OPTIMUM EXPERIMENTAL DESIGNS V, WITH APPLICATIONS TO SYSTEMATIC AND ROTATABLE DESIGNS

J. KIEFER
CORNELL UNIVERSITY

1. Introduction and summary

In this paper we continue to develop the theory of construction of optimum experimental designs along the lines of [15], [19], [16], and [17]. Section 2 of the paper considers further general developments in both the exact and approximate theories, while in section 3 we apply the theory to construct optimum designs in the settings where systematic designs (subsection 3.1) and rotatable designs (subsection 3.2) are often employed, and in the setting of linear regression on an arbitrary Euclidean subset (subsection 3.3). Open problems are mentioned throughout the paper.

2. Generalities

2.1. *Notation and preliminaries.* Throughout this paper we shall achieve brevity by considering mainly a linear model. Corresponding asymptotic results in nonlinear problems hold and are obtainable without serious difficulty. One example of such a problem will be found in subsection 3.1 (ρ unknown), and further examples of explicit computations in certain nonlinear problems will be found in Chernoff [6] and Box and Lucas [4], while complete classes of designs for such problems were treated by the author (see [16], pp. 290–291). We shall also be primarily concerned with nonsequential designs, although one sequential problem is treated below theorem 3.1.2. The main idea in the construction of many such asymptotic sequential designs goes back to Wald [24], while recent work can be found in the papers of Chernoff [7] and his students.

We assume, then that $f_1, f_2, \cdots, f_k$ are k given real functions on a space $\mathfrak{X}$. Write f for the column vector of functions f_i. Let θ denote a real unknown column k-vector. Corresponding to each x in $\mathfrak{X}$, there is a random variable Y_x for which

$$E_\theta Y_x = \theta' f(x). \tag{2.1.1}$$

(Throughout this paper transposes are denoted by primes and subscripts on E

Research sponsored by the Office of Naval Research.

or P refer to the distribution under which an expectation or probability is computed.)

In various applications there will be specified the possible distributions of Y_x, the dependence among various Y_x, and so forth.

An *exact* or *discrete* design will now be defined. An integer N, the total number of observations to be taken, is specified. An exact or discrete design d is a choice of N points $x_1, x_2, \cdots, x_N$ in $\mathfrak{X}$. Sometimes it will be permitted that several x_i are equal, in which case the Y_{x_i} corresponding to two equal x_i will usually not be the same random variable. It may also be the case that certain restrictions are imposed on the allowable choices of $x_1, \cdots, x_N$. For example, in the setting of two-way heterogeneity where $v \times v$ Latin square designs are customarily employed ($N = v^2$), we can take $\mathfrak{X}$ to be the space of v^3 triples (i, j, k), $1 \leqq i, j, k \leqq v$, a design being restricted to a choice of $x_1, \cdots, x_N$ for which no two x_i agree in both of their first two coordinates.

In many applications the Y_{x_i} are assumed to be uncorrelated and to have common (perhaps unknown) variance σ^2. In this case, for any design $d = (x_1, \cdots, x_N)$, the matrix

$$A_d = \sum_{i=1}^{N} f(x_i)f(x_i)' \tag{2.1.2}$$

is called the *information matrix* of the design d. If all components of θ are estimable under d, then $\sigma^2 A_d^{-1}$ is the covariance matrix of best linear estimators (b.l.e.). See [15] and [19] for a discussion of why it suffices to consider linear estimators.

The computation of optimum designs in the above setting will be discussed briefly in subsection 2.2. This is the *exact* or *discrete* theory.

Suppose that there is no restriction on the choice of the x_i in the above setting. Let $\xi_d(x)$ be the proportion of x_i for $1 \leqq i \leqq N$ which are equal to x when design d is used. Then ξ_d can be thought of as a probability measure on $\mathfrak{X}$. If ξ is any probability measure on $\mathfrak{X}$ (there will never be any measure-theoretic difficulties), we write

$$m_{ij}(\xi) = \int f_i(x)f_j(x)\xi(dx) \tag{2.1.3}$$

and $M(\xi) = ||m_{ij}(\xi)||$. Thus, for an exact design d, we have $A_d = NM(\xi_d)$. We shall call $M(\xi)$ the *information matrix of* ξ. A typical optimality criterion in the design of experiments is to choose d to minimize some simple real functional Q of A_d, as we shall discuss in the next two subsections. An exact design d is a probability measure ξ_d taking on only values which are integral multiples of $1/N$. Suppose we find a probability measure ξ^* which minimizes $Q[M(\xi)]$ over *all* probability measures ξ on $\mathfrak{X}$. Clearly, it can happen that ξ^* takes on values other than multiples of $1/N$, and thus does not correspond to an exact design. Nevertheless, we shall consider this problem of minimizing $Q[M(\xi)]$ over all ξ, calling this the *approximate* or *continuous* theory and calling any probability measure ξ on $\mathfrak{X}$ an *approximate* design.

There are three reasons for considering the approximate theory: (1) the exact theory will often exhibit a fine structure dependence on N, necessitating a lengthy table of optimum designs for a given problem, whereas one optimum approximate design is relevant for all N; (2) the optimum approximate design immediately yields an exact design for each N which is optimum to within order N^{-1}, and often will turn out to be an exact design for many N; (3) the exact theory often presents a difficult combinatorial problem admitting no simple method of solution, whereas the approximate theory admits simple computational algorithms such as those mentioned in subsection 2.3.

In the present paper we shall consider the computation of designs which are optimum with respect to certain specific criteria. The reader is referred to Kiefer [16] and to Elfving [12] for proofs and listings of results on such related topics in optimum design theory as admissibility and complete classes of designs, the role of randomized designs, the computations associated with other optimality criteria, and so forth.

2.2. *The exact theory.* Results in the exact theory have been obtained mainly in the settings where incomplete block designs, factorial designs, et cetera, are often employed, rather than in regression experiments where $\mathfrak{X}$ is a continuum. Many scattered results for various settings and optimality criteria were obtained by various authors (see [16] and [12] for listings), all of these results being obtainable from an elementary approach of the author [15], [16]. Let Q be a $v \times v$ orthogonal matrix whose first row Q_1 is constant, and let Q_2 denote its last $v - 1$ rows. If C_d is Bose's $v \times v$ information matrix of design d for the varieties in a block design setting where there are v varieties (of course, $v < k$, where k was defined in subsection 2.1), then $Q_2C_dQ_2'$ is proportional to the inverse of the covariance matrix V_d of best linear estimators of contrasts $\psi = Q_2\bar{\theta}$ of the variety effects $\bar{\theta}' = (\theta_1, \cdots, \theta_v)$. One proves easily that, if d^* *maximizes the trace of* C_d *and* C_{d^*} *has all diagonal elements equal and all off-diagonal elements equal*, then tr V_d is maximized by d^*, and V_{d^*} is a multiple of the identity. From this we conclude that any design d^* with the above italicized properties is optimum according to any of a wide variety of optimality criteria which were considered separately by various authors. These criteria include:

(a) D-optimality: minimizing the generalized variance, or det V_d;

(b) A-optimality: minimizing the average variance, or tr V_d;

(c) E-optimality, called minimaxity with respect to all standard parametric forms in [12]: minimizing the largest eigenvalue of V_d;

(d) called minimaxity with respect to single parameters in [12]: minimizing the maximum diagonal element of V_d;

(e) maximizing the average efficiency, that is, minimizing the average of the variances of best linear estimators of $\theta_i - \theta_i$, this average being easily proved to be proportional to tr V_d;

(f) L-optimality: maximizing, in the Gaussian case, the minimum power on spheres $\psi'\psi = c^2$ as $c \to 0$, for testing the hypothesis $\psi = 0$. If σ^2 is unknown, d^* must also maximize the number of degrees of freedom for error among designs for which ψ is estimable, to insure L-optimality.

As discussed in [15] and [16], properties such as (a) through (e) also have interpretations in terms of power properties for hypothesis testing problems in the Gaussian case; remarkably enough, these optimality properties for testing problems are no longer possessed by the standard symmetrical designs if they are compared with certain intuitively less appealing randomized designs.

The interpretation of these criteria with regard to confidence region problems in the Gaussian case, such as those considered by Scheffé [22], is also well known.

In a given design setting, for example, that where balanced incomplete block designs are customarily used, one need only verify that the design d^* has the property italicized above in order to conclude the optimality of d^* in all of these senses. The corresponding results in settings where $\bar{\theta}$ and not merely ψ is to be estimated, are usually even simpler to obtain. This often involves only elementary arithmetic. A comparatively difficult example is that of the generalized Youden square, for which optimality has not yet been proved if neither the number of rows nor the number of columns is divisible by v; in fact, the above method fails in this case (see [15]).

Of course, there are many design settings, especially where $\mathfrak{X}$ is a continuum, wherein the various criteria (a) through (f) above need not lead to the choice of the same design. In such cases, if, as is often the case, one does not have a well-specified loss function, one may want to choose one of these criteria. The criterion (a) seems to the author to have several appealing properties in such circumstances. For example, if the problem is that of polynomial regression of degree $k - 1$ on a given interval, this criterion alone among those listed yields a design which does not depend on the scale of measurement or on which k linear functions of coefficients are chosen as parameters. Another optimality criterion one might consider is

(g) G-optimality: minimize the maximum (over $\mathfrak{X}$) variance $N^{-1}\sigma^2\bar{d}(\xi)$ of the estimated regression function.

A second appealing property of (a) is that, in the approximate theory, it is equivalent to (g) [20]. These two criteria had both been considered often in the past, but as different criteria; see, for example, [4], p. 89. Further properties of the various optimality criteria are discussed in [15] and [16].

G-optimality is not generally equivalent to D-optimality in the exact theory. For example, in the problem of linear regression on the interval $[-1, 1]$, with $N = 3$, let ξ_1 be the exact design for which $\xi_1(-1) = \xi_1(0) = \xi_1(1) = 1/3$, and let ξ_2 be the exact design for which $\xi_1(-1) = 1/3$, $\xi_2(1) = 2/3$. It is easy to verify that $\det M(\xi_1) = 2/3$, $\bar{d}(\xi_1) = 5/2$, $\det M(\xi_2) = 8/9$, $\bar{d}(\xi_2) = 3$, and that ξ_1 is G-optimum among all exact designs, while ξ_2 is D-optimum among all exact designs. The same result holds if $\mathfrak{X}$ is replaced by the set consisting only of the three points $-1, 0, 1$. Nevertheless, it is often true, especially in settings where each f_i can only take on two values, that G-optimality and D-optimality are still equivalent, although this does not follow from theorem 2.3.1 below; it would be enlightening to investigate this relationship further (bounds like those of subsection 2.3 might prove useful). A related problem is that of making precise the idea that, in settings where an appropriate symmetrical design (for example, a *BIBD*) does not exist, that design which is "closest" to it in some sense will often have optimum properties. The exact design which is closest to an approximate design which is optimum, seems to possess similar properties.

In some examples of the discrete theory, one can also obtain the desired result by applying the methods of the next subsection and noting that an approximately optimum design turns out to be exactly optimum. For example, we shall describe two methods for proving the D-optimality of the Latin square design d^* in the setting described in subsection 2.1, and it is easy to verify that D-optimality implies the other types of optimality because of the structure of V_{d^*} in this case.

Let $\theta' = (\alpha', \beta', \gamma', \mu)$, where α, β, and γ are the v-vectors of variety, row, and column effects, respectively, and μ is the "grand mean." We may assume $\sum \alpha_i = \sum \beta_j = \sum \gamma_k = 0$. Thus, $EY_{ijk} = \alpha_i + \beta_j + \gamma_k + \mu$. Let Q be a $v \times v$ orthogonal matrix of the type described earlier. Let $\bar{\alpha} = Q_2\alpha$, $\bar{\beta} = Q_2\beta$, $\bar{\gamma} = Q_2\gamma$, and let $\phi' = (\bar{\alpha}', \bar{\beta}', \bar{\gamma}', \mu)$, so that ϕ has maximal dimension among estimable vectors. Since A_{d^*}, in terms of θ, consists of four diagonal blocks each of which is a multiple of an identity and six pairs of off-diagonal blocks each of which is a constant matrix, we obtain easily that A_{d^*}, in terms of ϕ, consists of four diagonal matrices each of which is a multiple of the identity, and is zero elsewhere. It follows at once that, if ξ^* is the measure corresponding to d^*, then the function $D_{\bar{\alpha}}(x, \xi^*)$ of the next subsection is constant on the v^3 points of $\mathfrak{X}$, so that d^* is D-optimum for estimating contrasts of the α_i. The D-optimality of d^* for estimating *all* parameters ϕ is a consequence of the even more obvious fact that the variance $N^{-1}\sigma^2 d(x, \xi^*)$ of the estimated regression is the same at each of the v^3 points x. We have purposely refrained from explicit computation of A_{d^*} here; it was unnecessary, only the form of A_{d^*} being important, in view of theorem 2.3.2!

A second method of proof, which uses the invariance results referred to in the next section, is even shorter. These results imply that the invariant design $\bar{\xi}$ which assigns measure $1/v^3$ to each point of $\mathfrak{X}$, is (in the approximate theory) D-optimum for $\bar{\alpha}$ and also for ϕ. Since $M(\xi^*) = M(\bar{\xi})$, we conclude that the Latin square design d^* is (exactly) D-optimum.

Unfortunately, these approaches do not work in all settings; for example, this is the case for a Youden square which is not a Latin square, and we would have to fall back on the earlier method of proof in that problem.

We end this subsection by mentioning a common misconception which has occurred repeatedly in the literature. Authors have often restricted their attention to designs for which the b.l.e. are orthogonal, or are orthogonal in or between intuitively appealing sets (for example, blocks and varieties), apparently assuming that optimum designs are to be found among such designs. While it is well known that some optimum designs possess such orthogonality properties, it is only orthogonality *in combination with some other property such as the trace maximization we have mentioned* which yields optimality. However, in experiments such as those concerned with polynomial regression, we rarely have such orthogonality, although the misconception persists even in that familiar setting (see, for example, the discussion of [16], p. 306, lines 9–13 and p. 316, lines 17–23). For another example, we note that it is common in designing experiments where a time trend is to be removed, to restrict attention to designs for which the b.l.e. of time trend is orthogonal to the b.l.e. of effects in which one is interested. In subsection 3.1, we shall give an example to illustrate the bad consequences of such an orthogonality restriction in this setting.

2.3. *The approximate theory; invariance; bounds; extensions.* The earliest general development of algorithms for computing optimum designs is due to Elfving [11] in the case of the average variance, further developments being due to Chernoff [6]. Kiefer and Wolfowitz [16] developed algorithms for various optimality criteria, of which we shall be concerned in the present paper with the generalized variance. Further results for that criterion were obtained by the author[17].

If ξ is an approximate design for which θ is estimable, that is, for which $M(\xi)$ is nonsingular, then the variance of the b.l.e. of the estimated regression at the point x is $\sigma^2 N^{-1} d(x, \xi)$, where

$$d(x, \xi) = f(x)' M^{-1}(\xi) f(x). \tag{2.3.1}$$

Assume for simplicity that f is continuous in a topology for which $\mathfrak{X}$ is compact, and write $\bar{d}(\xi) = \max_x d(x, \xi)$. The equivalence of D- and G-optimality alluded to in subsection 2.2 is contained in the following theorem [20]:

THEOREM 2.3.1. *ξ is D-optimum if and only if it is G-optimum, and if and only if $\bar{d}(\xi) = k$. For all D-optimum ξ, $M(\xi)$ is the same.*

This theorem has been the basis for computing optimum designs in problems of polynomial regression on a simplex or hypercube [17]. A generalization of theorem 2.3.1, of use when we are interested in a subset $\theta^{(1)} = (\theta_1, \cdots, \theta_s)'$ of the parameters, has been proved in [17]. Partition θ into $\theta^{(1)}$ and $\theta^{(2)}$[a $(k - s)$-vector], f into $f^{(1)}$ and $f^{(2)}$, and $M(\xi)$ into

$$M(\xi) = \begin{Vmatrix} M_1(\xi) & M_2(\xi) \\ M_2(\xi)' & M_3(\xi) \end{Vmatrix}, \tag{2.3.2}$$

where M_1 is $s \times s$. The information matrix for estimating $\theta^{(1)}$ can be written as $M^* = M_1 - M_2 M_3^{-1} M_2'$ if M is nonsingular. Thus, ξ is D-optimum for $\theta^{(1)}$ if it maximimizes $\det M^*(\xi)$. Write $D_{\theta^{(1)}}(x, \xi) = d(x, \xi) - f^{(2)}(x)' M_3^{-1}(\xi) f^{(2)}(x)$ and $D(\xi) = \max_x D_{\theta^{(1)}}(x, \xi)$. One of the results of [17] is

THEOREM 2.3.2. *Suppose $M(\xi^*)$ is nonsingular. Then ξ^* is D-optimum for $\theta^{(1)}$ if and only if ξ^* minimizes $D(\xi)$, and if and only if $D(\xi^*) = s$.*

The reader is referred to [17] for a determination of the structure of the class of matrices $M(\xi)$ for which ξ is D-optimum for $\theta^{(1)}$, for the modifications necessary when $\theta^{(1)}$ is estimable but all of θ is not, et cetera. When $s = k$ we obtain theorem 2.3.1, and when $s = 1$ we obtain a different proof of the algorithm of [19] for this case. When $1 < s < k$, the algorithm of [19] differs from that of theorem 2.3.2, the latter so far seeming to have yielded simpler arithmetic in examples (see [17]).

Invariance. The subject of invariant optimum designs, invariant complete classes for $\theta^{(1)}$, et cetera, has been dealt with extensively in [16]. Here we will only comment briefly on a slightly different approach to invariance where D- (hence, G-) optimality is involved. Instead of working with $\det M(\xi)$ or $\det M^*(\xi)$ as in [16], we can work with $d(x, \xi)$ or $D(x, \xi)$. For example, let $\mathcal{G}$ be a compact group of transformations on $\mathfrak{X}$ with Haar measure μ, $\mu(\mathcal{G}) = 1$. Suppose that, for g in $\mathcal{G}$, there is an associated transformation $\bar{g}$ on the space of ξ, such that

$$d(gx, \xi) = d(x, \bar{g}\xi). \tag{2.3.3}$$

Writing

$$\bar{\xi} = \int (\bar{g}\xi)\mu(dg), \tag{2.3.4}$$

the trivial fact that $\lambda A^{-1} + (1 - \lambda)B^{-1} - [\lambda A + (1 - \lambda)B]^{-1}$ is nonnegative definite for $0 \leqq \lambda \leqq 1$ if A and B are, yields

$$\begin{aligned} \sup_x d(x, \xi) &\geqq \sup_x \int d(gx, \xi)\mu(dg) = \sup_x \int d(x, \bar{g}\xi)\mu(dg) \\ &= \sup_x f(x)' \left\{\int M^{-1}(\bar{g}\xi)\mu(dg)\right\} f(x) \\ &\geqq \sup_x f(x)' \left\{\int M(\bar{g}\xi)\mu(dg)\right\}^{-1} f(x) = \sup_x d(x, \bar{\xi}). \end{aligned} \tag{2.3.5}$$

Since $\bar{\xi}$ is an invariant design, that is, $\bar{\xi}(gA) = \bar{\xi}(A)$ for all g and A, we conclude:

THEOREM 2.3.3. *Under the above conditions, there is a $\mathcal{G}$-invariant D- (and G-) optimum design for estimating θ.*

This result can be extended without serious difficulty to noncompact groups (just as with the usual invariance theory in statistics, for example, as in [18]), as well as to the estimation of $\theta^{(1)}$, in the same way that such extensions were obtained in [17].

Bounds. For computational purposes in obtaining designs which are almost optimum in complex settings, it is useful to know when one has obtained a design which is sufficiently close to optimality for practical purposes. For the criterion of G-optimality of θ, we have such a method: $[\bar{d}(\xi) - k]/k$ is, by theorem 2.3.1, the relative excess of $\bar{d}(\xi)$ over the minimum attainable. No such simple expression is available in terms of $\det M(\xi)$, but we can use this expression in terms of $\bar{d}(\xi)$ to obtain a bound on $\det M(\xi)/\max_{\xi'} \det M(\xi')$. Thus, one can in practice compute $\bar{d}(\xi)$ and, if it is close to k, conclude that $\det M(\xi)$ is within a bound we shall derive, from the maximum attainable. For completeness we shall give such bounds in both directions. For brevity we shall consider only the case $s = k$; the case $s < k$ can be treated similarly.

Write $\Delta = \max_\xi \det M(\xi)$. Suppose η is D-optimum. Fix ξ. An elementary computation of the first two derivatives of $q(\alpha) = \log\det M[\alpha\eta + (1-\alpha)\xi]$ for $0 < \alpha < 1$ (see [20] or [17] for details) yields

$$q'(0) = \frac{d}{d\alpha} q(\alpha)\Big|_{\alpha=0} \leqq \bar{d}(\xi) - k, \qquad \frac{d^2}{d\alpha^2} q(\alpha) \leqq 0, \tag{2.3.6}$$

and hence

$$\frac{d}{d\alpha} q(\alpha) \leqq \bar{d}(\xi) - k \qquad \text{for} \quad 0 \leqq \alpha \leqq 1. \tag{2.3.7}$$

Hence, $q(1) - q(0) \leqq \bar{d}(\xi) - k$, and thus

$$\frac{\det M(\xi)}{\Delta} \geqq \exp[k - \bar{d}(\xi)]. \tag{2.3.8}$$

Inequality (2.3.8) is easily seen to be strict unless ξ is D-optimum.

In the other (less useful) direction, suppose $\bar{d}(\xi) = k + \epsilon$. For simplicity, assume $\epsilon \leqq 1$. Again following the derivation of [20] or [17], we conclude that there is an η (not necessarily optimum) for which $q'(0) = \epsilon$. Let A be nonsingular and such that $AM(\xi)A'$ is the identity and $AM(\eta)A'$ is diagonal with diagonal entries d_i. Then, for $0 \leqq \alpha \leqq 1$,

$$\frac{d^2}{d\alpha^2} q(\alpha) = -\sum_{i=1}^{k} \frac{(d_i - 1)^2}{(1 - \alpha + \alpha d_i)^2}. \tag{2.3.9}$$

Also, the fourth derivative of $q(\alpha)$ is nonpositive, so that $d^2q(\alpha)/d\alpha^2$ is concave and thus attains its minimum L on $0 \leqq \alpha \leqq r \leqq 1$ at 0 or r. Thus,

$$L = -\max\left[\sum_i (d_i - 1), \sum_i \frac{(d_i - 1)^2}{(1 - r + rd_i)^2}\right]. \tag{2.3.10}$$

Now, $\sum(d_i - 1)^2$ is convex in the d_i on the set $B = \{(d_1, \cdots, d_k)|$ all $d_i \geqq 0$, $\sum d_i = k + \epsilon\}$, which must contain the actual $(d_1, \cdots, d_k)$ corresponding to η, since $q'(0) = \sum(d_i - 1)$. The maximum of $\sum(d_i - 1)^2$ on B, taken on at an extreme point, is $k - 1 + (k + \epsilon - 1)^2$.

Consider now the second term in the expression (2.3.10) for L. Let $h(u) = (u - 1)^2/(1 - r + ru)^2$. Then $d^2h(u)/du^2 \geqq 0$ if $0 \leqq u \leqq 1 + 1/2r$. Since $d_i \leqq k + \epsilon \leqq k + 1$ on B, we conclude that, if $r = 1/2k$, this second expression of (2.3.10) has its maximum over B at an extreme point, where this maximum is easily estimated to be less than k^2.

We conclude that, if $r = 1/2k$, we have $L \geqq -k(k+1)$. Hence, $q(\alpha) \geqq \epsilon\alpha - k(k+1)\alpha^2/2$ for $0 \leqq \alpha \leqq 1/2k$. Putting $\alpha = \epsilon^2/2k(k+1)$ and $\xi' = \alpha\eta + (1-\alpha)\xi$ for this α, we obtain

$$\log[\det M(\xi')/\det M(\xi)] \geqq \epsilon^2/2k(k+1), \tag{2.3.11}$$

where again the inequality is strict unless $\epsilon = 0$. Summarizing our results, we have

THEOREM 2.3.4. *For any* ξ,

$$\frac{\det M(\xi)}{\max_{\xi'} \det M(\xi')} \geqq \exp\,[k - \bar{d}(\xi)] \tag{2.3.12}$$

and, if $\bar{d}(\xi) - k \leqq 1$,

$$\frac{\det M(\xi)}{\max_{\xi'} \det M(\xi')} \leqq \exp\left\{-\frac{[\bar{d}(\xi) - k]^2}{2k(k+1)}\right\}, \tag{2.3.13}$$

with strict inequality unless ξ *is D-optimum.*

In the unlikely situation that one has computed a lower bound on $\det M(\xi)/\Delta$ and wants to obtain an upper bound on $\bar{d}(\xi)$, one would invert (2.3.13). Thus, if $\Delta/\det M(\xi) \leqq \exp\,(1/2k[k+1])$, we obtain, from (2.3.11),

$$\bar{d}(\xi) \leqq k + \left[2k(k+1)\log\frac{\Delta}{\det M(\xi)}\right]^{1/2}. \tag{2.3.14}$$

Extension to the vector case. Various extensions of our theory are possible, to models where Y_x is a vector. For example, if Y_x is an h-vector of components Y_{xi} and θ and f are hr-vectors ($k = hr$) of components θ_{ij} and f_{ij} with $E_\theta Y_{xi} = \sum_j \theta_{ij} f_{ij}(x)$, the Y_x being uncorrelated with $E_\theta Y_x Y_x' = \sigma^2 I$ (other cases are easily reduced to this form), then the information matrix of any ξ breaks up into r diagonal $h \times h$ blocks $M_j(\xi)$, and the previously discussed theory goes over into this setting with $d(x, \xi)$ becoming $\sum_j f^{(j)}(x)' M_j^{-1}(\xi) f^{(j)}(x)$ where $f^{(j)}$ is the h-vector of f_{ij}. Other vector models can also be treated.

The number of points supporting a design. In general, if H is the dimension of the linear space spanned by the functions $f_i f_j$, $i \leqq j$, then for any ξ there is a ξ' supported by no more than $H + 1$ points of $\mathfrak{X}$ and for which $M(\xi') = M(\xi)$. Often more can be said. For example, if H is the maximum possible $[k(k+1)/2]$, then there is a D-optimum design on H points, since the optimum design yields a boundary point in the convex body of $M(\xi)$. Various results for the number of points needed when $s < k$ can be found in [11], [6], [19], [16], [23]. In certain cases, improvements are possible because of the nature of the f_i. For example in the case of polynomial regression of degree $m = k - 1$ on a real interval, we have $H = 2m$, but for any ξ there is a ξ' on at most $m + 1$ points with $M(\xi') = M(\xi)$. The admissible ξ were characterized in this example by the author in [16]. An outstanding problem is the characterization, in terms of $\mathfrak{X}$ and the f_i, of the smallest number q such that, for any ξ, there is a ξ' supported by at most q points, and with $M(\xi') = M(\xi)$.

We cite an example to indicate the theoretical importance of the above considerations (the practical importance being obvious). The author [17] has computed D-optimum ξ for the problem of quadratic regression on an m-cube when $m \leqq 5$, these optimum designs being supported by the vertices, midpoints of edges, and midpoints of two-dimensional faces. When $m = 6$, a design with such support can not be optimum. Already for $m = 4$ and $m = 5$ the optimum designs of this structure are supported by more than $H + 1$ points, which, incidentally, exemplifies the fact that the optimum ξ is no longer unique as it is in the case $m = 1$. It is reasonable to suppose that, if one could obtain D-optimum designs supported by fewer points when $m = 4$ or 5, this might suggest a structure which would also work when $m \geqq 6$.

Note added in proof. Such designs for all m have been obtained recently by Dr. R. H. Farrell and the author. Details will appear elsewhere.

Another question is this: Suppose we restrict our attention, as has often been done in the literature (see, for example, [4]) to designs supported by k points. The best [for example, in terms of det $M(\xi)$] design of this structure need not be optimum among all designs, but by how much does it miss? Using the concavity of log det $M(\xi)$, one can develop bounds of the sort developed earlier in this section, but it would be useful if these could be improved by using the form of the f_i.

The results we have cited on the number of points needed to support designs do not apply to the exact theory as has sometimes, especially in the polynomial case, been assumed. However, the error introduced by assuming these results to apply is only of order $1/N$, as indicated in subsection 2.1.

3. Applications

3.1. *Systematic Designs.* Systematic designs arise in contexts where observations are taken over time. Thus, if at time $t = 1, 2, \cdots, T$, one or more observations can be taken at points in a space S, we can view $\mathfrak{X}$ as S^T; however, there will usually be restrictions on the number of observations which can be taken at each time point t. Thus, although it can happen that an approximate design, obtained without imposing any restrictions, either turns out to satisfy these practical restrictions, or is close to a design which satisfies them, this will usually not be the case, and the theory of subsection 2.3 is thus not always useful. What would often be useful here is a theory which is approximate with respect to S but exact with respect to T.

There are many recent papers in this area, some of which are those of Williams [25], Cox [8], [9], [10], Box [1], Box and Hay [2], and Patterson [21]; further references can be found in these papers.

In the simplest of these problems, the model is such that the dependence on the time variable can be treated as the block effect in a *BIBD* or as the row or column effect in a Youden square. The theory of subsection 2.2 can then be used to give optimum designs.

In more complex models, this device will not suffice. As mentioned in subsection 2.2, it is often erroneously assumed in such problems that an optimum design is to be found among those designs for which the b.l.e. of treatment effects are orthogonal to the b.l.e. of time effects (see, for example, [1], [2]). A simple counterexample is the following: Suppose that S is the interval $[-1, 1]$, that $T = 4$, and that the expected value of an observation at the point z in S at time t is $\alpha z + \beta_0 + \beta_1 t + \beta_2 t^2$, where α and the β_i are unknown. All observations have the same variance σ^2 and are independent. We wish to estimate α.

For simplicity of computation in this counterexample, we suppose that we are allowed one observation at each of the four times t; thus, a design is a quadruple (z_1, z_2, z_3, z_4), where z_t is the point at which an observation is to be taken at time t. One sees easily that the restriction to designs for which the b.l.e. of α is orthogonal to those of the β_i means that the design is of the form $(z_1, -3z_1, 3z_1, -z_1)$, and the variance of the b.l.e. of α is obviously a minimum among designs of this form, namely $9\sigma^2/20$, when $z_1 = \pm 1/3$. On the other hand, the design $(-1, 1, -1, 1)$, for which we do not have this orthogonality, yields a b.l.e. of α with variance $5\sigma^2/16$. The variance for the best orthogonal design is thus 44% larger!

For settings where S is higher-dimensional, the computation of optimum designs will usually be very tedious. The neat approach of Box [1], which yields designs with the above computationally useful orthogonality property, will not yield optimum designs, both because of the phenomenon illustrated in the previous paragraph, and also because the method of construction used in [1] is not geared to the specification of a fixed S with respect to which all designs are to be compared. For example, if S is replaced by the unit disc in the example of the previous paragraph with $T \geqq 5$, the method of [1] yields as a design a T-tuple $(z_1, z_2, \cdots, z_T)$, where $z_t = (z_{t1}, z_{t2})$ is a two-vector and the T-vectors $(z_{11}, z_{21}, \cdots, z_{T1})$ and $(z_{12}, z_{22}, \cdots, z_{T2})$ are chosen so as to yield the orthogonality property. But the method does not insure that $z_{t1}^2 + z_{t2}^2 \leqq 1$ for all t; thus, to use such a design for the S we have specified, it would be necessary to change the original design of [1] by dividing all z_{tj} by the factor $\max_t (z_{t1}^2 + z_{t2}^2)^{1/2}$. Hence, the various choices of $(z_1, \cdots, z_T)$ in [1] which yield the same information matrix there do not yield the same matrix when we scale them down to our specified S. We shall again encounter this need for a careful scaling of the designs, which are usually scaled in a different manner in the literature, in our consideration of rotatable designs in subsection 3.2.

Williams' model. We shall now consider in detail the model studied by Williams [25], wherein the effect of time appears entirely through a correlation among observations. There are k treatments, any one of which can be tested at time t for $t = 1, 2, \cdots, T$. Thus, a design d_T is a T-tuple $(v_1, v_2, \cdots, v_T)$, where v_t is the label number of the treatment tested at time t, $(1 \leqq v_t \leqq k)$. If treatment i is tested at time t, the expected value of the observation Y_t, say, is θ_i.

The first order model. We shall first consider Williams' first order model, wherein $Y_t = \rho Y_{t+1} + \epsilon_t$, the ϵ_t being uncorrelated with common variance σ^2 for $t < 1$. The variance σ_1^2 of Y_1 is assumed positive, but will be shown not otherwise to concern us, since we will be concerned with an asymptotic theory as $T \to \infty$. It is often customary to assume var (Y_1) to be such as to make $Y_t - EY_t$ stationary in the wide sense, but this is unnecessary. For the present, we assume ρ known, $-1 < \rho < 1$; actually, we shall see that it is only necessary to know that $-1 < \rho \leqq 0$ or $0 \leqq \rho < 1$ in order to choose an asymptotically optimum design. Later we shall consider a minimax approach when nothing is known about ρ.

The problem is to estimate all contrasts of the θ_i, and we shall find designs which are asymptotically optimum as $T \to \infty$, in any of the senses (a) through (f) of subsection 2.2. Our considerations are thus asymptotic, but for exact designs.

Williams discussed in detail two types of designs for this problem: A design is of type II(a) if (i) each unordered pair of distinct integers between 1 and k, inclusive, appears the same number of times among the pairs (v_1, v_2), (v_2, v_3), $\cdots$, (v_{T-1}, v_T), and (ii) $v_{t-1} \neq v_t$ for all t. A design is of type II(b) if (i) is satisfied and if each unordered pair of different integers appears twice as often as each pair of equal integers. Examples of these two types of designs when $k = 4$ are (2123423143142) and (11234413224133421). In constructing such designs, it is often convenient, as here, to have $v_{rk+2}, v_{rk+3}, \cdots, v_{(r+1)k+1}$ constitute a permutation of $1, 2, \cdots, k$ for each integer $r \geqq 0$. Williams showed that, between these two types of designs, as $T \to \infty$, II(a) is the better in the sense of criterion (e) of subsection 2.2 if $\rho > 0$, while II(b) is the better if $\rho < 0$. It has often been conjectured (see, for example, Cox [9]) that these designs are optimum among *all* designs in the two respective cases. This conjecture turns out to be true for II(a) but false for II(b) [although II(b) will turn out to have a different optimum property], as we shall now see.

Let $W_t = Y_t - \rho Y_{t-1}$ for $t > 1$. The W_j are uncorrelated with common variance σ^2. Since $\sigma_1^2 > 0$ and $|\rho| < 1$, it is easy to see that, for any design, T^{-1} times the information matrix associated with $Q_2\theta$ (in the notation of subsection 2.2) is asymptotically the same as $T \to \infty$, whether based on $Y_1, Y_2, \cdots, Y_T$ (or, which is equivalent, on $Y_1, W_2, \cdots, W_T$) or on $W_2, \cdots, W_T$, and approaches a positive definite limit for the asymptotically optimum designs obtained below. It follows that we can ignore Y_1 and base our considerations on $W_2, \cdots, W_T$ in the asymptotic considerations which follow.

Writing $W^{(T)\prime} = (W_2, \cdots, W_T)$ and $EW^{(T)} = J_T\theta$, we thus must consider the information matrix $J_T'J_T = A_{d_T}$, say, of the $W^{(T)}$ associated with any design $d_T = (v_1, \cdots, v_T)$ based on T observations. Write $A_d = \|a_{dij}\|$. We see at once that

$$
\begin{aligned}
(3.1.1) \qquad a_{dii} = {} & (\text{number of } r,\ 2 \leqq r \leqq T, \text{ for which } v_r = i) \\
& + \rho^2(\text{number of } r,\ 1 \leqq r \leqq T-1, \text{ for which } v_r = i) \\
& - 2\rho(\text{number of } r,\ 1 \leqq r \leqq T-1, \text{ for which } v_r = v_{r+1} = i)
\end{aligned}
$$

and, for $i \neq j$,

$$
\begin{aligned}
(3.1.2) \qquad a_{dij} = {} & -\rho[\text{number of } r,\ 1 \leqq r \leqq T-1, \text{ for which } (v_r, v_{r+1}) \\
& = (i, j) \text{ or } = (j, i)].
\end{aligned}
$$

For any d based on T observations, write

$$(3.1.3) \qquad \frac{1}{T} QA_dQ' = \begin{Vmatrix} b_d & h_d' \\ h_d & Z_d \end{Vmatrix},$$

where Z_d is $(k-1) \times (k-1)$. The covariance matrix of the b.l.e. of $Q_2\theta$ based on $W^{(T)}$ is $\sigma^2T^{-1}(Z_d - h_db_dh_d')^{-1}$. Since $h_db_dh_d'$ is nonnegative-definite, the discussion of subsection 2.2 shows that d_T^* is optimum in all of the senses (a) through (f) of subsection 2.2, provided that (1) $Z_{d_T^*}$ is a multiple of the identity and $h_{d_T^*} = 0$, and (2) tr Z_{d_T} is maximized by d_T^*. We shall exhibit a design d_T^* for which $Z_{d_T^*}$ approaches a positive definite limit as $T \to \infty$ and for which (1) and (2) are satisfied to within order T^{-1}; these designs are thus asymptotically optimum as $T \to \infty$.

Since Q_2 has row sums equal to zero, (1) is satisfied to within order T^{-1} by any sequence, with T, of designs $\{d_T^*\}$ for which (3) for all T and some finite constant c, independent of T, $|a_{d_Tii} - a_{d_Tjj}| < c$ and $|a_{d_Tij} - a_{d_Tef}| < c$ for all $i \neq j$ and $e \neq f$. Moreover, if we define

$$\text{(3.1.4)} \qquad \pi_{d_T} = T^{-1}(\text{number of } r,\ 1 \leqq r \leqq T-1, \text{ for which } v_r = v_{r+1}),$$

we obtain easily, since $b_{d_T} = k^{-1}(1 - T^{-1})(1 - \rho)^2$ for all d_T,

$$\text{(3.1.5)} \qquad \begin{aligned} \operatorname{tr} Z_{d_T} &= T^{-1} \operatorname{tr} A_{d_T} - b_{d_T} = (1 - T^{-1})(1 + \rho^2) - 2\rho\pi_{d_T} - b_{d_T} \\ &= (1 + \rho^2)(1 - T^{-1})(1 - k^{-1}) - 2\rho[\pi_{d_T} - k^{-1}(1 - T^{-1})]. \end{aligned}$$

Thus, condition (2) is satisfied to within order T^{-1} by any sequence of designs d_T^* for which (4a) $T\pi_{d_T^*} < c$ for all T, if $\rho \geqq 0$, or (4b) $T(1 - \pi_{d_T^*}) < c$ for all T, if $\rho \leqq 0$. Noting that our omission of Y_1 changed Z_{d_T} by order T^{-1}, we conclude:

THEOREM 3.1.1. *A sequence $\{d_T^*\}$ of designs is asymptotically optimum to within relative error T^{-1} as $T \to \infty$ in Williams' first order model provided $\{d_T^*\}$ satisfies condition* (3) *of the previous paragraph and also condition* 4(a) (*respectively* 4(b)) *if $\rho \geqq 0$* (*respectively, $\rho \leqq 0$*).

The statement of necessary and sufficient conditions for asymptotic optimality without having the error term of order T^{-1}, is obvious. Theorem 3.1.3, for the second order case, will be stated in that form.

Thus, when $\rho > 0$, Williams' designs of type II(a), or ones of approximately this structure (which exist for *any* T), are asymptotically optimum. *However, when $\rho < 0$, Williams' designs of type II(b), although better than those of type II(a), are not asymptotically optimum.* Rather, an approximately optimum design is now one which observes approximately T/k treatments of type 1, then T/k of type 2, et cetera (or which is almost of this structure): $d_T^* = (1, 1, \cdots, 1, 2, 2, \cdots, 2, \cdots, k, k, \cdots, k)$. For a design of type II(b), $\pi_{d_T} \to 1/k$ as $T \to \infty$; thus, by (3.1.5), the "relative efficiency" of such designs for $\rho < 0$, compared with optimum ones (in terms of the variance of the b.l.e. of any contrast) as $T \to \infty$, is $1 - 2(-\rho)/(1 - \rho)^2$.

It is clear that, for fixed T with the appropriate divisibility property, designs of the above forms will be *exactly* optimum if σ_1^2/σ^2 is sufficiently large. On the other hand, if σ_1^2/σ^2 is sufficiently small, an exactly optimum design will have $v_j \neq v_1$ for all $j > 1$. It would be interesting to delimit the sets of values of σ_1^2/σ^2 wherein these extreme designs and others between them are exactly optimum.

Of course, all designs for which the treatments appear with (approximately) equal frequency are asymptotically optimum when $\rho = 0$.

Suppose ρ were not known exactly, but that we only knew in advance of the experiment that $\rho \geqq 0$ (or, similarly, that $\rho \leqq 0$). Then ρ can be estimated consistently from the data (see [24] for details) and one finds that the same limiting formulas hold as before (of course, estimators are no longer linear). Thus, theorem 3.1.1 specifies asymptotically optimum designs for this situation, too.

Next, suppose we know in advance of the experiment only that $-1 < \rho < 1$. We seek a design which is minimax over ρ as $T \to \infty$, with respect to any of the criteria (a) through (f) of subsection 2.2. Since ρ can be estimated as indicated in the previous paragraph, the problem thus reduces to choosing d_T so as to maximize $\min_\rho \operatorname{tr} Z_{d_T}$ as $T \to \infty$. The result is clear:

Theorem 3.1.2. *If ρ is unknown, $\{d_T^*\}$ is asymptotically minimax over ρ for criteria (a) through (f) if and only if, for $i \neq j$ and $e \neq f$,*

$$\lim_{T\to\infty} T^{-1}(a_{d_T^{*\prime} ii} - a_{d_T^* jj}) = 0,$$

$$\lim_{T\to\infty} T^{-1}(a_{d_T^* ij} - a_{d_T^* ef}) = 0, \tag{3.1.6}$$

$$\lim_{T\to\infty} \pi_{d_T^*} = 1/k.$$

Williams' type II(b) designs satisfy these conditions.

Thus, the type II(b) designs do possess an optimum property, but not the one they are usually thought to possess.

It is easy to give sequential designs which improve upon those specified in theorem 3.1.2: It is only necessary, in standard fashion, to decide after $n(T)$ observations, where $n(T) \to \infty$ and $T^{-1}n(T) \to 0$ as $T \to \infty$, whether $\rho \geqq 0$ or $\rho < 0$, and to use the appropriate design of theorem 3.1.1 for the remaining $T - n(T)$ observations, as dictated by this decision. As $T \to \infty$, the resulting designs yield the same limiting information matrix as would have been obtained if ρ had been known.

The second order model. Williams' second order model differs from the first order model only in that we now assume $Y_t + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} = \epsilon_t$, $t > 2$, where the ϵ_t are again uncorrelated with common variance σ^2. The joint distribution of Y_1 and Y_2 is assumed to be nonsingular, and it is assumed that $-1 < \rho_2 < 1$ and $-(1 + \rho_2) < \rho_1 < (1 + \rho_2)$. The model for EY_t, and the notation for a design, are as before.

Williams considers in this setting designs of type III, which satisfy the conditions of a design of type II(a) and also the condition that $v_{t-2} \neq v_t$ for all t and that each unordered pair of integers between 1 and k, inclusive, appears the same number of times among the pairs (v_{t-2}, v_t), $3 \leqq t \leqq T$. Thus, $k \geqq 3$ for such a design to exist. An example of such a design for $k = 4$ is $d = (2, 4, 1, 2, 3, 4, 2, 1, 3, 4, 1, 3, 2, 4)$. There is no corresponding analogue of type II(b) designs here.

We now define π_{d_T} as before, and also

$$\gamma_{d_T} = T^{-1}(\text{number of } r,\ 1 \leqq r \leqq T - 2, \text{ for which } v_r = v_{r+2}). \tag{3.1.7}$$

We now replace our analysis in terms of the W_t by one in terms of $V_t = Y_t + \rho_1 Y_{t-1} + \rho_2 Y_{t-2}$ for $3 \leqq t \leqq T$. A simple computation now shows that

$$\text{(3.1.8)} \qquad \operatorname{tr} Z_{d_T} = T^{-1} \operatorname{tr} A_{d_T} - b_{d_T} = 1 + \rho_1^2 + \rho_2^2 + 2\{\rho_1(1 + \rho_2)\pi_{d_T} + \rho_2\gamma_{d_T}\} - k^{-1}[1 + \rho_1 + \rho_2]^2 + O(T^{-1}).$$

As before, we will obtain designs which are approximately symmetrical with respect to treatments; that is, d_T will be such as to make the $a_{d_T ii}$ almost equal and the $a_{d_T ij}$ for $i \neq j$ almost equal, while maximizing $\operatorname{tr} Z_{d_T}$.

Since $\rho_2 + 1 > 0$, the maximization of (3.1.8) breaks up into four cases according to the signs of ρ_1 and ρ_2. If ρ_1 and ρ_2 are both positive, optimality necessitates $\pi_{d_T} \to 1$ and $\gamma_{d_T} \to 1$ as $T \to \infty$; for example, one can take approximately T/k consecutive observations on treatment 1, then the same number on treatment 2, et cetera. If $\rho_1 < 0 < \rho_2$, we want $\gamma_{d_T} \to 1$ and $\pi_{d_T} \to 0$; this can be achieved while maintaining the approximate symmetry with respect to treatments by taking observations in consecutive blocks of approximately $2T/k(k - 1)$ observations each, each block being of the form $ijij \cdots ij$ for a different pair $i < j$.

If $\rho_1 < 0$ and $\rho_2 < 0$, there are two cases to consider, according to whether $k = 2$ or $k > 2$. If $k > 2$, we can obviously achieve optimality by using an approximately symmetric design with $\pi_{d_T} \to 0$ and $\gamma_{d_T} \to 0$; for example, we can construct a design consisting of $k(k - 1)(k - 2)$ blocks of approximately equal numbers of observations, each of the form $hijhij \cdots hij$ for a different triple of unequal integers h, i, j. On the other hand, when $k = 2$, we cannot achieve both $\pi_{d_T} \to 0$ and $\gamma_{d_T} \to 0$. The set of achievable points $(\pi_{d_T}, \gamma_{d_T})$ has as its limiting set (topologically, not set-theoretically) a set B in the plane which is clearly convex. We must determine the lower left hand boundary B' of B. It is easy to see that, in any approximately symmetric design containing the same number of 1's and 2's (which is all we need consider, since any d_T can be replaced by a d_{2T} with approximately the same π and γ and with the two treatments appearing approximately symmetrically), any appearance of blocks of three or more consecutive 1's can be broken up by exchanging some 1's with 2's from a corresponding block of 2's, without increasing π or γ. Thus, in determining B', it suffices to consider designs which never contain more than two consecutive 1's or 2's. It is now easy to see that B' consists of the line segment $\{2\pi + \gamma = 1, 0 \leqq \gamma \leqq 1\}$. It follows that an asymptotically optimum sequence $\{d_T\}$ satisfies (i) $\pi_{d_T} \to 0$, $\gamma_{d_T} \to 1$ if $\rho_1(1 + \rho_2) < 2\rho_2$, (ii) $\pi_{d_T} \to 1/2$, $\gamma_{d_T} \to 0$ if $\rho_1(1 + \rho_2) > 2\rho_2$, and (iii) $2\pi_{d_T} + \gamma_{d_T} \to 1$ if $\rho_1(1 + \rho_2) = 2\rho_2$. Examples of such designs in the respective cases are (i) $d = (1212 \cdots 12)$, (ii) $d = (11221122 \cdots 1122)$, and (iii) many designs, including the previous two and $d = (112112 \cdots 112; 221221 \cdots 221)$.

When $\rho_1 > 0 > \rho_2$, we must determine the lower right hand boundary B'' of B. This time if there are many occurrences in d of triples of the form 1a1, 1b1, 1c1, et cetera, it is easy to see how to combine them (for example, into 111abc111 here) so as not to increase γ or decrease π. Thus, we need only consider designs which

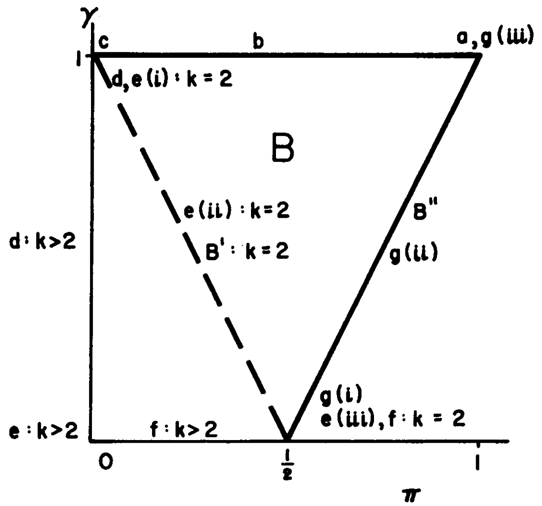

FIGURE 1

contain blocks of at least two i's in a row, every time any i occurs. Symmetrically arranged blocks of exactly $m \geqq 2$ equal integers yield, approximately, $\pi = (m-1)/m$, $\gamma = (m-2)/m$, and it is easy to verify that convex combinations of these points generate B''; thus, $B'' = \{\gamma + 1 = 2\pi, 0 \leqq \gamma \leqq 1\}$. We conclude that an optimum $\{d_T\}$ satisfies (i) $\pi_{d_T} \to 1/2$, $\gamma_{d_T} \to 0$ if $\rho_1(1+\rho_2) < -2\rho_2$, (ii) $\pi_{d_T} \to 1$, $\gamma_{d_T} \to 1$ if $\rho_1(1+\rho_2) > -2\rho_2$, and (iii) $2\pi_{d_T} - \gamma_{d_T} \to 1$ if $\rho_1(1+\rho_2)$

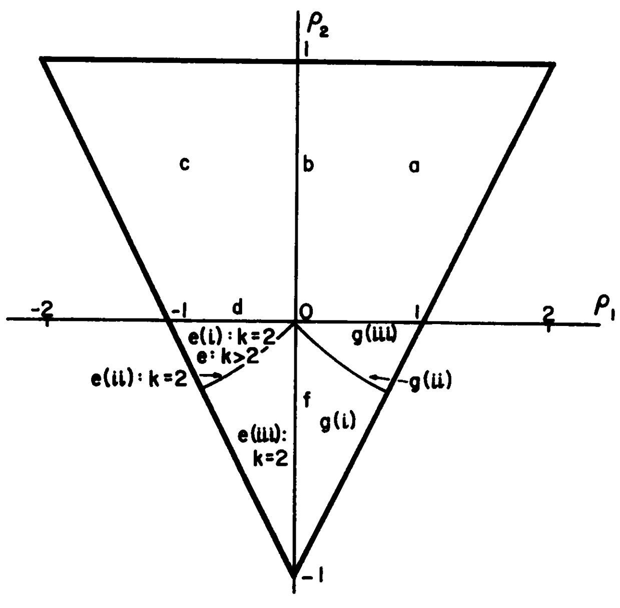

FIGURE 2

$= -2\rho_2$. Examples are (i) the design described above with $m = 2$; for example, $d = (112233113322 \cdots)$ for $k = 3$; (ii) the above form with $m \to \infty$ as $T \to \infty$; and (iii) the above form for any $m \geqq 2$.

We summarize our results in figures 1 and 2 and in the following statement, which traces counterclockwise around the boundary of B, starting with the point (1, 1), and around the (ρ_1, ρ_2) domain starting with the region $\rho_1 > 0$, $\rho_2 \geqq 0$:

THEOREM 3.1.3. *The sequence* $\{d_T\}$ *is asymptotically optimum in senses* (a) *through* (*f*) *for Williams' second order model, if and only if* (1) $T^{-1}A_{d_T}$ *approaches a limit with diagonal elements equal and off-diagonal elements equal, and* (2) $\pi_{d_T} \to \pi$ *and* $\gamma_{d_T} \to \gamma$, *where*

(a) $\pi = \gamma = 1$	*if* $\rho_1 > 0, \rho_2 \geqq 0$;
(b) $\gamma = 1$	*if* $\rho_1 = 0, \rho_2 > 0$;
(c) $\gamma = 1, \pi = 0$	*if* $\rho_1 < 0, \rho_2 > 0$;
(d) $\pi = 0$ (*hence, if* $k = 2, \gamma = 1$)	*if* $\rho_1 < 0, \rho_2 = 0$;
(e) $k > 2$: $\pi = 0, \gamma = 0$	*if* $\rho_1 < 0, \rho_2 < 0$;
$k = 2$: (i) $\pi = 0, \gamma = 1$	*if* $\rho_1(1 + \rho_2) < 2\rho_2 < 0$;
(ii) $2\pi + \gamma = 1$	*if* $\rho_1(1 + \rho_2) = 2\rho_2 < 0$;
(iii) $\pi = 1/2, \gamma = 0$	*if* $2\rho_2 < \rho_1(1 + \rho_2) < 0$;
(f) $\gamma = 0$ (*hence, if* $k = 2, \pi = 1/2$)	*if* $\rho_1 = 0, \rho_2 < 0$;
(g) (i) $\pi = 1/2, \gamma = 0$	*if* $0 < \rho_1(1 + \rho_2) < -2\rho_2$;
(ii) $2\pi - \gamma = 1$	*if* $0 < \rho_1(1 + \rho_2) = -2\rho_2$;
(iii) $\pi = \gamma = 1$	*if* $0 < -2\rho_2 < \rho_1(1 + \rho_2)$;

and where, in cases (*b*), (*d*) *for* $k > 2$, (e) (ii) *for* $k = 2$, (*f*) *for* $k > 2$, *and* (*g*) (*ii*), *the vector* $(\pi_{d_T}, \gamma_{d_T})$ *need not approach a limit, but only the designated line.*

The version of theorem 3.1.3 with error term T^{-1} (as in theorem 3.1.1) is obvious.

We note that Williams' type III designs are thus optimum only in cases (d), (e), (f) for $k > 2$.

When $\rho = 0$, the same designs are asymptotically optimum as in the case of the linear model.

The case where ρ_1 and ρ_2 are not known exactly, but are only known to fall in a specified one of the regions listed in theorem 3.1.3, is treated as in the linear case. So is the case where ρ_1 and ρ_2 are completely unknown, but where a sequential design can be used.

Finally, a nonsequential minimax (over ρ_1 and ρ_2) design is obtained by making the coefficients of terms other than 1, ρ_1^2, and ρ_2^2 in the expression (3.1.8), equal zero. We obtain:

THEOREM 3.1.4. *If* ρ_1 *and* ρ_2 *are unknown,* $\{d_T\}$ *is asymptotically minimax over* ρ_1 *and* ρ_2 *for criteria* (*a*) *through* (*f*) *if and only if it satisfies the first two lines of equation* (3.1.6) *and also*

$$\lim_{T\to\infty} \pi_{d_T} = \lim_{T\to\infty} \gamma_{d_T} = \frac{1}{k} \cdot \tag{3.1.9}$$

Since the point $(1/k, 1/k)$ is an interior point of B, there are many ways of obtaining asymptotically minimax designs. Two examples when $k = 3$ are $d = (112113112113 \cdots ; 221223221 \cdots ; 331332331 \cdots 332)$ and $d = [11 \cdots 122 \cdots 233 \cdots 3$ of length $T/3$ together with $123123 \cdots 123$ of length $2T/3]$.

Higher order dependence. Models with higher order dependence can be investigated in the same fashion.

3.2. *Polynomial regression on an m-ball; rotatable designs.* Let m and d be positive integers. We now treat the problem where $\mathfrak{X}$ is the unit m-dimensional ball, consisting of those points $x = (x^{(1)}, \cdots, x^{(m)})$ in Euclidean m-space for which $\sum_j (x^{(j)})^2 \leqq 1$, and where the f_i are all the functions $\prod_j (x^{(j)})^{r_j}$ for which the r_j are nonnegative integers satisfying $\sum_j r_j \leqq d$; thus, $k = \binom{d+m}{m}$. We refer to this as the problem of dth degree regression on the m-ball. We shall consider the approximate theory, and shall treat the problem of D-optimal estimation of all regression coefficients, which by theorem 2.3.1 is the same as G-optimal estimation of the entire regression function.

The optimum orthogonal-invariant approximate design characterized in theorem 3.2.1 below is not a discrete measure. In theorem 3.2.1 we also give an upper bound on the number of points needed to support a discrete measure which is D-optimum (see subsection 2.3). We shall not be concerned here with the actual construction of these discrete ξ, except for brief mention of the cases $d = 1$ and $d = 2$; these considerations when $d > 2$ and $m > 2$ are considerably more difficult, since it is no longer generally possible to replace uniformly distributed measure on an $(m - 1)$-sphere by a uniform discrete distribution on an appropriate finite set of "uniformly spaced" points as in the case $m = 2$. The construction of exact rotatable designs, which is the subject of much recent literature, is also not our concern here. In fact, such designs may be far from being optimum, as we shall exemplify below in the case $d = 2$, $m = 2$. In fact, our results do *not* imply that there exists an exact rotatable design which is within order N^{-1} of being optimum, but only that there is an exact design which is within order N^{-1} of being rotatable [in the value of its $M(\xi)$] and which is within order N^{-1} of being optimum. The term "rotatable" will be used by us in reference not only to exact, but also to approximate, designs.

The case $m = 1$, where $\mathfrak{X}$ is an interval, was treated in full by Guest [13] and Hoel [14]. Unfortunately neither of their elegant methods is applicable when $m > 1$, and we shall not obtain such explicit results for general d as those of Guest and Hoel.

Rotatable designs were invented by Box and Hunter [3], and their intuitive appeal (which has never been justified until the present paper) has attracted considerable interest and usage. As mentioned in the first part of subsection 3.1, the usual treatment in the literature of design problems where such designs are employed does not make a precise specification of $\mathfrak{X}$. Thus, in [3] and in subsequent papers on the subject, it is standard to employ the normalization $\sum_{i=1}^{N} x_i^{(j)}$

$= 0$, that is, $\int (x^{(j)})d\xi = 0$, and the normalization $\sum_{i=1}^{N} (x_i^{(j)})^2 = N$, that is, $\int (x^{(j)})^2 d\xi = 1$, and then to compare the variance of the estimated regression for various designs at each distance ρ from the origin. Now, the use of two different designs with the same such normalization will usually entail taking observations in balls of quite different radii, that is, $\max_i \sum_j (x_i^{(j)})^2$ will usually not be the same for two such designs. Thus, the comparison of various rotatable designs in this manner is of questionable practical value! If the set of points x at which one is interested in estimating the regression function and at which observations can be taken is actually a ball of radius R, it is meaningless to insist on the above normalization, since it can exclude many good admissible rotatable designs and can include many poor ones, namely, those taking all observations within a ball of radius $< R$, when $mN < R^2$. In fact, a trivial consequence of the behavior of $\mathfrak{X}$ (and the f_i) under multiplication by R/R' is that any design which takes all observations in a ball of radius R' with $R' < R$ yields a larger generalized variance and a larger $\bar{d}(\xi)$ than the corresponding design in the ball of radius R. We shall return to this point in the final paragraph of this subsection.

The reader is warned that certain formulas which are relevant to the material of this section will differ from corresponding formulas of [3] and other papers on the subject, because of the above-mentioned normalization used in these papers (see, for example, the second paragraph above theorem 3.2.2, below).

We shall use, in this section only, a notation which is more convenient in this particular setting than is the general notation employed in the rest of the paper. Instead of using a single subscript, we label the elements of θ and of the vector f, as well as the rows and columns of $M(\xi)$, by m-tuples $(r_1, r_2, \cdots, r_m)$, corresponding to the function $\Pi_j (x^{(j)})^{r_j}$.

Let $\mathcal{G}$ be the orthogonal group on m-space. For g in $\mathcal{G}$, the design $g\xi$ is as usual defined by $(g\xi)(A) = \xi(g^{-1}A)$. Since the problem (set of possible regression functions, et cetera) looks the same with respect to $g\xi$ as with respect to ξ, we clearly have $d(gx, g\xi) = d(x, \xi)$, or $d(gx, \xi) = d(x, g^{-1}\xi)$. Hence, theorem 2.3.3 is applicable and we conclude that there is an optimum ξ^* which is invariant under $\mathcal{G}$; that is, for which $\xi^*(A) = \xi^*(gA)$ for every orthogonal transformation g and Borel set A. Such a ξ^* can be factored into the form $\xi_1^* \times \xi_2^*$:

$$\xi^*(A) = \int_0^1 \xi_1^*(\rho^{-1}[A \cap S_\rho])\xi_2^*(d\rho), \tag{3.2.1}$$

where S_ρ is the $(m - 1)$-sphere of radius ρ, ξ_1^* is the uniform probability measure on S_1, the integrand is taken to be 1 or 0 when $\rho = 0$ according to whether or not $0 \in A$, and ξ_2^* is a probability measure on the interval [0, 1]. Write

$$\mu_s = \int_0^1 \rho^s \xi_2^*(d\rho) \tag{3.2.2}$$

and

$$F(s_1, s_2, \cdots, s_m) = \int_{S_1} \Pi_j (x^{(j)})^{s_j} \xi_2^*(d\rho). \tag{3.2.3}$$

Then the elements of $M(\xi^*)$ for $\xi^* = \xi_1^* \times \xi_2^*$ can be computed as follows:

(3.2.4) $$m_{(\alpha_1,\cdots,\alpha_m),(\beta_1,\cdots,\beta_m)}(\xi^*) = \mu_{\Sigma(\alpha_i+\beta_i)}F(\alpha_1 + \beta_1, \cdots, \alpha_m + \beta_m).$$

In particular, this is zero if $\alpha_i + \beta_i$ is odd for any i, and is positive otherwise, unless $\xi_2^*(0) = 1$.

Now, suppose ξ^* is D-optimum and invariant, as above. Then $d(x, \xi^*)$ is easily seen to be a polynomial in ρ^2, of degree $\leq d$ in ρ^2, say $d^*(\rho^2, \xi^*) = \sum_{j=0}^{d} q\, \rho^{2j}$. From the behavior as $x^{(1)} \to \infty$, we see that $q_d > 0$. Now, $\int q(x, \xi)\xi(dx) = k$ for any ξ (see subsection 2.3), so that, by theorem 2.3.1, $d^*(\rho^2, \xi^*) = k$ on a set of ξ^*-measure one if ξ^* is optimum. It is easy to prove for any invariant ξ^* that $M(\xi^*)$ is singular if ξ_2^* gives measure one to a set consisting of fewer than $(d + 1)/2$ points of the interval $0 \leq \rho \leq 1$, where the origin is counted as 1/2 point (the converse of this for general invariant ξ^* is also simple to prove, but will not be needed here, since it will follow from other considerations that ξ_2^* has exactly $(d + 1)/2$ points of support for a D-optimum invariant ξ^*). Hence, for an optimum ξ^*, the function $d^*(\tau, \xi^*)$ is a polynomial of degree d with $q_d > 0$ and such that $d^*(\tau, \xi^*) = k$ at least $(d + 1)/2$ points.

If d is even, say $d = 2b$, it follows from elementary algebraic considerations that $d^*(\tau, \xi^*)$ is equal to its maximum k on $0 \leq \tau \leq 1$ at $\tau = 0$ and at b other points, of which one is $\tau = 1$, since $d^*(\tau, \xi)$ must be increasing at $\tau = 1$. Similarly, if d is odd, say $d = 2b + 1$, then $d^*(\tau, \xi^*)$ is equal to its maximum k at $b + 1$ points, all different from zero, one of them being 1. In either case, write $\tau_1, \cdots, \tau_{b+1}$ for these values, so that ξ^* is supported by $\{\tau_1^{1/2}, \cdots, \tau_{b+1}^{1/2}\}$.

Thus, an invariant design can be optimum only if ξ_2^* has exactly $(d + 1)/2$ points of support, one of which is at 1.

According to theorem 2.3.1, $M(\xi)$ and $d(x, \xi)$ are the same for all D-optimum ξ, whether or not ξ is invariant. Thus, all D-optimum approximate designs ξ are rotatable, in that $d(x, \xi)$ is a function only of ρ for all of them. Since the function $d(x, \xi)$ is the same for all D-optimum designs, and since $d(x, \xi)$ attains its maximum on a set of ξ-measure one if ξ is D-optimum, we conclude that every D-optimum ξ (invariant or not) gives measure one to the union of the same $b + 1$ $(m - 1)$-spheres (one of which may degenerate to the origin) of radii $\tau_1^{1/2}, \cdots, \tau_{b+1}^{1/2}$.

Any D-optimum ξ can be integrated over $\mathcal{G}$ to yield an invariant D-optimum ξ^* with $\xi_2^*(\rho) = \xi(S_\rho)$ for all ρ. For fixed values of $\tau_1, \cdots, \tau_{b+1}$, the function log det $M(\xi)$ is strictly concave in the positive weights $\gamma_i = \xi_2(\tau_i^{1/2})$ among invariant designs, so that the optimum weights, as well as the τ_i, are unique. For any invariant designs ξ' for which ξ_2' is supported by $(d + 1)/2$ points $\lambda_1^{1/2}, \cdots, \lambda_{b+1}^{1/2}$ with $\lambda_{b+1} = 1$ and with $\lambda_1 = 0$ if d is even, write $d^*(\tau, \xi') = d(\tau; \gamma_1, \cdots, \gamma_{b+1}, \lambda_1, \cdots, \lambda_{b+1}) = d(\tau; \gamma, \lambda)$ to exhibit the support points and weights. Consider the equations

$$d(\lambda_j; \gamma, \lambda) = k;$$

$$\frac{\partial}{\partial\tau} d(\tau;\gamma,\lambda)|_{\tau=\lambda_j} = 0, \qquad \begin{cases} 1 \leqq j \leqq b \text{ if } d \text{ is odd,} \\ 2 \leqq j \leqq b \text{ if } d \text{ is even;} \end{cases} \tag{3.2.5}$$

$$\lambda_{b+1} = 1;$$

$$\lambda_1 = 0 \qquad \text{if } d \text{ is even.}$$

The first line contains only b independent equations. Clearly, the unique optimum choices of γ and λ satisfy these equations. Conversely, any positive probability vector γ and any λ, with $0 \leqq \lambda_i \leqq 1$, satisfying these equations clearly define an invariant ξ^* for which $\bar{d}(\xi^*) = k$, so that this ξ^* is D-optimum.

The integer H of Section 4 is the number of functions of the form $\Pi_j(x^{(j)})^{s_j}$ with $s_j \geqq 0$ and $1 \leqq \sum_j s_j \leqq 2d$, which is just $\binom{2d+m}{m} - 1$.

We summarize our results.

THEOREM 3.2.1. *For the problem of dth degree regression on the unit m-ball, there are numbers* $\lambda_1 < \lambda_2 < \cdots < \lambda_{b+1} = 1$ *with* $b =$ (*greatest integer* $\leqq d/2$) *and* $\lambda_1 = 0$ (*respectively*, > 0) *if d is even* (*respectively, odd*), *and positive numbers* $\gamma_1, \cdots, \gamma_{b+1}$ *whose sum is unity, such that* ξ *is D-optimum if and only if it satisfies*

(a) $\xi(S_{\lambda_j^{1/2}}) = \gamma_j, \qquad 1 \leqq j \leqq b + 1,$

(b) ξ *is rotatable, that is,* $d(x, \xi)$ *depends only on* $\rho^2 = \sum_{j=1}^m (x^{(j)})^2$.

In particular, there is a unique orthogonal-invariant design satisfying these conditions, and there are designs supported by at most $\binom{2d+m}{m}$ *points which satisfy these conditions.*

The γ_j *and* λ_j *can be obtained as the unique solution of* (3.2.5) *satisfying* $\gamma_j > 0$, $\sum\gamma_j = 1$, $0 \leqq \lambda_j \leqq 1$.

It should be noted that the common practice of combining nonrotatable designs on spheres in such a way as to yield a rotatable design will often lead to a design on more than $(d + 1)/2$ spheres, which thus cannot be optimum.

We now consider some examples of optimum designs.

When $d = 1$, we have the trivial result that $\xi(S_1) = 1$ if ξ is optimum. Examples of discrete ξ's which are D-optimum are the uniform distribution on the $m + 1$ vertices of an inscribed regular simplex or on the vertices of any other inscribed regular polygon.

When $d = 2$, we need only consider rotatable designs which assign measure δ, say, to S_1 and measure $1 - \delta$ to the origin. Using formula (3.2.4) (or, renormalizing formula (49) of [3] in the manner discussed earlier, using $V(\delta^{-1/2}\rho)$ where V is given by that formula), one can without difficulty write out the function $d^*(\tau, \xi)$ for such a ξ and use its convexity in τ (see the next paragraph) to compute the desired result. However, it is even unnecessary to do this in order to obtain the result, since a simpler argument suffices to compute $d^*(0, \xi)$ and $d^*(1, \xi)$ in this example, as we shall now see.

The average U of "observations" at the origin is an unbiased estimator of

$\theta_{(0,0,\cdots,0)}$ and yields no information about any other parameters. Moreover, U is the b.l.e. of $\theta_{(0,\cdots,0)}$, since otherwise it is easy to see that $M(\xi')$ would be nonsingular, where ξ' is the uniform measure on S_1, and this last is false by the argument leading up to theorem 3.2.1. We conclude that U is the b.l.e. of the regression function at the origin. Hence, $d(0, \xi) = (1 - \delta)^{-1}$. Now, if $d(0, \xi) = k$, we automatically obtain $d(1, \xi) = k$, since $\int d(x, \xi)\xi(dx) = k$ for all ξ; by the convexity of the second degree polynomial $d(\tau, \xi)$ with positive coefficient of τ^2, we will then have $\bar{d}(\xi) = k$, so that ξ will be D-optimum. Thus, we obtain

THEOREM 3.2.2. *When $d = 1$, ξ is D- and G-optimum if and only if it is rotatable and $\xi(S_1) = 1$. When $d = 2$, ξ is optimum if and only if it is rotatable and $\xi(0) = 1 - \xi(S_1) = 2/(m + 1)(m + 2)$.*

The authors of [3] say on page 215 that they do not claim any optimum properties for their designs, but we now see that some, but not all, of their designs are indeed D-optimum. Thus, when $m = 2$ and $d = 2$, their design which takes one observation at the origin and one at each of five equally spaced points on the unit circle (or any multiple of this design) is D-optimum. The reader is invited to consult table 1 of [3] to see how the considerations with the normalization used there are, as we have mentioned earlier, misleading for the problem we have considered unless one is careful about translating their meaning. For example, "$\rho = 1$" there does not mean "$\rho = 1$" for us, but rather "$\rho = \delta^{1/2}$" for designs of the structure considered above; thus, values of V for a given value of ρ there actually correspond to different points of $\mathfrak{X}$ for different designs, and one must look at the behavior of various designs on different domains of ρ there, in order to obtain their behavior on the same domain $\mathfrak{X}$ in our problem.

3.3. *Linear regression on a Euclidean subset.* Let $\mathfrak{X}$ be a compact subset of Euclidean m-space, which we assume not to lie in an $(m - 1)$-dimensional hyperplane (if it did, θ would not be estimable for any ξ). Writing $x = (x^{(1)}, \cdots, x^{(m)})$, we consider the problem of estimating all of θ, or the whole regression function, when $k = m + 1$ and $f_i(x) = x^{(i)}$ for $1 \leqq i \leqq m$ and $f_{m+1}(x) = 1$.

For any ξ for which θ is estimable, $d(x, \xi)$ is quadratic in x and is in fact strictly convex. Hence, if $\mathfrak{X}'$ is the convex closure of $\mathfrak{X}$, the function $d(x, \xi)$ achieves its maximum on $\mathfrak{X}'$ only on a set B of extreme points of $\mathfrak{X}'$, which are clearly points of $\mathfrak{X}$. Thus, $d(x, \xi)$ attains its maximum on $\mathfrak{X}$ precisely on the set B, and thus, as in previous examples, $\xi^*(B) = 1$ if ξ^* is D-optimum. Since $d(x, \xi^*)$ is quadratic and is equal to k on B if ξ^* is optimum, there must therefore exist an ellipsoid T (by which we mean the hypersurface, not the solid) which can be circumscribed about $\mathfrak{X}$ in such a way that $T \cap \mathfrak{X} = B$ and $d(x, \xi) = m + 1$ for x in T. Thus, $d(x, \xi^*) = (x - c)'C(x - c) + c_0$ for a suitable vector c, positive real c_0, and positive definite symmetric matrix C, T being the set where $(x - c)'C(x - c) = m + 1 - c_0$,

If A is an $m \times m$ matrix such that $A'A = (m + 1 - c_0)C$, the affine transformation $y = A(x - c)$ takes T onto the unit sphere $S = \{y|y'y = 1\}$. Under this transformation, ξ^* is transformed into a measure ξ' on S. Since $(x^{(1)}, \cdots,$

$x^{(m)}, 1)M^{-1}(\xi^*)(x^{(1)}, \cdots, x^{(m)}, 1)' = (x - c)'C(x - c) + c_0$, we conclude that

$$M^{-1}(\xi^*) = \begin{Vmatrix} C & -Cc \\ -c'C & c'Cc + c_0 \end{Vmatrix}, \tag{3.3.1}$$

and hence that

$$M(\xi^*) = \begin{Vmatrix} C^{-1} + c_0^{-1}cc' & c_0^{-1}c \\ c_0^{-1}c' & c_0^{-1} \end{Vmatrix}. \tag{3.3.2}$$

Hence,

$$\begin{aligned} C^{-1} + c_0^{-1}cc' &= \int xx'\xi^*(dx) = \int (A^{-1}y + c)(A^{-1}y + c)'\xi'(dy), \\ c_0^{-1}c &= \int x\xi^*(dx) = \int (A^{-1}y + c)\xi'(dy), \\ c_0^{-1} &= \int 1\xi^*(dx) = 1, \end{aligned} \tag{3.3.3}$$

from which we conclude at once that

$$\begin{aligned} \int y\xi'(dy) &= 0, \\ \int yy'\xi'(dy) &= m^{-1}I. \end{aligned} \tag{3.3.4}$$

Since $\sum_i \int x_i^2\xi(dx) = 1$ for any ξ on S, we easily compute that $H = (m^2 + 3m - 2)/2$.

We summarize our results:

THEOREM 3.3.1. *For the problem of linear regression on a compact subset* $\mathfrak{X}$ *of Euclidean m-space which does not lie in an* $(m - 1)$*-dimensional hyperplane, there is an affine transformation* t *of* $\mathfrak{X}$ *into the unit m-ball such that the intersection of* $t\mathfrak{X}$ *with the unit sphere supports a probability measure* ξ' *for which*

$$\begin{aligned} \int y_i\xi'(dy) &= 0, \qquad 1 \leqq i \leqq k, \\ \int y_iy_j\xi'(dy) &= m^{-1}\delta_{ij}, \qquad 1 \leqq i, j \leqq k; \end{aligned} \tag{3.3.5}$$

a D- and G-optimum ξ^* *is then defined by*

$$\xi^*(W) = \xi'(tW). \tag{3.3.6}$$

Conversely, each optimum ξ^* *can be obtained in this way from such a* ξ'*, for some* t*. Whenever such a* ξ' *exists for a given* t*, there exists such a* ξ' *supported by* $m(m + 3)/2$ *or fewer points.*

Of course, the optimum ξ^* need not be unique.

The ellipsoid T which was circumscribed about $\mathfrak{X}$ (or $\mathfrak{X}'$) above, and which is "closest" to it in the sense appropriate for our considerations, does not seem to have been considered in the literature on circumscribing figures about convex bodies. The explicit determination of T or t for a given $\mathfrak{X}$ or $\mathfrak{X}'$ poses an interesting geometric problem. In the case $m = 2$, using complex notation, ξ' must be such that

$$\int z\xi'(dz) = 0, \qquad \int z^2\xi'(dz) = 0. \tag{3.3.7}$$

If the subset $\mathfrak{X}$ of the unit circle supports such a measure, then a subset of $t\mathfrak{X}$ consisting of 5 or fewer points will also support such a measure. Fewer points may suffice, as when $\mathfrak{X}$ is a triangle, square, or circle; if $\mathfrak{X}$ is a regular pentagon, five points are needed.

REFERENCES

[1] G. E. P. Box, "Multi-factor designs of first order," *Biometrika*, Vol. 39 (1952), pp. 49–57.

[2] G. E. P. Box and W. A. Hay, "A statistical design for the efficient removal of trends occurring in a comparative experiment with an application in biological assay," *Biometrics*, Vol. 9 (1953), pp. 304–319.

[3] G. E. P. Box and J. S. Hunter, "Multi-factor experimental designs for exploring response surfaces," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 195–241.

[4] G. E. P. Box and H. L. Lucas, "Design of experiments in nonlinear situations," *Biometrika*, Vol. 46 (1959), pp. 77–90.

[5] G. E. P. Box and K. B. Wilson, "On the experimental attainment of optimum conditions," *J. Roy. Statist. Soc., Ser. B*, Vol. 13 (1951), pp. 1–45.

[6] H. Chernoff, "Locally optimum designs for estimating parameters," *Ann. Math. Statist.*, Vol. 24 (1953), pp. 586–602.

[7] ———, "Sequential design of experiments," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 755–770.

[8] D. R. Cox, "Some systematic experimental designs," *Biometrika*, Vol. 38 (1951), pp. 312–323.

[9] ———, "Some recent work on systematic experimental designs," *J. Roy. Statist. Soc., Ser. B*, Vol. 14 (1952), pp. 211–219.

[10] ———, "The design of an experiment in which certain treatment arrangements are inadmissible," *Biometrika*, Vol. 41 (1954), pp. 287–295.

[11] G. Elfving, "Optimum allocation in linear regression theory," *Ann. Math. Statist.*, Vol. 23 (1952), pp. 225–262.

[12] ———, "Design of linear experiments," *Probability and Statistics* (Cramér volume), New York, Wiley, 1959, pp. 58–74.

[13] P. G. Guest, "The spacing of observations in polynomial regression," *Ann. Math. Statist.*, Vol. 29 (1958), pp. 294–299.

[14] P. G. Hoel, "Efficiency problems in polynomial regression," *Ann. Math. Statist.*, Vol. 29 (1958), pp. 1134–1145.

[15] J. Kiefer, "On the nonrandomized optimality and randomized nonoptimality of symmetrical designs," *Ann. Math. Statist.*, Vol. 29 (1958), pp. 675–699.

[16] ———, "Optimum experimental designs," *J. Roy. Statist. Soc., Ser. B*, Vol. 21 (1959), pp. 273–319.

[17] ———, "Optimum designs in regression problems, II," to appear in *Ann. Math. Statist.*

[18] ———, "Invariance, sequential estimation, and continuous time processes," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 573–601.

[19] J. Kiefer and J. Wolfowitz, "Optimum designs in regression problems," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 271–294.

[20] ———, "The equivalence of two extremum problems," *Canad. J. Math.*, Vol. 12 (1960), pp. 363–366.

[21] H. D. PATTERSON, "The construction of balanced designs for experiments involving sequences of treatments," *Biometrika*, Vol. 39 (1952), pp. 32–48.
[22] H. SCHEFFÉ, "A method for judging all contrasts in the analysis of variance," *Biometrika*, Vol. 40 (1953), pp. 87–104.
[23] M. STONE, "Application of a measure of information to the design and comparison of regression experiments," *Ann. Math. Statist.*, Vol. 30 (1959), pp. 55–70.
[24] A. WALD, "Asymptotic minimax solutions of sequential point estimation problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 1–12.
[25] R. M. WILLIAMS, "Experimental designs for serially correlated observations," *Biometrika*, Vol. 39 (1952), pp. 151–167.