On the Duality between Locally Optimal Tests and Optimal Experimental Designs

U. Müller-Funk
Institut für Mathematische Stochastik
Albert-Ludwigs-Universität
D-7800 Freiburg im Breisgau, Federal Republic of Germany

F. Pukelsheim
Lehrstuhl für Stochastik und ihre Anwendungen
Universität Augsburg
D-8900 Augsburg, Federal Republic of Germany

and

H. Witting
Institut für Mathematische Stochastik
Albert-Ludwigs-Universität
D-7800 Freiburg im Breisgau, Federal Republic of Germany

ABSTRACT

Locally unbiased tests with maximal power curvature are determined as the solutions of an optimization problem which turns out to be of dual type as compared to the optimal design problem. In both cases the proper optimization problem is concerned with matrices only, and the transition from the matrix problem to the original variables is a separate second step. This approach provides a novel, statistical interpretation of the dual problem that arises with the optimal design problem.

1. INTRODUCTION AND SUMMARY

In this paper we present a duality relationship between the problem of finding tests with maximal power curvature, and the problem of characterizing optimal experimental designs. This duality seems to have gone unnoticed despite the fact that, for instance, Kiefer [5, p. 683] discusses the two problems side by side.

Duality has been applied to the optimal design problem for quite a while. To date, the interpretation of the ensuing dual problem has been geometrical, as a problem of minimal covering cylinders, due to Silvey and Titterington [12, 13]. Thus the present paper contributes a statistical interpretation in the sense that it is an optimal test problem which, together with the optimal design problem, forms a dual pair. The statistical appeal of this interpretation is evident: In the optimal test problem we fix a particular design, namely a differentiable model, and seek an optimal test procedure, whereas in the optimal design problem we fix the procedure, namely an F-test, and look for an optimal design.

Both problems share the feature that they come in two parts. The first part is completely in terms of the matrices which appear in the problem, and the second part is a transition to the original quantities. We now give a brief summary of this paper, and also point out some cross-references to other work in the literature.

The matrix part of the optimal test problem is given in Section 2. In Lemma 2 we derive for the matrix function $J(A)=(K'A^{-}K)^{-1}$ a novel closure property which is needed for duality considerations, and which complements the convexity and monotonicity behavior of J as set out by Pukelsheim and Styan [9]. (In the definition of J, a prime denotes transposition and a superscript minus sign generalized inversion.)

In Section 3 we discuss the duality relation between the matrix parts of the optimal test problem and the optimal design problem. This essentially parallels Section 3 of [7], but the additional closure operation on J calls for a few additional arguments, which are supplied for the convenience of the reader. The main result is the linearization of the problem inherent in the equivalence theorem (Theorem 5), with application to p-means (Corollary 5.2) and local admissibility (Corollary 5.3).

A very useful sufficient condition for the existence of an optimal matrix solution is obtained in Corollary 5.1; it is based on duality considerations and only requires the polar information functional to be strictly isotone. The latter assumption has been found useful also in Section 4 of [10]. Theorem 5 is a generalization of the particular case of maximizing Gaussian curvature, as discussed by Giri and Kiefer [3, p. 30].

In Section 4 we make the transition to the original quantities in the optimal test problem; the corresponding step for the optimal design problem was carried through in Section 4 of [7]. Theorem 6 gives the 0-1 form of a locally optimal test, thus generalizing the Neyman-Pearson fundamental lemma for the Gaussian curvature case of Isaacson [4, p. 221]. For this particular case a development along our lines was already given in [6], with applications to sequential and nonparametric testing problems. Here we conclude the paper with an example which is based on the smallest eigenvalue of the Hesse matrix of the power function, showing that a locally admissible test whose Hesse matrix is positive definite has an acceptance region which is, not just convex, but even ellipsoidal.

2. THE OPTIMAL TESTING PROBLEM

By $\mathrm{NND}(k)$ and $\mathrm{PD}(k)$ we denote the sets of all $k \times k$ matrices which are nonnegative definite and positive definite, respectively.

Let K be some $k \times s$ matrix which is kept fixed. When the null hypothesis $\Theta(K) = \{\vartheta \in \Theta \mid K'\vartheta = 0\}$ is of "linear type" and H is the $k \times k$ Hesse matrix of the power function of a test φ at $\vartheta \in \Theta(K)$, the part of interest is the $s \times s$ matrix $K'HK$. Nonnegative definiteness of this matrix still leaves open the possibility that at ϑ the power function of φ has a saddle point, rather than a minimum. The following notions seek to exclude this possibility.

Definition 1. Let K be a fixed $k \times s$ matrix of rank s. Define the set

$$\mathscr{B}(K) = \left\{ B \in \mathrm{NND}(k) \middle| \text{nullspace}\, B \cap \text{range}\, K = 0 \right\}.$$

In other words, $K'HK$ is positive definite if and only if H lies in $\mathscr{B}(K)$. It is easily seen that $\mathscr{B}(K)$ is a convex cone whose relative interior is $\mathrm{PD}(k)$, and whose closure is $\mathrm{NND}(k)$. If $s < k$ then $\mathrm{PD}(k) \subset \mathscr{B}(K) \subset \mathrm{NND}(k)$ and both inclusions are proper; if $s = k$ then $\mathrm{PD}(k) = \mathscr{B}(K)$. Obviously $\mathscr{B}(K)$ plays a role similar to the convex cone

$$\mathscr{A}(K) = \left\{ A \in \mathrm{NND}(k) \middle| \text{range}\, A \supset \text{range}\, K \right\}$$

in optimal design theory; we have $\mathscr{A}(K) \subset \mathscr{B}(K)$. The function I to be defined next yields the reduction from the grand Hesse matrix of order $k \times k$ to the smaller null hypothesis Hesse matrix of order $s \times s$.

Definition 2. The function I from $\mathrm{NND}(k)$ to $\mathrm{NND}(s)$ is defined by mapping $B \in \mathrm{NND}(k)$ into $K'BK$ if $B \in \mathscr{B}(K)$, and into 0 otherwise.

Evidently I is concave and isotone; in the testing model $I(H)$ can be interpreted as a measure of curvature off the null hypothesis at the point $\vartheta \in \Theta(K)$ and thus serves as a measure of information. In order to obtain a scalar measure of curvature we shall use the same information functionals j as in optimal design theory, i.e. real functions j on NND(s) which are

(1) nonnegative on NND(s) and positive on PD(s),
(2) positively homogeneous, and
(3) superadditive.

Thus assume $\mathscr{H}$ to be a compact convex subset of NND(k) which intersects $\mathscr{B}(K)$. Any member of $\mathscr{H}$ will be called a *Hesse matrix*. The matrix part of the optimal test problem then reads:

(P) Maximize $j \circ I(H)$ subject to $H \in \mathscr{H}$.

The *optimal value* $v = \sup_{H \in \mathscr{H}} j \circ I(H)$ is the maximal j-curvature at $\vartheta \in \Theta(K)$ which can be achieved over $\mathscr{H}$; any optimal matrix will be said to have *$\mathscr{H}$-maximal j-curvature at* $\vartheta \in \Theta(K)$.

LEMMA 1. *The composition $j \circ I$ is nonnegative on* NND(k), *positive on* $\mathscr{B}(k)$, *positively homogeneous, superadditive, concave, and isotone, and satisfies $j \circ I(0) = 0$. Furthermore $j \circ I$ is closed if and only if j vanishes outside* PD(s).

Proof. Only the last statement needs proof. Assuming $j \circ I$ to be closed, let $D_0 \in \text{NND}(s)$ be singular and for $\varepsilon > 0$ define $D_\varepsilon = D_0 + \varepsilon I_s$. The Moore-Penrose inverse K^+ satisfies $K^+ = (K'K)^{-1}K$. Thus we obtain $K'K^{+\prime}D_\varepsilon K^+K = D_\varepsilon$ and $K^{+\prime}D_\varepsilon K^+ \in \mathscr{B}(K)$, but $K^{+\prime}D_0K^+ \notin \mathscr{B}(K)$. Therefore

$$\lim_{\varepsilon \downarrow 0} j(D_\varepsilon) = \lim_{\lambda \uparrow 1} j \circ I((1-\lambda)K^{+\prime}K^+ + \lambda K^{+\prime}D_0K^+)$$

$$= j \circ I(K^{+\prime}D_0K^+) = j(0) = 0.$$

Conversely, assume j vanishes outside PD(s). If $B \in \mathscr{B}(K)$ then $K'BK$ is a continuity point of j, and therefore

$$\lim_{\varepsilon \downarrow 0} j(K'(B + \varepsilon I_k)K) = j(K'BK) = j \circ I(B).$$

If $B \notin \mathscr{B}(K)$ then $j \circ I(B) = j(0) = 0$. On the other hand the matrices $D_\varepsilon = K'(B + \varepsilon I_k)K$ tend to a singular limit matrix $D_0 = K'BK$, and so $\lim_{\varepsilon \downarrow 0} j \circ I(B + \varepsilon I_k) = \lim_{\varepsilon \downarrow 0} j(D_\varepsilon) = 0$. ■

The continuity behavior of $j \circ I$ has the following obvious consequence for the existence of optimal Hesse matrices.

THEOREM 1 (Existence). *If j vanishes outside* PD(s) *or if $\mathscr{H}$ is a subset of $\mathscr{B}(K)$, then there exists a Hesse matrix in $\mathscr{H}$ which has $\mathscr{H}$-maximal j-curvature at $\vartheta \in \Theta(K)$.*

An alternative and more interesting existence result will be established in Corollary 5.1, based on duality correspondences.

Simultaneous optimality of a matrix $H \in \mathscr{H}$ with respect to all information functionals is easily seen to be equivalent to *uniform optimality* in the sense that the matrix $K'HK$ is $\mathscr{H}$-maximal in the Loewner matrix ordering; cf. Theorem 1 in [7].

Our development in the sequel closely parallels that of the optimal design problem, as it will turn out that the optimal test problem and the optimal design problem are of dual type. However, dual problems are closed, and therefore we need the closures of the objective functions. The closure cl j of an information functional j is given by ordinary convex analysis through

$$(\mathrm{cl}\, j)(C) = \lim_{\varepsilon \downarrow 0} j(C + \varepsilon D), \qquad D \in \mathrm{PD}(s);$$

see [11, pp. 307, 57].

Though the function I from Definition 2 is not real-valued, linearity makes it so simple that the only sensible "closure" definition is

$$(\mathrm{cl}\, I)(B) = K'BK \qquad \text{for all} \quad B \in \mathrm{NND}(k).$$

Its counterpart in the optimal design problem needs slightly more attention. There, a function J is defined on NND(k) which maps A into $(K'A^{-}K)^{-1}$ if $A \in \mathscr{A}(K)$, and into 0 otherwise. Now define, for $A \in \mathrm{NND}(k)$,

$$(\mathrm{cl}\, J)(A) = \lim_{\varepsilon \downarrow 0} J(A + \varepsilon B), \qquad B \in \mathrm{PD}(k).$$

LEMMA 2. *The function* cl J *is well defined, i.e., for $A \in \mathrm{NND}(k)$ the limit $\lim_{\varepsilon \downarrow 0} J(A + \varepsilon B)$ exists and does not depend on $B \in \mathrm{PD}(k)$. If $A \in \mathscr{A}(K)$, then the matrix $(\mathrm{cl}\, J)(A)$ is equal to $J(A)$, and is nonsingular; and if $A \notin \mathscr{A}(K)$, then the matrix $(\mathrm{cl}\, J)(A)$ is singular.*

Proof. For $A \in \mathrm{NND}(k)$ and $\varepsilon > 0$ set $A_\varepsilon = A + \varepsilon I_k$. For $B \in \mathrm{PD}(k)$ monotonicity of J yields

$$J(A + \varepsilon\lambda_{\min}(B) I_k) \leqslant J(A + \varepsilon B) \leqslant J(A + \varepsilon\lambda_{\max}(B) I_k).$$

Thus if $J(A_\varepsilon)$ converges so does $J(A + \varepsilon B)$, and the limits coincide.

In case $A \in \mathscr{A}(K)$, Lemma 5.6.3 of Bandemer et al. [1] establishes $J(A_\varepsilon) \to J(A)$. It remains to consider the case $A \notin \mathscr{A}(K)$. The proof of Lemma 1 of [7] shows that

$$J(A_\varepsilon) \leqslant \frac{\Lambda + \varepsilon}{\varepsilon + z'z}\left(C_0 + \varepsilon(K'K)^{-1}\right) \to \frac{\Lambda}{z'z} C_0$$

for some $\Lambda > 0$, some nonzero $\mathbf{R}^s$-vector z, and some singular $s \times s$ matrix C_0.

It follows that $J(A_\varepsilon)$ stays bounded in norm and has singular cluster points as ε tends to 0. Along appropriate subsequences the monotonicity behavior of J shows that any two cluster points C_1 and C_2 satisfy $C_1 \leqslant C_2$ and $C_2 \leqslant C_1$, and therefore coincide. Thus $\lim_{\varepsilon \downarrow 0} J(A_\varepsilon)$ exists and is a singular matrix. ■

For the particular choice $A = KCK'$, with an arbitrary matrix $C \in \mathrm{NND}(s)$, we obtain $(\mathrm{cl}\, J)(KCK') = C$.

We note that $(\mathrm{cl}\, J)(A)$ does not, in general, coincide with $(K'A^+K)^+$; for instance, in Example 6.2.5 in [7] we have

$$(\mathrm{cl}\, J)(M_0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} = (K'M_0^+K)^{-1}.$$

Now we turn to duality results.

3. DUALITY RESULTS

The polar set $\mathscr{H}^\circ$ of $\mathscr{H}$ and the polar function j° of j are given by

$$\mathscr{H}^\circ = \left\{ A \in \mathbf{R}^{k \times k} \middle| \langle A, H \rangle \leqslant 1 \text{ for all } H \in \mathscr{H} \right\},$$

$$j^\circ(C) = \inf\left\{ \frac{\langle C, D \rangle}{j(D)} \middle| D \in \mathrm{PD}(s) \right\},$$

respectively, where $\langle A, H \rangle = \operatorname{trace} AH$. Because of the monotonicity properties inherent in the present problem it suffices to replace the set $\mathscr{H}^\circ$ by the smaller set

$$\mathscr{G} = \mathscr{H}^\circ \cap \mathrm{NND}(k).$$

We are now in a position to formulate the dual problem of the optimal test problem:

(D) Minimize $1/j^{\circ} \circ (\mathrm{cl}\, J)(G)$, subject to $G \in \mathscr{G}$.

Duality of these problems is established in two steps, by first showing that the two problems "bound each other" and then establishing Fenchel duality.

THEOREM 3 (Mutual boundedness). *For every Hesse matrix $H \in \mathscr{H}$ and for every matrix $G \in \mathscr{G}$ one has $j \circ I(H) \leqslant 1/j^{\circ} \circ (\mathrm{cl}\, J)(G)$, with equality if and only if H lies in $\mathscr{B}(K)$ and the following conditions are satisfied with $C = (\mathrm{cl}\, J)(G)$ and $D = K'HK$:*

(1) $\operatorname{trace} GH = 1$,
(2) $GH = KCK'H$,
(3) $j^{\circ}(C) j(D) = \operatorname{trace} CD$.

Proof. The proof of Theorem 3 in [7] carries over from J to cl J, with the following modifications. Suppose $H \in \mathscr{B}(K)$. Observe that the relative interior of $\mathscr{G}$ consists of positive definite matrices. Choose a matrix B in the relative interior of $\mathscr{G}$, and for $\varepsilon \in [0,1]$ define $G_{\varepsilon} = (1-\varepsilon)G + \varepsilon B \in \mathrm{PD}(k)$. Then

$$1 \geqslant \langle G_{\varepsilon}, H \rangle \geqslant \left\langle J(G_{\varepsilon}), K'HK \right\rangle \geqslant j^{\circ} \circ J(G_{\varepsilon}) \cdot j(K'HK),$$

and a transition to the limit as ε tends to 0 establishes

$$1 \geqslant \langle G, H \rangle \geqslant \langle C, D \rangle \geqslant j^{\circ}(C) \cdot j(D).$$

Equality in the first and third inequalities is equivalent to conditions (1) and (3). Obviously (2) implies $\langle G, H \rangle = \langle C, D \rangle$, and so it remains to establish the converse direction. Assume, then, that $\langle G, H \rangle = \langle C, D \rangle$. Then, as ε tends to 0,

$$0 = \lim \langle G_{\varepsilon}, H \rangle - \langle J(G_{\varepsilon}), K'HK \rangle$$

$$= \lim \left\| G_{\varepsilon}^{1/2} H^{1/2} - G_{\varepsilon}^{-1/2} K J(G_{\varepsilon}) K' H^{1/2} \right\|^{2},$$

and

$$0 = \lim G_{\varepsilon}^{1/2} H^{1/2} - G_{\varepsilon}^{-1/2} K J(G_{\varepsilon}) K' H^{1/2}.$$

Premultiplication with $G^{1/2} = \lim G_{\varepsilon}^{1/2}$ and postmultiplication with $H^{1/2}$ then yields (2). ∎

THEOREM 4 (Duality). $\sup_{H \in \mathcal{H}} j \circ I(H) = \min_{G \in \mathcal{G}} 1/j^\circ \circ (\text{cl}\, J)(G)$.

Proof. The result is a special case of the Fenchel duality theorem, as is Theorem 4 in [7]. ∎

The following reformulation eliminates the explicit appearance of the dual problem.

THEOREM 5 (Equivalence). *Let $H \in \mathcal{H}$ be a Hesse matrix which lies in $\mathcal{B}(K)$, and set $D = K'HK$. Then H has $\mathcal{H}$-maximal j-curvature at $\vartheta \in \Theta(K)$ if and only if there exists a matrix $C \in \text{NND}(s)$ with the properties that*

$$j^\circ(C)j(D) = \text{trace}\, CD = 1,$$

and

$$\text{trace}\, CK'BK \leqslant 1 \qquad \textit{for all} \quad B \in \mathcal{H}.$$

Proof. For the direct part, let $G \in \mathcal{G}$ be an optimal solution of the dual problem. Take $C = (\text{cl}\, J)(G)$. Then conditions (1), (2), (3) yield $j^\circ(C)j(D) = \text{trace}\, CD = 1$. The inequalities $\text{trace}\, CK'BK \leqslant \text{trace}\, GB \leqslant 1$, for all $B \in \mathcal{H}$, were established in the course of the proof of Theorem 3.

For the converse direction, define $G = KCK'$. By assumption then $G \in \mathcal{G}$; we have $(\text{cl}\, J)(G) = C$. But then $j^\circ \circ (\text{cl}\, J)(G) = j^\circ(C) = 1/j(D)$, whence both G and H are optimal solutions of their respective problems. ∎

We now draw a number of useful conclusions from these duality results. First we give a new sufficient condition for the existence of optimal solutions, based on duality.

COROLLARY 5.1 (Existence). *If j° is strictly isotone, then there exists a Hesse matrix in $\mathcal{H}$ which has $\mathcal{H}$-maximal j-curvature at $\vartheta \in \Theta(K)$.*

Proof. Consider the "closure" of the primal problem: There always exists some matrix $H \in \mathcal{H}$ which maximizes $(\text{cl}\, j)(K'BK)$ for $B \in \mathcal{H}$. We shall show that $D = K'HK$ is positive definite, whence $H \in \mathcal{B}(K)$. Then $(\text{cl}\, j)(K'HK) = j(K'HK) = j \circ I(H)$, and therefore H has $\mathcal{H}$-maximal j-curvature at $\vartheta \in \Theta(K)$, as asserted.

Recall that $(\text{cl}\, j)^\circ = j^\circ$ (see [8] Section 3). Thus for the "closure" of the primal problem the equivalence theorem states that for H to be optimal it is necessary that there exists a matrix $C \in \mathrm{NND}(s)$ such that $j^\circ(C)(\mathrm{cl}\, j)(D) = \mathrm{trace}\, CD$. Now suppose $z \in \mathbf{R}^s$ is a null vector of D, i.e., $Dz = 0$. Then, for all $\alpha > 0$,

$$j^\circ(C + \alpha zz')(\mathrm{cl}\, j)(D) \leqslant \langle C + \alpha zz', D\rangle = \langle C, D\rangle = j^\circ(C)(\mathrm{cl}\, j)(D)$$
$$\leqslant j^\circ(C + \alpha zz')(\mathrm{cl}\, j)(D).$$

But in view of strict monotonicity of j°, the function $h(\alpha) = j^\circ(C + \alpha zz')$ can be constant only if $z = 0$. Hence D is positive definite. ■

As an illustration consider the information functionals j_p, i.e. the generalized means of order p of the eigenvalues. Their polar function is sj_q, where q is determined from $pq = p + q$. Therefore j_p has a strictly isotone polar function provided $p \in [-\infty, +1)$. In other words, for $p < 1$ there always exists a Hesse matrix $H \in \mathscr{H}$ with $\mathscr{H}$-maximal j_p-curvature at $\vartheta \in \Theta(K)$.

Remark. A similar statement holds true for the existence of optimal information matrices in the optimal design problem. Indeed, the examples in [7] on the nonexistence of optimal designs all utilize the information functional j_1 whose polar function $sj_{-\infty}$ fails to be strictly isotone. Notice the difference in assumptions as compared to Corollary 5.3 in [7] on multiplicity of optimal designs: There strict monotonicity is a property which pertains to the primal objective functional, whereas here we have imposed it on the dual objective functional.

Corollary 5.2 (j_p-optimality). *Let H be a Hesse matrix which lies in $\mathscr{B}(K)$. If $p > -\infty$, then H has $\mathscr{H}$-maximal j_p-curvature at $\vartheta \in \Theta(K)$ if and only if*

$$\mathrm{trace}(K'HK)^{p-1}K'BK \leqslant \mathrm{trace}(K'HK)^p \qquad \text{for all} \quad B \in \mathscr{H}.$$

If $p = -\infty$ and S is the set of all $s \times s$ matrices of the form zz' such that z is a normalized eigenvector of $K'HK$ corresponding to $\lambda_{\min}(K'HK)$, then H has $\mathscr{H}$-maximal $j_{-\infty}$-curvature at $\vartheta \in \Theta(K)$ if and only if there exists a matrix $F \in \mathrm{conv}\, S$ such that

$$\mathrm{trace}\, FK'BK \leqslant \lambda_{\min}(K'HK) \qquad \text{for all} \quad B \in \mathscr{H}.$$

Simultaneous j_p-optimality and universal optimality in the optimal test problem may now be discussed quite similarly to those in the optimal design problem. Let H be a Hesse matrix which lies in $\mathscr{H}(K)$, and suppose that H is *balanced*, i.e., $K'HK = \rho I_s$ for some $\rho > 0$. If H has optimal j_{p_0}-curvature for some $p_0 > -\infty$, then it has optimal j_p-curvature simultaneously for all $p \in [-\infty, +1]$. If H has optimal j_1-curvature, i.e., trace-curvature, then it is universally optimal in a sense quite similar to Kiefer's notion of universal optimality for optimal designs.

Finally we list the corresponding results on admissibility. We shall call a Hesse matrix $H \in \mathscr{H}$ locally admissible if no other Hesse Matrix $B \in \mathscr{H}$ satisfies $B \geqslant H$.

COROLLARY 5.3 (Admissibility). *Let $H \in \mathscr{H}$ be a Hesse matrix. If there exists a positive definite matrix $G \in \mathscr{G}$ such that* $\operatorname{trace} GH = 1$, *then H is locally admissible. If H is locally admissible and has maximal rank in $\mathscr{H}$, then*

(i) *H has $\mathscr{H}$-maximal $j_{-\infty}$-curvature at $\vartheta \in \Theta(K)$ whenever $K^{+\prime}K^{+} = H$,*
(ii) *there exists a matrix $G \in \mathscr{G}$ such that* $\operatorname{trace} GH = 1$, *and*
(iii) *H has $\mathscr{H}$-maximal j_1-curvature at $\vartheta \in \Theta(K)$ whenever $KK' = G$.*

Proof. Suppose $G \in \mathrm{PD}(k)$ is such that $\operatorname{trace} GB \leqslant 1 = \operatorname{trace} GH$ for all $B \in \mathscr{H}$. If $B \geqslant H$, then $1 \geqslant \operatorname{trace} GB \geqslant \operatorname{trace} GH = 1$ entails $\operatorname{trace} G(B - H) = 0$ and $B = 0 = H$, whence H is locally admissible.

Now let H be locally admissible. (i): Let s be the rank of H. The $k \times s$ matrix K satisfies $K^{+\prime}K^{+} = H$ if and only if K satisfies $K'HK = I_s$. Evidently $H \in \mathscr{B}(K)$. Let $B \in \mathscr{H}$ have $\mathscr{H}$-maximal $j_{-\infty}$-curvature at $\vartheta \in \Theta(K)$, and let G be an optimal solution of the dual problem. Then

$$\frac{1}{\lambda_{\min}(K'BK)} = \frac{1}{j_{-\infty} \circ I(B)} = s j_1 \circ (\operatorname{cl} J)(G) \leqslant \operatorname{trace} GH \leqslant 1,$$

where the first inequality follows from

$$(\operatorname{cl} J)(G) = \lim_{\varepsilon \downarrow 0} \left(K'G_\varepsilon^{-1}K\right)^{-1} \leqslant \lim_{\varepsilon \downarrow 0} K^{+}G_\varepsilon K^{+\prime} = K^{+}GK^{+\prime}.$$

Hence $1 \leqslant \lambda_{\min}(K'BK)$, and $I_s \leqslant K'BK$. Postmultiplication by K^{+} and premultiplication by its transpose yields $H \leqslant KK^{+}BKK^{+} = HH^{+}BHH^{+} = B$, the last equality being a consequence of rank maximality. Admissibility forces $H = B$, whence H is $j_{-\infty}$-optimal at $\vartheta \in \Theta(K)$, and (ii) $\operatorname{trace} GH = 1$.

(iii): Let s be the rank of G. We have $j_1 \circ I(H) = (1/s) \operatorname{trace} K'HK = (1/s) \operatorname{trace} GH = 1/s$, while at the same time $J(G) = (K'G^{-}K)^{-1} =$

$(K'(KK')^{-}K)^{-1} = I_s$ entails $1/(sj_{-\infty} \circ J(G)) = 1/s$. Hence both H and G are optimal solutions of their respective problems. ■

The main contribution of the above results is that they provide a linearization of the original, possibly nonlinear problem. Accordingly, when the original information function j itself is linear already, as is the case with j_1 or, more generally, with $j_L(D) = \operatorname{trace} DL$, then the results above reduce to mere tautologies. For nonlinear functions j, however, this linearization greatly facilitates the checking of optimality. In the optimal test problem it opens the way for an application of the Neyman-Pearson lemma, and to this we turn next.

4. LOCALLY OPTIMAL TESTS

Let $\mathfrak{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$ be a k-parameter family of distributions on a sample space $\mathscr{X}$ with sigma algebra $\mathscr{B}$. Thus the parameter set Θ is taken to be a subset of $\mathbf{R}^k$; suppose the null hypothesis Θ_0 lies in the interior of Θ and is of "linear type," i.e., there exists some $k \times s$ matrix K of rank s such that

$$\Theta_0 = \{\vartheta \in \Theta | K'\vartheta = 0\} = \Theta(K), \qquad \text{say.}$$

Denote by Φ the set of all tests φ, ψ on $(\mathscr{X}, \mathscr{B})$.

We assume $\mathfrak{P}$ to be twice $L_1(P_\vartheta)$-differentiable, for all $\vartheta \in \Theta(K)$, with first derivative $\dot{L}_\vartheta(x)$, a k-vector of P_ϑ-integrable functions, and with second derivative $\ddot{L}_\vartheta(x)$, a symmetric $k \times k$ matrix of P_ϑ-integrable functions. Definitions and properties of this type of differentiability are given in [6] and discussed in full detail in [14], so we only quote its properties as needed in this paper:

$$E_\vartheta[\dot{L}_\vartheta(x)] = 0, \qquad E_\vartheta[\ddot{L}_\vartheta(x)] = 0,$$

and

$$\nabla E_\vartheta[\psi(x)] = E_\vartheta[\psi(x)\dot{L}_\vartheta(x)],$$
$$\nabla\nabla E_\vartheta[\psi(x)] = E_\vartheta[\psi(x)\ddot{L}_\vartheta(x)],$$

for all tests $\psi \in \Phi$, where ∇ denotes differentiation with respect to ϑ.

Let $\alpha \in (0,1)$ be a given level of significance, and choose an information functional j. A locally unbiased level-α test φ for $\Theta(K)$ is said to have *maximal j-curvature at* $\vartheta \in \Theta(K)$ if φ is a solution of the problem

$$\begin{aligned}
&\text{Maximize} \quad && j \circ I(E_\vartheta[\psi(x)\ddot{L}_\vartheta(x)]) \\
&\text{subject to} \quad && E_\vartheta[\psi(x)\ddot{L}_\vartheta(x)] \geqslant 0, \\
& && E_\eta[\psi(x)\dot{L}_\eta(x)] = 0 \quad \text{for all} \quad \eta \in \Theta(K), \\
& && E_\eta[\psi(x)] = \alpha \quad \text{for all} \quad \eta \in \Theta(K), \\
& && \psi \in \Phi.
\end{aligned}$$

If there happens to be an optimal solution φ which is the same for all $\vartheta \in \Theta(K)$, then φ will be called *locally j-optimal for* $\Theta(K)$; in the one-parameter case ($k = 1$) such a test φ is locally most powerful for $\vartheta = 0$.

The relation to the matrix maximization problem in Section 2 becomes evident upon choosing $\mathscr{H}$ to be the set of nonnegative definite matrices $E_\vartheta[\psi(x)\ddot{L}_\vartheta(x)]$ obtained from tests $\psi \in \Phi$ with

$$E_\eta[\psi(x)\dot{L}_\eta(x)] = 0 \quad \text{and} \quad \mathrm{E}_\eta[\psi(\mathrm{x})] = \alpha \qquad \textit{for all} \quad \eta \in \Theta(\mathrm{K}).$$

There exists at least one such test, namely the constant α, giving $0 \in \mathscr{H}$. Moreover, $\mathscr{H}$ is convex and bounded, and a weak compactness argument shows that it is also closed. We make the assumption that there exists at least one feasible test $\hat{\psi}$ whose Hesse matrix at ϑ is positive definite:

$$\mathscr{H} \cap \mathrm{PD}(k) \neq \varnothing.$$

This assumption is needed for Theorem 6 below, and it implies that $\mathscr{H}$ intersects $\mathscr{B}(K)$, as required in Section 2.

The basic linearization which helps to determine optimal tests φ is given by the equivalence theorem (Theorem 5). Namely, put

$$H = E_\vartheta[\varphi(x)\ddot{L}_\vartheta(x)];$$

assume $H \in \mathscr{B}(K)$, and with the matrix C from Theorem 5 define

$$\begin{aligned}
f_C(\psi) &= E_\vartheta[\psi(x)\,\mathrm{trace}\, CK'\ddot{L}_\vartheta(x)K] \\
&= \mathrm{trace}\, CK'E_\vartheta[\psi(x)\ddot{L}_\vartheta(x)]\,K.
\end{aligned}$$

Theorem 5 then requires $f_C(\psi) \leqslant 1 = f_C(\varphi)$. Thus φ has maximal j-curvature at $\vartheta \in \Theta(K)$ if and only if φ is a solution of

$$
\begin{aligned}
&\text{Maximize} && f_C(\psi) \\
&\text{subject to} && E_\vartheta\left[\psi(x)\ddot{L}_\vartheta(x)\right] \geqslant 0, \\
& && E_\eta\left[\psi(x)\dot{L}_\eta(x)\right] = 0 \qquad \text{for all} \quad \eta \in \Theta(K), \\
& && E_\eta\left[\psi(x)\right] = \alpha \qquad \text{for all} \quad \eta \in \Theta(K), \\
& && \psi \in \Phi.
\end{aligned}
$$

Since the constraints which force local unbiasedness depend on all of the composite hypothesis $\Theta(K)$, an appropriate version of the Neyman-Pearson lemma would call for some least favorable mixing distribution on $\Theta(K)$. To keep the exposition simple we restrict attention to the case of a simple null hypothesis $\Theta(I_k) = \{0\}$.

THEOREM 6 (Fundamental lemma). *A test $\varphi \in \Phi$ is a locally unbiased level-α test for $\vartheta = 0$ with maximal j-curvature at $\vartheta = 0$ if and only if the matrix $H = E_0[\varphi(x)\ddot{L}_0(x)]$ is positive definite and there exists a matrix $C \in \mathrm{NND}(k)$ with the properties that $j^0(C)j(H) = \operatorname{trace} CH = 1$ and that, for P_0-almost all x,*

$$
\varphi(x) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \qquad \text{for} \quad \operatorname{trace} C\ddot{L}_0(x) \begin{Bmatrix} > \\ < \end{Bmatrix} \lambda'\dot{L}_0(x) + \lambda_0,
$$

where the Lagrangian multipliers $\lambda \in \mathbf{R}^k$ and $\lambda_0 \in \mathbf{R}$ are determined from

$$
E_0\left[\varphi(x)\dot{L}_0(x)\right] = 0, \qquad E_0\left[\varphi(x)\right] = \alpha.
$$

Proof. As pointed out above, φ is optimal if and only if φ has a positive definite Hesse matrix H and φ solves the auxiliary problem:

$$
\begin{aligned}
&\text{Maximize} && f_C(\psi) = E_0\left[\psi(x) \operatorname{trace} C\ddot{L}_0(x)\right] \\
&\text{subject to} && E_0\left[\psi(x)\ddot{L}_0(x)\right] \geqslant 0, \\
& && E_0\left[\psi(x)\dot{L}_0(x)\right] = 0, \\
& && E_0\left[\psi(x)\right] = \alpha, \\
& && \psi \in \Phi.
\end{aligned}
$$

As a first constraint qualification we have assumed that there exists a test $\hat{\psi} \in \Phi$ such that $E_0[\hat{\psi}] = \alpha$ and $E_0[\hat{\psi}\dot{L}_0] = 0$, with positive definite Hesse matrix $E_0[\hat{\psi}\ddot{L}_0]$. As a second constraint qualification we need that $\binom{\alpha}{0}$ is an interior point of the set R of all points

$$r(\psi) = \begin{pmatrix} E_0[\psi] \\ E_0[\psi \dot{L}_0] \end{pmatrix} \qquad \text{for} \quad \psi \in \Phi.$$

This is true for $k = 1$: If $\dot{L}_0 = 0$, then it does not contribute to the constraints and there is nothing to show. If $\dot{L}_0 \neq 0$, then R contains the four points

$$r(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad r(1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$r(1[\dot{L}_0 > 0]) = \begin{pmatrix} \varepsilon \\ \delta \end{pmatrix}, \quad \text{say},$$

and

$$r(1[\dot{L}_0 < 0]) = \begin{pmatrix} 1 - \varepsilon \\ -\delta \end{pmatrix},$$

whence

$$\begin{pmatrix} \alpha \\ 0 \end{pmatrix} \in \operatorname{int} R.$$

For $k > 1$ we now make the mild assumption that $\dot{L}_0$ falls into any one of the 2^k orthants *with positive P_0-probability*. The foregoing argument then easily generalizes and proves $\binom{\alpha}{0}$ to be in the interior of R.

Now standard duality theory (e.g. [2, Theorem 6.2.4]) tells us that φ satisfies, for every $k \times k$ matrix $\Lambda \geqslant 0$, every vector $\lambda \in \mathbf{R}^k$, and every scalar $\lambda_0 \in \mathbf{R}$,

$$\begin{aligned} f_C(\varphi) &\leqslant E_0[\varphi \operatorname{trace} C\ddot{L}_0] + E_0[\varphi \operatorname{trace} \Lambda \ddot{L}_0] - E_0[\varphi \lambda' \dot{L}_0] + \lambda_0(\alpha - E_0[\varphi]) \\ &\leqslant \lambda_0 \alpha + \sup_{\psi \in \Phi} E_0[\psi \{\operatorname{trace}(C + \Lambda)\ddot{L}_0 - \lambda' \dot{L}_0 - \lambda_0\}] \\ &= \lambda_0 \alpha + E_0[\{\operatorname{trace}(C + \Lambda)\ddot{L}_0 - \lambda' \dot{L}_0 - \lambda_0\}_+] \\ &= g(\Lambda, \lambda, \lambda_0), \quad \text{say}. \end{aligned}$$

Furthermore, φ is an optimal solution of the auxiliary problem if and only if $f_C(\varphi) = g(\Lambda, \lambda, \lambda_0)$ for some $\Lambda, \lambda, \lambda_0$, which, in turn, is equivalent to $0 = E_0[\varphi \operatorname{trace} \Lambda \ddot{L}_0] = \operatorname{trace} \Lambda H$, i.e., $\Lambda = 0$, and φ being of 0-1 form as stated in the theorem. ■

COROLLARY 6.1 (j_p-optimality). *Let* $\varphi \in \Phi$ *be a locally unbiased level-*α *test for* $\vartheta = 0$, *with positive definite Hesse matrix* $H = E_0[\varphi(x)\ddot{L}_0(x)]$. *Fix* $p \in (-\infty, 1]$. *Then* φ *has maximal* j_p*-curvature at* $\vartheta = 0$ *if and only if, for some* $\lambda \in \mathbf{R}^k$, $\lambda_0 \in \mathbf{R}$ *and for* P_0*-almost all* x,

$$\varphi(x) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \qquad \textit{for} \quad \operatorname{trace} H^{p-1}\ddot{L}_0(x) \begin{Bmatrix} > \\ < \end{Bmatrix} \lambda'\dot{L}_0(x) + \lambda_0.$$

Applications to tests with maximal Gaussian (i.e., j_0-) curvature are given in [6]. We here only point out another interesting relationship, using $j_{-\infty}$ as a measure of curvature. Suppose the underlying family $\mathfrak{P}$ is exponential in ϑ and $T(x)$. Let τ_0 and Σ_0 be the mean vector and dispersion matrix of T under $\vartheta = 0$. Then

$$\dot{L}_0(x) = T(x) - \tau_0,$$

and

$$\ddot{L}_0(x) = [T(x) - \tau_0][T(x) - \tau_0]' - \Sigma_0.$$

Corollary 6.1 implies that the test with maximal j_p-curvature at $\vartheta = 0$ has ellipsoidal acceptance region consisting for some $a \in \mathbf{R}^k$ and $c \in \mathbf{R}$ of all x such that

$$[T(x) - a]'H^{p-1}[T(x) - a] < c.$$

If $p = -\infty$, then H^{p-1} has to be replaced by some matrix $C \in \operatorname{conv} S$ as in Corollary 5.2, but the acceptance region remains ellipsoidal. This, in conjunction with Corollary 5.3, proves that every locally admissible test with positive definite Hesse matrix at $\vartheta = 0$ has an acceptance region which is, not just convex, but ellipsoidal.

REFERENCES

1 H. Bandemer et al., *Theorie und Anwendung der Optimalen Versuchsplanung I*, Akademie-Verlag, Berlin, 1977.

2 M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming, Theory and Algorithms*, Wiley, New York, 1979.

3 N. Giri and J. Kiefer, Local and asymptotic minimax properties of multivariate tests, *Ann. Math. Statist.* 35:21–35 (1964).

4 S. L. Isaacson, On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters, *Ann. Math. Statist.* 22:217–234 (1951).

5 J. Kiefer, On the nonrandomized optimality and randomized nonoptimality of symmetrical designs, *Ann. Math. Statist.* 29:675–699 (1958).

6 U. Müller-Funk, F. Pukelsheim, and H. Witting, Locally most powerful tests for two-sided hypotheses, in *Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics*, Bad Tatzmannsdorf, 4–9 Sept. 1983.

7 F. Pukelsheim, On linear regression designs which maximize information, *J. Statist. Plann. Inference* 4:339–364 (1980).

8 F. Pukelsheim, On information functions and their polars, *J. Optim. Theory Appl.* 41:533–546 (1983).

9 F. Pukelsheim and G. P. H. Styan, Convexity and monotonicity properties of dispersion matrices of estimators in linear models, *Scand. J. Statist.* 10:145–149 (1983).

10 F. Pukelsheim and D. M. Titterington, General differential and Lagrangian theory for optimal experimental design, *Ann. Statist.* 11:1060–1068 (1983).

11 R. T. Rockafellar, *Convex Analysis*, Princeton U.P., Princeton, N.J., 1970.

12 S. D. Silvey and D. M. Titterington, A geometric approach to optimal design theory, *Biometrika* 60:21–32 (1973).

13 D. M. Titterington, Optimal design: Some geometrical aspects of *D*-optimality, *Biometrika* 62:313–320 (1975).

14 H. Witting, *Mathematische Statistik I, Parametrische Verfahren bei festem Stichprobenumfang*, Teubner, Stuttgart, 1984.