Roy V. Erickson; Václav Fabian; Jan Mařík

An optimum design for estimating the first derivative

Ann. Statist. 23 (4) (1995), 1234–1247

Persistent URL: http://dml.cz/dmlcz/502153

The Annals of Statistics
1995, Vol. 23, No. 4, 1234–1247

AN OPTIMUM DESIGN FOR ESTIMATING THE FIRST DERIVATIVE

BY ROY V. ERICKSON, VÁCLAV FABIAN[1] AND JAN MAŘÍK[2]

Michigan State University

An optimum design of experiment for a class of estimates of the first derivative at 0 (used in stochastic approximation and density estimation) is shown to be equivalent to the problem of finding a point of minimum of the function Γ defined by $\Gamma(x) = \det[1, x^3, \dots, x^{2m-1}]/\det[x, x^3, \dots, x^{2m-1}]$ on the set of all m-dimensional vectors with components satisfying $0 < x_1 < -x_2 < \cdots < (-1)^{m-1}x_m$ and $\prod |x_i| = 1$. (In the determinants, 1 is the column vector with all components 1, and x^i has components of x raised to the i-th power.) The minimum of Γ is shown to be m, and the point at which the minimum is attained is characterized by Chebyshev polynomials of the second kind.

1. Introduction. An optimum design of experiment is considered for estimating the first derivative at 0 of a function f on the real line when only observations, subject to error, of function values are available. In Section 2 the optimum design problem is described and shown to be equivalent to the problem of finding a point of minimum of a function Γ, defined below. The solution is described in Theorem 1.2.

We *assume* throughout that m is an integer, $m \geq 2$. We define Γ as a restriction of a function G; we will need G in Section 3.

DEFINITION 1.1. G is the function defined on the set

$$(1)\qquad \mathscr{D}G = \{x; x = \langle x_1, \dots, x_m\rangle, 0 < x_1 < -x_2 < \cdots < (-1)^{m-1}x_m\}$$

by the relation

$$(2)\qquad G(x) = \frac{\det[1, x^3, x^5, \dots, x^{2m-1}]}{\det[x, x^3, x^5, \dots, x^{2m-1}]}\left(\prod_{j=1}^{m}|x_j|\right)^{1/m},$$

where $1 = \langle 1, 1, \dots, 1\rangle$ and $x^i = \langle x_1^i, \dots, x_m^i\rangle$. [The denominator in (2) is different from 0 by properties of the Vandermonde determinants.] The function Γ is the restriction of G to the set

$$(3)\qquad \mathscr{D}\Gamma = \{x; x \in \mathscr{D}G, \prod |x_i| = 1\}.$$

Received April 1993; revised January 1995.

[1]Research partly supported by NSF Grant DMS-91-01396.

[2]Professor Jan Mařík died January 6, 1994; he is missed by his co-authors, colleagues, friends and former students in the United States, the Czech Republic and elsewhere. [Obituary to appear in *Real Analysis Exchange* **19** (2).]

AMS 1991 *subject classifications.* Primary 62K05, 62L20; secondary 15A15.

Key words and phrases. Stochastic approximation, determinants, linear independence, orthogonal polynomials, Chebyshev polynomials of second kind.

THEOREM 1.2. *The minimal value of* Γ *is* m *and is attained at exactly one point* x, *namely, at* x *defined by*

$$x_i = (-1)^{i-1} 2\cos\left(\frac{m+1-i}{2m+1}\pi\right) \quad \text{for } i = 1, \dots, m. \tag{4}$$

These x_i *are the roots of the polynomial* Q *defined by* $Q(t) = U_m(t/2) + (-1)^m U_{m-1}(t/2)$, *where* U_m *is the degree* m *Chebyshev polynomial of the second kind.*

Proof is given in Section 5.

REMARK 1.3 (Comments on the proof). The problem has arisen already in Fabian (1968a); at that time the solution was not found. At the beginning of a renewed effort, numerical studies and beginnings of the theoretical understanding gave a surprising result: for small m, the minimum of Γ was found to be close to m, and the coordinates of the minimal point x were roots of an integer-coefficient polynomial. This and similar properties discovered later helped us to believe there was a simple explicit answer to the problem and sustained us in the effort to find it.

Attempts to use existing results on optimum design of experiments failed. There, Wynn (1984) is a very useful review. Within the theory, there are strong and simple results [e.g., Kiefer and Wolfowitz (1959, 1960) and Karlin and Studden (1966)], but also problems where results have been obtained by numerical computations with attendant difficulties of local versus global minima [e.g., Mitchell (1974) and Galil and Kiefer (1980)].

The present solution proceeds as follows. First, in Section 3, it is shown that Γ attains its minimum and that the point at which it attains its minimum is stationary (see Definition 3.1). Section 4 describes the property $\Gamma(x) = \gamma$ by three polynomials P, Q and R and uses these to characterize the stationarity of x. This is used in Section 5, where these polynomials, corresponding to a stationary point x, are determined. From there, the unique stationary point x is determined, and this x is the point of minimum.

2. The optimum design problem.

REMARK 2.1. We consider the statistical problem of estimation of the first derivative at 0 of a smooth function f on $(-\infty, \infty)$. The estimates considered are linear combinations of estimates of the differences of f at 0 with the coefficients chosen in such a way that the bias of the estimate is not influenced by derivatives of order $2, \dots, 2m$ of f at 0. The expected squared error of the estimate depends on the steps used in the differences, and the problem studied here is that of the optimal choice of these steps. Such estimates have been used to obtain an improved rate of convergence for a modification of the Kiefer–Wolfowitz stochastic approximation method in Fabian (1967); later, the rate was shown to be optimal by Chen (1988). The estimate was also used by Koronacki (1987) for estimation of densities, and a similar estimate was used by Fabian (1990) in cubic spline estimation of nonparametric regression functions. Estimates of the first derivatives are also used in the response surface methods, and Box and Draper (1959) and Karson, Manson and Hader (1969) discuss optimum designs for problems with a motivation very similar to ours. However, the assumptions and the actual questions treated there differ from ours so that their results are not applicable in our case.

2.2. *The estimate.* Consider a function f on the real line with derivatives, at 0, of order $1, \dots, 2m+1$; denote these derivatives $\varphi_1, \dots, \varphi_{2m+1}$. We assume that it is possible to observe function values subject to error; thus also function differences. Averaging such independent observations we may control (at cost) the variances. This will be summarized later more formally. First, let us consider approximations of φ_1 by linear combinations of function differences:

$$\psi = \frac{1}{2c} \sum_{i=1}^{m} v_i\, d(cu_i), \tag{5}$$

where c is a positive number, $d(y) = f(y) - f(-y)$, $u \in D = \{u; u \in \mathbb{R}^m, 0 < u_1 < \cdots < u_m = 1\}$, U is the transpose of the m-by-m matrix $[u, u^3, \dots, u^{2m-1}]$ and

$$Uv = \langle 1, 0, \dots, 0 \rangle. \tag{6}$$

For $s = 1, \dots, 2m+1$, the sth derivative, at 0, of the function $c \to d(cu_i)$ is $2\varphi_s u_i^s$ if s is odd, and 0 if s is even. Using a Taylor formula and relation (6), we obtain that

$$|\psi - \varphi_1| = c^{2m} \frac{|\varphi_{2m+1} + r(c)|}{(2m+1)!} \sum_{i=1}^{m} v_i u_i^{2m+1} \quad \text{with } \lim_{c \to 0} r(c) = 0. \tag{7}$$

We assume our estimate Y is given by the right-hand side of (5), with $d(cu_i)$ replaced by Y_i for $i = 1, \dots, m$, where $Y_1, \dots, Y_m$ are uncorrelated random variables, and each Y_i has expectation $d(cu_i)$ and variance $2\sigma^2/\xi_i$, with $\sum_i \xi_i = 1$. Under these assumptions,

$$\mathbb{E}Y = \psi, \qquad \operatorname{var}(Y) = \frac{\sigma^2}{2c^2} \sum_{i=1}^{m} \frac{v_i^2}{\xi_i}. \tag{8}$$

The meaning of ξ_i is rather standard in the design of experiment theory since Kiefer and Wolfowitz (1959). Thus if, for each i, Y_i is the arithmetic mean of n_i uncorrelated random variables with expectation $d(cu_i)$ and variance $2\sigma_0^2$, then Y_i has variance $2\sigma_0^2/n_i = 2\sigma^2/\xi_i$, where $\xi_i = n_i/N$, $\sigma^2 = \sigma_0^2/N$ and $N = n_1 + \cdots + n_m$. If ξ_i and N are given, we can determine n_i to satisfy $n_i = N\xi_i$ approximately, or, conservatively, we may choose n_i the smallest integers such that $n_i \geq N\xi_i$.

Suppose we have a bound φ for $|\varphi_{2m+1} + r(c)|$. We then obtain an upper bound e for the expected squared error of Y:

$$e = \Psi^2 + \operatorname{var}(Y), \tag{9}$$

where Ψ is defined by the right-hand side of (7) with $\varphi_{2m+1} + r(c)$ replaced by φ. We are interested in the optimal (i.e., mimimizing e) choice of the vector u, provided c and ξ have been chosen optimally. A simplifying fact is that this optimal choice of u depends only on m, the dimension of the problem. It is easy to derive formulas specifying the optimal c and ξ (see Lemma 2.3 below), but the formula for c involves, implicitly, the bound φ. In most asymptotic considerations, the estimate is applied with σ approaching 0. Then c approaches 0 [cf. (10)] and the goal is to minimize (9) with $\varphi = |\varphi_{2m+1}|$ [as in, e.g., Fabian, (1968a, b)]; φ_{2m+1} can be estimated. The optimization here is done for a given m, but the solution is also useful for deciding which m to use.

In the lemma below, it is assumed that $\varphi > 0$. Consider thus, for a moment, the case $\varphi = 0$ (an unlikely case, in particular, if c is not restricted). If also $\sigma = 0$, then $e = 0$, and the optimization problem is trivial. If $\sigma > 0$, then the minimization concerns only var(Y); if c can be chosen any number, then e does not attain its infimum value 0. However, if c is specified, then var(Y) is minimized by minimizing $\sum_{i=1}^m v_i^2/\xi_i$, that problem has a known solution, see Remark 2.4.

LEMMA 2.3. *Assume $\varphi > 0$. The optimal choice of ξ and c is*

$$(10) \qquad \xi = \frac{\langle |v_1|, \dots, |v_m| \rangle}{\Sigma |v_i|} \quad \text{and} \quad c = \left[\frac{B}{sA}\right]^{1/(2s+2)},$$

where $s = 2m$,

$$(11) \qquad A = \left[\frac{\varphi}{(s+1)!} \sum_{i=1}^m v_i u_i^{s+1}\right]^2 \quad \text{and} \quad B = \frac{\sigma^2}{2}\left(\sum_{i=1}^m |v_i|\right)^2.$$

With these ξ and c,

$$(12) \quad e = (1+s)s^{-s/(s+1)}\left[\frac{\varphi}{(s+1)!}\right]^{2/(s+1)}\left[\frac{\sigma^2}{2}\right]^{s/(s+1)}[h(u)]^{2s/(s+1)},$$

where

$$(13) \qquad h(u) = \left[\sum_{i=1}^m |v_i|\right]\left[\prod_{i=1}^m u_i\right]^{1/m} = \Gamma(Tu)$$

and T is defined on D by $T(u) = (u_1 u_2 \cdots u_m)^{-1/m}\langle u_1, -u_2, \dots, (-1)^{m-1}u_2\rangle$.

PROOF. The optimal ξ minimizes $H(\xi) = \sum_{i=1}^m v_i^2/\xi_i$. By the Schwarz inequality, applied to $\langle |v_1|/\sqrt{\xi_1}, \dots, |v_m|/\sqrt{\xi_m}\rangle$ and $\langle \sqrt{\xi_1}, \dots, \sqrt{\xi_m}\rangle$, $H(\xi) \geq (\Sigma|v_i|)^2$. The ξ given in (10) is optimal because, for it, $H(\xi)$ equals the lower bound established.

With the optimal ξ, $e = Ac^{2s} + Bc^{-2}$. Differentiating gives the result concerning c. For this c,

$$e = A^{1/(s+1)}B^{s/(s+1)}s^{-s/(s+1)}(1+s),$$

and thus (12) holds with

$$(14)\qquad h(u) = \left[\sum_{i=1}^{m} |v_i|\right] \left|\sum_{i=1}^{m} v_i u_i^{2m+1}\right|^{1/s}.$$

Set $a = (\prod_i u_i)^{-1/m}$ and set $z = au$, $w = v/a$. Note that $h(u)$ is equal to (14) with u and v replaced by z and w. Let Z denote the matrix obtained from U by replacing u by z and note that (6) holds with U and v replaced by Z and w. Thus w is the first column of Z^{-1} and thus, with $Z_{1,i}$ the cofactor of the $\langle 1, i\rangle$ element in Z,

$$(15)\qquad w_i = \frac{Z_{1,i}}{\det(Z)} \quad \text{for } i = 1, \dots, m.$$

As a consequence, if c_i are numbers, then

$$(16)\qquad \sum_i c_i w_i = \frac{\det(C)}{\det(Z)},$$

where C is the determinant obtained from Z by changing its first row to $[c_1, \dots, c_m]$. Use (16) with $c_i = z_i^{2m+1}$ to obtain that the left-hand side of (16) is $(-1)^{m-1}\prod_i z_i^2 = (-1)^{m-1}$, and thus

$$(17)\qquad h(u) = \sum_{i=1}^{m} |w_i|.$$

Since $w = v/a$, the first equality in (13) holds.

By properties of Vandermonde determinants, $\det(Z)$ is positive and so are the $\langle 1, i\rangle$ minors of Z. By (15), the sign of w_i is $(-1)^{i-1}$; thus $h(u) = \sum_i (-1)^{i-1} w_i$ and, by (16),

$$(18)\qquad h(u) = \frac{\det(C)}{\det(Z)},$$

where C is Z with its first row replaced by $[1, -1, \dots, (-1)^{m-1}]$.

Next, multiply the even columns in both matrices in (18) by -1, transpose the matrices and set $x_i = (-1)^{i-1} z_i$. Then $h(u) = \Gamma(x)$ with $x = T(u)$ and the second equality in (13) holds.

REMARK 2.4 (A related result). Using results in Kiefer and Wolfowitz (1959), Fabian [(1968a), Theorem 5.1] found the solution to the problem of minimizing the expression $\sum_{i=1}^{m} v_i^2/\xi_i$, mentioned at the end of Section 2.2, giving a heuristic argument that the solution will approximately minimize e. Choosing ξ_i optimal makes the expression equal to $g(u)^2$, where $g(u) = \sum_{i=1}^{m} |v_i|$. The point u, minimizing $g(u)$, is given by $u_i = \cos[(m - i)\pi/(2m - 1)]$ for $i = 1, 2, \dots, m$. For this u, the value of h defined in (13) is given by $h(u) = 2^{1/m}(m - 1/2)$ [see (5.1.4) in Fabian (1968a) and note that v there is $\frac{1}{2}$ times the present v]. This $h(u)$ is slightly larger than the minimal value m. For four values of m, Table 1 gives the last factor in (12) for u optimal, the *suboptimal* u that minimizes $g(u)$ only, and the equidistant choice $u_i = i/m$ for $i = 1, \dots, m$.

TABLE 1
The values of $[h(u)]^{2s/(s+1)}$

	u		
m	**Optimal**	**Suboptimal**	**Equidistant**
2	3.0314	3.3310	3.3310
5	18.6576	19.8210	25.5865
10	80.3086	83.1132	137.5495
15	188.9321	193.4869	372.9216

3. Attainment of minimum.

DEFINITION 3.1. A point x is *stationary* if it is in $\mathscr{D}\Gamma$ and the gradient of G at x is 0.

LEMMA 3.2. *If x is in $\mathscr{D}G$ and a is a positive number, then $G(ax) = G(x)$. The function Γ attains its minimal value at a stationary point x; the minimal value is positive.*

PROOF. The first assertion follows easily from (2). Consider an x in $\mathscr{D}\Gamma$. Dividing each row of the matrix in the denominator D in (2) by the first element gives a Vandermonde matrix, and $D = \prod_i x_i \prod_{s<r}(x_r^2 - x_s^2)$. Expanding the numerator by the minors of the first column, we express it as a sum of terms M_i (for $i = 1,\dots,m$) with $M_i = (-1)^{i-1}\prod'_j x_j^3 \prod'_{s<r}(x_r^2 - x_s^2)$, where the prime means the subscripts in the product are different from i. Since $|x_i|^3 = (-1)^{i-1}x_i^3$ and $\prod_i |x_i| = 1$, we obtain an alternative expression for Γ:

$$\Gamma(x) = \sum_{i=1}^{m} \frac{1}{|x_i^3| \prod_{s<i}(x_i^2 - x_s^2) \prod_{i<r}(x_r^2 - x_i^2)}. \tag{19}$$

There exists a sequence $\langle x_n \rangle$ in $\mathscr{D}\Gamma$ such that $\langle \Gamma(x_n) \rangle$ converges to $\gamma = \inf\{\Gamma(x);\ x \in \mathscr{D}\Gamma\}$. The sequence is bounded, because, if not, $\Gamma(x_n) \to \infty$ by (19) and by (3). Thus, changing the sequence to a suitable subsequence, we obtain $x_n \to x_0$ for an x_0. It is easy to see from (19) that $x_0 \in \mathscr{D}\Gamma$, because, if not, again $\Gamma(x_n) \to \infty$. From the continuity of Γ, it follows that Γ attains its minimal value γ at x_0 and that γ is positive.

Because of the property $G(ax) = G(x)$, the ranges of G and Γ are equal. Thus Γ attains its minimal value γ at a point x, and G has a minimal value at x. Since the gradient of G obviously exists, x is stationary. □

4. A polynomial description.

LEMMA 4.1. *If $x \in \mathscr{D}\Gamma$ and $Q(t) = \prod_i (t - x_i)$, then there is a unique polynomial R of degree at most $m - 1$ such that, for $P = QR$, $P - 1$ is odd. For these P, Q and R, and for $\gamma = \Gamma(x)$, with $[m/2]$ the largest integer less or* *equal to* $m/2$,

$$Q(0) = R(0) = (-1)^{m+[m/2]} \quad \textit{and} \quad P'(0) = -\gamma. \tag{20}$$

PROOF. From $\gamma = \Gamma(x)$ it follows that $\det[1 - \gamma x, x^3, \dots, x^{2m-1}] = 0$. Since the columns 2 to m are linearly independent, it follows that $1 - \gamma x$ is a linear combination of $x^3, x^5, \dots, x^{2m-1}$. Consequently, there is a polynomial P of degree at most $2m - 1$ such that $P - 1$ is odd, and satisfies the second part of (20). The components of x are roots of P. Thus P is divisible by Q and equal to QR with a polynomial R of degree at most $m - 1$. The unicity of P follows from the linear independence of $x, x^3, \dots, x^{2m-1}$ and implies the unicity of R; $Q(0) = \prod_i(-x_i) = (-1)^{m+[m/2]}$ by (1) and (3); and $1 = P(0) = R(0)Q(0)$, proving the first part of (20). □

4.2. *Convention.* Within the context where an x in $\mathscr{D}\Gamma$ is specified, we shall use the notation γ and P, Q and R in the sense of Lemma 4.1.

LEMMA 4.3. *Let x be a point in $\mathscr{D}\Gamma$. Then x is stationary if and only if, for all t,*

$$R(t) = \tfrac{1}{2}\left[2Q(t) - \mu t Q(t) + (-1)^m \mu t Q(-t)\right] \ \textit{where}\ \mu = (-1)^m \gamma/m. \tag{21}$$

PROOF. Write (2) as $G = (f/h)g$. The functions f, h and g do not acquire value 0 at any point in $\mathscr{D}G$ (cf. Definition 1.1 and Lemma 3.2). Denote the partial derivative with respect to the kth coordinate by a subscript k. The phrase *for all k* will mean *for every $k = 1, \dots, m$*. Note that $g(x) = 1$. We have

$$g_k(x) = \frac{1}{m x_k} \tag{22}$$

and $hG_k = f_k g + (f/h)hg_k - (f/h)gh_k$. Thus x is stationary if and only if

$$f_k(x) - \gamma h_k(x) = -\frac{\gamma h(x)}{m x_k} \quad \text{for all } k. \tag{23}$$

The derivatives f_k and h_k are obtained by replacing the kth row of the defining determinant by its derivative and, consequently, the left-hand side in (23) is equal to the determinant of the matrix with rows $r_j = \langle 1 - \gamma x_j, x_j^3, \dots, x_j^{2m-1}\rangle$, for all j, except that $r_k = \langle -\gamma, 3x_k^2, \dots, (2m-1)x_k^{2m-2}\rangle$. Recall that the components of x are roots of the polynomial $P = QR$. Denote the coefficients of P by p_i. If we add, to the first column, the second column multiplied by $p_3, \dots$ and the last column multiplied by p_{2m-1}, then the changed matrix has the first column elements 0 except that the kth element is $P'(x_k)$. However, $P'(x_k) = Q'(x_k)R(x_k)$, since $Q(x_k) = 0$.

Expanding the determinant by the first column then gives

$$f_k(x) - \gamma h_k(x) = (-1)^{k-1} Q'(x_k) R(x_k) W_k \quad \text{for all } k, \tag{24}$$

where $W_k = \det[w^3, \dots, w^{2m-1}]$ and w is the vector obtained from x by deleting the kth component (w thus depends on k).

Next we obtain

$$h(x) = \Big(\prod_i x_i\Big)\Big(\prod_{i<j}(x_j^2 - x_i^2)\Big), \qquad W_k = \Big(\prod_i{}' x_i^3\Big)\Big(\prod_{i<j}{}'(x_j^2 - x_i^2)\Big),$$

where $\prod'$ indicates a product with subscript k omitted.

Thus W_k is not zero. Since $\prod_i |x_i| = 1$, we obtain

$$\begin{aligned}\frac{h(x)}{W_k} &= x_k^3 \prod_{i<k}(x_k^2 - x_i^2)\prod_{j>k}(x_j^2 - x_k^2)\\ &= (-1)^{m-k} x_k^3 \prod_j{}'(x_k + x_j)\prod_j{}'(x_k - x_j).\end{aligned}$$

On the right-hand side, the first product is $(-1)^m Q(-x_k)/(2x_k)$. The second product is nonzero (a fact used again soon below) and is equal to $Q'(x_k)$; this is easy to see when $Q(z)$ is written as $(z - x_k)C$ so that $Q'(z) = C + (z - x_k)C'$. Consequently,

$$\frac{h(x)}{W_k} = (-1)^k \frac{1}{2} x_k^2 Q(-x_k) Q'(x_k). \tag{25}$$

A condition equivalent to (23) is now obtained by dividing both sides of (23) by $W_k Q'(x_k)$ and using (24) and (25): the condition so obtained is

$$R(x_k) = \frac{\gamma}{2m} x_k Q(-x_k) \quad \text{for } k = 1, \dots, m. \tag{26}$$

The proof will be complete when we show that (26) and (21) are equivalent. That (21) implies (26) is immediate. Assume (26) holds. Since the highest terms of the polynomials $tQ(t)$ and $(-1)^m tQ(-t)$ are equal, the polynomials on both sides of (21) are of degree at most m. Next, (21) holds at each x_k and it holds at 0, because $Q(0) = R(0)$. Thus (21) holds. □

5. Unicity of the stationary points. The properties of the polynomials P, Q and R will be used in this section to prove that there is only one stationary point and to determine that point. For that, a rescaling of polynomials P, Q and R will simplify the proofs. In addition, we shall express the new polynomials in terms of polynomials S_k, renormalized Chebyshev polynomials of the second kind.

CONDITION 5.1. In the remainder of this section we assume that x is a stationary point in $\mathscr{D}\Gamma$, use Convention 4.2 and assume that μ is as in Lemma 4.3, that is,

$$\mu = (-1)^m \frac{\gamma}{m}. \tag{27}$$

REMARK 5.2 (Polynomials S_k). Define the polynomials S_k by $S_{-1} = 0$, $S_0 = 1$, $S_1(t) = t$ and

$$S_k = S_1 S_{k-1} - S_{k-2} \quad \text{for } k = 1, 2, \dots \tag{28}$$

[see Rivlin (1990), Exercise 1.5.54b]. It follows that each S_k is a monic polynomial of degree k and the same parity as k [Rivlin (1990), Exercise 1.5.53], and satisfies

$$S_k(2\cos\theta) = \frac{\sin[(k+1)\theta]}{\sin\theta} \quad \text{for every } \theta \in (0, \pi) \tag{29}$$

[Rivlin (1990), (1.23)]. For any i, j in $\mathbb{N}$, and for $i \wedge j$ the minimum of i and j,

$$S_i S_j = \sum_{s=0}^{i \wedge j} S_{i+j-2s}. \tag{30}$$

For $i = 1$, (30) follows immediately from (28) and the proof of (30) is completed by induction, applying the inductive assumption to (30) with the left-hand side replaced by $(S_1 S_{i-1} - S_{i-2})S_j$.

NOTATION 5.3. Define polynomials A, B and C by

$$B(t) = \mu^m Q\left(\frac{t}{\mu}\right), \qquad C(t) = \mu^m R\left(\frac{t}{\mu}\right), \qquad A = BC, \tag{31}$$

and express these in terms of S_k:

$$\begin{aligned} A &= a_0 S_0 + \cdots + a_{2m-1} S_{2m-1}, \qquad B = b_0 S_0 + \cdots + b_m S_m, \\ C &= c_0 S_0 + \cdots + c_{m-1} S_{m-1}; \end{aligned} \tag{32}$$

set a_i, b_i and c_i equal to zero for integer subscripts not appearing in the representations.

We shall use χ_{odd} and χ_{even} for indicator functions of the set of all odd and even integers, respectively, and set $\mathbb{N} = \{0, 1, \dots\}$.

The next lemma is an easy result needed in the proof of Theorem 1.2.

LEMMA 5.4. *The numbers*

$$x_i = (-1)^{i-1} 2\cos\left(\frac{m+1-i}{2m+1}\pi\right) \quad \textit{for } i = 1, \dots, m$$

are the roots of the polynomial $S_m + (-1)^m S_{m-1}$.

PROOF. Let Q denote the polynomial in the assertion, let α denote the argument of the function cosine in the formula for x_i and set $t = x_i$

and $\tau = |t|$. By the parity of S_k, $Q(t) = (-1)^{m(i-1)}S_m(\tau) + (-1)^{m+(m-1)(i-1)}S_{m-1}(\tau)$ and

$$(-1)^{m(i-1)}Q(t) = \frac{\sin\theta_1 + \sin\theta_2}{\sin\alpha} \tag{33}$$

with $\theta_1 = (m+1)\alpha$ and $\theta_2 = (-1)^{m+1-i}m\alpha$.

The numerator in (33) is 0 because, if $m + 1 - i$ is even, then $\theta_1 = -\theta_2$ [mod 2π], and, if $m + 1 - i$ is odd, then $\theta_1 = \pi + \theta_2$ [mod 2π]. □

LEMMA 5.5. *For each t,*

$$C(t) = \tfrac{1}{2}\left[2B(t) - tB(t) + (-1)^m tB(-t)\right]. \tag{34}$$

Also,

$$b_m = b_{m-1} = c_{m-1} = 1, \qquad a_k = 0 \quad \textit{for positive even } k, \tag{35}$$

and, for all k in $\mathbb{N}$,

$$\begin{aligned} c_k &= b_k - \chi_{\text{even}}(m-k)(b_{k+1} + b_{k-1}),\\ b_k &= c_k + \chi_{\text{even}}(m-k)(c_{k+1} + c_{k-1}). \end{aligned} \tag{36}$$

PROOF. Equality (34) follows from the assumed stationarity of x by Lemma 4.3. From (28) and the parity properties of S_k, we obtain that $tS_k(qt) = q^k[S_{k+1}(t) + S_{k-1}(t)]$ for $q \in \{-1, 1\}$. From (34), we obtain

$$\begin{aligned} C &= \tfrac{1}{2}\sum_{k=0}^{\infty} b_k\left[2S_k - (S_{k+1} + S_{k-1}) + (-1)^{m-k}(S_{k+1} + S_{k-1})\right]\\ &= \sum_{k=0}^{\infty} b_k\left[S_k - \chi_{\text{odd}}(m-k)(S_{k+1} + S_{k-1})\right]\\ &= \sum_{k=0}^{\infty} S_k\left[b_k - \chi_{\text{even}}(m-k)(b_{k+1} + b_{k-1})\right]. \end{aligned}$$

Since $C = \sum_{k=0}^{\infty} c_k S_k$, we obtain the first part of (36) for all $k \geq 0$.

When $m - k$ is odd, then both relations in (36) become $b_k = c_k$. For $m - k$ even, the first relation is $c_k = b_k - b_{k+,1} - b_{k-1} = b_k - c_{k+1} - c_{k-1}$, which is the second relation. This proves (36). Further, since B and S_m are monic, we have $b_m = 1$. Using (36) for b_m and then for b_{m-1} gives the first part in (35). The second part follows from the fact that, for k even, S_k is even and $A - 1$ is odd. □

NOTATION 5.6. For k in $\mathbb{N}$, set

$$e_k = \sum_{i=0}^{k} b_i c_{k-i}, \qquad f_k = \sum_{l=0}^{\infty} b_l c_{k+2+l}, \qquad g_k = \sum_{l=0}^{\infty} b_{k+2+l} c_l. \tag{37}$$

LEMMA 5.7. *Suppose k is even. Then*

$$e_k = \begin{cases} f_k + g_k, & \text{if } k \geq 2, \\ 0, & \text{if } k \geq m-1. \end{cases} \tag{38}$$

PROOF. From relations (31), (32) and (30), we obtain

$$A = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} b_i c_j S_i S_j = \sum_{l=0}^{\infty} \sum_{j=0}^{\infty} b_i c_j \sum_{l=0}^{i \wedge j} S_{i+j-2l} \quad \text{and} \quad a_k = \sum_{l=0}^{\infty} \sum_{i=0}^{k} b_{i+l} c_{k-i+l}.$$

Write a_{k+2} with the summands corresponding to $i = 0$ and $i = k+2$ separately, thus obtaining

$$a_{k+2} = f_k + g_k + \sum_{l=0}^{\infty} \sum_{i=1}^{k+1} b_{i+l} c_{k+2-i+l}. \tag{39}$$

Write a_k, treating the case $l = 0$ separately, and change the summation indices as indicated:

$$a_k = e_k + \sum_{l=1}^{\infty} \sum_{i=0}^{k} b_{(i+1)+(l-1)} c_{k+2-(i+1)+(l-1)} = e_k + \sum_{l=0}^{\infty} \sum_{i=1}^{k+1} b_{i+l} c_{k+2-i+l}.$$

This and (39) give the first part in (38), because $a_k = a_{k+2} = 0$ if k is positive even by (35). If k is even and $k \geq m-1$, then all summands in the expressions for f_k and g_k in (37) are trivially 0 and thus also the second part in (38) holds.

LEMMA 5.8. *We have*

$$-c_{i-1} = c_i = b_i = b_{i+1} \quad \text{if } m-i \text{ is odd} \tag{40}$$

and

$$e_k = 0 \quad \text{if } k \text{ is positive even.} \tag{41}$$

PROOF. Denote by I the set of integers i such that $m - i$ is odd. Let D_s denote the condition that (40) holds for all i in I such that $i \geq s$. Introduce a weaker condition:

C_s: (a) $c_i = b_i = b_{i+1}$ holds for every i in I such that $i \geq s$; (b) $c_i = -c_{i-1}$ holds for every i in I such that $i \geq s + 2$.

First we shall prove:

(i) if C_s holds for an s in I such that $s \leq m-1$, then D_s holds.

Assume the premise of the implication. It is enough to prove that $c_s = -c_{s-1}$. If $s \leq -1$, that relation holds trivially. Thus, it remains to consider $s \geq 0$. Set $k = m + s - 1$; k is even and satisfies $m - 1 \leq k \leq 2(m-1)$. By (38), $e_k = 0$. Thus, using (37), we obtain

$$0 = [b_s c_{m-1} + b_{s+1} c_{m-2}] + \cdots + [b_{m-3} c_{s+2} + b_{m-2} c_{s+1}] + [b_{m-1} c_s + b_m c_{s-1}].$$

In all the brackets except the last one, condition C_s applies and makes the sum in the bracket 0. Thus also the last bracket is zero. Since $b_{m-1} = b_m = 1$ by (35), we obtain $c_s = -c_{s-1}$ and D_s holds.

Next we shall prove:

(ii) if C_s holds for an s in I such that $1 \leq s \leq m-1$, then D_{s-2} holds.

Assume the premise of (ii). By (i), D_s holds. We have $b_{s-1} = c_{s-1} + c_s + c_{s-2}$ by (36) and $c_{s-1} + c_s = 0$ by D_s, so that $b_{s-1} = c_{s-2}$. By (36), we have $c_{s-2} = b_{s-2}$ and (a) of condition C_{s-2} holds. Part (b) of condition C_{s-2} follows from D_s and thus C_{s-2} holds. By (i), D_{s-2} holds. This proves (ii).

For $s \geq m+1$, condition D_s holds trivially. For $s = m-1$, part (a) of condition C_s follows from (35), and part (b) holds trivially. Properties (i) and (ii) and an induction show that D_s holds for all s in I such that $s \geq -1$. For $i \leq -2$, (40) is trivially satisfied. This proves the first assertion.

Consider the second assertion. If $k \geq m-1$, then $e_k = 0$ by (38). Consider the case $2 \leq k \leq m-2$. Then, by (37),

$$e_k = b_0 c_k + [b_1 c_{k-1} + b_2 c_{k-2}] + \cdots + [b_{k-1} c_1 + b_k c_0]. \tag{42}$$

Suppose m is even. We obtain, by (40), that $b_0 = b_{-1} = 0$ and that each of the sums in brackets is 0 by (40). Thus $e_k = 0$. Suppose m is odd. Then, by (40), $c_0 = -c_{-1} = 0$, and the last term on the right-hand side can be omitted; moving the brackets one term left and using the same argument as before gives $e_k = 0$ again. This proves the second assertion. □

LEMMA 5.9. *We have* $b_0 = \cdots = b_{m-2} = 0$ *and* $c_1 = \cdots = c_{m-3} = 0$.

PROOF. For even nonnegative k, consider the condition

C_k: $c_i = b_i = 0$ for all $i = 0, 1, \ldots, m-3-k$.

Note that C_k implies C_{k+2} for even nonnegative k and that C_k is trivially satisfied for $k \geq m-2$.

Suppose C_k holds for an even k such that $2 \leq k \leq m-2$. Equalities (38) and (41) give $f_k + g_k = 0$ for all even positive k. Relation $b_m c_{m-2-k} = 0$ is obtained from the defining formula (37), omitting terms that are trivially 0 and those that are 0 because of C_k. Since $b_m = 1$ by (35), it follows that $c_{m-2-k} = 0$. Applying (40) for $i = m-1-k$, we obtain that also

$$b_{m-k} = b_{m-1-k} = c_{m-1-k} = -c_{m-2-k} = 0. \tag{43}$$

In addition, (36) gives $c_{m-2-k} = b_{m-2-k} - b_{m-1-k} - b_{m-3-k}$ and thus, because of C_k and (43), we obtain $b_{m-2-k} = 0$.

It follows that C_k implies C_{k-2}. By induction, C_0 holds and $b_i = c_i = 0$ for $i = 1, \ldots, m-3$. By (40) also $b_{m-2} = 0$. This completes the proof. □

LEMMA 5.10. *We have*

$$B = S_m + S_{m-1}, \gamma = m \quad \textit{and} \quad Q = S_m + (-1)^m S_{m-1}. \tag{44}$$

PROOF. The first relation in (44) follows from (35) and Lemma 5.9. $|B(0)| = 1$ follows from the first relation in (44) and (28). Similarly, $|Q(0)| = 1$ by (20), and (31) then implies $|\mu| = 1$. Since γ is positive by Lemma 3.2, we obtain the second relation in (44) from (27). As a consequence, $\mu = (-1)^m$. By (31) and the first part in (44), $Q(t) = \mu(S_m(\mu t) + S_{m-1}(\mu t))$, and the last relation in (44) follows from the parity of S_k (see Remark 5.2). □

PROOF OF THEOREM 1.2. By Lemma 3.2, Γ attains its minimal value γ at a stationary point x. By Lemma 5.10, the components of x are the roots of Q specified in (44), $\gamma = m$. This is the same Q as in Theorem 1.2, and (4) follows from Lemma 5.4. □

REMARK 5.11. We do not need to determine the polynomials C and R, but it may be of interest that they are also of simple form. Indeed, using subscripts to indicate the independence of Q, R and P on m, we have

$$C_m = S_{m-1} - S_{m-2}, \quad R_m = (-1)^m Q_{m-1}, \quad P_m = (-1)^m S_{2m-1} + 1. \tag{45}$$

The first relation follows from Lemma 5.9 since $c_{m-1} = 1$ by (35) and $c_{m-2} = -c_{m-1}$ by (40). The second relation follows then from (27). To prove the last relation, note that

$$P_m = Q_m R_m = (-1)^m [S_m S_{m-1} - S_{m-1} S_{m-2}] + [S_{m-1}^2 - S_m S_{m-2}]$$

and use (29). It is also easy to show that $Q_m(t)$ is proportional to $P_m^{(1/2,-1/2)}(t/2)$ for m even and to $P_m^{(-1/2,1/2)}(t/2)$ for m odd, where $P_m^{(\alpha,\beta)}$ are the Jacobi polynomials [see Szegö (1939), Section 4.1].

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.

CHEN, H. (1988). Lower rate of convergence for locating a maximum of a function. *Ann. Statist.* **16** 1330–1334.

FABIAN, V. (1967). Stochastic approximation of minima with improved asymptotic speed. *Ann. Math. Statist.* **38** 191–200.

FABIAN, V. (1968a). On the choice of design in stochastic approximation methods. *Ann. Math. Statist.* **39** 457–466.

FABIAN, V. (1968b). On asymptotic normality in stochastic approximation. *Ann. Math. Statist.* **39** 1327–1333.

FABIAN, V. (1990). Complete cubic spline estimation of non-parametric regression functions. *Probab. Theory Related Fields* **85** 57–64.

GALIL, Z. and KIEFER, J. (1980). Time- and space-saving computer methods, related to Mitchell's DETMAX, for finding D-optimum designs. *Technometrics* **22** 301–313.

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

KARSON, M. J., MANSON, A. R. and HADER, R. J. (1969). Minimum bias estimation and experimental design for response surfaces. *Technometrics* **11** 461–476.

KIEFER, J. and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12** 363–366.

KORONACKI, J. (1987). Kernel estimation of smooth densities using Fabian's approach. *Statistics* **18** 37–47.

MITCHELL, T. J. (1974). An algorithm for the construction of "D-optimal" experimental designs. *Technometrics* **16** 203–210.

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

SZEGŐ, G. (1939). *Orthogonal Polynomials*. Amer. Math. Soc., Providence, RI.

WYNN, H. (1984). Jack Kiefer's contributions to experimental design. *Ann. Statist.* **12** 416–423.

DEPARTMENT OF STATISTICS AND PROBABILITY
AND DEPARTMENT OF MATHEMATICS
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824-1027