The Annals of Statistics
2001, Vol. 29, No. 3, 763–783

E-OPTIMAL DESIGNS FOR RATIONAL MODELS

BY LORENS A. IMHOF[1] AND WILLIAM J. STUDDEN[2]

Stanford University and Purdue University

E-optimal and standardized-E-optimal designs for various types of rational regression models are determined. In most cases, optimal designs are found for every parameter subsystem. The design points and weights are given explicitly in terms of Bernstein-Szegő polynomials. The analysis is based on a general theorem on E-optimal designs for Chebyshev systems.

1. Introduction. Rational functions have appealing approximation properties and are widely used in regression models; see Petrushev and Popov (1987) for an extensive review of rational approximation and Ratkowsky (1990) for a discussion from a more applied point of view. However, it is only recently that optimal designs for rational models have been found. Haines (1992) determined D-optimal designs for an inverse quadratic model and He, Studden and Sun (1996) and Dette, Haines and Imhof (1999) investigated D-optimal designs for more general rational models. This paper is concerned with the construction of E-optimal designs for several types of rational models. They include

$$E(Y(x)) = \theta_0 + \frac{\theta_1}{x-\alpha_1} + \cdots + \frac{\theta_n}{x-\alpha_n}, \qquad x \in [-1,1], \tag{1.1}$$

where $\alpha_1, \dots, \alpha_n \in \mathbb{R} \setminus [-1,1]$; and

$$E(Y(x)) = \theta_{-m} x^{-m} + \cdots + \theta_n x^n, \qquad x \in [-b,-a] \cup [a,b], \tag{1.2}$$

where $0 < a < b$ and $m, n \geq 0$. Here $Y(x)$ is the response to the controlled variable x and the θ_j are unknown parameters.

In Section 2 we prove a general theorem on E-optimal designs for linear regression models. The theorem extends results of Studden (1968) on c-optimal designs. It gives a complete solution to the E-optimal design problem for parameter subsystems provided that the underlying regression functions and certain subsets of these functions form Chebyshev systems. For example, every Descartes system is covered. The E-optimal designs are expressed in terms of the Chebyshev polynomial of the regression system. In addition to the ordinary E-optimal designs, the theorem provides also standardized-E-optimal designs [Dette (1997a,b)]. The standardization takes into account that the sizes of the variances and covariances of the least-squares estimator may be

Received February 2000; revised January 2001.
[1]Supported in part by DFG.
[2]Supported by NSF Grant DMS-96-6784.
AMS 2000 *subject classification*. 62K05.
Key words and phrases. Approximate designs, Bernstein-Szegő polynomials, Chebyshev systems, E-criterion, standardized criterion.

very different. In Sections 3 and 4 the results of Section 2 are applied to several rational models. Here we use Bernstein-Szegő orthogonal polynomials to obtain explicit representations of the corresponding Chebyshev polynomials.

For model (1.1), we obtain the E-optimal and standardized-E-optimal designs for any given parameter subset. We obtain also a complete solution for two other rational models on a compact interval. For model (1.2), our findings are less complete. Here the parameter subset has to satisfy some condition and the design space must not be too large. These restrictions should not come as a surprise in view of what is known about E-optimal designs for polynomial models [Dette (1993), Heiligers (1994, 1996), Pukelsheim (1993) and Pukelsheim and Studden (1993)]. On the other hand, the E-optimal designs derived in this paper have some properties that E-optimal designs for polynomial regression on an interval do not share. For example, the choice of the regression interval for model (1.1) is merely a matter of convenience; all arguments used are easily extended to arbitrary intervals $[a, b]$ provided that $\alpha_1, \dots, \alpha_n \notin [a, b]$. A similar remark applies to the other rational models considered in Section 3. Recall that E-optimal designs for polynomial regression depend critically on the regression interval: the well-known results for $[-1, 1]$ do not carry over to arbitrary intervals and even symmetric intervals require in general a different approach; see Melas (2000). Moreover, depending on m and n, the number of support points of the E-optimal designs for model (1.2) can be smaller or larger than the number of parameters, and the E-optimal designs are not always unique.

NOTATION. Let $\mathbf{f} = (f_0, \dots, f_n)^T$ be a vector of linearly independent continuous functions on a compact set $\mathscr{X}$. Suppose that for each $x \in \mathscr{X}$ a random variable $Y(x)$ can be observed with expectation

$$E(Y(x)) = \theta^T \mathbf{f}(x) \tag{1.3}$$

and variance σ^2 not depending on x. Here $\theta = (\theta_0, \dots, \theta_n)^T$ is a vector of unknown parameters. An approximate design ξ is a probability measure on $\mathscr{X}$. If ξ has finite support $\{x_0, \dots, x_k\}$, then the observations on $Y(x)$ are made at $x_0, \dots, x_k$ with frequencies approximately proportional to $\xi(x_0), \dots, \xi(x_k)$. All observations are assumed to be uncorrelated. For every design ξ, let $M(\xi) = \int \mathbf{f}(x)\mathbf{f}^T(x)\, d\xi(x)$. Suppose that $K^T\theta$ is the parameter system of interest, where $K = [\mathbf{k}_1, \dots, \mathbf{k}_s] \in \mathbb{R}^{(n+1)\times s}$ and $\operatorname{rank}(K) = s$. A design ξ is feasible for $K^T\theta$ if $\operatorname{range}(K) \subset \operatorname{range}(M(\xi))$ or, equivalently, if $K = M(\xi)M^+(\xi)K$, where $M^+(\xi)$ is the Moore-Penrose inverse of $M(\xi)$. For every feasible design ξ, the information matrix for $K^T\theta$ is

$$C_K(\xi) = C_K(M(\xi)) = (K^T M^+(\xi) K)^{-1}.$$

A design ξ is E-optimal for $K^T\theta$ if it maximizes $\lambda_{\min}(C_K(\xi))$, the minimum eigenvalue of $C_K(\xi)$. Let ξ_j^* be the optimal design for estimating the linear combination $\mathbf{k}_j^T\theta$, that is, ξ_j^* minimizes $\mathbf{k}_j^T M^+(\xi)\mathbf{k}_j$ among all feasible designs for $\mathbf{k}_j^T\theta$. Let Δ be the $s \times s$ diagonal matrix with entries $(\mathbf{k}_j^T M^+(\xi_j^*)\mathbf{k}_j)^{-1/2}$,

$j = 1, \ldots, s$. Then the standardized information matrix for $K^T\theta$ as introduced by Dette (1997a,b) is defined by

$$\hat{C}_K(\xi) = \hat{C}_K(M(\xi)) = (\Delta K^T M^+(\xi) K \Delta)^{-1}.$$

A standardized-E-optimal design for $K^T\theta$ maximizes $\lambda_{\min}(\hat{C}_K(\xi))$. Note that the standardized-E-optimal design for $K^T\theta$ is invariant with respect to multiplication of the columns of K by non-zero numbers [Dette (1997b), Theorem 3.2].

A set of continuous functions $f_0, \ldots, f_n : [a, b] \to \mathbb{R}$ is a weak Chebyshev system if there is an $\varepsilon \in \{+1, -1\}$ such that

$$\varepsilon \det \begin{bmatrix} f_0(x_0) \cdots f_0(x_n) \\ \vdots \qquad \vdots \\ f_n(x_0) \cdots f_n(x_n) \end{bmatrix} \geq 0 \tag{1.4}$$

for all $b \geq x_0 > \cdots > x_n \geq a$. If inequality (1.4) is always strict, then $\{f_0, \ldots, f_n\}$ is a Chebyshev system. For every Chebyshev system $\{f_0, \ldots, f_n\}$, there is a unique polynomial $\sum_{j=0}^n c_j f_j(x) = \mathbf{c}^T \mathbf{f}(x)$ such that:

(i) $|\mathbf{c}^T\mathbf{f}(x)| \leq 1$ for all x and

(ii) there are $n + 1$ points $s_0 > \cdots > s_n$ such that $\mathbf{c}^T\mathbf{f}(s_k) = (-1)^k$ for $k = 0, \ldots, n$;

see Karlin and Studden (1966), Theorem II.10.2. The polynomial $\mathbf{c}^T\mathbf{f}(x)$ is called the Chebyshev polynomial and $s_0, \ldots, s_n$ are the Chebyshev points. They need not be unique. They are unique if $1 \in \operatorname{span}\{f_0, \ldots, f_n\}$ and $n \geq 1$. In this case $s_0 = b$ and $s_n = a$.

2. *E*-optimal designs for Chebyshev systems. Theorem 2.1 shows for a wide class of Chebyshev systems that the E-optimal designs are supported by Chebyshev points and gives an explicit expression for the weights. Thus in applications one still has to determine the Chebyshev points, which can be done numerically with the Remez algorithm; cf. Studden and Tsay (1976), DeVore and Lorentz [(1993), Section 3.8] and Heiligers (1996).

THEOREM 2.1. *Consider model* (1.3). *Suppose $\{f_0, \ldots, f_n\}$ is a Chebyshev system on $\mathscr{X} = [a, b]$ and every subsystem that consists of n of these functions is a weak Chebyshev system. Let $\mathbf{c}^T\mathbf{f}(x)$ be the Chebyshev polynomial with Chebyshev points $s_0 > \cdots > s_n$. Suppose the Chebyshev points are unique. Set $F = (f_j(s_k))_{jk=0}^n$ and $J = \operatorname{diag}(1, -1, \ldots, (-1)^n)$.*

(a) *For every weighted parameter subsystem $K^T\theta = (k_{j_1}\theta_{j_1}, \ldots, k_{j_s}\theta_{j_s})^T$, there is a unique E-optimal design ξ; the design is concentrated on the Chebyshev points, the weights are given by*

$$(\xi(s_0), \ldots, \xi(s_n))^T = \frac{1}{\|K^T\mathbf{c}\|^2} J F^{-1} K K^T \mathbf{c}, \tag{2.1}$$

and $\lambda_{\min}(C_K(M(\xi))) = \|K^T\mathbf{c}\|^{-2}$. *If* ξ_{j_μ} *denotes the optimal design for estimating the single coefficient* θ_{j_μ}, $\mu = 1, \ldots, s$, *then*

$$\xi = \frac{1}{\|K^T\mathbf{c}\|^2} \sum_{\mu=1}^{s} k_{j_\mu}^2 c_{j_\mu}^2 \xi_{j_\mu}. \tag{2.2}$$

(b) *For every parameter subsystem* $(\theta_{j_1}, \ldots, \theta_{j_s})^T$, *there is a unique standardized-E-optimal design* ξ'; *the design is equal to the E-optimal design for* $(c_{j_1}^{-1}\theta_{j_1}, \ldots, c_{j_s}^{-1}\theta_{j_s})^T$, *and*

$$\xi' = \frac{1}{s} \sum_{\mu=1}^{s} \xi_{j_\mu}.$$

PROOF OF THEOREM 2.1. (a) Suppose that $\det F > 0$ and let $\varepsilon_0, \ldots, \varepsilon_n \in \{-1, +1\}$ be such that

$$\varepsilon_k \det \begin{bmatrix} f_0(x_0) & \cdots & f_{k-1}(x_0) & f_{k+1}(x_0) & \cdots & f_n(x_0) \\ \vdots & & \vdots & \vdots & & \vdots \\ f_0(x_{n-1}) & \cdots & f_{k-1}(x_{n-1}) & f_{k+1}(x_{n-1}) & \cdots & f_n(x_{n-1}) \end{bmatrix} \geq 0$$

for all $b \geq x_0 > \cdots > x_{n-1} \geq a$ and $k = 0, \ldots, n$. Then the sign of the element in position (j, k) of F^{-1} is $(-1)^{j+k}\varepsilon_k$ or zero. Thus

$$JF^{-1}JS \geq 0, \tag{2.3}$$

where $S = \operatorname{diag}(\varepsilon_0, \ldots, \varepsilon_n)$. Here and in what follows inequality signs between matrices or vectors apply componentwise. Let $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^{n+1}$. Since $\mathbf{c}^T F = \mathbf{1}^T J$ and since every column of F^{-1} must have at least one non-zero entry, it follows from (2.3) that

$$\mathbf{c}^T JS = \mathbf{1}^T JF^{-1}JS > \mathbf{0}. \tag{2.4}$$

Let ξ be defined by (2.1) and $\xi([a, b]\backslash\{s_0, \ldots, s_n\}) = 0$. Note that ξ is indeed a probability measure:

$$\sum_{j=0}^{n} \xi(s_j) = \frac{1}{\|K^T\mathbf{c}\|^2}\mathbf{1}^T JF^{-1}KK^T\mathbf{c} = \frac{\mathbf{c}^T KK^T\mathbf{c}}{\|K^T\mathbf{c}\|^2} = 1,$$

and by (2.3) and (2.4),

$$(\xi(s_0), \ldots, \xi(s_n))^T = \frac{1}{\|K^T\mathbf{c}\|^2} JF^{-1}JSKK^T SJ\mathbf{c} \geq \mathbf{0}.$$

If $\xi(s_j) = 0$, then the jth row of $JF^{-1}JSK$ must vanish, so that the jth row of $F^{-1}K$ must vanish, too. Thus $M(\xi)M^+(\xi)K = FWW^+F^{-1}K = K$, where $W = \operatorname{diag}(\xi(s_0), \ldots, \xi(s_n))$. That is, ξ is feasible for $K^T\theta$. Moreover, by (2.1), $WJ\mathbf{1} = \|K^T\mathbf{c}\|^{-2}F^{-1}KK^T\mathbf{c}$, and so

$$\begin{aligned} K^T\mathbf{c} &= K^T M^+(\xi)M(\xi)\mathbf{c} = K^T M^+(\xi)FWF^T\mathbf{c} = K^T M^+(\xi)FWJ\mathbf{1} \\ &= \frac{1}{\|K^T\mathbf{c}\|^2} K^T M^+(\xi)KK^T\mathbf{c} = \frac{1}{\|K^T\mathbf{c}\|^2} C_K^{-1}(\xi)K^T\mathbf{c}. \end{aligned} \tag{2.5}$$

Let U be the $s \times s$ diagonal matrix with diagonal entries ± 1 such that $UK^T = K^T SJ$. It then follows from (2.5) that $UK^T\mathbf{c}$ is an eigenvector of $UC_K^{-1}(\xi)U$ corresponding to the eigenvalue $\|K^T\mathbf{c}\|^2$. Assume without loss of generality that $K \geq 0$. Then, by (2.3),

$$UC_K^{-1}(\xi)U = K^T SJ(F^T)^{-1}JW^+JF^{-1}JSK \geq 0,$$

and by (2.4), $UK^T\mathbf{c} = K^T SJ\mathbf{c} > \mathbf{0}$. It therefore follows from Lemma 2.1 below that $\|K^T\mathbf{c}\|^2 = \lambda_{\max}(UC_K^{-1}(\xi)U) = \lambda_{\max}(C_K^{-1}(\xi))$.

Now let η be any feasible design for $K^T\theta$. Then $K = M^{1/2}(\eta)(M^+(\eta))^{1/2}K$, and so, by Cauchy's inequality,

$$\begin{aligned}(\mathbf{c}^T KK^T\mathbf{c})^2 &\leq \mathbf{c}^T M(\eta)\mathbf{c} \ \cdot \ \mathbf{c}^T KK^T M^+(\eta)KK^T\mathbf{c}\\ &= \int(\mathbf{c}^T\mathbf{f}(x))^2\,d\eta(x) \ \cdot \ \mathbf{c}^T KC_K^{-1}(\eta)K^T\mathbf{c}\\ &\leq 1 \ \cdot \ \lambda_{\max}(C_K^{-1}(\eta))\mathbf{c}^T KK^T\mathbf{c}.\end{aligned}$$

Hence

$$\lambda_{\max}(C_K^{-1}(\eta)) \geq \|K^T\mathbf{c}\|^2 = \lambda_{\max}(C_K^{-1}(\xi)). \tag{2.6}$$

That is, ξ is an E-optimal design. To show uniqueness, suppose $\lambda_{\max}(C_K^{-1}(\eta)) = \|K^T\mathbf{c}\|^2$. Then $\operatorname{supp}(\eta) \subset \{s_0, \dots, s_n\}$, and $M^{1/2}(\eta)\mathbf{c}$ and $(M^+(\eta))^{1/2}KK^T\mathbf{c}$ must be proportional. Thus $M\mathbf{c} = F\operatorname{diag}(\eta(s_0), \dots, \eta(s_n))F^T\mathbf{c}$ is proportional to $KK^T\mathbf{c}$. This determines the weights of η uniquely.

To prove the representation (2.2) note that

$$(\xi_{j_\mu}(s_0), \dots, \xi_{j_\mu}(s_n))^T = \frac{1}{c_{j_\mu}^2} JF^{-1}\mathbf{e}_{j_\mu}\mathbf{e}_{j_\mu}^T\mathbf{c},$$

where $\mathbf{e}_{j_\mu}$ denotes the j_μth unit vector in $\mathbb{R}^{n+1}$. Hence

$$\begin{aligned}(\xi(s_0), \dots, \xi(s_n))^T &= \frac{1}{\|K^T\mathbf{c}\|^2} JF^{-1}KK^T\mathbf{c}\\ &= \frac{1}{\|K^T\mathbf{c}\|^2} \sum_{\mu=1}^{s} k_{j_\mu}^2 c_{j_\mu}^2 (\xi_{j_\mu}(s_0), \dots, \xi_{j_\mu}(s_n))^T.\end{aligned}$$

(b) By (a), $\mathbf{e}_{j_\mu}^T M^+(\xi_{j_\mu})\mathbf{e}_{j_\mu} = c_{j_\mu}^2$. Thus for $K = [\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_s}]$, the standardized information matrix $\hat{C}_K$ is equal to the non-standardized information matrix $C_{\tilde{K}}$, where $\tilde{K} = [|c_{j_1}|^{-1}\mathbf{e}_{j_1}, \dots, |c_{j_s}|^{-1}\mathbf{e}_{j_s}]$. □

LEMMA 2.1. *A positive eigenvector of an elementwise non-negative symmetric matrix corresponds to the maximum eigenvalue.*

PROOF. Let A be a non-negative symmetric matrix and let $\mathbf{x} = (x_0, \dots, x_n)^T \neq \mathbf{0}$ be such that $A\mathbf{x} = \lambda_{\max}(A)\mathbf{x}$. Set $\mathbf{y} = (|x_0|, \dots, |x_n|)^T$. Then

$$\lambda_{\max}(A) = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{y}^T\mathbf{y}} \leq \frac{\mathbf{y}^T A\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \leq \lambda_{\max}(A).$$

It follows that $\mathbf{y}$ is an eigenvector of A corresponding to $\lambda_{\max}(A)$. Now let $\mathbf{z}$ be a positive eigenvector of A. Since $\mathbf{y}$ is non-negative, $\mathbf{y}$ and $\mathbf{z}$ are not orthogonal and must therefore correspond to the same eigenvalue, that is, to $\lambda_{\max}(A)$. □

REMARK 2.1. The weights of the E-optimal design ξ and the standardized-E-optimal design ξ' can be expressed as

$$\xi(s_\kappa) = \frac{(-1)^\kappa}{\sum_{\mu=1}^{s} k_{j_\mu}^2 c_{j_\mu}^2} \sum_{\mu=1}^{s} l_{j_\mu}^{(\kappa)} k_{j_\mu}^2 c_{j_\mu}, \qquad \xi'(s_\kappa) = \frac{(-1)^\kappa}{s} \sum_{\mu=1}^{s} \frac{l_{j_\mu}^{(\kappa)}}{c_{j_\mu}},$$

where the $l_j^{(k)}$ are the coefficients of the Lagrange polynomials for the nodes $s_0, \dots, s_n$; that is, $L_k(x) = \sum_{j=0}^{n} l_j^{(k)} f_j(x)$ and $L_k(s_k) = 1$, $L_k(s_j) = 0$ if $j \neq k$.

REMARK 2.2. If the Chebyshev points in Theorem 2.1 are not unique, then the designs given there are still optimal but not unique anymore. For example, let $f_0(x) = 1 + x/2$ and $f_1(x) = 2 + x/2 - x^2 - x^3$, $x \in [-1, 1]$. The Chebyshev polynomial is $T(x) = f_0(x) - f_1(x)$ and $T(-1) = T(0) = -1$ and $T(1) = 1$. So there is an E-optimal design with support $\{-1, 1\}$ and another one with support $\{0, 1\}$.

Theorem 2.1 is easily extended to design spaces $\mathscr{X} \subset \mathbb{R}$ which are not intervals. In that case, the E-optimal designs may be not unique even if $1 \in \operatorname{span}\{f_0, \dots, f_n\}$. Indeed, the assumption that $1 \in \operatorname{span}\{f_0, \dots, f_n\}$ does not ensure uniqueness of the Chebyshev points when $\mathscr{X}$ is not an interval. If $f_j(x) = x^j$, $j = 0, 1, 2$, $x \in \mathscr{X} = [0, 1] \cup [2, 3]$, then there is an E-optimal design with support $\{0, 1, 3\}$ and another one with support $\{0, 2, 3\}$.

REMARK 2.3. The condition that $\{f_0, \dots, f_n\}$ be a Chebyshev system and every subsystem consisting of n functions be a weak Chebyshev system is satisfied if $\{f_0, \dots, f_n\}$ is a Descartes system [Karlin and Studden (1966), page 25]. The condition is also satisfied if, for all $x_0 > \cdots > x_n$, the matrix $(f_j(x_k))_{jk=0}^{n}$ is non-singular and totally non-negative. See Heiligers (2001) for some recent results on E-optimal designs for totally non-negative regression.

But consider the functions $f_0(x) = 1$, $f_1(x) = x$, $f_2(x) = (1 - x)^2$, $x \in [-1, 1]$. The matrix $(f_j(x_k))_{jk=0}^{2}$ is not totally non-negative for $x_0 = -1$, $x_1 = 0$, $x_2 = 1$; and $\{f_0, f_1, f_2\}$ is not a Descartes system either. Nevertheless, $\{f_0, f_1, f_2\}$ does satisfy the condition of Theorem 2.1.

The next lemma provides a simple means to check whether a given Chebyshev system satisfies the assumption of Theorem 2.1 on the subsytems.

LEMMA 2.2. *Let $\{f_0, \dots, f_n\}$ be a Chebyshev system. Then the following two conditions are equivalent*:

(i) *Every subsystem that consists of n of the functions $f_0, \dots, f_n$ is a Chebyshev system.*

(ii) *If $\sum_{j=0}^{n} a_j f_j(x) \not\equiv 0$ has n zeros, then all the coefficients $a_0, \dots, a_n$ are non-zero.*

PROOF. This follows from the fact that if $g(x) = \sum_{j=0}^{n} a_j f_j(x)$ vanishes at $x_1, \ldots, x_n$, then $g(x) = \rho \det[\mathbf{f}(x_1), \ldots, \mathbf{f}(x_n), \mathbf{f}(x)]$. $\square$

3. *E*-optimal designs for rational models on compact intervals. Let $U_m(x)$ denote the usual Chebyshev polynomial of the second kind, thus $U_m(x) = \sin((m+1)\arccos x)/\sqrt{1-x^2}$. Set $U_{-1}(x) = 0$ and $U_{-m}(x) = -U_{m-2}(x)$ for $m \geq 2$.

THEOREM 3.1. *Let $\alpha_1, \ldots, \alpha_n$ be distinct real numbers outside* $[-1, 1]$. *Consider the model*

$$E(Y(x)) = \theta_0 + \frac{\theta_1}{x-\alpha_1} + \cdots + \frac{\theta_n}{x-\alpha_n}, \qquad x \in [-1, 1]. \tag{3.1}$$

Define $\beta_1, \ldots, \beta_n \in (-1, 1)$ and $\gamma_0, \ldots, \gamma_{2n} \in \mathbb{R}$ by $(\beta_k + \beta_k^{-1}) = 2\alpha_k$ and

$$\sum_{k=0}^{2n} \gamma_k x^k = (x-\beta_1)^2 \cdots (x-\beta_n)^2.$$

Then the E-optimal design ξ and the standardized-E-optimal design ξ' for the weighted parameter subsystem $K^T\theta = (k_{j_1}\theta_{j_1}, \ldots, k_{j_s}\theta_{j_s})^T$ concentrate mass at the zeros $s_0 > \cdots > s_n$ of $(1-x^2)\sum_{k=0}^{2n} \gamma_k U_{-n+k-1}(x)$. The weights are given by

$$\xi(s_\kappa) = \frac{(-1)^\kappa}{\sum_{\mu=1}^{s} k_{j_\mu}^2 c_{j_\mu}^2} \sum_{\mu=1}^{s} l_{j_\mu}^{(\kappa)} k_{j_\mu}^2 c_{j_\mu} \quad \textit{and} \quad \xi'(s_\kappa) = \frac{(-1)^\kappa}{s} \sum_{\mu=1}^{s} \frac{l_{j_\mu}^{(\kappa)}}{c_{j_\mu}}, \tag{3.2}$$

where

$$c_0 = \frac{(-1)^n}{2} \left(\beta_1 \cdots \beta_n + \beta_1^{-1} \cdots \beta_n^{-1}\right),$$

$$c_j = -\left(\frac{\beta_j - \beta_j^{-1}}{2}\right)^2 \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{1-\beta_j\beta_k}{\beta_j - \beta_k}, \qquad j = 1, \ldots, n$$

and

$$\begin{aligned} l_0^{(k)} &= \frac{\prod_{\nu=1}^{n}(s_k - \alpha_\nu)}{\prod_{\substack{\nu=0 \\ \nu\neq k}}^{n}(s_k - s_\nu)}, \\ l_j^{(k)} &= (s_k - \alpha_j) \prod_{\substack{\nu=1 \\ \nu \neq j}}^{n} \frac{s_k - \alpha_\nu}{\alpha_j - \alpha_\nu} \prod_{\substack{\nu=0 \\ \nu \neq k}}^{n} \frac{\alpha_j - s_\nu}{s_k - s_\nu}, \qquad j = 1, \ldots, n, \; k = 0, \ldots, n. \end{aligned} \tag{3.3}$$

PROOF. Let $f_0(x) = 1$, and $f_j(x) = (x-\alpha_j)^{-1}$ for $j = 1, \ldots, n$. It is easily seen that $\{f_0, \ldots, f_n\}$ and all subsystems thereof are Chebyshev systems. By Proposition 4.1 of Borwein, Erdélyi and Zhang (1994), the Chebyshev polynomial for $\{f_0, \ldots, f_n\}$ is $\sum_{j=0}^{n} c_j f_j(x)$. By Theorem A.2 in Appendix A,

$s_0, \ldots, s_n$ are the Chebyshev points. Using the formula for the Cauchy determinant [Achieser (1956), pages 19-20], one may verify that the inverse of $(f_j(s_k))_{jk=0}^n$ is $(l_j^{(k)})_{kj=0}^n$. Now the assertions follow from Theorem 2.1 and Remark 2.1. □

The next theorem describes how the E-optimal design points vary as the poles of the regression model (3.1) vary and shows that the E-optimal designs converge to an arcsin support design [Pukelsheim (1993), page 217] when the poles approach $\pm\infty$.

THEOREM 3.2. *Let $s_k = s_k(\alpha_1, \ldots, \alpha_n)$, $k = 0, \ldots, n$, denote the support points of the E-optimal design ξ given in Theorem* 3.1.

(a) *If any one of the poles α_j moves to the right and the other poles remain fixed, then every interior design point s_k, $k = 1, \ldots, n-1$, moves to the left.*

(b) *If every pole moves toward $\pm\infty$, then $s_k \to \cos(k\pi/n)$ and $\xi(s_k) \to \frac{1}{n}$ for $k = 1, \ldots, n-1$ and $\xi(s_0) \to \frac{1}{2n}$, $\xi(s_n) \to \frac{1}{2n}$. In particular, the limiting behavior of ξ does not depend on K and is the same for the standardized-E-optimal design.*

PROOF. (a) By Theorems 2.1 and A.1, $s_1, \ldots, s_{n-1}$ are the zeros of the $(n-1)$st monic orthogonal polynomial with respect to

$$w(x) = \frac{\sqrt{1-x^2}}{(x-\alpha_1)^2 \cdots (x-\alpha_n)^2}.$$

For every $j = 1, \ldots, n$,

$$\frac{1}{w(x)} \frac{d}{d\alpha_j} w(x) = \frac{2}{x - \alpha_j}$$

is a decreasing function of $x \in [-1, 1]$. The assertion follows now from Markov's theorem on the variation of the zeros of orthogonal polynomials [Szegő (1975), page 115].

(b) Let β_k and γ_k be defined as in Theorem 3.1. If $|\alpha_j| \to \infty$ for all $j = 1, \ldots, n$, then $\beta_k \to 0$ for all k, and so $\gamma_k \to 0$ for $k \le 2n-1$ and $\gamma_{2n} = 1$. Thus $s_1, \ldots, s_{n-1}$ move toward the zeros of $U_{n-1}(x)$; that is, $s_k \to x_k := \cos(k\pi/n)$, $k = 0, \ldots, n$.

In view of representation (2.2) it is sufficient to prove the limit assertion for the weights in the special case where ξ is optimal for estimating a single parameter, say θ_j. Set $u(x) = (1-x^2)U_{n-1}(x)$. Then, from (3.3),

$$\lim_{|\alpha_1|, \ldots, |\alpha_n| \to \infty} \frac{l_j^{(k)}}{l_j^{(k')}} = \frac{\prod_{\nu \ne k'} (x_{k'} - x_\nu)}{\prod_{\nu \ne k} (x_k - x_\nu)} = \frac{u'(x_{k'})}{u'(x_k)}, \qquad k, k' = 0, \ldots, n.$$

But $u'(x) = -xU_{n-1}(x) - nT_n(x)$, where $T_n(x) = \cos(n \arccos x)$. Thus $u'(x_k) = (-1)^{k+1} n$ for $k = 1, \ldots, n-1$, and $u'(x_k) = (-1)^{k+1} 2n$ for $k = 0, n$. It now

TABLE 1
Optimal designs ξ for model (3.1) *and efficiencies of the arcsin support design ξ^*. The support points are given with the corresponding weights in parentheses*

$\boldsymbol{\alpha}_1$	$\boldsymbol{\alpha}_2$	$\boldsymbol{\alpha}_3$			ξ						$\mathbf{eff}(\xi^*)$
2	4	6	−1	(0.189)	−0.228	(0.356)	0.706	(0.311)	1	(0.144)	0.518
12	14	16	−1	(0.167)	−0.444	(0.334)	0.552	(0.333)	1	(0.166)	0.966
−2	4	6	−1	(0.125)	−0.552	(0.304)	0.494	(0.375)	1	(0.196)	0.952
−12	14	16	−1	(0.158)	−0.488	(0.325)	0.513	(0.342)	1	(0.175)	0.999

follows from (3.2) that

$$\frac{\xi(s_k)}{\xi(s_{k'})} = (-1)^{k+k'} \frac{l_j^{(k)}}{l_j^{(k')}} \to 1, \qquad k, k' = 1, \dots, n-1$$

and

$$\frac{\xi(s_0)}{\xi(s_1)} \to \frac{1}{2}, \qquad \frac{\xi(s_n)}{\xi(s_1)} \to \frac{1}{2}. \qquad \square$$

Theorem 3.2(b) suggests that the arcsin support design ξ^* with $\xi^*(\cos(k\pi/n)) = \frac{1}{n}$, $k = 1, \dots, n-1$, $\xi^*(-1) = \xi^*(1) = \frac{1}{2n}$ should be a good approximation to the E-optimal design ξ for model (3.1), provided that the poles are far enough away from the design space. Specifically, the efficiency of ξ^*, $\mathrm{eff}(\xi^*) = \lambda_{\min}(M(\xi^*))/\lambda_{\min}(M(\xi))$, should, then, be high. The numerical examples in Table 1 confirm this.

REMARK 3.1. He, Studden and Sun (1996) observed a relation between D-optimal designs for rational and polynomial models which is similar to that in Theorem 3.2(b).

THEOREM 3.3. *Let $\alpha_1 > \cdots > \alpha_m > 0$. Consider the model*

$$E(Y(x)) = \frac{\theta_{-m}}{x+\alpha_m} + \cdots + \frac{\theta_{-1}}{x+\alpha_1} + \theta_0 + \theta_1 x + \cdots + \theta_n x^n, \qquad x \in [0,1].$$

The E-optimal design ξ and the standardized-E-optimal design ξ' for the weighted parameter subsystem $K^T\theta = (k_{j_1}\theta_{j_1}, \dots, k_{j_s}\theta_{j_s})^T$ concentrate mass at $s_0 = 1$, $s_{n+m} = 0$ and the zeros $s_1 > \cdots > s_{n+m-1}$ of the orthogonal polynomial of degree $n+m-1$ with respect to

$$\frac{\sqrt{x(1-x)}\,dx}{(x+\alpha_1)^2 \cdots (x+\alpha_m)^2}, \qquad x \in [0,1].$$

The weights are given by

$$(\xi(s_0), \dots, \xi(s_{n+m}))^T = \frac{1}{\|K^T\mathbf{c}\|^2} JF^{-1}KK^T\mathbf{c},$$

$$(\xi'(s_0), \dots, \xi'(s_{n+m}))^T = \frac{1}{s} JF^{-1}K(K^TK)^{-1}K^T(c_0^{-1}, \dots, c_{m+n}^{-1})^T,$$

where $J = \operatorname{diag}(1, -1, \ldots, (-1)^{m+n})$, $\mathbf{c} = (F^T)^{-1}J\mathbf{1}$ *and* F^T *is the Cauchy–Vandermonde matrix*

$$F^T = \begin{bmatrix} \dfrac{1}{s_0+\alpha_m} & \cdots & \dfrac{1}{s_0+\alpha_1} & 1 & s_0 & \cdots & s_0^n \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \dfrac{1}{s_{m+n}+\alpha_m} & \cdots & \dfrac{1}{s_{m+n}+\alpha_1} & 1 & s_{m+n} & \cdots & s_{m+n}^n \end{bmatrix}.$$

PROOF. This follows from Theorem 2.1 and Theorem A.1 in Appendix A. That Theorem 2.1 can be applied to the model at hand is ensured by Lemma B.1 in Appendix B and Lemma 2.2. □

Theorems 3.1 and 3.3 dealt with E-optimal designs for regression functions with several poles of order 1. In the next theorem, the regression function may have a pole whose order is larger than 1, but may have no other poles. It seems to be more difficult to obtain E-optimal designs when there are several poles whose order is larger than 1; see the discussion at the end of Appendix B.

THEOREM 3.4. *Let* $\alpha > 0$. *Consider the model*

$$E(Y(x)) = \frac{\theta_{-m}}{(x+\alpha)^m} + \cdots + \frac{\theta_{-1}}{(x+\alpha)^1} + \theta_0 + \theta_1 x + \cdots + \theta_n x^n, \qquad x \in [0, 1].$$

The E-*optimal design* ξ *and the standardized-*E-*optimal design* ξ' *for the weighted parameter subsystem* $K^T\theta = (k_{j_1}\theta_{j_1}, \ldots, k_{j_s}\theta_{j_s})^T$ *concentrate mass at* $s_0 = 1$, $s_{n+m} = 0$ *and the zeros* $s_1 > \cdots > s_{n+m-1}$ *of the orthogonal polynomial of degree* $n+m-1$ *with respect to*

$$\frac{\sqrt{x(1-x)}\,dx}{(x+\alpha)^{2m}}, \qquad x \in [0, 1].$$

The weights are given by

$$(\xi(s_0), \ldots, \xi(s_{n+m}))^T = \frac{1}{\|K^T\mathbf{c}\|^2} JF^{-1}KK^T\mathbf{c},$$

$$(\xi'(s_0), \ldots, \xi'(s_{n+m}))^T = \frac{1}{s} JF^{-1}K(K^TK)^{-1}K^T(c_0^{-1}, \ldots, c_{m+n}^{-1})^T,$$

where $J = \operatorname{diag}(1, -1, \ldots, (-1)^{m+n})$, $\mathbf{c} = (F^T)^{-1}J\mathbf{1}$ *and* F^T *is the Cauchy–Vandermonde matrix*

$$F^T = \begin{bmatrix} \dfrac{1}{(s_0+\alpha)^m} & \cdots & \dfrac{1}{s_0+\alpha} & 1 & s_0 & \cdots & s_0^n \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \dfrac{1}{(s_{m+n}+\alpha)^m} & \cdots & \dfrac{1}{s_{m+n}+\alpha} & 1 & s_{m+n} & \cdots & s_{m+n}^n \end{bmatrix}.$$

PROOF. Theorems 2.1, A.1 and Lemmas B.3 and 2.2. $\square$

Explicit representations of the inverse matrices F^{-1} in Theorems 3.3 and 3.4 can be derived from the formulas in Mühlbach (1996).

4. E-optimal designs for a rational model on two disjoint intervals. This section addresses the problem of determining the E-optimal designs for the model

$$E(Y(x)) = \theta_{-m}x^{-m} + \cdots + \theta_n x^n, \qquad x \in [-b, -a] \cup [a, b], \tag{4.1}$$

where $0 < a < b$ and $m, n \geq 0$. A complete solution for the case where interest is in coefficients of odd powers only or of even powers only is given in Theorem 4.1. Theorem 4.2 treats more general parameter subsystems under the assumption that $b \leq 1$. It turns out that the E-optimal designs are not always unique and the number of their support points may be larger or smaller than the number of parameters. This is different from corresponding results for ordinary polynomial regression on a compact interval. Indeed, E-optimal designs for homoscedastic polynomial regression on $[-1, 1]$ are always unique and the number of support points is never larger than the number of parameters; see Studden (1968), Pukelsheim and Studden (1993) and Heiligers (1994). On the other hand, Chang and Heiligers (1996) give an example where E-optimal designs for a heteroscedastic polynomial regression model are not unique.

THEOREM 4.1. *Consider model* (4.1). *Let* $K^T\theta = (k_{j_1}\theta_{j_1}, \ldots, k_{j_s}\theta_{j_s})^T$ *and suppose that either every index* j_μ *is even or every* j_μ *is odd. If the* j_μ *are even, let*

$$m_0 = \max\{j : j \text{ even},\ j \leq m\}, \qquad n_0 = \max\{j : j \text{ even},\ j \leq n\}.$$

If the j_μ *are odd, let*

$$m_0 = \max\{j : j \text{ odd},\ j \leq m\}, \qquad n_0 = \max\{j : j \text{ odd},\ j \leq n\}.$$

Suppose that $m_0 \geq 0$, $n_0 \geq 0$ *and* $m_0 + n_0 > 0$. *Let* $t(x) = \sum_{j=-m_0}^{n_0} c_j x^{j+m_0}$ *be the orthogonal polynomial of degree* $m_0 + n_0$ *with respect to*

$$\frac{dx}{|x|^{2m_0-1}\sqrt{(x^2-a^2)(b^2-x^2)}}, \qquad x \in [-b, -a] \cup [a, b], \tag{4.2}$$

normalized so that $t(b) = b^{m_0}$. *Let* $s_0 > \cdots > s_{m_0+n_0+1}$ *denote the points in* $[-b, -a] \cup [a, b]$ *where* $|t(x)/x^{m_0}| = 1$. *Let*

$$F = \begin{bmatrix} s_0^{-m} & \cdots & s_{m_0+n_0+1}^{-m} \\ \vdots & & \vdots \\ s_0^{n} & \cdots & s_{m_0+n_0+1}^{n} \end{bmatrix} \quad \text{and} \quad J = \operatorname{diag}\left(\frac{t(s_0)}{s_0^{m_0}}, \ldots, \frac{t(s_{m_0+n_0+1})}{s_{m_0+n_0+1}^{m_0}}\right).$$

Let $\mathbf{c} = (c_{-m}, \ldots, c_n)^T$, *where* $c_{-m} = 0$ *if* $m > m_0$ *and* $c_n = 0$ *if* $n > n_0$.

(a) *If $m+n > m_0+n_0$, then there is a unique E-optimal design ξ for $K^T\theta$, it is supported on the points $s_0, \dots, s_{m_0+n_0+1}$, and the weights are given by the unique solution of*

$$\|K^T\mathbf{c}\|^2 FJ(\xi(s_0), \dots, \xi(s_{m_0+n_0+1}))^T = KK^T\mathbf{c}. \tag{4.3}$$

(b) *If $m+n = m_0+n_0$, then every E-optimal design for $K^T\theta$ is supported on $s_0, \dots, s_{m_0+n_0+1}$ and the weights satisfy* (4.3). *Moreover, there is a unique symmetric E-optimal design.*

(c) *A design is standardized-E-optimal for $(\theta_{j_1}, \dots, \theta_{j_s})^T$ if and only if it is E-optimal for $(c_{j_1}^{-1}\theta_{j_1}, \dots, c_{j_s}^{-1}\theta_{j_s})^T$.*

This theorem is proved by first deriving optimal designs for a simpler model, a model that contains only powers of the same parity. These designs are then shown to be optimal for the original model. This second step relies on the following lemma, which is an immediate consequence of Theorem 1 of Imhof and Krafft (1999).

LEMMA 4.1. *Consider the initial model*

$$E(Y(x)) = \theta^T\mathbf{f}(x) + \vartheta^T\mathbf{g}(x), \qquad x \in \mathscr{X},$$

and the reduced model

$$E(Y_0(x)) = \theta^T\mathbf{f}(x), \qquad x \in \mathscr{X}.$$

Let ξ be an E-optimal design for estimating $K_0^T\theta$ in the reduced model. Set

$$K^T = \left[K_0^T, K_0^T\left(\int \mathbf{f}(x)\mathbf{f}^T(x)\,d\xi(x)\right)^+\left(\int \mathbf{f}(x)\mathbf{g}^T(x)\,d\xi(x)\right)\right].$$

Then ξ is also E-optimal for estimating $K^T\begin{bmatrix}\theta\\ \vartheta\end{bmatrix}$ in the initial model and the optimal information matrices of ξ for the two models coincide. Moreover, every E-optimal design for $K^T\begin{bmatrix}\theta\\ \vartheta\end{bmatrix}$ is E-optimal for $K_0^T\theta$ as well.

PROOF OF THEOREM 4.1. (a) and (b) Let $f_j(x) = x^j$ for $j = -m, \dots, n$, and let $\mathscr{J} = \{-m_0, -m_0+2, \dots, n_0\}$. The system $\{f_j\}_{j\in\mathscr{J}}$ and all its subsystems are Chebyshev systems on $[a,b]$; see DeVore and Lorentz [(1993), page 69]. Since the weight (4.2) is even, $t(x)$ is even as well, so that $t(x)/x^{m_0} = \sum_{j\in\mathscr{J}} c_j f_j(x)$. That $\sum_{j\in\mathscr{J}} c_j f_j(x)$ is the Chebyshev polynomial for $\{f_j(x)\}_{j\in\mathscr{J}}$, $x \in [a,b]$, follows from Theorem A.3. The Chebyshev points are $s_0 > \cdots > s_{(m_0+n_0)/2}$. Let K_0 be the matrix obtained from K by deleting every second row beginning with the first or second row according as $m > m_0$ or $m = m_0$. It now follows from Theorem 2.1 that there is a unique E-optimal design ξ_0 for estimating $K_0^T(\theta_{-m_0}, \theta_{-m_0+2}, \dots, \theta_{n_0})^T$ in the model

$$E(Z(x)) = \sum_{j\in\mathscr{J}} \theta_j f_j(x), \qquad x \in [a,b].$$

The support points of ξ_0 are $s_0, \dots, s_{(m_0+n_0)/2}$ and $\lambda_{\max}(C_{K_0}^{-1}(M_0(\xi_0))) = \|K^T\mathbf{c}\|^2$. Here $M_0(\xi) = \left(\int x^{j+k}\, d\xi(x)\right)_{j,k\in\mathscr{J}}$. This matrix contains only even moments of ξ. Therefore, a design ξ on $[-b,-a]\cup[a,b]$ is E-optimal for estimating $K_0^T(\theta_{-m_0}, \theta_{-m_0+2}, \dots, \theta_{n_0})^T$ in the model

$$E(Y_0(x)) = \sum_{j\in\mathscr{J}} \theta_j f_j(x), \qquad x \in [-b,-a]\cup[a,b],$$

if and only if $\xi(x)+\xi(-x) = \xi_0(x)$ for all $x\in[a,b]$. In particular, the symmetric design $\xi_1(x) = \xi_0(|x|)/2$ is optimal. Clearly, $\int f_j(x) f_k(x)\, d\xi_1(x) = 0$ for all $j \in \mathscr{J}$ and $k \in \{-m,\dots,n\}\setminus\mathscr{J}$. It therefore follows by Lemma 4.1 that ξ_1 is also E-optimal for $K^T\theta$ in the initial model (4.1) and $\lambda_{\max}(C_K^{-1}(M(\xi_1))) = \|K^T\mathbf{c}\|^2$. Moreover, if ξ is E-optimal for $K^T\theta$ in (4.1), then $\xi(x)+\xi(-x) = \xi_0(x)$ for all $x\in[a,b]$. Thus the support of ξ must be contained in $\{s_0,\dots,s_{m_0+n_0+1}\}$, so that $\mathbf{c}^T M(\xi)\mathbf{c} = \int(\mathbf{c}^T\mathbf{f}(x))^2\, d\xi(x) = 1$. It now follows from Cauchy's inequality that

$$\|K^T\mathbf{c}\|^2 = \lambda_{\max}(C_K^{-1}(M(\xi))) \ge \frac{\mathbf{c}^T KK^T M^+(\xi) KK^T\mathbf{c}}{\|K^T\mathbf{c}\|^2} \ge \frac{\mathbf{c}^T KK^T\mathbf{c}}{\mathbf{c}^T M(\xi)\mathbf{c}} = \|K^T\mathbf{c}\|^2.$$

Thus $M(\xi)\mathbf{c}$ and $KK^T\mathbf{c}$ must be proportional. This is equivalent to (4.3). If $m+n > m_0+n_0$, then (4.3) has exactly one solution.

(c) If ξ_j is an optimal design for $\theta_j = \mathbf{e}_j^T\theta$, $j\in\mathscr{J}$, then, by (a) and (b), $\mathbf{e}_j^T M^+(\xi_j)\mathbf{e}_j = c_j^2$. $\square$

The following example shows that there may be infinitely many E-optimal designs for model (4.1) and that not every design whose weights satisfy (4.3) must be E-optimal.

EXAMPLE 4.1. Consider the model

$$E(Y(x)) = \frac{\theta_{-1}}{x} + \theta_0 + \theta_1 x, \qquad x\in[-1,-1/2]\cup[1/2,1].$$

Let $K^T = \left[\begin{smallmatrix} 1 & 0 & 0\\ 0 & 0 & 1/2\end{smallmatrix}\right]$. Then, with the notation from Theorem 4.1, $m = n = m_0 = n_0 = 1$, $t(x) = -1+2x^2$, $\mathbf{c} = (-1,0,2)^T$, $s_0 = 1$, $s_1 = 1/2$, $s_2 = -1/2$, $s_3 = -1$, and condition (4.3) means that

$$\xi(s_0) = \tfrac14 - \rho, \quad \xi(s_1) = \tfrac14 - \rho, \quad \xi(s_2) = \tfrac14 + \rho, \quad \xi(s_3) = \tfrac14 + \rho,$$

where $-1/4 \le \rho \le 1/4$. For these ξ, the eigenvalues of $C_K(\xi)$ are $1/2$ and $9/2(1-16\rho^2)$. That is, if $|\rho| \le 1/(3\sqrt{2}) \approx 0.236$, then $\lambda_{\min}(C_K(\xi)) = 1/2$ and ξ is E-optimal for $K^T\theta$. If $|\rho| > 1/(3\sqrt{2})$, then $\lambda_{\min}(C_K(\xi)) < 1/2$.

On the other hand, if $K^T = (0,0,1)$, then every design which satisfies (4.3) is optimal for $K^T\theta$.

The next theorem extends Theorem 4.1 to parameter subsystems that satisfy a certain neighborhood condition similar to those in Pukelsheim and Studden (1993) and Heiligers (1994).

THEOREM 4.2. *Consider model* (4.1). *Suppose that* $0 < a < b \leq 1$. *Let* $K^T\theta = (\theta_{j_1}, \dots, \theta_{j_s})^T$, $-m \leq j_1 < \cdots < j_s \leq n$, *and suppose that for every* $\mu = 1, \dots, s-1$,

$$\text{if } j_s - j_\mu \text{ is odd, then } j_{\mu+1} = j_\mu + 1.$$

Let K_0 *be the matrix obtained from* $K = [\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_s}]$ *by deleting those columns* $\mathbf{e}_{j_\mu}$ *for which* $j_s - j_\mu$ *is odd. Let* ξ^* *be the unique* (*symmetric*) *E-optimal design for* $K_0^T\theta$ *as given in Theorem* 4.1. *Then* ξ^* *is also E-optimal for* $K^T\theta$. *If* $n + m + 1 \geq |\operatorname{supp}(\xi^*)|$, *then* ξ^* *is the only E-optimal design for* $K^T\theta$.

PROOF. Let ξ be a feasible design for $K^T\theta$. Then ξ is also feasible for $K_0^T\theta$ and

$$\lambda_{\max}(C_{K_0}^{-1}(M(\xi^*))) \leq \lambda_{\max}(C_{K_0}^{-1}(M(\xi))).$$

If $n + m + 1 \geq |\operatorname{supp}(\xi^*)|$, then the inequality is strict unless $\xi = \xi^*$; see Theorem 4.1. Let K_1 denote the matrix that consists of those columns of K which are not in K_0 and let $\mathbf{x}$ be a normalized eigenvector of $C_{K_0}^{-1}(M(\xi))$ corresponding to the maximum eigenvalue. Then

$$\begin{aligned} \lambda_{\max}(C_{K_0}^{-1}(M(\xi))) &= [\mathbf{x}^T, \mathbf{0}^T] \begin{bmatrix} K_0^T \\ K_1^T \end{bmatrix} M^+(\xi)[K_0, K_1] \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} \\ &\leq \lambda_{\max}(C_K^{-1}(M(\xi))). \end{aligned} \tag{4.4}$$

The proof is now complete by showing that there is equality in (4.4) if $\xi = \xi^*$. This can be done by arguments similar to those in Pukelsheim and Studden [(1993), pages 410–411] or Heiligers [(1994), pages 922–923]. We omit the details. $\square$

APPENDIX A:
CHEBYSHEV POLYNOMIALS FOR RATIONAL SYSTEMS

The next three theorems provide explicit expressions for the Chebyshev polynomial of a Chebyshev system $\{f_0, \dots, f_n\}$ whose span satisfies

$$\operatorname{span}\{f_0(x), \dots, f_n(x)\} = \left\{ \frac{q(x)}{\sqrt{p(x)}} : q(x) \in \mathscr{P}_n \right\}. \tag{A.1}$$

Here $\mathscr{P}_n$ denotes the set of all algebraic polynomials of degree less than or equal to n and $p(x)$ is a fixed polynomial. The Chebyshev polynomial and the Chebyshev points will be expressed in terms of Bernstein-Szegő orthogonal polynomials [Szegő (1975), Section 2.6].

The first theorem is a consequence of Theorem 2.6 in Szegő (1975) and the argument principle; see also Achieser [(1956), pages 58–60, 249–251], Karlin and Studden [(1966), pages 287–289] and Krein and Nudel'man [(1977), pages 373–374].

THEOREM A.1. *Let $p \in \mathscr{P}_\nu$ be positive on $[a, b]$. Let $\{f_0, \dots, f_n\}$, $n \geq \nu/2$, be a Chebyshev system on $[a, b]$ satisfying* (A.1). *Let $t(x) \in \mathscr{P}_n$ be the nth Bernstein-Szegő orthogonal polynomial with respect to $dx/(p(x)\sqrt{(x-a)(b-x)})$ normalized so that $t(b) = \sqrt{p(b)}$ and let $u(x) \in \mathscr{P}_{n-1}$ be the $(n-1)$st monic Bernstein-Szegő orthogonal polynomial with respect to $\sqrt{(x-a)(b-x)}dx/p(x)$. Then the Chebyshev polynomial for $\{f_0, \dots, f_n\}$ is given by $t(x)/\sqrt{p(x)}$ and the Chebyshev points are the zeros of $(x-a)(b-x)u(x)$. Moreover, p(x) has the representation*

$$p(x) = t^2(x) + c(x-a)(b-x)u^2(x),$$

where $c > 0$.

If the zeros of p are known and $[a, b] = [-1, 1]$, then the Bernstein-Szegő polynomials have a simple representation as linear combinations of the ordinary Chebyshev polynomials of the first and second kind, $T_m(x) = \cos(m \arccos x)$, $U_m(x) = \sin((m+1)\arccos x)/\sqrt{1-x^2}$.

THEOREM A.2. *Let p be an algebraic polynomial of degree $\nu \geq 1$ which is positive on $[-1, 1]$. Let $\alpha_1, \dots, \alpha_\nu$ denote the zeros of p. Let $\{f_0, \dots, f_n\}$, $n \geq \nu/2$, be a Chebyshev system on $[-1, 1]$ satisfying* (A.1). *Define $\beta_1, \dots, \beta_\nu \in \{z \in \mathbb{C} : |z| < 1\}$ by*

$$\frac{1}{2}\left(\beta_k + \frac{1}{\beta_k}\right) = \alpha_k,$$

and define $\gamma_0, \dots, \gamma_\nu \in \mathbb{R}$ by

$$\sum_{k=0}^{\nu} \gamma_k x^k = (x - \beta_1) \cdots (x - \beta_\nu).$$

Then the Chebyshev polynomial for $\{f_0, \dots, f_n\}$ is a multiple of

$$\frac{\sum_{k=0}^{\nu} \gamma_k T_{|n-\nu+k|}(x)}{\sqrt{p(x)}},$$

and the Chebyshev points are the zeros of

$$(1 - x^2) \sum_{k=0}^{\nu} \gamma_k U_{n-\nu+k-1}(x).$$

Here $U_{-1}(x) = 0$ and $U_{-m}(x) = -U_{m-2}(x)$ for $m \geq 2$.

PROOF. Set $h(z) = \sum_{k=0}^{\nu} \gamma_k z^k$. For $x \in [-1, 1]$, let $z = z(x) = e^{i\theta}$, where $\theta = \arccos x \in [0, \pi]$ and $i = \sqrt{-1}$. Consider

$$t(x) := \Re\{z^{n-\nu}h(z)\} = \sum_{k=0}^{\nu} \gamma_k \Re\left\{z^{n-\nu+k}\right\} = \sum_{k=0}^{\nu} \gamma_k T_{|n-\nu+k|}(x),$$

$$u(x) := \frac{\Im\{z^{n-\nu}h(z)\}}{\sqrt{1-x^2}} = \sum_{k=0}^{\nu} \gamma_k U_{n-\nu+k-1}(x).$$

Obviously

$$t^2(x) + (1-x^2)u^2(x) = |h(z)|^2 = \prod_{k=1}^{\nu} (z-\beta_k)(\overline{z} - \overline{\beta_k}).$$

Now the non-real β_k occur in pairs of complex conjugate numbers and $z\overline{z} = 1$ and $z + \overline{z} = 2x$. Therefore,

$$\prod_{k=1}^{\nu} (z-\beta_k)(\overline{z} - \overline{\beta_k}) = \prod_{k=1}^{\nu} 2\beta_k(\alpha_k - x) = c^2 p(x),$$

for some $c > 0$. Hence $|t(x)/\sqrt{p(x)}| \leq c$ for all $x \in [-1, 1]$. By the argument principle applied to $z^{n-\nu}h(z)$, there exist $n+1$ zeros $s_0 > \cdots > s_n$ of $(1-x^2)^{1/2}u(x)$ and $t(s_k) = (-1)^k c\sqrt{p(s_k)}$. $\square$

The next theorem gives the Chebyshev polynomials for rational systems on two disjoint intervals.

THEOREM A.3. *Let p be an even algebraic polynomial of degree ν which is positive on $\mathscr{X} = [-b, -a] \cup [a, b]$, $0 < a < b$. Let $f_0, \ldots, f_n$, $n \geq \max\{\nu/2, 2\}$, be continuous functions on $\mathscr{X}$ that satisfy* (A.1). *If n is odd, suppose also that $p(0) = 0$. Let $t(x) \in \mathscr{P}_n$ be the nth orthogonal polynomial with respect to*

$$\frac{|x|\,dx}{p(x)\sqrt{(x^2-a^2)(b^2-x^2)}}, \quad x \in \mathscr{X},$$

normalized so that $t(b) = \sqrt{p(b)}$. Let $u(x) \in \mathscr{P}_{n-2}$ be the $(n-2)$nd monic orthogonal polynomial with respect to

$$\frac{|x|\sqrt{(x^2-a^2)(b^2-x^2)}\,dx}{p(x)}, \qquad x \in \mathscr{X}.$$

Set $T(x) = t(x)/\sqrt{p(x)}$. Then $T \in \operatorname{span}\{f_0, \ldots, f_n\}$, $|T(x)| \leq 1$ for all $x \in \mathscr{X}$, and

(a) *if n is odd, then $(x^2-a^2)(b^2-x^2)u(x)$ has $n+1$ zeros $s_0 > \cdots > s_n$ in $\mathscr{X}$, $s_k = -s_{n-k}$ and $T(s_k) = (-1)^k$ for $k = 0, \ldots, n$;*

(b) *if n is even, then $(x^2-a^2)(b^2-x^2)u(x)$ has $n+2$ zeros $s_0 > \cdots > s_{n+1}$ in $\mathscr{X}$, $s_k = -s_{n+1-k}$ and $T(s_k) = (-1)^k$ for $k = 0, \ldots, n/2$ and $T(s_k) = (-1)^{k+1}$ for $k = n/2+1, \ldots, n+1$.*

Moreover, $p(x)$ *has the representation*

$$p(x) = t^2(x) + c(x^2 - a^2)(b^2 - x^2)u^2(x),$$

where $c > 0$.

PROOF. This follows from Theorem A.1 applied to the interval $[a^2, b^2]$ and the weight $p(\sqrt{x})$ if n is even and $p(\sqrt{x})/x$ if n is odd. □

REMARK A.1. The condition that $p(0) = 0$ if n is odd is not superfluous. If $n = 5$, $\mathscr{X} = [-1, -1/2] \cup [1/2, 1]$ and $p(x) \equiv 1$, then $|t(x)| < 1$ for all $x \in \mathscr{X} \setminus \{\pm 1\}$. However, $t(1/2) \approx 0.99$ and $t(s_1) \approx -0.99$, where s_1 is the positive zero of $u(x)$.

The Chebyshev polynomial for $\{1, x, \dots, x^n\}$, $x \in [-1, -a] \cup [a, 1]$, n odd, is given in terms of elliptic functions in Achieser [(1956), page 287]. Peherstorfer (1995) investigated the case where the two intervals have different lengths.

REMARK A.2. Spruill (1987) calculates optimal interpolation designs for polynomial regression models with design spaces that consist of two intervals. The designs are based on Chebyshev points, too. Theorem A.3 can be used to extend the results to rational systems.

APPENDIX B: AUXILIARY RESULTS

LEMMA B.1. *Let* $\alpha_1 > \cdots > \alpha_m > 0$ *and let*

$$\begin{aligned} f_{-j}(x) &= \frac{1}{x + \alpha_j}, \quad && j = 1, \dots, m, \\ f_j(x) &= x^j, && j = 0, \dots, n. \end{aligned}$$

If $f = \sum_{j=-m}^{n} a_j f_j \not\equiv 0$ *has* $m + n$ *zeros in* $[0, \infty)$, *then* $a_j a_{j+1} < 0$ *for* $j = -m, \dots, n-1$.

The proof of Lemma B.1 uses a subsidiary result on symmetric functions. Let σ_k denote the kth elementary symmetric function of $\alpha = (\alpha_1, \dots, \alpha_m)$, that is,

$$\sigma_k = \sigma_k(\alpha) = \sum_{1 \le j_1 < \cdots < j_k \le m} \alpha_{j_1} \cdots \alpha_{j_k}, \qquad k = 1, \dots, m,$$

$\sigma_0 = 1$, and $\sigma_k = 0$ for $k > m$ and for $k < 0$. Moreover, let

$$\bar{\sigma}_k = \bar{\sigma}_k(\alpha) = \sum_{1 \le j_1 \le \cdots \le j_k \le m} \alpha_{j_1} \cdots \alpha_{j_k}, \qquad k = 1, 2, \dots,$$

and $\bar{\sigma}_0 = 1$.

LEMMA B.2. *For every $n \in \mathbb{N}$,*

$$\begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_n \\ 0 & \sigma_0 & \sigma_1 & \cdots & \sigma_{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & \sigma_0 \end{bmatrix}^{-1} = \begin{bmatrix} \bar\sigma_0 & -\bar\sigma_1 & \bar\sigma_2 & \cdots & (-1)^n\bar\sigma_n \\ 0 & \bar\sigma_0 & -\bar\sigma_1 & \cdots & (-1)^{n-1}\bar\sigma_{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & \bar\sigma_0 \end{bmatrix}.$$

This follows from Stanley [(1999), equation (7.13), page 296].

PROOF OF LEMMA B.1. Suppose that $f = \sum_{j=-m}^{n} a_j f_j$ has $m+n$ nonnegative zeros. Write $\omega(x) = (x+\alpha_1)\cdots(x+\alpha_m)$ and define $b_0,\dots,b_{m+n}$ by $\omega(x)f(x) = \sum_{k=0}^{m+n} b_k x^k$. Then, by Descartes' rule of signs,

$$b_0 b_1 \le 0 \quad \text{and} \quad b_k b_{k+1} < 0 \qquad \text{for } k = 1,\dots,m+n-1. \tag{B.1}$$

Suppose that $b_1 < 0$. Then, for $j = 1,\dots,m$,

$$0 < \sum_{k=0}^{m+n} b_k(-\alpha_j)^k = a_{-j}(\alpha_1-\alpha_j)\cdots(\alpha_{j-1}-\alpha_j)(\alpha_{j+1}-\alpha_j)\cdots(\alpha_m-\alpha_j),$$

and so $(-1)^{m-j}a_{-j} > 0$. To determine the signs of $a_0,\dots,a_n$ observe that for $k = m,\dots,m+n$,

$$k!b_k = \frac{d^k}{dx^k}\sum_{j=0}^{m+n} b_j x^j\bigg|_{x=0} = \frac{d^k}{dx^k}\omega(x)f(x)\bigg|_{x=0} = \sum_{j=0}^{\min\{k,n\}}\binom{k}{j}\omega^{(k-j)}(0)j!a_j.$$

As $\omega^{(l)}(0) = l!\sigma_{m-l}(\alpha_1,\dots,\alpha_m) = l!\sigma_{m-l}$ for $l \ge 0$, and $\sigma_l = 0$ for $l > m$, one obtains the equations

$$\begin{aligned} b_m &= \sigma_0 a_0 + \sigma_1 a_1 + \cdots + \sigma_n a_n, \\ b_{m+1} &= \sigma_0 a_1 + \cdots + \sigma_{n-1} a_n, \\ &\cdots\cdots\cdots\cdots \\ b_{m+n} &= \sigma_0 a_n. \end{aligned}$$

According to Lemma B.2, this system of equations is equivalent to the system

$$\begin{aligned} a_0 &= \bar\sigma_0 b_m - \bar\sigma_1 b_{m+1} \pm \cdots + (-1)^n\bar\sigma_n b_{m+n}, \\ a_1 &= \bar\sigma_0 b_{m+1} \mp \cdots + (-1)^{n-1}\bar\sigma_{n-1} b_{m+n}, \\ &\cdots\cdots\cdots\cdots \\ a_n &= \bar\sigma_0 b_{m+n}. \end{aligned}$$

Since $\bar\sigma_l > 0$ for every l, it therefore follows by inequalities (B.1) that $(-1)^{m+j}a_j > 0$ for $j = 0,\dots,n$. □

LEMMA B.3. *Let $\alpha > 0$,*

$$\begin{aligned} f_{-j}(x) &= \frac{1}{(x+\alpha)^j}, \qquad j = 1,\dots,m, \\ f_j(x) &= x^j, \qquad j = 0,\dots,n. \end{aligned}$$

If $f = \sum_{j=-m}^{n} a_j f_j \not\equiv 0$ has $m+n$ zeros in $[0,\infty)$, then $a_j a_{j+1} < 0$ for $j = -m,\dots,n-1$.

PROOF. Suppose for simplicity that $\alpha = 1$. Let $b_0, \ldots, b_{m+n}$ be such that $(x+1)^m f(x) = \sum_{k=0}^{m+n} b_k x^k$. If f has $m+n$ zeros in $[0, \infty)$, then, by Descartes' rule of signs,

$$b_0 b_1 \leq 0 \quad \text{and} \quad b_k b_{k+1} < 0 \qquad \text{for } k = 1, \ldots, m+n-1. \tag{B.2}$$

To express the a_j in terms of the b_k note first that

$$[A|B] \begin{pmatrix} a_{-m} \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_0 \\ \vdots \\ b_{m+n} \end{pmatrix},$$

where

$$A = \left[\binom{k}{j} \right]_{j=0,\ k=0}^{m+n,\ m-1} \quad \text{and} \quad B = \left[\binom{m}{j-k} \right]_{j=0,\ k=0}^{m+n,\ n}.$$

The inverse of $[A|B]$ is given by $\left[\begin{smallmatrix} C \\ D \end{smallmatrix}\right]$ with

$$C = \left[(-1)^{j+k} \binom{k}{j} \right]_{j=0,\ k=0}^{m-1,\ m+n} \quad \text{and} \quad D = \left[(-1)^{j+k+m} \binom{k-j-1}{k-j-m} \right]_{j=0,\ k=0}^{n,\ m+n}.$$

This is seen by verifying that $\left[\begin{smallmatrix} C \\ D \end{smallmatrix}\right] [A|B] = I$. Since $\left[\begin{smallmatrix} C \\ D \end{smallmatrix}\right]$ and $[A|B]$ are upper triangular matrices, $DA = 0$. Furthermore, it follows from Feller [(1968), equation (12.15), page 65] that, for $j = 1, \ldots, m-1$ and $k = 0, \ldots, n$,

$$(CB)_{jk} = \sum_l (-1)^{j+l} \binom{l}{j} \binom{m}{l-k} = (-1)^{j+k+m} \binom{k}{m+k-j} = 0.$$

That $(CB)_{0k} = 0$ follows from Feller [(1968), equation (12.14)]. Thus $CB = 0$. Similarly, $CA = I_m$ and $DB = I_{n+1}$. Hence

$$\begin{pmatrix} a_{-m} \\ \vdots \\ a_n \end{pmatrix} = \begin{bmatrix} C \\ D \end{bmatrix} \begin{pmatrix} b_0 \\ \vdots \\ b_{m+n} \end{pmatrix}.$$

In view of the sign pattern of C and D the assertion now follows from (B.2). □

It is not possible to extend Lemmas B.1 and B.3 to cover Chebyshev systems of the form

$$\frac{1}{x+\alpha_1}, \ldots, \frac{1}{(x+\alpha_1)^{n_1}}, \frac{1}{x+\alpha_2}, \ldots, \frac{1}{(x+\alpha_2)^{n_2}}, \ldots, 1, x, \ldots, x^n.$$

Consider, for example, the system

$$f_{-3}(x) = \frac{1}{x+1}, \quad f_{-2}(x) = \frac{1}{x+3}, \quad f_{-1}(x) = \frac{1}{(x+3)^2}, \quad f_0(x) = 1,$$

where $x \in \mathscr{X} = [1, 7]$. Then $f = 1f_{-3} + 0f_{-2} - 7f_{-1} - \frac{1}{18} f_0$ has three zeros in $\mathscr{X}$ even though one of the coefficients is zero. In particular, Theorem 2.1

cannot be applied to $\{f_{-3}, \dots, f_0\}$ as the subsystem $\{f_{-3}, f_{-1}, f_0\}$ is not a weak Chebyshev system.

Acknowledgments. We thank the referees and an Associate Editor for their constructive comments. The paper was initiated when L. Imhof visited the Department of Statistics, Purdue University. This author is grateful to the department for its hospitality.

REFERENCES

ACHIESER, N. I. (1956). *Theory of Approximation.* Ungar, New York.

BORWEIN, P., ERDÉLYI, T. and ZHANG, J. (1994). Chebyshev polynomials and Markov-Bernstein type inequalities for rational spaces. *J. London Math. Soc.* **50** 501–519.

CHANG, F.-C. and HEILIGERS, B. (1996). E-optimal designs for polynomial regression without intercept. *J. Statist. Plann. Inference* **55** 371–387.

DETTE, H. (1993). A note on E-optimal designs for weighted polynomial regression. *Ann. Statist.* **21** 767–771.

DETTE, H. (1997a). Designing experiments with respect to 'standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59** 97–110.

DETTE, H. (1997b). E-optimal designs for regression models with quantitative factors—a reasonable choice? *Can. J. Statist.* **25** 531–543.

DETTE, H., HAINES, L. M. and IMHOF, L. (1999). Optimal designs for rational models and weighted polynomial regression. *Ann. Statist.* **27** 1272–1293.

DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation*. Springer, New York.

FELLER, W. (1968). *An Introduction to Probability and Its Applications* **1**, 3rd ed. Wiley, New York.

HAINES, L. M. (1992). Optimal design for inverse quadratic polynomials. *South African Statist. J.* **26** 25–41.

HE, Z., STUDDEN, W. J. and SUN, D. (1996). Optimal designs for rational models. *Ann. Statist.* **24** 2128–2147.

HEILIGERS, B. (1994). E-optimal designs in weighted polynomial regression. *Ann. Statist.* **22** 917–929.

HEILIGERS, B. (1996). Computing E-optimal polynomial regression designs. *J. Statist. Plann. Inference* **55** 219–233.

HEILIGERS, B. (2001). Totally positive regression: E-optimal designs. *Metrika.* To appear.

IMHOF, L. and KRAFFT, O. (1999). Extending design-optimality from an initial model to augmented models. *Metrika* **49** 19–26.

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

KREIN, M. G. and NUDEL'MAN, A. A. (1977). *The Markov Moment Problem and Extremal Problems.* Amer. Math. Soc., Providence, RI.

MELAS, V. B. (2000). Analytical theory of E-optimal designs for polynomial regression. In *Advances in Stochastic Simulation Methods* (N. Balakrishnan, V. B. Melas and S. Ermakov, eds.) 85–115. Birkhäuser, Boston.

MÜHLBACH, G. (1996). On Hermite interpolation by Cauchy-Vandermonde systems: the Lagrange formula, the adjoint and the inverse of a Cauchy-Vandermonde matrix. *J. Comput. Appl. Math.* **67** 147–159.

PEHERSTORFER, F. (1995). Elliptic orthogonal and extremal polynomials. *Proc. London Math. Soc.* **70** 605–624.

PETRUSHEV, P. P. and POPOV, V. A. (1987). *Rational Approximation of Real Functions.* Cambridge Univ. Press.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

PUKELSHEIM, F. and STUDDEN, W. J. (1993). E-optimal designs for polynomial regression. *Ann. Statist.* **21** 402–415.

RATKOWSKY, D. A. (1990). *Handbook of Nonlinear Regression Models.* Dekker, New York.
SPRUILL, M. C. (1987). Optimal designs for interpolation. *J. Statist. Plann. Inference* **16** 219–229.
STANLEY, R. P. (1999). *Enumerative Combinatorics* **2**. Cambridge Univ. Press.
STUDDEN, W. J. (1968). Optimal designs on Tchebycheff points. *Ann. Math. Statist.* **39** 1435–1447.
STUDDEN, W. J. and TSAY, J.-Y. (1976). Remez's procedure for finding optimal designs. *Ann. Statist.* **4** 1271–1279.
SZEGŐ, G. (1975). *Orthogonal Polynomials*, 4th ed. Amer. Math. Soc., Providence, RI.

INSTITUT FÜR STATISTIK
UND WIRTSCHAFTSMATHEMATIK
RWTH AACHEN
D-52056 AACHEN
GERMANY

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907-1399
E-MAIL: studden@stat.purdue.edu