OPTIMALITY CRITERIA FOR EXPERIMENTAL DESIGNS

Friedrich PUKELSHEIM, Cornell University*

BU-943-M† 14 October 1987

ABSTRACT

Performance measures for experimental designs and their moment matrices are discussed. On top ranks optimality of information matrices under the Loewner ordering. This concept coincides with dispersion optimality and uniform optimality. An equivalence theorem is presented, but Loewner optimal designs mostly fail to exist. Information functions are introduced as weaker criteria that are isotonic relative to the Loewner ordering, concave, and positively homogeneous. The need for these properties is developed in some detail, as are the technical aspects of upper semicontinuity and polar information functions. An existence theorem is proved which provides three sufficient conditions for the existence of optimal moment matrices that are also feasible. The general theory is substantiated by the family of p-means, which are the means of order p of the eigenvalues of the information matrix of the moment matrix of the design. The polar function of the p-mean is proportional to the q-mean, with p and q being conjugate numbers.

NOTE

This is a sequel to the reports

A Experimental Designs in Linear Models. Technical Report No. 234. Department of Statistics, Stanford University. July 1987. 20 pp.

B Optimal Designs for One-Dimensional Parameter Systems. Technical Report No. 235. Department of Statistics, Stanford University. July 1987. 13 pp.

C Information Matrices. Technical Report No. 240. Department of Statistics, Stanford University. August 1987. 20 pp.

Results from those reports are quoted with a preceding letter *A, B,* or *C*. Complete references will be included in a subsequent report.

---

* On leave from the Institut für Mathematik der Universität Augsburg. The work of this author is partially supported by the Mathematical Sciences Institute, Cornell University.

† Paper BU-943-M in the Biometrics Unit, Cornell University.

Optimality Criteria for Experimental Designs

D1 Sets of Competing Moment Matrices
D2 Moment Matrices with Maximum Range and Rank
D3 Maximum Range for the Twoway Classification
D4 Loewner Optimality
D5 Dispersion Optimality and Uniform Optimality
D6 Equivalence Theorem for Loewner Optimality
D7 Nonexistence of Loewner Optimal Designs
D8 Loewner Optimal Designs for the Twoway Classification
D9 Real Criterion Functions
D10 Monotonicity, Concavity, and Homogeneity
D11 Nonnegativity and Positivity
D12 Information Functions
D13 Unit Level Sets
D14 Functional Operations
D15 Polar Information Functions
D16 Polarity Correspondence
D17 Semicontinuity and Regularization
D18 Compositions with the Information Matrix Mapping
D19 The General Design Problem
D20 Existence Theorem
D21 c-Optimality Revisited
D22 D-, A-, E-, T-Optimality
D23 The p-Mean Family
D24 The p-Means as Information Functions
D25 Rearrangement Inequalities
D26 Polarity Lemma
D27 Polarity Equation

Optimality Criteria for Experimental Designs

1 In its general form the design problem calls for finding a design ξ that is optimal, relative to some criterion, in a specified subclass of the set Ξ of all designs. A point has been made in Chapter C that statistically reasonable criteria depend on the design ξ only through its information matrix $C\big(M(\xi)\big)$ which in turn is a function of the moment matrix $M(\xi)$. Restricting attention to subclasses of designs therefore means to work with subsets $\mathcal{M}$ of the set of *all* moment matrices,

$$\mathcal{M} \subseteq M(\Xi).$$

Such a subset $\mathcal{M}$ in which optimal moment matrices are sought is called a *set of competing moment matrices.*

Throughout the sequel we make the *grand assumption* that there exists at least one competing moment matrix that is feasible for the subsystem $K'\theta$ under investigation,

$$\mathcal{M} \cap \mathcal{A}(K) \neq \emptyset.$$

Furthermore this subset $\mathcal{M}$ will often be convex, two simple consequences of which are the following.

2 LEMMA. *Let $\mathcal{M}$ be a convex set of nonnegative definite $k \times k$ matrices. Then there exists a matrix $M \in \mathcal{M}$ whose range and rank are a maximum,*

$$\operatorname{range} A \subseteq \operatorname{range} M, \quad \operatorname{rank} A \leq \operatorname{rank} M \qquad \textit{for all } A \in \mathcal{M}.$$

Moreover the information matrix mapping C permits regularization along straight lines in $\mathcal{M}$, that is, whenever B lies both in $\mathcal{M}$ and the feasibility cone $\mathcal{A}(K)$ then

$$C(A) = \lim_{\alpha \to 0} C\big((1-\alpha)A + \alpha B\big) \qquad \textit{for all } A \in \mathcal{M}.$$

PROOF. Of course we assume $\mathcal{M}$ to be nonempty. Choose a matrix $M \in \mathcal{M}$ that attains the maximal rank,

$$\operatorname{rank} M = \max\{\operatorname{rank} A : A \in \mathcal{M}\}.$$

We show that this matrix also has a range as large as possible. Otherwise there is a matrix $A \in \mathcal{M}$ whose range is not included in the range of M. Because of convexity $\mathcal{M}$ contains $B = \frac{1}{2}A + \frac{1}{2}M$. By Lemma *B3* the matrix B has a larger rank than M, contrary to assumption.

Furthermore $\mathcal{M}$ contains the path $(1-\alpha)A + \alpha B$ connecting $A, B \in \mathcal{M}$. Positive homogeneity of C, established in Theorem *C13*, permits to extract the factor $1-\alpha$, giving

$$C\big((1-\alpha)A + \alpha B\big) = (1-\alpha)C(A + \frac{\alpha}{1-\alpha}B).$$

This has limit $C(A)$ as α tends to zero, by upper semicontinuity. $\Diamond$

In many cases, though not in all, there exist positive definite competing moment matrices, and the maximal rank in $\mathcal{M}$ then equals k.

3 The maximal rank of moment matrices in a twoway classification is $a+b-1$. As outlined in Section *A26* we identify designs ξ with their weight matrices W and write $W \in \Xi$ etc. We have already seen in Section *A26* that a weight matrix $W \in \Xi$ has a degenerate moment matrix,

$$M(W)\begin{pmatrix} +1_a \\ -1_b \end{pmatrix} = \begin{pmatrix} \Delta_r & W \\ W' & \Delta_s \end{pmatrix}\begin{pmatrix} +1_a \\ -1_b \end{pmatrix} = \begin{pmatrix} r-r \\ s-s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hence the maximal rank is at most $a+b-1$. This value is actually attained for the *uniform design* with weight matrix $\bar{1}_a\bar{1}_b'$, assigning uniform mass $1/(ab)$ to every point (i,j) in the experimental domain $\mathcal{U} = \{1,\dots,c\} \times \{1,\dots,b\}$. It suffices to show that

$$M(\bar{1}_a\bar{1}_b')\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{a}I_a & \frac{1}{ab}1_a1_b' \\ \frac{1}{ab}1_b1_a' & \frac{1}{b}I_b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$$

occurs only when $x = \delta 1_a$ and $y = -\delta 1_b$, for some $\delta \in \mathbb{R}$. But $\frac{1}{a}x + \frac{1}{ab}(\sum y_j)1_a = 0$ entails $x = \alpha 1_a$, while $y = \beta 1_b$ follows from $\frac{1}{ab}(\sum x_i)1_b + \frac{1}{b}y = 0$. Replacing x and y by their new expressions leads to $\alpha = -\beta = \delta$, say.

Another important class of designs for the twoway classification are the *product designs.* By definition their weights w_{ij} are the product of the row-sum r_i and the column-sum s_j. In matrix terms this means that the weight matrices W are of the form

$$W = rs',$$

for some row-sum vector r and some column-sum vector s. We call a vector r *positive* and write $r > 0$ when all its entries are positive. For positive row- and column-sum vectors r and s the product design rs' again attains the maximal rank, that is,

$$\operatorname{rank} M(rs') = a+b-1 \qquad \text{when } r,s > 0.$$

The uniform design is just one distinguished instance of this.

The twoway classification also serves as a prime example that subsets $\mathcal{M}$ of the full set $M(\Xi)$ of all moment matrices are indeed of interest. For instance, fixing the row-sum vector means fixing the proportions of replication on levels $1, \ldots, a$ of factor A. Considering all weight matrices that achieve the given row-sum vector r leads to the subsets

$$\Xi(r) = \{ W \in \Xi : W 1_b = r \}, \qquad \mathcal{M}(r) = M(\Xi(r)).$$

When the prescribed vector r is positive the maximal rank in $\mathcal{M}(r)$ again is $a + b - 1$. In any case this set is convex. It is also compact being a closed subset of the compact set $M(\Xi)$, see Lemma *A25*.

It is these three properties that sets of competing moment matrices must enjoy in order to permit successful application of the theory to be developed: They should be convex, compact, and fulfill the grand assumption of Section *1* that they intersect the feasibility cone $\mathcal{A}(K)$ in question.

4 It is time now to be more specific about the meanings of optimality. The most satisfying criterion is Loewner optimality, that is, optimality relative to the Loewner ordering among information matrices. For the sake of transparency we assume that the coefficient matrix K has full column rank s. Then a moment matrix $M \in \mathcal{M}$ is called *Loewner optimal for $K'\theta$ in $\mathcal{M}$* when

$$C(M) \geq C(A) \qquad \text{for all } A \in \mathcal{M}.$$

(If K is rank deficient we simply require $M_K \geq A_K$.) Designs inherit the properties by their moment matrices. Given a subclass of designs, $\Xi' \subseteq \Xi$, a design $\xi \in \Xi'$ is called *Loewner optimal for $K'\theta$ in Ξ'* when its moment matrix $M(\xi)$ is Loewner optimal for $K'\theta$ in $M(\Xi')$.

Sections *C4–C10* leave no doubt that Loewner optimality is desirable for every inferential procedure: estimation, testing, and parametric modelling. The following theorem summarizes the various facets of Loewner optimality, the emphasis being on information matrices, dispersion matrices, and simultaneous c-optimality. Perhaps the last property comes closest to the notion of 'uniform optimality' that some authors prefer.

5 THEOREM. *Let $\mathcal{M}$ be a convex set of competing moment matrices. Then for every moment matrix $M \in \mathcal{M}$ the following statements are equivalent:*

a (Information optimality) *M is Loewner optimal for $K'\theta$ in $\mathcal{M}$.*

b (Dispersion optimality) *$M \in \mathcal{A}(K)$, and $K'M^-K \leq K'A^-K$ for all $A \in \mathcal{M} \cap \mathcal{A}(K)$.*

c (Uniform optimality) *M is optimal for $c'\theta$ in $\mathcal{M}$ for all vectors $c \neq 0$ in the range of K.*

PROOF. First we present a proof of a implying b when K has full column rank s. Assume a. By our grand assumption the intersection $\mathcal{M} \cap \mathcal{A}(K)$ contains a matrix B, say. Then $C(M) \geq C(B) > 0$ whence M lies in the feasibility cone $\mathcal{A}(K)$. For every other matrix $A \in \mathcal{M} \cap \mathcal{A}(K)$ the order reversing property of matrix inversion discussed in Section *A11* turns $(K'M^-K)^{-1} = C(M) \geq C(A) = (K'A^-K)^{-1}$ into $K'M^-K \leq K'A^-K$.

When K does not have full column rank we invoke Theorem *C15* to conclude that M lies in $\mathcal{A}(K)$, by way of

$$(\operatorname{range} M) \cap (\operatorname{range} K) = \operatorname{range} M_K \supseteq \operatorname{range} B_K = (\operatorname{range} B) \cap (\operatorname{range} K) = \operatorname{range} K.$$

The inclusion follows from $M_K \geq B_K$ and Lemma *B3*. Now Lemma *C22* applies to M and $A \in \mathcal{M} \cap \mathcal{A}(K)$, and yields $K'M^-K \leq K'A^-K$.

The converse implication from b to a uses the same order reversing properties on the intersection $\mathcal{M} \cap \mathcal{A}(K)$. The extension to all of $\mathcal{M}$ follows by means of regularization along straight lines in $\dot{\mathcal{M}}$, permitted according to Lemma *2*.

Next we show that b implies c. Take any vector in the range of K, $c = Kz$, say. The point is that optimality of M for $c'\theta$ refers to the feasibility cone, not $\mathcal{A}(K)$, but $\mathcal{A}(c)$,

$$c'M^-c \leq c'A^-c \qquad \text{for all } \mathcal{M} \cap \mathcal{A}(c).$$

Again regularization is used. Choose an arbitrary moment matrix $A \in \mathcal{M} \cap \mathcal{A}(c)$. Then $(1-\alpha)A + \alpha M$ lies in $\mathcal{M} \cap \mathcal{A}(K)$ whence, by assumption,

$$c'M^-c = z'K'M^-Kz \leq z'K'\big((1-\alpha)A + \alpha M\big)^-Kz = c'\big((1-\alpha)A + \alpha M\big)^-c.$$

Letting α tend to zero we obtain $c'M^-c \leq c'A^-c$.

Finally we establish that c implies b. If M lies in the feasibility cone $\mathcal{A}(c)$ for all $c \in \operatorname{range} K$, then it also lies in $\mathcal{A}(K)$. The inequalities $z'K'M^-Kz \leq z'K'A^-Kz$, with arbitrary vector z, evidently prove the required matrix inequality. $\diamondsuit$

Property c enables us to deduce an equivalence theorem for Loewner optimality, similar in nature to the Equivalence Theorem *B22* for c-optimality. There we concentrated on the set $M(\Xi)$ of all moment matrices; we now use—and prove later—that Theorem *B22* remains valid with the set $M(\Xi)$ replaced by a set $\mathcal{M}$ of competing matrices that is convex and compact.

6 THEOREM. *Let $\mathcal{M}$ be a convex and compact set of competing moment matrices, and let $M \in \mathcal{M}$ have maximal rank. Then M is Loewner optimal for $K'\theta$ in $\mathcal{M}$ if and only if*

$$K'M^-AM^-K \leq K'M^-K \qquad \textit{for all } A \in \mathcal{M}.$$

PROOF. First we verify that the product $K'GA$ is invariant to choice of generalized inverse $G \in M^-$. Due to the maximality assumption the range of M includes the range of K as well as the range of every other competing moment matrix A. Thus we can write $K = MW$ and $A = MH$, and obtain

$$K'GA = W'MGMH = W'MH \qquad \text{for all } G \in M^-.$$

It follows that the left hand side $K'M^-AM^-K$ is well defined.

The converse part provides a sufficient condition for optimality. When K has full column rank s its proof can be arranged just as the corresponding part of the proof of Theorem *B22*. Instead we present an argument which does not include the rank of K, utilizing the matrices A_K from Section *C21*. The proof of Lemma *C22* has shown that every generalized inverse G of M also is a generalized inverse of M_K. Thus M_KM^- projects onto the range of M_K. The proof of Theorem *C15* shows that the latter is equal to the range of K. Hence we obtain $QK = K$ for $Q = M_KM^- = K(K'M^-K)^-K'M^-$, and

$$\begin{aligned} A_K &= \min\nolimits_{QK=K} QAQ' \\ &\leq K(K'M^-K)^-K'M^-AM^-K(K'M^-K)^-K' \\ &\leq K(K'M^-K)^-K'M^-K(K'M^-K)^-K' \\ &= K(K'M^-K)^-K' = M_K. \end{aligned}$$

The second inequality invokes the inequality from the theorem; the next equality uses $K'M^-K(K'M^-K)^-K' = K'$ which follows from Lemma *A17*.

The direct part, necessity of the condition, is based on Theorem *5c*. This part invokes Theorem *B22* with $\mathcal{M}$ replacing $M(\Xi)$. Given a vector $c = Kz$ in the range of K there exists a generalized inverse $G \in M^-$—according to Theorem *B22* possibly dependant on c— such that

$$z'K'GAG'Kz \leq z'K'M^-Kz \qquad \text{for all } A \in \mathcal{M}.$$

But because of the maximality assumption the expression $z'K'GA$ is invariant to the choice of the generalized inverse G, and the left hand side becomes $z'K'M^-AM^-Kz$. Since z is arbitrary the desired matrix inequality follows. ◊

We reiterate that the necessity part of the proof hinges on Theorem *B22* which covers the case $\mathcal{M} = M(\Xi)$, only. It is an ironic twist at this stage of the development that the present theorem, with $\mathcal{M} = M(\Xi)$, is vacuous.

7 COROLLARY. *No moment matrix in $M(\Xi)$ is Loewner optimal for $K'\theta$ in $M(\Xi)$, except when the coeffcient matrix K has rank one.*

PROOF. Assume that $M(\xi)$ is a moment matrix which is Loewner optimal for $K'\theta$ in $M(\Xi)$, and let $x_1, \ldots, x_\ell$ be the support points of ξ. The matrix M necessarily has maximal rank, whence in Theorem *B22* we are allowed to replace G by M^-. Applying that theorem with $c = Kz$ to $A = x_i x_i'$ we obtain

$$c'M^- x_i x_i' M^- c \leq c'M^- c.$$

Here equality must obtain, since otherwise

$$c'M^- c = c'M^- M M^- c = \sum \xi(x_i) c'M^- x_i x_i' M^- c < \sum \xi(x_i) c'M^- c = c'M^- c.$$

Thus we have $z'K'M^- x_1 x_1' M^- Kz = z'K'M^- Kz$ for all $z \in \mathbb{R}^s$, and therefore the two matrices are equal. Lemma *A17* now yields the assertion,

$$\operatorname{rank} K = \operatorname{rank} K'M^- K = \operatorname{rank} K'M^- x_1 x_1' M^- K = 1. \qquad \diamondsuit$$

The destructive nature of this corollary is deceiving. Firstly, and above all, an equivalence theorem gives necessary and sufficient conditions for a design or a moment matrix to be optimal, and this is genuinely distinct from an existence statement. Indeed, the statements

—'If a design is optimal then it must look like this.'

—'If it looks like this then it must be optimal.'

in no way assert that an optimal design exists. Even if existence fails to hold they are logically true, though vacuous.

Secondly, however, we have exploited the knowledge from the Equivalence Theorem *6* to deduce nonexistence: Equivalence theorems do provide an indispensable and constructive tool to study the existence problem. This is complemented by the topological Existence Theorem *20* which refers to continuity and compactness properties. Therefore the present corollary ought to be taken as a manifestation of the constructive contribution that an equivalence theorem adds to the theory. We shall see later how to deduce from it valuable partial insight about the number of support points of an optimal design, their location, and their weights. The Elfving Theorem *B14* serves as a prototype, though results of a similar completeness are not available for the general design problem.

Thirdly the corollary stresses the role of the set $\mathcal{M}$ of competing moment matrices, saying that the set $M(\Xi)$ is too large to permit Loewner optimal moment matrices. A maximum is achieved only for certain smaller subsets $\mathcal{M}$. For instance if the subset consists of a single moment matrix, $\mathcal{M}_0 = \{M_0\}$, then it is trivially true that M_0 is Loewner optimal for $K'\theta$ in $\mathcal{M}_0$.

Subsets of competing moment matrices tend to be of interest for the reason that they show more structure. It has been the experience of the author that this structure often permits a direct derivation of Loewner optimality, circumventing Theorem *6*.

8 This section continues the discussion of Sections *3* and *C23* for the twoway classification. We concentrate on the symmetrized contrasts of factor A,

$$\begin{pmatrix} \alpha_1 - \bar{\alpha}. \\ \vdots \\ \alpha_a - \bar{\alpha}. \end{pmatrix} = (K_a, 0)\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = K\theta, \qquad K = \begin{pmatrix} K_a \\ 0 \end{pmatrix}.$$

Let r be a fixed positive rum-sum vector. We claim the following.

The product designs rs' with arbitrary column-sum vector s are the unique Loewner optimal designs for the symmetrized contrasts of factor A in the set of all designs with row-sum vector equal to r; the optimal information matrix for this parameter system is $\Delta_r - rr'$.

The information matrix for the symmetrized contrasts is determined by the C-matrix

$$\Delta_r - W\Delta_s^- W'.$$

Product designs have weight matrices $W = rs'$. Now $\Delta_s^- s$ is a vector with i^{th} entry equal to one or zero according as s_i is positive or vanishes. Therefore $s'\Delta_s^- s = \sum_{i:s_i>0} s_i = 1$, and

$$W\Delta_s^- W' = rs'\Delta_s^- sr' = rr'. \tag{†}$$

This proves that all product designs have the same C-matrix. Hence there is no loss in generality if from now on we assume the column-sum vector s to be positive, thus securing a maximal rank for the moment matrix $M(rs')$.

The moment matrix M, a generalized inverse G for it, and the product $K'G$ are then given by

$$M = \begin{pmatrix} \Delta_r & rs' \\ sr' & \Delta_s \end{pmatrix}, \qquad G = \begin{pmatrix} \Delta_r^{-1} & 0 \\ 0 & \Delta_s^{-1} - 1_b 1_b' \end{pmatrix}, \qquad K'G = (K_a\Delta_r^{-1}, 0).$$

In approaching Loewner optimality let A be an arbitrary moment matrix, partioned into four blocks A_{11} etc. The left hand side in the inequality of Theorem *6* turns into

$$K'M^-AM^-K = (K_a\Delta_r^{-1}, 0)\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} \Delta_r^{-1}K_a \\ 0 \end{pmatrix} = K_a\Delta_r^{-1}A_{11}\Delta_r^{-1}K_a.$$

The right hand side equals $K'M^-K = K_a\Delta_r^{-1}K_a$. Hence the two sides coincide when $A_{11} = \Delta_r$. With the notation of Section *3* this means that A lies in $\mathcal{M}(r)$, the subset of moment matrices corresponding to the class $\Xi(r)$ of weight matrices with prescribed row-sum vector r. Thus Theorem *6* proves that rs' is Loewner optimal in $\Xi(r)$,

$$\Delta_r - rr' \geq \Delta_r - W\Delta_s^- W' \qquad \text{for all } W \in \Xi(r).$$

Every optimal weight matrix W must satisfy equation (†), forcing W to have rank one and hence being of the form $W = rs'$. This establishes our claim.

Brief contemplation opens a more direct route to this result. For an arbitrary weight matrix $W \in \Xi(r)$ with column-sum vector s we not only have $s'\Delta_s^- s = 1$ but also $W\Delta_s^- s = r$. This yields the inequality

$$\begin{aligned} 0 &\leq (W - rs')\Delta_s^-(W' - sr') \\ &= W\Delta_s^- W' - W\Delta_s^- sr' - rs'\Delta_s^- W' + rs'\Delta_s^- sr' \\ &= W\Delta_s^- W' - rr', \end{aligned}$$

with equality if and only if $W = rs'$. This, too, establishes our claim.

When a moment matrix is feasible for the symmetrized contrasts its weight matrices must have positive row-sum vectors. For if $r_i = 0$ then the i^{th} row and column of the C-matrix $\Delta_r - W\Delta_s^- W'$ vanish; since its nullity is larger than one its range cannot include the range of K_a. The class of all designs with feasible moment matrices for the symmetrized contrasts thus decomposes into the cross sections $\Xi(r)$ with positive row-sum vectors r. Within each cross section the information matrix $\Delta_r - rr'$ is Loewner optimal, while Loewner optimality between cross sections is ruled out by Corollary *7*. How does one go on from here?

There are essentially two ways leading out of the dilemma. Either determine the subset of admissible information matrices, that is, those information matrices that cannot be improved upon in the Loewner ordering. Or select a realvalued optimality criterion ϕ, and look for ϕ-optimal information matrices. The discussion of optimality criteria ϕ is taken up first, in the remainder of this chapter. The specific criterion of eigenvalue optimality will be instrumental in attacking the admissibility problem.

9 A realvalued criterion is a function ϕ from the cone of nonnegative definite matrices into the real line, with properties that make it eligible to measure largeness of information matrices. A passage from a highdimensional matrix cone to the onedimensional real line can salvage only partial aspects and the question is, which.

There is no such thing as a single optimality criterion that fits all experimental situations and pleases every human mind.

When the coefficient matrix K is of full column rank we take ϕ to be defined on the cone $\mathrm{NND}(s)$,

$$\phi : \mathrm{NND}(s) \mapsto \mathbb{R}.$$

When K is rank deficient the information matrices A_K are of order $k \times k$ and their range is included in the range of K, compare Theorem *C15*. Then the domain of definition for ϕ ought to be the subcone of nonnegative definite $k \times k$ matrices whose range is included in the range of K. But this subcone is clearly isomorphic to the cone $\mathrm{NND}(\operatorname{rank} K)$, making a distinction of the two cases more or less superficial. Therefore we concentrate on the matrix K having full column rank s.

10 Optimality criteria are essentially reduced to their associated *function induced ordering* among information matrices, that is, under the criterion ϕ an information matrix C is at least as good as another information matrix D when $\phi(C) \geq \phi(D)$. It is then indispensable that a reasonable criterion is *isotonic* relative to the Loewner ordering,

$$C \geq D \quad \Rightarrow \quad \phi(C) \geq \phi(D) \qquad \text{for all } C, D \geq 0.$$

A second property, similarly obliging, is *concavity*,

$$\phi\big(\alpha C + (1-\alpha)D\big) \geq \alpha\phi(C) + (1-\alpha)\phi(D) \qquad \text{for all } \alpha \in (0,1),\ C, D \geq 0.$$

For otherwise the situation $\phi\big(\alpha C + (1-\alpha)D\big) < \alpha\phi(C) + (1-\alpha)\phi(D)$ will occur: Rather than carrying out the experiment belonging to $\alpha C + (1-\alpha)D$ itself, one can achieve greater information through interpolation from two other experiments. This is absurd, information cannot be increased by interpolation.

A third property is *positive homogeneity*,

$$\phi(\delta C) = \delta\phi(C) \qquad \text{for all } \delta > 0,\ C \geq 0.$$

For Section *C5* has shown that the true information matrix is $(n/\sigma^2)C$, being directly proportional to the number of observations n and inversely proportional to the model variance σ^2. A criterion that is positively homogeneous maps this into $(n/\sigma^2)\phi(C)$, and the common factor n/σ^2 can rightly be omitted from further comparisons.

These properties evidently resemble those met for the information matrix mapping in Theorem *C13*. We need not go much further. Given positive homogeneity, concavity is equivalent with *superadditivity*,

$$\phi(C+D) \geq \phi(C) + \phi(D) \qquad \text{for all } C, D \geq 0.$$

Indeed, starting from concavity we have $\phi(C+D) = 2\phi(\frac{1}{2}D + \frac{1}{2}D) \geq 2\big(\frac{1}{2}\phi(C) + \frac{1}{2}\phi(D)\big) = \phi(C) + \phi(D)$. Conversely, starting from superadditivity we get $\phi\big(\alpha C + (1-\alpha)D\big) \geq \phi(\alpha C) + \phi\big((1-\alpha)D\big) = \alpha\phi(C) + (1-\alpha)\phi(D)$. Furthermore a close analysis of this reasoning shows that *strict concavity*,

$$\phi\big(\alpha C + (1-\alpha)D\big) > \alpha\phi(C) + (1-\alpha)\phi(D) \qquad \text{for all } \alpha \in (0,1),\ C, D \geq 0,\ C \neq D,$$

is equivalent with *strict superadditivity*,

$$\phi(C+D) > \phi(C) + \phi(D) \qquad \text{for all } C, D \geq 0,\ C \not\propto D.$$

The latter means that $\phi(C+D) = \phi(C) + \phi(D)$ holds just when C and D are proportional. The cases we interest us have strict concavity and strict superadditivity restricted to the open cone of positive definite matrices.

Next we show that isotonicity is the same as nonnegativity, in the presence of concavity and homogeneity.

11 LEMMA. *For every concave and positively homogeneous criterion ϕ on* NND(s) *the following statements are equivalent:*

- *a* (Monotonicity) *ϕ is isotonic: $\phi(C) \geq \phi(D)$ for all $C \geq D \geq 0$.*
- *b* (Nonnegativity) *ϕ is nonnegative: $\phi(C) \geq 0$ for all $C \geq 0$.*
- *c* (Positivity) *ϕ is nonnegative on the closed cone* NND(s) *and positive on its interior* PD(s)*: $\phi(C) > 0$ for all $C > 0$, or ϕ is constant.*

PROOF. First we establish the equivalence of *a* and *b*. If ϕ is isotonic then for $C \geq 0$ homogeneity yields

$$\phi(C) \geq \phi(0C) = 0\phi(C) = 0.$$

If ϕ is nonnegative then we apply superaddivity to $C \geq D \geq 0$ to obtain

$$\phi(C) = \phi(C - D + C) \geq \phi(C - D) + \phi(D) \geq \phi(D).$$

Now we turn to *c*. Assume *a*. In view of homogeneity ϕ can be constant only if it vanishes identically. Otherwise there exists a matrix $D \geq 0$ with $\phi(D) > 0$. Since D cannot be zero its largest eigenvalue, $\lambda_{\max}(D)$, is positive. For an arbitrary matrix $C \in \mathrm{PD}(s)$ its smallest eigenvalue, $\lambda_{\min}(C)$, is positive as well. Hence the eigenvalues of $C/\lambda_{\min}(C)$ and of $D/\lambda_{\max}(D)$ are separated by one. The inequality

$$\frac{1}{\lambda_{\min}(C)} C \geq I_s \geq \frac{1}{\lambda_{\max}(D)} D$$

gives $C \geq \big(\lambda_{\min}(C)/\lambda_{\max}(D)\big) D$. Monotonicity now leads to

$$\phi(C) \geq \phi\left(\frac{\lambda_{\min}(C)}{\lambda_{\max}(D)}\, D\right) = \frac{\lambda_{\min}(D)}{\lambda_{\max}(D)}\, \phi(D) > 0.$$

Thus *a* implies *c*. That *c* implies *b* is obvious. $\diamondsuit$

For the purpose of comparing numerical values of different criteria we usually *normalize* a nonconstant criterion according to

$$\phi(I_s) = 1.$$

Because of homogeneity this does not change the function induced ordering among information matrices.

12 Finally we require upper semicontinuity, meaning that the criterion behaves smoothly for nonnegative definite matrices that are singular. Criteria that enjoy all the properties discussed so far are called information functions.

- An *information function* on $\mathrm{NND}(s)$ is a criterion $\phi : \mathrm{NND}(s) \mapsto \mathbb{R}$ that is isotonic, concave, positively homogeneous, nonconstant, and upper semicontinuous.

The class of all information functions on $\mathrm{NND}(s)$ is denoted by Φ. Every convex function is continuous on the interior of its domain of definition, whence information functions on $\mathrm{NND}(s)$ are continuous on $\mathrm{PD}(s)$. The role of semicontinuity is scrutinized in Lemma *17*.

13 There exists a bewildering multitude of information functions. The easiest way to make this visible is to look at the *unit level set*

$$\mathcal{C} = \{\, C \geq 0 : \phi(C) \geq 1 \,\}.$$

This is a closed convex subset of the cone $\mathrm{NND}(s)$ which does not contain zero and recedes in all directions of $\mathrm{NND}(s)$. The latter property means that

$$C + \mathrm{NND}(s) \subseteq \mathcal{C} \qquad \text{for all } C \in \mathcal{C},$$

that is, when the cone $\mathrm{NND}(s)$ is translated so that its tip comes to lie in $C \in \mathcal{C}$ then all of the translate is included in $\mathcal{C}$.

Conversely, every closed convex subset $\mathcal{C}$ of $\mathrm{NND}(s)$ that does not contain zero and recedes in all directions of $\mathrm{NND}(s)$ determines an information function, namely

$$\phi(C) = \sup\{\, \delta > 0 : C \in \delta\mathcal{C} \,\} \cup \{\, 0 \,\} \qquad \text{for all } C \geq 0.$$

Moreover, this function has unit level set $\mathcal{C}$.

In other words, there is a one-to-one correspondence between information functions ϕ on $\mathrm{NND}(s)$ and closed convex subsets $\mathcal{C}$ of $\mathrm{NND}(s)$ that do not contain zero and recede in all directions of $\mathrm{NND}(s)$. Evidently there are plenty of such subsets, and so there are plenty of information functions.

We skip the formal proof of this correspondence. It simply parallels the well known correspondence between norms and their unit balls. The functional dinstinction is that norms are convex and information functions are concave. This finds its counterpart in the geometric orientation of the sets considered: Unit balls contain the point zero and exclude the point infinity, unit level sets exclude the point zero and 'contain' the point infinity. This difference in orientation also becomes manifest when we construct new information functions by means of polarity, in Section *15*.

14 New information functions can be constructed from elementary operations. We show that the class Φ of all information functions is closed under formation of nonnegative combinations and least upper bounds of finite subclasses, and under pointwise infima of arbitrary subclasses.

Firstly, given a finite family of information functions, $\phi_1, \ldots, \phi_\ell$, every nonnegative combination produces a new information function,

$$\sum_{i \leq \ell} \delta_i \phi_i, \qquad \min \delta_i \geq 0,$$

unless it degenerates to the constant zero when all coefficients δ_i vanish. In particular, sums and averages of finitely many information functions are information functions.

Secondly, the pointwise minimum of a finite family of information functions is an information function. Monotonicity, concavity, and homogeneity are easily verified. By Lemma *11c* the individual functions are positive for positive definite matrices, so that the minimum is positive, too. And upper semicontinuity follows since the minimum of upper semicontinuous functions is itself upper semicontinuous. Moreover, these arguments carry over to the pointwise infimum

$$\inf\{\, \phi_i : i \in I \,\}$$

of an arbitrary family $(\phi_i)_{i\in I}$ of information functions, unless it degenerates to the constant zero.

Thirdly, this allows to construct least upper bounds for a finite family, $\phi_1, \ldots, \phi_\ell$, namely

$$\mathrm{lub}_{i\le\ell}\, \phi_i = \inf\{\, \phi : \phi \in \Phi,\ \phi \ge \max_{i\le\ell} \phi_i \,\}.$$

The set over which the infimum is sought is nonempty, containing at least the sum $\sum \phi_i$. And being bounded for instance by ϕ_1 the infimum cannot degenerate to the constant zero.

These structural properties of the class Φ suggest that our definition of information functions is not only statistically reasonable, but also theoretically appropriate. The most important functional operation, however, is still to come: polarity.

15 Polarity is a special case of a duality relationship, and as such based on the underlying inner product for symmetric matrices,

$$\langle C, D\rangle = \mathrm{trace}\, CD \qquad \text{for all } C, D \in \mathrm{Sym}(s).$$

This is the restriction of the Euclidean matrix inner product discussed in Section *A7*; the angular brackets provide some notational relief in the development to follow.

The polar function ϕ^∞ of a given information function ϕ is best thought of as the largest function to satisfy the (generalized) *Hölder inequality**

$$\phi(C)\, \phi^\infty(D) \le \langle C, D\rangle \qquad \text{for all } C, D \ge 0.$$

The precise definition is as follows.

- The *polar function* ϕ^∞ of an information function ϕ on $\mathrm{NND}(s)$ is defined by

$$\phi^\infty(D) = \inf_{C>0} \frac{\langle C, D\rangle}{\phi(C)} \qquad \text{for all } D \in \mathrm{NND}(s).$$

The validity of the Hölder inequality for $C > 0$ is evident from the definition, for $C \ge 0$ it follows through regularization.

The next lemma states the most important relationships: Polars of information function are themselves information functions. And polarity is an idempotent operation, that is, second polars recover the original information function.

* Otto Hölder 1859–1937

16 LEMMA. *For every information function* ϕ *on* NND(s) *its polar function* ϕ^∞ *is again an information function on* NND(s) *whose polar function in turn is* ϕ,

$$\phi^{\infty\infty} = \phi.$$

PROOF. Monotonicity holds since Lemma *A18* yields $\langle C, D-E\rangle \geq 0$ for $D \geq E \geq 0$, and this entails

$$\phi^\infty(D) = \inf_{C>0} \frac{\langle C, D\rangle}{\phi(C)} \geq \inf_{C>0} \frac{\langle C, E\rangle}{\phi(C)} = \phi^\infty(E).$$

Concavity follows from the polar function being an infimum over linear functions:

$$\begin{aligned}\phi^\infty\big(\alpha D + (1-\alpha)E\big) &= \inf_{C>0} \frac{\langle C, \alpha D + (1-\alpha)E\rangle}{\phi(C)} \\ &\geq \alpha \inf_{C>0} \frac{\langle C, D\rangle}{\phi(C)} + (1-\alpha)\inf_{C>0} \frac{\langle C, E\rangle}{\phi(C)} \\ &= \alpha\phi^\infty(D) + (1-\alpha)\phi^\infty(E).\end{aligned}$$

Homogeneity rests on the bilinearity of the inner product, according to

$$\phi^\infty(\delta D) = \inf_{C>0} \frac{\langle C, \delta D\rangle}{\phi(C)} = \delta \inf_{C>0} \frac{\langle C, D\rangle}{\phi(C)} = \delta\phi^\infty(D).$$

For nonconstancy we verify that $\phi^\infty(I_s) > 0$. Multiplying numerator and denominator by $\|C\|$ gives

$$\phi^\infty(I_s) = \inf_{C>0} \frac{\langle C, I_s\rangle/\|C\|}{\phi(C)/\|C\|} \geq \frac{\min_{C\geq 0: \|C\|=1} \operatorname{trace} C}{\sup_{C>0} \phi(C)/\|C\|} > 0.$$

The minimum in the numerator is attained and positive since the set $\{C \geq 0 : \|C\| = 1\}$ is compact and the function $\operatorname{trace} C$ is continuous. The supremum in the denominator is positive since ϕ is nonzero.

That the second polar reproduces the original function is a standard result from convex analysis. ◇

Polar information functions form a principal tool in the sequel, a first instance is encountered in the following lemma. It casts the semicontinuity of information functions into a form similar to the matrix semicontinuity met in Theorem *C13*.

17 LEMMA. *For every isotonic, concave and positively homogeneous criterion* ϕ *on* NND(s) *the following statements are equivalent:*

a (Real semicontinuity) ϕ *is upper semicontinuous.*

b (Regularization) $\lim_{n\to\infty} \phi(C + \frac{1}{n}D) = \phi(C)$ *for all* $C \geq 0$ *and* $D > 0$.

c (Matrix semicontinuity) $\lim_{n\to\infty} \phi(C_n) = \phi(C)$ *for all sequences* $(C_n)_{n\geq 1}$ *in* NND(s) *with* $\lim_{n\to\infty} C_n = C$ *and* $\phi(C_n) \geq \phi(C)$ *for all* $n \geq 1$.

PROOF. The equivalence of *a* and *b* is a general result from convex analysis, valid for every criterion function ϕ that is concave.

Next we show that *b* implies *c*, using the double polarity property

$$\phi(C)f = \phi^{\infty\infty}(C) = \inf_{D>0} \frac{\langle C, D\rangle}{\phi^{\infty}(D)}.$$

For every $\epsilon > 0$ there exists a matrix $D_\epsilon > 0$ such that

$$\frac{\langle C, D_\epsilon\rangle}{\phi^{\infty}(D_\epsilon)} \leq \epsilon + \inf_{D>0} \frac{\langle C, D\rangle}{\phi^{\infty}(D)} = \epsilon + \phi(C).$$

As in the proof of Theorem *C13* we conclude that

$$\phi(C) \leq \phi(C_n) = \inf_{D>0} \frac{\langle C_n, D\rangle}{\phi^{\infty}(D)} \leq \frac{\langle C_n, D_\epsilon\rangle}{\phi^{\infty}(D_\epsilon)} \longrightarrow \frac{\langle C, D_\epsilon\rangle}{\phi^{\infty}(D_\epsilon)} \leq \epsilon + \phi(C),$$

where the limit is taken as n tends to ∞. Since ϵ is arbitrary the sequence $\phi(C_n)$ converges to $\phi(C)$.

Conversely, if *c* is given then *b* follows upon setting $C_n = C + \frac{1}{n}D$. ◇

The theory centers around moment matrices. Hence our genuine interest is in the composition of an information function ϕ with the information matrix mapping C,

$$\phi \circ C : \text{NND}(k) \mapsto \mathbb{R}.$$

It is now easy to put the two pieces together.

18 LEMMA. *For every information function* ϕ *on* NND(s) *the composition* $\phi \circ C$ *with the information matrix mapping* C *is an information function on* NND(k) *whose polar is given by*

$$(\phi \circ C)^{\infty}(B) = \phi^{\infty}(K'BK) \qquad \textit{for all } B \geq 0.$$

PROOF. The composition $\phi \circ C$ is isotonic, concave, positively homogeneous, and nonconstant. While among real functions composition preserves semicontinuity, we here have for C only matrix upper semicontinuity as detailed in Theorem *C13*. Yet semicontinuity follows by verifying Lemma *17b* for $\phi \circ C$. To this end we must show that

$$\lim_{n\to\infty} \phi \circ C(A + \frac{1}{n}B) = \phi \circ C(A) \qquad \text{for all } A \in \text{NND}(k),$$

regardless of the choice $B \in \mathrm{PD}(k)$. According to Theorem *C13* the matrices $C_n = C(A+\frac{1}{n}B)$ converge to $C(A)$ while obeying $C_n \geq C(A)$. Monotonicity of ϕ yields $\phi(C_n) \geq \phi\big(C(A)\big)$. An application of Lemma *17c* to ϕ shows that $\phi(C_n)$ converges to $\phi\big(C(A)\big)$. Hence $\phi \circ C$ is an information function on $\mathrm{NND}(k)$.

The polarity formula is established in two steps both of which make use of the inequality

$$\langle C(A), K'BK \rangle = \operatorname{trace} KC(A)K'B \leq \operatorname{trace} AB = \langle A, B \rangle. \tag{$\dagger$}$$

The inequality follows from Lemma *A8*, and the relations

$$A_K = KC(A)K' = \min\nolimits_{QK=K} QAQ' \leq A$$

which are discussed at length in Section *C21*. In a first step the candidate $\phi^\infty(K'BK)$ is seen to satisfy the Hölder inequality, since the polarity relation between ϕ and ϕ^∞, and the inequality ($\dagger$) yield

$$\phi \circ C(A)\, \phi^\infty(K'BK) \leq \langle C(A), K'BK \rangle \leq \langle A, B \rangle.$$

If A is positive definite so is $C(A)$, and $\phi\big(C(A)\big)$ is positive. This leads to the inequality

$$\phi^\infty(K'BK) \leq \inf\nolimits_{A>0} \frac{\langle A, B \rangle}{\phi \circ C(A)} = (\phi \circ C)^\infty(B).$$

In the second step we utilize inequality ($\dagger$) to obtain the lower bound

$$(\phi \circ C)^\infty(B) \geq \inf\nolimits_{A>0} \frac{\langle KC(A)K', B \rangle}{\phi \circ C(A)} \geq \inf\nolimits_{D>0} \frac{\langle D, K'BK \rangle}{\phi(D)} = \phi^\infty(K'BK).$$

Here the second inequality holds since if A is positive definite then so is $C(A)$, see Theorem *C15*. The two steps establish the polarity formula. $\diamondsuit$

A first use of the polar function of the composition $\phi \circ C$ is made in Theorem *20* on the existence of optimal moment matrices that are also feasible. First we define the design problem in its full generality.

19 Recall that we are interested in a parameter subsystem $K'\theta$ with a coefficient matrix K of full column rank s, and that the set $\mathcal{M}$ of competing moment matrices is assumed to intersect the feasibility cone $\mathcal{A}(K)$. Given an information function ϕ the general *design problem* then reads:

- *Maximize* $\quad \phi\big(C(M)\big) \quad$ *subject to* $\quad M \in \mathcal{M}$.

This calls for maximizing information as measured by the information function ϕ, in the set $\mathcal{M}$ of competing moment matrices. The optimal value of this problem is, by definition,

$$v(\phi) = \sup_{M \in \mathcal{M}} \phi\big(C(M)\big).$$

A moment matrix $M \in \mathcal{M}$ is said to be an *optimal solution of the design problem* when $\phi\big(C(M)\big)$ attains the optimal value $v(\phi)$; if, in addition, the matrix M lies in the feasibility cone $\mathcal{A}(K)$ then M is called *ϕ-optimal for $K'\theta$ in $\mathcal{M}$.* Again designs inherit the optimality properties by their moment matrices. Given a subclass Ξ' a design $\xi \in \Xi'$ is called *ϕ-optimal for $K'\theta$ in Ξ'* whenever its moment matrix $M(\xi)$ is ϕ-optimal for $K'\theta$ in $M(\Xi')$.

In this form the design problem makes no reference to the feasibility cone $\mathcal{A}(K)$. Indeed, pathological instances are quite possible wherein optimal moment matrices are *not* feasible! Such moment matrices, though solving a well formulated optimization problem, are statistically useless. The appropriate tool has been shaped in Section *C15*: In order to make sure that the moment matrix M is feasible for $K'\theta$ one computes the information matrix $C(M)$ and checks its rank.

The following theorem singles out three sufficient conditions under which every optimal solution matrix has maximal rank and hence is feasible. When an information function is zero for singular nonnegative definite matrices we briefly say that it *vanishes for singular matrices.* Furthermore a criterion ϕ is called *strictly isotonic* on the cone of positive definite matrices when $C \geq D > 0$ and $C \neq D$ imply $\phi(C) > \phi(D)$.

20 EXISTENCE THEOREM. *Let the set $\mathcal{M}$ of competing moment matrices be compact. Then there exists a moment matrix $M \in \mathcal{M}$ that is an optimal solution of the design problem. Moreover, in order that every such matrix lies in the feasibility cone $\mathcal{A}(K)$—and thus is ϕ-optimal for $K'\theta$ in $\mathcal{M}$—any one of the following conditions is sufficient:*

a (Condition on $\mathcal{M}$) *The set $\mathcal{M}$ is included in the feasibility cone $\mathcal{A}(K)$.*

b (Condition on ϕ) *The information function ϕ vanishes for singular matrices.*

c (Condition on ϕ^∞) *The polar information function ϕ^∞ vanishes for singular matrices, is strictly isotonic for positive definite matrices, and there exists a matrix $D \in \mathrm{NND}(s)$ such that*

$$\phi\big(C(M)\big)\,\phi^\infty(D) = \operatorname{trace} C(M)D = 1.$$

PROOF. By Lemma *18* the function $\phi \circ C$ is upper semicontinuous, and thus attains its supremum over the compact set $\mathcal{M}$. Hence an optimal solution to the design problem exists. Because of our grand assumption in Section *1* the intersection $\mathcal{M} \cap \mathcal{A}(K)$ contains at least one matrix B, say. Its information matrix $C(B)$ is positive definite and has a positive information value $\phi(C(B))$, by Theorem *C15* and Lemma *11c*. Therefore the optimal value is positive, $v(\phi) > 0$.

Under condition a all competing moment matrices are contained in $\mathcal{A}(K)$, including the optimal ones. Under condition b the criterion ϕ vanishes for singular matrices, whence any optimal solution matrix M has a nonsingular information matrix $C(M)$. Then M must lie in $\mathcal{A}(K)$, by Theorem *C15*.

Finally we turn to condition c. Let z be a nullvector of $C(M)$. Since ϕ^∞ vanishes for singular matrices, the equality $\phi(C(M))\, \phi^\infty(D) = 1$ forces D to be positive definite. For $\delta > 0$ we have

$$\begin{aligned}\phi(C(M))\, \phi^\infty(D + \delta zz') &\le \langle C(M), D + \delta zz'\rangle = \langle C(M), D\rangle \\ &= \phi(C(M))\, \phi^\infty(D) \le \phi(C(M))\, \phi^\infty(D + \delta zz'),\end{aligned}$$

as follows from the Hölder inequality, the property $C(M)z = 0$, the formula in condition c, and monotonicity of ϕ^∞. Therefore we obtain for all $\delta > 0$ the constant value

$$\phi^\infty(D + \delta zz') = \frac{\langle C(M), D\rangle}{\phi(C(M))}.$$

Because of strict monotonicity of ϕ^∞ on $\mathrm{PD}(s)$ the vector z must vanish. Again this yields positive definiteness of $C(M)$, and feasibility of M. $\diamondsuit$

Before introducing the important class of p-mean criteria we specilize the present results to subsystems of rank one.

21 The preceding sections on Loewner optimality and information functions come to bear only when the coefficient matrix K has rank larger than one. We briefly digress to see how these concepts simplify when the parameter system of interest is onedimensional.

For a system $c'\theta$ the information 'matrix' mapping C is scalar, positive on the feasibility cone $\mathcal{A}(c)$,

$$C(A) = (c'A^-c)^{-1} > 0,$$

and zero outside. Hence the Loewner ordering among information matrices coincides with the ordering among variances that is induced by the optimality criterion $c'A^-c$ of Section *B7*.

Next we consider an information function ϕ. In accordance with homogeneity we can extract the information number $C(A)$,

$$\phi(C(A)) = C(A)\phi(1).$$

The positive constant $\phi(1) > 0$ does not change the function induced ordering. Hence the information numbers $C(A)$ are essentially the *only* criteria of importance. (However, in general they fail to be normalized.)

Onedimensional subsystems also illustrate condition *b* of the Existence Theorem *20*, as $C(A)$ vanishes for singular matrices A.

Finally consider polar information functions. We have

$$\phi^\infty(D) = \inf_{C>0} \frac{CD}{\phi(C)} = \frac{1}{\phi(1)} D \qquad \text{for all } D \in [0, \infty),$$

since here C and D simply are scalars. In accordance with Lemma *18* the composition with the information 'matrix' mapping then has a polar function given by

$$(\phi \circ C)^\infty(B) = \phi^\infty(c'Bc) = \frac{1}{\phi(1)} c'Bc \qquad \text{for all } B \in \mathrm{NND}(k).$$

Apart from the constant $1/\phi(1)$ the polar function recovers the criterion funtion of the dual problem of Section *B12*. This, in fact, soon emerges as the general rule.

The following list of particular information functions is therefore of interest only when the dimensionality of the parameter subsystem, s, is larger than one.

22 Before turning to the full family of the p-mean criteria we introduce its four most prominent members,
—the *determinant criterion* $\phi_0(C) = (\det C)^{1/s}$,
—the *average-variance criterion* $\phi_{-1}(C) = (\frac{1}{s} \operatorname{trace} C^{-1})^{-1}$ (assuming C to be positive definite for the time being),
—the criterion of the smallest eigenvalue $\phi_{-\infty}(C) = \lambda_{\min}(C)$, called the *eigenvalue criterion*, and
—the *trace criterion* $\phi_1(C) = \frac{1}{s} \operatorname{trace} C$.
The subscripting looks strange at this point, but emerges as the natural choice within the p-mean family. All these criteria are normalized, as they assign value one to the identity matrix I_s.

The determinant criterion, and the determinant itself only differ by taking the s^{th} root. This being a monotone operation both functions produce the same function induced ordering among information matrices. From a practical point of view one may therefore dispense with taking the s^{th} root and consider the determinant directly. However, the determinant is positively homogeneous of degree s rather than one. For comparing different criteria, and for applying the theory to be developed, only the version ϕ_0 is appropriate.

Maximizing the determinant of information matrices is the same as minimizing the determinant of variance–covariance matrices, due to the determinant formula

$$(\det C)^{-1} = \det(C^{-1}).$$

Indeed, in Section *C15* the inverse C^{-1} of an information matrix was identified to be the (normalized) variance–covariance matrix of the optimal estimator for the parameter system of interest. Its determinant is called the *generalized variance*, and is a familiar measure from multivariate analysis for the size of a variance–covariance matrix. This is the origin of the great popularity that the determinant criterion enjoys in applications.

A closely related aspect is based on the determinant formula

$$\det(H'CH) = (\det H^2)(\det C),$$

with some nonsingular $s \times s$ matrix H. Suppose that the parameter system $K'\theta$ is reparametrized according to $H'K'\theta$. This is a special case of 'iterating on information' for which Theorem *C19* provides the identity

$$C_{KH}(A) = C_H\big(C_K(A)\big) = \Big(H'\big(C_K(A)\big)^{-1}H\Big)^{-1} = H^{-1}C_K(A)H'^{-1}.$$

Thus the determinant assigns proportional values to $C_K(A)$ and $C_{KH}(A)$, with proportionality factor $1/\det H^2$, and the two function induced ordering of information matrices are identical.

In other words, the determinant induced ordering is invariant under reparametrizations. It can be shown that the determinant is the unique criterion whose function induced ordering has this invariance property.

Another invariance property pertains to the determinant function itself, rather than to its induced ordering: It is invariant under reparametrizations with matrices H that fulfill $\det H = \pm 1$. This simply follows from $\det C_{KH}(A) = (1/\det H^2)\det C_K(A) = \det C_K(A)$. We verify in the next chapter that again the determinant is uniquely characterized by this invariance property.

Invariance under reparametrization loses its appeal when the parameters of interest have a definite physical meaning, as is often the case. Then the average-variance criterion provides an alternative. When the coefficient matrix is partitioned into its columns, $K = (c_1, \ldots, c_s)$, the inverse $1/\phi_{-1}$ can be represented as

$$\frac{1}{s}\operatorname{trace} C(A)^{-1} = \frac{1}{s}\operatorname{trace} K'A^-K = \frac{1}{s}\sum\nolimits_{j\le s} c_j'A^-c_j.$$

This is the average of the normalized variances for $c_j'\theta$, for $j = 1, \ldots, s$. Again we can pass back and forth between the information point of view and the and dispersion point of view: Maximizing the average-variance criterion among information matrices is the same as minimizing the average of the variances given above.

The third criterion, the smallest eigenvalue, also gains in understanding by a passage to variances. It is the same as minimizing the largest eigenvalue of the inverse matrix,

$$\lambda_{\max}\big(C(A)^{-1}\big) = \max\nolimits_{z:\|z\|=1} z'K'A^{-}Kz.$$

Minimizing this expression guards against the worst possible variance among all onedimensional subsystems $z'K'\theta$ with z being normalized to have length one. In terms of variances this is a minimax approach, in terms of information a maximin approach. This will play a special role in the admissibility discussion.

The eigenvalue criterion $\phi_{-\infty}$ is the leftmost member of the p-mean family, corresponding to the order $p = -\infty$. The rightmost member is trace optimality ϕ_1. Nothing is gained for its interpretation if we try the information–dispersion transition.

By itself the trace criterion is rather weak. For example, in twoway classification models it assigns a constant value to *all* moment matrices,

$$\phi_1\big((M(W)\big) = \frac{1}{a+b-1}\operatorname{trace}\begin{pmatrix} \Delta_r & W \\ W' & \Delta_s \end{pmatrix} = \frac{2}{a+b-1} \qquad \text{for all } W \in \Xi.$$

Also we have stressed in Section *10* that a criterion ought to be concave so that information cannot be increased through interpolation. The trace criterion is actually linear, and this is too weak to *not* prevent interpolation from being useful. Yet trace optimality has its place in the theory, often being accompanied by further conditions which prevent it from going astray.

Trace optimality furnishes a simple illustration that optimal moment matrices are not automatically feasible. As an example take the quadratic regression model of Section *A6*, with all three parameters being of interest. Assume that the experimental domain is the symmetric unit interval $\mathcal{U} = [-1, 1]$. As seen in Section *A27* a distribution v on $\mathcal{U}$ induces a design ξ such that

$$\phi_1\big(M(\xi)\big) = \frac{1}{3}\operatorname{trace}\begin{pmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{pmatrix} = \frac{1+\mu_2+\mu_4}{3} \leq 1.$$

The moments $\mu_j = \int_{[-1,1]} u^j\, dv$ for $j = 2, 4$ attain the maximal value one if and only if the distribution v is concentrated on the set ± 1. Thus every optimal design ξ must be supported by the two vectors $(1, \pm 1, 1)'$, whence its moment matrix has at most rank two. No such moment matrix can be feasible for a threedimensional parameter system.

Trace optimality is an exception; we see presently that the other members of the p-mean family fulfill at least on of the conditions b or c of the Existence Theorem *20*, so that for them every optimal moment matrix is feasible.

23 The four criteria met in the preceding section are the four outstanding members of the oneparameter family of *matrix means* ϕ_p with order $p \in [-\infty, 1]$. It is instructive to introduce these means for all orders $p \in [-\infty, \infty]$, and contrast their concavity–convexity behaviour for $p \le 1$ and $p \ge 1$. For positive definite $s \times s$ matrices the definition is

$$\phi_p(C) = \begin{cases} \lambda_{\max}(C) & \text{for } p = \infty; \\ (\frac{1}{s} \operatorname{trace} C^p)^{1/p} & \text{for } 0, p \neq \pm\infty; \\ (\det C)^{1/s} & \text{for } p = 0; \\ \lambda_{\min}(C) & \text{for } p = -\infty. \end{cases}$$

The extension to singular matrix $C \ge 0$ will be carried out by regularization.

Before doing so we recall that real powers of positive definite matrices are defined through spectral decompositions $C = \sum_{j \le s} \lambda_j z_j z_j'$, say, according to

$$C^p = \sum_{j \le s} \lambda_j^p \, z_j z_j' \qquad \text{for all } p \in \mathbb{R}.$$

Application of the trace operator then yields $\operatorname{trace} C^p = \sum \lambda_j^p \operatorname{trace} z_j z_j' = \sum \lambda_j^p$. Therefore $\phi_p(C)$ is the generalized mean of order p of the eigenvalues λ_j of the matrix C.

It is known from calculus that for a fixed set of positive numbers λ_j the *real means*

$$m_p(\lambda_1, \ldots, \lambda_s) = \begin{cases} \max_{j \le s} \lambda_j & \text{for } p = +\infty; \\ (\frac{1}{s} \sum_{j \le s} \lambda_j^p)^{1/p} & \text{for } p \neq 0, \pm\infty; \\ (\prod_{j \le s} \lambda_j)^{1/s} & \text{for } p = 0; \\ \min_{j \le s} \lambda_j & \text{for } p = -\infty. \end{cases}$$

are continuous and increasing as a function of $p \in [-\infty, +\infty]$. Hence $\phi_p(C)$, too, is a continuous and increasing function of p when the argument matrix $C > 0$ is kept fixed.

We now utilize this monotonicity property for varying order p to compute the semicontinuous extension of the function ϕ_p to singular matrices. Let the matrix C be nonnegative definite and singular. With some positive definite matrix D and $\delta > 0$ we have

$$0 \le \phi_p(C + \delta D) \le \phi_0(C + \delta D) \qquad \text{for all } p < 0.$$

As δ tends to zero so does $\phi_0(C + \delta D)$ since the determinant is a continuous function on the whole space $\mathrm{Sym}(s)$. For orders $p \in [-\infty, 0)$ the upper semicontinuous extension therefore is

$$\phi_p(C) = 0 \qquad \text{for all } C \in \mathrm{NND}(s),\ \operatorname{rank} C < s.$$

For orders $p \in (0, \infty]$ no inversion is involved, and the matrix means ϕ_p are continuous on the full space $\mathrm{Sym}(s)$; they vanish in C if and only if all eigenvalues of C are zero, that is, $C = 0$.

This concludes the definition of the matrix means ϕ_p as functions on the closed cone $\mathrm{NND}(s)$. In its course we have established that they vanish for singular matrices if and only if their order lies in the interval $[-\infty, 0]$. Another consequence of the isotonic behaviour in p is that the optimal value $v(p) = v(\phi_p)$ is an increasing function of p.

We now turn to the functional properties of the matrix means ϕ_p, amplifying their close relation with the real means m_p. Two numbers $p, q \in [-\infty, \infty]$ are called *conjugate* when $p + q = pq$. For a convex function ψ that is positive on $\mathrm{PD}(s)$ its *polar function* is defined by

$$\psi^0(D) = \sup\nolimits_{C>0} \frac{\langle C, D\rangle}{\psi(C)} \qquad \text{for all } D \in \mathrm{NND}(s).$$

24 THEOREM. *For every order $p \in [-\infty, 1]$ the mean ϕ_p is an information function on the cone $\mathrm{NND}(s)$; on $\mathrm{PD}(s)$ it is strictly isotonic when $p \neq -\infty$ and strictly concave when $p \neq -\infty, 1$; its polar function is $s\phi_q$ where $q \in [-\infty, 1]$ is conjugate to p.*

For every order $\tilde{p} \in [1, \infty]$ the mean $\phi_{\tilde{p}}$ has all properties of an information function on the cone $\mathrm{NND}(s)$ except that convexity replaces concavity; on $\mathrm{PD}(s)$ it is strictly isotonic when $\tilde{p} \neq \infty$ and strictly convex when $\tilde{p} \neq 1, \infty$; its polar function is $s\phi_{\tilde{q}}$ where $\tilde{q} \in [1, \infty]$ is conjugate to $\tilde{p}$.

PROOF. Upper semicontinuity of ϕ_p on the cone $\mathrm{NND}(s)$ is built into the definition when $p < 0$; when $p \geq 0$ the function ϕ_p is continuous on all of $\mathrm{Sym}(s)$. Nonconstancy and homogeneity are evident. Also monotonicity is fairly straightforward. For when $C \geq D \geq 0$ then the decreasingly ordered eigenvalues $\lambda_{\downarrow 1} \geq \cdots \geq \lambda_{\downarrow s}$ and $\mu_{\downarrow 1} \geq \cdots \geq \mu_{\downarrow s}$ of C and D satisfy

$$\lambda_{\downarrow j} \geq \mu_{\downarrow j} \qquad \text{for all } j \leq s.$$

Since the real means m_p are invariant under permutations of their arguments and increasing in each of them we find

$$\phi_p(C) = m_p(\lambda_{\downarrow 1}, \ldots, \lambda_{\downarrow s}) \geq m_p(\mu_{\downarrow 1}, \ldots, \mu_{\downarrow s}) = \phi_p(D).$$

Furthermore when $p \neq \pm\infty$ then the real mean m_p is strictly isotonic for positive arguments, and so the matrix mean ϕ_p is strictly isotonic for positive definite matrices.

It remains to establish concavity for $p \leq 1$ and convexity for $r \geq 1$. Instead we prove, in Section *26*, the polarity relations

$$\phi_p^\infty(D) = s\phi_q(D), \quad \phi_{\tilde{p}}^0(D) = s\phi_{\tilde{q}}(D) \qquad \text{for all } D \geq 0.$$

For then ϕ_q—being the infimum over linear functions— is concave, and $\phi_{\tilde{q}}$ is convex. But when p runs through $[-\infty, 1]$ so does its conjugate number q, and when $\tilde{p}$ varies over $[1, \infty]$ so does its conjugate number $\tilde{q}$. Thus the concavity–convexity properties hold true as claimed. Strict concavity and convexity will be derived in Section *26*. ◇

For the proof of polarity we need an auxiliary result which says with what advantage or disadvantage the eigenvalues λ_j may be arranged relative to the eigenvectors z_j. We call two s-dimensional vectors λ and μ *discordant (concordant)* if for $j \leq s$ the j^{th} largest entry of λ matches the j^{th} smallest (largest) entry of μ. For instance, $(2,9,4)'$ and $(11,1,5)'$ are discordant, while $(2,9,4)'$ and $(1,11,5)'$ are concordant. The *decreasing (increasing) rearrangement* of λ is denoted by $\lambda_\downarrow$ $(\lambda_\uparrow)$, the entries $\lambda_{\downarrow j}$ $(\lambda_{\uparrow j})$ being the j^{th} largest (smallest) component of λ.

25 REARRANGEMENT INEQUALITITES. *Any two sequences $\lambda = (\lambda_j)_{j\leq s}$ and $\mu = (\mu_j)_{j\leq s}$ of positive numbers satisfy the inequalities*

$$\sum\nolimits_{j\leq s} \lambda_{\downarrow j}\mu_{\uparrow j} \leq \sum\nolimits_{j\leq s} \lambda_j \mu_j \leq \sum\nolimits_{j\leq s} \lambda_{\downarrow j}\mu_{\downarrow j};$$

equality holds in the left (right) inequality if and only if the vectors λ and μ are discordant (concordant).

Moreover if $(\lambda_j)_{j\leq s}$ and $(\mu_j)_{j\leq s}$ are the eigenvalues of two positive definite $s \times s$ matrices C and D, respectively, then

$$\sum\nolimits_{j\leq s} \lambda_{\downarrow j}\mu_{\uparrow j} \leq \operatorname{trace} CD \leq \sum\nolimits_{j\leq s} \lambda_{\downarrow j}\mu_{\downarrow j};$$

equality holds in the left (right) inequality if and only if there exists a common set of normalized eigenvectors $z_1, \ldots, z_s$ in $\mathbb{R}^s$ such that

$$C = \sum\nolimits_{j\leq s} \lambda_{\downarrow j} z_j z_j', \qquad D = \sum\nolimits_{j\leq s} \mu_{\uparrow j} z_j z_j'.$$

$$\Big(C = \sum\nolimits_{j\leq s} \lambda_{\downarrow j} z_j z_j', \qquad D = \sum\nolimits_{j\leq s} \mu_{\downarrow j} z_j z_j'.\Big)$$

PROOF. With the inner product notation $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^s$ the assertion reads

$$\langle\lambda_\downarrow, \mu_\uparrow\rangle \leq \langle\lambda, \mu\rangle \leq \langle\lambda_\downarrow, \mu_\downarrow\rangle.$$

The proof is by induction on s. If $s = 2$ and $\lambda \leq \Lambda$ and $\mu \leq \mathrm{M}$ then

$$0 \leq (\Lambda - \lambda)(\mathrm{M} - \mu) = \Lambda\mathrm{M} + \lambda\mu - (\Lambda\mu + \lambda\mathrm{M}),$$

which is all there is to prove. We now step up from s to $s+1$. For the left inequality let j_0 and j_1 be such that

$$\max_{j \leq s+1} \lambda_j = \lambda_{j_0} = \Lambda, \quad \min_{j \leq s+1} \mu_j = \mu_{j_1} = \mu, \quad \text{say.}$$

We distinguish the cases whether j_0 and j_1 coincide or not. In case $j_0 = j_1$ the induction hypothesis yields

$$\langle (\lambda_{\downarrow j})_{j \neq j_0}, (\mu_{\uparrow j})_{j \neq j_0} \rangle \leq \langle (\lambda_j)_{j \neq j_0}, (\mu_j)_{j \neq j_0} \rangle,$$

and adding $\Lambda\mu$ establishes the left inequality.

In case $j_0 \neq j_1$ we use the inequalities for $s = 2$, proved in the beginning, and the induction hypothesis to obtain

$$\begin{aligned}
\langle (\lambda_j)_{j \leq s+1}, (\mu_j)_{j \leq s+1} \rangle &= \Lambda\mu_{j_0} + \lambda_{j_1}\mu + \langle (\lambda_j)_{j \neq j_0, j_1}, (\mu_j)_{j \neq j_0, j_1} \rangle \\
&\geq \Lambda\mu + \lambda_{j_1}\mu_{j_0} + \langle (\lambda_j)_{j \neq j_0, j_1}, (\mu_j)_{j \neq j_0, j_1} \rangle \\
&= \Lambda\mu + \big\langle \big(\lambda_{j_1}, (\lambda_j)_{j \neq j_0, j_1}\big), \big(\mu_{j_0}, (\mu_j)_{j \neq j_0, j_1}\big) \big\rangle \\
&\geq \Lambda\mu + \langle (\lambda_{\downarrow j})_{j \neq j_0}, (\mu_{\uparrow j})_{j \neq j_1} \rangle \\
&= \langle (\lambda_{\downarrow j})_{j \leq s+1}, (\mu_{\uparrow j})_{j \leq s+1} \rangle.
\end{aligned}$$

This establishes the left inequality, and the equality condition. The proof of the right inequality is similar and omitted.

For the matrix statement we choose spectral decompositions $C = \sum \lambda_i z_i z_i'$ and $D = \sum \mu_j y_j y_j'$. We then have

$$\operatorname{trace} CD = \sum\nolimits_i \sum\nolimits_j \lambda_i \mu_j (z_i' y_j)^2 = \lambda' S \mu, \text{ say.}$$

The matrix S has nonnegative entries, $s_{ij} = \langle z_i, y_j \rangle^2 \geq 0$, and its rows and columns add to one since $\sum_j s_{ij} = \sum_j z_i' y_j y_j' z_i = z_i' z_i = 1$. Such matrices are called doubly stochastic, and are known to admit a representation as an average over permutation matrices. Let Q_π be the permutation matrix corresponding to a permutation π in the *symmetric group* $\mathcal{S}$ of permutations of s elements, that is, with Euclidean unit vectors $e_j \in \mathbb{R}^s$ the matrix Q_π admits the representation

$$Q_\pi = \sum\nolimits_{j \leq s} e_j e_{\pi(j)}'.$$

With nonnegative coefficients α_π which add to one we can thus write

$$S = \sum\nolimits_{\pi \in \mathcal{S}} \alpha_\pi Q_\pi.$$

Applying the first part of the lemma we finally obtain

$$\operatorname{trace} CD = \sum\nolimits_{\pi \in \mathcal{S}} \alpha_\pi \lambda' Q_\pi \mu = \sum\nolimits_{\pi \in \mathcal{S}} \alpha_\pi \sum\nolimits_{j \le s} \lambda_j \mu_{\pi(j)} \begin{cases} \le \sum \lambda_{\downarrow j} \mu_{\downarrow j} \\ \ge \sum \lambda_{\downarrow j} \mu_{\uparrow j} \end{cases}.$$

To study the equality condition notice first that the entries $\mu_{\pi(j)}$ together form the vector $Q_\pi \mu = \sum \mu_{\pi(j)} e_j$. Thus equality holds in the left inequality if and only if the vectors λ and $Q_\pi \mu$ are discordant whenever $\alpha_\pi > 0$. For the sake of simplicity assume that the entries of λ and μ are pairwise distinct. Then there is a unique permutation π for which λ and $Q_\pi \mu$ are discordant, forcing $\alpha_\pi = 1$ and $S = Q_\pi$. Hence the $\big(j, \pi(j)\big)^{\text{th}}$ entries of S are one and the others are zero. This entails equality in the Cauchy inequality,

$$1 = s_{j,\pi(j)} = \langle z_j, y_{\pi(j)} \rangle^2 \le \langle z_j, z_j \rangle \langle y_{\pi(j)}, y_{\pi(j)} \rangle = 1.$$

hence the vectors z_j and $y_{\pi(j)}$ are proportional. They are also normalized, so that that the constant of proportionality is one and $z_j z_j' = y_{\pi(j)} y_{\pi(j)}'$. Thus besides $C = \sum \lambda_j z_j z_j'$ we obtain

$$D = \sum \mu_j y_j y_j' = \sum \mu_{\pi(j)} y_{\pi(j)} y_{\pi(j)}' = \sum \mu_{\pi(j)} z_j z_j'.$$

This means that the j^{th} largest eigenvalue of C has the same eigenvevtor z_j as the j^{th} smallest eigenvalue of D, as claimed in the assertion. $\diamondsuit$

The Hölder inequelities and the polarity formulas are proved in the next lemma. Again the close relation with the Hölder inequality in $\mathbb{R}^s$ is emphasized.

26 POLARITY LEMMA. *Let p and q be conjugate numbers in $[-\infty, 1]$, and let $\tilde{p}$ and $\tilde{q}$ be conjugate numbers in $[1, \infty]$. Then we have*

$$\phi_p(C)\, \phi_q(D) \le \frac{1}{s} \operatorname{trace} CD \le \phi_{\tilde{p}}(C)\, \phi_{\tilde{q}}(D) \qquad \textit{for all } C, D \ge 0;$$

equality holds if and only if C^p and D^q are proportional, provided $p \ne -\infty, 1$ and the matrices C and D are positive definite. Moreover,

$$\phi_p^\infty(D) = s\phi_q(D), \quad \phi_{\tilde{p}}^0(D) = s\phi_{\tilde{q}}(D) \qquad \textit{for all } D \ge 0.$$

PROOF. We prove the concavity case, for orders $p, q \ne -\infty, 0, 1$; the orders $p, q = -\infty, 0, 1$ then follow by continuity in p. Also we can assume $C, D > 0$ and then pass to singular matrices by semicontinuity. Let λ_j and μ_j be the eigenvalues of C and D, respectively. The Hölder inequality is established as follows,

$$\begin{aligned}\phi_p(C)\,\phi_q(D) &= \left(\tfrac{1}{s}\operatorname{trace} C^p\right)^{\frac{1}{p}} \left(\tfrac{1}{s}\operatorname{trace} D^q\right)^{\frac{1}{q}} \\ &= \tfrac{1}{s}\left(\sum(\lambda_{\downarrow j})^p\right)^{\frac{1}{p}} \left(\sum(\mu_{\uparrow j})^q\right)^{\frac{1}{q}} \\ &\le \tfrac{1}{s}\sum \lambda_{\downarrow j}\mu_{\uparrow j} \\ &\le \tfrac{1}{s}\operatorname{trace} C\end{aligned}$$

The first inequality is the real Hölder inequality for $p, q < 1$, the second is the rearrangement inequality from the preceding section.

Equality holds if and only if there is a constant $\delta > 0$ such that $(\lambda_{\downarrow j})^p = \delta(\mu_{\uparrow j})^q$ for all $j \le s$, and $C = \sum \lambda_{\downarrow j} z_j z_j'$ and $D = \sum \mu_{\uparrow j} z_j z_j'$. This means that C^p must be proportional to D^q.

The Hölder inequality proves one half of the polarity formula,

$$s\phi_q(D) \le \inf_{C>0} \frac{\operatorname{trace} CD}{\phi_p(C)} = \phi_p^{\infty}(D).$$

The other half follows from inserting for C the matrix $D^{q/p}$: Its p^{th} power is D^q and fulfills the equality condition of the previous paragraph. Thus we have

$$\inf_{C>0} \frac{\operatorname{trace} CD}{\phi_p(C)} \le \frac{\operatorname{trace} D^{q/p}D}{\phi_p(D^{q/p})} = \frac{s\phi_p(D^{q/p})\,\phi_q(D)}{\phi_p(D^{q/p})} = s\phi_q(D). \qquad \diamondsuit$$

The lemma also allows to establish strict superadditivity of the matrix mean ϕ_p when the order p lies in $(-\infty, 1)$, on the open cone PD(s). Indeed, we have

$$\begin{aligned}\phi_p(C+D) &= \inf_{E>0} \frac{\operatorname{trace}(C+D)E}{s\phi_q(E)} \\ &\ge \inf_{C>0} \frac{\operatorname{trace} CE}{s\phi_q(E)} + \inf_{C>0} \frac{\operatorname{trace} DE}{s\phi_Q(E)} \\ &= \phi_p(C) + \phi_p(D).\end{aligned}$$

with strict inequality unless the last two infima are attained at a common matrix $E > 0$. According to the lemma E^q must then be proportional to both C^p and D^p, which in turn forces C and D to be proportional as well.

According to Section *10* strict superadditivity entails strict concavity; in particular, this completes the proof of Theorem *24*.

The following corollary singles out the conditions for equality in the Hölder inequality, evidently related to statements like condition c in the Existence Theorem20.

27 COROLLARY. *Let C be a positive definite $s \times s$ matrix. Then a nonnegative definite $s \times s$ matrix D solves the equations*

$$\phi_p(C)\,\phi_p^\infty(D) = \operatorname{trace} CD = 1$$

if and only if

$$D \begin{cases} = C^{p-1}/\operatorname{trace} C^p & \text{for } p \neq -\infty; \\ \in \frac{1}{\lambda_{\min}(C)} \operatorname{conv} S & \text{for } p = -\infty \end{cases}$$

where the set S consists of all matrices zz' such that z is a normalized eigenvector of C corresponding the smallest eigenvalue.

$\diamondsuit$

Notice that in all of this discussion the determinant criterion ϕ_0 plays a special role, being the only member of the p-mean family that is *selfpolar*, that is, ϕ_0 and its polar function $\phi_0^\infty = s\phi_0$ are proportional. Also we have made a step forward to verifying the conditions of the Existence Theorem *20*. The means ϕ_p with order $p \in (-\infty, 0]$ vanish for singular matrices, thus satisfying condition *b*. The means ϕ_p with $p \in (0,1)$ have polar functions $s\phi_q$ with $q \in (-\infty, 0)$; they vanish for singular matrices and are strictly isotonic, as demanded by condition *c*. The trace criterion ϕ_1 has polar function proportional to the eigenvalue criterion $\phi_{-\infty}$ which is *not* strictly isotonic.

Optimal moment matrices for a fixed information function ϕ will be characterized through equivalence theorems similar to Theorem *6*. Such theorems are based on calculus, they do not provide a systematic route how a given design might be improved to an optimal design. The latter can be partly achieved through invariance considerations—exploiting symmetries—, and leads to what in design of experiments is commonly called *balancedness*. For finite experimental domains such invariance considerations sometimes even lead to optimal designs in a much more transparent way then does the general equivalence theory. We therefore choose to first study invariant design problems, and the information increasing orderings that are generated by invariance considerations.