A QUICK INTRODUCTION TO MATHEMATICAL PROGRAMMING WITH
APPLICATIONS TO MOST POWERFUL TESTS, NONNEGATIVE
VARIANCE ESTIMATION, AND OPTIMAL DESIGN THEORY

BY
FRIEDRICH PUKELSHEIM

TECHNICAL REPORT NO. 128
MAY 21, 1978

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 75-09450

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A QUICK INTRODUCTION TO MATHEMATICAL PROGRAMMING WITH
APPLICATIONS TO MOST POWERFUL TESTS, NONNEGATIVE
VARIANCE ESTIMATION, AND OPTIMAL DESIGN THEORY

by

FRIEDRICH PUKELESHIM

TECHNICAL REPORT NO. 128
MAY 21, 1978

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 75-09450

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A QUICK INTRODUCTION TO MATHEMATICAL PROGRAMMING WITH APPLICATIONS TO MOST POWERFUL TESTS, NONNEGATIVE VARIANCE ESTIMATION, AND OPTIMAL DESIGN THEORY

by Friedrich Pukelsheim

Albert-Ludwigs-Universität Freiburg im Breisgau

and Stanford University

Author's mailing address: Friedrich Pukelsheim
Department of Statistics
Stanford University
Stanford, California 94305

Abbreviated title: PROGRAMMING EXAMPLES IN STATISTICS

AMS 1970 subject classification: Primary 62F10; 62J10; 62K05;
Secondary 90C50.

Key words and phrases: Level α tests; Neyman-Pearson lemma; MINQUE;
D-optimality; Equivalence Theorem of Kiefer & Wolfowitz.
SUMMARY

A Quick Introduction to Mathematical Programming With Applications to Most Powerful Tests, Nonnegative Variance Estimation, and Optimal Design Theory

An approach to mathematical programming is presented which requires neither topological assumptions nor convexity of objective functions and constraints. Its usefulness is exhibited with three examples from different statistical fields:

(i) The familiar characterizations for a test to be most powerful among all level α tests, or among all Neyman-Pearson tests are derived.

(ii) C. R. Rao's MINQUE for variance components is modified to obey a nonnegativity constraint, and a sufficient optimality condition for this type of estimator is obtained.

(iii) The equivalence of D- and G-optimal designs is entirely based on the primal-dual relationship of the D-optimality program and Silvey's minimal ellipsoid problem; this is done for the case of a general linear parameter, and the variances of comparison are interpreted through an appropriately chosen linear model.
1. Introduction and Summary

The purpose of this paper is to show how the theory of mathematical programming can be used to obtain results from various statistical fields. In Section 2, we formulate a 'minimal' programming approach. No convexity or topological assumptions are made; nevertheless, a basis for applications, as well as an overview of the theory, are obtained.

Each of the remaining three sections exhibits a statistical problem where advantage can be taken of the approach suggested. Section 3 treats the classical example of the zero-one-form of most powerful level $\alpha$ tests, Neyman-Pearson tests, and maximin tests; Section 4 modifies C. R. Rao's MINQUE to obey a nonnegativity constraint; and Section 5 is concerned with determinant-optimal statistical designs.

The novelty of this paper lies in the simplicity of the proposed approach. Its transparency leads almost automatically to new results and desirable generalization. In Section 4, for example, this approach shows that in the model with heteroscedastic variances the subgroup sample variance $s_k^2$ is actually optimal when the MINQUE method is required to produce nonnegative estimates.

Section 5 deals with D-optimal designs for estimating a linear parameter $C'b$, where $C$ is a $p \times s$ matrix of rank $s$. As it stands now, this section can hardly be considered a short example; but it illuminates the power and appeal of the general programming theory. Our primal program (DO) stresses the underlying reduction
by unbiasedness; more amenable to analysis is the equivalent program (DOA). In a straightforward manner, the dual program is found to be Silvey's and Sibson's minimal ellipsoid problem (ME). After an appropriate definition and interpretation of the variances of comparison $d(x, \xi)$, the Equivalence Theorem of Kiefer & Wolfowitz is generalized as expected; the proof is entirely based on the primal-dual relationship of (DO) and (ME). However, $C$-optimality has to be paired with a condition (?) that ensures existence of at least one $D$-optimal design measure $\xi^*$ whose information matrix $M(\xi^*)$ is positive-definite.

Detailed references are listed at the end of each section. For a general account of programming methods in statistics and probability see the survey of Krafft (1970) and the textbook of Rustagi (1976).
2. Mathematical Programming

By definition, a minimization program is a problem of the following type:

(P) Minimize \( f(x) \) subject to \( x \in C \) and \( r(x) \succcurlyeq 0 \).

Here \( C \) is a subset of a set \( X \), the restriction function \( r \) maps \( C \) into an ordered real vectorspace \( (U, \succcurlyeq) \), and the objective function \( f \) is a real valued function defined on the set \( \{x \in C | r(x) \succcurlyeq 0\} \) of feasible solutions of (P). An optimal solution \( x^* \) is a feasible solution such that \( f(x^*) \) attains the optimal value \( \inf\{f(x) | x \in C, r(x) \succcurlyeq 0\} \) of (P). The ordering \( \succcurlyeq \) is identified with the order cone \( K = \{u \in U | u \succcurlyeq 0\} \) via the relation \( u \succcurlyeq v \Longleftrightarrow u - v \in K \). By definition, \( K \) is an order cone if it is convex, a cone (i.e., \( \alpha K \subseteq K \) for \( \alpha > 0 \)), and \( 0 \in K \).

Now choose \( Y \) to be a real vectorspace of linear forms on \( U \), i.e., \( Y \) is a subspace of the algebraic dual of \( U \). Then \( (u, y) \leftrightarrow y(u) \) is a bilinear form on \( U \times Y \), called pairing, and denoted by \( \langle u, y \rangle \).

The order cone \( K \) in \( U \) induces the dual order cone \( K^d \) in \( Y \) that consists of all order preserving linear forms \( y \),

\[
K^d = \{ y \in Y | y(u) \geq 0 \text{ for } u \succcurlyeq 0 \},
\]

and the corresponding dual ordering \( y \succeq z \Longleftrightarrow y(z) \geq 0 \in K^d \).

We now approach the construction of a program dual to (P).

Program Generating Inequalities 2.1. Using the above notation:
\[
\inf_{x \in C; r(x) \geq 0} f(x) \geq \inf_{x \in C} \sup_{y \geq 0} f(x) - \langle r(x), y \rangle \geq \sup_{y \geq 0} \inf_{x \in C} f(x) - \langle r(x), y \rangle.
\]

Proof. In the first inequality, equality holds if the right infimum is taken over \(\{x \in C | r(x) \geq 0\}\); on the larger set \(C\), the infimum can at most be smaller. The second inequality is immediate. \(\square\)

The last line of 2.1 suggests the maximization program

\[(D)\quad \text{Maximize } g(y) \text{ subject to } y \geq 0.\]

Here the objective function \(g\) is defined on \(\mathbb{K}^d\) by

\[g(y) := \inf_{x \in C} f(x) - \langle r(x), y \rangle.\]

(D) is termed a dual candidate for (P), and the notions of optimal value and feasible and optimal solution are introduced as for (P).

(D) is called a dual program if the optimal values of (P) and (D) are equal. For clarity, (P) is referred to as the primal program.

The Program Generating Inequalities 2.1 have an immediate corollary on the relation of (P) and (D).

Optimality Check 2.2. Let \(x^*\) and \(y^*\) be feasible solutions of (P) and (D), respectively. Then \(f(x^*) = g(y^*)\) if and only if \(x^*\) is an optimal solution of (P), \(y^*\) is an optimal solution of (D), and (D) is a dual program of (P). In this case necessarily \(\langle r(x^*), y^* \rangle = 0\).
Proof. The necessity assertion follows from
\[ f(x^*) = g(y^*) \leq f(x^*) - \langle r(x^*), y^* \rangle \leq f(x^*). \]

The work that is needed before the Optimality Check can be
brought to bear on a particular minimization problem comprises
computation of both the dual order cone \( K_d \) and the dual objective
function \( g \) under a choice of \( Y \) that eases this task as far as
possible.

The present setup, although not presuming convexity or
invoking topological properties, nevertheless touches upon essential
aspects of the more elaborate programming theory:

(a) **Duality.** Primal programs \( (P) \) are accompanied by dual
candidates \( (D) \), and optimality characterizations may be obtained by
relating these two programs appropriately. There is considerable
liberty in choosing \( (D) \), and a sensible choice should favor computa-
tions as well as further insight into the underlying problem.

(b) **Minimax Theorems.** The interplay of \( (P) \) and \( (D) \) includes a
minimax problem. If \( U \) is a locally convex topological vectorspace
with a closed order cone \( K \), and if \( Y \) is chosen to be the topological
dual of \( U \), then equality holds in the first inequality of 2.1; and
the question of whether \( (D) \) is a dual program of \( (P) \) is equivalent
to a minimax theorem concerning the kernel \( K(x, y) = f(x) - \langle r(x), y \rangle \).

(c) **Conjugate Functions.** Functions like \( g \) in \( (D) \) play a promi-
inent role in the theory and motivate the notion of conjugate
functions. There exists a powerful machinery for conjugates of
convex functions, providing the basis for the theory of convex
programming, in general, and contributing to the computation of \( g \), in particular.

(d) **Kuhn-Tucker Conditions.** The necessary condition 
\[ \langle r(x^*), y^* \rangle = 0 \] 
for optimal \( x^* \), \( y^* \) is a Kuhn-Tucker type condition. It is of obvious interest how this condition can be augmented in order to obtain optimality characterizations that are also sufficient.

(e) **Existence Problems.** The above setting leads to optimality characterizations rather than existence statements. Typically, the latter involve further hypotheses, as on continuity and compactness, or on convexity and recession cones.

Our presentation has been extracted from the texts of Gol'šteĭn (1972) and Rockafellar (1974) which provide an introduction into the field of mathematical programming.
3. Most Powerful Tests

For a measurable sample space \((\mathcal{Y}, \mathfrak{A})\), consider a class \(\{p_\theta d\mu | \theta \in H + \{0\}\}\) of probability densities with dominating non-negative \(\sigma\)-finite measure \(\mu\); choose \(\alpha \in ]0,1[\). A most powerful level \(\alpha\) test for testing \(H\) versus \(\{\theta_1\}\) is an optimal solution \(\phi^*\) of the program

\[
\text{(L\alpha T)} \quad \text{Minimize } - E_{\theta_1} \phi \text{ subject to } \phi \in \Phi \text{ and } E_{\theta} \phi \leq \alpha \text{ for } \theta \in H.
\]

Here \(\Phi\) is the set of all \(\mathfrak{A}\)-measurable functions \(\phi: \mathcal{Y} \rightarrow [0,1]\), and

\[
E_{\theta} \phi = \int_{\mathcal{Y}} \phi(x) p_{\theta}(x) d\mu.
\]

Assume that \(\mathcal{H}\) is a \(\sigma\)-field on \(H\) such that \((\theta,x) \mapsto p_{\theta}(x)\) is \(\mathcal{H} \otimes \mathfrak{A}\)-measurable. By Fubini's theorem, \(r(\phi) := (\alpha - E_{\theta} \phi)_{\theta \in H}\) is a \(\mathcal{H}\)-measurable function on \(H\), whence \(U\) is taken to be the vectorspace \(B(H,\mathcal{H})\) of all bounded \(\mathcal{H}\)-measurable real functions \(h\) on \(H\), and \(K = B_+(H,\mathcal{H}) := \{h \in B(H,\mathcal{H}) | h \geq 0\}\). Choose \(Y\) to be the space \(ca(H,\mathcal{H})\) of all signed measures \(\lambda\) on \(H\), the pairing being \(\langle h, \lambda \rangle = \int_{H} hd\lambda\).

Lemma 3.1. A dual candidate for \((L\alpha T)\) is

\[
\text{(LFD)} \quad \text{Maximize } - \alpha \lambda(H) - \int_{\mathcal{Y}^+} d_+(x) d\mu \text{ subject to } \lambda \geq 0.
\]

Here \(d_+ := \max\{d,0\}\), and \(d\) is defined by

\[
d(x) := d(x,\lambda) = p_{\theta_1}(x) - \int_{H} p_{\theta}(x) d\lambda.
\]

Proof. The cone \(ca_+(H,\mathcal{H})\) of nonnegative finite measures is the dual cone of \(B_+(H,\mathcal{H})\). Fubini's theorem yields the dual objective function
\[ g(\lambda) = \inf_{\phi \in \Phi} - \int_{\mathcal{X}} \phi(x) p_{\theta_1} \, d\mu - \int_{H} (\alpha - \int_{\mathcal{X}} \phi(x) p_{\theta}(x) \, d\mu) \, d\lambda \]

\[ = -\alpha \lambda(H) - \sup_{\phi \in \Phi} \int_{\mathcal{X}} \phi(x) d(x) \, d\mu = -\alpha \lambda(H) - \int_{L} d_+(x) \, d\mu , \]

which proves the assertion. \[ \square \]

The Optimality Check 2.1 now leads easily to the familiar zero-one-form of most powerful level \( \alpha \) tests.

**Theorem 3.2.** Let \( \phi^* \) be a feasible solution of (LAT). If there exists a nonnegative finite measure \( \lambda^* \) on \((H,H)\) such that, except on a \( H \)-measurable \( \lambda^* \)-nullset, \( E_\theta \phi = \alpha \), and, except on a \( \Theta \)-measurable \( \mu \)-nullset,

\[ \phi^*(x) = \begin{cases} 
1 & \text{for } p_{\theta_1}(x) > \int_{H} p_{\theta}(x) \, d\lambda^* \\
0 & \text{for } < 
\end{cases} \]

then \( \phi^* \) is a most powerful level \( \alpha \) test for testing \( H \) versus \( \{ \theta_1 \} \).

**Proof.** Using the difference \( d \) from Lemma 3.1, we rewrite the primal objective function as

\[ f(\phi) = -E_{\theta_1} \phi = -\int_{\mathcal{X}} \phi(x)(d(x) + \int_{H} p_{\theta}(x) \, d\lambda) \, d\mu \]

\[ = -\int_{\mathcal{X}} \phi(x)(d_+(x) - d_-(x)) \, d\mu - \int_{H} E_\theta \phi \, d\lambda \ . \]

This gives

\[ f(\phi) - g(\lambda) = \int_{H} (\alpha - E_\theta \phi) \, d\lambda + \int_{d > 0} (1 - \phi(x)) d_+(x) \, d\mu \]

\[ + \int_{d < 0} \phi(x) d_-(x) \, d\mu \ . \]
For feasible $\phi$ and $\lambda$ this vanishes if and only if all three integrands are zero almost everywhere. □

The proof includes that any $\lambda^*$ of Theorem 3.2 is an optimal solution of (LFD). If the critical value $c^* := \lambda^*(\mathcal{H})$ is positive, then $\lambda^*/c^*$ is a least favorable a priori distribution for testing $\mathcal{H}$ versus $\{\theta_0\}$; and the test $\phi^*$ from above amounts to comparing the quotient $p_{\theta_1}(x)/\int_{\mathcal{H}} p_{\theta}(x)d(\lambda^*/c^*)$ with the critical value $c^*$.

Existence of an optimal $\phi^*$ follows from the sequential compactness of $\phi$ in the $L_1$-topology of $L_\infty(\mathcal{Y},\mathcal{R},\mu)$. An optimal $\lambda^*$ need not exist, but is easily exhibited in case of a simple hypothesis $\mathcal{H} = \{\theta_0\}$. Moreover, Krafft & Witting (1967) prove the general result that (LFD) is, in fact, a dual program of (LoT).

Next, consider the program for Neyman-Pearson tests:

\[(\text{NPT}) \quad \text{Minimize } -E_{\theta_1} \phi \text{ subject to } \phi \in \Phi \text{ and } E_{\theta} \phi = \alpha(\theta) \text{ for } \theta \in \mathcal{H}.\]

Here $\alpha: \mathcal{H} \to ]0,1[$ is a prescribed $\mathcal{H}$-measurable function. The equality constraint $E_{\theta} \phi = \alpha(\theta)$ corresponds to the trivial order cone $K = \{0\} \subset B(\mathcal{H},\mathcal{H})$ whose dual is $K^d = c_a(\mathcal{H},\mathcal{H})$. With obvious changes, the preceding calculations yield the following.

**Neyman-Pearson Lemma 3.3.** Let $\phi^*$ be a feasible solution of (NPT). If there exists a signed measure $\lambda^*$ on $(\mathcal{H},\mathcal{H})$, i.e., a countably additive set function $\lambda^*: \mathfrak{M} \to \mathbb{R}$, such that, except on a $\Theta$-measurable $\mu$-nullset,
\[ \phi^*(x) = \begin{cases} 1 & \text{for } \frac{p_{\theta_1}}{p_{\theta}}(x) > \int_H p_{\theta}(x) d\lambda^* \\ 0 & \text{for } \frac{p_{\theta_1}}{p_{\theta}}(x) < \end{cases} \]

then \( \phi^* \) is most powerful among all tests \( \phi \in \Phi \) which satisfy \( E_\theta \phi = \alpha(\theta) \) for \( \theta \in H \). □

The extension of the above to maximin level \( \alpha \) tests is straightforward, see Krafft & Witting (1967). For further references, the reader is referred to the textbooks of Schmetterer (1974, pp. 160-185), or Witting (1966, pp. 69-75, 92-98).
4. Nonnegative Variance Estimation

Consider the estimation of a nonnegative linear combination
\[ \sum q_k^2 \sigma_k^2, \quad q_k \geq 0, \] of variance components \( \sigma_k^2 \), when for a random
\( \mathbb{R}^n \)-vector \( Y \) a linear model \( Y \sim (Xb; \sigma_1^2 V_1 + \ldots + \sigma_k^2 V_k) \) is assumed.

C. R. Rao's (1973, pp. 302-305) MINQUE method yields a quadratic
form \( Y^* A Y \) which has several desirable properties, but is not safeguarded against producing negative estimated values for the nonnegative parameter \( \sum q_k \sigma_k^2 \). This defect is evaded by adjoining a
restriction to nonnegative-definite (NND) forms \( Y^* A Y \), and we call
\( Y^* A Y \) the best nonnegative estimate for \( \sum q_k \sigma_k^2 \) when \( A^* \) is the
optimal solution of

\[ \text{(BNE)} \quad \text{Minimize } \|A\|^2 \text{ subject to } A \in \text{Unb}(q) \text{ and } A \succeq 0. \]

This program is defined on the space \( X = \text{Sym}(n) \) of symmetric \( n \times n \)
matrices, with \( \|A\|^2 = \text{trace } A^2 \), and order cone NND(\( n \)) of NND
matrices. By definition, \( A \in \text{Unb}(q) \) if \( Y^* A Y \) is unbiased for
\( \sum q_k \sigma_k^2 \). Choose \( Y = \text{Sym}(n) \), the pairing being the Euclidean matrix
inner product \( \langle A, B \rangle = \text{trace } AB' \). Define \( N : \text{Sym}(n) \rightarrow \text{Sym}(n) \) to be
the orthogonal projector onto the subspace \( \{A \in \text{Unb}(0) | AX = 0\} \) of
all unbiased invariant quadratic estimators of zero. The following
uses the MINQUE \( \hat{A} \).

Lemma 4.1. A dual candidate for (BNE) is

\[ \text{(LFR)} \quad \text{Maximize } \|\hat{A}\|^2 - \langle \hat{A}, B \rangle - \frac{1}{q} \|N(B)\|^2 \text{ subject to } B \succeq 0. \]
Proof. \( \text{NND}(n) \) is selfdual. We may impose
\[
A \in \{ \hat{A} + N(Z) | Z \in \text{Sym}(n) \}, \text{ rather than } A \in \text{Unb}(q), \text{see Pukelsheim (1977, Lemma 3.2)}. \text{ Then the fact } N(\hat{A}) = 0 \text{ implies }
\]
\[
\| \hat{A} + N(Z) \|^2 + \langle \hat{A} + N(Z), B \rangle = \| \hat{A} \|^2 - \langle \hat{A}, B \rangle - \frac{1}{4} \| N(B) \|^2 + \| N(Z - \frac{1}{2} B) \|^2, \text{ from which the assertion follows.} \]

The following sufficiency condition is obtained using the Optimality Check 2.2.

**Lemma 4.2.** If there exists a \( \text{NND} \) \( B \) such that \( N(B) \neq 0 \), and
\[
A^* := \hat{A} - \langle \hat{A}, B \rangle \| N(B) \|^{-2} N(B) \text{ is NND, then } A^* \text{ is the best non-negative estimator for } \sum q_k \sigma_k^2.
\]

Proof. \( A^* \) and \( B^* := -2 \langle \hat{A}, B \rangle \| N(B) \|^{-2} B \) are feasible solutions of (BNE) and (LFR), and both objective functions are equal to
\[
\| \hat{A} \|^2 + \langle \hat{A}, B \rangle^2 \| N(B) \|^{-2}. \]

Clearly \( \hat{A} = A^* \) if and only if \( \hat{A} \) is NND. Otherwise its negative part
\[
\hat{A}_- := \sum_{\lambda \in \text{spectrum}(\hat{A})} \lambda_- E(\lambda)
\]
does not vanish, where \( \lambda_- = \max\{0, -\lambda\} \) and \( E(\lambda) \) is the orthogonal projector onto the eigenspace of \( \lambda \). As a special case of Lemma 4.2, we obtain

**Corollary 4.3.** If \( N(\hat{A}_-) \neq 0 \), and
\[
A^* := \hat{A} + \| \hat{A} \|^2 \| N(\hat{A}_-) \|^{-2}
\]
\( N(\hat{A}_-) \) is \( \text{NND} \), then \( A^* \) is the best nonnegative estimator for
\[
\sum q_k \sigma_k^2. \]

13
Example. In the model of heteroscedastic variances
\[ EV_{\kappa \nu} = \mu; \ Var Y_{\kappa \nu} = \sigma^2_{\kappa}; \ \nu = 1, \ldots, n_\kappa; \ \kappa = 1, \ldots, k \]
with uncorrelated \( Y_{\kappa \nu} \)'s, the MINQUE of a single \( \sigma^2_\kappa \) is
\[ Y'AY = \frac{n_\kappa}{n_\kappa(n-2)} \sum_{\nu=1}^{n_\kappa} (Y_{\kappa \nu} - \overline{Y}_\kappa)^2 - \frac{s^2}{n-2}, \ s^2 = \frac{1}{n-1} \sum_{\lambda=1}^{k} \sum_{\nu=1}^{n_\lambda} (Y_{\lambda \nu} - \overline{Y}_\cdot)^2. \]
The only negative eigenvalue of \( \hat{A} \) is \( -(n-1)^{-1} (n-2)^{-1} \), and after some lengthy computations one obtains for the \( A^* \) in Corollary 4.3 that
\[ Y'A^*Y = \frac{1}{n_{\kappa-1}} \sum_{\nu=1}^{n_\kappa} (Y_{\kappa \nu} - \overline{Y}_{\kappa .})^2 = s^2_\kappa. \]
Hence, the sample variance of the \( \kappa \)th group provides not only the customary unbiased estimator (J.N.K. Rao 1973, p. 13), but is also optimal. □

T-weighted MINQUE can be discussed analogously, or by applying the program (BNE) to the transformed model
\[ Z \sim (Xb; \ \Sigma \sigma^2_{\kappa} T^2_k \nu_k T_k^2). \]—Existence and uniqueness of an optimal \( A^* \) follows from a convexity argument; for other detailed results see Pukelsheim (1977).
5. D-Optimal Designs

Let \( C' \) be a real \( s \times p \) matrix of rank \( s \), and consider the estimation of a linear parameter \( C'b \) when from a family of linear models \( Y \sim (Xb; \sigma^2 I_n) \) a particular member may be chosen for experimental realization. More precisely, the design matrix \( X \) arises from a design space \( \mathcal{X} \) through a function \( f: \mathcal{X} \to \mathbb{R}^p \) according to \( X' = [f(x_1): \ldots : f(x_n)] \), in this case \( (x_1, \ldots, x_n) \) is called a discrete design. For a fixed design matrix \( X \), the procedure considered is the least squares estimate \( C'X'Y \). Its dispersion matrix

\[
(5.1) \quad \sigma^2 C'(X'X)^+ C = \frac{c^2}{n} C' \left( \frac{1}{n} X'X \right)^+ C
\]

is an accepted measure of performance if \( C'X'Y \) is unbiased for \( C'b \).

The unbiasedness requirement is equivalent to

\[
(5.2) \quad X'X(X'X)^+ C = C
\]

and subject to this constraint one may wish to minimize a meaningful functional of (5.1), preferably the generalized variance

\[
\det C' \left( \frac{1}{n} X'X \right)^+ C.
\]

One has \( \frac{1}{n} \ X'X = \sum \frac{1}{n} f(x_v)f(x_v)' = \int f(x)f(x)'d\xi \)

where \( \xi \) is the uniform distribution over the discrete design \( (x_1, \ldots, x_n) \), and in generalization of this---the end justifies the means---\( \xi \) is taken to vary over all probability measures on \( \mathcal{X} \). This background motivates the following.
Definitions and Notations 5.1. (a) Let $(\mathcal{X}, \mathcal{B})$ be a measurable space, called design space. The vectorspace of all real countably additive set functions on $\mathcal{B}$, i.e., signed measures on $\mathcal{X}$, is denoted by $\text{ca}(\mathcal{X}, \mathcal{B})$. If $\xi$ is a member of the cone $\text{ca}_+(\mathcal{X}, \mathcal{B})$ of nonnegative finite measures, we shall write $\xi \geq 0$. Let $\Xi$ stand for the set of all probability measures on $(\mathcal{X}, \mathcal{B})$, any $\xi \in \Xi$ is called a design measure. Assume that in $\mathcal{X}$ all one-point sets are $\mathcal{B}$-measurable, and let $\delta(x) \in \Xi$ denote the one-point measure at $x$.

(b) Introduce $\text{NND}(p)$ and $\text{PD}(p)$ for the cones of all nonnegative-definite and positive-definite matrices in the vectorspace $\text{Sym}(p)$ of all real symmetric $p \times p$ matrices. Let $C$ be a fixed real $p \times s$ matrix of rank $s$, with $1 \leq s \leq p$. Define

$$
(5.3) \quad G := \{A \in \text{NND}(p) \mid AA^+ C = C\} .
$$

(c) Let $f : \mathcal{X} \to \mathbb{R}^p$ be a bounded $\mathcal{B}$-measurable function not identically zero, whose image $f(\mathcal{X})$ is closed in $\mathbb{R}^p$. Define the linear operator $M : \text{ca}(\mathcal{X}, \mathcal{B}) \to \text{Sym}(p)$ by

$$
(5.4) \quad M(\xi) := \int_{\mathcal{X}} f(x)f(x)' \, d\xi .
$$

For a design measure $\xi \in \Xi$, $M(\xi)$ is called its information matrix. A design measure $\xi \in \Xi$ is termed feasible (for estimating $C\beta$) if $M(\xi) \in G$, and the set of all feasible $\xi$ is denoted by

$$
(5.5) \quad \Xi(C) := \{\xi \in \Xi \mid M(\xi)M(\xi)^+ C = C\} .
$$

A feasible design measure $\xi^* \in \Xi(C)$ is called a determinant optimal design (for estimating $C\beta$) if it minimizes the function
det \( C' M(\xi)^+ C \) over the set \( \mathcal{E}(C) \) of all feasible designs.

Remarks 5.2. (a) In Sym(p), NND(p) forms a closed convex cone with interior PD(p). With respect to the Euclidean matrix inner product \( \langle A, B \rangle = \text{trace } AB' \), the cone NND(p) is selfdual.

(b) The operator \( M \) is order preserving, in that \( \xi \geq 0 \) implies \( M(\xi) \in \text{NND}(p) \). The boundedness of \( f \) ensures that \( M \) is well-defined, the closedness of \( f(\lambda) \), actually its compactness, will be utilized to prove closedness of the cone \( \{ M(\xi) | \xi \geq 0 \} \). Our assumptions are satisfied, in particular, if \( \lambda \) is compact, and \( f \) is continuous.

(c) The definition (5.3) of \( G \) originates with the unbiasedness requirement (5.2); we wish to stress that this definition does not involve any design measures. One has \( A \in G \) if and only if range \( C \subset \text{range } A \), or, equivalently, nullspace \( A \subset \text{nullspace } C' \). Hence, \( G \) is a convex cone, but it does not contain 0. \( G \) comprises all positive-definite matrices; in fact, \( \text{PD}(p) = \text{int } G \subset G \subset \text{cl } G = \text{NND}(p) \). \( G \) is, in general, neither open nor closed.

A determinant optimal design is characterized as an optimal solution of the program

\[
\text{(DO)} \quad \text{Minimize } \det C' M(\xi)^+ C \text{ subject to } \xi \in \mathcal{E}(C) .
\]

A decisive problem reduction is found by appropriately replacing the 'big' space \( \mathcal{L}(\mathcal{X}, \mathcal{B}) \) by the 'small' space \( \mathbb{R} \times \text{Sym}(p) \). Change the objective function to \( \ln \det C' M(\xi)^+ C + s(\xi(X) - 1) \); the logarithm serves for mathematical convenience, and the second term allows to drop the constraint \( \xi(X) = 1 \). The same information as in (DO) is
now conveyed by the program

\[
\begin{align*}
\text{Minimize} & \quad \ln \det C' A^+ C + s(u - 1) \\
\text{(DOA)} & \quad \text{subject to} \quad (u, A) \in \mathbb{R} \times \mathbb{G} \\
& \quad \text{and} \quad (u, A) \in \{(\xi(I), M(\xi)) | \xi \geq 0\}
\end{align*}
\]

For the natural dual space \( \mathbb{R} \times \text{Sym}(p) \) we choose the pairing

\[
(u, A), (v, N) := uv - \text{trace} \, AN
\]

\[
= uv - \langle A, N \rangle
\]

Derivation of the dual candidates, now, requires no more than simple calculations.

**Candidate Lemma 5.3.** (a) A dual candidate for (DOA) is

\[
\begin{align*}
\text{Maximize} & \quad \ln \det C'NC \\
\text{(MEN)} & \quad \text{subject to} \quad N \in \text{NND}(p), C'NC \in \text{PD}(s), \; \nu = s \\
& \quad \text{and} \quad f(x)' N f(x) \leq \nu \quad \text{for} \; x \in \mathcal{X}
\end{align*}
\]

(b) A dual candidate for (DO) is

\[
\begin{align*}
\text{Maximize} & \quad \det C'NC \\
\text{(ME)} & \quad \text{subject to} \quad N \in \text{NND}(p) \\
& \quad \text{and} \quad f(x)' N f(x) \leq s \quad \text{for} \; x \in \mathcal{X}
\end{align*}
\]

(c) Both (MEN) and (ME) have feasible solutions.

**Proof.** (c) Since \( f \) is bounded but not zero, one has

\[
\alpha = \sup_{x \in \mathcal{X}} \| f(x) \| \in (0, +\infty] 
\]

Hence, \( \frac{s}{\alpha} \mathbb{I}_p \) and \( (s, \frac{s}{\alpha} \mathbb{I}_p) \) are
feasible solutions of (ME) and (MEN), respectively, and the transition from one program to the other causes no problem. We shall prove statement (a).

(a) \( K := \{(\xi(\lambda), M(\xi)) | \xi \geq 0\} \) is an order cone, since it is the image of the order cone \( ca_+(\lambda, \beta) \) under the linear mapping \( \xi \rightarrow (\xi(\lambda), M(\xi)) \). By definition, one has \((v, N) \in K^d\) if and only if \(0 \leq uv - \langle A, N \rangle\) for all \((u, A) \in K\). Equivalently, \(\langle M(\xi), N \rangle \geq v\) for all \(\xi \in \Xi\); and consideration of one point measures yields the constraint as asserted.

Computation of the dual objective function is done in several steps. If we set

(5.7) \[ h(A) := \ln \det C' A^+ C + \langle A, N \rangle \]

then the problem is one of finding

(5.8) \[ g(v, N) = -s + \inf_{u \in R; A \in G} h(A) + u(s - v) \]

I. If \( v \neq s \), then \( g(v, N) = -\infty \).

II. Using \((I + \beta b'b')^{-l} = I - \beta b'b'/(1 + \beta b'b')\), and \(\det(I + bc') = 1 + c'b\), we have for \(\beta \geq 0\) and \(b \in R^p\) that \(I_p + \beta b'b' \in G\), and

(5.9) \[ h(I_p + \beta b'b') = \ln \det C'C + \text{trace } N + \beta b'Nb \]

\[ + \ln (1 + \beta b'(I_p - CC^t)b) - \ln (1 + \beta b'b') \]

III. If \(N \notin \text{NND}(p)\), then there exists a \(b \in R^p\) with \(b'Nb < 0\), and (5.9) leads to \(g(s, N) = -\infty\).
IV. If \( C'NC \notin PD(s) \), then there exists a \( b \in \mathbb{R}^p \) with \( 0 \neq b \in \text{range } C \cap \text{nullspace } N \); and \( h(I_p + \beta bb') = \ln \det C'C + \text{trace } N - \ln(1 + \beta b'b) \) yields again \( g(s,N) = -\infty \).

V. We claim that

\[
\{ C'A^+C \mid A \in G \} = PD(s).
\]

First, fix \( A \in G \). Then \( C'A^+Cb = 0 \Rightarrow A^+Cb = 0 \Rightarrow 0 = C^+AA^+Cb = b \), hence \( C'A^+C \in PD(s) \). Second, fix \( B \in PD(s) \), and define \( A := CB^{-1}C' \).

The four Penrose criteria show that \( A^+ = C'^+B^{-1}C'B \); hence \( C'A^+C = B \), \( AA^+C = C \), and \( B \in \{ C'A^+C \mid A \in G \} \).

VI. For the \( N \) under consideration, the last argument applies to \( B := C'NC \); hence \( C(C'NC)^{-1}C' \in G \), and

\[
\inf_{A \in G} h(A) \leq h(C(C'NC)^{-1}C') = s + \ln \det C'NC.
\]

VII. If \( N \in NNDD(p) \) and \( A \in G \), then

\[
<A,N> \geq <(C'A^+C)^{-1}, C'NC>.
\]

Since \( HR^+ \) is an orthogonal projector, this is an application of the Pythagorean Theorem:

\[
<A,N> = \left\| N^{1/2} A^{1/2} \right\|^2 \geq \left\| N^{1/2} A^{1/2} \right\|^2.
\]

\[
A^{1/2} + C(A^{1/2})^+ C^+ \right\|^2 = \text{trace } AA^+C(C'A^+C)^{-1} (C'A^+C)(C'A^+C)^{-1} C'AA^+N = <(C'A^+C)^{-1}, C'NC>.
\]

VIII. The final argument runs as follows: \( \inf_{A \in G} h(A) \geq \inf_{A \in G} \ln \det C'A^+C + <(C'A^+C)^{-1}, C'NC> = \inf_{B \in PD(s)} \ln \det B + <B^{-1}, C'NC> = s + \ln \det C'NC \). The last equality follows from
differential calculus. Together with (5.11) this completes the proof.

The program (ME) is Silvey's minimal ellipsoid problem (Wynn 1972, p. 174; Sibson 1974). By calling (ME) a dual candidate for (DO) we admittedly trespass the notational conventions as set up in Section 2.

The Optimality Check 2.2, augmented by the arithmetic-geometric mean inequality, provides the central technical result.

Optimality Characterization 5.4. Let $\xi^*$ and $N^*$ be feasible solutions of (DO) and (ME), respectively. Then $\xi^*$ and $N^*$ are optimal and (ME) is a dual program of (DO), if and only if

$$C'M(\xi^*)^+C = C'N^*C.$$  \hspace{1cm} (5.13)

In this case, necessarily, $<M(\xi^*),N^*> = s$,

$$N^*M(\xi^*) = N^*C(C'M(\xi^*)^+C)^{-1}C'.$$  \hspace{1cm} (5.14)

and, except for a $\emptyset$-measurable $\xi^*$-nullset, $f(x)'N^*f(x) = s$.

Proof. For greater legibility, we write $A = M(\xi^*)$, and $N$ instead of $N^*$. The proof will show that the matrices in (5.13) are equal if their determinants coincide, the converse being obvious. The argument is based on the following inequality string:

$$s \geq <A,N> = \|N^{1/2}A^{1/2}\|^2$$

$$\geq \|N^{1/2}A^{1/2}(A^{1/2}+C)(A^{1/2}+C)^+\|^2 = <(C'A+C)^{-1}, C'NC>$$

$$\geq s \det^{1/s}(C'A+C)^{-1} C'NC = s.$$
The first inequality is obtained from integrating \( f(x)'Nf(x) \leq s \), the second is (5.12) above, and the third is the arithmetic-geometric mean inequality \( \text{trace } B/s \geq \det^{1/s} B \) for \( B \in \mathbb{NND}(s) \), with equality only for \( B = \beta I_S \). Hence (5.15) takes care of (5.13), \( \langle A, N \rangle = s \), and the last statement. Also \( \|N^{1/2}A^{1/2}[I-(A^{1/2}+C)(A^{1/2}+C)^+]\| = 0 \), and \( NA = NA^{1/2}A^{1/2}+C(C'A'C)^{-1}C'A^{1/2}A^{1/2} \) provides the argument for (5.14). \( \Box \)

Next we list an existence statement. Its proof goes beyond our exposition in Section 2, but relies on standard minimax or programming arguments.

**Duality Theorem 5.5.** Assume that \( M(\xi) \) is positive-definite for at least one \( \xi \in \Xi \). Then the optimal solutions of (ME) form a nonempty compact convex set, and (ME) is a dual program of (DO).

**Proof.** Pick a design measure \( \xi \in \Xi \) so that \( A := M(\xi) \) is positive-definite. Integration and the general fact \( \|G \cdot H\| \leq \|G\| \cdot \|H\| \) yield \( s \geq \langle A, N \rangle = \|A^{1/2}N^{1/2}\|^2 \geq \beta^{-1} \|N\|^2 \geq \beta^{-1} \|N\| \), with \( \beta := \|A^{-1/2}\|^2 > 0 \). Hence those \( N^* \) that maximize the continuous concave function \( \det^{1/s} C'NC \) over the compact convex set \( \{N \in \mathbb{NND}(p)/f(x)'Nf(x) \leq s \text{ for } x \in \mathcal{X}\} \) form a nonempty compact convex set as asserted.

The duality statement follows from Theorem 18(d) in Rockafellar (1974, p. 42). Using his setup (op. cit., p. 26) we have to verify (I) convexity of the primal objective function of (MEN) over the convex set \( \mathbb{R} \times \mathcal{G} \), and (II) closedness of the order cone \( K \) in (MEN).
I. It suffices to show that the concavity of

\[(5.16) \quad \gamma(A) := (C' A^+ C)^+\]
on PD(p) = int G extends to all of G. This is true once

\[(5.17) \quad \gamma(A) = \lim_{\varepsilon \to 0} \gamma(A + \varepsilon I_p)\]

holds for A \in G. But writing A = X'X we have (Albert 1971, p. 19) that
\[\lim \gamma(A + \varepsilon I_p) = \lim (C'(X'X + \varepsilon I_p)^{-1} X'X A^+ C)^{-1} = (C'(X'X)^+ X'X A^+ C)^{-1} = \gamma(A).\]

II. Let \((\xi_i)_{i \in \mathbb{N}} \in \text{ca}_+ (\mathcal{L}, \mathcal{B})^{\mathbb{N}}\) be a sequence such that \(\xi_i \to u_\infty\), and \(M(\xi_i) \to A_\infty\). If \(u_\infty = 0\), then \(A_\infty = 0\), and \((u_\infty, A_\infty) \in K\); if \(u_\infty > 0\), we may assume \(u_\infty = 1\) and \(\xi_i \in \mathcal{E}\) for all \(i \in \mathbb{N}\). This leaves to prove that \(M(\mathcal{E})\) is closed. But \(H = \{bb' \in \text{Sym}(p) \mid b \in f(\mathcal{L})\}\) is compact since \(f(\mathcal{L})\) is compact, and \(\text{cl conv } H = \text{conv } H \subseteq M(\mathcal{E}) \subseteq \text{cl conv } H\); see Sibson (1974, p. 676).

Note that our slightly bizarre pairing (5.6) is still compatible with the standard topology of \(\mathbb{R} \times \text{Sym}(p)\), cf. Rockafellar (1974, p. 13).--In view of remark (b) in Section 2, the duality statement also follows from von Neumann's minimax theorem, as given in Gol'stein (1972, Thm. 1, p. 8).--The concavity of \(\gamma\) in (5.16) does not, in general, extend to \(\text{cl } G = \text{NND}(p)\).

In our approach to the Equivalence Theorem, we start with delineating the rationale underlying Kiefer & Wolfowitz's discovery. Estimating the full parameter \(b\), \(D\)-optimal design measures \(\xi^*\) that minimize the generalized variance \(\det M(\xi)^{-1}\) of the \(R^p\)-valued least
squares estimate $X^+Y$ coincide with G-optimal or minimax design measures that minimize the maximal variance in a family $\mathcal{F}$ of real valued estimates closely related to $X^+Y$. More precisely, the family $\mathcal{F}$ consists of all $f(x)'X^+Y$, $x \in \mathcal{Y}$; these are interpreted as unbiased estimates of $f(x)'b$, but enter into consideration only through their variances

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

When estimating the first $s$ components of $b$, i.e., $C' = [I_s : 0]$, a generalization of $d(x, \xi)$ is available, using a corresponding partitioning $M(\xi) = \begin{bmatrix} M_1 & M_2 \\ M_2' & M_3 \end{bmatrix}$, and $f = \begin{bmatrix} f(1) \\ f(2) \end{bmatrix}$:

$$d(x, \xi) = [f(1)'(x) - M_2M_3^+[f(2)'(x)]\left(M_3^+[f(1)'(x) - M_2M_3^+[f(2)'(x)]\right].$$

Here $C, C^+$, and $CC^+$ are hardly distinguishable, but the nature of the program (DO) suggests that only the range of $C$, and hence $CC^+$, should play any role.

Thus we are led to defining the function $d: \mathcal{X} \times \mathcal{S} \rightarrow \mathbb{R}$, for the fixed $p \times s$ matrix $C$ of rank $s$, by

$$d(x, \xi) := \|A^{1/2}C^+[I_p - A(QAQ)^+]f(x)\|^2$$

where $A = M(\xi)$, and $Q := I_p - CC^+$. It is readily checked that definition (5.20) specializes to (5.18) and--using $(QAQ)^+ = Q(QAQ)^+Q$--to (5.19).

Definition (5.20) is intimidatedly interwoven with rank-free BLUE formulas. For when a random $\mathbb{R}^p$-vector $f(x)$ follows a linear model
(5.21) \[ f(x) \sim (Ca; \sigma^2 A), \]

the BLUE of \( a \) is (Albert 1971, p. 90)

(5.22) \[ \hat{a}(x) := C^+[I_p - A(QAQ)^+]f(x) \]

With this abbreviation, \( d \) attains the form

(5.23) \[ d(x, \xi) = \hat{a}(x)' C' A^+ C \hat{a}(x), \]

and for a given \( \xi \), these are the variances of the family 
\[ \mathcal{F} = \{ \hat{a}(x)' C' X^+ Y | x \in \mathcal{X} \} \]
with which \( C' X^+ Y \) is to be compared.

If \( \xi \in \mathcal{E}(C) \) is feasible, then \( AA^+ C = C \). Under this condition (and only then) the BLUE (5.22) agrees with weighted least squares estimation, as is well known. More exactly, one has the matrix equality

(5.24) \[ C^+ [I_p - A(QAQ)^+] = (C'A^+ C)^{-1} C'A^+ \]

[Sketch of proof: Put \( A = V^2 \). Show \( (V'C)^+ = C^+ V[I_p - (QV)^+ QV] \).
Conclude \( (V'C)^+ V^+ = C^+ [I_p - V(QV)^+] \).

Hence on \( \mathcal{X} \times \mathcal{E}(C) \), the function \( d \) has the form

(5.25) \[ d(x, \xi) = f(x)' A^+ C(C'A^+ C)^{-1} C'A^+ f(x), \quad A = M(\xi) \]

This is the representation we shall use, cf., Kiefer (1974, eq. (4.23)).

The following theorem carefully respects the hierarchy governing the various statements.

Main Theorem 5.6. Assume that there exists at least one design measure which has a positive-definite information matrix.

25
(a) If \( \xi^* \in \Xi(C) \) is a feasible design and
\[ \sup_{x \in \mathcal{X}} d(x, \xi^*) = s, \]
then \( \xi^* \) is optimal. Moreover, the supremum in (5.26) is not only attained, but also one has, except for a \( B \)-measurable \( \xi^* \)-nullset, \( d(x, \xi^*) = s \).

(b) If \( \xi^* \) is optimal and has a positive-definite information matrix, then \( \sup_{x \in \mathcal{X}} d(x, \xi^*) = s \).

(c) If \( \xi^* \) is optimal and has a \( n \)-point support \( x_1, \ldots, x_n \in \mathcal{X} \), then \( d(x, \xi^* ) = s \) for all \( \nu = 1, \ldots, n \).

(d) There exists a sequence \( (\xi_i)_{i \in \mathbb{N}} \) of feasible designs such that
\[ A_i := M(\xi_i) \in \text{PD}(p) \] for \( i \in \mathbb{N} \),
and \( \det C' A_i^+ C \) converges to the optimal value of program (DO).

(e) If the sequence in (d) can be chosen such that, in addition, \( (A_i^{-1})_{i \in \mathbb{N}} \) is bounded, then
\[ \inf_{\xi \in \Xi(C)} \sup_{x \in \mathcal{X}} d(x, \xi) = s. \]

Proof. (a) Define \( A := M(\xi^*) \). Under the hypothesis, \( N := A^+ C (C' A^+ C)^{-1} C' A^+ \) is a feasible solution of (ME) to which (5.13), and hence the conclusions of Theorem 5.4, apply.
(b) Choose an optimal $N$ so that with $A = M(\xi^*)$ we have
$$\det C'A^+C = \det C'NC,$$ by Theorem 5.5. Premultiplying (5.14) by $N^{1/2}^+$ and postmultiplying by $A^{1/2}^+$, one obtains
\begin{equation}
N^{1/2}A^+A^+ = N^{1/2} C(C'A^+C)^{-1} C'A^+.
\end{equation}
In view of (5.13) and (5.25) we get for $x \in \mathcal{X}$
\begin{equation}
d(x, \xi^*) = f(x)' A^+ A^+ A A^+ f(x).
\end{equation}
Using the nonsingularity of $A$, (5.30) and feasibility of $N$ show that $\sup_{x \in \mathcal{X}} d(x, \xi^*) \leq s$. Again the converse inequality is due to the fact that $d(x, \xi^*)$ integrates to $s$.

(c) $\xi^* = \sum p_\nu \delta(x_\nu)$ is a convex combination of one point measures, $p_\nu > 0$, $\Sigma p_\nu = 1$. $A^+$ projects onto the range of $A$, but range $\sum p_\nu f(x_\nu)f(x_\nu)' = \Sigma$ range $f(x_\nu)f(x_\nu)' = \text{span} \{f(x_1), \ldots, f(x_n)\}$, whence $A^+ f(x_\nu) = f(x_\nu)$. (5.30) and again an integration argument prove the assertion.

(d) Recall that $M(\Xi)$ is convex. Choose any $A$ in its relative interior $\text{ri} M(\Xi)$. Select one of the hypothesized $\xi \in \Xi$ which has a nonsingular $M(\xi)$. Then $\frac{1}{2} A + \frac{1}{2} M(\xi)$ lies in both $\text{ri} M(\Xi)$, and $\text{ri} G = \text{int} G = \text{PD}(p)$, whence $\text{ri} M(\Xi(G)) = \text{ri} M(\Xi) \cap G = \text{ri} M(\Xi(\cap G) = \text{ri} M(\Xi(\cap \text{int} G) \subseteq \text{PD}(p)$; see Rockafellar (1970, Thm. 6.1, p. 45; Thm. 6.5, p. 47). This ensures the existence of optimizing sequences $(\xi_i)_{i \in \mathbb{N}}$ with positive-definite information matrices $A_i$.

(e) Fix an optimal $N$ and a sequence $(A_i)_{i \in \mathbb{N}}$ from (5.27) for which $(A_i^{-1})_{i \in \mathbb{N}}$ is bounded. The following version of (5.15) is true:
\[(5.31) \quad s \geq \langle A_{1}N \rangle \geq (C'A_{1}^{+}C)^{-1}C'NC \geq s \det^{-\frac{1}{s}}(C'A_{1}^{+}C)^{-1}C'NC \geq s .\]

The first two inequalities ensure that \((C'A_{1}^{+}C)_{i \in \mathbb{N}}\) is bounded, hence there exist convergent subsequences. Any limit other than \(C'NC\) conflicts, however, with (5.31); therefore,

\[(5.32) \quad C'A_{1}^{+}C \to C'NC .\]

As in (5.15), we infer also that

\[(5.33) \quad N^{1/2}A_{1}^{1/2} - N^{1/2}C(C'A_{1}^{+}C)^{-1}C'A_{1}^{-1/2} \to 0 .\]

Now \(\|(A_{1}^{-1/2}) \| \leq \|A_{1}^{-1}\| \cdot \|A_{1}^{1/2}\| \leq \text{trace} \cdot \sup_{A_{1}}\|A_{1}^{-1}\|\); and since \((A_{1})_{i \in \mathbb{N}}\) is in the bounded set \(M(\mathbb{X})\), we find that \((A_{1}^{-1/2})_{i \in \mathbb{N}}\) is bounded, too. Postmultiplication of (5.33) is thus permissible, and because of the nonsingularity of \(A_{1}\) we obtain \(N^{1/2}C(C'A_{1}^{+}C)^{-1}C'A_{1}^{+} \to N^{1/2}\), giving

\[(5.34) \quad A_{1}^{+}C(C'A_{1}^{+}C)^{-1}C'NC(C'A_{1}^{+}C)^{-1}C'A_{1}^{+} \to N .\]

The boundedness of \((A_{1}^{-1})_{i \in \mathbb{N}}\), and (5.32), force the remainder term \(\|A_{1}^{-1}C(C'A_{1}^{+}C'[C'A_{1}^{+}C-C'NC](C'A_{1}^{+}C)C'A_{1}^{-1}\|\) to converge to zero. We may conclude then, from (5.25) and (5.34), that for all \(x \in \mathbb{X}\)

\[(5.35) \quad d(x, \xi_{1}) \to f(x)'Nf(x) .\]

Since \(f(\mathbb{X})\) is compact, the convergence in (5.35) is uniform in \(x\).

This and again an integration argument show that for all \(\varepsilon > 0\) there exists a \(i \in \mathbb{N}\) such that
(5.36) \[ s \leq \sup_{x \in \mathcal{L}} d(x, \xi) \leq s + \varepsilon, \]

thus establishing (5.28). \[ \square \]

We conclude our investigations by extracting the equivalence of D- and G-optimality for design measures $\xi^*$. 

**Equivalence Theorem 5.7.** Let $\mathcal{E}$ be the set of design measures $\xi$, i.e., probability measures, on a measurable design space $(\mathcal{L}, \mathcal{B})$ whose one-point sets are measurable. Let $f: \mathcal{L} \to \mathbb{R}^P$ be a $\mathcal{B}$-measurable function with compact image $f(\mathcal{L})$, inducing the information matrices $M(\xi) := \int f(x)f(x)' \, d\xi$. Assume there exists a discrete design $(x_1, \ldots, x_n) \in \mathcal{L}^n$ whose design matrix $X' := [f(x_1) : \ldots : f(x_n)]$ has full rank $p$. Let $C$ be a $p \times s$ matrix of rank $s$, and $E(C) := \{ \xi \in \mathcal{E} | M(\xi)M(\xi)^+ C = C \}$ the set of design measures feasible for estimating the linear parameter $C'b$. Define the variances of comparison $d$ on $\mathcal{L} \times E(C)$ by \[ d(x, \xi) := f(x)'M(\xi)^+ C(C'M(\xi)^+ C)^{-1} C'M(\xi)^+ f(x). \]

Then for every feasible design $\xi^* \in E(C)$ with nonsingular information matrix $M(\xi^*)$ the characterizations (1) through (4) are equivalent:

1. (D-Optimality) $\xi^*$ minimizes $\det C'M(\xi)^+ C$ among all $\xi \in E(C)$;

2. (G-Optimality) $\xi^*$ minimizes $\sup_{x \in \mathcal{L}} d(x, \xi)$ among all $\xi \in E(C)$, and

3. there exists a sequence $(\xi_1)_{\xi \in \mathcal{N}} \in E(C)^N$ of feasible designs with nonsingular information matrices such that
\[ \sup_{\xi \in \mathbb{N}} \| M(\xi) \|^{-1} \| < \infty \text{ and such that } \det C'M(\xi)^{-1} C \text{ converges to } \inf_{\xi \in \mathbb{N}} \det C'M(\xi)^+ C; \]

(3) (Normality) \( \sup_{x \in X} d(x, x^*) = s; \)

(4) (Duality) there exists a nonnegative-definite \( p \times p \) matrix \( N \) such that \( C'NC = C'M(\xi^*)^{-1} C \) and \( f(x)'Nf(x) \leq s \) for all \( x \in X. \)

Proof. (1) \( \Rightarrow \) (4), whenever \( \xi^* \) is feasible, by Theorems 5.4 and 5.5.

(3) \( \Rightarrow \) (1), whenever \( \xi^* \) is feasible, by Theorem 5.6(a).

(1) \( \Rightarrow \) (3), whenever \( \xi^* \) is feasible and \( M(\xi^*) \) nonsingular, by Theorem 5.6(b).

(2) \( \Rightarrow \) (1), whenever \( \xi^* \) is feasible, by Theorem 5.6, parts (e) and (a).

(1) \( \Rightarrow \) (2), whenever \( \xi^* \) is feasible and \( M(\xi^*) \) is nonsingular; (1), via (3), implies G-optimality. Condition (?) is satisfied by the constant sequence \( \xi_1 = \xi^*. \)

The name "normality" for condition (3) points towards an interconnection with the general theory of convex analysis that we found surprising and reassuring. The primal program (DOA) may equivalently be expressed as follows: Minimize the convex function (5.37)

\[ h(u, A) := \begin{cases} 
\ln \det C' A^+ C + s(u - 1) & \text{for } (u, A) \in \mathbb{R} \times \mathbb{G} \\
+ \infty & \text{for } A \notin \mathbb{G} 
\end{cases} \]

over the cone \( K = \{(\xi, M(\xi)) | \xi \geq 0\}. \) If \( u \in \mathbb{R} \) and \( A \in \text{int } \mathbb{G} = \text{PD}(p), \) then \( h \) is differentiable having gradient
(5.38) \[ \nabla h(u, A) = (s, -A^{-1} C(C^{'A^{-1} C}^{-1} C^{'A^{-1}}) \]; 

for a general \( A \in \mathcal{G} \) we can still exhibit a subgradient at \((u, A)\):

(5.39) \[ (s, -A^{+} C(C^{'A^{+} C}^{-1} C^{'A^{+}}) \in \partial h(u, A) \]; 

the defining subgradient inequality (Rockafellar 1970, p. 214) may be verified with the limiting argument following (5.17).

Now a major theorem of convex analysis (Rockafellar 1970, Thm. 27.4, p. 270) asserts the following. In order that 
\((u, A) \in \mathbb{R} \times \mathcal{PD}(p)\) be a point where the infimum of \( h \) relative to \( K \) is attained, it is necessary and sufficient that \(-\nabla h(u, A)\) is normal to \( K \) at \((u, A)\). In order that a general \((u, A) \in \mathbb{R} \times \mathcal{G}\) be optimal, it is still sufficient that \((-s, A^{+} C(C^{'A^{+} C}^{-1} C^{'A^{+}})\) is normal to \( K \) at \((u, A)\). The latter requirement is the same as condition (3), by definition of normality, and using the structure of \( K \).

**Comments and Previous Work.** What is wanted most in a practical situation is a good discrete design, rather than an optimal design measure. Approximations are most often based on the Equivalence Theorem, and in view of computational stability the restriction to feasible designs with nonsingular information matrices, as above, cuts off cases that are anyhow of limited interest; see Fedorov (1972) and Wynn (1972). Condition (?), however, is very disagreeable indeed. It would vanish if (5.28) could be established under weaker assumptions, or if existence of a D-optimal design measure with nonsingular information matrix could generally be ascertained.
Existence statements for optimal solutions of (DOA) might be delicate to obtain because of the ambiguous character of the boundary of $G$. On the other hand, $G$ cannot be omitted from the constraint list of (DOA), as shows the following.

Counterexample 5.8. If $C' = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $f(L) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, then $M(\Xi) = \left\{ \begin{pmatrix} 1 & -\alpha \\ -\alpha & \alpha \end{pmatrix} \middle| \alpha \in [0,1] \right\}$. One has $\inf \{ C'A^+C \middle| A \in M(\Xi) \} = 0$, although (DO) and (ME) share the optimal value 3, with optimal solutions $A^* = \frac{1}{3} \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, and $N^* = \frac{1}{2} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

The Equivalence Theorem originates with Kiefer & Wolfowitz (1960) who were concerned with estimating the full parameter $b$. The case $C' = [I_8 : 0]$ was adjoined by Kiefer (1961), rederived with game theoretic methods by Karlin & Studden (1966), and further complemented by Atwood (1969).

The minimal ellipsoid problem was introduced by Silvey, and Sibson readily proved duality of (DO) and (ME), see the discussion after Wynn (1972, pp. 174, 181). In a second contribution, Sibson (1974) established duality for arbitrary $C'$ by showing that the primal (DO) is a dual program of the dual (ME). Silvey & Titterington (1973) propose a thinnest cylinder problem as an alternate dual program.

The papers mentioned above derive their technical results by extensive use of partitioned matrices. This is felt the main reason why the basic inequality string (5.15) was not recognized, neither in its existence nor its usefulness to obtain the equality (5.14), although the missing argument is one of the most celebrated results in mathematical history: the Pythagorean Theorem.
Appendix

In the Appendix we elaborate remark 2(b), the existence of Neyman-Pearson tests, and the programs for maximin tests.

Lemma 2.3. Let $U$ be a real locally convex topological vector-space with a closed order cone $K$. Let $Y$ be the space of all continuous linear functionals $y : U \to \mathbb{R}$. Then

$$\inf \sup_{x \in C \setminus \{0\}} f(x) = \inf_{x \in C} f(x) - \langle r(x), y \rangle.$$  

Proof. Put $F(x, y) := f(x) - \langle r(x), y \rangle$. If $r(x) \not\in 0$, then

$$\sup_{y \not\in 0} F(x, y) = f(x).$$  

Assume now $r(x) \in 0$. Since $\{r(x)\}$ is compact and $K$ is closed, the set $S := \{y \in Y / \langle r(x), y \rangle \leq \inf_{u \in K} \langle u, y \rangle\}$ of separating functionals is nonempty (Dunford-Schwarz 1964, Thm. V.2.9, p. 417). But $\inf_{u \in K} \langle u, y \rangle \leq 0$ because of $0 \in K$, and $0 \leq \inf_{u \in K} \langle u, y \rangle$ whenever $y$ is bounded from below on the cone $K$. Hence $\emptyset \neq S = \{y \not\in 0 / \langle r(x), y \rangle < 0\}$, and

$$\sup_{y \not\in 0} F(x, y) = \begin{cases} f(x) & \text{for } r(x) \not\in 0 \\ +\infty & \text{for } r(x) \in 0 \end{cases},$$

which implies the assertion. □

It follows from the proof that both infima, when their common value is $< +\infty$, are attained by the same $x$'s.

Section 3. The following is a sufficiency version of the Neyman-Pearson Lemma 3.3. Define the set

$$Q := \{(E_0 \phi)_{\theta \in \mathcal{H}} / \phi \in \phi\},$$

33
obtained from restricting the power functions to \( H \). Assume that the objective function in \((\text{NPT})\) attains more than one value, i.e., the minimization problem is nontrivial.

Theorem 3.4. Assume there exists a nonnegative \( \sigma \)-finite measure \( \chi \) on \( (H, \mu) \) such that \( \alpha \) is in the interior of \( Q \) in the \( L_1 \)-topology of \( L_\infty(H, \mu, \chi) \).

Let \( \phi^* \) be a feasible solution of \((\text{NPT})\). Then \( \phi^* \) is an optimal solution of \((\text{NPT})\) if and only if there exists a signed measure \( \lambda^* \) on \( (H, \mu) \) such that \( \phi^* \) has the zero-one-form of Theorem 3.3.

Proof. The \( L_1 \)-topology of \( L_\infty \) is obtained by identifying \( L_\infty \) and \( L_1^* \) (Dunford-Schwarz 1964, Thm. IV. 8.5, p. 289, and p. 420). In the topological space \( L_\infty \times \mathbb{R} \), the set

\[
\mathcal{Q} := \{ (E_\theta \phi) \in \mathbb{H}^+ \mid \theta \in \mathcal{H}, \phi \in \phi \}
\]

of power functions is convex and its interior is not empty. The latter follows since \( \text{int} \ Q \neq \emptyset \), and the set \( \{ E_\theta \phi \in \mathcal{Q} \mid \theta \in \mathcal{H}, \phi \in \phi \} \) is an interval not consisting of one point only.

If \( -\beta \) is the optimal value of \((\text{NPT})\), then \( (\alpha, \beta) \) lies on the boundary of \( \mathcal{Q} \); hence there exists a closed hyperplane supporting \( \mathcal{Q} \) at \( (\alpha, \beta) \); i.e., there exists a continuous linear functional \( (h, k) : L_\infty \times \mathbb{R} \to \mathbb{R} \) such that

\[
\sup \{ h(\tilde{\alpha}) + k \cdot \tilde{\beta} \mid (\tilde{\alpha}, \tilde{\beta}) \in \mathcal{Q} \} = h(\alpha) + k \cdot \beta
\]

(Dunford-Schwarz 1964, Cor. V.9.6, p. 449; Thm. V.2.1, p. 413; Lemma
V.2.7, p. 417; or Valentine 1964, Thm. 2.15, p. 27). The hypothesis \( \alpha \in \text{int } Q \) entails \( k \neq 0 \). Since \( L_\infty \) bears the \( L_1 \)-topology, its continuous functionals are \( \tilde{\alpha} \rightarrow \int_H \tilde{\alpha} h(\theta) d\chi \), with \( h \in L_1(H, \mathcal{M}, \chi) \).

(Dunford-Schwarz 1964, Thm. V.3.9, p. 421). Now \( d\lambda^* := -\frac{1}{k} h(\theta) d\chi \) gives a well-defined signed measure, for which

\[
\sup \{ \beta - \int_H \tilde{\alpha} d\lambda^* \mid (\alpha, \beta) \in Q \} = \beta - \int_H \alpha d\lambda^* .
\]

Given an optimal \( \phi^* \) this means that

\[
\sup_{\phi \in \Phi} \int \phi(x)(\rho_\theta(x) - \int_H \rho_\theta(x) d\lambda^*) d\mu
\]

\[
= \int \phi^*(x)(\rho_\theta(x) - \int_H \rho_\theta(x) d\lambda^*) d\mu ,
\]

from which the direct part of the assertion follows. The converse part is Lemma 3.3. \( \square \)

For an account of a finite hypothesis \( H \), see Dantzig & Wald (1951), or Witting (1966, Satz 2.35).

The generalization of Lemma 3.1 to maximin level \( \alpha \) tests for a measurable hypothesis \( (H, \mathcal{M}) \) versus a (composite) measurable alternative \( (K, \mathcal{K}) \) is obtained using the primal program

(MmT) Maximize \( \inf_{\theta \in K} E_\theta \phi \) subject to \( \phi \in \Phi \) and \( E_\theta \phi \leq \alpha \) for \( \theta \in H \),

and the dual candidate
Minimize $\alpha c + \int_X d_+(x) \, d\mu$ subject to $c \geq 0$,

\[
\begin{align*}
\lambda &\in ca_+(H,\mathcal{M}) \text{ and } \lambda(H) = 1, \text{ and} \\
\nu &\in ca_+(K,\mathcal{M}) \text{ and } \nu(K) = 1;
\end{align*}
\]

where $d$ is given by $d(x) = \int_K p_0(x) \, d\nu - \int_H p_0(x) \, d\lambda$. See Krafft & Witting (1967).

References in the Appendix


Acknowledgements

I wish to thank Professor H. Witting, Freiburg im Breisgau, whose guidance and encouragement was vital in initiating this work. Professor O. Krafft, Aachen, pointed out to me that Sibson's paper (1974) includes an appealing further example for programming methods in statistics. Most of the research was carried out while I was on a postdoctoral fellowship of the Deutsche Forschungsgemeinschaft, Bonn, with the Department of Statistics, Stanford University.

References


