Journal of Statistical Planning and Inference 4 (1980) 339–364.

ON LINEAR REGRESSION DESIGNS WHICH MAXIMIZE INFORMATION

Friedrich PUKELSHEIM
Universität Freiburg im Breisgau, Fed. Rep. Germany

Received 27 December 1979; revised manuscript received 4 June 1980
Recommended by J. Srivastava and J. Kiefer.

Abstract: Necessary and sufficient conditions are established when a continuous design contains maximal information for a prescribed s-dimensional parameter in a classical linear model. The development is based on a thorough study of a particular dual problem and its interplay with the optimal design problem, extending partial results and earlier approaches based on differential calculus, game theory, and other programming methods. The results apply in particular to a class of information functionals which covers c-, D-, A-, L-optimality, they include a complete account of the non-differentiable criterion of E-optimality, and provide a constructive treatment of those situations in which the information matrix is singular. Corollaries pertain to the case of s out of k parameters, simultaneous optimality with respect to several criteria, multiplicity of optimal designs, bounds on their weights, and optimality which is induced by admissibility.

AMS 1980 Subject Classification: Primary 62K05; Secondary 65K05.

Key words: Approximate Design Theory; General Equivalence Theorems; c-, D-, A-, L-, E-optimality; Polar Information Functionals; Minimal Cylinder Problems; Singular Information Matrices; Admissible Designs.

1. Introduction

Convex programming methods are applied to the approximate, or continuous design theory of classical linear models. Emphasis is on characterizing those designs which provide maximal information on the unknown parameter, in contrast to the customary approach which prefers to minimize some kind of loss. Although the distinction between maximizing information and minimizing loss seems marginal, the information point of view has led to a consistency of exposition which I failed to gain otherwise.

Our investigations synthesize partial results and previous approaches based on differential calculus, game theory, and programming methods. In particular, they resolve the two open problems of characterizing optimality of singular information matrices, and of characterizing optimality with respect to non-differentiable criteria, such as E-optimality. More general, our approach covers Kiefer's (1974) Φ_p-optimality, Fedorov's (1972) L-optimality, and other familiar concepts such as c-, D-, A-optimality for all, or for s out of k parameters. We shall now give a short summary, and then list the notation to be used in the sequel.

Section 2 introduces the optimal design problem (P) as one of maximizing an information functional over a compact convex set $\mathfrak{M}$ of information matrices, for a fixed s-dimensional linear parameter $K'\beta$. More precisely, the objective function is constructed by first reducing any $k \times k$ information matrix $M \in \mathfrak{M}$ to a $s \times s$ matrix $J(M)$, called information matrix for $K'\beta$, and then mapping $J(M)$ into a non-negative real number by means of a positively homogeneous and concave function j, called information functional. Simultaneous optimality with respect to all information functionals leads to uniform optimality (Theorem 1), existence of optimal information matrices is settled by semi-continuity (Theorem 2). Section 3 characterizes optimality using duality theory of convex analysis. The dual problem (D) amounts to maximizing the polar information functional over the polar set of information matrices, this problem being considerably smoother than the primal problem (P) in that the new objective function does not require any matrix inversion and always is semi-continuous. Our main result is the detailed duality relation, established in Theorems 3 and 4. An optimality characterization closer to the well known Kiefer–Wolfowitz type equivalence theorems is deduced in Theorem 5, with corollaries on c-, and U-optimality, and on possible multiplicity of optimal information matrices.

Section 4 further specifies the optimal design problem in the usual manner, i.e., by generating the information matrices M as moment matrices of design measures ξ. The interplay of design measures and their information matrices is studied by bounding the number of support points of ξ given M (Theorem 6), identifying possible support points of design measures which are optimal (Theorem 7), and computing the weights that an optimal design assigns to its support points (Corollary 7.1).

In Section 5 these results are applied to the j_p-family of information functionals which correspond to Kiefer's Φ_p-criteria (Theorem 8). Corollaries pertain to the case of s out of k parameters, linear parameters $K'\beta$ which do not have full rank s, simultaneous optimality with respect to all j_p-criteria, and optimality induced by admissibility. Section 6 concludes the paper with some examples.

In this paper all matrices are real matrices. The following notation will be used throughout:

$\mathbb{R}^n$	Euclidean n-space of column vectors
$\mathbb{R}^{k\times s}$	the linear space of all $k \times s$ matrices
A', A^-, A^+	the transpose, an arbitrary g-inverse, the Moore–Penrose inverse of a matrix A
$\langle A, B\rangle$	trace $A'B$, the Euclidean matrix inner product
$\lVert A\rVert$	$\sqrt{\text{trace } A'A}$, its associated norm
$\mathrm{Sym}(k)$	the linear space of all symmetric $k \times k$ matrices
$\lambda_{\min}(A)$, $\lambda_{\max}(A)$	the smallest and largest eigenvalue of a symmetric matrix A
$\mathrm{NND}(k)$	the closed convex cone of all symmetric non-negative definite $k \times k$ matrices

$\left.\begin{array}{l} A \geqslant B \\ B \leqslant A \end{array}\right\}$	if and only if $A-B$ is non-negative definite, the Loewner ordering of symmetric matrices
$\mathrm{PD}(k)$	the relatively open convex cone of all symmetric positive definite $k \times k$ matrices

A real function j is called super-additive if $j(C+D) \geqq j(C)+j(D)$; a concave function j is said to be closed if j is upper semi-continuous.

2. Information functionals

Consider a classical linear model $Y \sim (X\beta; \sigma^2 I_n)$ in which n real observations $Y_1, \ldots, Y_n$ form a random $\mathbb{R}^n$-vector Y which has mean vector $X\beta$ and dispersion matrix $\sigma^2 I_n$. The $n \times k$ matrix X may be chosen by the experimenter prior to drawing the observations, as will be specified in Section 4. Among all linear functions $K'\beta$ of the vector parameter $\beta \in \mathbb{R}^k$ only those are of interest which are identifiable in the model $Y \sim (X\beta; \sigma^2 I_n)$. This means that when the mean vector of Y may be represented as both $X\beta$ and $X\gamma$ with two values $\beta, \gamma \in \mathbb{R}^k$, then $K'\beta$ and $K'\gamma$ must coincide. The following definition is adapted to the present situation in which K is fixed and X may vary.

Definition 1. Let K be a fixed $k \times s$ matrix of rank s. Then the set $\mathfrak{A}(K)$ is defined *to consist of all those matrices* $A \in \mathrm{NND}(k)$ *whose range contains the range of* K.

Hence $K'\beta$ is identifiable in the model $Y \sim (X\beta; \sigma^2 I_n)$ if and only if $X'X$ lies in $\mathfrak{A}(K)$. As is easily seen the set $\mathfrak{A}(K)$ is a convex cone, its relative interior is $\mathrm{PD}(k)$, and its closure is $\mathrm{NND}(k)$. If $s<k$ then both inclusions $\mathrm{PD}(k) \subset \mathfrak{A}(K) \subset \mathrm{NND}(k)$ are proper and $\mathfrak{A}(K)$ is neither relatively open nor closed, if $s=k$ then $\mathfrak{A}(K)$ equals $\mathrm{PD}(k)$. The following notions aim to distinguish one model from another by the different amount of information they contain about $K'\beta$.

Definition 2. Define J to be *the function from* $\mathrm{NND}(k)$ *into* $\mathrm{NND}(s)$ *which maps* A *into* $(K'A^-K)^{-1}$ *if* $A \in \mathfrak{A}(K)$, *and into* 0 *otherwise.*

It is well known that the function J is well defined, concave, and isotone, see Pukelsheim and Styan (1979, Theorem 1), Gaffke and Krafft (1979, Theorem 4.8). In a normal model $Y \sim \mathfrak{N}_n(X\beta; \sigma^2 I_n)$ the *information matrix for* $K'\beta$ is $J(X'X)/\sigma^2$, in the sense that it provides the Cramér–Rao bound for unbiased estimation of $K'\beta$, see Rao (1973, Section 5a.3), and that it determines the power of the F-test for $K'\beta=0$, see Krafft (1978, Satz 20.1). Information matrices associated with different design matrices X need not be comparable in the Loewner-ordering $\leqslant$, and this suggests to study real functions j of J, besides J itself. Certainly j should have properties which appropriately relate to the concept of information.

Definition 3. A real function j on NND(s) will be called an *information functional* if j is

(a) non-negative on NND(s), and positive on PD(s),
(b) positively homogeneous, and
(c) super-additive.

As a consequence every information functional j is concave and isotone, and satisfies $j(0)=0$. If j is strictly concave then it is also strictly isotone, i.e., $C \leqslant D$ and $C \neq D$ imply $j(C) < j(D)$. However, $j(C) = \text{trace } C$ is a strictly isotone information functional which fails to be strictly concave. Other information functionals are:

$$\begin{aligned} j_L(C) &= \text{trace } CL, && (L \in \text{NND}(s), L \neq 0). \\ j_p(C) &= (\text{trace } C^p/s)^{1/p}, && (p \leqq 1, p \neq 0), \\ j_0(C) &= (\det C)^{1/s}, \\ j_{-\infty}(C) &= \lambda_{\min}(C). \end{aligned}$$

The family $\{j_p \mid p \in [-\infty, +1]\}$ leads to the Φ_p-criteria of Kiefer (1974, eq. (4.18)), j_1 is also implicit in Silvey and Titterington (1974, p. 301), Kiefer (1975, p. 338), and Titterington (1975), (1980). For other discussions of general optimality criteria see Kiefer (1974), Silvey (1978), Gaffke (1979).

In general, the average of an arbitrary collection of information functionals, and the minimum of a finite collection again are information functionals. Even the infimum of an arbitrary collection is an information functional, provided it is positive on PD(s). This applies in particular to the *polar function* j^0 of an information functional j, which is given by

$$j^0(D) = \inf\{\langle C, D\rangle / j(C) \mid C \in \text{PD}(s)\}.$$

For when D is positive definite then $j^0(D)$ is bounded from below by

$$\inf\{\langle C, D\rangle \mid C \in \text{NND}(s), \|C\| = 1\} / \sup\{j(C) \mid C \in \text{PD}(s), \|C\| = 1\} > 0.$$

Moreover j^0 is closed, being the infimum of a collection of upper semi-continuous functions.

More is needed than just closedness of j, in order to ensure closedness of the composition $j \circ J$. An information functional j will be said to *vanish outside* PD(s) if $\lim_{\varepsilon \searrow 0} j(C_0 + \varepsilon I_s) = 0$ for all singular matrices $C_0 \in \text{NND}(s)$.

Lemma 1. *The composition $j \circ J$ is non-negative on* NND(k), *positive on* $\mathfrak{A}(K)$, *positively homogeneous, super-additive, concave and isotone, and satisfies* $j \circ J(0) = 0$. *Furthermore $j \circ J$ is closed if and only if j vanishes outside* PD(s).

Proof. The first set of properties is immediate. Assume, then, that $j \circ J$ is closed, let $C_0 \in \mathrm{NND}(s)$ be singular, and define $C_\varepsilon = C_0 + \varepsilon I_s$. Verify $J(KC_\varepsilon K') = \{K'(KC_\varepsilon K')^+K\}^{-1} = C_\varepsilon$. Now Corollary 7.5.1 in Rockafellar (1970) yields

$$\lim_{\varepsilon \searrow 0} j(C_\varepsilon) = \lim_{\lambda \nearrow 1} j \circ J((1-\lambda)KK' + \lambda KC_0K') = j \circ J(KC_0K') = 0.$$

Conversely, assume that j vanishes outside $\mathrm{PD}(s)$, and for $A \in \mathrm{NND}(k)$ define $A_\varepsilon = A + \varepsilon I_k$. By Theorem 7.5 in Rockafellar (1970) closedness of $j \circ J$ follows provided $\lim_{\varepsilon \searrow 0} j \circ J(A_\varepsilon) = j \circ J(A)$, for all $A \in \mathrm{NND}(k)$. This limit formula certainly holds for $A \in \mathfrak{A}(K)$, since then even $K'A_\varepsilon^{-1}K$ tends to $K'A^-K$, see Lemma 5.6.3 in Bandemer et al. (1977). Hence consider $A \notin \mathfrak{A}(K)$, and let Λ be its maximal eigenvalue. Since $K'(I_k - AA^+)K \neq 0$, there exists a non-zero $\mathbb{R}^s$-vector z such that $\Lambda(K'K)^{-1/2}K'(I_k - AA^+)K(K'K)^{-1/2} \geqslant zz'$. The following estimate is then easy to derive:

$$\begin{aligned}K'(A + \varepsilon I_k)^{-1}K &\geqslant \varepsilon^{-1}K'(I_k - AA^+)K + (\Lambda + \varepsilon)^{-1}K'AA^+K \\ &= (\Lambda + \varepsilon)^{-1}(K'K)^{1/2}\{I_s + \varepsilon^{-1}\Lambda(K'K)^{-1/2}K'(I_k - AA^+)K(K'K)^{-1/2}\}(K'K)^{1/2} \\ &\geqslant (\Lambda + \varepsilon)^{-1}(K'K)^{1/2}\{I_s + \varepsilon^{-1}zz'\}(K'K)^{1/2}.\end{aligned}$$

With the singular matrix $C_0 = z'z(K'K)^{-1} - (K'K)^{-1/2}zz'(K'K)^{-1/2}$ this yields

$$\begin{aligned}J(A_\varepsilon) &\leqslant (\Lambda + \varepsilon)(K'K)^{-1/2}\{I_s - (\varepsilon + z'z)^{-1}zz'\}(K'K)^{-1/2} \\ &= (\Lambda + \varepsilon)\{C_0 + \varepsilon)(K'K)^{-1}\}/(\varepsilon + z'z).\end{aligned}$$

Monotonicity of j, and again Theorem 7.5 in Rockafellar (1970) finally give $\lim_{\varepsilon \searrow 0} j \circ J(A_\varepsilon) \leqq (\Lambda / z'z) \lim_{\varepsilon \searrow 0} j(C_0 + \varepsilon(K'K)^{-1}) = 0$, as desired.

Now assume $\mathfrak{M}$ to be a compact convex subset of $\mathrm{NND}(k)$ which intersects $\mathfrak{A}(K)$. Any member of $\mathfrak{M}$ will be called an *information matrix*. The optimal design problem then reads:

(P) Maximize $j \circ J(M)$,

subject to $M \in \mathfrak{M}$.

The *optimal value* $v = \sup_{M \in \mathfrak{M}} j \circ J(M)$ is the maximal j-information for $K'\beta$ in $\mathfrak{M}$, any information matrix $M \in \mathfrak{M}$ for which $j \circ J(M)$ attains this value will be said to have *$\mathfrak{M}$-maximal j-information for $K'\beta$.*

A special case arises when $s = 1$. Then K may be identified with a $\mathbb{R}^k$-vector c, and the concept of information functionals trivializes, since $j \circ J(A) = j(1)/c'A^-c$ whenever $A \in \mathfrak{A}(c)$. Hence in this case an optimal solution of (P) will simply be said to have $\mathfrak{M}$-maximal information for $c'\beta$. The situations in which use of information functionals is redundant are described in the following theorem; if an information matrix satisfies any one of its four statements it will be called *uniformly optimal for $K'\beta$ in $\mathfrak{M}$*, cf., Kurotschka (1978, p. 1367).

Theorem 1 (U-optimality). *For every information matrix $M \in \mathfrak{M}$ the following four* *statements are equivalent:*

(a) M *has* $\mathfrak{M}$*-maximal* j*-information for* $K'\beta$, *for all information functionals* j.

(b) $J(M) \geqslant J(A)$, *for all* $A \in \mathfrak{M}$.

(c) $K'M^-K \leqslant K'A^-K$, *for all* $A \in \mathfrak{M} \cap \mathfrak{A}(K)$, *and* $M \in \mathfrak{A}(K)$.

(d) M *has* $\mathfrak{M}$*-maximal information for* $c'\beta$, *for all* c *in the range of* K.

Proof. Apply (a) to $j_{zz'}(C) = z'Cz$ to obtain (b). Conversely, (b) implies (a). Equivalence of (b) and (c) is immediate. Now assume (c). If $c = Kz$, with $z \in \mathbb{R}^s$, then $M \in \mathfrak{A}(c)$. Choose any competing information matrix $A \in \mathfrak{M} \cap \mathfrak{A}(c)$. From $\lambda M + (1-\lambda)A \in \mathfrak{M} \cap \mathfrak{A}(K)$ it follows that $c'M^-c = z'K'M^-Kz \leqq z'K'\{\lambda M + (1-\lambda)A\}^-Kz \leqq \lambda c'M^-c + (1-\lambda)c'A^-c$. Letting λ tend to 0 shows that M has $\mathfrak{M}$-maximal information for $c'\beta$. Conversely, (d) implies (c).

The topological assumptions underlying the optimal design problem (P) have an immediate consequence concerning the existence of optimal information matrices.

Theorem 2 (Existence). *If* j *vanishes outside* $\mathrm{PD}(s)$ *or if* $\mathfrak{M}$ *is a subset of* $\mathfrak{A}(K)$ *then there exists an information matrix in* $\mathfrak{M}$ *which has* $\mathfrak{M}$*-maximal* j*-information for* $K'\beta$.

Proof. If j vanishes outside $\mathrm{PD}(s)$ then $j \circ J$ is upper semi-continuous, by Lemma 1, and hence attains its supremum over the compact set $\mathfrak{M}$. In fact, the proof of Lemma 1 shows that if $\mathfrak{M}$ is a subset of $\mathfrak{A}(K)$ the same argument applies to $h(M) = j \circ J(M)$ if $M \in \mathfrak{M}$, $h(M) = -\infty$ otherwise.

More and proper use of convexity will be made in the following section on necessary and sufficient conditions for optimality.

3. Duality theorems

The optimal design problem (P) will be paired with a dual problem (D) which effectively amounts to maximizing the polar information functional j^0 over the polar set $\mathfrak{M}^0$. For the set $\mathfrak{M}$ of information matrices its polar set is the closed convex set given by

$$\mathfrak{M}^0 = \{B \in \mathbb{R}^{k \times k} \mid \langle M, B \rangle \leqq 1, \quad \text{for all } M \in \mathfrak{M}\}.$$

Because of the monotonicity behaviour of j^0 it suffices, in fact, to study the smaller set $\mathfrak{N}$ defined by

$$\mathfrak{N} = \mathfrak{M}^0 \cap \mathrm{NND}(k).$$

With the convention $1/0 = +\infty$ the dual of the optimal design problem is of the following type:

(D) Minimize $1/j^0(K'NK)$,
subject to $N \in \mathfrak{N}$.

The next two theorems relate problems (P) and (D) in the expected manner: they bound each other, and they share the same optimal value.

Theorem 3 (Mutual boundedness). *For every information matrix $M \in \mathfrak{M}$ and for every matrix $N \in \mathfrak{N}$ one has $j \circ J(M) \leqq 1/j^0(K'NK)$, with equality if and only if M lies in $\mathfrak{A}(K)$ and Conditions* (1), (2), (3) *are satisfied with $C = J(M)$ and $D = K'NK$:*

(1) trace $MN = 1$,
(2) $MN = KCK'N$,
(3) $j(C) \cdot j^0(D) = \text{trace } CD$.

Proof. If $M \notin \mathfrak{A}(K)$ then $j \circ J(M) = 0 < 1/j^0(K'NK)$, and equality is impossible. If $M \in \mathfrak{A}(K)$ the assertion follows from the triple inequality

$$1 \geqq \langle M, N\rangle \geqq \langle C, D\rangle \geqq j(C) \cdot j^0(D).$$

Conditions (1), (2), (3) correspond to equality in the first, second, and third inequality, respectively.

The definition of $\mathfrak{M}^0$ gives $1 \geqq \langle M, N\rangle$, and (1). Since N is taken to be non-negative definite we may continue $\langle M, N\rangle = \|M^{1/2}N^{1/2}\|^2$. Now $T(A) = M^{1/2+}KCK'M^{1/2+}A$ is an orthogonal projection on $\mathbb{R}^{k\times k}$, since $C = (K'M^+K)^{-1}$. The Pythagorean Theorem yields $\|M^{1/2}N^{1/2}\|^2 \geqq \|T(M^{1/2}N^{1/2})\|^2$, then, with equality if and only if $M^{1/2}N^{1/2} = T(M^{1/2}N^{1/2})$. The fact $M \in \mathfrak{A}(K)$ entails $M^{1/2}M^{1/2+}K = K$, Condition (2), and

$$\begin{aligned}\|T(M^{1/2}N^{1/2})\|^2 &= \text{trace } N^{1/2}M^{1/2}M^{1/2+}KCK'M^{1/2+}M^{1/2+}KCK'M^{1/2+}M^{1/2}N^{1/2}\\ &= \langle C, D\rangle.\end{aligned}$$

The last inequality follows from the definition of j^0.

The value of Theorem 3 lies in the explicit information obtained in Conditions (1)–(3). Also it makes evident that for a matrix $M \in \mathfrak{M}$ to have $\mathfrak{M}$-maximal j-information for $K'\beta$ it is sufficient to find a matrix $N \in \mathfrak{N}$ which satisfies $j \circ J(M) \cdot j^0(K'NK) = 1$. Theorem 4 now shows that this condition is necessary as well.

Theorem 4 (Duality). *In order that an information matrix $M \in \mathfrak{M}$ have $\mathfrak{M}$-maximal j-information for $K'\beta$ it is necessary and sufficient that there exists a matrix $N \in \mathfrak{N}$ such that $j \circ J(M) = 1/j^0(K'NK)$. More generally, one has*

$$\sup_{M \in \mathfrak{M}} j \circ J(M) = \min_{N \in \mathfrak{N}} 1/j^0(K'NK).$$

Proof. With the conventions $\log(0)=-\infty$ and $\log(+\infty)=+\infty$, the assertion may be rephrased with $\log\circ j\circ J(M)$ in place of $j\circ J(M)$, and $-\log j^0(K'NK)$ in place of $1/j^0(K'NK)$. On $\mathbb{R}^{k\times k}$ define the functions

$$\begin{aligned} f(A)&=0 && \text{if } A\in\mathfrak{M},\\ &=+\infty && \text{otherwise;}\end{aligned}$$

$$\begin{aligned} g(A)&=\log\circ j\circ J(A) && \text{if } A\in\mathfrak{A}(K),\\ &=-\infty && \text{otherwise.}\end{aligned}$$

The primal problem then reads equivalently: Maximize $g(A)-f(A)$ over $\mathbb{R}^{k\times k}$.

For the first part of the proof assume that $\mathfrak{M}$ intersects $\mathrm{PD}(k)$. Then the relative interior of the effective domain of f is contained in $\mathrm{PD}(k)$, by Lemma 2 in LaMotte (1977), and thus meets the relative interior of the effective domain of g. Hence Fenchel's Duality Theorem applies (Rockafellar 1970, Theorem 31.1) and states that

$$\sup_A\{g(A)-f(A)\}=\min_B\{f^*(B)-g^*(B)\},$$

where f^* and g^* are the functions conjugate to f and g, respectively. We now verify that the right minimization problem is nothing but a disguised version of the dual problem (D). By definition,

$$g^*(B)=\inf_{A\in\mathfrak{A}(K)}\{\langle A,B\rangle-\log\circ j\circ J(A)\}.$$

In particular, g^* has the same value at B and $\frac{1}{2}B+\frac{1}{2}B'$, so that B may be taken to be symmetric. Steps 1–4 will show that $g^*(B)=1+\log j^0(K'BK)$.

Step 1. If $B\notin\mathrm{NND}(k)$, then $g^*(B)=-\infty$. For choose a $\mathbb{R}^k$-vector u with $u'Bu<0$. Along the path $I_k+\alpha uu'$ monotonicity of $\log\circ j\circ J$ gives

$$g^*(B)\leqq\inf_{\alpha>0}\{\text{trace } B+\alpha u'Bu-\log\circ j\circ J(I_k+\alpha uu')\}=-\infty.$$

Step 2. If $B\in\mathrm{NND}(k)$ and $K'BK=0$, then $g^*(B)=-\infty$. For along the path $\alpha KK'$ one has

$$g^*(B)\leqq\inf_{\alpha>0}\{-\log\alpha-\log\circ j\circ J(KK')\}=-\infty.$$

Step 3. If $B\in\mathrm{NND}(k)$ and $K'BK\neq 0$, then $g^*(B)\leqq 1+\log j^0(K'BK)$. For when $C\in\mathrm{PD}(s)$, then $\langle KCK',B\rangle=\langle C,K'BK\rangle>0$, and $J(KCK')=C$. Define $A_C=KCK'/\langle C,K'BK\rangle$, then $A_C\in\mathfrak{A}(K)$, and

$$\begin{aligned} g^*(B)&\leqq\inf_{C\in\mathrm{PD}(s)}\{\langle A_C,B\rangle-\log\circ j\circ J(A_C)\}\\ &=\inf_C\{1-\log(j(C)/\langle C,K'BK\rangle)\}\\ &=1+\log j^0(K'BK).\end{aligned}$$

Step 4. If $B \in \mathrm{NND}(k)$, then $g^*(B) \geqq 1 + \log j^0(K'BK)$. For this is trivially true if $j^0(K'BK) = 0$. Otherwise use $\langle A, B\rangle \geqq j \circ J(A) \cdot j^0(K'BK)$ from the proof of Theorem 3 to obtain

$$g^*(B) \geqq \inf_{A \in \mathfrak{A}(K)} \{j \circ J(A) \cdot j^0(K'BK) - \log \circ j \circ J(A)\}$$
$$\geqq \inf_{\alpha > 0} \{\alpha j^0(K'BK) - \log \alpha\} = 1 + \log j^0(K'BK).$$

By definition, $f^*(B) = \sup_{M \in \mathfrak{M}} \langle M, B\rangle$, so that $f^*(B)$ is positive for $B \in \mathrm{NND}(k)$, $B \neq 0$. For fixed B, consider the function

$$h(\alpha) = f^*(\alpha B) - g^*(\alpha B) = \alpha f^*(B) - 1 - \log \alpha - \log j^0(K'BK), \quad \text{for } \alpha > 0.$$

Unless $g^*(B) = -\infty$, the unique minimum of h occurs at $1/f^*(B)$ and is equal to $\log f^*(B) - \log j^0(K'BK) = -\log j^0(K'NK)$, with $N = B/f^*(B) \in \mathfrak{N}$. This completes the first part of the proof.

The second part merely uses the fact that $\mathfrak{M}$ intersects $\mathfrak{A}(K)$. Choose a matrix $M \in \mathfrak{M}$ which has maximal rank m, say, and choose a matric $U \in \mathbb{R}^{k \times m}$ which has the same range as M and which satisfies $U'U = I_m$. Then $A = UU'AUU'$, for every $A \in \mathfrak{M}$, and $UU'K = K$. Verify $(UU'AUU')^+ = U(U'AU)^+U'$, so that $J(A) = (K'A^-K)^{-1} = (K'UE^-U'K)^{-1} = H(E)$, say, with $E = U'AU$. Thus problem (P) is 'equivalent' with the problem

(P′) Maximize $j \circ H(E)$,
subject to $E \in U'\mathfrak{M}U$.

The first part of this proof applies to (P′) and the problem

(D′) Minimize $1/j^0(K'UFU'K)$,
subject to $F \in \{U'\mathfrak{M}U\}^0 \cap \mathrm{NND}(m)$.

It is not hard to see that $\{U'\mathfrak{M}U\}^0 \cap \mathrm{NND}(m) = U'\mathfrak{N}U$. Replacing F by $U'NU$ and using $UU'K = K$ establishes the relation between (D′) and (D) which completes the proof.

Traditionally convex analysis prefers minimization of convex functions to maximization of concave functions. However, the Φ_p-criteria of Kiefer (1974) all happen to be, not only convex, but even log–convex. In the present setting log–convexity no longer appears to be accidental but finds a natural explanation, namely, any function Φ given by $\Phi(A) = 1/j \circ J(A)$ is log–convex since $-\log \Phi = \log \circ j \circ J$ certainly is concave.

Theorems 3 and 4 contain sufficient information to formulate yet another optimality characterization which does not explicitly refer to the dual problem (D). As a motivation for Theorem 5 postmultiply both sides of Condition (2) first by $N^{1/2+}$ and then by their respective transposes. Replacing $K'NK$ by D this yields the equation

[*] $MNM = KCDCK'$.

If an optimal information matrix M is non-singular it follows that $M^{-1}KCDCK'M^{-1}$ must be a member of $\mathfrak{N}$, it is for singular matrices that the notion of contracting g-inverses is now introduced.

Definition 4. A g-inverse G of an information matrix $M \in \mathfrak{M}$ is said to be *contracting* if there exists an optimal solution N of the dual problem (D) such that

$$u'Nu = u'G'MNMGu, \quad \text{for all } u \in \mathbb{R}^k.$$

If G is a contracting g-inverse of M then the projection MG of $\mathbb{R}^k$ is a contraction of the cylinder $Z = \{u \in \mathbb{R}^k \mid u'Nu \leqq 1\}$, i.e., $MG(Z) \subset Z$. Moreover, the projection $(MG \otimes MG)(A) = MGAG'M$ of $\mathbb{R}^{k \times k}$ is a contraction of the polar set $\{N\}^0 = \{A \in \mathbb{R}^{k \times k} \mid \langle A, N\rangle \leqq 1\}$. The construction of contracting g-inverses of M is geometric in nature, based on complementary subspaces of the range of M as sought and discussed by Silvey (1978, p. 557).

Lemma 2. *For every information matrix $M \in \mathfrak{M}$ which lies in $\mathfrak{A}(K)$ there exists a positive definite contracting* g*-inverse G of M.*

Proof. For $N \in \mathrm{NND}(k)$ define $N_K = NK(K'NK)^+K'N$. Since the nullspace of $K'N_K = K'N$ is contained in the nullspace of N_K it follows that nullspace $K' \cap$ range $N_K = \{0\}$, so that $\mathbb{R}^k =$ range $K +$ nullspace N_K. If N lies in $\mathfrak{N}$ so does N_K, since

$$1 \geqq \langle M, N\rangle = \|N^{1/2}M^{1/2}\|^2 \geqq \|N^{1/2}K(K'NK)^+K'N^{1/2}N^{1/2}M^{1/2}\|^2 = \langle M, N_K\rangle.$$

And if N is optimal then so is N_K, since $K'N_KK = K'NK$. Fix one such optimal solution N_K.

The assumption $M \in \mathfrak{A}(K)$ implies $\mathbb{R}^k =$ range $M +$ nullspace N_K. Let r be the rank of M, and choose some $k \times (k-r)$ matrix H such that its columns span a subspace of the nullspace of N_K which is complementary to the range of M, define $G = (M + HH')^{-1}$. Then G is a positive definite g-inverse of M, see Rao (1973, p. 34). From $I_k = MG + HH'G$ and $N_KH = 0$ one gets $u'N_Ku = u'G'MN_KMGu$, whence G is also contracting.

Theorem 5 (Equivalence). *Let $M \in \mathfrak{M}$ be an information matrix which lies in $\mathfrak{A}(K)$, and let C be the matrix $J(M) = (K'M^-K)^{-1}$. Then M has $\mathfrak{M}$-maximal j-information for $K'\beta$ if and only if there exist a* g*-inverse G of M and a matrix $D \in \mathrm{NND}(s)$ with the properties that*

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

and that G and *D jointly satisfy the system of inequalities*

$$\text{trace } K'GAG'KCDC \leqq 1, \quad \textit{for all } A \in \mathfrak{M}.$$

Moreover, if M has $\mathfrak{M}$-maximal j-information for $K'\beta$ then for every contracting *g-inverse G of M there exists a matrix $D \in \mathrm{NND}(s)$ such that G and D have the stated properties. And if G and D have the stated properties, then actually* trace $K'GAG'KCDC = 1$ *whenever A has $\mathfrak{M}$-maximal j-information for $K'\beta$.*

Proof. For the direct part assume M to be optimal. Let G be a contracting g-inverse of M, with associated optimal solution N of the dual problem, define $D = K'NK$. Conditions (1)–(3) yield $j(C)j^0(D) = \text{trace } CD = \text{trace } MN = 1$. Equation [$*$] before Definition 4 leads to $G'MNMG = G'KCDCK'G$. The contraction argument shows that $\langle A, G'KCDCK'G\rangle = \langle MGAG'M, N\rangle = \langle A, N\rangle \leqq 1$, as desired.

For the converse part assume that G and D have the stated properties. Define $N = G'KCDCK'G$, then $N \in \mathfrak{N}$, and $j^0(K'NK) = j^0(D) = 1/j(C)$. Hence both M and N are optimal solutions of their respective programs, and trace $AN = 1$ whenever A is optimal as well, by Condition (1).

Theorem 5 splits the characterization of optimal solutions of the design problem (P) into two parts, according to the fact that the objective function is a composition of the functions j and J. The first part is in terms of the $s \times s$ matrices C and D; in many cases the solutions D of the equations $j(C) \cdot j^0(D) = \text{trace } CD = 1$ can be described explicitly. The second part mainly concerns the $k \times k$ matrices G and A; inversions are required only of the matrix M which poses as a candidate for optimality, whereas the inequalities are *linear* in the competing information matrices A. As expected, the matrices C and D disappear completely in case of c-optimality.

Corollary 5.1 (c-optimality). *Let $M \in \mathfrak{M}$ be an information matrix which lies in $\mathfrak{A}(c)$, $c \in \mathbb{R}^k$. Then M has $\mathfrak{M}$-maximal information for $c'\beta$ if and only if there exists a g-inverse G of M such that $c'GAG'c \leqq c'M^-c$, for all $A \in \mathfrak{M}$.*

If the rank of M is maximal then the expression $c'GAG'c$ is invariant to the choice of the g-inverse G, and may be written as $c'M^-AM^-c$. This, in conjunction with Theorem 1(d), gives the following characterization of uniform optimality.

Corollary 5.2 (U-optimality). *Let $M \in \mathfrak{M}$ be an information matrix with maximal rank. Then M is uniformly optimal for $K'\beta$ in $\mathfrak{M}$ if and only if $K'M^-AM^-K \preccurlyeq K'M^-K$, for all $A \in \mathfrak{M}$.*

The design problem (P) need not have a unique optimal solution, but at least when j is strictly concave only a surprisingly special type of multiplicity is possible: given one optimal information matrix all others are obtained as solutions of an inhomogeneous *linear* matrix equation.

Corollary 5.3 (Multiplicity). *Suppose the information functional j is strictly concave. Let $M \in \mathfrak{M}$ be an information matrix which has $\mathfrak{M}$-maximal j-information for $K'\beta$, and let G be a contracting g-inverse of M. Then any other information matrix $A \in \mathfrak{M} \cap \mathfrak{A}(K)$ also has $\mathfrak{M}$-maximal j-information for $K'\beta$ if and only if $AG'K = K$.*

Proof. For the direct part assume A to be optimal. Strict concavity of j implies $J(M) = J(A) = C$, say. Let $D \in \mathrm{NND}(s)$ satisfy $j(C)j^0(D) = \operatorname{trace} CD$, and suppose $Dz = 0$, $z \in \mathbb{R}^s$. It is easily seen that for $\alpha > 0$

$$j(C+\alpha zz')j^0(D) \leqq \langle C+\alpha zz', D\rangle = \langle C, D\rangle = j(C)j^0(D) \leqq j(C+\alpha zz')j^0(D).$$

But since j is strictly isotone $h(\alpha) = j(C+\alpha zz')$ can be constant only if $z = 0$. Hence D must be positive definite. Apply Condition (2) to the optimal solutions A and $G'KCDCK'G$ of (P) and (D), and postmultiply by K. Cancel CD in the resulting equation $AG'KCD = KCD$. For the converse part premultiply $AG'K = K$ by $K'A^-$ to obtain $J(M) = J(A)$.

Theorem 5 comes closest to the classical Kiefer–Wolfowitz type equivalence theorems, see Kiefer and Wolfowitz (1960) and Kiefer (1974). It should be clear, however, that Theorem 5 is no substitute for Theorems 3 and 4 on duality, the latter also allow to determine the optimal value v, to identify optimizing sequences $j \circ J(M_i) \to v$, and to establish non-existence of optimal designs. Examples will be given in Section 6.

The system of inequalities in Theorem 5 involves one inequality for each matrix $A \in \mathfrak{M}$. In many cases fewer inequalities will do, namely, when S is a subset of information matrices whose convex hull is $\mathfrak{M}$ then only the inequalities for $A \in S$ need be considered. Furthermore, if an optimal information matrix M is written as a convex combination of $A_1, \ldots, A_l \in S$, then necessarily trace $K'GA_iG'KCDC = 1$, for all $i = 1, \ldots, l$. More can be said if more is known of the structure of S, as in the following section.

4. Design measures

The statistical assumptions underlying the design problem typically go further than outlined at the beginning of Section 2. Usually the expectation of the ith observation Y_i is taken to be of the form $f(x_i)'\beta$, so that the $\mathbb{R}^k$-vectors $f(x_1), \ldots, f(x_n)$ appear as the rows of the matrix X. In such a linear regression model an experimental design for a sample size n simply consists of some n-tupel $(x_1, \ldots, x_n)$ where every level x_i lies in a specified experimental domain $\mathfrak{X}$, telling the experimenter to draw the ith observation Y_i at level x_i. Since the numbering of the observations is immaterial one may as well quote of the n-tuple $(x_1, \ldots, x_n)$ only its distinct levels $x_1, \ldots, x_l$ and their associated standardized frequencies $n_1/n, \ldots, n_l/n$. Accordingly a *design of size n* is a probability measure

ξ_n on $\mathfrak{X}$ which to a finite number of support points assigns weights that are multiples of $1/n$. For such a design ξ_n the matrix $X'X$ attains the form

$$X'X = n\sum_{i=1}^{l}(n_i/n)f(x_i)f(x_i)' = n\int_{\mathfrak{X}} f(x)f(x)'\,\mathrm{d}\xi_n.$$

Large sample considerations suggest the following definition, see Elfving (1952, p. 256), Kiefer (1959, p. 281). The assumptions on f pertain to its range rather than to its domain of definition, see Kiefer and Wolfowitz (1960, p. 363), Silvey and Titterington (1973, p. 23).

Definition 5. Let f be a $\mathbb{R}^k$-valued function defined on a set $\mathfrak{X}$ such that its image $f(\mathfrak{X})$ is compact. The set Ξ is defined to consist of all probability measures ξ on (the σ-algebra of all subsets of) $\mathfrak{X}$ which have a finite support, any such ξ is called a *design measure.* The *information matrix of* ξ is defined to be $M(\xi) = \int_{\mathfrak{X}} f(x)f(x)'\,\mathrm{d}\xi$, and the set of all these matrices is denoted by $M(\Xi)$.

Obviously $M(\Xi)$ is the convex hull of the set S when S consists of all rank 1 information matrices $f(x)f(x)'$, with $x\in\mathfrak{X}$, hence $M(\Xi)$ is also compact, see Rockafellar (1970, Theorem 17.2). It is customary to assume that the image $f(\mathfrak{X})$ spans all of $\mathbb{R}^k$, then there exist positive definite matrices in $M(\Xi)$; this property is not fullfilled in all models of interest. In the sequel we merely assume that $M(\Xi)$ meets $\mathfrak{A}(K)$, when $K'\beta$ is the parameter under investigation. With these assumptions the set $M(\Xi)$ is a feasible choice for the set $\mathfrak{M}$ which enters into the design problem (P). However, since all results in Sections 2 and 3 are given in terms of information matrices $M(\xi)$ rather than design measures ξ, the following questions suggest themselves: Given an information matrix $M\in M(\Xi)$, possibly optimal, how can one recover the number of support points, the support points themselves, and the weights of those design measures which have information matrix M?

The following bound on the number of support points is due to Fellman (1974, Theorem 4.1.4) and generalizes earlier results of Elfving (1952, p. 260) and Chernoff (1953, p. 590). However, it does not depend on any optimality criterion as the proof of those authors suggest, but simply is a property of the $s\times s$ information matrices J for $K'\beta$.

Theorem 6 (Support points). *For every information matrix $A\in M(\Xi)$ which lies in $\mathfrak{A}(K)$ there exists a design measure $\xi\in\Xi$ with not more than $s(s+1)/2+s(\operatorname{rank}A-s)$ support points such that $\alpha J(A)=J(M(\xi))$ for some $\alpha\geqq 1$.*

Proof. Let r be the rank of A. Choose a g-inverse G of A, a $k\times r$ matrix U which has the same range as A and which satisfies $U'U=I_r$, and define $Q=I_r-U'GK(U'GK)^+$. Then $A\in\mathfrak{A}(K)$ entails

$$K=AGK=U(U'AU-QU'AUQ)U'GK=UT(A)U'GK,$$

when the linear operator T on the space Sym(k) is defined by $T(B)=U'BU-QU'BUQ$. The range of T has dimension $s(s+1)/2+s(r-s)=d$, say, since it is spanned by $T(U\begin{bmatrix} v \\ w \end{bmatrix}\begin{bmatrix} v \\ w \end{bmatrix}'U')=\begin{bmatrix} vv' & vw' \\ wv' & 0 \end{bmatrix}$ when Q is in its simplest form $\begin{bmatrix} 0 & 0 \\ 0 & I_{r-s} \end{bmatrix}$. Choose a design measure $\eta\in\Xi$ such that $A=M(\eta)$, and define C to be the convex hull of all matrices $f(x)f(x)'$ with x being a support point of η. Hence the image $T(C)$ is a compact convex set which contains the matrix $T(A)$, let α be the largest number $\geqq 1$ with $\alpha T(A)\in T(C)$. Then $\alpha T(A)$ lies on the boundary of $T(C)$, and by Caratheodory's Theorem there exist d points $x_1,\ldots,x_d$ in the support of η such that with new weights $\xi(x_i)$ one has $\alpha T(A)=T(M(\xi))$. Thus range $M(\xi)\subset$ range A, and $UU'M(\xi)UU'=M(\xi)$. This entails

$$M(\xi)GK=UT(M(\xi))U'GK=\alpha UT(A)U'GK=\alpha K,$$

implying $M(\xi)\in\mathfrak{A}(K)$, $K'A^-K=\alpha K'M(\xi)^-K$, and $\alpha J(A)=J(M(\xi))$.

For the present approach it seems natural that a design measure ξ inherit its optimality properties from its information matrix $M(\xi)$. By Theorem 2, then, there exists a design measure which has $M(\Xi)$-maximal j-information for $K'\beta$ provided j vanishes outside PD(s). In fact, the topological arguments of Theorem 2 carry over to designs of size n, since the associated set $M(\Xi_n)$ of information matrices is the continuous image of the n-fold Cartesian product of $f(\mathfrak{X})$ and hence compact. This clearly illustrates that the sets $M(\Xi)$ and $M(\Xi_n)$ are distinguished, not by compactness, but by convexity.

The constraints in the dual problem (D) now lend themselves to an appealing interpretation. When a matrix $N\in$ NND(k) is identified with the cylinder (including ellipsoids) that it defines in $\mathbb{R}^k$, i.e., with the set $\{u\in\mathbb{R}^k \mid u'Nu\leqq 1\}$, then $\mathfrak{N}$ consists precisely of all $f(\mathfrak{X})$ *covering cylinders.*

$$N\in M(\Xi)^0 \Leftrightarrow f(x)'Nf(x)\leqq 1, \quad \text{for all } x\in\mathfrak{X}.$$

This idea dates back to Elfving (1952, p. 260) and is adopted *expressis verbis* by Silvey (Wynn 1972, p. 174), Sibson (1974, p. 684) and Silvey and Titterington (1973, p. 25). In fact, further geometric considerations lead to a direct proof of duality of (P) and (D) in case of c-optimality, see Pukelsheim (1979), and also yield the following bound on the optimal value v of problem (P). Define the *regression ball* $\mathfrak{R}$ to be the convex hull of the image $f(\mathfrak{X})$ and its reflection $-f(\mathfrak{X})$ and assume that $f(\mathfrak{X})$ spans all of $\mathbb{R}^k$, then $\mathfrak{R}$ is a compact convex set which contains 0 in its interior. The bound will be given in terms of an in-ball radius $r=\|d\|$ of $\mathfrak{R}$, i.e., d is a boundary point of $\mathfrak{R}$ and d has minimal Euclidean norm r among all other boundary points of $\mathfrak{R}$. In other words, the hyperplane $\{u\in\mathbb{R}^k \mid d'u=r^2\}$ supports $\mathfrak{R}$ in d. Therefore $r^{-4}dd'$ is a $f(\mathfrak{X})$ covering cylinder, and $v\leqq r^4/j^0(K'dd'K)$. Equality holds if $s=1$ (*op. cit.*); if $s>1$ the matrix $K'dd'K$ is singular and the proof of Corollary 5.3 shows that equality cannot hold if j is strictly isotone. Identification of possible support points is part of the following reformulation of Theorem 5 and Condition (1) in Theorem 3.

Theorem 7. *Let $\xi \in \Xi$ be a design measure for which $M(\xi)$ lies in $\mathfrak{A}(K)$, and let C be the matrix $(K'M(\xi)^-K)^{-1}$. Then ξ has $M(\Xi)$-maximal j-information for $K'\beta$ if and only if there exist a g-inverse G of $M(\xi)$ and a matrix $D \in \mathrm{NND}(s)$ such that $j(C) \cdot j^0(D) = \operatorname{trace} CD = 1$, and*

$$f(x)'G'KCDCK'Gf(x) \leqq 1, \quad \text{for all } x \in \mathfrak{X}.$$

If ξ is optimal then actually $f(x)'G'KCDCK'Gf(x) = 1$ for all support points x of every design measure which has $M(\Xi)$-maximal j-information for $K'\beta$; more general, such points x must satisfy $f(x)'Nf(x) = 1$ whenever N is an optimal $f(\mathfrak{X})$ covering cylinder.

For instance, Theorem 7 provides proof of the following statement in Wald (1943, p. 136): *No design measure is uniformly optimal for $K'\beta$ in $M(\Xi)$, unless $s = 1$*, For assume that ξ is uniformly optimal for $K'\beta$, and let x be one of its support points. Then for every vector c in the range of K the design measure ξ is optimal for $c'\beta$, by Theorem 1(d), and for all choices of $M(\xi)^-$ one has $c'M(\xi)^-f(x)f(x)'M(\xi)^-c = c'M(\xi)^-c$, by Theorem 7 or Corollary 5.1; comparing ranks in the resulting equation $K'M(\xi)^-f(x)f(x)'M(\xi)^-K = K'M(\xi)^-K$ proves $s = 1$. Theorem 7 also leads to a rather strong statement concerning the weights of optimal design measures. Its proof is based on the same idea as in Sibson and Kenny (1975, p. 290), namely, to expand the quadratic form in the inequalities of Theorem 7 until the matrix $M(\xi)$ appears in the middle.

Corollary 7.1 (Weights). *Let $M \in M(\Xi)$ be an information matrix which has $M(\Xi)$-maximal j-information for $K'\beta$, and let C, D, G be as in Theorem 7. Suppose M is obtained as the information matrix of a design measure which assigns weights w_i to l support points $x_i \in \mathfrak{X}$, and choose a root E of C, i.e., $C = EE'$. The weight vector $w = (w_1, \ldots, w_l)'$ then solves the equation $Aw = 1_l$, where $1_l = (1, \ldots, 1)' \in \mathbb{R}^l$ and where the entries of the matrix $A \in \mathrm{NND}(l)$ are given by*

$$a_{hi} = \{g(x_h)'(E'DE)^{1/2}g(x_i)\}^2, \qquad g(x) = E'K'Gf(x).$$

A single weight w_i is bounded by $1/a_{ii}$, and no weight is larger than $\lambda_{\max}(CD)$.

Proof. Theorem 7 gives $g(x_h)'E'DEg(x_h) = 1$. Expand $E'DE$ into $(E'DE)^{1/2}E'K'GMG'KE(E'DE)^{1/2}$ and use $M = \sum_i w_i f(x_i)f(x_i)'$ to obtain $\sum_i a_{hi}w_i = 1$, for all h. This proves $Aw = 1_l$, and $a_{ii}w_i \leqq 1$. But

$$1 = \{g(x_i)'E'DEg(x_i)\}^2 \leqq \{g(x_i)'(E'DE)^{1/2}g(x_i)\lambda_{\max}(E'DE)^{1/2}\}^2$$
$$= a_{ii}\lambda_{\max}(CD).$$

Hence $a_{ii} > 0$, and $w_i = \leqq 1/a_{ii} \leqq \lambda_{\max}(CD)$. Since A is the Hadamard (component-wise) square of a Gramian matrix, A is non-negative definite.

Notice that the bound $\lambda_{\max}(CD)$ is not absurd, since $\lambda_{\max}(CD) \leqq \operatorname{trace} CD = 1$.

The function $g(x)$ defined in Corollary 7.1 plays a particular role in the function space $L_2(\xi)$, for special cases this is discussed in Kiefer and Wolfowitz (1960, p. 364), Kiefer (1962, p. 597), Karlin and Studden (1966, Theorem 6.2).

Corollary 7.2 ($L_2(\xi)$-version). *Let $\xi \in \Xi$ be a design measure for which $M(\xi)$ lies in $\mathfrak{A}(K)$, and let C be the matrix $(K'M(\xi)^- K)^{-1}$. Then ξ has $M(\Xi)$-maximal j-information for $K'\beta$ if and only if there exist matrices $T \in \mathbb{R}^{k\times s}$ and $F \in \mathbb{R}^{s\times s}$ such that $j(C) \cdot j^0(FF') = 1$, and such that the transformed function $g(x) = T'f(x)$ has the properties:*

(a) *the components of g form an orthonormal system with respect to ξ,*
(b) *the components of g are orthogonal to the system $(I_k - KK^+)f$,*
(c) *$F'CK'Tg$ takes its values in the closed Euclidean unit ball of $\mathbb{R}^s$.*

Proof. For the direct part choose C, D, G as in Theorem 7, and define $T = G'KC^{1/2}$ and $F = D^{1/2}$. For the converse part notice first that (a) is equivalent to $T'M(\xi)T = I_s$, and that (b) is the same as $M(\xi)T = K(K'K)^{-1}K'M(\xi)T$. Therefore

$$K'T = K'M(\xi)^- M(\xi)T = K'M(\xi)^- K(K'K)^{-1}K'M(\xi)T$$

and

$$I_s = T'M(\xi)T = T'K(K'K)^{-1}K'M(\xi)T = T'KCK'T.$$

Hence $T'K$ is non-singular and satisfies $CK'TT'K = I_s$. Define $N = TT'KCFF'CK'TT'$. Then N is in $\mathfrak{N}$, $K'NK = FF'$, and $j(C) = 1/j^0(K'NK)$ proves that both $M(\xi)$ and N are optimal solutions of their respective problems.

Next we turn to particular choices of the information functional j.

5. Special criteria

Our investigations apply in particular to the family $\{j_p \mid p \in [-\infty, +1]\}$ introduced in Section 2. It is readily verified that j_p vanishes outside PD(s) if and only if $p \in [-\infty, 0]$ or $s = 1$, whence in these cases j_p-optimal information matrices exist, according to Theorem 2. For $s = 1$ existence follows much simpler from Elfving's (1952) geometric argument, see Chernoff (1972, p. 12), Pukelsheim (1979); for $p = -1$ existence of optimal solutions is established by Fellman (1974, Theorem 4.1.3), for $p = -1, -2, \ldots$ by Pázman (1980, Propositions 2, 4). In case $p = 0$ (D-optimality) existence poses no problem when $s = k$, nor has it been doubted when $s < k$, see Kiefer (1961, p. 306), Atwood (1973, p. 343). However, compactness of $M(\Xi)$ alone does *not* suffice, and in Section 6 examples are given where no j_1-optimal design measure exists. In particular, the proposed proof of Theorem 1 in Whittle (1973, p. 125) is complete only if one assumes the existence of an optimal ξ^* or closedness of the objective function ϕ. Precisely this closedness is destroyed by Gribik and Kortanek (1977, p. 243) by setting the objective function equal to ∞ for *all* singular information matrices. Optimality will be characterized using the following lemma.

Lemma 3. *The polar function of the information functional j_p is sj_q, provided $p, q \in [-\infty, +1]$ and $p+q=pq$. If a matrix $C \in \mathrm{PD}(s)$ is given, then a matrix $D \in \mathrm{NND}(s)$ solves the equations $j_p(C) \cdot (j_p)^0(D) = \mathrm{trace}\, CD = 1$ if and only if $D = C^{p-1}/\mathrm{trace}\, C^p$ in case $p > -\infty$, or $\lambda_{\min}(C) \cdot D \in \mathrm{conv}\, S$ in case $p = -\infty$. Here* conv S *denotes the convex hull of all $s \times s$ matrices of the form zz' such that z is an eigenvector of C corresponding to $\lambda_{\min}(C)$ with Euclidean norm 1.*

Proof. Since polar functions are closed it suffices to compute $(j_p)^0(D) = \inf_{C \in \mathrm{PD}(s)} \langle C, D\rangle / j_p(C)$ only for $D \in \mathrm{PD}(s)$. In case $p \notin \{-\infty, 0, +1\}$ Theorem 5.10 in Gaffke and Krafft (1979) gives $\langle C, D\rangle \geqq j_p(C) sj_q(D)$. Hence

$$sj_q(D) \leqq \inf_C \langle C, D\rangle / j_p(C) \leqq \langle D^{q/p}, D\rangle / j_p(D^{q/p}) = sj_q(D),$$

therefore $sj_q = (j_p)^0$. Given C, the solution to $j_p(C)(j_p)^0(D) = \langle C, D\rangle$ then is $D = \alpha C^{p/q}$, and α is determined from $\alpha\langle C, C^{p/q}\rangle = 1$. In case $p = 0$ the same argument leads to the familiar arithmetic-geometric-mean inequality as in Karlin and Studden (1966, p. 795). In case $p = 1$ one certainly has $\langle C, D\rangle \geqq (\mathrm{trace}\, C)\lambda_{\min}(D)$. Hence

$$\begin{aligned} sj_{-\infty}(D) \leqq \inf_C \langle C, D\rangle / j_1(C) &\leqq \inf_{\|z\|=1, \varepsilon>0} s\langle zz' + \varepsilon I_s, D\rangle / \mathrm{trace}(zz' + \varepsilon I_s) \\ &= s \inf_{\|z\|=1} z'Dz = sj_{-\infty}(D). \end{aligned}$$

Given C, the desired solution is $D = \alpha I_s$ with $\alpha\, \mathrm{trace}\, C = 1$. In case $p = -\infty$ one similarly has $\langle C, D\rangle \geqq \lambda_{\min}(C)(\mathrm{trace}\, D)$. Hence $sj_1(D) \leqq \inf_C \langle C, D\rangle / j_{-\infty}(C) \leqq \langle I_s, D\rangle / j_{-\infty}(I_s) = sj_1(D)$. Given C, let $\lambda_1 > \cdots > \lambda_r > 0$ be its distinct eigenvalues, with associated projectors $E(\lambda_i)$. Write D as $\lambda_r^{-1} \sum_{j=1}^s \alpha_j z_j z_j'$, with $\mathbb{R}^s$-vectors z_j of Euclidean norm 1 and with coefficients $\alpha_j \geqq 0$. Then

$$\langle C, D\rangle = \sum_i \sum_j \lambda_i \lambda_r^{-1} \alpha_j z_j' E(\lambda_i) z_j \geqq \sum_j \alpha_j = \lambda_{\min}(C)(\mathrm{trace}\, D),$$

with equality if and only if for all $i = 1, \ldots, r-1$ and $j = 1, \ldots, s$ one has $\alpha_j z_j' E(\lambda_i) z_j = 0$, i.e., $\alpha_j = 0$ or z_j lies in the range of $E(\lambda_r)$. This completes the proof.

With Lemma 3 the previous results simplify considerably and completely resolve the two classical problems of j_p-optimality of singular information matrices, and of optimality with respect to the non-differentiable E-, i.e., $j_{-\infty}$-criterion.

Theorem 8 (j_p-optimality). *Let $M \in \mathfrak{M}$ be an information matrix which lies in $\mathfrak{A}(K)$. If $p > -\infty$, then M has $\mathfrak{M}$-maximal j_p-information for $K'\beta$ if and only if there exists a g-inverse G of M such that*

$$\operatorname{trace} K'GAG'K(K'M^-K)^{-p-1} \leqq \operatorname{trace}(K'M^-K)^{-p}, \quad \textit{for all } A \in \mathfrak{M}.$$

If $p = -\infty$ and conv S *is the set defined in Lemma 3, then M has $\mathfrak{M}$-maximal $j_{-\infty}$-information for $K'\beta$ if and only if there exist a g-inverse G of M and a matrix $E \in$* conv *such that*

$$\operatorname{trace} K'GAG'KE \leqq \lambda_{\max}(K'M^-K), \quad \textit{for all } A \in \mathfrak{M}.$$

Proof. For finite p Theorem 5 and Lemma 3 give $CDC = (K'M^-K)^{-p-1}/\operatorname{trace}(K'M^-K)^{-p}$, for $p = -\infty$ one obtains $CDC = E/\lambda_{\max}(K'M^-K)$.

When the set $\mathfrak{M}$ of information matrices is induced by the set of all design measures, i.e., $\mathfrak{M} = M(\Xi)$, Theorem 8 allows an obvious modification in order to parallel Theorem 7. The resulting version is intimately related to the Theorem in Silvey (1978, p. 555) who proves the sufficiency part, and conjectures the necessity part. Fedorov and Malyutov (1972, p. 286) seem to imply that G may be chosen to be the Moore–Penrose inverse M^+, Bandemer et al. (1977, Section 5.6.3) claim that an arbitrary g-inverse M^- may replace G and only the inequalities of those $x \in \mathfrak{X}$ with $f(x) \in$ range M need be considered. Either of these versions allows counterexamples, see Pukelsheim (1979, Examples 1, 2). Kiefer (1974, Theorem 6) has a partial result on $j_{-\infty}$-optimality covering the least complicated situation: when the eigenvalue $\lambda_{\max}(K'M^-K)$ is simple there is a unique matrix E which satisfies the conditions of Theorem 8. Of course, Theorem 8 also includes the differentiable cases, such as formula (4.19) of Kiefer (*op. cit.*), or the original Equivalence Theorem of Kiefer and Wolfowitz (1960, p. 364).

Duality approaches to the optimal design problem are first mentioned in the discussion of Wynn (1972), a duality theorem on j_0-optimality is presented by Sibson (1974, p. 685). For s out of k parameters a dual problem different from ours is chosen by Silvey and Titterington (1973, p. 25), their dual variable consists of a pair $(D, B) \in \mathrm{PD}(s) \times \mathbb{R}^{s \times (k-s)}$; the following corollary extends their results (*op. cit.*) to all $p \in [-\infty, +1]$.

Corollary 8.1 (s out of k parameters). *For every $p \in [-\infty, +1]$ there exists a $s \times (k-s)$ matrix B such that for all information matrices $M \in M(\Xi)$ whose range contains the leading s-dimensional coordinate subspace the following holds: If $p > -\infty$, then M has $M(\Xi)$-maximal j_p-information for $(\beta_1, \ldots, \beta_s)'$ if and only if $C = ([I_s : 0]M^-[I_s : 0]')^{-1}$ satisfies*

$$f(x)'[I_s : B]'C^{p-1}[I_s : B]f(x) \leqq \operatorname{trace} C^p, \quad \textit{for all } x \in \mathfrak{X}.$$

If $p = -\infty$ and conv S *is the set defined in Lemma 3, then M has $M(\Xi)$-maximal*

$j_{-\infty}$-information for $(\beta_1, \ldots, \beta_s)'$ if and only if there exists a matrix $E \in \operatorname{conv} S$ which satisfies

$$f(x)'[I_s : B]'E[I_s : B]f(x) \leqq \lambda_{\min}(C), \quad \text{for all } x \in \mathfrak{X}.$$

In case of optimality $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$ satisfies $BM_{22} = -M_{12}$ if $p > -\infty$, or $EBM_{22} = -EM_{12}$ if $p = -\infty$.

Proof. Set $K = [I_s : 0]'$, and choose an optimal solution N of the dual problem. Then $N_K = NK(K'NK)^+K'N$ is optimal as well, see the proof of Lemma 2, and suitable choices of $D \in \mathrm{NND}(s)$ and $B \in \mathbb{R}^{s \times (k-s)}$ give $N_K = [I_s : B]'D[I_s : B]$. Thus M is optimal if and only if $j_p(C) \cdot (j_p)^0(D) = \operatorname{trace} CD = 1$, by Theorems 3 and 4. By Lemma 3, then, $D = C^{p-1}/\operatorname{trace} C^p$ if $p > -\infty$, and $D \in \{1/\lambda_{\min}(C)\} \operatorname{conv} S$ if $p = -\infty$. In case of optimality Condition (2) implies $MNK = K'CD$, and $BM_{22} = -M_{12}$ and $EBM_{22} = -EM_{12}$, respectively.

A prominent application pertains to j_0-, i.e., D-optimality for s out of k parameters, see Kiefer (1961, Theorem 2), Karlin and Studden (1966, Theorem 6.1), Atwood (1969, Theorem 3.2). While the candidates for the matrix B are restricted to be solutions of $BM_{22} = -M_{12}$ the multiplicity which is possible when M is singular has caused a great many difficulties, see Atwood (1969, p. 1579). In the present context it is easy to see that the matrices G of Theorem 8, and B of Corollary 8.1 are connected through $(K'M^-K)^{-1}K'G = [I_s : B]$.

Since j_p is strictly concave for $p \in]-\infty, +1[$ multiplicity of j_p-optimal information matrices is discussed in Corollary 5.3. When $p = -\infty$ the proof of Corollary 5.3, in conjunction with Lemma 3, still yields the necessary condition that $AG'KE = KE$ must hold in order that both M and A be $j_{-\infty}$-optimal. Notice that Theorem 3 of Kiefer (1961) is close in spirit to our Corollary 5.3.

In Corollary 7.1 the matrix A determining the weights of a j_p-optimal design measure turns out to have entries, for $p > -\infty$,

$$a_{hi} = \{f(x_h)'G'KC^{1+p/2}K'Gf(x_i)\}^2/\operatorname{trace} C^p;$$

the uniform bound $\lambda_{\max}(CD)$ becomes $\lambda_{\max}(C^p)/\operatorname{trace} C^p$. This generalizes the bound $1/s$ for j_0-optimality for $K'\beta$, due to Atwood (1973, Theorem 4). For $p = -\infty$ use the matrix E of Theorem 8 to obtain

$$a_{hi} = \{\lambda_{\min}(C) \cdot f(x_h)'G'KE^{1/2}K'Gf(x_i)\}^2,$$

and uniform bound $\lambda_{\max}(CD) = \lambda_{\max}(E)$. Notice that $\lambda_{\max}(E) < 1$ unless rank $E = 1$, the latter necessarily being the case when $\lambda_{\max}(K'M^-K)$ is a simple eigenvalue.

The discussion clearly illustrates that in the j_p-family the member j_0, though best known, is least representative: It hides the conjugacy correspondence between p and q as in Lemma 3, and it obscures the fact that the bound $\lambda_{\max}(CD)$ for optimal weights depends, in general, on a predetermined information matrix.

Another typical instance is encountered in Corollary 7.2, where for j_0 property (c) simply requires g itself to take its values in the closed Euclidean unit ball of $\mathbb{R}^s$.

Up to this point no use has been made of the fact that the j_p-criteria are orthogonally invariant. However, if $j_p(C)$ is redefined to be the generalized mean of order p of the *positive* eigenvalues of C we may replace $(K'A^-K)^{-1}$ in Definition 2 by $(K'A^-K)^+$, and thus dispense with the hypothesis that the $k\times s$ matrix K must have full column rank s. While the set $\mathfrak{A}(K)$ remains unchanged, some provision is needed in order to circumvent negative powers of the now possibly singular matrix $K'M^-K$.

Corollary 8.2 (Arbitrary K). *Suppose K is a non-zero $k\times t$ matrix of rank $s\leqq t$; let $M\in\mathfrak{M}$ be an information matrix which lies in $\mathfrak{A}(K)$, and let C be the matrix $(K'M^-K)^+$. If $p>-\infty$, then M has $\mathfrak{M}$-maximal j_p-information for $K'\beta$ if and only if there exists a g-inverse G of M such that*

$$\operatorname{trace} K'GAG'KC(K'M^-K)^{1-p}C\leqq\operatorname{trace} C(K'M^-K)^{1-p},\quad \textit{for all } A\in\mathfrak{M}.$$

If $p=-\infty$ and $\operatorname{conv} S$ denotes the convex hull of all $t\times t$ matrices of the form zz' such that z lies in the range of K' and is an eigenvector of $K'M^-K$ corresponding to $\lambda_{\max}(K'M^-K)$ with Euclidean norm 1, then Theorem 8 holds verbatim.

Proof. Decompose K into HV' with some $k\times s$ matrix H and some $t\times s$ matrix V which satisfies $V'V=I_s$. Then $\mathfrak{A}(K)=\mathfrak{A}(H)$, and $H'M^-H$ is positive definite having the same positive eigenvalues and multiplicities as $K'M^-K$. Thus Theorem 8 holds with H replacing K. Since $V(H'M^-H)^{-1}V'=(K'M^-K)^+$, the assertion follows.

The matrix $C=(K'M^-K)^+$ is often called the $\mathfrak{C}$-matrix associated with M, see e.g., Kraft (1978, p. 200). For block design models, Kiefer (1958, Lemma 2.2; 1975, Proposition 1') investigates simultaneous optimality with respect to a large class of criteria, the present situation allows the following analogue.

Corollary 8.3 (Simultaneous optimality). *Suppose K is a non-zero $k\times t$ matrix of rank $s\leqq t$, and $p_0>-\infty$; let $M\in\mathfrak{M}$ have $\mathfrak{M}$-maximal j_{p_0}-information for $K'\beta$. If $(K'M^-K)^+$ equals ρK^+K for some $\rho>0$, then M has $\mathfrak{M}$-maximal j_p-information ρ for $K'\beta$, for all $p\in[-\infty,+1]$.*

Proof. The inequalities in Corollary 8.2 simplify to $\operatorname{trace} K'GAG'K\leqq s/\rho$; for $p=-\infty$ choose $E=K^+K/s$.

Another application of rank deficient matrices K pertains to optimality criteria which are linear in the sense of Fedorov (1972, Section 2.9). When a matrix $L\in\mathrm{NND}(k)$ is fixed, then an information matrix $M\in M(\Xi)$ is said to be *L-optimal* if it minimizes $\operatorname{trace} M^-L$ among all information matrices in $M(\Xi)$ whose range contains the range of L. Thus M is L-optimal if and only if M has $M(\Xi)$-maximal

j_{-1}-information for $K'\beta$ and $KK'=L$. Hence there exists an L-optimal information matrix, by Theorem 2, and $M\in M(\Xi)\cap\mathfrak{A}(L)$ is L-optimal if and only if there exists a g-inverse G of M which satisfies $f(x)'G'LGf(x)\leqq \text{trace } M^-L$, for all $x\in\mathfrak{X}$, by Corollary 8.2. This improves Theorem 8.2 of Karlin and Studden (1966), and Theorem 2.9.2 of Fedorov (1972). If L is replaced by KK' then j_{-1}-optimality for $K'\beta$ is characterized; since $R=G'K$ solves $MR=K$ this rederives Theorem 4.3.1 of Fellman (1974). Moreover, Theorem 7.1(ii) of Karlin and Studden (1966) may be extended and connected with Theorem 4.3.2 of Fellman (1974) in the following way. Recall that an information matrix $M\in\mathfrak{M}$ is called admissible if no other information matrix $A\in\mathfrak{M}$ satisfies $A\geqslant M$.

Corollary 8.4 (Admissibility). *Let $M\in\mathfrak{M}$ be an information matrix. If M is admissible, then* (i) *M has $\mathfrak{M}$-maximal $j_{-\infty}$-information for $K'\beta$, whenever $KK'=M$,* (ii) *there exists a matrix $N\in\mathfrak{M}^0\cap\mathrm{NND}(k)$ such that* trace $MN=1$, *and* (iii) *M has $\mathfrak{M}$-maximal j_{-1}-information for $K'\beta$ and M is L-optimal, whenever $L=KK'=MNM$.*

Proof. Let $A\in\mathfrak{M}$ have $\mathfrak{M}$-maximal $j_{-\infty}$-information for $K'\beta$ with $KK'=M$, and let N be an optimal solution of the dual problem. Then Theorem 4 and Lemma 3 yield $\lambda_{\max}(K'A^-K)=1/j_{-\infty}\circ J(A)=sj_1(K'NK)=\text{trace } MN\leqq 1$. This entails $K'A^-K\leqslant I_t$, and $K'^+K'A^-KK^+\leqslant M^+$. Hence

$$M\leqslant (K'^+K'A^-KK^+)^+=K(K'A^-K)^+K'\leqslant A,$$

the first inequality follows from Theorem 3.1 in Milliken and Akdeniz (1977), the second equality is immediate, and the third inequality is the second step in the proof of Theorem 3, with M replaced by A, and N replaced by uu'. Since admissibility forces M and A to coincide M is the unique $j_{-\infty}$-optimal information matrix for $K'\beta$, and trace $MN=1$. Assertion (iii) follows from $j_{-1}\circ J(M)=s=1/\{sj_{1/2}(K'NK)\}$ and Theorem 4, since $(K'NK)^{1/2}=K'M^-K$.

Notice that the j_p-family is related to the j_L-family, also introduced in Section 2, through $sj_1=j_{I_s}$, and $j_{-\infty}=\inf_{\|z\|=1} j_{zz'}$. The polar functions $(j_L)^0$ furnish yet another class of information functionals. For $L\in\mathrm{NND}(s)$, $L\neq 0$, let $\mathfrak{A}(L)$ denote the set of those matrices $D\in\mathrm{NND}(s)$ whose range contains the range of L.

Lemma 4. *The polar function of the information functional j_L is given by $(j_L)^0(D)=1/\lambda_{\max}(D^-L)$ if $D\in\mathfrak{A}(L)$, $(j_L)^0(D)=0$ otherwise. If a matrix $C\in\mathrm{PD}(s)$ is given, then the unique matrix $D\in\mathrm{NND}(s)$ which solves $j_L(C)\cdot(j_L)^0(D)=$* trace $CD=1$ *is $D=L/\text{trace } CL$.*

Proof. Let r be the rank of L, and the choose a matrix $K\in\mathbb{R}^{s\times r}$ such that $L=KK'$. In case $D\in\mathfrak{A}(L)$ the proofs of Theorem 3 and Lemma 3 give $\langle C,D\rangle\geqq$

$\langle (K'D^-K)^{-1}, K'CK\rangle \geqq \lambda_{\min}\{(K'D^-K)^{-1}\} j_L(C)$. Hence

$$\begin{aligned}\lambda_{\min}\{(K'D^-K)^{-1}\} &\leqq \inf_C \langle C, D\rangle / j_L(C)\\ &\leqq \inf_{\|z\|=1, \varepsilon>0} \langle D^+K(K'D^-K)^{-1}zz'(K'D^-K)^{-1}K'D^+\\ &\quad + \varepsilon I_s, D\rangle/(1+\varepsilon \operatorname{trace} KK')\\ &\leqq \inf_{\|z\|=1} z'(K'D^-K)^{-1}z = \lambda_{\min}\{(K'D^-K)^{-1}\},\end{aligned}$$

and $(j_L)^0(D) = 1/\lambda_{\max}(D^-L)$.
In case $D \notin \mathfrak{A}(L)$ choose a vector z in the nullspace of D which is not in the nullspace of L, then

$$(j_L)^0(D) \leqq \inf_{\varepsilon>0} \langle I_s + \varepsilon zz', D\rangle / \langle I_s + \varepsilon zz', L\rangle = 0.$$

Given $C \in \mathrm{PD}(s)$ and $D \in \mathfrak{A}(L)$, equality holds in $\langle C, D\rangle \geqq \langle (K'D^-K)^{-1}, K'CK\rangle$ if and only if $D = KEK'$ for some matrix $E \in \mathrm{PD}(r)$; for Condition (2) entails $D = K(K'D^-K)^{-1}K'$, while the converse follows upon choosing for KEK' the g-inverse $K(K'K)^{-1}E^{-1}(K'K)^{-1}K'$. Since $\langle E, K'CK\rangle = \lambda_{\min}(E)(\operatorname{trace} K'CK)$ necessitates $E = \alpha I_r$, the proof is complete.

The final section will illustrate some of the results above.

6. Examples

6.1. Trigonometric design

One of the smoothest examples is provided by the trigonometric regression function $f(x) = (1, \cos x, \ldots, \cos kx, \sin x, \ldots, \sin kx)'$ on the 'unit circle' $\mathfrak{X} = [0, 2\pi[$, see Fedorov (1972, Section 2.4), Kraft (1978, Section 19(c)). For every sample size $n > 2k+1$ every design measure which assigns weight $1/n$ to n equidistant points on the unit circle has $(2k+1)\times(2k+1)$ information matrix $M = \operatorname{diag}(1, \frac{1}{2}, \ldots, \frac{1}{2})$, with $M(\Xi)$-maximal j_0-information for β. Moreover, any such design is j_p-optimal for β, for all $p \in [-\infty, +1]$, by Corollary 8.1. Hence for simultaneous optimality the sufficient condition of Corollary 8.3 is not, in general, necessary. It *is* necessary, though, in order that the optimal value function $v(p)$ be constant; in the present example v varies from $v(-\infty) = \frac{1}{2}$ to $v(1) = \frac{1}{2} + 1/(4k+2)$. Also notice that *every* design measure with non-singular information matrix is j_1-optimal for β, demonstrating the particularly poor performance of j_1, and that many choices for E are feasible to verify $j_{-\infty}$-optimality in Theorem 8 or Corollary 8.1.

6.2. *Quadratic design*

One of the simplest examples is provided by the quadratic regression function $f(x)=(1, x, x^2)'$ over the symmetric interval $\mathfrak{X}=[-1,+1]$. Let ξ_α be the design measure $\xi_\alpha(0)=\alpha$, $\xi_\alpha(-1)=\xi_\alpha(+1)=(1-\alpha)/2$, with information matrix

$$M_\alpha=\begin{pmatrix} 1 & 0 & 1-\alpha \\ 0 & 1-\alpha & 0 \\ 1-\alpha & 0 & 1-\alpha \end{pmatrix},$$

$$M_\alpha^{-1}=\begin{pmatrix} 1/\alpha & 0 & -1/\alpha \\ 0 & 1/(1-\alpha) & 0 \\ -1/\alpha & 0 & 1/\{\alpha(1-\alpha)\} \end{pmatrix}.$$

6.2.1

There exists no j_1-optimal design measure for β. For it follows from the equation $\{3j_{-\infty}(I_3/3)\}^{-1}=1=\lim_{\alpha\searrow 0} j_1(M_\alpha)$ that $I_3/3$ is an optimal solution of the dual problem. The points ± 1, determined from Condition (1), cannot support a design measure which has a non-singular information matrix.

6.2.2

The unique $j_{-\infty}$-optimal design measure for β is $\xi_{3/5}$, and $1/\sqrt{5}=\|c\|$, $c=(-\frac{1}{5}, 0, \frac{2}{5})'$, is an in-ball radius of the regression ball $\mathfrak{R}$. For one has $\frac{1}{5}=\lambda_{\min}(M_{3/5})\leqq v(-\infty)\leqq r^2$, where $r^2=r^4/(j_{-\infty})^0(dd')$ is the bound derived before Theorem 7. On the other hand r is connected to the regression norm ρ (Pukelsheim 1979, equation (2.3), Theorem 1) through $r^2\leqq\|c\|^2/\{\rho(c)\}^2\leqq\|c\|^2/c'Nc=\frac{1}{5}$, when the $f(\mathfrak{X})$ covering cylinder N is chosen to be $(-1,0,2)'(-1,0,2)$. Hence $r=1/\sqrt{5}$, and $\rho(c)=1$, so that c is a boundary point of $\mathfrak{R}$, and $M_{3/5}$ is $j_{-\infty}$-optimal for β (Kiefer 1974, p. 868). Furthermore, $M_{3/5}$ is optimal for $c'\beta$, since $c'M_{3/5}^{-1}c=1=\{\rho(c)\}^2$, and c is an eigenvector of $M_{3/5}$ corresponding to the eigenvalue $r^2=\frac{1}{5}$. For $j_{-\infty}$-optimality Corollary 8.2 with $E=5cc'$ determines -1, 0, $+1$ as only possible points of support; Corollary 5.3, then, proves uniqueness of $M_{3/5}$.

6.2.3

The unique j_0-optimal design measure for $(\beta_1, \beta_2)'$ is $\xi_{1/2}$, by Theorem 8 and Corollary 5.3. In fact, this design is j_p-optimal for $(\beta_1, \beta_2)'$, for all $p\in[-\infty, +1]$, by Corollary 8.3.

6.2.4

If $p\in\,]-\infty, +1[$ then the unique j_p-optimal design measure for $(\beta_2, \beta_3)'$ is ξ_α with α being defined implicitly by $\alpha^{1-p}+2\alpha=1$. For when Theorem 8 is applied to any design ξ_α one obtains the condition on α, and the possible support points -1, 0, $+1$; again Corollary 5.3 proves uniqueness. The weight function $\alpha(p)$ is strictly decreasing, with derivative $\alpha'(p)=(\log\alpha)\alpha^{1-p}/\{\alpha^{-p}(1-p)+2\}<0$. For the five arguments '$-\infty$', -1, 0, $\frac{1}{2}$, '$+1$' the values of $\alpha(p)$ are $\frac{1}{2}$, $\sqrt{2}-1$, $\frac{1}{3}$ (Kiefer 1961, p. 312), $\frac{1}{4}$, 0, while the respective optimal j_p-information $v(p)$ is $\frac{1}{4}$, $2/(3+\sqrt{8})$, $2/\sqrt{27}$, $\frac{27}{64}$, $\frac{1}{2}$. Note that $\frac{1}{4}$ also is the lower bound for $v(-\infty)$ derived before Theorem 7. In fact, the unique $j_{-\infty}$-optimal design measure for $(\beta_2, \beta_3)'$ is $\xi_{1/2}$, with $j_{-\infty}$-information $\frac{1}{4}$.

6.2.5

The j_1-optimal value for $(\beta_2, \beta_3)'$ is $\frac{1}{2}$, but there exists no j_1-optimal design measure for $(\beta_2, \beta_3)'$. For the $f(\mathfrak{X})$ covering cylinder

$$N = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

satisfies $\{2j_{-\infty}(K'NK)\}^{-1} = \frac{1}{2} = \lim_{\alpha \searrow 0} j_1 \circ J(M_\alpha)$, hence N is an optimal solution of the dual problem and determines the support points $-1, 0, +1$. If a measure ξ with this support is optimal then it must have information matrix M_0, by comparing the first row of either side in Condition (2), although M_0 does not lie in $\mathfrak{A}(K)$.

6.3. A contraction example

Consider a projection MG where G is a contracting g-inverse of $M \in M(\Xi)$. If $k = 2$ then MG is a contraction of the regression ball $\mathfrak{R}$; in this case there is no need to go through the construction of Lemma 2, but G may be visualized from the geometry of $\mathfrak{R}$, see Pukelsheim (1979). If $k > 2$ then MG need not be a contraction of $\mathfrak{R}$: Let $\mathfrak{R}$ be the regression ball generated by the identity function and

$$\mathfrak{X} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -1 \\ 1 \end{pmatrix}; \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Then the cross-section of $\mathfrak{R}$ with the (x, y)-plane is the unit square, with vertices on the coordinate axes. Therefore the measure which assigns weight $\frac{1}{2}$ to the last two levels in $\mathfrak{X}$ is optimal for $\frac{1}{2}\beta_1 + \frac{1}{2}\beta_2$, by Theorem 1 in Pukelsheim (1979). However, no projection onto the (x, y)-plane can be a contraction of $\mathfrak{R}$ since two of the edges of $\mathfrak{R}$ have been twisted out of their 'natural' vertical positions.

Acknowledgement

It is a pleasure to acknowledge valuable and helpful discussions with Professors J. Kiefer, O. Krafft, D.M. Titterington, W. Weil and H. Witting. In particular it was Professor Krafft who drew my attention to the optimal design problem during the December 1976 Oberwolfach conference. Work on this paper begun while I was a postdoctoral fellow of the Deutsche Forschungsgemeinschaft with the Department of Statistics, Stanford University. An initial version of parts of this paper appeared as Technical Report No. 128 (May 1978), Department of Statistics, Stanford University.

References

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

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

Bandemer, H. et al. (1977). *Theorie und Anwendung der optimalen Versuchsplanung I.* Berlin, Akademie-Verlag.

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

Chernoff, H. (1972). *Sequential Analysis and Optimal Design.* Philadelphia, PA, Society for Industrial and Applied Mathematics.

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

Fedorov, V.V. (1972). *Theory of Optimal Experiments.* New York, Academic Press.

Fedorov, V.V. and M.B. Malyutov (1972). Optimal designs in linear regression problems. *Math. Operationsforsch. Statist.* 3, 281–308.

Fellman, J. (1974). On the allocation of linear observations. *Commentationes Physico-Mathematicae (Finska Vetenskaps Societeten)* 44, 27–77.

Gaffke, N. (1979). Some classes of optimality criteria and optimal designs for complete two-way-layouts. *Ann. Statist.*, to appear.

Gaffke, N. and O. Krafft (1979). Matrix inequalities in the Löwner ordering. In: B. Korte, ed., *Modern Applied Mathematics: Optimization and Operations Research*, Amsterdam, North-Holland, to appear.

Gribik, P.R. and K.O. Kortanek (1977). Equivalence theorems and cutting plane algorithms for a class of experimental design problems. *SIAM J. Appl. Math.* 32, 232–259.

Karlin, S. and W.J. Studden (1966). Optimal experimental designs. *Ann. Math. Statist.* 37, 783–815.

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

Kiefer, J. (1959). Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* 21, 272–304.

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

Kiefer, J. (1962). An extremum result. *Canad. J. Math.* 14, 597–601.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* 2, 849–879.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In: J. N. Srivastava, (ed.), *A Survey of Statistical Design and Linear Models.* Amsterdam, North-Holland, 333–353.

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

Krafft, O. (1978). *Lineare statistische Modelle und optimale Versuchspläne.* Göttingen, Vandenhoeck & Ruprecht.

Kurotschka, V. (1978). Optimal design of complex experiments with qualitative factors of influence. *Comm. Statist. A—Theory Methods* 7, 1363–1378.

Lamotte, L.R. (1977). A canonical form for the general linear model. *Ann. Statist.* 5, 787–789.

Milliken, G.A. and F. Akdeniz (1977). A theorem on the difference of the generalized inverses of two nonnegative matrices. *Comm. Statist. A—Theory Methods* 6, 73–79.

Pázman, A. (1980). Singular experimental designs (standard and Hilbert-space approaches). *Math. Operationsforsch. Statist. Ser. Statist.* 11, 137–149.

Pukelsheim, F. (1979). On c-optimal design measures. *Math. Operationsforsch. Statist. Ser. Statist.*, to appear.

Pukelsheim, F. and G.P.H. Styan (1979). Convexity and monotonicity properties of dispersion matrices in linear models. Typescript.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications. Second Edition.* New York, John Wiley.

Rockafellar, R.T. (1970). *Convex Analysis.* Princeton, NJ, Princeton University Press.

Sibson, R. (1974). D_A-optimality and duality. In: J. Gani, K. Sarkadi and I. Vincze, ed., *Progress in Statistics. European Meeting of Statisticians, Budapest (Hungary), 1972. Colloquia Mathematica Societatis Janos Bolyai, 9,* Vol. III. Amsterdam: North-Holland, 677–692.

Sibson, R. and A. Kenny (1975). Coefficients in D-optimal experimental design. *J. Roy. Statist. Soc. Ser. B* 37, 288–292.

Silvey, S.D. (1978). Optimal design measures with singular information matrices. *Biometrika* 65, 553–559.

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

Silvey, S.D. and D.M. Titterington (1974). A Lagrangian approach to optimal design. *Biometrika* 61, 299–302.

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

Titterington, D.M. (1980). Geometric approaches to design of experiment. *Math. Operationsforsch. Statist. Ser. Statist.* 11, 151–163.

Wald, A. (1943). On the efficient design of statistical investigations. *Ann. Math. Statist.* 14, 134–140.

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

Wynn, H.P. (1972). Results in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* 34, 133–147. Discussion of Dr. Wynn's and of Dr. Laycock's papers. *J. Roy. Statist. Soc. Ser. B* 34, 170–186.