ON A THEOREM OF HOEL AND LEVINE ON EXTRAPOLATION DESIGNS

By J. Kiefer[1] and J. Wolfowitz[2]

Cornell University

0. Summary. Recent results [5] of Hoel and Levine (1964), which assert that designs on $[-1, 1]$ which are optimum for certain polynomial regression extrapolation problems are supported by the "Chebyshev points," are extended to cover other nonpolynomial regression problems involving Chebyshev systems. In addition, the large class of linear parametric functions which are optimally estimated by designs supported by these Chebyshev points is characterized.

1. Introduction. Let $f = (f_0, f_1, \cdots, f_m)$ be a row vector of $m + 1$ continuous real-valued functions on a compact set X, where $m > 0$. (Unprimed vectors will ordinarily denote row vectors, and the transpose of a vector or matrix will be denoted by a prime.) The f_i are assumed to be linearly independent on X. A design is a probability measure ξ (which can always be assumed discrete) on X. (For a discussion of this see [8].)

Write

$$m_{ij}(\xi) = \int f_i f_j \, d\xi, \qquad M(\xi) = \{m_{ij}(\xi), 0 \leqq i, j \leqq m\}. \tag{1.1}$$

If N uncorrelated observations with equal variance σ^2 (known or unknown) are made, taking $N\xi(x)$ observations at x for each x in X, and if the expected value of an observation at x is $\theta f(x)' = \sum_0^m \theta_i f_i(x)$ where $\theta = (\theta_0, \cdots, \theta_m)$ with the θ_i unknown real parameters, then, if $M(\xi)$ is nonsingular, $\sigma^2 N^{-1} M^{-1}(\xi)$ is the covariance matrix of best linear estimators of the vector θ. Moreover, setting

$$V(a, \xi) = aM^{-1}(\xi)a' \tag{1.2}$$

where $a = (a_0, a_1, \cdots, a_m)$ with the a_i real, $\sigma^2 N^{-1} V(a, \xi)$ is the variance of the best linear estimator of the linear parametric function $\theta a'$. The function V of (1.2) is defined to have the same meaning even if M is singular; in particular, $V(a, \xi) = \infty$ if $\theta a'$ is not estimatable under ξ.

As has been discussed in other papers, we do not restrict ξ to take on values which are integral multiples of N^{-1}. This allows us to obtain optimum design characterizations which cannot be obtained under that restriction, and at the same time yields designs which can be implemented in practice through the use of closely related ξ's which do take on only values which are integral multiples of N^{-1}.

Received 21 April 1965.

[1] Research supported by the Office of Naval Research under Contract No. Nonr 266(04) (NR 047-005).

[2] Research supported by the U. S. Air Force under Contract No. AF 18(600)-685, monitored by the Office of Scientific Research.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics*. www.jstor.org

One problem of interest is to characterize, for each a, a design ξ, termed *a-optimum*, which minimizes $V(a, \xi)$; work on this problem has been done in [4] Elfving (1952), [3] Chernoff (1953), and [8] Kiefer and Wolfowitz (1959). A particular way in which this problem can arise, and which is of considerable practical importance, is that f is extended continuously so as to be defined on the set $Y \cup X$, and that for some point e in Y it is required to choose a design which estimates optimally the regression $\theta f(e)'$ at the point e, the design ξ still being restricted to be a probability measure on the set X of points at which we are permitted to take observations. For e in $Y - X$ (resp., $Y \cap X$) this may be called the problem of *extrapolation* (resp., *interpolation*) of the estimated regression to the point e of the set Y on which the regression function is of interest to us. This is clearly the problem of finding an a-optimum design when $a = f(e)$. (Other extrapolation and interpolation problems, such as that of minimizing $\max_{y \varepsilon Y} V(f(y), \xi)$ for certain sets Y other than those of the type mentioned in the next two paragraphs, are considered in [9] Kiefer and Wolfowitz (1964a, b); when $Y = X$, this problem has been considered extensively in other papers.)

Hoel and Levine (1964) [5] have considered this extrapolation problem in the important case of univariate polynomial regression, where $X = [-1, 1]$, $Y = (-\infty, \infty)$, $f_i(y) = y^i$ for y in Y. Their elegant main result is that, for $|e| > 1$, there is an $f(e)$-optimum design which, for every e, is supported by the same set of $m + 1$ points (with weights which depend on e), namely, the "Chebyshev points" which were shown in [8] to be the support of the a-optimum design when $a = (0, 0, \cdots, 0, 1)$. (It will often be helpful to think of this a as $\lim_{e \to \pm\infty} [f_m(e)]^{-1} f(e)$.) It is well known (see Example 2 of Section 5) that this conclusion cannot be extended to the $f(x)$-optimum design for $|x| \leqq 1$.

(We are indebted to Dr. T. J. Rivlin for pointing out that part of the above-mentioned development on p. 1556 of [5] which shows that $\sum_0^m |L_i(e)|$ is maximized by the Chebyshev points, where the L_i are Lagrange interpolation polynomials, rediscovers a result proved by S. Bernstein on p. 186 of [2].)

A second result of Hoel and Levine is that, for e sufficiently large, their $f(e)$-optimum design also minimizes $\max_{-1 \leqq y \leqq e} V(f(y), \xi)$, which is proportional to the maximum over the interval $[-1, e]$ of the variance of the estimated regression. (In most of the literature to which we refer, $V(f(y), \xi)$ is denoted by $d(y, \xi)$.)

It is natural to ask then, whether in this polynomial case there are vectors a other than constant multiples of those in the one-dimensional set $\{f(e), |e| > 1\}$, such that there is an a-optimum design supported by this same set of Chebyshev points. One of the results of the present paper (Theorems 1 and 2 of Section 4 and Example 2 of Section 5) is that there are many such a, in fact, an $(m + 1)$-dimensional set of them in the $(m + 1)$-dimensional space of all a. This apparent anomaly (in view of the low dimensionality of these designs in the space of all admissible designs) is discussed after the introduction of necessary nomenclature, in the first paragraph of Section 4, and is illustrated in Example 2(b) of Section 5.

It is also natural to ask whether the two Hoel-Levine results and the result discussed in the previous paragraph can be extended to examples other than that of polynomial regression. Theorems 4, 5, 1 and 2 give such extensions under assumptions stated in the next section; the main assumptions (1 and 2) are related to the behavior of the f_i for a related Chebyshev approximation problem. (The proof of Theorem 4 reduces, in the polynomial case, to one which differs from that of [5].) It seems interesting to determine, under these assumptions, the set of vectors a for which there is an optimum design supported by (that is, which assigns positive probability to, and only to) the Chebyshev points, and Theorems 1 and 2 show that this is the set T^* defined in the next section. Also of interest is the set $\bar{T}^*$ of vectors for which there is an optimum design supported by some subset of the Chebyshev points. $\bar{T}^*$, which is not generally merely the closure of T^* (as will be seen in Example 2(b)), does not permit as simple an analysis as T^*. For the most part we are concerned with a subset R^* of T^*, where a related Chebyshev approximation problem has a solution of a particular form, and where the optimum design is unique (as it need not be in $T^* - R^*$). The sets R^* and $T^* - R^*$ are not difficult to characterize explicitly (see (2.20), Theorem 2, and (2.27)), but Theorem 3 describes the inclusion in R^* of a set A^* of vectors which is sometimes also easy to compute, namely, the set of vectors a for which $a'\theta$ is not estimatable for any design on fewer than $m + 1$ points. Theorem 3 is used in establishing Theorem 4.

Section 2 contains nomenclature, definitions, assumptions, and statements of results from previous papers which we shall use. Section 3 contains proofs of auxilliary lemmas which are used in the proofs of our main results (Theorems 1–5) in Section 4; Remark 5 in the latter section describes some further extensions of our results. Finally, Section 5 contains some examples which illustrate the relationship among T^*, R^*, A^*, and other sets we shall consider.

The preliminary propositions, examples, and theorems of Section 2 will be numbered decimally (2.x). All other theorems and examples, and all lemmas and remarks, will be numbered consecutively without indication of section.

2. Preliminaries. Our basic model of f, X, and Y will be as described in Section 1. Throughout this paper, except when explicitly stated to the contrary (as, in particular, in the extensions of Remark 5), X will be $[-1, 1]$ and Y will be $(-\infty, \infty)$. We shall usually denote points of X, Y and $Y - X$ by x, y, and e, respectively. As described in Section 1, we always assume, without further statement in the sequel,

ASSUMPTION 0. $f_0, f_1, \cdots, f_m$ are linearly independent on $[-1, 1]$ and continuous on $(-\infty, \infty)$.

The functions $f_0, \cdots, f_k$ are called a *Chebyshev system* on the set U if every linear combination $\sum_0^k c_i f_i$, with not all of the real constants c_i zero, has k or fewer zeros on U. This condition can be rephrased in any of several equivalent forms which will be useful in what follows: For distinct $x_0, x_1, \cdots, x_k$ in U,

(i) the vector $\sum_0^k c_i(f_i(x_0), \cdots, f_i(x_k))$ is not the zero vector unless all c_i are 0;

(ii) the matrix $\{f_i(x_j), 0 \leqq i, j \leqq k\}$ has rank $k + 1$;

(iii) the vector $(f_0(x_k), \cdots, f_k(x_k))$ cannot be represented as $\sum_0^{k-1} c_i(f_0(x_i), \cdots, f_k(x_i))$.

In the lemmas and theorems of Sections 3 and 4, we shall make use of the following five assumptions, stating explicitly where any of them is made. The first two assumptions are used in all five of the theorems.

Assumption 1. The functions $f_0, f_1, \cdots, f_{m-1}$ constitute a Chebyshev system on $[-1, 1]$.

If F is a continuous real-valued function on $[-1, 1]$, we shall say that F *changes direction* at x_0 if $-1 < x_0 < 1$ and if F has a local maximum or minimum at x_0. In particular, if F is constant on any open subinterval of $[-1, 1]$, it will be said to have infinitely many changes of direction.

Assumption 2. For each q in E^{m+1} (Euclidean $(m + 1)$-space), the function fq' on $[-1, 1]$ either has fewer than m changes of direction, or else is constant on $[-1, 1]$.

In proving our generalizations of the Hoel-Levine results we shall also use

Assumption 3. The continuous functions $f_0, f_1, \cdots, f_m$ are a Chebyshev system on $(-\infty, \infty)$;

Assumption 4. For $0 \leqq i < m$, we have $\lim_{e \to \pm\infty} f_i(e)/f_m(e) = 0$;

Assumption 5. $|f_m(e)|$ is strictly increasing (resp., decreasing) when e(resp., $-e$) is sufficiently large; $\lim_{e \to \pm\infty} |f_m(e)| = +\infty$; and, for $0 \leqq i < m$, the quantity

$$\sup_{0<|\Delta|\leqq 1} |f_i(e + \Delta) - f_i(e)|/|f_m(e + \Delta) - f_m(e)|$$

remains bounded as $e \to \pm\infty$.

Assumption 5 can of course be phrased in terms of derivatives, if they exist. We next remark briefly on Assumptions 1 and 2.

Remark 1(a). We shall prove in Lemma 2 that Assumption 2 implies that $f_0, f_1, \cdots, f_m$ constitute a Chebyshev system on $[-1, 1]$ with $\sum_0^m d_j f_j(x) \equiv 1$ for some $d_0, \cdots, d_m$. In the other direction we have the obvious

Proposition 2.1. *If $f_0 \equiv 1$, the f_i are differentiable, and $\{df_i(x)/dx, 1 \leqq i \leqq m\}$ constitute a Chebyshev system on $[-1, 1]$, then Assumption 2 holds.*

However, simple examples show that Assumption 2 is stronger than $\{f_i, 0 \leqq i \leqq m\}$ being a Chebyshev system. An illustration is

Example 2.1. If $\{f_i, 0 \leqq i \leqq m\}$ satisfies Assumption 2 and h is positive and continuous on $[-1, 1]$, then $\{hf_i, 0 \leqq i \leqq m\}$ is Chebyshev but need not satisfy Assumption 2 (e.g., let $f_i(x) = x^i$ and $h(x) = 2 + \sin 10x$).

Remark 1(b). The form of Assumptions 1 and 2 appears to single out f_m for special treatment. This asymmetric form of the first two assumptions can be replaced by a more symmetric form. For example, Assumptions 1 and 2 can be replaced by the hypotheses of the following:

Proposition 2.2 *Assumption 2 and the assumption that $f_0, f_1, \cdots, f_m$ is Chebyshev on a set $[-1, 1] \cup \{v\}$ for some point $v \not\in [-1, 1]$ imply that Assumptions 1 and 2 are satisfied for $\{\bar{f}_i, 0 \leqq i \leqq m\}$ where $\bar{f}' = Qf'$ for some nonsingular Q. In particular, this conclusion is implied by Assumptions 2 and 3.*

(The second assumption of Proposition 2.2, without the v, is the conclusion of Lemma 2 obtained under Assumptions 1 and 2. Example 2.1 above shows that Assumption 3 alone does not imply Assumption 2.)

PROOF OF PROPOSITION 2.2. We clearly need only prove that Assumption 1 is satisfied for $\bar{f}_0, \cdots, \bar{f}_{m-1}$ with some nonsingular Q. By linear independence there is a nonsingular Q such that $\bar{f}(v)' \equiv Qf(v)' = (0, 0, \cdots, 0, 1)'$. But then, writing $x_m = v$, if $0 \leqq x_0 < x_1 < \cdots < x_{m-1} \leqq 1$ we have

$$\det \{\bar{f}_i(x_j), 0 \leqq i, j < m\} = \det \{\bar{f}_i(x_j), 0 \leqq i, j \leqq m\}$$
$$= \det Q \det \{f_i(x_j), 0 \leqq i, j \leqq m\} \neq 0,$$

the last by the Chebyshev assumption for $[-1, 1] \cup \{v\}$. This proves the desired result.

The following example shows that the Chebyshev nature of $\{f_0, f_1, \cdots, f_m\}$ on $[-1, 1]$ does not by itself imply that $\{\bar{f}_0, \cdots, \bar{f}_{m-1}\}$ is Chebyshev for some Q, even if $f_0(x) \equiv 1$:

EXAMPLE 2.2. Suppose $f_0(x) \equiv 1$ on $[-1, 1]$, and let $f_1(x) = \cos \pi x$ and

$$f_2(x) = \sin \pi x \qquad \text{if} \quad -1 \leqq x \leqq \tfrac{1}{2},$$
$$= (1 + \sin \pi x)/2 \qquad \text{if} \quad \tfrac{1}{2} \leqq x \leqq 1.$$

This system $\{f_0, f_1, f_2\}$ was given in [11] Volkov (1958) as an example of a Chebyshev system on $[-1, 1]$ which cannot be extended to be Chebyshev on a larger interval (our example must of course be of this nature by Proposition 2.2). To see that this system is indeed Chebyshev, it is only necessary to graph the function $c_1f_1 + c_2f_2$ (with c_1 and c_2 not both zero) in each of the four cases $0 \leqq \pm c_1 \leqq c_2$ and $0 \leqq c_2 < \pm c_1$ and to note that this function assumes each value at most twice. If now $\{\bar{f}_0, \bar{f}_1\}$ were Chebyshev on $[-1, 1]$ with $\bar{f}_i = \sum_j q_{ij}f_j$, some linear combination of $\bar{f}_0$ and $\bar{f}_1$ would be of the form $g_1 = c_1f_1 + c_2f_2$ and (by the Chebyshev property) would have at most one zero on $[-1, 1]$. An examination of the four cases described above shows that any function of this form has two zeros on $[-1, 1]$, except for multiples of $f_1 + cf_2$ with $c > 2$. For g_1 of this last form we have $g_1'(r) = 0$ for a single r satisfying $-1 < r < -\frac{1}{2}$, where we use primes to denote derivatives in this paragraph. Let g_0 be a linear combination of $\bar{f}_0$ and $\bar{f}_1$ with g_0 positive throughout $[-1, 1]$; the existence of such a g_0 is shown in the first paragraph of Example 1 of Section 5. The development of that paragraph also shows (since $g_1(-1) < 0 < g_1(1)$) that g_1/g_0 is nondecreasing on $[-1, 1]$. Since $(g_1/g_0)' = g_1'/g_0 - g_1g_0'/g_0^2$, we see that $(g_1/g_0)'(r) \geqq 0$ if and only if $g_0'(r) \geqq 0$. If $g_0'(r) > 0$, then, since $g_0'(-1) < 0$ for g_1/g_0 to be increasing (because $g_1(-1) < 0, g_1'(-1) < 0$), there is an s with $-1 < s < r$ and $h'(s) = 0$; but then $(g_1/g_0)'(s) < 0$, a contradiction. Hence $g_0'(r) = 0$. But then, since $\{1, g_1, f_1\}$ span the same vector space as $\{f_0, f_1, f_2\}$ and $f_1'(r) \neq 0$, we conclude that $g_0 = k_0 + k_1g_1$ for some constants k_i with $k_0 \neq 0$. But then $g_1 + k_3(g_0 - k_1g_1)$ is a linear combination of $\bar{f}_0$ and $\bar{f}_1$ which, for a suitable choice of the constant k_3, has two zeros in a small neighborhood of r (since $g_1''(r) \neq 0$).

For reference in the proofs of Sections 3 and 4, we now summarize in Theorems 2.1–2.3 the known results of Chebyshev approximation theory and optimum design theory which will be used. Suppose $h_0, h_1, \cdots, h_m$ are continuous real-valued functions on $[-1, 1]$. Then $\sum_0^{m-1} c_i^* h_i$ is called a *best Chebyshev (uniform) approximation* to h_m on $[-1, 1]$ if $\max_{-1 \leq x \leq 1} |h_m(x) - \sum_0^{m-1} c_i h_i(x)|$ is minimized, over all choices of the real constants c_i, by the choice $c_i = c_i^*$ $(0 \leq i < m)$. The vector $c^* = (c_0^*, \cdots, c_{m-1}^*)$ is then called a *Chebyshev vector*. A classical result of approximation theory ([1] Achieser (1956), p. 74) is

THEOREM 2.1. *If $h_0, \cdots, h_{m-1}$ form a Chebyshev system, then there is a unique Chebyshev vector c^*, and it is characterized by the fact that there are at least $m + 1$ points at which the residual $h_m - \sum_0^{m-1} c_i^* h_i$ attains its maximum in absolute value, this maximum being assumed with at least $m + 1$ successive alternations in sign.*

If c^* is a Chebyshev vector, we shall denote by $B(c^*)$ the set where $|h_m - \sum_0^{m-1} c_j^* h_j|$ attains its maximum on $[-1, 1]$.

We shall be concerned both with cases where $h_0, \cdots, h_{m-1}$ are a Chebyshev system, and with cases where they are not (see Lemma 8, etc.). It will be shown in Lemma 3 that, if Assumptions 1 and 2 hold, which in particular imply that Theorem 2.1 holds with $h_i = f_i$ $(0 \leq i \leq m)$, the set $B(c^*)$ consists of exactly $m + 1$ points $-1 = x_0^* < x_1^* < \cdots < x_m^* = 1$. We shall denote by X_m^* the set of these *Chebyshev points*.

If $h_0, h_1, \cdots, h_m$ are continuous real-valued linearly independent functions on $[-1, 1]$, and if $\sum_0^m \psi_i h_i(x)$ is the expected value of an observation at x, where the ψ_i are unknown real parameters and the observations are uncorrelated with common variance σ^2, we are led to consider the game with payoff function

$$K(\xi, q) = \int_{-1}^{1} [h_m(x) - \sum_0^{m-1} q_j h_j(x)]^2 \xi(dx), \tag{2.1}$$

where the minimizing player may (by convexity of K in q) be restricted to pure strategies which are vectors $q = (q_0, \cdots, q_{m-1})$ of real components, while the maximizing player has probability measures ξ (which can be taken to be discrete) for his mixed strategies. From Section 2 of [8], we have the following results:

THEOREM 2.2. *The variance $\sigma^2 N^{-1} v_m(\xi)$ of the best linear estimator of ψ_m when the design ξ is used satisfies*

$$v_m(\xi) = 1/\min_q K(\xi, q). \tag{2.2}$$

Hence, ξ^* is optimum for estimating ψ_m if and only if it is maximin for the game with payoff K; i.e., *if and only if*

$$\min_q K(\xi^*, q) = \max_\xi \min_q K(\xi, q). \tag{2.3}$$

THEOREM 2.3. *The game with payoff K is determined, the minimizing player has a pure minimax strategy q^*, and the maximizing player has a maximin strategy*

ξ^* on $m + 1 - p$ points, where p is the dimensionality of the convex set of Chebyshev vectors for approximating h_m by $h_0, \cdots, h_{m-1}$ on $[-1, 1]$. The minimax strategies coincide with the Chebyshev vectors. (Hence, from standard game theory, ξ is maximal relative to q and q is minimal with respect to ξ, if and only if ξ is optimum and q is Chebyshev.) The design ξ^* is maximin if and only if, for any Chebyshev vector q^*, $\xi^*(B(q^*)) = 1$ and ξ^* satisfies the "orthogonality relations"

$$\int_{-1}^{1} [h_m(x) - \textstyle\sum_0^{m-1} q_j^* h_j(x)] h_i(x) \xi^*(dx) = 0, \qquad 0 \leqq i < m. \tag{2.4}$$

In the setting of Section 1 and the first paragraph of the present section, we considered linear parametric functions $\theta a'$ for a in $B^* = E^{m+1} - \{0\}$ where E^k denotes Euclidean k-space and 0 denotes the origin in whatever Euclidean space is under consideration (or, where appropriate, a matrix of zeros). Clearly, for every a and every real $\lambda \neq 0$, ξ is a-optimum if and only if it is (λa)-optimum. Thus, from the point of view of characterizing a-optimum designs, one could replace B^* by the real projective m-space P^m of its equivalence classes under the equivalence $a \sim \lambda a$ for all $\lambda \neq 0$. Throughout most of the developments of Sections 3 and 4, it is not profitable to do this, and therefore *all starred sets, such as A^*, R^*, and T^* will be regarded as subsets of B^**. However, because of the special role of the last coördinate θ_m in Assumption 1 (where f_m, of the $m + 1$ functions f_i, is absent), it is sometimes more convenient, especially in explicit representations of the optimum designs for various parametric functions, to consider the a's in terms of two disjoint sets $B = E^m$ and $B_0 = P^{m-1}$, as follows: To each a with $a_m = 1$ we make correspond the vector $b = (b_0, \cdots, b_{m-1})$ in B defined by $b_i = a_i$ $(0 \leqq i < m)$, and conversely; we shall thus think of the m-vector b as corresponding to the linear parametric function $\theta_m + \sum_0^{m-1} b_i\theta_i$ and shall call a design which is optimum for this linear parametric function *b-optimum*. Of course, for any a with $a_m \neq 0$ (whether or not $a_m = 1$), the a-optimum designs will then coincide with the b-optimum designs with

$$b = (a_0/a_m, \cdots, a_{m-1}/a_m).$$

Similarly, the a's with $a_m = 0$ but not all $a_i = 0$ correspond to points in B_0. We let Γ denote the mapping of B^* onto $B \cup B_0$ under this identification. Throughout this paper we shall use a to mean an element of B^* and b to mean an element of B. The disadvantages of working in terms of B and B_0 rather than B^* will be seen in Examples 1 and 2(c) of Section 5.

It will sometimes be useful in the developments which follow to work not with the space Ξ_m of probability $(m + 1)$-vectors with positive components (to be thought of as being on X_m^*), but rather with the set $\Xi_m^* = \{\eta : \eta = \lambda\xi$ for some ξ in Ξ_m and some real $\lambda \neq 0\}$, which can be regarded as the union of two congruent convex cones in E^{m+1}.

We shall apply Theorems 2.2–2.3 to the problem of determining a-optimum designs in the setting of the first paragraph of this section by using the following simple reduction of [7] Kiefer (1962), p. 795: For fixed a and i_0 with $a_{i_0} \neq 0$, write

$$\begin{aligned} \varphi_{i_0} &\equiv \varphi_{i_0,a} = \textstyle\sum_0^m a_i\theta_i/a_{i_0}, & \\ \varphi_i &= \theta_i, & i \neq i_0, \\ g_{i_0} &= f_{i_0}, & \\ g_i &\equiv g_{i,a} = f_i - (a_i/a_{i_0})f_{i_0}, & i \neq i_0. \end{aligned} \tag{2.5}$$

Then $\sum_0^m \varphi_i g_i = \sum_0^m \theta_i f_i$, and the problem of estimating $a\theta'$ when the regression is $\theta f(x)'$ for $x \varepsilon [-1, 1]$ is the same as that of estimating φ_{i_0} when the regression is $\varphi g(x)'$ for $x \varepsilon [-1, 1]$. Thus, we can apply Theorems 2.2–2.3 to the problem of finding a-optimum designs by setting

$$\begin{aligned} \psi_m &= \varphi_{i_0}, & \{\psi_i, 0 \leqq i < m\} &= \{\varphi_i, i \neq i_0\}, \\ h_m &= g_{i_0}, & \{h_i, 0 \leqq i < m\} &= \{g_i, i \neq i_0\}. \end{aligned} \tag{2.6}$$

In particular, the payoff function (2.1) becomes

$$K(\xi, q) = \int_{-1}^1 [f_{i_0}(x) - \textstyle\sum_{i \neq i_0} q_i(f_i(x) - (a_i/a_{i_0})f_{i_0}(x))]^2 \xi(dx), \tag{2.7}$$

the orthogonality relations (2.4) become

$$\begin{aligned} \int_{-1}^1 [a_{i_0}f_{i_0}(x) - \textstyle\sum_{j \neq i_0} q_j^*(a_{i_0}f_j(x) - a_j f_{i_0}(x))] \\ \cdot [a_{i_0}f_i(x) - a_i f_{i_0}(x)]\xi^*(dx) = 0, \quad i \neq i_0, \end{aligned} \tag{2.8}$$

and the related Chebyshev problem is that of approximating f_{i_0} on $[-1, 1]$ by $\{f_i - (a_i/a_{i_0})f_{i_0}, i \neq i_0\}$. All of these depend on the a and i_0 under consideration (although the latter dependence will be seen to be irrelevant).

The functions g_i of course satisfy Assumption 0 if the f_i do. However, the same is not true of Assumption 1 (as will be seen in Lemma 8). Thus, although Theorem 2.1 can be applied under Assumption 1 to help characterize a 0-optimum design (for estimating θ_m, where $b = 0$), and to find X_m^*, we cannot apply Theorem 2.1 in the same way to find other a-optimum designs.

We now define the sets whose study will be the chief concern of this paper. A parallel notation will be used throughout: A starred symbol D^* (say) will always be defined as a subset of B^* which is invariant under multiplication by a nonzero scalar, and we will then always write

$$D = \Gamma(D^*) \cap B, \qquad D_0 = \Gamma(D^*) \cap B_0; \tag{2.9}$$

the starred sets D^* will be our main objects of interest in Sections 3 and 4, but it will be useful to consider the unstarred sets in the examples of Section 5. Under Assumptions 1 and 2, as we have mentioned, there is, corresponding to $\{f_0, f_1, \cdots, f_m\}$, the set of $m + 1$ Chebyshev points $\{x_0^*, x_1^*, \cdots, x_m^*\} = X_m^*$. We define, as used in Section 1,

$$T^* = \{a : a \varepsilon B^* \text{ and there is an } a\text{-optimum design supported by the entire set } X_m^*\}. \tag{2.10}$$

One could instead study the set $\bar{T}^*$ (say) where there is an a-optimum design supported by a subset of X_m^*, but the set $\bar{T}^* - T^*$ is less susceptible to study by our methods. Let U be the subset of E^m defined by

$$(2.11)\qquad U = \{(x_0, x_1, \cdots, x_{m-1}) \colon \text{ all } |x_i| \leqq 1 \text{ and all } x_i \text{ different}\},$$

and write, for $\tilde{x} = (x_0, \cdots, x_{m-1})$ in E^m and a in B^*,

$$(2.12)\qquad P^*(\tilde{x}, a) = \det \begin{pmatrix} f_0(x_0) & \cdots & f_0(x_{m-1}) & a_0 \\ \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ f_m(x_0) & \cdots & f_m(x_{m-1}) & a_m \end{pmatrix}.$$

We define

$$(2.13)\qquad A^* = \{a \colon a \,\varepsilon\, B^* \text{ and } P^*(\tilde{x}, a) \neq 0 \text{ for all } \tilde{x} \text{ in } U\}.$$

We also define

$$(2.14)\qquad N^* = \{a \colon a \,\varepsilon\, B^* \text{ and } a\theta' \text{ is only estimatable for designs supported by at least } m + 1 \text{ points of } [-1, 1]\}.$$

It is easy to see (Lemma 1 below) that $N^* = A^*$. The usefulness of this set is of course that, for a in A^*, any a-optimum design has at least $m + 1$ points of support, so that X_m^* is at least a possible candidate for this support.

For fixed a in B^*, suppose $a_{i_0} \neq 0$, and consider the system of m linear equations in the $m + 1$ unknowns η_j,

$$(2.15)\qquad \sum_{j=0}^m (-1)^j \eta_j [a_{i_0} f_i(x_j^*) - a_i f_{i_0}(x_j^*)] = 0, \qquad 0 \leqq i \leqq m, \quad i \neq i_0.$$

We shall also consider the related system

$$(2.16)\qquad \begin{aligned} &\sum_{j=0}^m (-1)^j \xi_j [a_{i_0} f_i(x_j^*) - a_i f_{i_0}(x_j^*)] = 0, \qquad i \neq i_0; \\ &\sum_{j=0}^m \xi_j = 1. \end{aligned}$$

For $i_0 = m$, putting $b = \Gamma(a)$, (2.16) becomes

$$(2.17)\qquad \begin{aligned} &\sum_{j=0}^m (-1)^j \xi_j [f_i(x_j^*) - b_i f_m(x_j^*)] = 0, \qquad 0 \leqq i < m; \\ &\sum_{j=0}^m \xi_j = 1. \end{aligned}$$

(It will be clear from the derivation of (2.20) and (2.27) that for fixed a, an η in Ξ_m^* is a solution of (2.15) or (2.23) for some i_0 for which $a_{i_0} \neq 0$, if and only if it is a solution for every such i_0.) The form (2.15) will be of chief concern in Sections 3 and 4. We shall use (2.17) extensively in Example 2(b) of Section 5. We now define

$$(2.18)\qquad \begin{aligned} R^* &= \{a \colon a \,\varepsilon\, B^* \text{ and for some } i_0 \text{ with } a_{i_0} \neq 0 \text{ (2.15) has a solution } \eta \text{ in } \Xi_m^*\} \\ &= \{a \colon a \,\varepsilon\, B^* \text{ and for some } i_0 \text{ with } a_{i_0} \neq 0 \text{ (2.16) has a solution } \xi \text{ in } \Xi_m\}. \end{aligned}$$

Write

$$F_{R^*} = \{(-1)^j f_i(x_j^*), \qquad 0 \leqq i, \ j \leqq m\}, \tag{2.19}$$
$$F_{S^*} = \{f_i(x_j^*), \qquad 0 \leqq i, \ j \leqq m\}.$$

By Lemma 2, Assumption 2 implies that these matrices are nonsingular, a fact which we shall use repeatedly. Suppose $a = \eta F'_{R^*}$ for some η in Ξ_m^*. Since F_{R^*} is nonsingular by Assumption 2, $a_{i_0} \neq 0$ for some i_0, and a clearly satisfies (2.15), since the latter can be written as $a_{i_0}(F_{R^*}\eta')_i = a_i(F_{R^*}\eta')_{i_0}$. Also, this form of (2.15) shows that every a in R^* can be obtained in this way. Thus,

$$\begin{aligned} R^* &= \{a : a = \eta F'_{R^*} \quad \text{for some} \quad \eta \quad \text{in} \quad \Xi_m^*\} \\ &= (\Xi_m^*)F'_{R^*}. \end{aligned} \tag{2.20}$$

We thus have an explicit representation of R^* as a pair of congruent open convex cones obtained from the linear mapping F_{R^*} acting on Ξ_m^*. One of these cones is spanned by the $m + 1$ half-lines consisting of the positive multiples of column vectors of F_{R^*}; the other, of the negative multiples.

From (2.18) we also have

$$\begin{aligned} R &= \{b : b \in B \quad \text{and (2.17) has a solution in} \quad \Xi_m\} \\ &= \{b : b_i = \textstyle\sum_{j=0}^m (-1)^j \xi_j f_i(x_j^*) / \sum_{j=0}^m (-1)^j \xi_j f_m(x_j^*), \\ &\qquad 0 \leqq i < m, \quad \text{for some} \quad \xi \quad \text{in} \quad \Xi_m\}. \end{aligned} \tag{2.21}$$

The importance of the set R^* is given in Theorem 1. For a in R^*, it will turn out that the residual of the best Chebyshev approximation of f_{i_0} by $\{f_i - (a_i/a_{i_0})f_{i_0}, i \neq i_0\}$, mentioned below (2.8), is *oscillatory* (that is, satisfies the condition of Theorem 2.1) even though (Lemma 8) $\{f_i - (a_i/a_{i_0})f_{i_0}, i \neq i_0\}$ is not a Chebyshev system for $a \in R^* - A^*$. It will also turn out that this residual attains its maximum in absolute value at the x_j^*. Hence, the residual at x_j^* (first factor of the integrand of (2.8)) is a multiple of $(-1)^j$ and thus, writing

$$\xi_j = \xi^*(x_j^*), \tag{2.22}$$

the orthogonality relations (2.8) will be seen to reduce to (2.16). The application of the game theory of Theorem 2.3 will be used, in the proof of Theorem 1, to show that $R^* \subset T^*$.

We have mentioned in the previous paragraph that $a \in R^*$ implies the oscillatory nature of a certain Chebyshev approximation problem. One could also study the designs in $T^* - R^*$, which are also supported by X_m^*, but for which (by Lemma 5 and Theorem 1) the solution to this Chebyshev approximation problem has *constant* nonzero residual. Paralleling the development indicated in the previous paragraph, we now consider, in place of (2.15), (2.16), (2.17), the systems

$$\textstyle\sum_{j=0}^m \eta_j[a_{i_0}f_i(x_j^*) - a_i f_{i_0}(x_j^*)] = 0, \qquad 0 \leqq i \leqq m, \quad i \neq i_0; \tag{2.23}$$

$$(2.24)\qquad \sum_{j=0}^{m} \xi_j[a_{i_0} f_i(x_j^*) - a_i f_{i_0}(x_j^*)] = 0, \qquad i \neq i_0,$$
$$\sum_{j=0}^{m} \xi_j = 1;$$

$$(2.25)\qquad \sum_{j=0}^{m} \xi_j[f_i(x_j^*) - b_i f_m(x_j^*)] = 0, \qquad 0 \leqq i < m,$$
$$\sum_{j=0}^{m} \xi_j = 1.$$

In place of (2.18) we now define

$$(2.26)\qquad \begin{aligned} S^* &= \{a\colon a \varepsilon B^* \text{ and for some } i_0 \text{ with } a_{i_0} \neq 0 \text{ (2.23) has a solution } \eta \text{ in } \Xi_m^*\} \\ &= \{a\colon a \varepsilon B^* \text{ and for some } i_0 \text{ with } a_{i_0} \neq 0 \text{ (2.24) has a solution } \xi \text{ in } \Xi_m\}. \end{aligned}$$

Using the second half of (2.19) we obtain in place of (2.20),

$$(2.27)\qquad \begin{aligned} S^* &= \{a\colon a = \eta F'_{S^*} \text{ for some } \eta \text{ in } \Xi_m^*\} \\ &= (\Xi_m^*)F'_{S^*}. \end{aligned}$$

In place of (2.21) we now have

$$(2.28)\qquad \begin{aligned} S &= \{b\colon b \varepsilon B \text{ and (2.25) has a solution in } \Xi_m\} \\ &= \{b\colon b_i = \textstyle\sum_{j=0}^{m} \xi_j f_i(x_j^*)/\sum_{j=0}^{m} \xi_j f_m(x_j^*), \quad 0 \leqq i < m, \\ &\qquad \text{for some } \xi \text{ in } \Xi_m\}. \end{aligned}$$

Using the results indicated at the outset of the present paragraph, we shall show in Theorem 2 that $T^* - R^* = S^*$. A major difference between R^* and S^*, which will be illustrated in Example 2(b) of Section 5, is that the a-optimum design is unique for $a \varepsilon R^*$, while for $a \varepsilon S^*$ we can have other a-optimum designs whose supports are not X_m^*.

While the set $\bar{T}^*$ defined just below (2.10) will be illustrated in Example 2(b), we shall not analyze $\bar{T}^*$ in general. Such an analysis would be more complicated than that of T^* because of the variety of forms the residual can now have and the necessity of determining when the orthogonality relations parallel to (2.15) or (2.23) do indeed correspond to an a for which the residual attains its maximum absolute value on the relevant subset of X_m^*. In particular, for $m > 1$ the set $\bar{T}^*$ is not generally the closure of T^*.

The closure of R^* or S^* (and, hence, of T^*) in $E^{m+1} - \{0\}$ is obviously obtained by replacing Ξ_m^* in (2.20) or (2.27) by the closure of Ξ_m^* in $E^{m+1} - \{0\}$; that is, by the set of all $(m + 1)$-vectors not all of whose components are zero, but whose nonzero components all have the same sign.

The definition of the generalization of the set considered by Hoel and Levine is

$$(2.29)\qquad H^* = \{a\colon a \varepsilon B^* \text{ and } a = \lambda f(e)' \text{ for some real } \lambda \neq 0 \text{ and some real } e \text{ with } |e| > 1\}.$$

(If $f(e) = 0$, $\theta f(e)'$ can of course be estimated without error; such e are excluded from H^*, and under Assumption 3 they obviously can not exist.) In particular, if $f_m(e) \neq 0$ for $|e| > 1$, H_0 is empty and

$$(2.30) \quad \Gamma(H^*) = H = \{b: b_i = f_i(e)/f_m(e), \quad 0 \leqq i < m, \quad \text{for some } e \text{ with } |e| > 1\}.$$

The examples of Section 5 illustrate the sets defined in this section.

We add the definition of a concept which arises in the first paragraph of Section 4 and in Example 2(b) of Section 5, that of an *admissible design* ξ, which is a design such that for no ξ' is $M(\xi') - M(\xi)$ nonnegative definite and not the zero matrix. The meaning of this concept is discussed in [6] Kiefer (1959).

In Example 2(a) we shall introduce some additional material from the literature, which is used only there.

3. Auxiliary lemmas. The lemmas of this section will be used in proving our main results in the next section.

LEMMA 1. $N^* = A^*$.

PROOF. Since $\theta f(x_j)'$ is the expected value of an observation at x_j, any linear parametric function which is estimatable under a design supported at $\{x_0, x_1, \cdots, x_{m-1}\}$ or a subset thereof must be of the form $\sum_0^{m-1} \gamma_j \theta f(x_j)'$ for some real $\gamma_0, \cdots, \gamma_{m-1}$. Hence, $\theta a'$ is estimatable under a design on $\{x_0, x_1, \cdots, x_{m-1}\}$ if and only if there exist $\gamma_0, \cdots, \gamma_{m-1}$ such that $\sum_0^{m-1} \gamma_j \theta f(x_j)' = \theta a'$ for all θ; that is, such that $\sum_0^{m-1} \gamma_j f(x_j) = a$. This last is equivalent to $P^*(\bar{x}, a) = 0$. This completes the proof.

LEMMA 2. *Under Assumption 2, $f_0, f_1, \cdots, f_m$ form a Chebyshev system on $[-1, 1]$ and there are numbers d_j such that $\sum_0^m d_j f_j(x) \equiv 1$ for x in $[-1, 1]$.*

PROOF. If $f_0, f_1, \cdots, f_m$ are not a Chebyshev system, there is a vector q other than the zero vector such that $q'f$ has at least $m + 1$ zeros on $[-1, 1]$. Since $q'f$ clearly has at least one change of direction at some point strictly between any two successive zeros, it follows that $q'f$ has at least m changes of direction.

To prove the second assertion, write $h(x) = f(x) - f(0)$. Since $ch(0)' = 0$ for all c, Assumption 2 implies that, for each c, ch' either has fewer than m changes of direction or else is identically zero on $[-1, 1]$. If the latter holds for some c which is not the zero vector, we must have $cf(0)' \neq 0$ (since otherwise $cf(x)' \equiv 0$, contradicting Assumption 0), and $d_j = c_j/cf(0)'$ then yields the desired result. If $ch(x)'$ is not identically zero for all nonzero c, the h_i are linearly independent so that, by continuity, there are $m + 1$ points x_j $(0 \leqq j \leqq m)$ in $[-1, 1]$, none of them zero, such that $H = \{h_i(x_j), 0 \leqq i, j \leqq m\}$ is nonsingular. The $m + 1$ linear equations $cH = (0, \cdots, 0, 1)$ then have a solution $c = \bar{c}$ (say), and $\bar{c}h'$ then vanishes at the $m + 1$ points $0, x_0, x_1, \cdots, x_{m-1}$ and thus has at least m changes of direction, which is a contradiction.

LEMMA 3. *Under Assumptions 1 and 2 the residual $f_m - \sum_0^{m-1} c_j^* f_j$ of the unique best Chebyshev approximation to f_m on $[-1, 1]$ of the form $\sum_0^{m-1} c_j f_j$, attains its maximum in absolute value at exactly $m + 1$ points $-1 = x_0^* < x_1^* \cdots < x_m^* = 1$, the residual alternating in sign at successive x_j^*.*

PROOF. The alternating nature of the residual on $m + 1$ points follows from Assumption 1 and Theorem 2.1. It then follows from Assumption 2 that the residual cannot take on its maximum in absolute value at more than $m + 1$ points, and that $x_0^* = -1$, $x_m^* = 1$.

The reader is reminded of the definition which follows Theorem 2.1, according to which $\{x_0^*, x_1^*, \cdots, x_m^*\}$ will be called the *Chebyshev points of* $\{f_0, f_1, \cdots, f_m\}$.

LEMMA 4. *Assumption* 1 *implies that* $0 \varepsilon A$.

PROOF. The proof of Lemma 1, with $a = (0, 0, \cdots, 0, 1) = a^*$ (say), shows that θ_m is estimatable for a design on $\{x_0, x_1, \cdots, x_{m-1}\}$ if and only if $0 = P^*(\tilde{x}, a^*) = \det \{f_i(x_j), 0 \leqq i, j < m\}$. The latter is not zero if the x_i are distinct, by Assumption 1.

LEMMA 5. *Under Assumptions* 1 *and* 2, *if an a-optimum design is supported by at least* $m + 1$ *points, then either* $a \varepsilon R^*$ *or else, for each* i_0 *for which* $a_{i_0} \neq 0$, *every best Chebyshev approximation of* f_{i_0} *by* $\{f_i - (a_i/a_{i_0})f_{i_0}, i \neq i_0\}$ *on* $[-1, 1]$ *has constant nonzero residual.*

PROOF. Suppose there is a best approximation $\sum_{i \neq i_0} c_i'[f_i - (a_i/a_{i_0})f_{i_0}]$ such that the residual $r(x) = f_{i_0}(x) - \sum_{i \neq i_0} c_i'[f_i(x) - (a_i/a_{i_0})f_{i_0}(x)]$ is not constant. By Theorem 2.3 and the hypothesis of the lemma, there are $m + 1$ points $x_0 < x_1 < \cdots < x_m$ in the support of ξ at which $|r(x)|$ attains its maximum on $[-1, 1]$. It follows easily from Assumption 2 that, if $r(x)$ is not constant, $r(x)$ alternates in sign at $x_0, x_1, \cdots, x_m$ and thus has m zeros on $[-1, 1]$. In that case the coefficient of f_m in r is not zero, since, if it were, r would be a linear combination of $f_0, \cdots, f_{m-1}$ which has m zeros but which is not identically zero, contradicting Assumption 1.

Writing $r(x) = q[f_m(x) - \sum_0^{m-1} h_i f_i(x)]$, it follows from the oscillation property of $q^{-1}r(x)$ at $x_0, x_1, \cdots, x_m$ and Theorem 2.1 that $\sum_0^{m-1} h_i f_i$ is the best Chebyshev approximation of f_m by $f_0, \cdots, f_{m-1}$. Hence $x_0, x_1, \cdots, x_m$ are the Chebyshev points and by Lemma 3 they in fact constitute the entire support of ξ. Thus, $a \varepsilon R^*$.

Finally, if $r(x) \equiv 0$, then by Theorem 2.2 the variance of the best linear estimator of $\theta a'$ is infinite, contradicting the fact that any design on $m + 1$ points yields an estimator with finite variance.

LEMMA 6. *For fixed a with* $a_{i_0} \neq 0$, *there are real constants* c_i *and* K *such that*

$$f_{i_0}(x) - \sum_{i \neq i_0} c_i[f_i(x) - (a_i/a_{i_0})f_{i_0}(x)] \equiv K \quad \text{for} \quad x \varepsilon [-1, 1], \tag{3.1}$$

if and only if $K \neq 0$ *and there are unique numbers* d_i *such that*

$$\sum_0^m d_i f_i(x) \equiv 1 \quad \text{for} \quad x \varepsilon [-1, 1], \tag{3.2}$$

and

$$K = a_{i_0}[\sum_{i=0}^m a_i d_i]^{-1}, \qquad c_i = -K d_i . \tag{3.3}$$

For fixed a with $a_{i_0} \neq 0$ *and fixed reals* c_i', *there are real constants* c_i *and* K *such that*

$$\begin{aligned} f_{i_0}(x) - \sum_{i \neq i_0} c_i[f_i(x) - (a_i/a_{i_0})f_{i_0}(x)] \\ \equiv K \sum_{i=0}^m c_i' f_i(x) \quad \text{for} \quad x \varepsilon [-1, 1] \end{aligned} \tag{3.4}$$

if and only if $\sum_0^m a_i c_i' \neq 0$ *and*

$$K = a_{i_0}[\sum_0^m a_i c_i']^{-1}, \qquad c_i = -Kc_i'. \tag{3.5}$$

PROOF. First suppose (3.1) holds. Since it is impossible for all c_i to be 0 while $1 + \sum_{i \neq i_0} a_i c_i / a_{i_0}$ is also 0, the left side of (3.1) cannot be identically 0, by the linear independence of the f_i. Hence $K \neq 0$. The existence of numbers d_i satisfying (3.2) now follows, and their uniqueness is a consequence of the linear independence of the f_i. Substituting $K \sum_0^m d_i f_i(x)$ for K in (3.1), for each i $(0 \leqq i \leqq m)$ the coefficients of f_i on both sides must be the same, again by linear independence. This yields (3.3). The converse is obvious.

Finally, assuming (3.4), equality of the coefficients of f_i on both sides yields (3.5). The converse is again clear.

LEMMA 7. *Under Assumptions* 1 *and* 2, *suppose that* $\sum_0^{m-1} c_j^* f_j$ *is the best Chebyshev approximation of* f_m *by* $\{f_0, f_1, \cdots, f_{m-1}\}$ *and write* $c_m^* = -1$; *furthermore, let* $d_0, \cdots, d_m$ *be the numbers whose existence is guaranteed by Lemma* 2. *Then* $a \,\varepsilon\, R^*$ *implies* $\sum_0^m a_i c_i^* \neq 0$, *and* $a \,\varepsilon\, S^*$ *implies* $\sum_0^m a_i d_i \neq 0$.

PROOF. Suppose $a \,\varepsilon\, R^*$ but that $\sum_0^m a_i c_i^* = 0$. Multiply the ith orthogonality relation (2.15) by c_i^* and sum over $i \neq i_0$. We obtain

$$\sum_{j=0}^m (-1)^j \eta_j [\sum_{i=0}^m c_i^* f_i(x_j^*)] = 0. \tag{3.6}$$

Since the term in square brackets in (3.6) is some nonzero constant times $(-1)^j$, this leads to a contradiction. Similarly, if $a \,\varepsilon\, S^*$ but $\sum_0^m a_i d_i = 0$, multiplying the ith relation of (2.23) by d_i and summing over $i \neq i_0$ yields

$$\sum_{j=0}^m \eta_j [\sum_{i=0}^m d_i f_i(x_j^*)] = 0, \tag{3.7}$$

which yields a contradiction since $\sum_0^m d_i f_i \equiv 1$.

LEMMA 8. *Suppose* $a_{i_0} \neq 0$. *Then* $\{a_{i_0} f_i - a_i f_{i_0}, i \neq i_0\}$ *is a Chebyshev system on* $[-1, 1]$ *if and only if* $a \,\varepsilon\, A^*$.

PROOF. For $0 \leqq i \leqq m$ and $i \neq i_0$, subtract a_i / a_{i_0} times the i_0th row of the matrix of (2.12) from the ith row. We obtain

$$P^*(\tilde{x}, a) = \pm a_{i_0} \det \{f_i(x_j) - (a_i/a_{i_0}) f_{i_0}(x_j), i \neq i_0, 0 \leqq j < m\},$$

which at once yields the conclusion.

LEMMA 9. *Under Assumption* 3, $H^* \subset A^*$.

The proof is immediate.

REMARK 2. Lemma 9 really uses something weaker than Assumption 3, namely, the nonvanishing of det $\{f_i(x_j), 0 \leqq i, j \leqq m\}$ when m different x_i's are in $[-1, 1]$ and one x_i is outside $[-1, 1]$.

4. Principal results. Our first result is that $a \,\varepsilon\, R^*$ implies that the unique a-optimum design is supported by the Chebyshev points, and that R is m-dimensional (and hence R^* is $(m + 1)$-dimensional, which we already knew from (2.20)). This last is perhaps surprising in view of the fact that, in such a simple example as that of polynomial regression, the designs on the Chebyshev points are only an m-parameter family out of the $(2m - 1)$-parameter family of designs on $m + 1$ points including ± 1, all of which are admissible (see [6]), and the Hoel-Levine set H is one-dimensional. We shall see why this is possible in Example 2(b), where it will be seen that infinitely many different supporting sets may yield a-optimum designs for the same a.

THEOREM 1. *Under Assumptions* 1 *and* 2, *if* $a \in R^*$ *the orthogonality relations* (2.16) *have a unique solution* $(\xi_0, \cdots, \xi_m)$, *and the corresponding design* (2.22) (*which is supported by the Chebyshev points*) *is the unique a-optimum design.* (*Thus,* $R^* \subset T^*$.) *Furthermore, R contains a neighborhood of the origin in* E^m (*that is,* R^* *contains a neighborhood of* $(0, 0, \cdots, 0, 1)$ *in* E^{m+1}).

PROOF. Suppose $a \in R^*$. Let $c_0^*, \cdots, c_m^*$ be as in Lemma 7. By Lemma 7, $\sum_0^m a_i c_i^* \neq 0$. Hence, by Lemma 6 with $c_i' = -c_i^*$ and i_0 such that $a_{i_0} \neq 0$, there are values $c_i = \bar{c}_i$ (say) and $K \neq 0$ (by (3.5)) such that, on $[-1, 1]$,

$$(4.1) \qquad f_{i_0} - \textstyle\sum_{i \neq i_0} \bar{c}_i[f_i - (a_i/a_{i_0})f_{i_0}] \equiv K[f_m - \sum_0^{m-1} c_i^* f_i].$$

By Lemma 3 the right member of (4.1) attains its maximum in absolute value on $[-1, 1]$ at, and only at, the Chebyshev points $x_0^*, \cdots, x_m^*$, and this is therefore true of the left member. Hence, ξ^* is maximal with respect to $\bar{c} = \{\bar{c}_i, i \neq i_0\}$ for the game with payoff (2.7) if and only if the support of ξ^* is a subset of X_m^*. On the other hand, since $f_{i_0}(x_j^*) - \sum_{i \neq i_0} \bar{c}_i[f_i(x_j^*) - (a_i/a_{i_0})f_{i_0}(x_j^*)]$ is some nonzero constant times $(-1)^j$, by (2.22) the orthogonality relations (2.8) reduce to (2.16), and thus $\bar{c}$ is minimal relative to any nonnegative solution to (2.16). One such strictly positive solution is of course guaranteed by the definition of R^*. Since ξ^* is maximal with respect to $\{\bar{c}_i, i \neq i_0\}$, and the latter is minimal with respect to ξ^*, standard game theory results, mentioned in Theorem 2.3, assert that ξ^* is maximin for the game with payoff (2.7). Hence ξ^* is a-optimum. Moreover, these same results assert that the left member of (4.1) is the residual of the Chebyshev approximation of f_{i_0} on $[-1, 1]$ by a linear combination of the functions $\{(f_i - (a_i/a_{i_0})f_{i_0}), i \neq i_0\}$, a fact of which we shall make use in a moment.

We now turn to the proof of uniqueness. According to Theorem 2.3, every a-optimum design ξ is maximal relative to the $\bar{c}$ of the previous paragraph, and hence, as deduced there, is supported by some subset of X_m^* and (by Theorem 2.3) satisfies (2.16) (with the interpretation (2.22)), with all ξ_i nonnegative (perhaps some zero). If there were more than one such solution, the $(m + 1) \times (m + 1)$ matrix L_a whose (i, j)th element is

$$(4.2) \qquad \begin{aligned} (L_a)_{i,j} &= f_i(x_j^*) - (a_i/a_{i_0})f_{i_0}(x_j^*), \quad 0 \leq i \leq m, \quad i \neq i_0, \quad 0 \leq j \leq m, \\ &= (-1)^j, \qquad\qquad\qquad\qquad\qquad\qquad i = i_0, \quad 0 \leq j \leq m, \end{aligned}$$

would be singular, since (2.16) can be written as $L_a(\xi_0, -\xi_1, \xi_2, \cdots, (-1)^m\xi_m)' = (0, \cdots, 0, 1, 0, \cdots, 0)'$, with the 1 in i_0th place in the last vector. Now $(-1)^j$ can be written as $K_0 \sum_0^m c_i^* f_i(x_j^*)$ for some nonzero constant K_0. Hence (4.2) yields

$$(4.3) \qquad L_a = \begin{pmatrix} I_{i_0} & \rho_1' & 0 \\ & K_0 c^* & \\ 0 & \rho_2' & I_{m-i_0} \end{pmatrix} F_{S^*}$$

where I_r is the $r \times r$ identity matrix, $c^* = (c_0^*, c_1^*, \cdots, c_m^*)$, and $\rho_1 = (-a_0/a_{i_0}, \cdots, -a_{i_0-1}/a_{i_0})$, $\rho_2 = (-a_{i_0+1}/a_{i_0}, \cdots, -a_m/a_{i_0})$. The second factor on the right side of (4.3) is nonsingular by Lemma 2. The determinant of the first factor, which can be computed by adding $-K_0c_i^*$ times the ith row to the i_0th row for each $i \neq i_0$, is $K_0a_{i_0}^{-1} \sum_0^m a_ic_i^*$, which is nonzero by Lemma 7. Hence, L_a is nonsingular and there is a unique a-optimum design.

It remains to show that R contains an open neighborhood of the origin. When $b = 0$, there is an optimum design ξ^* on the Chebyshev points by Lemma 3 and Theorem 2.3 with $h_i = f_i$ and $\psi_i = \theta_i$, and all ξ_i^* are then positive, since otherwise θ_m would be estimatable on fewer than $m + 1$ points, in violation of Lemmas 1 and 4. This shows that $0 \varepsilon R$. Moreover, if $a = (b, 1)$ in the first factor on the right side of (4.3) is varied by varying b in a small enough neighborhood of 0, $L_{(b,1)}$ remains nonsingular (since $\sum_0^{m-1} b_jc_j^* \neq -c_m^*$ for b near 0) and the coordinates ξ_j of the solution to (2.16), which will vary continuously with b, will remain positive as they were when $b = 0$. (Alternatively, this last sentence may be replaced by (2.20).) This completes the proof of Theorem 1.

THEOREM 2. *Under Assumptions* 1 *and* 2, $T^* - R^* = S^*$; *and, if* $a \varepsilon S^*$, *the orthogonality relations* (2.24) *have a unique solution which corresponds to the design* (2.22) *on the entire set* X_m^*. *There is no other a-optimum design supported by* X_m^* *or a subset thereof.* (*There may be other a-optimum designs.*)

PROOF. The proof parallels that of Theorem 1, so we merely outline the differences. Suppose $a \varepsilon S^*$. By Lemma 7, $\sum_0^m a_i d_i \neq 0$. By Lemmas 2 and 6 with $a_{i_0} \neq 0$, there are constants c_i^0 and $K \neq 0$ such that, on $[-1, 1]$,

$$f_{i_0}(x) - \sum_{i \neq i_0} c_i^0[f_i(x) - (a_i/a_{i_0})f_{i_0}(x)] \equiv K. \tag{4.4}$$

Hence, every ξ^* is maximal with respect to $c^0 = \{c_i^0, i \neq i_0\}$ for the game with payoff (2.7). By (2.22) and (4.4), the orthogonality relations (2.8) become (2.24). Therefore c^0 is minimal relative to any nonnegative solution of (2.24), while, as we have already seen, the latter is maximal relative to c^0. Hence, by the standard game theory results cited in the proof of Theorem 1, any nonnegative solution of (2.24) is a-optimum. One strictly positive solution of (2.24) is guaranteed by the definition of S^*, and this is surely a-optimum.

If there were two a-optimum designs with subsets of X_m^* for support, there would be more than one solution to (2.24), which can be written as $M_a(\xi_0, \xi_1, \cdots, \xi_m)' = (0, \cdots, 0, 1, 0, \cdots, 0)$, where M_a is obtained from the L_a of (4.2) by replacing $(-1)^j$ by 1 in the i_0th row. Since $\sum_0^m d_if_i(x) \equiv 1$, the equation for M_a corresponding to (4.3) is obtained by replacing the i_0th row of the first factor on the right side by $(d_0, d_1, \cdots, d_m)$. Adding $-d_i$ times the ith row of this factor to the i_0th row for $i \neq i_0$, we obtain $a_{i_0}^{-1}\sum_0^m a_i d_i$ for the determinant of this factor, which is thus nonzero by Lemma 7. Hence there is only one a-optimum design supported by a subset of X_m^*.

By Lemma 5, $T^* = R^* \cup S^*$, so that it remains to show that R^* and S^* are disjoint. If, to the contrary, there were an a in $R^* \cap S^*$, then for $a_{i_0} \neq 0$ there would, by our previous development, be two different Chebyshev approximations to f_{i_0} by $\{f_i - (a_i/a_{i_0})f_{i_0}, i \neq i_0\}$, one with constant residual and one with oscillatory residual. The Chebyshev vectors are thus at least one-dimensional. Applying Theorem 2.3 with $p \geqq 1$, we conclude that there is an a-optimum design supported by m or fewer points. Since $a \varepsilon R^*$, this contradicts the conclusion of Theorem 1. The proof of Theorem 2 is now complete.

REMARK 3. Example 2(b) of Section 5 will illustrate the lack of uniqueness of a-optimum designs for $a \varepsilon S^*$, as well as the fact mentioned in Section 2 that the set $\bar{T}^*$, defined just below (2.10) and discussed above (2.29), has a more complicated structure than T^* (in particular, that $\bar{T}^*$ is not merely the closure of T^*).

THEOREM 3. *Under Assumptions* 1 *and* 2, $A^* \subset R^*$.

PROOF. Suppose $a \varepsilon A^*$. Let i_0 be such that $a_{i_0} \neq 0$. By Lemma 8, $\{(a_{i_0}f_i - a_if_{i_0}), i \neq i_0\}$ is Chebyshev, and hence by Theorem 2.1 the best Chebyshev approximation of $a_{i_0}f_{i_0}$ by $\{(a_{i_0}f_i - a_if_{i_0}), i \neq i_0\}$ has an oscillatory residual. Since $a \varepsilon A^*$, any a-optimum design is, by Lemma 1, supported by at least $m + 1$ points. Lemma 5 now yields $a \varepsilon R^*$.

The next (and last) two theorems of this section are direct generalizations of the Hoel-Levine results discussed in Section 1, since their example of polynomial regression satisfies the assumptions of these theorems.

THEOREM 4. *Under Assumptions* 2 *and* 3, $H^* \subset R^*$. *If also* $f_m(e) \neq 0$ *for* $|e| > 1$, *then* $\Gamma(H^*) = H \subset R$.

PROOF. By Proposition 2.2 of Remark 1(b), Assumptions 2 and 3 imply that Assumptions 1 and 2 are satisfied for $\{\bar{f}_i, 0 \leqq i \leqq m\}$ where $\bar{f}' = Qf'$, for some nonsingular Q. Let A_1^* and R_1^* be the sets defined by (2.13) and (2.18) if f is replaced by $\bar{f}$ and θ is replaced by $\bar{\theta} = \theta Q^{-1}$ (so that $\theta f' = \bar{\theta}\bar{f}'$). Then $a\bar{\theta}' = aQ'^{-1}\theta'$, so that the vector a in A_1^* must be multiplied by Q'^{-1} to give the corresponding vector in A^*; that is, $A^* = A_1^*Q'^{-1}$, and similarly $R^* = R_1^*Q'^{-1}$. Since $A_1^* \subset R_1^*$ by Theorem 3, we thus obtain $A^* \subset R^*$. Lemma 9 now completes the proof that $H^* \subset R^*$. The remainder of the theorem follows from (2.30).

REMARK 4. Assumptions 2 and 3 may be replaced in Theorem 4 by Assumptions 1 and 2 and the assumption indicated in Remark 2.

A consequence of our conclusion that $H^* \subset R^* \subset T^*$ under Assumptions 1–3 is the result of Hoel and Levine [5] mentioned in Section 1, that $H^* \subset T^*$ if $f_i(x) \equiv x^i$.

The last theorem of this section concerns $V(f(y), \xi) = f(y)M^{-1}(\xi)f(y)'$ which, we recall, is proportional to the variance of the best linear estimator of the regression function $\theta f(y)'$ at the point y of Y when the design ξ on $[-1, 1]$ is used.

THEOREM 5. *Under Assumptions* 1, 2, 4, *and* 5, *if k is a real function of e*, $1 < e < \infty$, *such that always* $k(e) \leqq e$ *and* $\lim\inf_{e\to+\infty} k(e) > -\infty$, *then for e (resp., $-e$) sufficiently large, the unique $f(e)$-optimum design* $\xi^{(e)}$ *minimizes* $\max_{k(e)\leqq y\leqq e} V(f(y), e)$ *(resp.,* $\max_{e\leqq y\leqq -k(-e)} V(f(y), e)$*).*

PROOF. We shall prove only the first conclusion (as $e \to +\infty$), the case $e \to -\infty$ being treated similarly.

By Assumption 4, $f(e) = f_m(e)(o(1), \cdots, o(1), 1)$ as $e \to +\infty$, so that for e

sufficiently large the second part of Theorem 1 shows that the $f(e)$-optimum design $\xi^{(e)}$ is unique and is supported by X_m^*. Moreover, the proof of Theorem 1 shows that

$$\lim_{e\to+\infty} \xi^{(e)}(x_i^*) = \xi^*(x_i^*), \qquad 0 \leqq i \leqq m, \tag{4.5}$$

where ξ^* is the unique optimum design for estimating θ_m.

Since $\xi^{(e)}$ minimizes $V(f(e), \xi)$, the theorem will follow if we prove that, for some real e_0,

$$\max_{k(e)\leqq y\leqq e} V(f(y), \xi^{(e)}) = V(f(e), \xi^{(e)}) \qquad \text{for } e > e_0. \tag{4.6}$$

It was shown in the proof of Theorem 1 that the unique optimum design ξ^* for estimating θ_m has $\xi^*(x_i^*) > 0$ for $0 \leqq i \leqq m$. It follows from Lemma 2 that $M(\xi^*)$ is nonsingular. Hence, from (4.5), $M(\xi^{(e)})$ is nonsingular for sufficiently large, and we can write

$$M^{-1}(\xi^{(e)}) = M^{-1}(\xi^*) + \{o(1)\} \qquad \text{as } e \to +\infty \tag{4.7}$$

where $\{o(1)\}$ is a matrix whose elements approach 0 as $e \to +\infty$. Since f is continuous and $V(f(y), \xi) = f(y)M^{-1}(\xi)f(y)'$, it follows from the nonsingularity of $M(\xi^*)$ and (4.7) that, for every compact set K,

$$\max_{y\epsilon K} |V(f(y), \xi^{(e)}) - V(f(y), \xi^*)| \to 0 \qquad \text{as } e \to +\infty. \tag{4.8}$$

We shall show below that there is an $\epsilon > 0$ and real k_0 and e_1, $k_0 < e_1$, such that $e > e_1$ implies

(4.9)
(a) $V(f(y), \xi^{(e)})$ is strictly increasing in y for $y \geqq e_1$;
(b) $V(f(e_1 + 1), \xi^{(e)}) - V(f(e_1), \xi^{(e)}) > \epsilon$;
(c) $k(e) \geqq k_0$;
(d) $V(f(e_1), \xi^*) = \max_{k_0 \leqq y \leqq e_1} V(f(y), \xi^*)$.

If we let K be the interval $[k_0, e_1]$, we can by (4.8) find an e_2 such that the left side of (4.8) is $<\epsilon/2$ for $e > e_2$. Then (4.9) implies that $V(f(e_1 + 1), \xi^{(e)}) = \max_{k_0 \leqq y \leqq e_1+1} V(f(y), \xi^{(e)})$ if $e > \max(e_1 + 1, e_2) = e_3$ (say); consequently, from (4.9) (a), we obtain (4.6) for $e_0 = e_3$.

We now prove (4.9). The hypothesis of the theorem on k implies (c) if e_1 is sufficiently large. (4.9) (d) follows from the validity of (4.9) (a) with $\xi^{(e)}$ replaced by ξ^*, which will be proved below, and from the fact that

$$\lim_{e\to+\infty} V(f(e), \xi^*) = \infty;$$

the latter follows from Assumption 4, according to which $V(f(e), \xi^*) = f_m^2(e)v_m(\xi^*)(1 + o(1))$ as $e \to +\infty$, where $v_m(\xi^*)$ is the lower right element of $M^{-1}(\xi^*)$, and from the fact (Assumption 5) that $f_m^2(e)$ approaches $+\infty$ with e. Next, we note that

$$\begin{aligned} V(f(y + \Delta), \xi) &- V(f(y), \xi) \\ &= [f(y + \Delta) + f(y)]M^{-1}(\xi)[f(y + \Delta) - f(y)]'. \end{aligned} \tag{4.10}$$

By Assumption 4, $f(y + \Delta) + f(y) = (f_m(y + \Delta) + f_m(y))(o(1), o(1), \cdots, o(1), 1)$ as $y \to +\infty$, with the $o(1)$ terms uniform for positive Δ. Similarly, by Assumption 5, $f(y + \Delta) - f(y) = (f_m(y + \Delta) - f_m(y))(O(1), O(1), \cdots, O(1), 1)$ as $y \to +\infty$, with the $O(1)$ terms uniform for $0 < \Delta \leqq 1$. We also note, from Assumption 5, that

$$f_m^2(y + \Delta) - f_m^2(y) > 0 \tag{4.11}$$

for y sufficiently large and all $\Delta > 0$. From these and (4.7), we have

$$\begin{aligned} V(f(y + \Delta), \xi^{(e)}) &- V(f(y), \xi^{(e)}) \\ &= [f_m^2(y + \Delta) - f_m^2(y)]v_m(\xi^*)(1 + o(1)) \end{aligned} \tag{4.12}$$

as $\min(y, e) \to +\infty$, uniformly for $0 < \Delta \leqq 1$. (4.11) and (4.12) yield (4.9) (a) and (b) for e_1 sufficiently large. (4.9) (a) with $\xi^{(e)}$ replaced by ξ^* is proved in the same way. This completes the proof of Theorem 5.

REMARK 5. *Extensions.* The results of this section can be extended by altering the nature of X and Y. For example, it is well known that much of the Chebyshev approximation theory, in particular Assumption 1 and Theorem 2.1, apply if X is a subset of the 1-sphere (boundary of the unit circle). Without going into further detail, we note that a case of practical importance which can be treated by our methods is that where there are open intervals in $[-1, 1]$ where observations are prohibited for technological reasons; $[-1, 1]$ is then replaced by a union of closed intervals. Similarly, Y can be altered from $(-\infty, \infty) - [-1, 1]$; for example, it may be that it only makes sense to define f on $[-1, \infty)$ because $x + 1$ is inherently nonnegative; for another example, if X is a union of disjoint intervals as mentioned just above, Y might be $(-\infty, \infty) - X$. For the required approximation theory results, see, e.g., [10a], section 2.3.3. These results apply, in particular, to the polynomial csae where X is two intervals, studied independently by [4a] Hoel (1965), some of whose arguments can be simplified by use of this theory.

5. Examples.

EXAMPLE 1. *The case* $m = 1$. For the sake of completeness (and for use in Example 2.2) we first determine the possible Chebyshev systems when $m = 1$, and then, in the next paragraph, show that under the stronger Assumptions 1 and 2 there is essentially only one example. We first show, then, that *if* $\{f_0, f_1\}$ *is a continuous Chebyshev system on* $[-1, 1]$, *then, for some nonsingular* D *and* $g = fD$, *we have* $g_0(x) > 0$ *and* $h(x) = g_1(x)/g_0(x)$ *strictly increasing for all* x. This result is probably known (although we did not succeed in finding a reference), and the proof is quite simple: We first show there is a linear combination $g_0 = fa'$ which is positive throughout $[-1, 1]$. By the Chebyshev assumption, there are linear combinations $G_i = fa^{(i)\prime}$, $i = \pm 1$, with $G_i(i) = 0$, $G_i(-i) = 1$. If $G_0 = G_1 + G_2$, then either (i) G_0 is such a g_0; or else (ii) G_0 has at least two zeros (contradicting the Chebyshev assumption); or else (iii) G_0 has a single zero at q (say) with $-1 < q < 1$, in which case $G_0 - (\operatorname{sgn} G_1(q))G_1/2$ has at least two zeros. With the existence of a g_0 thus established, we need only observe that the Chebyshev nature of $\{1, h\}$ follows from that of $\{f_0, f_1\}$, and that h is hence strictly monotone and can be taken as increasing by making a change of sign if necessary.

If now we also impose Assumptions 1 and 2, since $G_0(\pm 1) = 1$ we see that $G_0(x) \equiv 1$, because otherwise G_0 would be a nonconstant function with at least one change of direction. Hence, in this case we can find D such that $g_0(x) \equiv 1$ on $[-1, 1]$ and $g_1(\pm 1) = \pm 1$, with g_1 strictly increasing on $[-1, 1]$. Write $f\theta' = g\psi'$. Now map X onto another copy Z of $[-1, 1]$ using the mapping g_1. The regression problem on Z with regression $\psi_0 + \psi_1 z$ then corresponds to that on X with regression $\psi_0 + \psi_1 g_1(x)$ in such a manner that if ξ' is a-optimum on Z, then an a-optimum design ξ on X is defined by $\xi(x) = \xi'(g_1(x))$.

For the linear regression problem on Z just described, it is easily verified that $A = \{b_0 : |b_0| < 1\}$ and that A_0 is empty (since ψ_0 can be estimated by an observation at $z = 0$). The Chebyshev points are $x_0^* = -1$, $x_0^* = 1$, so that

$$F_{R^*} = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}.$$

Hence, writing $\alpha = \eta_0 - \eta_1$ and $\sigma = -\eta_0 - \eta_1$, we have from (2.20) that a general element a of R^* has the form (α, σ). Since $\sigma \neq 0$ for $\eta \in \Xi_1^*$, we see that R_0 is empty. Since the range of α/σ for η in Ξ_1^* is the interval $(-1, 1)$, we conclude that $R = A$. The points $b_0 = \pm 1$ of B correspond to optimum designs on one point: $\xi(\pm 1) = 1$. All admissible designs in this problem are well known to be supported on X_1^* or a subset thereof, and from this or (2.27) we see that S^* consists of all points of B^* except R^* and $\Gamma^{-1}(b_0)$ for each of the two additional points $b_0 = \pm 1$ of B. (As Example 2 (b) and (c) shows, no such simple result holds when $m > 1$.) The point $S_0 = T_0$ corresponds to estimating ψ_0, which can be done optimally both by the design ξ' (say) for which $\xi'(1) = \xi'(-1) = \frac{1}{2}$ and also by any of an infinite number of inadmissible designs, the simplest of which is the design ξ'' (say) for which $\xi''(0) = 1$; it is easily verified that $M(\xi') - M(\xi'')$ is nonnegative definite, of rank 1.

The above characterizations also hold for the regression $g\psi'$ on X. The linear transformation which took f into g can then be used to characterize the corresponding sets for the original problem with regression $f\theta'$ on X. For example, as in the proof of Theorem 4, if $g' = Qf'$, then $(A^* \text{ for } g)Q'^{-1} = A^*$ for f. However, R need no longer be connected. For example, if $f_0(x) = 1 + x/2$, $f_1(x) \equiv 1$, so that

$$Q = \begin{pmatrix} 0 & 1 \\ 2 & -2 \end{pmatrix},$$

one obtains $(a_0, a_1) = (\bar{a}_0 + \bar{a}_1/2, \bar{a}_0)$, where (a_0, a_1) refers to f and $(\bar{a}_0, \bar{a}_1)$ refers to g (treated in the previous paragraph). Thus, for f we obtain

$$R = A = (-\infty, \tfrac{1}{2}) \cup (\tfrac{3}{2}, \infty),$$

with $R_0 = A_0 =$ "point at infinity." This unnecessary complication points up the advantage of working in terms of R^* (as described by (2.20)), whose geometric characteristics are unchanged by the linear transformation Q.

As for H, suppose we extend the map of $X \to Z$ to $(-\infty, \infty) \to (-\infty, \infty)$ by the identity map on $(-\infty, \infty) - [-1, 1]$. Write $\psi h(z)'$ for the regression function on $(-\infty, \infty)$ as extended from Z, so that $h_i(z) = z^i$ for $z \in Z$. Under Assumption 3 it is easy to see that the graph of h_1 crosses (and is not merely tangent to) that of h_0 at 1 and at no other point of $(-\infty, \infty)$, and that $h_1(z) = 0$ only at $z = 0$. Under Assumption 4 (for example, if $h_i(z) = z^i$ on $(-\infty, \infty)$) we obtain $H = (-1, 0) \cup (0, 1) = A - \{0\}$ (another result which does not hold if $m > 1$); if Assumption 4 does not hold (for example, if $h_0(z) = 1$ and $h_1(z) = 2z/[1 + |z|]$ for $|z| > 1$), H is a proper subset of $A - \{0\}$.

EXAMPLE 2. *Polynomial regression* ($f_i(x) = x^i$), $m > 1$.

(a) *General results.* The Chebyshev points in the polynomial case are well known (for example, see [1]) to be $x_j^* = -\cos(j\pi/m)$, $0 \leqq j \leqq m$. Thus, R^* and S^* can be described explicitly from (2.20) and (2.27), as we shall do in detail below for $m = 2, 3$.

The set A^* has been characterized in [10] as follows: Define real-valued functions S_j and Q_h on E^m (whose points we write as $x = (x_0, \cdots, x_{m-1})$) by

$$S_j(x) = (-1)^j \textstyle\sum^{(j)} x_{i_1} x_{i_2} \cdots x_{i_j}, \qquad 1 \leqq j \leqq m, \tag{5.1}$$

where $\sum^{(j)}$ denotes summation over the set $0 \leqq i_1 < i_2 < \cdots < i_j < m$, and

$$Q_h(x) = \textstyle\sum^{(h)} (1 - x_{i_1})(1 - x_{i_2}) \cdots (1 - x_{i_h})(1 + x_{i_{h+1}}) \cdots (1 + x_{i_m})/\binom{m}{h}, \quad 0 \leqq h \leqq m, \tag{5.2}$$

where the m subscripts in the summand are distinct ($\sum^{(0)}$ consists of one term). Define the points $b^{(h)} = (b_0^{(h)}, \cdots, b_{m-1}^{(h)})$, $0 \leqq h \leqq m$, by

$$Q_h(x) = 1 + \textstyle\sum_{j=1}^m b_{m-j}^{(h)} S_j(x). \tag{5.3}$$

In particular, $b^{(m)} = (1, 1, \cdots, 1)$ and $b^{(0)} = ((-1)^m, (-1)^{m-1}, \cdots, (-1)^1)$. The points $b^{(0)}, \cdots, b^{(m)}$ can be shown not to lie in any hyperplane of E^m, so that they span an m-dimensional simplex. Let Δ_m denote this simplex minus the closed edge containing $b^{(0)}$ and $b^{(m)}$. The main result of [10] is

THEOREM 6. *For polynomial regression with $m > 1$, $A = \Delta_m$ and A_0 is empty.*

As we shall see in Example 2(c), R_0 is not generally empty.

We note that H is the twisted curve $\{(t^m, t^{m-1}, \cdots, t^1) : 0 < |t| < 1\}$, whose two open components have end-points $b^{(0)}, b^{(m)}$, and (in common) 0.

(b) *The case $m = 2$.* As in the case $m = 1$, a complete analysis of the a-optimum designs, for a in B^*, is possible here, but would be much more complicated as m increases, as will be seen in (c). We begin by describing the structure of b-optimum designs for all b in the (b_0, b_1)-plane B. From Theorem 6, we have

$$A = \text{triangle with vertices } (-1, 0), (1, 1), (1, -1),$$

minus closed segment joining the latter two.

Recalling the first paragraph of Example 2(a), we have $X_2^* = \{-1, 0, 1\}$ and

$$F_{R^*} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \qquad F_{S^*} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Setting $\alpha = \eta_0 + \eta_2$ and $\beta = \eta_2 - \eta_0$, we have from (2.20) that a general element a of R^*, obtained as a linear combination of the columns of F_{R^*}, is of the form $(\alpha - \eta_1, \beta, \alpha)$. Since $\alpha \neq 0$ for $\eta \varepsilon \Xi_m^*$, we obtain that R_0 is empty. Moreover, R is the set of points of the form $(b_0, b_1) = (1 - \eta_1/\alpha, \beta/\alpha)$. For $\eta \varepsilon \Xi_m^*$, the variables η_1/α and β/α can vary independently of each other over domains $(0, \infty)$ and $(-1, 1)$, respectively. Hence, $R = \{(b_0, b_1): b_0 < 1, |b_1| < 1\}$. Similarly, by (2.27), a point of S^* is of the form $(\alpha + \eta_1, \beta, \alpha)$, so that S_0 is empty, $S = \{(b_0, b_1): b_0 > 1, |b_1| < 1\}$. Thus, T_0 is empty and $\Gamma(T^*) = T = \{(b_0, b_1): b_0 \neq 1, |b_1| < 1\}$.

In subdividing the plane B into regions where the b-optimum designs are of various forms, we shall encounter repeatedly the parabola $b_0 = b_1^2$, which consists of 0, the set $H = \{(t^2, t): 0 < |t| < 1\}$, and the set J (say) where $b_0 = b_1^2 \geqq 1$. The point (t^2, t) of J with $|t| \geqq 1$ corresponds to the linear parametric function $f(t^{-1})\theta'$. Since $|t^{-1}| \leqq 1$, this linear parametric function can be estimated by the design $\xi^{(t)}$ (say) for which $\xi^{(t)}(t^{-1}) = 1$. It was shown in [6] that $\xi^{(t)}$ is admissible for $|t| \leqq 1$. Since $f(t^{-1})\theta'$ and its multiples are the only linear parametric functions estimatable under $\xi^{(t)}$, it follows that $\xi^{(t)}$ must be $f(t^{-1})$-optimum for $0 < |t| \leqq 1$. ($\xi^{(0)}$ will be discussed with B_0.) No two of these $\xi^{(t)}$'s allow estimation of the same linear parametric function. Hence, if there were a design other than $\xi^{(t)}$ which was also $f(t^{-1})$-optimum, it would have to be supported by at least two points, and thus it would allow estimation of some linear parametric function not estimatable under $\xi^{(t)}$, from which it follows easily that $\xi^{(t)}$ would be inadmissible. Thus, we have shown that

$$J \equiv \{(t^2, t): |t| \geqq 1\} = \{\text{points of } B \text{ where there is an optimum design supported by one point}\}.$$

(The analogue of this holds for general m, with $J = \{(t^m, t^{m-1}, \cdots, t): |t| \geqq 1\}$.)

In analyzing B further we shall use the fact that, for a design supported by *more than one point*, the residual to the best Chebyshev approximation of x^2 on $[-1, 1]$ by $\{1 - b_0x^2, x - b_1x^2\}$, being quadratic and attaining its maximum in absolute value at the points of support, must be of one of the following forms:

(i) a multiple of $x^2 - \frac{1}{2}$, with support a subset of X_2^*;

(ii) a constant;

(iii) a quadratic with derivative 0 at q, $0 < |q| \leqq 1$,

and with values of equal magnitude and opposite sign, at -1 and q if $q > 0$, and at 1 and q if $q < 0$, these two values being the support in the respective cases (the case $q = 0$ is covered by (i) above);

(iv) a multiple of $x^2 - L$ with $L < \frac{1}{2}$ and support $\{-1, 1\}$;

(v) a quadratic or linear function with nonzero derivatives of the same sign at ± 1, and with values of equal magnitude and opposite sign at the two points of support -1, 1.

Corresponding to these, there are three forms of the orthogonality relation (2.4) which we shall consider:

(I) the Equations (2.17), where we no longer demand that all ξ_j be positive, but only that two be positive and one nonnegative; this corresponds to (i) and, with $\xi(0) = 0$, to (iv) and the case of (ii) where the support is $\{-1, 1\}$;

(II) corresponding to (iii) and (v), the equations (a) and (b) for $q > 0$ and $q < 0$, respectively:

$$(5.4)\qquad \begin{aligned} &\text{(a)} \quad \begin{aligned} \xi(-1)(1 - b_0) &- \xi(q)(1 - q^2 b_0) = 0, \\ \xi(-1)(-1 - b_1) &- \xi(q)(q - q^2 b_1) = 0, \\ \xi(-1) &+ \xi(q) = 1; \end{aligned} \\ &\text{(b)} \quad \begin{aligned} \xi(1)(1 - b_0) &- \xi(q)(1 - q^2 b_0) = 0, \\ \xi(1)(1 - b_1) &- \xi(q)(q - q^2 b_1) = 0, \\ \xi(1) &+ \xi(q) = 1; \end{aligned} \end{aligned}$$

(III) corresponding to the part of case (ii) not covered in (I), equations which will be discussed below, and which lead to (5.8).

It is trivial that for each fixed b there exists a vector $c = (c_0, c_1)$ which yields a residual with each of the possible sets of extrema and oscillations of sign represented by (I), (II), and (III). Hence, in each of these three cases, any ξ with the given support is maximal relative to any c yielding such a residual, and if the orthogonality relations are satisfied then c is minimal with respect to ξ. From Theorem 2.3 it then follows that ξ is b-optimum; it is unnecessary to go back to (i)–(v) and compare residuals to find which approximation is best, where the best approximation is not unique, etc.

The regions where these three forms hold can be described as follows: partition B into disjoint sets B_{I}, B_{II}, B_{III}, J, defined by

$$(5.5)\qquad \begin{aligned} B_{\text{I}} &= \{(b_0, b_1) : b_0 \leqq 1, |b_1| \leqq 1\} - \{(1, 1)\} - \{(1, -1)\}, \\ B_{\text{II}} &= \{(b_0, b_1) : |b_1| > 1, b_0 < b_1^2\}, \\ B_{\text{III}} &= \{(b_0, b_1) : b_0 > \max(b_1^2, 1)\}. \end{aligned}$$

We shall show that, for L = I, II, III, the orthogonality relations of case L have a solution on two or more points if and only if $b \,\varepsilon\, B_L$.

Case I is treated by the same computation which yields R; in fact, the Equations (2.17) have a nonnegative solution on the closure of R,

$$(5.6)\qquad \begin{aligned} cl(R) &= B_{\text{I}} \cup \{(1, 1) \cup \{(1, -1)\} \\ &= R \cup \{(b_0, 1) : b_0 < 1\} \cup \{(b_0, -1) : b_0 < 1\} \cup \{(1, b_1) : |b_1| < 1\} \\ &\quad \cup \{(1, 1)\} \cup \{(1, -1)\}. \end{aligned}$$

In this last partition of (5.6) we have, respectively, none of the three $\xi(x_j^*)$'s zero, only $\xi(-1) = 0$, only $\xi(1) = 0$, only $\xi(0) = 0$, $\xi(1) = 1$, and $\xi(-1) = 1$, the last two being points of J. (The point corresponding to $\xi(0) = 1$, which does not arise from (2.17), will be seen later to be in B_0.)

In describing Case II, we shall use the partition of B_{II} into disjoint sets L_s, $-\infty < s < \infty$, defined as follows:

$$(5.7) \qquad \begin{aligned} L_s &= \{(b_0, b_1): b_0 - sb_1 = 1 + s, b_0 < b_1^2\} \qquad &\text{if } s \geqq 0, \\ &= \{(b_0, b_1): b_0 - sb_1 = 1 - s, b_0 < b_1^2\} \qquad &\text{if } s \leqq 0. \end{aligned}$$

Thus, L_s is that portion not in $cl(B_{III})$ of a line passing through $(1, -1)$ if $s \geqq 0$ and through $(1, 1)$ if $s \leqq 0$; in particular, $L_0 = \{(1, b_1): |b_1| > 1\}$. Consider now the orthogonality relations (5.4) (a) in the case $0 < q < 1$. Equating the ratios $\xi(-1)/\xi(q)$ in the first two equations, one obtains $b_0 - sb_1 = s + 1$ where $s = (1 - q)/q > 0$; from the positivity condition $0 < \xi(-1)/\xi(q) < \infty$ one obtains $0 < (1 - b_0q^2)/(1 - b_0) < \infty$ or $\{b_0 > q^{-2}\} \cup \{b_0 < 1\}$, which with $b_0 - sb_1 = s + 1$ yields L_s as the subset of B for which the b-optimum design is supported by the two points -1, q and the residual has opposite signs at these two points. (The support $\{-1, q\}$ arises in case III with constant residual.) The case $-1 < q < 0$ of (5.4) (b) similarly yields L_s with $s = (1 + q)/q < 0$. Finally, the case $q = 1$ of (5.4) (a) coincides with $q = -1$ in (5.4) (b) and yields L_0 as the subset of B for which the b-optimum design is supported by the two points 1, -1 with residual of opposite sign at the two points. (The set $\{(1, b_1): |b_1| < 1\}$ encountered in case I also has support $\{1, -1\}$, but with residual of the same sign at the two points.)

For any $b \varepsilon S$ every best Chebyshev approximation has constant residual (Theorem 2 and Lemma 5). Since for b in $B_I \cup B_{II}$ the residual is not constant, as we have seen, it follows that $S \subset B_{III}$. On $B_I \cup B_{II}$ the optimum design is unique because the orthogonality relations have a unique solution. For b in B_{III} there is no uniqueness of the b-optimum design. In fact, while the b-optimum design for each b in $B - B_{III}$ is unique and hence admissible, for each b in B_{III} there are infinitely many different supporting sets of admissible b-optimum designs, and also infinitely many different supporting sets (including supporting sets with an arbitrarily large finite number, or an infinite number, of points) of inadmissible b-optimum designs. Since the admissible b-optimum designs are of greater theoretical and practical interest, we shall exhibit only the totality of these, for each b in B_{III}. We shall then indicate by an example the existence of inadmissible b-optimum designs.

It was shown in [6] that the supports of admissible designs are of the form $\{-1, q, 1\}$ with $-1 < q < 1$, or subsets thereof, and conversely. The orthogonality relations (2.4) for the set $\{-1, q, 1\}$ in case III are

$$(5.8) \qquad \begin{aligned} &\xi(-1)(1 - b_0) + \xi(q)(1 - b_0q^2) + \xi(1)(1 - b_0) = 0, \\ &\xi(-1)(-1 - b_1) + \xi(q)(q - b_1q^2) + \xi(1)(1 - b_1) = 0, \\ &\xi(-1) + \xi(q) + \xi(1) = 1. \end{aligned}$$

We seek a nonnegative solution to these for which $0 < \xi(q) < 1$; this condition is equivalent to that of finding a solution for which at least two of the components are positive (to eliminate J) and for which $\xi(q) > 0$ (to eliminate the part of (ii) included in case I, namely, the interval $\{(1, b_1): |b_1| < 1\}$ denoted by L_0' below). In describing such solutions, it is convenient to write

$$(5.9)\quad \begin{aligned} L_s' &= \{(b_0, b_1): b_0 - sb_1 = 1 + s, b_0 > b_1^2\} && \text{if } s \geqq 0,\\ &= \{(b_0, b_1): b_0 - sb_1 = 1 - s, b_0 > b_1^2\} && \text{if } s \leqq 0;\\ M_r' &= \{(b_0, b_1): b_0 - rb_1 = 1 - r, b_0 > b_1^2\} && \text{if } r > 1,\\ &= \{(b_0, b_1): b_0 - rb_1 = 1 + r, b_0 > b_1^2\} && \text{if } r < -1,\\ &= \{(b_0, b_1): |b_1| = 1, b_0 > 1\} && \text{if } r = \infty. \end{aligned}$$

Thus, L_s', L_s and the two points (q^{-2}, q^{-1}) and $(1, -\text{sign } q)$ of J (or $(1, 1)$ and $(1, -1)$ if $s = 0$) constitute a partition of the line encountered in conjunction with (5.7). For $r \neq \infty$, M_r' is the intersection with B_{III} of a line of slope $1/r$ through $(1, 1)$ if $r > 0$ and through $(1, -1)$ if $r < 0$, while M_∞' consists of two half-lines in B_{III}.

The Equations (5.8) have the formal solution

$$(5.10)\quad \begin{aligned} \xi(q) &= (b_0 - 1)/b_0(1 - q^2),\\ \xi(1) &= [1 + b_1(1 - q) - b_0 q]/2b_0(1 - q),\\ \xi(-1) &= [1 - b_1(1 + q) + b_0 q]/2b_0(1 + q). \end{aligned}$$

The condition $0 < \xi(q) < 1$ is equivalent to $b_0 > 1$. We also require the nonnegativity of the numerators of $\xi(1)$ and $\xi(-1)$ in (5.10), with at least one being positive. It is easy to verify that $\xi(1) = 0$ on the line through $(1, -1)$ of slope $q/(1 - q)$, and that $\xi(-1) = 0$ on the line through $(1, 1)$ of slope $q/(1 + q)$. We conclude that, for $-1 < q < 1$, (5.8) has a solution for which $0 < \xi(q) < 1$ for b in the set V_q defined by

$$(5.11)\quad \begin{aligned} V_q &= \{\text{triangle with vertices } (1, 1)(1, -1), (q^{-2}, q^{-1})\} \cap B_{\text{III}} \qquad \text{if } 0 < |q| < 1,\\ V_0 &= \{(b_0, b_1): b_0 > 1, |b_1| \leqq 1\}. \end{aligned}$$

In each case, all three components of ξ are positive if b is in the interior of V_q, while two components are positive on that part of the boundary which is in V_q. The latter is M_∞' if $q = 0$ and, if $q \neq 0$, consists of the two open line segments L_s' and M_r', where $s = (1 - q)/q$ and $r = (1 + q)/q$ if $q > 0$, and $s = (1 + q)/q$ and $r = (1 - q)/q$ if $q < 0$. The rest of the boundary of V_q of course consists of the interval L_0' of B_{I} and the points $(1, 1)$, $(1, -1)$, and (if $q \neq 0)(q^{-2}, q^{-1})$ of J. Thus, for any point $b = (b_0, b_1)$ in B_{III} there is an admissible design supported by q and one or both of the points $1, -1$, provided that $b \in V_q$. From the condition of nonnegativity of $\xi(1)$ and $\xi(-1)$ in (5.10), this interval of q-values, always of positive length for b in B_{III}, is

$$\{q: (b_1 - 1)/(b_0 - b_1) \leqq q \leqq (b_1 + 1)/(b_1 + b_0)\},$$

the endpoints corresponding to designs for which ξ has only two nonzero components. Hence, for each b in B_{III}, there are infinitely many different supporting sets of admissible b-optimum designs.

As an illustration of inadmissible a-optimum designs for $\Gamma(a)$ in B_{III}, consider (2, 0, 1)-optimality, that is, optimality for estimating $2\theta_0 + \theta_2$. Among the admissible designs for this problem, obtained above, two examples are $q = 0$, $\xi(-1) = \xi(1) = \frac{1}{4}$, $\xi(0) = \frac{1}{2}$, for which

$$(5.12) \qquad M(\xi) = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \qquad M^{-1}(\xi) = \begin{pmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{pmatrix},$$

for which $V(a, \xi) = (2, 0, 1)M^{-1}(\xi)(2, 0, 1)' = 4$, and the design with $q = \frac{1}{2}$, $\xi(1) = 0$, $\xi(-1) = \frac{1}{3}$, $\xi(\frac{1}{2}) = \frac{2}{3}$, for which $M(\xi)$ is singular, but for which $V(a, \xi)$ is again 4. Among the many inadmissible designs are symmetric designs supported by $\{(1 - \epsilon), -(1 - \epsilon), 0\}$ with $0 < \epsilon < 1 - 2^{-\frac{1}{2}}$ and with $\xi'(\pm(1 - \epsilon)) = \frac{1}{4}(1 - \epsilon)^2$. For such a design

$$M(\xi') = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & (1-\epsilon)^2/2 \end{pmatrix},$$

$$M^{-1}(\xi') = [4/(1 - 4\epsilon + 2\epsilon^2)] \begin{pmatrix} (1-\epsilon)^2/2 & 0 & -\frac{1}{2} \\ 0 & (1 - 4\epsilon + 2\epsilon^2)/2 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix},$$

so that again $V(a, \xi) = 4$. The inadmissibility of such designs is exhibited in the fact that, for ξ given by (5.12), $M(\xi) - M(\xi')$ (or $M^{-1}(\xi') - M^{-1}(\xi)$) is nonnegative definite of rank one. It is not difficult to obtain inadmissible b-optimum ξ's here supported by any number of points $\geqq 2$ (a 2-point design being given by ξ' just above when $\epsilon = 1 - 2^{-\frac{1}{2}}$), or even with ξ absolutely continuous with positive Lebesgue density on $[-1, 1]$.

It remains to consider B_0. The unique optimum design for estimating θ_0 is, by the same argument used in discussing J, that for which $\xi(0) = 1$. For $-\infty < s < \infty$, the unique optimum design for estimating $s\theta_0 + \theta_1$, to which corresponds the problem of approximating x on $[-1, 1]$ by $\{x^2, 1 - sx\}$, is easily found by calculations parallel to those for B (solutions to the orthogonality relations now existing only in the case corresponding to II above). We obtain an optimum design supported by $\{-1, q\}$ if $q > 0$ and $s = (1 - q)/q$, and by $\{1, q\}$ if $q < 0$ and $s = (1 + q)/q$. In particular, for $s = 0$ we obtain the unique optimum design for estimating θ_1, for which $\xi(-1) = \xi(1) = \frac{1}{2}$. We note, then, that if we think of $B_0 = P^1 = \{b_{(s)}, -\infty < 1/s \leqq \infty\}$ in the usual manner, $b_{(s)}$ being the "point at infinity" of all lines in B of slope $1/s$, $-\infty < 1/s \leqq \infty$, then the optimum designs for these points $b_{(s)}$ of B_0 can be obtained, for

$-\infty < 1/s < \infty$, as limits of the corresponding designs for the family L_s ; for $s = \infty$ we have the optimum design for estimating θ_0 considered at the outset of this paragraph, which can be thought of conveniently as the limit of designs as $b \to \infty$ in R or in B_{I} or in B_{III} or in J.

We note also that, in the notation of and the sentence following (2.10),

$$\bar{T} = \{(b_0, b_1): |b_1| \leqq 1\} \cup \{(1, b_1): -\infty < b_1 < \infty\},$$
$$\bar{T}^* = \Gamma^{-1}(\bar{T}) \cup \{b_{(0)}\} \cup \{b_{(\infty)}\}.$$

As an example of the explicit computation of how large the "sufficiently large" of Theorem 5 is, we consider the case $k(e) \equiv -1$, $e > 1$. For $b = \Gamma(f(e)) = (e^{-2}, e^{-1})$, writing $e^{-1} = t$, (2.17) yields

$$\xi^{(e)} = (\xi_0^{(e)}, \xi_1^{(e)}, \xi_2^{(e)}) = [2(2 - t^2)]^{-1}(1 - t, 2(1 - t^2), 1 + t),$$

so that

$$M(\xi^{(e)}) = [1/(2 - t^2)]\begin{pmatrix} 2 - t^2 & t & 1 \\ t & 1 & t \\ 1 & t & 1 \end{pmatrix},$$

$$M^{-1}(\xi^{(e)}) = [(2 - t^2)/(1 - t^2)]\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -t \\ -1 & -t & 2 \end{pmatrix},$$

and thus

$$V(f(y), \xi^{(e)})\cdot(1 - t^2)/(1 - 2t^2) = 1 - y^2 - 2ty^3 + 2y^4 = p_t(y) \quad \text{(say)}.$$

The function p_t is easily seen to have local minima at $y = [3t \pm (9t^2 + 16)^{\frac{1}{2}}]/8$ and a local maximum at $y = 0$, all three of these points being in $[-1, 1]$. Thus, $\max_{-1 \leqq y \leqq e} p_t(y) = \max\,(p_t(-1), p_t(0), p_t(t^{-1}))$. Since $p_t(-1) > p_t(0)$, we seek t such that $p_t(-1) - p_t(t^{-1}) \leqq 0$, that is, such that $2t^4 - t^3 + t^2 + 2t - 2 \leqq 0$. This last is satisfied for $t \leqq .694$, or $t^{-1} \geqq 1.44$. Thus, for $e \geqq 1.44$ the design $\xi^{(e)}$ minimizes $\max_{-1 \leqq y \leqq e} V(f(y), \xi)$. We remark that it is even easier to conclude that, since $p_t(y)$ is increasing for $y \geqq 1$, the design $\xi^{(e)}$ minimizes $\max_{1 \leqq y \leqq e} V(f(y), \xi)$ for $e \geqq 1$.

(c) *The case* $m = 3$. The set $A = \Delta_3$ of Theorem 6 is determined by the points

$$b^{(0)} = (-1, 1, -1),$$
$$b^{(1)} = (1, -\tfrac{1}{3}, -\tfrac{1}{3}),$$
$$b^{(2)} = (-1, -\tfrac{1}{3}, \tfrac{1}{3}),$$
$$b^{(3)} = (1, 1, 1).$$

The set X_3^* is $\{-1, -\frac{1}{2}, \frac{1}{2}, 1\}$, and

$$F_{R^*} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & \frac{1}{2} & \frac{1}{2} & -1 \\ 1 & -\frac{1}{4} & \frac{1}{4} & -1 \\ -1 & \frac{1}{8} & \frac{1}{8} & -1 \end{pmatrix}.$$

Setting $\alpha = \eta_0 + \eta_3$, $\beta = \eta_0 - \eta_3$, $\gamma = \eta_1 - \eta_2$, $\sigma = \eta_0 + \eta_1 + \eta_2 + \eta_3$, we obtain from (2.20) that a general element of R^* has the form

$$a = (\beta - \gamma, (\sigma - 3\alpha)/2, (4\beta - \gamma)/4, (\sigma - 9\alpha)/8). \tag{5.13}$$

Thus, R_0 is no longer empty as it was when $m = 2$. To obtain R, we consider η to be in Ξ_m and thus $\sigma = 1$, and find

$$\begin{aligned} R = \{(8[\beta - \gamma]/[1 - 9\alpha], 4[1 - 3\alpha]/[1 - 9\alpha], \\ 2[4\beta - \gamma]/[1 - 9\alpha]) : (\alpha, \beta, \gamma) \in \{0 < \alpha < 1, \\ \alpha \neq \tfrac{1}{9}\} \cap \{-1 < \beta/\alpha < 1\} \cap \{-1 < \gamma/(1 - \alpha) < 1\}\}; \end{aligned} \tag{5.14}$$

the value $\alpha = \frac{1}{9}$ yields points in R_0, discussed below. The variables α, β/α $\gamma/(1 - \alpha)$ vary independently in (5.14), so that for b in R the range of b_1 is $(-\infty, 1) \cup (4, \infty)$. For each fixed value k of b_1 (that is, for each fixed value of α) the range of (b_0, b_2) in (5.14) is an open parallelogram $R(k)$ (say) in the plane $b_1 = k$, symmetric about $(0, k, 0)$, but whose dimensions and angles depend on k. Thus,

$$R = \bigcup_{b_1 < 1 \text{ or } > 4} R(b_1)$$

is no longer connected as it was when $m = 2$.

The set R_0 can be obtained as the set of elements of (5.13) with $\sigma = 9\alpha = 1$ and $b_1 = (\sigma - 3\alpha)/2 = \frac{1}{3}$; this is, by an analysis similar to that of (5.14), the set of ratios $(b_0/b_1, b_2/b_1) = (3(\beta - \gamma), 3(\beta - \gamma/4))$ in the region $|\beta| < \frac{1}{9}$, $|\gamma| < \frac{8}{9}$. This can be thought of as a "parallelogram at infinity" corresponding to the ratios $(b_0/b_1, b_2/b_1)$ of (5.14) as $\alpha \to \frac{1}{9}$.

The set S^* can be analyzed similarly. As was the case with R, the sets S and T no longer have the simple structure of the case $m = 2$. As in the next to last paragraph of Example 1, this again points up the greater simplicity of working with R^*, S^*, and T^*. The convexity of the cones which constitute half of $R^*$and S^* can of course be carried over to R and S in a different parametrization, one in which R_0 and S_0 are empty so that $R = \Gamma(R^*)$ can be thought of as a base (section) of a cone which constitutes half of R^*, and similarly for S. Thus, in place of x^3 we seek a function $\tilde{f}_3(x) = \sum_0^3 \lambda_i x^i$ such that, if we work in terms of $\tilde{f} = (\tilde{f}_0, \tilde{f}_1, \tilde{f}_2, \tilde{f}_3)$ instead of f, the quantity $\sum_0^3 (-1)^j \eta_j \tilde{f}_3(x_j^*)$, which corresponds to the last element of (5.13), and the quantity $\sum_0^3 \eta_j \tilde{f}_3(x_j^*)$ for the corresponding development for S^*, are never 0 for $\eta \in \Xi_3^*$. Writing $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, this says that all non-zero elements (there is at least one such) of $(\zeta_0, \zeta_1, \zeta_2, \zeta_3) = \lambda F_{R^*}$ must be of the same sign, and all non-zero elements of $(\zeta_0, -\zeta_1, \zeta_2, -\zeta_3) = \lambda F_{S^*}$ must be of the same sign. Hence, either (i) ζ_0 or $\zeta_2 \neq 0$, $\zeta_0\zeta_2 \geqq 0$, $\zeta_1 = \zeta_3 = 0$, or else (ii) ζ_1 or $\zeta_3 \neq 0$, $\zeta_1\zeta_3 \geqq 0$, $\zeta_2 = \zeta_4 = 0$. Since

$$\begin{pmatrix} -1 & 1 & 4 & -4 \\ -1 & 2 & 1 & -2 \\ 1 & 2 & -1 & -2 \\ 1 & 1 & -4 & -4 \end{pmatrix} F_{R^*} = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & 0 \\ 0 & 0 & \frac{3}{2} & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix},$$

the solutions in case (i) are easily seen to be $\lambda = \pm[k_0(-1, 1, 4, -4) + k_2(1, 2, -1, -2)]$ with $k_0 \geqq 0$, $k_2 \geqq 0$, $k_0 + k_2 > 0$, and in case (ii) they are $\lambda = \pm[k_1(-1, 2, 1, -2) + k_3(1, 1, -4, -4)]$ with $k_1 \geqq 0$, $k_3 \geqq 0$, $k_1 + k_3 > 0$. For any such λ and $\tilde{f}_3$ we can, for example, take $\tilde{f}_i = x^i$ for $0 \leqq i \leqq 2$ and the transformation from f to $\tilde{f}$ will be nonsingular.

A development analogous to that of the previous paragraph can be carried out for general m.

We shall not analyze B^* further in the manner of Example 2(b). The number of cases to be treated and the complexity of the resulting regions increase with m, as is evident even from the above characterization of R.

EXAMPLE 3. Other Chebyshev systems are discussed in the literature of approximation theory. As illustrated in Example 2.1 of Remark 1, Assumption 2 is somewhat stronger than the assumption that $\{f_i, 0 \leqq i \leqq m\}$ is Chebyshev. The sufficient condition for Assumption 2 which is given in Proposition 2.1 of Remark 1(a) is useful in applications, as is the condition of Proposition 2.2.

REFERENCES

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

[2] BERNSTEIN, S. (1926). *Lecons sur les Propriétés Extrémales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réele*. Gauthier-Villars, Paris.

[3] CHERNOFF, H. (1953). Locally optimum designs for estimating parameters. *Ann. Math. Statist.* **24** 586–602.

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

[4a] HOEL, P. G. (1965). Optimum designs for polynomial extrapolation. *Ann. Math. Statist.* **36** 1483–1493.

[5] HOEL, P. G. and LEVINE, A. (1964). Optimal spacing and weighting in polynomial prediction. *Ann. Math. Statist.* **35** 1553–1560.

[6] KIEFER, J. (1959) Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **21** 272–319.

[7] KIEFER, J. (1962). Two more criteria equivalent to *D*-optimality of designs. *Ann. Math. Statist.* **33** 792–796.

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

[9] KIEFER, J. and WOLFOWITZ, J. (1964 a, b). Optimum extrapolation and interpolation designs, I and II. *Ann. Inst. Statist. Math.* **16** 79–108 and 295–303.

[10] KIEFER, J. and WOLFOWITZ, J. (1965). On a problem connected with the Vandermonde determinant. *Proc. Amer. Math. Soc.* **16**.

[10a] TIMAN, A. F. (1963). *Theory of Approximation of Functions of a Real Variable*, Macmillan, N.Y.

[11] VOLKOV, V. I. (1958). Some properties of Chebyshev systems. *Kalinin Gos. Ped. Inst. Uč. Zap.* **26** 41–48.