}\right)^2 = \frac{1}{\min_{z\neq 0}(\|z\|/\rho(Kz))^2}.$$

This proves (1).

Now assume that ξ and z satisfy (1) with equality. Attaining the upper bound $(\|z\|/\rho(Kz))^2$, the design ξ is E-optimal for $K'\theta$. For every E-optimal design $\tilde{\xi}$ for $K'\theta$ we get

$$(3) \quad \left(\frac{\rho(Kz)}{\|z\|}\right)^2 \le \frac{1}{\|z\|^2} z'K'M(\tilde{\xi})^{-1}Kz \le \lambda_{\max}\left(C_K\left(M(\tilde{\xi})\right)^{-1}\right) = \left(\frac{\rho(Kz)}{\|z\|}\right)^2.$$

Hence $(\rho(Kz))^2 = z'K'M(\tilde{\xi})^{-1}Kz$, showing that $\tilde{\xi}$ is optimal for $z'K'\theta$. □

COROLLARY 2.3. *Let ξ be an E-optimal design for $K'\theta$ such that the smallest eigenvalue of its information matrix C for $K'\theta$ has multiplicity one, and let z be a corresponding eigenvector. Then C and z jointly satisfy* (1) *with equality.*

PROOF. The design ξ is also optimal for $z'K'\theta$, by Theorem 2.1. Hence we have $(\rho(Kz))^2 = z'K'M(\xi)^{-1}Kz = \|z\|^2/\lambda_{\min}(C)$. □

If ξ and z satisfy (1) with equality, then (3) implies, for all vectors $d \in \mathbb{R}^s$ with $\|d\| = 1$ and for all designs η with $M(\eta) > 0$,

$$d'K'M(\xi)^{-1}Kd \le \frac{1}{\|z\|^2} z'K'M(\xi)^{-1}Kz \le \frac{1}{\|z\|^2} z'K'M(\eta)^{-1}Kz.$$

That is, the pair $(z/\|z\|, M(\xi))$ is a saddle-point of the kernel $(d, M) \mapsto d'K'M^{-1}Kd$ that underlies (2). Notice, however, that this kernel is *convex in both variables*, see 16.E.7.f in Marshall and Olkin [(1979), page 469].

Theorem 2.2 applies to polynomial regression of degree 2, see Kiefer [(1974), page 868] or Pukelsheim [(1980), Example 6.2.2]. There, the vector $c = (-1, 0, 2)'$ fulfills $\|c\|/\rho(c) = 1/\sqrt{5}$. The moment matrix of the optimal design ξ for $c'\theta$ has smallest eigenvalue $1/5$. Hence ξ and c satisfy equality in (1), and ξ is E-optimal for θ.

Thus Theorem 2.2 suggests a reverse approach to the problem, to begin with a vector $z \neq 0$. Next find an optimal design ξ for $z'K'\theta$, the minimum variance provides the norm $\rho(Kz)$. Finally check whether for ξ and z equality holds in (1). If so, ξ is E-optimal for $K'\theta$.

This poses the question with which vector z to begin. For $K = I_k$, the minimum of the right-hand side in (1) is given by the squared in-ball radius, $\min_{c \neq 0}(\|c\|/\rho(c))^2 = r^2$. The minimum is attained by a vector c which defines a direction where the in-ball touches the boundary of the Elfving set $\mathscr{R}$.

3. *c*-Optimal polynomial regression designs. In a polynomial regression model of degree $d \ge 1$ the regression function is $f(x) = (1, x, \dots, x^d)'$. The experimental domain is taken to be $\mathscr{X} = [-1, 1]$. It is evident from Studden (1968) and the references given there that the Chebyshev polynomial $T_d(x) = \sum_{i=0}^d c_i x^i = c'f(x)$ plays a central role in the discussion. We call $c = (c_0, c_1, \dots, c_d)' \in \mathbb{R}^{d+1}$ the *Chebyshev coefficient vector*, and $s_i = \cos((d-i)\pi/d)$, for $i = 0, 1, \dots, d$, the *Chebyshev points*.

The Lagrange polynomials with nodes $s_0, s_1, \dots, s_d$ are

$$L_i(x) = \frac{\prod_{k \neq i}(x - s_k)}{\prod_{k \neq i}(s_i - s_k)} = \sum_{j=0}^{d} v_{ij} x^j = v_i' f(x), \quad \text{for all } i = 0, 1, \dots, d,$$

say. The two $(d+1) \times (d+1)$ matrices

$$A' = (v_0, v_1, \dots, v_d), \qquad B = (f(s_0), f(s_1), \dots, f(s_d))$$

are inverse to each other, see Karlin and Studden [(1966), page 336].

Any polynomial $P(x) = \sum_{j=0}^d a_j x^j$ satisfies $P(x) = \sum_{i=0}^d P(s_i) L_i(x)$. A comparison of coefficients yields

$$a_j = \sum_{i=0}^{d} P(s_i) v_{ij}, \quad \text{for all } j = 0, 1, \dots, d. \tag{4}$$

For the Chebyshev polynomial T_d we obtain $c_{d-2j} = \sum_{i=0}^{d}(-1)^{d-i}v_{i,d-2j}$. The sign pattern of the Chebyshev vector is known to be $c_{d-2j} = (-1)^j|c_{d-2j}|$. For our purposes the sign pattern of $v_{i,d-2j}$ becomes essential.

Multiplying out the products in the definition of the Lagrange polynomial we find, for all $i = 0, 1, \dots, d$ and $j = 0, 1, \dots, [d/2]$,

$$(5) \qquad v_{i,d-2j} = (-1)^{d-i+j} \frac{\sum_{I \subseteq \{0,1,\dots,[(d-1)/2]\} \setminus \{i,d-i\}, \#I=j} \prod_{k \in I} s_k^2}{\prod_{k \in \{0,1,\dots,d\} \setminus \{i\}} |s_i - s_k|}.$$

Thus the sign pattern of $v_{i,d-2j}$ is $(-1)^{d-i+j}$. This and (4) yield

$$(6) \qquad |c_{d-2j}| = \sum_{i=0}^{d} |v_{i,d-2j}|.$$

The numerator in (5) is nonzero unless the summation is empty; this occurs only if d is even and $j = d/2 \neq i$. Since numerator and denominator of (5) are symmetrical in i and $d-i$ we finally get

$$(7) \quad (-1)^{d-i+j} v_{i,d-2j} = |v_{i,d-2j}| = |v_{d-i,d-2j}| \begin{cases} = 0, & \text{for } j = d/2 \neq i, \\ > 0, & \text{otherwise.} \end{cases}$$

These properties were derived in Studden (1968) as a consequence of design optimality. Conversely, availability of these properties greatly facilitates the proof of design optimality.

Let the parameter system of interest be $\theta_I = \{\theta_i : i \in I\}$, where I is the s element index set $I = \{i_1, \dots, i_s\}$. With the Euclidean unit vectors $e_0, e_1, \dots, e_d$ of $\mathbb{R}^{d+1}$ and the $(d+1) \times s$ matrix $K = (e_{i_1}, \dots, e_{i_s})$, the parameter system of interest is represented as $K'\theta$. The matrix K fulfills $K'K = I_s$ and $KK' = \sum_{l=1}^{s} e_{i_l} e_{i_l}'$, the latter is a diagonal matrix D with d_{ii} being 1 or 0 according as i belongs to the set I or not.

We will relate the set I to the *Chebyshev index set* $\{d - 2j : j = 0, 1, \dots, [d/2]\}$, that is, the indices where the Chebyshev coefficients are nonzero. Our first essential assumption is that the parameters of interest contain at least one member from this set,

$$(8) \qquad J = \{j = 0, 1, \dots, [d/2] : d - 2j \in I\} \neq \emptyset.$$

This happens only if $KK'c \neq 0$. Notice that the vector $KK'c = \sum_{j \in J} c_{d-2j} e_{d-2j}$ depends on I only through J. The following theorem gives the optimal design for $c'KK'\theta = \sum_{j \in J} c_{d-2j} \theta_{d-2j}$.

THEOREM 3.1. *Under assumption* (8), *the only design ξ_I that is optimal for $c'KK'\theta$ has the Chebyshev points s_i for its support points, with weights*

$$(9) \qquad w_i = w_{d-i} = \frac{(-1)^{d-i} u_i}{\|K'c\|^2}, \quad \textit{for all } i = 0, 1, \dots, d,$$

and minimum variance $(\rho(KK'c))^2 = \|K'c\|^4$, *where the coefficients*

$u_0, u_1, \dots, u_d$ are determined from

$$(10) \qquad \sum_{i=0}^{d} u_i f(s_i) = KK'c.$$

If $J \neq \{d/2\}$, then all the weights are positive; if $J = \{d/2\}$, then $c'KK\theta = \theta_0$ and $w_{d/2} = 1$, that is, the one-point design at 0 is optimal for θ_0.

PROOF. The vector $u = (u_0, u_1, \dots, u_d)'$ solves $Bu = KK'c$. We get $u = A'KK'c$, that is, $u_i = \sum_{j \in J} v_{i,d-2j} c_{d-2j}$. This sum is nonempty by (8). Because of the sign pattern of $v_{i,d-2j}$ the weights are nonnegative,

$$(11) \quad (-1)^{d-i} u_i = (-1)^{d-i} \sum_{j \in J} v_{i,d-2j} c_{d-2j} = \sum_{j \in J} |v_{i,d-2j}|\, |c_{d-2j}| \geq 0.$$

Now (7) implies that the weights are all positive, unless $J \neq \{d/2\}$, as well as symmetric. Using $(-1)^{d-i} = T_d(s_i) = c'f(s_i)$ we find that they sum to one,

$$\sum_{i=0}^{d} (-1)^{d-i} u_i = c' \sum_{i=0}^{d} u_i f(s_i) = c'KK'c = \|K'c\|^2.$$

Let M be the moment matrix of ξ_I. The key step is the following:

$$(12) \quad Mc = \sum_{i=0}^{d} w_i f(s_i) f(s_i)'c = \frac{1}{\|K'c\|^2} \sum_{i=0}^{d} u_i f(s_i) = \frac{1}{\|K'c\|^2} KK'c.$$

Premultiplication by $c'KK'M^-$ gives $c'KK'M^-KK'c = \|K'c\|^4$.

The design problem, of minimizing $c'KK'M(\xi)^-KK'c$ over all designs ξ, is paired with the dual problem of maximizing $c'KK'NKK'c$, over those nonnegative definite matrices N that satisfy $f(x)'Nf(x) \leq 1$ for all $x \in [-1, 1]$. In particular, the choice $N = cc'$ satisfies $f(x)'cc'f(x) = (T_d(x))^2 \leq 1$, and the dual objective function takes the value $\|K'c\|^4$. Therefore ξ_I and N are optimal solutions of their respective problems.

Furthermore only points x with $f(x)'cc'f(x) = 1$, that is, the Chebyshev points s_i, can support an optimal design. Corollary 1 of Pukelsheim and Torsney (1991) provides the unique optimal weights, $w_i = |u_i| / \sum_{k=0}^{d} |u_k|$. □

We remark that two index sets I and H that contain the same members from the Chebyshev index set lead to the same set J in (8), and hence yield identical designs in the theorem, $\xi_I = \xi_H$.

The case $I = \{0, 1, \dots, d\}$ gives the optimal design ξ_c for $c'\theta$, with minimum variance $\|c\|^4$. The cases $I = \{d - 2j\}$ yield the optimal designs ξ_{d-2j} for the individual parameters θ_{d-2j}, with minimum variance $(\rho(e_{d-2j}))^2 = c_{d-2j}^2$, given in Section 4 of Studden (1968). One has the relation

$$\xi_c(s_i) = \frac{(-1)^{d-i} v_i'c}{\|c\|^2} = \sum_{j=0}^{[d/2]} \frac{c_{d-2j}^2}{\|c\|^2} \frac{(-1)^{d-i} v_{i,d-2j}}{c_{d-2j}} = \sum_{j=0}^{[d/2]} \frac{c_{d-2j}^2}{\|c\|^2} \xi_{d-2j}(s_i).$$

Therefore ξ_c is a mixture of the designs ξ_{d-2j}, with mixing weights $c_{d-2j}^2 / \|c\|^2$.

From (12) we obtain the Elfving representation

$$KK'c/\rho(KK'c) = \sum_{i=0}^{d} (-1)^{d-i} w_i f(s_i), \qquad \rho(KK'c) = \|K'c\|^2.$$

Thus the coefficient vector $KK'c$ penetrates the Elfving set $\mathscr{R}$ through the face generated by the "alternating regression vectors" $(-1)^d f(s_0)$, $(-1)^{d-1} f(s_1), \dots, -f(s_{d-1})$, $f(s_d)$. Theorem 5.1 of Studden (1968) provides a related result.

The following corollary sets the stage for E-optimality when the parameters of interest are $K'\theta$.

COROLLARY 3.2. *The design ξ_I has an information matrix C for $K'\theta$ that has $K'c$ as an eigenvector corresponding to the eigenvalue $\|K'c\|^{-2}$.*

PROOF. We introduce the residual projector $R = I_{d+1} - KK'$. Let M be the moment matrix of ξ_I. Then we can represent the information matrix as

$$C = K'MK - K'MR(RMR)^- RMK.$$

Postmultiplication by $K'c$ gives, using $MKK'c = Mc - MRc$ and replacing Mc according to (12),

$$CK'c = \frac{1}{\|K'c\|^2} K'c. \qquad \square$$

4. *E*-Optimal polynomial regression designs. For E-optimality we need to ensure that the smallest eigenvalue comes from the block corresponding to the Chebyshev index set. Our second essential assumption is sufficient to secure this, demanding that any non-Chebyshev index is accompanied by the Chebyshev index following it,

$$(13) \qquad d - 1 - 2j \in I \quad \Rightarrow \quad d - 2j \in I,$$

for all $j = 0, 1, \dots, [d/2]$. Since the scalar case is taken care of by Theorem 3.1, we assume that I contains at least two indices. Because of (13) the set J from (8) then cannot degenerate to $\{d/2\}$. Our main result is that the design ξ_I from Theorem 3.1 is E-optimal for $K'\theta = \theta_I$.

THEOREM 4.1. *Under assumptions* (8) *and* (13), *the design ξ_I from Theorem* 3.1 *is the only E-optimal design for $K'\theta$, with optimal value $\|K'c\|^{-2}$. If $d > 1$, then the smallest eigenvalue of the information matrix $C = C_K(M(\xi_I))$ has multiplicity one.*

PROOF. (i) From Theorem 3.1 we know $\rho(KK'c) = \|K'c\|^2$. In view of (1) of Theorem 2.2 we show that $(\|K'c\|/\rho(KK'c))^2 = \|K'c\|^{-2}$ is the smallest eigenvalue of C,

$$z'Cz \geq \frac{\|z\|^2}{\|K'c\|^2}, \quad \text{for all } z \in \mathbb{R}^s,$$

with equality if and only if z is proportional to $K'c$. Let M be the moment matrix of ξ_I. There exists a left inverse L of K such that the information matrix can be represented as $C = LML'$. Hence we wish to show that

$$\|K'c\|^2 a'Ma \geq \|z\|^2, \quad \text{where } a = L'z. \tag{14}$$

Because of symmetry of the design ξ_I the odd moments in M vanish, and M decomposes into two interlacing block matrices. Therefore $\|K'c\|^2 a'Ma$ can be written as

$$\sum_{i=0}^{d}(-1)^{d-i}u_i(a'f(s_i))^2 = \sum_{i=0}^{d}(-1)^{d-i}u_i\big((P(s_i))^2 + (Q(s_i))^2\big),$$

where P and Q are the polynomials associated with the Chebyshev index set and its complement,

$$P(x) = \sum_{j=0}^{[d/2]} a_{d-2j}x^{d-2j}, \qquad Q(x) = \sum_{j=0}^{[(d-1)/2]} a_{d-1-2j}x^{d-1-2j}.$$

The contributions from P^2 and Q^2 are discussed separately.

(ii) From (11) and (6) we get

$$\begin{aligned}
\sum_{i=0}^{d}(P(s_i))^2(-1)^{d-i}u_i &= \sum_{i=0}^{d}(P(s_i))^2\sum_{j\in J}|v_{i,d-2j}|\,|c_{d-2j}| \\
&= \sum_{j\in J}\left(\sum_{i=0}^{d}(P(s_i))^2|v_{i,d-2j}|\right)\left(\sum_{i=0}^{d}|v_{i,d-2j}|\right) \\
&\geq \sum_{j\in J}\left(\sum_{i=0}^{d}P(s_i)\sqrt{|v_{i,d-2j}|}\,(-1)^{d-i+j}\sqrt{|v_{i,d-2j}|}\right)^2 \qquad (15)\\
&= \sum_{j\in J}\left(\sum_{i=0}^{d}P(s_i)v_{i,d-2j}\right)^2 \\
&= \sum_{j\in J}a_{d-2j}^2,
\end{aligned}$$

where in the last line we have applied (4).

If equality holds in the Cauchy inequalities in (15) then we need exploit only any one index $j \neq d/2$ to obtain proportionality, for $i = 0, 1, \ldots, d$, of $P(s_i)\sqrt{|v_{i,d-2j}|}$ and $(-1)^{d-i}\sqrt{|v_{i,d-2j}|}$. Because of $v_{i,d-2j} \neq 0$ this entails $P(s_i) = \alpha(-1)^{d-i}$, for some $\alpha \in \mathbb{R}$. Hence equality holds in (15) if and only if $P = \alpha T_d$, that is, $a_{d-2j} = \alpha c_{d-2j}$ for all $j = 0, 1, \ldots, [d/2]$.

(iii) The argument for Q^2 is reduced to that in part (ii), by introducing

$$xQ(x) = \sum_{j=0}^{[(d-1)/2]} a_{d-1-2j} x^{d-2j} = \tilde{P}(x),$$

say. From $s_i^2 \le 1$ and (15) we get the two estimates

$$(16) \qquad \sum_{i=0}^{d} (Q(s_i))^2 (-1)^{d-i} u_i \ge \sum_{i=0}^{d} s_i^2 (Q(s_i))^2 (-1)^{d-i} u_i,$$

$$(17) \qquad \sum_{i=0}^{d} (\tilde{P}(s_i))^2 (-1)^{d-i} u_i \ge \sum_{j \in J} a_{d-1-2j}^2.$$

If d is even and $j = d/2$, then the last sum involves the coefficient of x^0 in $\tilde{P}$ which is $a_{-1} = 0$.

Equality holds in (17) only if, for some $\beta \in \mathbb{R}$, we have $\tilde{P} = \beta T_d$. In case $d > 1$ equality holds in (16) only if $(Q(s_i))^2 = s_i^2(Q(s_i))^2$. Any one index $i = 1, \dots, d-1$ entails $Q(s_i) = 0 = \tilde{P}(s_i) = (-1)^{d-i}\beta$. Hence equality obtains in (16) as well as in (17) if and only if $a_{d-1-2j} = 0$ for all $j = 0, 1, \dots, [(d-1)/2]$.

(iv) Combining (15), (16) and (17), and now utilizing assumption (13) we obtain

$$\begin{aligned}\|K'c\|^2 a'Ma &\ge \sum_{j \in J} \left(a_{d-2j}^2 + a_{d-1-2j}^2\right)\\ &\ge \sum_{i \in I} a_i^2\\ &= a'KK'a.\end{aligned}$$

With $a = L'z$ and $LK = I_s$ we get $a'KK'a = \|z\|^2$. Hence (14) is established. If $d > 1$, then equality holds in (14) if and only if $a = \alpha c$, that is, $z = \alpha K'c$.

(v) If $\tilde{\xi}$ is another E-optimal design for $K'\theta$, then $\tilde{\xi}$ is also optimal for $c'KK'\theta$, by Theorem 2.2. Hence the uniqueness statement of Theorem 3.1 carries over to E-optimality for $K'\theta$. □

It is an immediate consequence of the theorem that the design ξ_c that is optimal for $c'\theta$ is the unique E-optimal design for the full parameter vector θ, and that the smallest eigenvalue of its moment matrix is $\|c\|^{-2}$.

5. Discussion. It follows from Theorems 2.2 and 4.1 that for polynomial regression on $[-1, 1]$ the Elfving set $\mathscr{R}$ has in-ball radius $\|c\|^{-2}$, and that the Chebyshev coefficient vector c defines the direction where the in-ball touches the boundary of $\mathscr{R}$. This can also be obtained from the following extremal property of the Chebyshev polynomial.

LEMMA 5.1. *For every vector* $a \in \mathbb{R}^{d+1}$ *with* $\max_{i=0,1,\dots,d}(a'f(s_i))^2 \le 1$ *we have the inequality* $\|a\|^2 \le \|c\|^2$. *If* $d > 1$, *then equality holds if and only if* $a = \pm c$.

PROOF. Owing to a result of Erdös [(1947), pages 1175, 1176], the polynomial $P(x) = a'f(x)$ satisfies $|P(z)| \leq |T_d(z)|$, for all complex numbers z with $|z| \geq 1$. Hence integration relative to Lebesgue measure on the complex unit sphere yields

$$\|a\|^2 = \frac{1}{2\pi}\int_0^{2\pi}|P(e^{i\varphi})|^2\,d\varphi \leq \frac{1}{2\pi}\int_0^{2\pi}|T_d(e^{i\varphi})|^2\,d\varphi = \|c\|^2.$$

Equality necessitates $|T_d(z)| = |P(z)|$, for all z in the complex unit sphere. With $z = \sqrt{-1}$, the arguments of Erdös show that equality holds only if $a = \pm c$. □

The supporting hyperplanes to $\mathscr{R}$ are given by the vectors $0 \neq a \in \mathbb{R}^{d+1}$ such that $\max_{x\in[-1,1]}|a'f(x)| = 1$. Lemma 5.1 yields $\|a\|^2 \leq \|c\|^2$. The hyperplane $\{v \in \mathbb{R}^{d+1}: a'v = 1\}$ has distance $1/\|a\|$ to the origin. Therefore the supporting hyperplane closest to the origin is given by the Chebyshev coefficient vector c, and has distance $r \doteq 1/\|c\|$. This and Theorem 2.2 provided the starting point to show that the design ξ_c that is optimal for $c'\theta$ is also E-optimal for θ.

However, in retrospect, Theorem 4.1 provides a result stronger than Lemma 5.1, in that it refers P^2 over $s_0, s_1, \dots, s_d$ only to its average with respect to the measure ξ_c, rather than to its maximum.

COROLLARY 5.2. *For every vector $a \in \mathbb{R}^{d+1}$ with $\int_{[-1,1]}(a'f(x))^2\,d\xi_c \leq 1$, we have the inequality $\|a\|^2 \leq \|c\|^2$. If $d > 1$, then equality holds if and only if $a = \pm c$.*

PROOF. For $K = I_{d+1}$, the eigenvalue property (14) is $a'Ma \geq \|a\|^2/\|c\|^2$. This is equivalent to $1 \geq \|a\|^2/\|c\|^2$, for all vectors $a \in \mathbb{R}^{d+1}$ satisfying $a'Ma \leq 1$. □

Yet another form of our result pertains to the polynomials $\tilde{P}(x) = a'f(x)/\|a\|$ that are standardized so that their coefficient vector has Euclidean norm one. Then (14) becomes

$$\int_{[-1,1]}\left(\tilde{P}(x)\right)^2 d\xi_c \geq \int_{[-1,1]}\left(\tilde{T}_d(x)\right)^2 d\xi_c,$$

exhibiting a least squares property relative to ξ_c of the standardized Chebyshev polynomial which is complementary to the usual least square property relative to the arcsin distribution, see, for instance, Rivlin [(1990), page 42].

Moreover, Theorem 3.1 rederives and extends the classical extremum property of Chebyshev polynomials. Namely, one has

$$\begin{aligned} \|K'c\|^{-4} &= \max_{\xi:\, M(\xi)>0} \left(c'KK'M(\xi)^{-1}KK'c\right)^{-1} \\ &= \max_{\xi} \min_{a'KK'c=1} a'M(\xi)a \\ &= \min_{a'KK'c=1} \max_{\xi} a'M(\xi)a \\ &= \min_{a'KK'c=1} \max_{x\in[-1,1]} (a'f(x))^2. \end{aligned}$$

Hence among all polynomials $P(x) = a'f(x)$ that satisfy $a'KK'c = 1$, the sup-norm $\max_{x\in[-1,1]}|P(x)|$ has minimum value $\|K'c\|^{-2}$, and this minimum is attained only by the standardized Chebyshev polynomial $T_d/\|K'c\|^2$. For the highest index d this result is due to Chebyshev, for $d - 2j$ to Markoff (1916), see Natanson [(1955), pages 36, 50] or Rivlin [(1990), pages 67, 112]. Our formulation also allows for combinations of those coefficients.

It is not hard to provide an analogue of Theorem 3.1 that covers optimality for the remaining individual coefficients θ_{d-1-2j}, see Pukelsheim (1992). As pointed out by Studden (1968), the optimal design for θ_{d-1-2j} is supported by the Chebyshev points of one degree lower.

For large degree regression $d \to \infty$, the E-optimal designs ξ_c converge weakly to the distribution with Lebesgue density

$$\frac{\sqrt{2}}{\pi(1+x^2)\sqrt{1-x^2}}, \qquad -1 < x < 1.$$

If $m(d)$ is the index of the maximum absolute entry of the Chebyshev coefficient vector, $|c_{m(d)}| = \max_{j=0,1,\dots,[d/2]}|c_{d-2j}|$, then $\lim_{d\to\infty} m(d)/d = 1/\sqrt{2}$. The limiting density is the member for $q = 1/\sqrt{2}$ of the family on page 1395 of Studden (1978).

That some assumption like (13) is needed is evidenced by Preitschopf (1989). He quotes $d = 3$ and $I = \{0, 1, 2\}$ as an instance where the E-optimal design for $K'\theta$ is *not* supported by the Chebyshev points.

The majority of Preitschopf's tabulations are for the interval $[0, 1]$. They illustrate that our Theorem 4.1 does not carry over to this interval. For example, for $d = 4$ and the four subsets $\{0, 1, 2, 3, 4\}$, $\{0, 1, 2, 4\}$, $\{0, 2, 3, 4\}$, $\{0, 2, 4\}$—all having the same set $J = \{0, 2, 4\}$—the E-optimal designs for $K'\theta$ are distinct on $[0, 1]$, whereas on $[-1, 1]$ our Theorem 4.1 proves them to be identical.

However, on $[0, 1]$ or any other positive interval, the solution is more transparent since the oscillatory polynomial $W_d(x) = \sum_{i=0}^{d} w_i x^i = w'\theta$ that replaces T_d has all nonzero alternating sign coefficients and the analogue of the Chebyshev index set is simply the full set $\{0, 1, \dots, d\}$. It can be shown [see also Theorem 5.1 of Murty (1971)] that the optimal design ξ_I for $w'KK'\theta$ for every index set I (including the designs ξ_l for the individual coefficients) is supported by the extreme points of W_d. The same will apply to the E-optimal design ξ_I for θ_I where now the interlacing block structure of $M(\xi_I)$ needed in the proof of our Theorem 4.1 is no longer prevalent. All these designs are supported on the full set of extreme points of W_d with the lone exception of the single coefficient θ_0 where the design concentrates all its mass at zero when the interval [0, 1] is under consideration.

Acknowledgments. We would like to thank T. J. Rivlin for the proof of Lemma 5.1, A. Wilhelm for supplying numerical evidence toward Theorem 4.1 and N. Gaffke for alerting us to the work of B. Heiligers.

REFERENCES

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255–262.

ELFVING, G. (1959). Design of linear experiments. In *Probability and Statistics. The Harald Cramér Volume* (U. Grenander, ed.) 58–74. Almgrist & Wiksell, Stockholm.

ERDÖS, P. (1947). Some remarks on polynomials. *Bull. Amer. Math. Soc.* **53** 1169–1176.

FELLMAN, J. (1974). On the allocation of linear observations. *Soc. Sci. Fenn. Comment. Phys.-Math.* **44** 27–77.

GAFFKE, N. (1987). Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression. *Ann. Statist.* **15** 942–957.

HEILIGERS, B. (1991). *E-optimal Polynomial Regression Designs.* Habilitationsschrift, RWTH Aachen.

HOEL, P. G. (1958). Efficiency problems in polynomial estimation. *Ann. Math. Statist.* **29** 1134–1145.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Wiley, New York.

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

MARKOFF, W. (1916). Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen. *Math. Ann.* **77** 213–258.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

MURTY, V. N. (1971). Optimal designs with a Tchebycheffian spline regression function. *Ann. Math. Statist.* **42** 643–649.

NATANSON, I. P. (1955). *Konstruktive Funktionentheorie.* Akademie-Verlag, Berlin.

PÁZMAN, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

PREITSCHOPF, F. (1989). Bestimmung optimaler Versuchspläne in der polynomialen Regression. Dissertation, Univ. Augsburg.

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* **4** 339–364.

PUKELSHEIM, F. (1981). On c-optimal design measures. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 13–20.

PUKELSHEIM, F. (1992). *Optimality Theory of Experimental Design in Linear Models.* Wiley, New York. To appear.

PUKELSHEIM, F. and TORSNEY, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Ann. Statist.* **19** 1614–1625.

RIVLIN, T. J. (1990). *Chebyshev Polynomials. From Approximation Theory to Algebra and Number Theory*, 2nd ed. Wiley, New York.

STUDDEN, W. J. (1968). Optimal designs on Tchebycheff points. *Ann. Math. Statist.* **39** 1435–1447.

STUDDEN, W. J. (1978). Designs for large degree polynomial regression. *Comm. Statist. A—Theory Methods* **7** 1301–1397.

INSTITUT FÜR MATHEMATIK
UNIVERSITÄT AUGSBURG
D-8900 AUGSBURG
GERMANY

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907