A-OPTIMALITY FOR REGRESSION DESIGNS,

BY

N. N. CHAN

TECHNICAL REPORT NO. 49
JAN 1981

PREPARED UNDER CONTRACT N00014-75-C-0442 (NR-042-034)
OFFICE OF NAVAL RESEARCH
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A-OPTIMALITY FOR REGRESSION DESIGNS

by

N. N. CHAN

The Chinese University of Hong Kong

TECHNICAL REPORT NO. 49

JANUARY 1981

PREPARED UNDER CONTRACT NOOO-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government. Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Also issued as Technical report No. 163 under National Science Foundation Grant MCS 78-07736 – Dept. of Statistics, Stanford University.
1. Introduction.

Consider the linear regression model

\[ y = X\beta + \varepsilon, \]

where \( y \) is an \( m \times 1 \) vector of observations, \( X \) is an \( m \times n \) matrix to be called the design matrix, \( \beta \) is an \( n \times 1 \) vector of unknown parameters, and \( \varepsilon \) is an \( m \times 1 \) vector of random variables with mean the \( m \times 1 \) zero vector and known covariance matrix \( \Lambda \). We assume that \( m \geq n \) and denote the eigenvalues of \( \Lambda \) in ascending order of magnitude by

\[ \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \cdots \leq \lambda_m. \]

For later use denote the diagonal matrices with diagonal elements \( \lambda_1, \ldots, \lambda_i \) by \( \Lambda_i \), \( i = n \) and \( m \).

For a given design matrix \( X \) of rank \( n \), an unbiased estimate of the parameter \( \beta \) based on the observation \( y \) is the simple least squares estimate

\[ (X'X)^{-1}X'y, \]

whose covariance matrix is given by

\[ (X'X)^{-1}X'\Lambda X(X'X)^{-1}. \]

(1)
One of the design problems is to choose $X$ from a given experimental region such that the trace of the matrix in (1) is minimal. This is a problem in the A-optimal designs of regression experiments and was considered by Dorogovcev (1971) under the more general setting that the observations are the realization of stochastic processes. Earlier work on A-optimal designs was given by Elfving (1952) and Chernoff (1953).

In this paper the experimental region under consideration is taken to be the set $H$ of all $m \times n$ real matrices of rank $n$ whose $i^{th}$ column has a Euclidean norm not exceeding $c_i$, $i = 1, \ldots, n$, where the $c_i$ are given positive numbers. In section 2, it is shown that for any matrix $X$ in $H$ the trace of the matrix in (1) has as a lower bound of

$$\left( \sum_{i=1}^{n} c_i^2 \right)^{-1} \left( \sum_{i=1}^{n} \lambda_i \right)^2 .$$

In section 3, a necessary and sufficient condition for the existence of an $X$ in $H$ to attain the lower bound is derived. For the case in which all the $c_i$ are equal, a partial result was given in Chan and Wong (1981). Dorogovcev (1971) obtained the lower bound for the special case $n = 2$ and $c_1 = c_2$.

It is worth noting that in the regression model if one considers the best linear unbiased estimate $(X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} y$ and its covariance matrix $(X' \Lambda^{-1} X)^{-1}$, by minimizing the trace of the latter for all $X$ in $H$, the corresponding optimal design problem has a simple solution, as is given in Rao (1973, p. 236). On the other hand, if one wishes to minimize the determinant of $(X' \Lambda^{-1} X)^{-1}$, there is the so called D-optimal design.
problem, of which comprehensive reviews can be found in St. John and Draper (1975) and Kiefer and Galil (1980).

2. An Inequality.

For the regression model and the set H as given in section 1, we note that in minimizing the trace of the matrix in (1) with respect to X in H, the matrix \( \Lambda \) in (1) can be replaced by the diagonal matrix \( \Lambda_m \) without loss of generality, in view of the existence of an orthogonal matrix P such that

\[
\Lambda = P\Lambda_m P^T
\]

and the following equality

\[
(X'X)^{-1}X'AX(X'X)^{-1} = (Y'Y)^{-1}Y'\Lambda_m Y(Y'Y)^{-1}
\]

where \( Y = PX \) which is again in H. The following lemma of Fan (1949) will be required in the proof of our main inequality.

Lemma 1. Let \( B \) be a real \( m \times n \) matrix whose \( n \) columns form an orthonormal set. Then

\[
\text{tr } B'\Lambda B \geq \text{tr } \Lambda_n
\]

where \( \text{tr} \) represents the trace operation.

Theorem 1. For any X in H,

\[
\text{tr}((X'X)^{-1}X'AX(X'X)^{-1}) \geq \left( \sum_{i=1}^{n} c_i^2 \right)^{-1} \left( \sum_{i=1}^{n} \lambda_i \right)^{\frac{1}{2}}^2
\]

3
Proof. By the Cauchy-Schwarz inequality applied to the trace inner product $\text{tr}(X'Y)$ between two real $m \times n$ matrices $X$ and $Y$, we have

$$\text{(2)} \quad \text{tr}(X'X) \times \text{tr}((X'X)^{-1}X'A X(X'X)^{-1}) \geq \text{tr}^2(X'A X(X'X)^{-1})$$

But the trace on the right-hand side is

$$\text{(3)} \quad \text{tr}((X'X)^{-1}X'A X(X'X)^{-1}),$$

which is not less than $\sum_{i=1}^{n} \lambda_i^2$ by Lemma 1 on noting that the $n$ columns of the matrix $X(X'X)^{-1}$ are orthonormal. By the definition of the set $H$,

$$\text{(4)} \quad \text{tr}(X'X) \leq \sum_{i=1}^{n} c_i^2,$$

Hence the main inequality follows.


The main result of this work is to obtain a necessary and sufficient condition on $(\lambda_1, \ldots, \lambda_m)$ and $(c_1, \ldots, c_n)$ for the existence of a matrix in $H$ such that the lower bound in Theorem 1 is attained. For this we need the following lemmas.

Lemma 2. Let $D$ be an $n \times n$ real diagonal matrix with diagonal elements $d_1 \leq d_2 \leq \cdots \leq d_n$, and $a_1, \ldots, a_n$ be $n$ real numbers such that $a_1 \leq a_2 \leq \cdots \leq a_n$ and
Then there exists an $n \times n$ orthogonal matrix $P$ such that the $n$ diagonal elements of $P'DP$ are $a_1, \ldots, a_n$ if and only if

$$\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} d_i, \quad k = 1, 2, \ldots, n-1.$$ 

This lemma is a version of a result by Horn (1954) and a proof is given by Mirsky (1958). See also Marshall and Olkin (1979, p. 220).

**Lemma 3.** Let $D$ be as in Lemma 2 and $B$ be an $n \times k$ matrix whose $k$ columns form an orthonormal set. Arrange the eigenvalues of the $k \times k$ matrix $B'DB$ in ascending order $b_1 \leq b_2 \leq \cdots \leq b_k$. Then $b_i \geq d_i$, $i = 1, \ldots, k$.

This is the Poincaré separation theorem and can be found for example in Rao (1973, p. 64).

**Theorem 2.** Suppose that the positive numbers $c_i$, $i = 1, \ldots, n$, are arranged in ascending order of magnitude and that the smallest eigenvalue $\lambda_1$ of the covariance matrix $\Lambda$ is positive. Then there is an $X$ in $H$ such that

$$\text{tr}((X'X)^{-1}X'\Lambda X(X'X)^{-1}) = \left( \sum_{i=1}^{n} c_i^2 \right)^{-1} \left( \sum_{i=1}^{n} \lambda_i^2 \right)^2,$$

if and only if

$$\sum_{i=1}^{n} c_i^2 = \sum_{i=1}^{n} \lambda_i^2.$$
\[
\left( \sum_{i=1}^{n} c_i^2 \right)^{-1} \sum_{i=1}^{k} c_i^2 \geq \left( \sum_{i=1}^{n} \lambda_i^k \right)^{-1} \sum_{i=1}^{k} \lambda_i^k, \quad k = 1, \ldots, n-1.
\]

**Proof.** Sufficiency: Consider the diagonal matrix \( \Lambda_n \) whose diagonal elements are \( \lambda_i^k \), \( i = 1, \ldots, n \). By Lemma 2 there exists an orthogonal matrix \( P \) of order \( n \) such that the \( i \)th diagonal element of \( P' \Lambda_n \) is \( b \), \( i = 1, \ldots, n \), where

\[
b = \left( \sum_{i=1}^{n} c_i^2 \right)^{-1} \left( \sum_{i=1}^{n} \lambda_i^k \right).
\]

Denote by \( X \) the \( m \times n \) matrix

\[
b^{-k} \begin{bmatrix} \Lambda_n^k & 0 \\ 0 & 0 \end{bmatrix},
\]

where \( 0 \) is an \( (m-n) \times n \) submatrix of zeros. Note that \( X \) is of rank \( n \) as \( \lambda_1 > 0 \) and that the \( i \)th diagonal element of \( X'X \) equals \( c_i^2 \) as we have

\[
X'X = b^{-1} P' \Lambda_n^k P.
\]

Hence \( X \) is a member of the set \( H \). Moreover, for the diagonal matrix \( \Lambda_m \) of order \( m \), we have

\[
X' \Lambda_m X = b^{-1} P' \Lambda_n^k \begin{bmatrix} 0' \end{bmatrix} \Lambda_m \begin{bmatrix} \Lambda_n^k \\ 0 \end{bmatrix} = b^{-1} P' \Lambda_n^k \Lambda_n^k P = b^{-1} P' \Lambda_n^k P.
\]


and so

\[ \text{tr}((X'X)^{-1}X'A X(X'X)^{-1}) = \text{tr}(b(P'A_n^{-1}p)P'A_n^{1/2}P(P'A_n^{-1}p)) \]

\[ = b \text{tr}(P'A_n^{1/2}p) \]

\[ = \left( \sum_{i=1}^{n} c_i^2 \right)^{-1} \left( \sum_{i=1}^{n} \lambda_i^b \right)^2. \]

The proof for sufficiency is completed by replacing \( A \) by \( A \) as remarked at the beginning of Section 2.

**Necessity.** Suppose that \( X, \) a member of \( H, \) is such that the inequality in Theorem 1 becomes an equality. Then the three inequalities in the proof of Theorem 1 reduce to equalities. First, note that the \( i^{th} \) diagonal element of the matrix \( X'X \) equals \( c_i^2, i = 1, \ldots, n, \) because it cannot exceed \( c_i^2 \) (as \( X \) is in \( H \)) and from (4)

\[ \text{tr}(X'X) = \sum_{i=1}^{n} c_i^2. \]

By Lemma 2, it is then enough to show that \( \lambda_1^b, \ldots, \lambda_n^b \) are the eigenvalues of the \( n \times n \) matrix \( bX'X. \) For this, note that the Cauchy-Schwarz inequality (2) becoming an equality implies that there is a nonzero real number \( d \) such that

\[ X = d\lambda^b X(X'X)^{-1}. \]

So we have

\[ X'X = dX'A^b X(X'X)^{-1}. \]
The equality corresponding to (3) then implies that

\[ \sum_{i=1}^{n} \lambda_i^{k_i} = \text{tr}(X'\Lambda X(X'X)^{-1}) \]

(5)

\[ = d^{-1}\text{tr}(X'X) \]

\[ = d^{-1} \sum_{i=1}^{n} c_i^2. \]

Therefore, \( d = b^{-1} \), and so

\[ bX'X = X'\Lambda X(X'X)^{-1}. \]

It remains to show that the \( n \times n \) matrix

(6)

\[ (X'X)^{-1}X'\Lambda^{1/2}X(X'X)^{-1/2} \]

has \( \lambda_1^{1/2}, \ldots, \lambda_n^{1/2} \) as its eigenvalues. In fact, by replacing \( \Lambda \) by \( \Lambda_m \) and using Lemma 3, we see that the \( i \)th smallest eigenvalue of the matrix in (6) is not less than \( \lambda_i^{1/2}, i = 1, \ldots, n \), and, in view of the first equality in (5), must be equal to \( \lambda_1^{1/2} \), completing the proof.

**Acknowledgement.** I am thankful to Professor T. W. Anderson for helpful discussions and suggestions.
References


TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations,"

18. "On the Interrelationships among Sufficiency, Total Sufficiency and


Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models

    Raul Pedro Mentz, September 8, 1975.

22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting,"
    Mituaki Huzii, February 1976.

23. "Estimating Means when Some Observations are Classified by Linear
    Discriminant Function," Chien-Pai Han, April 1976.


    September 1976.


28. "Estimation of the Parameters of Finite Location and Scale Mixtures,"
    Javad Behboodian, October 1976.

29. "Identification of Parameters by the Distribution of a Maximum

30. "Discrimination Between Stationary Gaussian Processes, Large Sample

31. "Principal Components in the Nonnormal Case: The Test for Sphericity,"
    Christine M. Waternaux, October 1977.

32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a
Use is made of a result of Horn (1954) on the existence of a symmetric matrix with prescribed diagonal elements and eigenvalues. A necessary and sufficient condition is then given for the existence of an A-optimal design for a regression experiment in the Dorogovcev (1971) setting.