The Annals of Statistics
1974, Vol. 2, No. 5, 849–879

GENERAL EQUIVALENCE THEORY FOR OPTIMUM DESIGNS (APPROXIMATE THEORY)[1]

BY J. KIEFER

Cornell University

For general optimality criteria Φ, criteria equivalent to Φ-optimality are obtained under various conditions on Φ. Such equivalent criteria are useful for analytic or machine computation of Φ-optimum designs. The theory includes that previously developed in the case of D-optimality (Kiefer–Wolfowitz) and L-optimality (Karlin–Studden–Fedorov), as well as E-optimality and criteria arising in response surface fitting and minimax extrapolation. Multiresponse settings and models with variable covariance and cost structure are included. Methods for verifying the conditions required on Φ, and for computing the equivalent criteria, are illustrated.

1. Introduction. Let $f' = (f_1, f_2, \cdots, f_k)$ where the f_i are continuous real functions on a compact set $\mathscr{X}$. The expected value of an observation "at the level x in $\mathscr{X}$" is $\sum_1^k \theta_i f_i(x) = \theta' f(x)$. Observations are uncorrelated and have variance independent of x (an assumption relaxed in Section 5). We are concerned with the approximate design theory wherein the designs are a class Ξ of probability measures on $\mathscr{X}$ including all discrete measures, and the information matrix of a design ξ is $M(\xi) = \int_{\mathscr{X}} f(x)f(x)'\xi(dx)$. This has the usual meaning that $M^{-1}(\xi)$ is proportional to the covariance matrix of best linear estimators of θ (with the obvious analogue if M is singular). See Kiefer [17] or Fedorov [12] for further remarks on interpretation. We let $\mathscr{M} = \{M(\xi) : \xi \in \Xi\}$.

Let Φ be a function which is real or $+\infty$ on $\mathscr{M}$. One problem of optimum design theory is the characterization of designs ξ^* which are Φ-*optimum*; that is, for which

$$\Phi(M(\xi^*)) = \min_{\xi \in \Xi} \Phi(M(\xi)) . \tag{1.1}$$

The most common examples of optimality criteria are

$$\Phi_0(M) = \det M^{-1} \quad (D\text{-optimality}) , \tag{1.2}$$

$$\Phi_{1,C}(M) = \operatorname{tr} CM^{-1} \quad (L\text{-optimality; } A\text{-optimality if } C = I) , \tag{1.3}$$

$$\Phi_\infty(M) = \text{maximum eigenvalue of } M^{-1} \quad (E\text{-optimality}) ; \tag{1.4}$$

here C is a given nonnegative definite symmetric matrix, and (1.2) and (1.4) are to be regarded as infinite if M is singular (with the obvious analogue for (1.3)). The significance of the subscripts will be seen in Sections 4C—D. (We shall consider other criteria, later.)

Received January 1973; revised July 1973.

[1] Research supported by NSF Grant GP24438.

AMS 1970 *subject classification.* 62K05.

Key words and phrases. Optimum experimental designs, equivalence theory of designs, D-optimality, A-optimality, E-optimality, iterative design optimization, large eigenvalues.

The desired characterization just mentioned should aid in the computation of Φ-optimum designs. Thus, writing $\bar{d}_0(\xi) = \max_{x \in \mathscr{X}} f(x)'M^{-1}(\xi)f(x)$, Kiefer and Wolfowitz [23] showed in the case (1.2) that $M(\xi^*)$ is the same for all Φ_0-optimum ξ^* and that

$$(1.5) \qquad \xi^* \text{ is } \Phi_0\text{-optimum} \Leftrightarrow \bar{d}_0(\xi^*) = \min_\xi \bar{d}_0(\xi) \Leftrightarrow \bar{d}_0(\xi^*) = k\,.$$

This is also useful because, for a given ξ' which one guesses to be nearly optimum, one cannot usually assess the departure of det $M^{-1}(\xi')$ from the unknown minimum of det $M^{-1}(\xi)$; while the last statement of (1.5) gives both a verifiable condition for optimality and an indication (made precise in Section 6C) that $\bar{d}_0(\xi')$ near k implies det $M^{-1}(\xi')$ near the minimum. This last fact has also been implemented by a number of authors to obtain iterative schemes for computing ξ^*. This will be discussed in Section 6B. Another useful aspect of (1.5), implemented in [18], [19], [9], is that $f(x)'M^{-1}(\xi^*)f(x) = k$ on the support of ξ^*; this and the form of f often enable one to limit drastically the possible supports among which one must search, as indicated in Section 6A.

Subsequently Karlin and Studden [14], Theorems 8.1–8.2, and Fedorov [10], [12], studying what the latter called linear (in M^{-1}) optimality criteria, showed in the case (1.3) that, if $M(\xi^*)$ is nonsingular,

$$(1.6) \qquad \xi^* \text{ is } \Phi\text{-optimum} \Leftrightarrow \bar{d}_1(\xi^*) = \min_\xi \bar{d}_1(\xi)$$
$$\Leftrightarrow \bar{d}_1(\xi^*) = \operatorname{tr} CM^{-1}(\xi^*)\,,$$

where $\bar{d}_1(\xi) = \max_x f(x)'M^{-1}(\xi)CM^{-1}(\xi)f(x)$. The analogy between (1.5) and (1.6) is obvious, and Fedorov's presentation makes it clear that the steps in his proof of (1.6) parallel those of the proof of (1.5).

Since neither Φ_0 nor Φ_1 is a special case of the other, this suggests that there is a larger class of Φ for which one can obtain equivalent characterizations of Φ-optimality analogous to the last two statements of (1.5) and of (1.6). This has undoubtedly occurred to a number of workers in the field, and I have mentioned it in talks and an abstract [21 a]; but perhaps the intuitive appeal and computational tractability of D- and A-optimality have continued to make them the main topics of concentration.

A number of people have asked me for my results on this subject, and the present paper is a selection of some óf the material on general Φ-optimality characterizations which I have collected over the years. Since the completion of a comprehensive monograph (now in progress) seems several years off, it seemed appropriate to publish a collection of material which could be useful to other research workers *now*. While the present paper is long, many details, ramifications, and examples of a type previously published, have been omitted to yield, it is hoped, the best immediate guide the author can offer to aid others in solving optimum design problems. The basic Equivalence Theorem 1 is simple, but applying it can entail analytic labor.

Available literature on verifying convexity of such functions Φ is not too easy to find in applicable form, and therefore some space has been used to list and illustrate useful tools for this purpose. Development of equivalence criteria for Φ's which permit our treatment is the content of the longest section of the paper, Section 4. In addition to Φ_0 and Φ_1, they include the Φ_∞ of (1.4) and certain criteria which arise in response-surface problems in which one purposely fits biased surfaces of a simpler form than $\theta' f(x)$, discussed in [4], [16], [21], and in Section 4F.

For the sake of simplicity, Theorem 1 of the next section presents the basic theory in the case so far described, wherein (i) observations are univariate, (ii) observations have equal variance and cost, (iii) Φ is continuously differentiable at the $M(\xi^*)$ under consideration (which is automatic for some Φ, such as those of (1.2), (1.4) and (1.3) if C is nonsingular, for each of which any optimum $M(\xi^*)$ is nonsingular, implying differentiability). The modifications required to extend the results to other cases are described in Theorem 3 of Section 2, Section 3K, Section 4E, Section 5 and Section 7. A sequel to the present paper will treat illustrations of such modifications in detail.

We conclude this section by recording some additional notation and elementary results of matrix calculus. All matrices considered here have entries in the reals, R^1. We denote by $\mathscr{R}_{k_1,k_2}$ the $k_1 \times k_2$ matrices and by $\mathscr{S}_k$ the symmetric nonnegative definite $k \times k$ matrices. By $\mathscr{M}^+$, $\mathscr{R}^+_{k,k}$, and $\mathscr{S}^+$ we denote the nonsingular members of the corresponding classes without superscript; we write $(\mathscr{M}^+)^- = \{D\colon D^{-1} \in \mathscr{M}^+\}$.

In practice the criterion function Φ will often be defined not merely on the $\mathscr{M}$ of the application at hand, but on $\mathscr{S}_k$ or even an open subset of $\mathscr{R}_{k,k}$. The computations in this paper are typically carried out for a Φ defined on one of these larger domains. Where such a Φ is written in such fashion as to be defined on an open subset of $\mathscr{R}_{k,k}$, care must be taken in using the calculus. Specifically, in this case denote by $\Phi^{[s]} = \Phi \mid \mathscr{S}_k$ (s for "symmetric") the function Φ restricted to $\mathscr{S}_k$. It is usually easier to differentiate Φ with respect to the k^2 variables on $\mathscr{R}_{k,k}$ rather than to differentiate $\Phi^{[s]}$ with respect to the $k(k+1)/2$ variables on the submanifold $\mathscr{S}_k$ which is really of interest, but one must then note that, when the derivatives are meaningful,

$$
\begin{aligned}
(1.7)\qquad \frac{\partial}{\partial m_{ij}}\bigg|\,\mathscr{S}_k &= \left(\frac{\partial}{\partial m_{ij}} + \frac{\partial}{\partial m_{ji}}\right)\bigg|\,\mathscr{R}_{k,k} \qquad &\text{if } i \neq j\,, \\
&= \frac{\partial}{\partial m_{ii}}\bigg|\,\mathscr{R}_{k,k} \qquad &\text{if } i = j\,.
\end{aligned}
$$

Consequently, a first derivative of $\Phi^{[s]}$ is a sum of one or two derivatives of Φ on $\mathscr{R}_{k,k}$, while a second derivative is a sum of one, two, or four.

This is particularly important for checking convexity of $\Phi^{[s]}$ (condition (2.11) below), since convexity of Φ on $\mathscr{R}_{k,k}$, which is easier to check in terms of the Hessian of second derivatives, may not hold although $\Phi^{[s]}$ is convex; this is the case for such a simple Φ as tr AM^{-1}. One must handle such cases by using (1.7);

or, less conveniently (and hence not hereafter used), by making sure that Φ has been chosen as that extension of $\Phi^{[s]}$ to $\mathscr{R}_{k,k}$ which satisfies $\Phi(M) = \Phi(M')$ (e.g., as 2 tr $A(M + M')^{-1}$ in place of tr AM^{-1} noted above) and by restricting consideration to nonnegative definite $M + M'$.

In Section 2 we will have to consider, for $\bar{M} \in \mathscr{M}$ and $M \in \mathscr{M}$, the function

$$(1.8) \qquad -\sum_{i \le j} m_{ij}\, \partial\Phi^{[s]}(\bar{M})/\partial\bar{m}_{ij} = -\sum_{i,j} m_{ij}\, \partial\Phi(\bar{M})/\partial\bar{m}_{ij}\,,$$

the relation holding by (1.7) if Φ is an extension of $\Phi^{[s]}$ on an open subset of $\mathscr{R}_{k,k}$ corresponding to $\mathscr{M}$. The right side of (1.8) is more convenient computationally; when Φ is defined and differentiable on an open subset of $\mathscr{R}_{k,k}$, we define thereon the $k \times k$ matrix $\nabla\Phi$ (which is a shorter notation than the more proper grad Φ) by

$$(1.9) \qquad (\nabla\Phi)_{ij} = \partial\Phi(M)/\partial m_{ij}\,;$$

then (1.8) attains its most useful form, $-\mathrm{tr}\, M\,\nabla\Phi(\bar{M})$. If $\Phi^{[s]}$ is only defined on $\mathscr{S}_k$, it will still be convenient to use this form, by letting $(\nabla\Phi^{[s]})_{ij} = \frac{1}{2}(1 + \delta_{ij})\,\partial\Phi^{[s]}/\partial m_{ij}$ for all i, j.

Let E_{ij} be the $k \times k$ matrix with 1 in the (i, j)th place and 0 elsewhere. If b is a positive integer, it is well known (chain rule, or see [6]) that, on $\mathscr{R}^{+}_{k,k}$,

$$(1.10) \qquad \frac{\partial A^b}{\partial a_{ij}} = \sum_{h=0}^{b-1} A^h E_{ij} A^{b-h-1}\,,$$
$$\frac{\partial A^{-b}}{\partial a_{ij}} = -\sum_{h=0}^{b-1} A^{-h-1} E_{ij} A^{-b+h}\,.$$

Thus,

$$(1.11) \qquad \frac{\partial^2 A^b}{\partial a_{ij}\,\partial a_{st}} = \sum_{p,q,r \ge 0;\, p+q+r=b-2} A^p[E_{ij} A^q E_{st} + E_{st} A^q E_{ij}]A^r\,,$$
$$\frac{\partial^2 A^{-b}}{\partial a_{ij}\,\partial a_{st}} = \sum_{p,q,r \ge 0;\, p+q+r=b-1} A^{-p-1}[E_{ij} A^{-q-1} E_{st} + E_{st} A^{-q-1} E_{ij}]A^{-r-1}\,.$$

Noted during revision: We indicated earlier the likelihood that general equivalence theory has also occurred to others, and this is born out by some current publications of which the author has recently been made aware. We now describe the relationship between these results and those of the present paper, using the notation of the latter. The survey paper of Fedorov and Malyutov [12a], Theorem 2.2, states our Theorem 1 omitting the conclusion (2.17)(c), for convex Φ; nonsingularity of $M(\xi^*)$ seems to be assumed, but no differentiability assumption is stated (see our Theorem 3 and comments on it). Peter Whittle [26b], treating Φ as a convex function of ξ rather than of M as we do, proves essentially our Theorem 3, with a saddle point interpretation of $\mathscr{D}(M(\xi^*), M(\xi^*))$, and also (2.18) in the differentiable case; thus (Section 3G, below), further assumptions are needed to derive (2.17)(c) from Whittle's results. He also presents material like that of our Section 6B. A geometric duality approach of Silvey and Titterington [26a], which originally treated D-optimality and yielded iterative methods, is, according to a letter from the former, extendable to other criteria. For D-optimality, other duality cosiderations have been given by R. Sibson (discussion of [28] and a forthcoming paper). These papers all contain additional material, but no overlap with our main results, the development of explicit equivalence theory criteria for such new particular cases as those of Section 4.

2. Basic equivalence results. We examine the proofs of [23] and [10], [12] and see which portions extend to other Φ. The natural analogues of the three statements in each of (1.5) and (1.6) will be stated in (2.1), (2.10) and (2.9). The first is of course

(2.1) $$\xi^* \text{ is } \Phi\text{-optimum.}$$

The third statement is obtained by computing the obvious necessary condition for ξ^* to yield a local minimum,

(2.2) $$\frac{\partial}{\partial\alpha}\Phi(M([1-\alpha]\xi^*+\alpha\xi))\Big|_{\alpha=0^+} \geqq 0 \qquad \forall\xi \text{ in } \Xi\,;$$

we will have to assume that the differentiation in (2.2) makes sense as a right-hand derivative in α (automatic if Φ is convex), and in our derivation of Theorem 1 we in fact assume Φ continuously differentiable. Since optimality for many common Φ's, such as those of (1.2)—(1.4), entails $M(\xi^*)$ nonsingular, we will sometimes find it convenient to work in terms of the function ψ defined on $(\mathscr{M}^+)^-$ by

(2.3) $$\psi(D) = \Phi(D^{-1})\,.$$

In order to state analogues of (1.5) and (1.6), we must define the function $\mathscr{D}$ of (2.4), which requires Φ to be defined and differentiable in a neighborhood in $\mathscr{S}_k$ of the $M(\xi^*)$ under consideration. This is no restriction in many examples, where dim $(\mathscr{M}) = k(k+1)/2$; but it might be, in some isolated cases of Φ and of small and discrete $\mathscr{X}$. For the latter cases, the modifications needed in the proof of Theorem 1 are obvious, and for completeness the conclusions are stated as Theorem 2. But the most interesting, natural, and useful examples of Φ (and the most meaningful ones, from the viewpoint of "equivalence theory" parallel to (1.5) and (1.6)) are not so restricted. Hence,

THROUGHOUT THIS PAPER, UNLESS EXPLICITLY STATED TO THE CONTRARY, WE ASSUME Φ DEFINED AND DIFFERENTIABLE ON A NEIGHBORHOOD IN $\mathscr{S}_k$ OF THE $M(\xi^*)$ OR $\bar{M}$ UNDER CONSIDERATION.

For $\bar{M}$ as just described, and for $M \in \mathscr{M}$, define

(2.4) $$\begin{aligned}\mathscr{D}(M,\bar{M}) &= -\frac{\partial}{\partial\alpha}\Phi(\bar{M}+\alpha M)\Big|_{\alpha=0^+} = -\mathrm{tr}\,[M\,\nabla\Phi(\bar{M})] \\ &= \mathrm{tr}\,[\bar{M}^{-1}M\bar{M}^{-1}\,\nabla\psi(\bar{M}^{-1})]\,;\end{aligned}$$

the last two forms require Φ to be defined on an open subset of $\mathscr{R}_{k,k}$, as described in connection with (1.9), and the last form requires $\bar{M}$ to be nonsingular; if Φ is only defined on an open subset of $\mathscr{S}_k$, (2.4) becomes the first form of (1.8); the last equality of (2.4) depends on (1.7) and the chain rule $\nabla\Phi = -\bar{M}'^{-1}(\nabla\psi)\bar{M}'^{-1}$, which in turn depends on (1.10) for A^{-1}; of course, $\bar{M} + \alpha M \in \mathscr{S}_k$ for $\alpha > 0$. We also abbreviate, for ξ_x the measure assigning unit probability to the single point x, and for $M(\xi')$ of the form $\bar{M}$ described above (2.4) (the last two forms below holding when they did in (2.4)),

$$
\begin{aligned}
d(x, \xi') &= \mathscr{D}(M(\xi_x), M(\xi')) \\
&= -f(x)' \nabla\Phi(M(\xi'))f(x) \\
&= f(x)'M^{-1}(\xi') \nabla\psi(M^{-1}(\xi'))M^{-1}(\xi')f(x) ,
\end{aligned} \tag{2.5}
$$

and

$$\bar{d}(\xi') = \sup\nolimits_{x \in \mathscr{X}} d(x, \xi') . \tag{2.6}$$

We also abbreviate

$$
\begin{aligned}
d^\sharp(\xi') &= \mathscr{D}(M(\xi'), M(\xi')) \\
&= -\operatorname{tr} M(\xi') \nabla\Phi(M(\xi')) \\
&= \operatorname{tr} M^{-1}(\xi') \nabla\psi(M^{-1}(\xi')) .
\end{aligned} \tag{2.7}
$$

We shall append a subscript Φ to $\mathscr{D}$, $\bar{d}$, or $d^\sharp$ whenever ambiguity might otherwise arise.

We can now rewrite (2.2) as

$$\mathscr{D}(M(\xi), M(\xi^*)) \leqq d^\sharp(\xi^*) \qquad \forall \xi \text{ in } \Xi . \tag{2.8}$$

Since $\mathscr{D}(M, \bar{M})$ is linear in M, (2.8) is valid if and only if it is valid for all ξ of the form ξ_x. Also, there is equality in (2.8) when $\xi = \xi^*$. Thus, (2.2) is equivalent to

$$\bar{d}(\xi^*) = d^\sharp(\xi^*) . \tag{2.9}$$

(Of course, the linearity of $\mathscr{D}(M, \bar{M})$ in M implies $\bar{d} \geqq d^\sharp$.)

The relation (2.9) generalizes the third statement of (1.5) and (1.6). The extension of the second statement is

$$\bar{d}(\xi^*) = \inf\nolimits_\xi \bar{d}(\xi) . \tag{2.10}$$

We now study the implications among (2.1), (2.9), and (2.10). Of course, we have already seen that (2.1) implies (2.9), which we hereafter call *local Φ-optimality* of ξ^*. The most useful general condition on Φ such that local Φ-optimality implies global Φ-optimality is that there exist a strictly increasing function G on $\Phi(\mathscr{M})$ which is continuously differentiable at $\Phi(M(\xi^*))$ and such that

$$G \circ \Phi \quad \text{is convex on} \quad \mathscr{M} , \tag{2.11}$$

which means $G(\Phi([1 - \alpha]\bar{M} + \alpha M))$ convex in α, $0 < \alpha < 1$. (See also the first parts of Remark 3B.) If $\Phi(M(\xi')) < \Phi(M(\xi^*))$, then (2.11) implies

$$(2.12) \qquad 0 > \frac{\partial}{\partial\alpha} G(\Phi([1-\alpha]M(\xi^*) + \alpha M(\xi'))\Big|_{\alpha=0+}$$
$$= G'(\Phi(M(\xi^*)))\frac{\partial}{\partial\alpha}\Phi([1-\alpha]M(\xi^*) + \alpha M(\xi'))\Big|_{\alpha=0+},$$

violating (2.2). Thus, (2.11) and (2.2) imply (2.1).

We require a preliminary result before turning to (2.10). We shall find it convenient to invoke the condition on $\mathscr{M}' = \{M\colon M \in \mathscr{M}, \Phi(M) < \infty\}$,

$$(2.13) \qquad \Phi(M) = P(H(M)),$$

where H is positive and is *homogeneous of positive degree h*, and P is strictly decreasing and continuously differentiable on $H(\mathscr{M}')$, and such that $\log P^{-1}(\phi)$ is convex in ϕ. We will discuss these assumptions and the consequences of their violation (e.g., of $h < 0$) in Section 3. For the moment, we note that, abbreviating $M_0 = M(\xi_0)$ and $\phi_0 = \Phi(M(\xi_0))$, we have under (2.13),

$$(2.14) \qquad d^\sharp(\xi_0) = \mathscr{D}(M_0, M_0) = -hH(M_0)P'(H(M_0))$$
$$= -hP'(P^{-1}(\phi_0))P^{-1}(\phi_0)$$
$$= -h/[d\log P^{-1}(\phi)/d\phi|_{\phi=\phi_0}].$$

From the fact that $\log P^{-1}$ is decreasing and convex, we conclude that, on $\mathscr{M}'$,

$$(2.15) \qquad d^\sharp(\xi) \text{ is a non-decreasing function of } \Phi(M(\xi)).$$

The role of (2.13) is to insure (2.15); without knowing that Φ and $d^\sharp$ are functionally related, we cannot hope to relate (2.9) and (2.10).

Now assume ξ^* is Φ-optimum and hence optimum in the sense of (2.9), and that ξ^{**} satisfies (2.10) (with ξ^{**} for ξ^* there). Then, assuming (2.13) and using in order (2.10), (2.9), (2.15), and the trivial $d^\sharp \leqq \bar{d}$, we have

$$(2.16) \qquad \bar{d}(\xi^{**}) \leqq \bar{d}(\xi^*) = d^\sharp(\xi^*) \leqq d^\sharp(\xi^{**}) \leqq \bar{d}(\xi^{**}),$$

so that all members of (2.16) are equal. We conclude that ξ^* satisfies (2.10) and ξ^{**} satisfies (2.9) and (by (2.15), since $d^\sharp(\xi^*) = d^\sharp(\xi^{**})$) (2.1). Thus, under (2.13) we have shown that (2.1) and (2.10) are equivalent.

Two details remain to be treated, both just as in the D-optimality proof. Firstly, it is clear that any design ξ^* satisfying (2.9) assigns all measure to $\{x\colon d(x, \xi^*) = \bar{d}(\xi^*)\}$, since otherwise one obtains $d^\sharp(\xi^*) < \bar{d}(\xi^*)$. Secondly, if Φ satisfies (2.11), then any convex linear combination of designs satisfying (2.1) is also $G \circ \Phi$-optimum and hence Φ-optimum; and if (2.11) is strengthened by demanding that $G \circ \Phi$ be *strictly* convex, then all matrices $M(\xi)$ must be identical for Φ-optimum ξ.

We summarize:

THEOREM 1 ("Equivalence Theorem"). *For Φ continuously differentiable in a* *neighborhood of* $M(\xi^*)$,

$$\text{(2.17)}\qquad \begin{aligned}&\text{(a)}\quad (2.1) \Rightarrow (2.9)\,;\\ &\text{(b)}\quad \textit{Under}\quad (2.11): (2.9) \Rightarrow (2.1)\,;\\ &\text{(c)}\quad \textit{Under}\quad (2.13): (2.1) \Leftrightarrow (2.10)\,.\end{aligned}$$

Furthermore,

$$\text{(2.18)}\qquad \xi^* \quad \textit{satisfies} \quad (2.9) \Leftrightarrow \xi^*\{x: d(x, \xi^*) = \bar{d}(\xi^*)\} = 1\,.$$

Under (2.11), *the* Φ*-optimum* ξ^**'s (and corresponding* $M(\xi^*)$*'s) are convex; if* $G \circ \Phi$ *is strictly convex in a neighborhood in* $\mathscr{M}$ *of an optimum* $M(\xi^*)$, *then* $M(\xi^*)$ *is the unique optimum* M.

Recalling the remarks just below (2.4), we also state:

THEOREM 2. *If we do not assume* Φ *is defined in a neighborhood (in* $\mathscr{P}_k$*) of* $\mathscr{M}$, *then Theorem* 1 *is valid with the following alterations: In* (2.17) (a) *and* (b), *replace* (2.9) *by* (2.2) *or by*

$$\text{(2.19)}\qquad \inf_{x \in \mathscr{X}} \frac{\partial}{\partial \alpha} \Phi(M([1-\alpha]\xi^* + \alpha\xi_x))\Big|_{\alpha=0^+} \geqq 0\,;$$

delete (2.17)(c); *replace* (2.18) *by*

$$\text{(2.20)}\qquad (2.19) \Rightarrow \xi^* \left\{x: \frac{\partial}{\partial \alpha} \Phi(M([1-\alpha]\xi^* + \alpha\xi_x))\Big|_{\alpha=0^+} = 0\right\} = 1\,.$$

Finally, we turn to the modification of our theory in the event that Φ is not everywhere differentiable. A simple assumption with which to work is

(2.21) H is continuous on a neighborhood of $\mathscr{M}$, where Φ satisfies (2.11) and (2.13); Φ is no longer assumed differentiable.

Since Φ is convex and continuous, for fixed M and $\bar{M}$, the right-hand derivative of $\Phi((1-\alpha)\bar{M} + \alpha M)$ exists at $\alpha = 0$, the derivative $(\partial/\partial\alpha)\Phi((1-\alpha)\bar{M} + \alpha M)$ exists for almost all α, and the derivative at a convergent sequence of non-exceptional α's converges. This differentiability conclusion is seen also to hold for $H(\bar{M} + \alpha M) = (1+\alpha)^{-h} H((1+\alpha)^{-1}\bar{M} + \alpha(1+\alpha)^{-1}M)$, upon differentiating $P^{-1} \circ \Phi$. In conformity with the first relation of (2.4), we use the same definition with the derivative understood to be right-hand; or, alternatively, we can use

$$\text{(2.22)}\qquad \begin{aligned}\mathscr{D}(M, \bar{M}) &= -\lim_{\alpha \downarrow 0} \frac{\partial}{\partial \alpha} \Phi(\bar{M} + \alpha M)\\ &= -P'(H(\bar{M})) \lim_{\alpha \downarrow 0} \frac{\partial}{\partial \alpha} H(\bar{M} + \alpha M)\,,\end{aligned}$$

the limit being taken on an unexceptional sequence. Also, $d^*(\xi) = \mathscr{D}(M(\xi), M(\xi)) = -hP'(H(M(\xi)))H(M(\xi))$, as before. In (2.2) the evaluation at $\alpha = 0^+$ is replaced by taking $\lim_{\alpha \downarrow 0}$ on a non-exceptional sequence (which depends on

ξ, ξ^*), or by taking right-hand derivative at $\alpha = 0$. In place of (2.9) we obtain, using $H((1-\alpha)\bar{M} + \alpha M) = (1-\alpha)^{-h} H(\bar{M} + \alpha(1-\alpha)^{-1}M)$,

$$\sup_\xi \mathscr{D}(M(\xi), M(\xi^*)) = d^*(\xi^*) . \tag{2.23}$$

With the left side of (2.23) replacing $\bar{d}(\xi^*)$, the demonstration of (2.16) still holds. Thus, we obtain:

THEOREM 3. *Under Assumption* (2.21), *Theorem* 1 *is valid with the following alterations: Delete* (2.18); *in* (2.17), *replace* (2.9) *by* (2.23) *and* (2.10) *by*

$$\xi' = \xi^* \quad \textit{minimizes} \quad \sup_\xi -\mathscr{D}(-M(\xi), M(\xi')) , \qquad \xi' \in \mathscr{M}' . \tag{2.24}$$

For comments on this theorem, see Section 3K.

3. Remarks on complements on equivalence theorems.

A. The useful natural partial ordering on $\mathscr{M}$ is well known [17] to be

$$M_1 \succ M_2 \Leftrightarrow M_1 - M_2 \quad \text{is nonnegative definite.} \tag{3.1}$$

Most sensible Φ are non-increasing in this ordering; that is,

$$M_1 \succ M_2 \Rightarrow \Phi(M_1) \leqq \Phi(M_2) ; \tag{3.2}$$

in fact, if Φ did not satisfy (3.2), one would have the anomaly of a less informative experiment being preferred to a more informative one (without any consideration of experimental costs). One could often implement this anomaly by "throwing away" some of the information in M_1 so as to obtain M_2, as discussed on page 286 of [17]; in terms of optimality, this amounts to replacing Φ by $\tilde{\Phi}(M) = \min \{\Phi(M') : M' \prec M\}$ and solving the $\tilde{\Phi}$-optimality problem where $\tilde{\Phi}$ now satisfies (3.2). Nevertheless, we must be careful to distinguish that a Φ, which is not sensible for *all* problems because it violates (3.2) when $\mathscr{M}$ is replaced by $\mathscr{P}_k$, can be useful in a particular problem where (3.2) is satisfied. An example is $\Phi(M) = \operatorname{tr} M^2$ discussed in Section 4H.

B. The use of convexity of Φ is simply to make local optimality imply optimality, and more general conditions of *unimodality* can be used instead; such conditions are of course not generally as easy to verify as convexity.

Regarding the form of (2.11) used in the proof of Theorem 1, Φ-optimality obviously coincides with $G \circ \Phi$-optimality if G is strictly increasing. For $k = 1$, Φ strictly decreasing (a strict form of (3.2)) implies (2.11). For $k > 1$, (2.11) is not so automatic. Let $Q_\phi = \{M : M \in \mathscr{M}, \Phi(M) = \phi\}$. If, for any value ϕ, some convex mixture $\bar{M}$ of elements of Q_ϕ has $\Phi(\bar{M}) > \phi$, then clearly no rescaling $G \circ \Phi$ can be convex. On the other hand, if no such $\bar{M}$ exists for any ϕ and if (3.2) is strengthened by adding that $\Phi(\alpha M) < \Phi(M)$ for $\alpha > 1$ and $M \in \mathscr{M}'$ (so that the Q_ϕ are not k-dimensional), then it is easy to see that such a G exists. A simple example where no G exists is $\Phi(M(\xi)) = \sum_1^k [1 + \lambda_i^2(\xi)]^{-1}$ where the $\lambda_i(\xi)$ are the eigenvalues of $M(\xi)$; this satisfies the strengthened (3.2), but the condition on Q_ϕ fails for small diagonal M in Q_ϕ.

C. If we assume the strengthened form of (3.2) with $\mathscr{M}'$ replaced by $\mathscr{P}_k^+$ then the condition (2.13) is actually necessary as well as sufficient for (2.15) to hold on $\mathscr{P}_k^+$ (or on any $\mathscr{M}^+$ of full dimension). This is seen by defining g_1 by fixing M_0 and writing $g_1(\Phi(tM_0)) = \operatorname{tr}(tM_0)\nabla\Phi(tM_0)$ and then solving (for Φ) the univariate differential equation $\operatorname{tr}(tM)\nabla\Phi(tM) = g_1(\Phi(tM))$ in t for each fixed M. One obtains $\Phi(tM) = g_2^{-1}(\log[tc(M)])$ where $c(M)$ is an integration constant and $g_2(u)$ is an indefinite integral of $1/g_1$. This is easily translated into the form (2.13).

D. We now consider the relationship between (2.13) and (2.11). Convexity and monotonicity of Φ are not sufficient for (2.13), as is illustrated for $k = 1$, $\mathscr{M} = \mathscr{P}_1$, by $\Phi(x) = e^{-x}$, for which $d^\# = -x\Phi'(x) = xe^{-x}$ is not a monotone function of Φ. In the other direction, convexity and monotonicity of Φ are also not necessary for (2.13), as is illustrated by $\Phi(M) = (m_{33}^2 + m_{11}^2)^{-1} + (m_{33}^2 + m_{22}^2)^{-1}$ on the diagonal 3×3 matrices; as in the example at the end of Remark B (hold m_{33} fixed), no $G \circ \Phi$ is convex; but, with $h = 1$ and $H = \phi^{-2} = P^{-1}(\phi)$, we have $P\downarrow$ and $\log P^{-1}(\phi)$ convex. We also note that, under (3.2) and (2.13), it is easily verified that, except in degenerate cases, H cannot take on both positive and negative values, since that would make P non-monotone; it would also violate (2.11).

E. A problem may sometimes be studied most conveniently in terms of the ϕ of (2.3) rather than in terms of Φ. Since (see, e.g., [17] and Section 4B1 below)

$$[(1-\alpha)M_1 + \alpha M_2]^{-1} \prec (1-\alpha)M_1^{-1} + \alpha M_2^{-1}, \tag{3.3}$$

we see at once that (2.11) on $\mathscr{M}^+$ is a consequence of (3.2) and

$$G \circ \phi(D) \quad \text{is convex in} \quad D \quad \text{on} \quad (\mathscr{M}^+)^- . \tag{3.4}$$

In the cases treated in [23] and [10], (3.4) is satisfied; in fact, in [10] ϕ is linear. But, in general, (2.11) is weaker than (3.4). For example, if $\mathscr{M}^+$ consists only of diagonal matrices, then convexity of $\Phi(M) = \operatorname{tr}(M^{-\frac{1}{2}})$ is obvious; but no rescaling $G \circ \phi$ of $\phi(D) = \operatorname{tr} D^{\frac{1}{2}}$ can be convex on $\mathscr{T}^+$ when $k > 1$ because, in analogy with the example of Remark B above, ϕ is not convex on mixtures of two diagonal matrices with the same (permuted) set of non-identical diagonal entries.

F. Of course, the use of (2.13) is unchanged if we make *h negative* and *P increasing*. On the other hand, if we reverse only *one* of the two conditions $h > 0$, $P\downarrow$, we obtain $d^\#$ *decreasing* in Φ, in place of (2.15). The argument of (2.16) fails, and in general we cannot hope for the equivalence (2.17)(c) to be valid. For example, the function $\Phi(M) = \operatorname{tr} M^2$ treated in Section 4H will not generally be minimized by the same M that minimizes $\bar{d} = \min_x f(x)'Mf(x)$; these work in opposite directions. What is more interesting is the study of such criteria in settings which satisfy an additional restriction such as that of Section 4H ($\operatorname{tr} M =$ constant). As illustrated in the example there, it is then often possible to achieve (2.17)(c). Since this is a rather special consideration, we shall return to it elsewhere.

G. If (2.11) is satisfied but (2.13) is violated (as in examples like that of Section 4F or, less important, in F above), the first two parts of (2.17) remain valid, and these are the most important parts of the equivalence theorem. For, with a rare exception such as the criterion of G-optimality [22], $\bar{d}_\Phi$ does not arise as an optimality criterion of interest in itself, but only as a tool for proving Φ-optimality; it serves the latter role in (2.9) and (2.17)(a)-(b) (and in the resulting computational techniques and bounds of Section 6), rather than in (2.10) and (2.17)(c).

H. Simple examples show that (2.13) cannot be completely dispensed with in proving (2.17)(c); one such example is given in Section 4F. There are many simple and natural settings in which minimizing Φ, $d^\sharp$, and $\bar{d}$ can lead to three different designs in the absence of (2.13). (The criterion function $\bar{d} - d^\sharp$, which a Φ-optimum design still trivially minimizes, is of even less general intrinsic significance than $\bar{d}$.) There are special cases where (2.17)(c) is satisfied in the absence of (2.13), but we do not know definitive results. Thus, it is obvious that (2.15) can be replaced in our proof that (2.1) $\Rightarrow$ (2.10) (respectively, the opposite), by the condition that the Φ-optimum design is $d^\sharp$-optimum (respectively, the opposite), but this last is not necessary. Further conditions will be discussed elsewhere. It is interesting from the game-theoretic point of view to note that, if a $\bar{d}$-optimum design ξ is not Φ-optimum, $d(x, \xi)$ cannot achieve its maximum on the support of ξ; this does not contradict usual "minimax behavior," since ξ is not the maximin strategy of the other player for the game with payoff d.

I. The fact that constancy of $d^\sharp$ in the case (1.2) makes (2.9) easier to verify there than in the case (1.3) suggests we delimit those Φ for which there is a regular G_1 such that $d_{\tilde{\Phi}}^\sharp$ is constant, where $\tilde{\Phi} = G_1 \circ \Phi$. (If some $G \circ \Phi$ satisfies (2.11), so will $(G \circ G_1^{-1}) \circ \tilde{\Phi}$, so we need not worry about convexity.) Solving the differential equation tr $M\,\nabla\tilde{\Phi}(M) =$ constant (as in C above) yields $\tilde{\Phi}(M) = c_1 \log [c_2 H(M)]$ where H is homogeneous of degree h and $d_{\tilde{\Phi}}^\sharp = c_1 h$. This is exactly (2.13) with $P_\Phi(H) = G_1^{-1}(c_1 \log [c_2 H])$. (In terms of (2.14)—(2.15), $\log P_{\tilde{\Phi}}^\sharp$ is linear rather than *strictly* convex, and $d^\sharp$ is constant rather than strictly increasing.) Now, for any positive criterion function Φ^*, the chain rule always yields $\mathscr{D}_{\log \Phi^*}(M, \bar{M}) = \mathscr{D}_{\Phi^*}(M, \bar{M})/\Phi^*(\bar{M})$, and thus $\bar{d}_{\log \Phi^*}(\xi) = \bar{d}_{\Phi^*}(\xi)/\Phi^*(M(\xi))$. Hence, if the original Φ satisfies (2.13) and we put $\Phi^* = 1/P_\Phi^{-1}(\Phi)$ so that Φ^* is equivalent to Φ in the sense of (2.1) and Φ^* is homogeneous of degree $-h$, we have (by (2.14) and the first sentence of F, taking $P_{\Phi^*}(H) = H$) $d_{\Phi^*}^\sharp(\xi) = h\Phi^*(M(\xi))$, and thus $\bar{d}_{\log \Phi^*}(\xi) = h\bar{d}_{\Phi^*}(\xi)/d_{\Phi^*}^\sharp(\xi)$. This whole process, then, amounts to replacing Φ (satisfying (2.13)) by Φ^*, writing (2.9) for Φ^*, and dividing both sides by $d_{\Phi^*}^\sharp$, and it does not distinguish why (2.9) was genuinely simpler in the case (1.2). The reason for the latter is the multiplicative form of the determinant, which yields $\mathscr{D}_{\Phi_0}(M, \bar{M}) = \text{tr } \bar{M}^{-1}M/\det \bar{M}$.

J. In the manipulations of Section 3I it became evident that (2.17)(c) could be altered by replacing $\bar{d}_\Phi$ by $\bar{d}_{\tilde{\Phi}}$ in (2.10) for certain $\tilde{\Phi}$ equivalent to Φ in the sense of (2.1). This is quite a general result: if $\tilde{\Phi} = \tilde{G} \circ \Phi$ for some strictly increasing $\tilde{G}$, *and* $\tilde{\Phi}$ *(in place of* Φ*) satisfies* (2.13), then the derivation from (2.13) through the paragraph containing (2.16) proceeds as before, but with the criterion function Φ replaced by $\tilde{\Phi}$ everywhere. Since $\tilde{\Phi}$- and Φ-optimality in the sense of (2.1) are equivalent, we have proved

THEOREM 4. *If $\tilde{G}$ is strictly increasing and $\tilde{\Phi} = \tilde{G} \circ \Phi$ satisfies* (2.13) *then* (2.17) (c) *can be replaced by*

$$\xi^* \quad \text{is } \Phi\text{-optimum} \Leftrightarrow \xi^* \quad \text{minimizes} \quad \bar{d}_{\tilde{\Phi}}(\xi)\,. \tag{3.5}$$

Of course, we cannot replace $\bar{d}_{\tilde{\Phi}}$ in (3.5) by the $\bar{d}_\Phi$ of Theorem 1 (full equivalence) without the original (2.13) for Φ itself.

Although simultaneous minimization by the same ξ^* of $\bar{d}_{\tilde{G}\circ\Phi}$ for various $\tilde{G} \circ \Phi$ satisfying (2.13) is not a completely obvious result, replacement of (2.9) by $\bar{d}_{\tilde{\Phi}} = d_{\tilde{\Phi}}^{\#}$ in (2.17)(a)-(b) requires no proof, whether or not $\tilde{G} \circ \Phi$ satisfies (2.13). Thus, we have the interesting possibility of varying $\tilde{G}$ to achieve the most useful possible computational form of $\bar{d}_{\tilde{\Phi}} = d_{\tilde{\Phi}}^{\#}$. (In the discussion of Section 3I we disposed of the possibility of making $d_{\tilde{\Phi}}^{\#}$ constant.) This will be illustrated in the examples of Section 4C.

K. Theorem 3 will be used for such criteria as E-optimality ((1.4) and Section 4E). The critical difference from Theorem 1 is of course that the supremum in (2.23) need not be the same if we restrict ξ to the form ξ_x as we did in going from (2.8) to (2.9); that is, $\mathscr{D}(-M, \bar{M})$ is no longer linear in M. Since $\Phi((1-\alpha)\bar{M} + \alpha M)$ is *convex* in M, this non-linearity will be in the direction of possibly making (2.2) valid if ξ is restricted to ξ_x but not valid for certain other ξ. This will be illustrated in Section 4E.

Of course, (2.23) is not as useful a criterion as (2.9), and may sometimes be as difficult to implement as (2.1) directly. The example of Section 4E illustrates how particular properties of ξ^* can simplify (2.23). In all cases, violation of (2.23) when ξ is restricted to the form ξ_x obviously implies non-Φ-optimality of ξ^*.

The right-hand derivative in the definition of $\mathscr{D}$ can easily differ from the left-hand one, and so the expression $(\partial/\partial\alpha)\Phi(\bar{M} - \alpha M)|_{\alpha=0^+}$, which is equal to those of (2.4) under the differentiability assumption of Theorem 1, cannot be used in Theorem 3.

L. Much of the preceding material can be developed along game-theoretic lines in the manner of Karlin and Studden [14]. However, the present treatment seems more elementary and also seems to separate more clearly the conditions needed for the "minimax" criterion (2.10) to coincide with (2.1). Another aspect of the game-theoretic development will be mentioned in Section 7.

4. Computations and illustrations.

4A. *Transformations.* We now discuss the simple consequence of linear transformation on M and monotone transformation on Φ. Suppose $\mu = AMA' + B$ where μ is $k' \times k'$ and A is $k' \times k$, and where B is $k' \times k'$ non-negative definite. This can be thought of as relating the given problem in terms of M to another problem with regression function k'-vector g on $\mathscr{X}$, where $\mu = B + \int gg'\xi(dx)$ and where $g = Af$, and where B is the information matrix available from previous experimentation (suitably normalized relative to $\xi(\mathscr{X}) = 1$). Suppose

$$\Phi(M) = G(\tilde{\Phi}(AMA' + B)) \tag{4.1}$$

relates the optimality function Φ on $\mathscr{M}$ to that, $\tilde{\Phi}$, on $\{\mu\}$. Then, since $\partial\Phi(\bar{M})/\partial\bar{m}_{ij} = G'(\tilde{\Phi}(\bar{\mu}))[A' \nabla\tilde{\Phi}(\bar{\mu})A]_{ij}$, we obtain

$$\begin{aligned} \mathscr{D}_\Phi(M, \bar{M}) &= -\operatorname{tr}\{M \nabla\Phi(\bar{M})\} = -G'(\tilde{\Phi}(\bar{\mu})) \operatorname{tr}\{MA' \nabla\tilde{\Phi}(\bar{\mu})A\} \\ &= G'(\tilde{\Phi}(A\bar{M}A' + B))\mathscr{D}_{\tilde{\Phi}}(AMA', A\bar{M}A' + B)\,. \end{aligned} \tag{4.2}$$

This allows $\mathscr{D}_\Phi$ to be computed in terms of $\mathscr{D}_{\tilde{\Phi}}$.

Similarly, if $\Phi(M)$ is rewritten $\phi(D)$ with $D = M^{-1}$, as in (2.3) and the last form of (2.4), and if

$$\phi(D) = G(\tilde{\phi}(ADA' + B))\,, \tag{4.3}$$

we obtain for $\mathscr{D}$

$$\operatorname{tr}\{\bar{D}M\bar{D}\, \nabla\phi(\bar{D})\} = G'(\tilde{\phi}(A\bar{D}A' + B)) \operatorname{tr}\{A\bar{D}M\bar{D}A' \nabla\tilde{\phi}(A\bar{D}A' + B)\}\,. \tag{4.4}$$

4B. *Convexity tools.* We are mainly interested in verifying (2.11) and in computing $\mathscr{D}$ in the examples below. Usually (2.13) is of secondary interest, as we have mentioned earlier, and the status of (3.2) will usually be evident from the computation of $\mathscr{D}$, since (3.2) can be proved by showing that, if $M \in \mathscr{P}_k$ and $\Phi(M + \delta M)$ is defined for δ small positive and for $\delta = 0$,

$$0 \leqq \frac{\partial}{\partial\delta}\Phi(\bar{M} + \delta M)\Big|_{\delta=0} = \operatorname{tr} M \nabla\Phi(\bar{M})\,. \tag{4.5}$$

See [25b] for alternatives to (4.5).

Often the computation of the Hessian of $(k^2 + k)(k^2 + k + 2)/8$ second derivatives of Φ with respect to the $(k^2 + k)/2$ m_{ij}'s, in order to verify convexity, is tedious even in simple examples, as we shall illustrate briefly in C below, so that it is expeditious to invoke general convexity results instead. These are scattered in the literature (a recent list of some being in [2]) in such a way that the optimum design practitioner will often have difficulty finding what he needs. A forthcoming monograph by Marshall and Olkin [25a] should remedy this. Meanwhile, it seems useful to list three of the more useful tools in our setting.

1. If $\Gamma: \mathscr{M} \to \mathscr{P}_{k'}$ is convex in the ordering (3.2) on $\mathscr{P}_{k'}$ and Φ: convex span $(\Gamma(\mathscr{M})) \to R^1$ is convex and increasing, then $\Phi \circ \Gamma$ is convex. Also, R^1 may be replaced by $\mathscr{P}_{k''}$ in this statement. A familiar example on $\mathscr{P}_k^+$ is $\Gamma(M) = M^{-1}$, as described in Section 3E. The paper by Bendat and Sherman [3] discusses the classical work of Loewner (and his students Dobsch and Kraus)

on particular kinds of monotone matrix-valued functions $\bar{\gamma}$ induced by functions $\gamma: R^1 \to R^1$ as

$$\bar{\gamma}(M) = Q \operatorname{diag}[\gamma(\delta_1), \gamma(\delta_2), \cdots, \gamma(\delta_k)]Q' \tag{4.6}$$

where $\Delta = \operatorname{diag}[\delta_1, \cdots, \delta_k]$ is the diagonal matrix with diagonal entries $\{\delta_i\}$, Q is orthogonal, and $M = Q\,\Delta Q'$. These authors also extend work of Kraus on corresponding convex functions $\Gamma: \mathscr{P}_k \to \mathscr{P}_k$, which is our present interest.

2. There are known results on norms (convex by definition) on $\mathscr{P}_k$. For example, $(\operatorname{tr} A^p)^{1/p}$, $1 \leqq p < \infty$ (and maximum eigenvalue of A for $p = \infty$) is the L^p-norm.

3. A result of Ky Fan [8] (also related to results of von Neumann) states that, if $\lambda_1(A) \geqq \lambda_2(A) \geqq \cdots \geqq \lambda_k(A)$ are the eigenvalue of A, then $\sum_1^m \lambda_i(A)$ is convex on $\mathscr{P}_k$ $\forall m$. Hence, for $A, B \in \mathscr{P}_k$ and $0 < \alpha < 1$, if we define $x_i = \alpha\lambda_i(A) + (1-\alpha)\lambda_i(B)$ and $y_i = \lambda_i(\alpha A + (1-\alpha)B)$, we have the x_i and y_i nonnegative and increasing in i, and

$$\begin{aligned} \textstyle\sum_1^m x_i &\geqq \textstyle\sum_1^m y_i\,, \qquad 1 \leqq m < k\,, \\ \textstyle\sum_1^k x_i &= \textstyle\sum_1^k y_i\,, \end{aligned} \tag{4.7}$$

the last by linearity of the trace. But the "majorization" of $\{y_i\}$ by $\{x_i\}$ given by (4.7) is well known (e.g., [25c], [2]) to be equivalent to $\sum_1^k \gamma(x_i) \geqq \sum_1^k \gamma(y_i)$ for each real continuous convex function γ defined on some real interval. For each such γ, we conclude that $\sum_1^k \gamma \circ \lambda_i$ is convex on $\mathscr{P}_k$.

4C. *Simple trace criteria.* We recall the definition $M^m = Q\,\Delta^m Q'$ of (4.6) for $m > 0$ and $M \in \mathscr{P}_k$, or $m < 0$ and $M \in \mathscr{P}_k^+$. (To use (1.9), define M^m on $R_{k,k}^+$ through the exponential mapping.) We now consider the following parameters and family of functions on $\mathscr{P}_k$:

$$\begin{aligned} &r \text{ and } m \text{ are real and nonzero;} \\ &k' \text{ is a positive integer;} \quad k' \leqq k \quad \text{if} \quad m < 0\,; \\ &C \in \mathscr{P}_{k'}\,; \quad A \in \mathscr{R}_{k',k}\,, \quad \text{of rank } k' \quad \text{if} \quad m < 0\,; \\ &\Phi_{m,r,A,C}(M) = [\operatorname{tr} C(AMA')^m]^r \quad (\text{possibly } +\infty)\,. \end{aligned} \tag{4.8}$$

(A related, sometimes more useful family, will be introduced in (4.17).) As indicated in Remark 3B, changing the exponent r in (4.8) from 1 or -1 to a positive multiple thereof can only exhibit the G of (2.11), but does not change the optimum designs or validity of (2.17) (b). In what follows it will suffice to write $AMA' = \mu$ in (4.8) and to consider $[\operatorname{tr} \mu^m]^r$, since at worst this can mean *strict* convexity (not convexity itself) does not carry over from a function of μ to that of M; the question of strictness in M is then easy to answer.

We now illustrate the use of various techniques such as those mentioned in B above, to verify (2.11) in several cases. We assume $k' > 1$ since $k' = 1$ is trivial.

Case 1. Assume m integral and $r = 1$. First suppose $m > 0$, and write $\bar{E}_{ij} = E_{ij} + E_{ji}$ if $i \neq j$ and $\bar{E}_{ii} = E_{ii}$. From (1.11) and (1.7) we have, on $\mathscr{P}_{k'}^+$, for

$i \leqq j$ and $s \leqq t$,

$$\frac{\partial^2}{\partial \mu_{ij}\, \partial \mu_{st}} \operatorname{tr} C\mu^m = 2 \operatorname{tr} \textstyle\sum_{p,q,r \geqq 0; p+q+r=m-2} C\mu^p \bar{E}_{ij}\, \mu^q \bar{E}_{st}\, \mu^r . \tag{4.9}$$

Let x_{ij}, $i \leqq j$, be real, and write $\bar{X} = \sum_{i \leqq j} x_{ij} \bar{E}_{ij}$. Note that $\bar{X}$ is symmetric. We obtain, with the same right-side summation as in (4.9),

$$\textstyle\sum_{i \leqq j; s \leqq t} x_{ij}\, x_{st} \dfrac{\partial^2}{\partial \mu_{ij}\, \partial \mu_{st}} \operatorname{tr} C\mu^m = 2 \sum \operatorname{tr} C\mu^p \bar{X} \mu^q \bar{X} \mu^r . \tag{4.10}$$

Convexity as in (2.11) requires nonnegativity of (4.10) for all symmetric $\bar{X}$ in $\mathscr{R}_{k',k'}$. If $m = 1$, this is of course trivial. If $m = 2$, (4.10) becomes $2 \operatorname{tr} C\bar{X}^2$, which is nonnegative since C and $\bar{X}^2$ are in $\mathscr{P}_{k'}$. A similar derivation holds for $m = -1$. For other integral m and $C \neq$ const. $\times$ I, convexity on $\mathscr{P}_{k'}$ fails, and changing r does not alter this. There is an open subset of $\mathscr{P}_{k'}$ (depending on C) where $\operatorname{tr}(C\mu^m)$ is convex, which may be relevant in particular examples; we shall not consider this further, here.

If $C = I$, the summand on the right side of (4.10) becomes $\operatorname{tr}[(\mu^{p+r})(\bar{X}\mu^q\bar{X})]$, which is nonnegative because each parenthesized matrix is in $\mathscr{P}_{k'}$. This and the analogue for $m < 0$ yields *convexity of* $\operatorname{tr}(\mu^m)$ *for all integers* m. (See also Case 3.)

Case 2. Assume m and C arbitrary, $|r| = 1$. A theorem in the work of Bendat and Sherman [3] alluded to in B1 above yields convexity of the function $\Gamma(\mu) = \mu^m$ on $\mathscr{P}_k^+$ for all k if and only if the function $(z^m - 1)/(z - 1)$ has nonnegative imaginary part in the upper half of the complex plane. This can be verified to be true if and only if $-1 \leqq m \leqq 0$ or $1 \leqq m \leqq 2$. Hence, $\Phi_{m,1,A,C}$ *satisfies* (2.11) *for all* k, k', A, *and* C, *provided* $-1 \leqq m \leqq 0$ *or* $1 \leqq m \leqq 2$. The example at the end of Section 3E indicates why, even when $C = I$, changing r *to another positive value* does not yield an increased range of m unless $k' = 1$. However, the Bendat-Sherman tool yields convexity of $-\mu^m$ for $0 \leqq m \leqq 1$; since $-1/\operatorname{tr}\mu$ is convex and increasing, the result at the start of B1 yields that $\Phi_{m,-1,A,C}$ *satisfies* (2.11) *for all* k, k', A, *and* C, *provided* $0 \leqq m \leqq 1$.

Case 3. Assume $C = I$, m arbitrary. (Integral m have also been treated in Case 1.) Using the tool of B3 above, we have that $\Phi_{m,1,A,I}$ *satisfies* (2.11) *for all* k, k', *and* m, *provided* $m \leqq 0$ *or* $m \geqq 1$. Alternatively, B2 can be used with the convex increasing nature of x^p, $p \geqq 1$, to yield (by B1) convexity of $\operatorname{tr}\mu^p$. Then convexity of μ^{-1} and B1 yields convexity of $\operatorname{tr}\mu^{-p}$; for the additional range $0 < p < 1$ this last was obtained in Case 2. Again, the example at the end of Section 3E shows why changing r cannot help in the case $0 < m < 1$ unless $k' = 1$.

We now abbreviate $k'^{-1/p}\Phi_{-p,1/p,A,I}$ on $\mathscr{P}_k^+$ (where we recall, from (4.8), that we are still treating the case $k' \leqq k$, rank $A = k'$) by

$$\Phi^*_{p,A}(M) = [k'^{-1} \operatorname{tr}(AMA')^{-p}]^{1/p} . \tag{4.11}$$

The techniques of the previous paragraph show that (4.11) satisfies (2.11) for all $p > 0$, even with G the identity. The normalization of (4.11) is convenient for comparing the effects of using various trace criteria, as we shall illustrate elsewhere. (Similarly, $(\operatorname{tr} \mu^p)^{-1/p}$ is convex for $0 < p \leqq 1$; for $p > 1$, this fails, since $(\operatorname{tr} \mu^m)^{-r}$ is easily seen not to be convex on diagonal matrices for $k = 2$, $m > 1$, $r > 0$.)

It is not hard to see that in all the previous cases of convexity except $m = 1$, the convexity is strict if C and A have rank k.

We now turn to considerations other than (2.11), for the functions Φ of (4.8).

As for (2.13), if $h > 0$ we have $H = \Phi^{h/mr} = P^{-1}(\Phi)$ homogeneous of degree h; P is decreasing and $\log P^{-1}$ is convex if $mr < 0$. Remark 3F covers the case $mr > 0$, where we cannot in general expect (2.17)(c) to hold.

We now compute $\nabla\Phi$, recalling the convention adopted just below (1.9) when such a computation is carried out using $\mathscr{S}_k^+$ alone. From (1.10) we have, for integral $m > 0$, using the symmetry of μ,

$$\begin{aligned} \nabla \operatorname{tr}(C\mu^m) &= \textstyle\sum_{h=0}^{m-1} \mu^{m-h-1} C \mu^h , \\ \nabla \operatorname{tr}(C\mu^{-m}) &= -\textstyle\sum_{h=0}^{m-1} \mu^{-m+h} C \mu^{-h-1} . \end{aligned} \tag{4.12}$$

In particular, for all integral m,

$$\nabla \operatorname{tr} \mu^m = m\mu^{m-1} . \tag{4.13}$$

The expression (4.13) *is in fact valid for all real m*, which is known from the theory of matrix functions (power series). The expressions corresponding to (1.10) for non-integral b, and thus to (4.10) for non-integral m, require more space tc develop than is warranted here. Finally, from (4.2),

$$\nabla\Phi^{(k)}_{m,r,A,C}(M) = r[\Phi^{(k')}_{m,1,I,C}(AMA')]^{r-1} A' \{\nabla\Phi^{(k')}_{m,1,I,C}(AMA')\} A , \tag{4.14}$$

where the superscript (j) denotes domain $\mathscr{S}_j^+$.

It remains to consider (3.2). From (4.5), (4.13) and (4.14), we see that, when $C = I$, the Φ of (4.8) satisfies (3.2) for all m and r for which $mr < 0$.

If $C \neq I$ and $m = -r = \pm 1$, the expression (4.12) is positive definite, so that (3.2) again holds. Otherwise, (4.12) is not positive definite for *all* μ if $C \neq \text{const.} \times I$, and (as for (2.13) in Case 1 above) (3.2) holds only on a proper subset of $\mathscr{S}_k^+$, which depends on C.

We note that, in particular, the family (4.11) satisfies (2.11), (2.13) and (3.2) for all $p > 0$, and convexity is strict if A has rank k. For $p = 1$ and C non-singular, putting $A = C^{-\frac{1}{2}}$ in (4.11) yields the $\Phi_{1,C}$ of (1.3), which can be obtained for general C as $\Phi_{-1,+1,I,C}$, or from (4.18) below with $p = 1$.

To compute the $\mathscr{D}$ corresponding to (4.11), we use (4.13) and (4.14) to obtain

$$\begin{aligned} d(x, \xi) &= k'^{-1/p}[\operatorname{tr}(AM(\xi)A')^{-p}]^{-1+1/p} f(x)'A'(AM(\xi)A')^{-p-1}Af(x) , \\ d^*(\xi) &= [k'^{-1} \operatorname{tr}(AM(\xi)A')^{-p}]^{1/p} , \end{aligned} \tag{4.15}$$

in terms of which (2.9) and (2.10) can be written. If, instead of using (4.11), we use the equivalent $p^{-1}\operatorname{tr}(AM(\xi)A')^{-p}$ (which also satisfies (2.13)), we obtain

$$(4.16)\qquad \begin{aligned} d(x,\xi) &= f(x)'A'(AM(\xi)A')^{-p-1}Af(x)\,, \\ d^{*}(\xi) &= \operatorname{tr}(AM(\xi)A')^{-p}\,. \end{aligned}$$

Of course, (2.9) from (4.15) is obviously the same as from (4.16), although our result on the equivalence of the two forms of (2.10) is not so obvious, as mentioned in Section 3J. Although one would presumably use (4.16) in practice, the more complex (4.15) has also been stated, for use in Section 4D below.

A variant of (4.8), sometimes more useful when $k' < k$ in view of the meaning of $AM^{-1}A'$ as proportional to a covariance matrix, is

$$(4.17)\qquad [\operatorname{tr} C(AM^{-1}A')^{m}]^{r}\,,$$

for which the conditions analogous to those of (4.8) should be obvious; the expressions (4.17), (4.18), and (4.21), can of course be meaningful if $M \notin \mathscr{P}_k^{+}$, but then we must use Theorem 3 (see also Section 7) rather than the formulas we shall develop here to implement Theorem 1. The general discussion of (2.11) is similar to that for (4.8); a main difference is that $AM^{-1}A'$ is not the matrix inverse of a linear function of M. We shall not take the space for a full discussion, except in the important case (analogous to (4.11))

$$(4.18)\qquad \Phi_{p,A}^{**}(M) = [k'^{-1}\operatorname{tr}(AM^{-1}A')^{p}]^{1/p}\,,$$

for which (as in the alternate proof of Case 2) we can use the convex increasing structure of $(\operatorname{tr} D^{p})^{1/p}$ and the convexity in M of $D = AM^{-1}A'$ to obtain (2.11) for (4.18); (2.13) is again obvious. We now use (4.4) in computing $\mathscr{D}$ for $M \in \mathscr{P}_k^{+}$ (recalling the comment below (4.17) for $M \notin \mathscr{P}_k^{+}$). The simpler equivalent $p^{-1}\operatorname{tr}(AM^{-1}A')^{p}$ to (4.18) yields, in analogy with (4.16),

$$(4.19)\qquad \begin{aligned} d(x,\xi) &= f(x)'M^{-1}(\xi)A'(AM^{-1}(\xi)A')^{p-1}AM^{-1}(\xi)f(x)\,, \\ d^{*}(\xi) &= \operatorname{tr}(AM^{-1}(\xi)A')^{p}\,; \end{aligned}$$

and the analogue of (4.15), obtained by using (4.18) itself, is achieved by multiplying the functions of (4.19) by $k'^{-1/p}[\operatorname{tr}(AM^{-1}(\xi)A')^{p}]^{-1+1/p}$.

In (4.35) we will discuss (4.19) further.

Of course, for $k' = 1$ all criteria satisfying (3.2) coincide. The Chebyshev equivalence criterion and related matters in this case have been treated extensively in [7], [22], [25], [13], [14], [15], and require no further discussion here.

4D. *D-optimality*. It is natural to define, for $AMA' \in \mathscr{P}_{k'}^{+}$,

$$(4.20)\qquad \Phi_{0,A}^{*}(M) = \lim_{p\to 0}\Phi_{p,A}^{*}(M) = [\det(AMA')]^{-1/k'}$$

and, for $M \in \mathscr{P}_k^{+}$,

$$(4.21)\qquad \Phi_{0,A}^{**}(M) = \lim_{p\to 0}\Phi_{p,A}^{**}(M) = [\det(AM^{-1}A')]^{1/k'}\,.$$

(Recall the remark below (4.17) for $M \notin \mathscr{P}_k^{+}$.) These automatically satisfy (2.11), (2.13), (3.2), and $A \in \mathscr{R}_{k',k}^{+}$ implies strict convexity throughout $\mathscr{P}_k^{+}$. For

$A \in \mathscr{R}_{k,k}^{+}$ these criteria are of course equivalent to the D-optimality criterion (1.2). In terms discussed in Section 3I, a more useful $\bar{d}$, that of G-optimality when $A = I$, is obtained if we use $k' \log \Phi$ in place of each of (4.20) and (4.21). Clearly (2.13) and (3.2) are still satisfied, and it is well known that (2.11) can be obtained directly or by using $\log \det \mu = \lim_{p \to 0} p^{-1}(\operatorname{tr} \mu^p - k')$ appropriately. We obtain, from $\nabla \log \det \mu' = \mu^{-1}$ and (4.2) or (4.4),

$$d(x, \xi) = f(x)'A'(AM(\xi)A')^{-1}Af(x) \tag{4.22}$$

in the case of (4.20), and

$$d(x, \xi) = f(x)'M^{-1}(\xi)A'(AM^{-1}(\xi)A')^{-1}AM^{-1}(\xi)f(x) \tag{4.23}$$

in the case of (4.21), with $d^*(\xi') = k'$ in both cases. Note that these coincide with the results obtained formally by taking limits in (4.16) and (4.19).

If f is partitioned as $\binom{f_1}{f_2}$ with f_1 having k' components and M and $M^{-1} = D$ (say) are partitioned correspondingly, then "D-optimality for the first k' parameters" corresponds to putting $A = [I \,\vdots\, 0]$ in (4.21), from which (4.23) becomes

$$\begin{aligned} d(x, \xi) &= (D_{11}f_1 + D_{12}f_2)'D_{11}^{-1}(D_{11}f_1 + D_{12}f_2) \\ &= f'M^{-1}f - f_2'(M_{22})^{-1}f_2 \,. \end{aligned} \tag{4.24}$$

The latter form was first given in [19].

4E. *E-optimality*. For the sake of explicitness, we relabel the eigenvalues of 4B3 as $\lambda_{\max}(A) = \lambda_1(A)$ and $\lambda_{\min}(A) = \lambda_k(A)$. In analogy with (4.21)—(4.22), we define, for A of rank k',

$$\Phi_{\infty,A}^{*}(M) = \lim_{p \to \infty} \Phi_{p,A}^{*}(M) = \lambda_{\max}([AMA']^{-1}) \tag{4.25}$$

and

$$\Phi_{\infty,A}^{**}(M) = \lim_{p \to \infty} \Phi_{p,A}^{**}(M) = \lambda_{\max}(AM^{-1}A') \,. \tag{4.26}$$

When $A = I$ these both reduce to the E-optimality criterion (1.4). Both of them satisfy (2.11), (2.13), and (3.2), but at almost all points of $\mathscr{S}_k$ strict convexity is not satisfied, even when $A = I$. Since (4.25) and (4.26) are not differentiable, we must now use Theorem 3 rather than Theorem 1.

We first consider (4.25). If $\bar{\mu} \in \mathscr{S}_{k'}^{+}$ with $\lambda_{\min}(\bar{\mu})$ of multiplicity q, let the rows of $Q_1(q \times k')$ be orthonormal eigenvectors of $\bar{\mu}$ corresponding to $\lambda_{\min}(\bar{\mu})$, so that $Q_1 \bar{\mu} Q_1' = \lambda_{\min}(\bar{\mu})I_q$. A simple computation of $\det[\bar{\mu} + \varepsilon\mu - (\lambda_{\min}(\bar{\mu}) + \delta)I_{k'}]$ as $\varepsilon \downarrow 0$ verifies the well-known result that this determinant vanishes when $\delta = \varepsilon \times$ (any eigenvalue of $Q_1 \mu Q_1'$) $+ O(\varepsilon^2)$. Hence,

$$\lim_{\alpha \downarrow 0} \frac{\partial}{\partial \alpha} \lambda_{\min}(\bar{\mu} + \alpha\mu) = \lambda_{\min}(Q_1 \mu Q_1') \,. \tag{4.27}$$

If, in place of (4.25), we use the equivalent $-\lambda_{\min}(AMA')$, we thus obtain, for (2.23), with Q_1 corresponding to $\bar{\mu} = AM(\xi^*)A'$,

$$\sup_\xi \lambda_{\min}(Q_1 AM(\xi)A'Q_1') = \lambda_{\min}(AM(\xi^*)A') \,; \tag{4.28}$$

and, for (2.24), that ξ^* minimizes the left side of (4.28). The corresponding results for the original (4.25) are obtained by dividing both members of (4.28) by $\lambda^2_{\min}(AM(\xi^*)A')$, a less useful form.

In an important special case, (4.28) simplifies considerably: if $\lambda_{\min}(\bar{\mu})$ is simple, $\lambda_{\min}(Q_1\mu Q_1') = Q_1\mu Q_1'$, linear in μ, yielding

THEOREM 5. *If $\lambda_{\min}(AM(\xi^*)A')$ is positive and simple ($q = 1$), with normalized row eigenvector Q_1, then ξ^* satisfies* (2.1), (2.23) *and* (2.24) *for $\Phi(M) = -\lambda_{\min}(AMA')$, iff*

$$\sup_x [Q_1 Af(x)]^2 = \lambda_{\min}(AM(\xi^*)A') . \tag{4.29}$$

Otherwise, (4.29) *is replaced by* (4.28).

It is easy to give examples which demonstrate the insufficiency of restricting consideration to M of rank 1 as in (2.9) (or, in fact, to any rank $< q$) when $q > 1$. Most obvious is the extreme case $q = k' \geqq 2$: for any $\bar{\mu}$ of the form $cI_{k'}$ with c a positive scalar, the minimum eigenvalue of $\bar{\mu} + \alpha\mu$ is again c if μ has rank $< k'$. For such a $\bar{\mu}$, the left side of (4.28) when ξ is restricted to the form ξ_x is even less than the right side! See also Section 3K.

We note that formally letting $p \to \infty$ in (4.15) yields an incorrect result in place of (2.23), both because of the incorrect restriction to ξ_x and also because of interchanging passage to the limit with differentiation.

We now turn to (4.26). Let $\lambda_{\max}(AM^{-1}(\xi^*)A')$ have multiplicity $\bar{q}$ and let $AM^{-1}(\xi^*)A'$ have corresponding normalized row eigenvectors $\bar{Q}_1(\bar{q} \times k')$. An analysis like that above (using also $(\bar{M} + \varepsilon M)^{-1} = \bar{M}^{-1} - \varepsilon\bar{M}^{-1}M\bar{M}^{-1} + O(\varepsilon^2)$) yields, from (4.26), for (2.23),

$$\sup_\xi \lambda_{\max}(\bar{Q}_1 AM^{-1}(\xi^*)M(\xi)M^{-1}(\xi^*)A'\bar{Q}_1') = \lambda_{\max}(AM^{-1}(\xi^*)A') , \tag{4.30}$$

with (2.24) being obtained from the left side. The seemingly more complex form obtained by dividing both sides of (4.30) by $\lambda^2_{\max}(AM^{-1}(\xi^*)A')$ also has a left side which can be used for (2.24), since this is the form obtained from the equivalent $\Phi = -\lambda_{\min}([AM^{-1}A']^{-1})$. We mention this alternate form because it, rather than (4.30) as it stands, is a more direct analogue of (4.28), since this alternate form reduces to the latter when $k = k'$, upon writing A'^{-1} for A here. Similarly, the alternate form mentioned below (4.28) is the analogue of (4.30). Which form is more convenient in each of the cases (4.25) and (4.26) may depend on the example at hand, but the nonanalogous forms (4.28) and (4.30) as stated seem the simplest choices unless $k' = k$, in which case (4.30) is less convenient.

As an illustration of (4.30) when $k' < k$, suppose we are interested in the accuracy of "standard" linear combinations of the first k' parameters without scale change; that is, we take $A = [I \,\vdots\, 0]$ as in (4.24), and we partition M and $D = M^{-1}$ correspondingly. Write $N_{12}(\xi^*) = D_{11}^{-1}(\xi^*)D_{12}(\xi^*) = -M_{12}(\xi^*)M_{22}^{-1}(\xi^*)$. Then (4.30) reduces to

$$\sup_\xi \lambda_{\max}(\bar{Q}_1[I \,\vdots\, N_{12}(\xi^*)]M(\xi)[I \,\vdots\, N_{12}(\xi^*)]'\bar{Q}_1') = \lambda_{\max}(D_{11}(\xi^*)) . \tag{4.31}$$

Thus, we have

THEOREM 6. *If* $\lambda_{\max}(AM^{-1}(\xi^*)A')$ *is positive and simple* ($\bar{q} = 1$), *with normalized row eigenvector* $\bar{Q}_1$, *then* ξ^* *satisfies* (2.1), (2.23), *and* (2.24) *for the* Φ *of* (4.26), *iff*

$$\sup_x [\bar{Q}_1 AM^{-1}(\xi^*)f(x)]^2 = \lambda_{\max}(AM^{-1}(\xi^*)A') , \tag{4.32}$$

which reduces in the case of (4.31) *to*

$$\sup_x \{\bar{Q}_1[f_1(x) + N_{12}(\xi^*)f_2(x)]\}^2 = \lambda_{\max}(D_{11}(\xi^*)) . \tag{4.33}$$

Otherwise, (4.32) *and* (4.33) *are replaced by* (4.30) *and* (4.31).

It is interesting to compare the reductions of (4.19), (4.24), and (4.33) in the case $A = [I \mid 0]$. Since (4.24) is the most familiar of these, we express the other functions in terms of

$$\delta(x, \xi^*) = D_{11}^{\frac{1}{2}}(\xi^*)f_1(x) + D_{11}^{-\frac{1}{2}}(\xi^*)D_{12}(\xi^*)f_2(x) . \tag{4.34}$$

We then have

$$\begin{aligned} &D\text{-optimality } \bar{d} \text{ of } (4.19) = \sup_x \delta'\delta ; \\ &\bar{d} \text{ of } (4.24)(A\text{-optimality if } p = 1) = \sup_x \delta' D_{11}^{p}(\xi^*)\delta ; \\ &(4.33)(E\text{-optimality criterion if } \bar{q} = 1) = \lambda_{\max}^{-1}(D_{11}) \sup_x (\bar{Q}_1'\delta)^2 . \end{aligned} \tag{4.35}$$

(If we had used the logarithm of (4.26) for Φ, we would have obtained simply $\sup_x (\bar{Q}_1'\delta)^2$ in the last line of (4.35), but this does not seem simpler in applications than the other two forms discussed above.)

An example. Although one often encounters parametric families of matrices M over which $\lambda_{\min}(M)$ is maximized when $q > 1$, the simpler forms of Theorems 5 and 6 have frequent applicability. For example, in the case of quadratic regression $[-1, 1]$, with $f(x)' = (1, x, x^2)$ and $A = I$, either Theorem 5 or Theorem 6 can be used to show that the unique E-optimum design is given by $\xi^*(1) = \xi^*(-1) = \frac{1}{5}$, $\xi^*(0) = \frac{3}{5}$. For this design,

$$M(\xi^*) = \begin{pmatrix} 1 & 0 & .4 \\ 0 & .4 & 0 \\ .4 & 0 & .4 \end{pmatrix} \quad \text{and} \quad Q_1 = (5^{-\frac{1}{2}}, 0, -2(5^{-\frac{1}{2}}))$$

corresponding to the simple $\lambda_{\min} = \frac{1}{5}$; thus, (4.29) becomes the trivially verified $\sup_x (1 - 2x^2)^2/5 = \frac{1}{5}$.

More complex examples will be treated elsewhere, including detailed computations in higher-dimensional simplex experiments (generalizing the results obtained for dimensions 1 and 2 in the present example and that of Section 6A).

4F. *Trace criteria modified to include previous information.* For any of the trace criteria $\tilde{\Phi}$ of Sections 4C—4E, and for fixed B in $\mathscr{P}_k$, replacing the argument M by $M + B$ yields a new criterion $\Phi(M) = \tilde{\Phi}(M + B)$. Such Φ's arise in at least four different contexts of applications:

(i) The experimenter has available an information matrix B from a previous experiment (scaled relative to $\xi(\mathscr{X}) = 1$) and wants to combine it with the information $M(\xi)$ from a new experiment so as to minimize $\tilde{\Phi}(M(\xi) + B)$.

(ii) For certain normal Bayesian models which have appeared in the literature, minimizing the total expected loss for an appropriate loss function is often equivalent to minimization of such a corresponding $\tilde{\Phi}(B + M)$, where B is a parameter of the prior distribution.

(iii) In response surface theory (e.g., [4], [16], [21]) where one purposely fits surface of incorrect form for the sake of simplicity (fewer nonzero parameters in the fitted curve), the form $\tilde{\Phi}(B + M)$ arises with B a matrix in terms of which an assumption on the bias of the fit is expressed.

(iv) In some iterative methods for minimizing $\tilde{\Phi}(M)$, if B_n is the approximation to the minimizer after n stages, one tries, approximately, to minimize $\tilde{\Phi}((1 - \varepsilon_n)B_n + \varepsilon_n M_n)$ at the next stage, for a suitable sequence $\{\varepsilon_n\}$. (See Section 6B.)

We shall not pursue these uses here. The arithmetic of Sections 4C—4E can be modified for use here in accordance with the formula obtained from (4.2): $\mathscr{D}_\Phi(M, \bar{M}) = \mathscr{D}_{\tilde{\Phi}}(M, \bar{M} + B)$, which is valid also for the nondifferentiable E-optimality criteria. *We note that the condition for differentiability is now weakened*; for example, whenever we previously required $M(\xi^*) \in \mathscr{P}_k^+$, we now require only $B + M(\xi^*) \in \mathscr{P}_k^+$.

Here, then, we will only take the space to discuss the fact that (2.13) does not hold, so that one cannot expect (2.17)(c) to hold in general. (See also Section 3H.) We now give one such example, chosen for arithmetical simplicity. Suppose $\mathscr{X} = \{1, 2\}$, $k = 2, f(1)' = (1, 5), f(2)' = (3, 16)$, and $B = \begin{pmatrix} 1 & 5 \\ 5 & 29 \end{pmatrix}$. The problem is to minimize $\Phi(M) = \Phi_{1,I}^*(B + M) = \operatorname{tr}(B + M)^{-1}$. As before, let ξ_i assign measure 1 to the point i, and abbreviate $f(i)f(i)' = M(\xi_i)$ by M_i. The criterion (2.9), by (4.15) or (4.16), is

$$\max_i \{\operatorname{tr}[B + M(\xi^*)]^{-2}M_i\} = \operatorname{tr}[B + M(\xi^*)]^{-2}M(\xi^*)\,, \tag{4.36}$$

and $\bar{d}(\xi^*)$ is the left side of (4.36). A direct computation yields

$$(B + M_1)^{-2} = \frac{1}{8}\begin{pmatrix} 377 & -70 \\ -70 & 13 \end{pmatrix}, \tag{4.37}$$

$$\operatorname{tr}(B + M_1)^{-2}M_2 = \tfrac{1}{8}\,, \qquad d^*(\xi_1) = \operatorname{tr}(B + M_1)^{-2}M_1 = \tfrac{1}{4} = \bar{d}(\xi_1)\,,$$

from which (4.36) yields Φ*-optimality of* ξ_1. On the other hand,

$$(B + M_2)^{-2} = \frac{1}{1681}\begin{pmatrix} 84034 & -15635 \\ -15635 & 2909 \end{pmatrix},$$

$$d^*(\xi_2) = \operatorname{tr}(B + M_2)^{-2}M_2 = \tfrac{50}{1681}\,, \tag{4.38}$$

$$\operatorname{tr}(B + M_2)^{-2}M_1 = \tfrac{409}{1681} = \bar{d}(\xi_2)\,,$$

so that $\bar{d}(\xi_1) > \bar{d}(\xi_2)$ and $d^*(\xi_1) > d^*(\xi_2)$. Thus, without taking the space to compute the $\bar{d}$- or the d^*-optimum design, we see that neither can be the Φ-optimum design ξ_1.

As remarked in Section 3H, it is also not hard to find examples where $\bar{d}$-, $d^\sharp$-, and Φ-optimality coincide; for example, let $B = I$ and $f(i)' = (2 - i, i - 1)$ above. More generally, under (2.11) if $\mathscr{X}$ has k points and a Φ-optimum design has $\xi^*(i) > 0$ for all i, then ξ^* is $\bar{d}$- and $d^\sharp$-optimum even without (2.13).

4G. *Compound criteria.* This heading will be used, loosely, to describe criteria "built up" from simpler criteria, usually for one of the following three reasons: (1) uncertainty about the loss or covariance structure; (2) incomparability of various parts of M^{-1}, in terms of simple loss considerations, (3) the desire to combine features of several Φ's (illustrated also in many examples which fall under both of the previous headings). These descriptions are of course imprecise, intended to give the rationale behind adoption of certain criteria, rather than to categorize them taxonomically. We shall concentrate on such rationale, rather than on detailed analysis of these criteria, here.

Ideally, the criterion Φ is known exactly, at least to either decision theorists or subjectivists. In practice, the customer is often vague or confused about his objectives and relative losses, and the statistician's discussion of the positive and negative features of various loss structures may aid in the choice of a criterion and resulting design which reflect the customer's aims. There are arguments (e.g., [5]) that only criteria of the $\Phi^{**}_{1,A}$ variety need be considered, and it is certainly arguable from a decision-theoretic point of view that expectation of a non-quadratic function of errors may in principle be more meaningful than are our possibly non-linear functions Φ of covariances. Nevertheless, we feel the general discussion of such Φ's can be a fruitful basis for constructing a variety of designs among which the practitioner can find one at least approximately achieving his goals.

Regarding (1), suppose $\{\Phi_\tau\}$ is a family of criteria, indexed by some set $T = \{\tau\}$. If various possible customers of a single experiment have different loss functions Φ_τ, the designer may want to consider, for example, $\sup_{\tau \in T} \Phi_\tau$ or some convex average $\int_T \Phi_\tau \eta(d\tau)$ as his optimality criterion. (We omit mention of other such combinations, and discussion of "rational behavior axioms" which justify using or not using any of these.) Such a combination could also arise because of the experimenter's uncertainty about which Φ_τ is appropriate for a single user; equivalently, the same mathematical framework arises even in the case of a single Φ, if there is uncertainty about the covariance structure (thus far assumed to be that specified in the first paragraph of Section 1), as we shall discuss in (5.1)—(5.2).

If each of the Φ_τ satisfies (2.11), or (3.2), or (2.13) with the same degree h of homogeneity, then so does a supremum or convex average. If the Φ_τ are equi-differentiable, then their convex average is differentiable and Theorem 1 applies to it. This is not the case for the supremum, for which the treatment of Section 4E is typical.

In order to introduce (2), we describe the rationale some designers have given for their choice of certain criteria. On the one hand, some of these designers find D-optimality does not reflect their aims because the choice of design may be governed to such a great extent by a few characteristic values of M, that other possible advantages are sacrificed. On the other hand, use of a criterion such as A-optimality often seems unsatisfactory because in a multifactor setting it may mean adding squared units of apples to those of dung, or even to those of water $\times$ sunshine. Ideally the matrix A in $\Phi_{1,A}^{**}$ reflects such differences in units and their relative importance; in practice, again, A is usually not even approximately known, and the practitioner may feel uneasy about treating different factors, or interactions of different orders, in additive terms as in A-optimality.

One possible response to the above considerations is to use a criterion which combines, in the determinental manner of D-optimality, the losses from factors measured in different units; but which combines the contributions from different levels of the same factor through a criterion such as A- or E-optimality among items measured in the same units. This means that if θ' is decomposed as $(\theta^{(1)\prime}, \theta^{(2)\prime}, \cdots, \theta^{(r)\prime})$ where $\theta^{(i)}$ is a k_i-vector, with a corresponding decomposition of M, and if A_i is a $k_i \times k$ matrix of 0's except for an I_{k_i} in the $(1 + \sum_1^{i-1} k_j)$th to $(\sum_1^i k_j)$th columns, we use a criterion such as

$$\Phi(M) = \prod_{i=1}^r \Phi_{p,A_i}^{**}(M) , \tag{4.39}$$

where the simplest choice of p is 1 or ∞. Modifications of (4.39) will be obvious.

Thus, for example, if $p = 1$ in (4.39) and we compute $\mathscr{D}_{\log\Phi}$ from (4.19) and (4.2), we obtain

$$\begin{aligned} d(x, \xi) &= \textstyle\sum_{i=1}^r (\mathrm{tr}\, A_i M^{-1}(\xi) A_i')^{-1} f(x)' M^{-1}(\xi) A_i' A_i M^{-1}(\xi) f(x) , \\ d^*(\xi) &= r ; \end{aligned} \tag{4.40}$$

of course, $A_i M^{-1} A_i'$ is proportional to the covariance matrix of best linear estimators of $\theta^{(i)}$.

We mention two other illustrations of compound criteria. Firstly, in problems of extrapolation, one can think of Φ_τ as the variance of estimated response at the point τ. The average of Φ_τ's has then been considered frequently in response surface design considerations, and $\max_\tau \Phi_\tau$ has arisen in extrapolation to more than one point [24] as well as in the formulation of G-optimality [22]. Secondly, the criterion

$$\Phi(M) = \max_i \Phi_{p,A_i}^{**}(M) \tag{4.41}$$

includes, in the form $\max_i (M^{-1})_{ii}$ when all $k_i = 1$, the criterion Elfving [7a] described as that of "minimaxing over single parameter variances," to distinguish it from E-optimality, which is "minimaxing over variances of standard parametric functions."

4H. *Shah's criterion*. As a final illustration of our theory, we consider the criterion $\Phi_{-2,I}^*(M) = \mathrm{tr}\, M^2$ mentioned in Sections 3A and F. Suppose $\mathrm{tr}\, M$ is a constant kc on $\mathscr{M}$. Then minimizing $\mathrm{tr}\, M^2$ may not be so foolish. To see this, suppose the optimality criterion $\tilde{\Phi}$ of primary interest, but which leads to more difficult computations than tr M^2, has its minimum over $\mathscr{N} = \{M: M \in \mathscr{P}_k, \text{tr } M = kc\}$ (which includes $\mathscr{M}$) at $M = cI_k$, which is close to but not in $\mathscr{M}$. If $\tilde{\Phi}$ on $\mathscr{N}$ is a twice-differentiable function of only the eigenvalues of M (i.e., is orthogonal-invariant), then the terms through second degree of its Taylor series development about $M = cI_k$ are $\Phi(cI_k) + c_1 \text{ tr } (M - cI_k)^2 = c_2 + c_3 \text{ tr } M^2$, where the c_i are constants. Thus, if cI_k is sufficiently close to $\mathscr{M}$, minimizing tr M^2 will come close to minimizing $\Phi(M)$. Explicit bounds can be given, but we shall not take the space to do so here. (There are obvious modifications of the above where cI_k is replaced by another matrix. This occurs in the example below.) The minimization of tr M^2 was first considered by Shah [26], and we hereafter call it *S-optimality*. The assumption that tr $M = ck$ on $\mathscr{M}$ is most applicable in incomplete block design settings where, also, the approximate theory is of negligible usefulness (and where J. Eccleston [6a] has recently made use of the tr M^2 criterion for exact theory optimality). It does occur in some reasonable regression settings, illustrated below.

We first note, from (3.13), that

$$d(x, \xi) = -2f(x)'M(\xi)f(x)\,, \qquad d^*(\xi) = -2 \text{ tr } M^2(\xi)\,. \tag{4.42}$$

Thus, (2.9) is

$$\inf_x f(x)'M(\xi)f(x) = \text{tr } M^2(\xi)\,. \tag{4.43}$$

An example. A simple setting where $f'f$ is constant on $\mathscr{X}$, so that tr M is constant on $\mathscr{M}$, is that of linear regression on a subset of the unit $(k-2)$-sphere; there is an obvious trigonometric reformulation, which we consider at the same time. Here we treat in detail the case $k = 3$, with $\mathscr{X}$ the arc $\{(x_1, x_2): x_1 = \cos\theta, x_2 = \sin\theta, |\theta| \leqq \theta_0\}$ where θ_0 is specified, $0 < \theta_0 \leqq \pi$. Also, $f(x)' = (1, x_1, x_2)$. On grounds of symmetry, we try designs of the form $\xi_{(\alpha)}(\theta = 0) = 1 - 2\alpha$, $\xi_{(\alpha)}(\theta = \pm\theta_0) = \alpha$. Then

$$M(\xi_{(\alpha)}) = \begin{pmatrix} 1 & 1 - 2\alpha(1 - \cos\theta_0) & 0 \\ 1 - 2\alpha(1 - \cos\theta_0) & 1 - 2\alpha\sin^2\theta_0 & 0 \\ 0 & 0 & 2\alpha\sin^2\theta_0 \end{pmatrix},$$

$$\begin{aligned} \text{tr } M^2(\xi_{(\alpha)}) = 4\{&2\alpha^2[\sin^4\theta_0 + 2(1 - \cos\theta_0)^2] \\ &- \alpha[2(1 - \cos\theta_0) + \sin^2\theta_0] + 1\}\,. \end{aligned} \tag{4.44}$$

Let the positive value q satisfy $2q^3 + 2q^2 + q - 1 = 0$, and let $\theta^* = \cos^{-1} q$. If $\theta_0 \geqq \theta^*$, so that $\cos\theta_0 \leqq q$, the value

$$\alpha = \frac{2(1 - \cos\theta_0) + \sin^2\theta_0}{4[(1 - \cos^2\theta_0)^2 + (1 - \cos\theta_0)^2]} = \frac{3 + \cos\theta_0}{4[2 - \cos^2\theta_0 - \cos^3\theta_0]} \tag{4.45}$$

is $\leqq \frac{1}{2}$ (so $1 - 2\alpha \geqq 0$), and this value minimizes tr $M(\xi_{(\alpha)})$; and (4.45) is always $\geqq 0$. Since $f(x)'M(\xi_{(\alpha)})f(x) = \text{tr } M^2(\xi_{(\alpha)})$ for $\theta = 0, \pm\theta_0$ (this in fact being equivalent to (4.45) and thus another way of deriving it), and since $f(x)'M(\xi_{(\alpha)})f(x)$

can be written as a quadratic in x_1 for $1 \geqq x_1 \geqq \cos\theta_0$, we see that (4.43) is satisfied by $\xi_{(\alpha)}$ if and only if the coefficient of x_1^2 in this quadratic is $\leqq 0$. This coefficient, from (4.44), is $1 - 4\alpha\sin^2\theta_0$, which is $\leqq 0$ (from (4.45)) if and only if $\theta_0 \leqq 2\pi/3$. We conclude, as part of our solution, that $\xi_{(\alpha)}$ as given by (4.45) is S-optimum if $\theta^* \leqq \theta_0 \leqq 2\pi/3$.

The remaining parts of the solution are easy. If $2\pi/3 \leqq \theta_0 \leqq \pi$, assigning probability $\frac{1}{3}$ to each of three points $2\pi/3$ apart (and in $\mathscr{X}$) yields a constant $f(x)'Mf(x)$ and thus satisfies (4.43). If $\theta_0 \leqq \theta^*$, the design $\xi_{(\frac{1}{2})}$ which assigns probability $\frac{1}{2}$ to each of the values $\theta = \pm\theta_0$ is seen to be S-optimum upon checking that the quadratic $f(x)'Mf(x) = (1 - x_1^2)\sin^2\theta_0 + (1 + x_1\cos\theta_0)^2$ attains its minimum on $\mathscr{X}$ at $x_1 = \cos\theta_0$.

We turn to the considerations of the first paragraph of the present subsection. Although $cI_3 \notin \mathscr{M}$ for any θ_0, the fact that $m_{11} = 1$ and $\operatorname{tr} M = 2$ for all designs makes $\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2}I_2 \end{pmatrix} = M_0$ (say) the analogue of I_k in our initial discussion; thus, when $\theta_0 \geqq 2\pi/3$, it is easily checked that the uniform 3-point design of the previous paragraph yields this M_0 and is also D-optimum (as well as A-optimum, etc.). If $\theta_0 < 2\pi/3$, assigning probability $\frac{1}{3}$ to each of the 3 points $\theta = 0, \pm\theta_0$ achieves $\bar{d} = 3$ and, thus, D-optimality. For $\theta^* < \theta_0 < 2\pi/3$, the efficiency for D-optimality of the S-optimum design (in terms of ratio of approximate numbers of observations needed to achieve the same generalized variance) is

$$[\det M(\xi_{S\text{-OPT}})/\det M(\xi_{D\text{-OPT}})]^{\frac{1}{3}} = 3[\alpha^2(1 - 2\alpha)]^{\frac{1}{3}}, \tag{4.46}$$

where α is given by (4.45). Thus, when M_0 is close to $\mathscr{M}$, θ_0 is close to $2\pi/3$, α is close to $\frac{1}{3}$, and the efficiency of (4.46) is $1 - O((\theta_0 - 2\pi/3)^2)$. Of course, $\theta_0 \leqq \theta^*$ yields efficiency 0, not surprising in view of the distance of M_0 from $\mathscr{M}$ in such cases.

5. Modification to vector observations and variable or unknown covariance or cost structure. The previous theory applies to multiresponse problems with changes that are essentially only notational, as described in [19] and [11], [12]. In terms of the first paragraph of Section 1, f is now a $k \times m$ matrix of continuous real component functions on compact $\mathscr{X}$. The expectation of a row m-vector (single multiresponse observation) corresponding to level x is $\theta' f(x)$. The m-vectors are uncorrelated, and, for the moment, each has the same covariance matrix $\sigma^2 I_m$ (altered below). Then the entire previous development of this paper is valid with only one obvious alteration: In (2.5) and all subsequent expressions $f'Bf$, a trace operation must be inserted at the beginning; note that the formula for $M(\xi)$ in terms of f is still valid.

Now suppose a multiresponse m-vector Y_x at level x costs an amount $c_x > 0$ and has positive definite covariance matrix Q_x; again, observational m-vectors are uncorrelated with each other. Assume c_x and Q_x vary continuously in x. If V_x^{-1} is the positive definite symmetric square-root of $c_x Q_x$, then replacing Y_x by $Y_x V_x$ and $f(x)$ by $\bar{f}(x) = f(x)V_x$ makes $\bar{M}(\eta) = \int \bar{f}(x)\bar{f}(x)'\eta(dx)$ the appropriate information matrix for the approximate theory problem of minimizing a functional Φ of the inverse of the covariance matrix of best linear estimators, subject to a given restriction on the total cost (rather than number) of observations. This $\bar{f}$ and $\bar{M}$ then replace f and M in the previous paragraph and in the expressions of Theorem 1. For an optimum design subject to an upper bound C on total cost, one then sets $\xi(dx) = c_x^{-1}\eta(dx)/\int_{\mathscr{X}} c_x^{-1}\eta(dx)$ in the original framework, and takes a number of observations costing approximately C.

This type of substitution is considered in [17] and [22]; the formulas in [12] for the case of variable Q_x (and constant c_x) are given in terms of f and Q, and can be obtained by making the substitution for $\bar{f}$ given above.

Next, suppose that V_x is unknown but that it is known that it is a member of some class $\{V_x^{(\tau)}, \tau \in T\}$ for some index set T. Thus, we have $\bar{M}^{(\tau)}$ in place of $\bar{M}$ in the previous paragraph. For simplicity, assume c_x constant; it will be clear how to alter this. In the language of Section 4G, $\Phi(\bar{M}^{(\tau)})$ cannot in general be rewritten as $\Phi^{(\tau)}(M)$. Rather, even when $m = 1$, we must consider $\Phi^{(\tau)}$ as a function of ξ or, what is slightly more convenient, of the vector measure $\mu_\xi(dx) = f(x)f(x)'\xi(dx)$. We then write

$$\Phi^{(\tau)}(\mu) = \Phi(\textstyle\int_{\mathscr{X}} V_x^{(\tau)}\mu(dx)V_x^{(\tau)}) \,. \tag{5.1}$$

For a criterion such as $\max_\tau \Phi^{(\tau)}(\mu)$, which is convex if Φ is, a development like that of Theorem 6 and (4.30) is now possible, but it is even harder to apply than these tools of Section 4E because of the dependence on μ (or ξ) rather than on a single matrix $M(\xi)$. However, simplification is possible for the criterion $\Phi^*(\mu) = \int_T \Phi^{(\tau)}(\mu)\eta(d\tau)$ where η is a probability measure and $\Phi^{(\tau)}([1-\alpha]\bar{\mu} + \alpha\mu)$ is such that integration over τ and differentiation with respect to α (near $\alpha = 0$) commute if $\bar{\mu}$ is Φ^*-optimum. (This condition is satisfied for many of the criteria Φ considered previously, e.g., the $\Phi^{**}_{p,A}$ for $p < \infty$, under natural assumptions on $\{V^{(\tau)}, \tau \in T\}$.) We observe that $\partial\Phi^*((1-\alpha)\bar{\mu} + \alpha\mu)/\partial\alpha|_{\alpha=0}$ is linear in μ; thus, writing $M^{(\tau)}(\xi) = \int_{\mathscr{X}} V_x^{(\tau)}\mu_\xi(dx)V_x^{(\tau)}$, we have for the equivalent (2.9) to Φ^*-optimality,

$$\begin{aligned}\max_x -\operatorname{tr} f(x)f(x)' \textstyle\int_T V_x^{(\tau)}\, \nabla\Phi(M^{(\tau)}(\xi^*))V_x^{(\tau)}\eta(d\tau) \\ = -\operatorname{tr} \textstyle\int_T M^{(\tau)}(\xi^*)\, \nabla\Phi(M^{(\tau)}(\xi^*))\eta(d\tau) \,.\end{aligned} \tag{5.2}$$

6. Computational techniques.

6A. *Analytical demonstrations*. The tools used elsewhere (e.g., [17], [18], [19], [9], [1]) to prove designs optimum for particular criteria such as D-optimality can often be employed in the same manner for general Φ of the type we have considered. A common approach is to minimize Φ over a promising finite-dimensional subset of $\{\xi\}$ and then to use (2.9) to prove optimality of the design so obtained, relative to *all* competitors. But these computations can often be shortened greatly by the use of one or more of the following.

(i) *Invariance*. Let $G = \{g\}$ be a compact group of measurable transformations of $\mathscr{X}$ onto $\mathscr{X}$, and define ξ_g by $\xi_g(A) = \xi(gA)$. Suppose Φ has the invariance property

$$\Phi(M(\xi)) = \Phi(M(\xi_g)) \tag{6.1}$$

for all ξ and g. Then, if $\bar{\xi} = \int_G \xi_g \mu(dg)$ where μ is Haar probability measure, the design $\bar{\xi}$ is *invariant* under $G(\bar{\xi}(gA) = \bar{\xi}(A)$ for all g and $A)$ and, assuming some increasing function of Φ is convex on $\mathscr{M}$, we obtain $\Phi(M(\bar{\xi})) \leqq \Phi(M(\xi))$. Thus, *there exists an invariant Φ-optimum design.* An alternate approach which is sometimes useful is to replace (6.1) and convexity by the single assumption that

$$\bar{d}(\xi_g) = \bar{d}(\xi) \, . \tag{6.2}$$

We conclude that *there is a $\bar{d}$-optimum invariant design*, and under (2.13) it is also Φ-optimum. See [17], [18], [19], [9], [1] for discussion and examples.

(ii) *Nature of* $\{x: d(x, \xi) = \bar{d}\}$. Sometimes (2.18) and the nature of d can be used to describe limitations on the nature of Φ, especially if f consist of polynomials or functions with similar oscillatory properties. This has been used extensively (e.g., [18], [19], [9], [1]) in the case of polynomial regression on Euclidean sets $\mathscr{X}$, as illustrated below.

(iii) *Special properties of certain supports.* If the f_i are linearly independent on a set $B = \{x_1, \cdots, x_k\}$ of cardinality k, we can find A in $\mathscr{R}^+_{k,k}$ such that $(Af)_i(x_j) = \delta_{ij}$. This often simplifies greatly the computation of a Φ-optimum design *among those with support B*; simplest is Φ_0 (D-optimality), for which of course $\xi(x_i) = 1/k$. Similar computations can sometimes be made for B of larger cardinality.

(iv) *Uniqueness.* Suppose we are in a setting where the optimum $M(\xi)$ is unique. There is then the question of whether the optimum ξ is unique. Sometimes this has a trivial negative answer because one knows an optimum design whose support has cardinality $> 1+$ the dimension of $\mathscr{M}$. If one knows all optimum designs are supported by subsets of a set B, then the question is that of uniqueness of a nonnegative solution to the linear equations $\sum_{x \in B} \xi(x) f_i(x) f_j(x) = m_{ij}(\xi_{\mathrm{OPT}})$ in the variables $\xi(x)$, and this has yielded uniqueness results in some cases [9].

An example. Suppose $\mathscr{X}$ is the 2-simplex $\{(x_1, x_2, x_3): \sum_1^3 x_i = 1, \text{all } x_i \geqq 0\}$, that $k = 6$, and that the components of f are the functions x_i and $x_i x_j$, $i < j$: quadratic regression. If Φ is convex and invariant under permutations of the three variables x_i, we conclude by (i) that there is a symmetric Φ-optimum design. Now note (ii) that $d(x, \xi)$ is a quartic. It is often easy to see, as in the case of $\Phi^{**}_{p,I}$ for $p < \infty$, that this quartic, extended to the plane, approaches $+\infty$ at ∞. Consequently, on the line segment which is the intersection of any line with $\mathscr{X}$, we can have $d(x, \xi) = \bar{d}(\xi)$ at no more than one interior point of the segment. For an invariant design this means the support of an optimum design is a subset of the set B consisting of the three vertices, three midpoints of edges, and center $\bar{x}$ (say). In the case of D-optimality, we can try the six points other than the center; applying (iii), if this supports a D-optimum ξ^* we know $\xi^* = \frac{1}{6}$ at each of these other points. It is then automatic (D-optimality among designs on the 6-point set) that $d(x, \xi^*) = 6$ at each of these six points, and the single explicit computation required is to check that $d(\bar{x}, \xi^*) < 6$, which is true. Finally, (iv) uniqueness is obvious here. Thus, the computations needed to characterize the D-optimum designs in general have been reduced considerably by applying (i), (ii), (iii). For A- or E-optimality (the latter having support in B by a simple limiting argument), slightly more computation is needed. It is perphaps quickest to minimize $\Phi(M)$ in these cases with respect to the two variables ξ(vertex), $\xi(\bar{x})$. The interesting feature of the result is that the optimum ξ is now positive on all seven points of B, unlike the D-optimum ξ^*. This example and its higher-dimensional extensions will be treated elsewhere; I am indebted to R. J. Walker and Z. Galil for carrying out the computations.

6B. *Iterative methods.* There is a considerable literature on the determination of a sequence $\{\xi^{(n)}\}$ which converges to an optimum design, in the case of A- and D-optimality; e.g., [1a], [27], [28], [12]. As this aspect of the subject has developed, these authors have given increasing attention to the difficult problem of improving the obvious iterative techniques. The latter are all we will comment on here: if Φ is convex, we have at our disposal all the computational techniques for minimizing a convex function on a compact finite dimensional set whose extreme points $M(\xi_x)$ are readily available.

Simplest (and most used as a basis for D- and A-optimality in the past) are the descent methods for which $M(\xi^{(0)})$ is nonsingular and

$$\xi^{(n+1)} = (1 - \varepsilon_n)\xi^{(n)} + \varepsilon_n \xi_{x_n} \tag{6.3}$$

where x_n maximizes (or approximately maximizes) $d(x, \xi^{(n)}) - d^*(\xi^{(n)})$ (see (2.2)—(2.9)). There are many possible choices for the ε_n. Most general in applicability is any fixed sequence for which $1 > \varepsilon_n \downarrow 0$ and $\sum \varepsilon_n = +\infty$. For example, $\Phi(M(\xi^{(n)}))$ converges to the minimum for convex Φ with two bounded derivatives on a closed convex set to which (6.3) is limited by truncation and in whose interior all minima lie; under Theorem 1, if Φ is strictly convex, $M(\xi^{(n)})$ must also converge to the unique optimum value, but convergence of $\xi^{(n)}$ depends on the choice of x_n or considerations of 6A (iv).

By letting ε_n depend on $\xi^{(n)}$ and x_n, one can weaken regularity assumptions and speed convergence, but uses more computation. For D-optimality, det $M(\xi^{(n+1)})$ is easily minimized analytically with respect to ε_n, and this is the basis for procedures in the literature cited above. For A-optimality, Fedorov [12] obtains an upper bound b_n on the optimum choice of ε_n, and chooses $\varepsilon_n = cb_n$ where $0 < c < 1$. For general convex Φ a corresponding prescription is not always so easy, but the fact that

$$M^{-1}(\xi^{(n+1)}) = (1 - \varepsilon_n)^{-1}[I_k + \varepsilon_n(1 - \varepsilon_n)^{-1}M^{-1}(\xi^{(n)})f(x_n)f(x_n)']^{-1}M^{-1}(\xi^{(n)}) \tag{6.4}$$

yields useful bounds on the optimum ε_n, and hence reasonable analytic prescriptions for ε_n, for many common Φ. For example, a lower bound on $\partial^2\Phi(M(\xi^{(n+1)}))/(\partial\varepsilon_n)^2$, together with the evaluation (using (2.4)) of the first derivative at $\varepsilon_n = 0$, yields an upper bound on the optimum choice of ε_n.

With obvious modifications, the above comments can be applied to such non-differentiable cases as E-optimality, where the procedures are altered by replacing ξ_{x_n} by a design on more than one point; non-convex Φ of course require additional care. Further modifications are discussed in depth in [28] and [1 a]. A principal need appears to be an efficient smoothing routine for $\xi^{(n_j)}$ to consolidate the information at certain stages n_j so that the support of $\xi^{(n)}$ is not of cardinality unbounded in n.

6C. *Bounds on departure from optimality.* The approach of [18], [1] in the case of D-optimality, for bounding the (relative) departure from optimality of a given ξ' (for example, in order to know when to stop the iterative scheme of B above), extends easily to general convex Φ. Thus, in the differentiable case one can use estimates of derivatives just as for Φ_0; for example, (2.4) and convexity yield at once the roughest (but useful) upper bound

$$\Phi(M(\xi')) - \min_\xi \Phi(M(\xi)) \leqq \bar{d}(\xi') - d^*(\xi') . \tag{6.5}$$

Thus, if $\Phi > 0$ and $\bar{d}(\xi^{(n)}) - d^*(\xi^{(n)}) < \varepsilon\Phi(M(\xi^{(n)}))$, one can stop an iterative process with the assurance that $\Phi(M(\xi^{(n)}))/\min_\xi \Phi(M(\xi)) < (1 - \varepsilon)^{-1}$.

7. **The singular case.** We have already seen, in Theorem 3 and its application to Theorems 5 and 6 (especially (4.30)—(4.31)), the complicated form by which the Φ-optimality equivalent $d^* = \bar{d}$ must be replaced if Φ is not differentiable. Of special interest are cases where the $M(\xi^*)$ being tested for optimality is singular; for it then often occurs that a convex Φ which is differentiable on $\mathscr{M}^+$ is not differentiable (where finite) on $\mathscr{M}$, and $\mathscr{D}(M, \bar{M})$ is not linear in M. In particular, if $\Phi^{(k')}$ on $\mathscr{P}_{k'}^+$ denotes as smooth a criterion as $\Phi_{p,A}^{**}$, and (partitioning as in (4.24)) $M^* = M_{11} - M_{12}M_{22}^- M_{21}$ is the "information matrix for the first k' parameters" (well-defined even if M_{22} is singular), the criterion $\Phi(M) = \Phi^{(k')}(M^*)$ has this nature (if M is singular), due to the non-linearity in M of M^*.

In such cases Theorem 3 is still valid but is of course difficult to use. The problem, then, is to translate (2.24) into more useful terms. One would hope for an analogue of (4.32) of the nonsingular non-differentiable case, but we might sometimes expect to obtain an analogue of the less satisfactory (4.30).

In the case $\Phi^{(k')} = \Phi_0$, it was shown by Kiefer [20] and by Karlin and Studden [13] (as corrected in [1]) that, in rough terms for brevity, if $\bar{d}(\xi)$ is computed from (4.24), then the sufficient condition $\bar{d} = k'$ (for D-optimality) may not be realizable, but that it is always realizable for some transformed system Af, for some nonsingular A for which $(A'^{-1}\theta)^{(1)} = \theta^{(1)}$. In both treatments it is necessary to solve an auxiliary game to find the right A; or, equivalently, to take the infimum of $\bar{d}$ over all choices of A. In any event, it is too unwieldy an approach to be useful in many problems, although it has sometimes been applied with success [1]. The analogue of this approach for general Φ will be treated in a sequel to the present paper.

In the case of D-optimality for k' parameters, an extremely useful and simple sufficient condition was obtained by Atwood, and we can duplicate it for general convex Φ which is differentiable on $\mathscr{M}^+$: Suppose we want to demonstrate the Φ-optimality of a singular $\bar{M}$ at which Φ is continuous, and that M is any element of $\mathscr{M}$ such that $M + \bar{M}$ has rank k. Define $M(\xi_\varepsilon) = (1-\varepsilon)\bar{M} + \varepsilon M$ for $0 < \varepsilon < 1$. Suppose

$$\lim_{\varepsilon \downarrow 0} [\bar{d}(\xi_\varepsilon) - d^*(\xi_\varepsilon)] = 0 . \tag{7.1}$$

Then, by (6.5) and continuity at $\bar{M}$, we conclude that $\bar{M}$ is Φ-optimum. The interchange of the operations $\lim_\varepsilon$ and $\sup_x$ in (7.1) can be treated as in [1]. Illustrations of the use of (7.1) will appear in the sequel.

8. Acknowledgments. Several people, over several years, contributed to the ideas in this paper. Leonard Gross helped, with several conversations about the analysis used. Bob Walker performed the computations mentioned in Section 6. Mark Brown suggested the problem of unknown V_x described above (5.1). A conversation with J. N. Srivastava about the rationale behind choosing a reasonable Φ, led me to consideration of criteria like (4.39). Ingram Olkin and Bill Studden contributed helpful suggestions. To these colleagues, the others whose help I may have forgotten, and the students who have had to suffer through my presentation of this material, go my sincere thanks.

REFERENCES

[1] Atwood, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist.* **40** 1570–1602.

[1a] Atwood, C. L. (1973). Sequences converging to D-optimal designs of experiments. *Ann. Statist.* **1** 342–352.

[2] Beckenbach, E. F. and Bellman, R. (1965). *Inequalities.* Springer-Verlag, Berlin.

[3] Bendat, J. and Sherman, S. (1955). Monotone and convex operator functions. *Trans. Amer. Math. Soc.* **79** 58–00.

[4] Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54** 622–654.

[5] Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24** 586–602.

[6] Dwyer, P. S. and MacPhail, M. S. (1948). Symbolic matrix derivatives. *Ann. Math. Statist.* **19** 517–534.

[6a] Eccleston, J. A. (1972). On the theory of connected designs. Thesis, Cornell Univ.

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

[7a] Elfving, G. (1959). Design of linear experiments. *Cramér Festschrift Volume.* Wiley, New York. 58–00.

[8] Fan, K. (1959). On a theorem of Weyl concerning eigenvalues of linear transformations. *Proc. Nat. Acad. Sci. USA* **35** 652–000.

[9] Farrell, R., Kiefer, J. and Walbran, A. (1965). Optimum multivariate designs. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 113–138.

[10] FEDOROV, V. V. (1969). Design of experiments for linear optimality criteria. *Theor. Probability Appl.* **16** 189–000.

[11] FEDOROV, V. V. (1969). The design of experiments in the multiresponse case. *Theor. Probability Appl.* **16** 323–000.

[12] FEDOROV, V. V. (1972). *Theory of Optimal Experiments* (Transl. and ed. by E. M. Klienko and W. J. Studden). Academic press, New York.

[12a] FEDOROV, V. V. and MALYUTOV, M. B. (1972). Optimal designs in regression problems. *Math. Operationsforsch. und Statist.* **3** 281–308.

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

[14] KARLIN, S. and STUDDEN, W. J. (1966a). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

[15] KARLIN, S. and STUDDEN, W. J. (1966b). *Tchebycheff Systems.* Interscience, New York.

[16] KARSON, M. J., MANSON, A. R. and HADER, R. J. (1969). Minimum bias estimation and experimental designs for response surfaces. *Technometrics* **11** 461–000.

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

[18] KIEFER, J. (1960). Optimum experimental designs *V*, with applications to systematic and rotatable designs. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 381–405.

[19] KIEFER, J. (1961). Optimum designs in regression problems, II. *Ann. Math. Statist.* **32** 298–325.

[20] KIEFER, J. (1962). An extremum result. *Canad. J. Math.* **12** 597–601.

[21] KIEFER, J. (1972). Optimum designs for fitting biased multiresponse surfaces. *Proc. Multivar. Symp. III.* To appear.

[21a] KIEFER, J. (1972). General optimality equivalence theory for approximate designs (abstract). *Bull. Inst. Math. Statist.* **1** 258.

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

[23] KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Matth.* **14** 363–366.

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

[25] KIEFER, J. and WOLFOWITZ, J. (1965). On a theorem of Hoel and Levine on extrapolation. *Ann. Math. Statist.* **36** 1627–1655.

[25a] MARSHALL, A. W. and OLKIN, I. (1974). Majorization in multivariate distributions. To appear in *Ann. Statist.*

[25b] MARSHALL, A. W., WALKUP, D. W. and WETS, R. J.-B. (1967). Order-preserving functions; applications to majorization and order statistics. *Pacific J. Math.* **23** 569–584.

[25c] MIRSKY, L. (1963). Results and problems in the theory of doubly stochastic matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **1** 319–334.

[26] SHAH, K. R. (1960). Optimality criteria for incomplete block designs. *Ann. Math. Statist.* **31** 791–794.

[26a] SILVEY, S. D. and TITTERINGTON, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60** 21–32.

[26b] WHITTLE, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35** 123–130.

[27] WYNN, H. P. (1970). The sequential generation of *D*-optimum experimental designs. *Ann. Math. Statist.* **41** 1655–1664.

[28] WYNN, H. P. (1972). Results in the theory and construction of *D*-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **34** 133–147.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14850