Statistica Sinica **8**(1998), 535-542

E-OPTIMALITY FOR REGRESSION DESIGNS UNDER CORRELATIONS

Kim-Hung Li, Nai N. Chan and Koon-Sun Wong

Chinese University of Hong Kong

Abstract: In a regression setting, the greatest lower bound for the largest eigenvalue of the covariance matrix of the generalized least squares estimator when the experimental errors are correlated is derived under the experimental region considered by Chan and Li (1989). A neat and efficient algorithm for constructing an E-optimal design matrix via a CL vector is then achieved. It is also shown that for the E-optimal design matrix the generalized least squares and the ordinary least squares estimators are identical, and thus the two estimators have the same E-optimal design matrix.

Key words and phrases: A-optimal design, CL vector, correlated error, E-optimal design, generalized least squares, majorization, ordinary least squares.

1. Introduction

Consider the linear regression model

$$y = X\beta + \epsilon,$$

where y is an $n \times 1$ vector of observations, X is an $n \times p$ $(n \geq p)$ real matrix to be called the design matrix, β is a $p \times 1$ vector of unknown parameters, and ϵ is an $n \times 1$ random vector with mean the $n \times 1$ zero vector and covariance matrix $\sigma^2\Lambda$, where σ is an unknown parameter and Λ is a known $n \times n$ positive definite matrix. Denote the transpose of X by X'. For a given design matrix X of rank p, the best linear unbiased estimator of the parameter β based on the observation y is the generalized least squares estimator $(X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}y$, whose covariance matrix is given by $\sigma^2\Sigma$, where

$$\Sigma = (X'\Lambda^{-1}X)^{-1}. \tag{1}$$

Another unbiased estimator of the parameter β is the ordinary least squares estimator $(X'X)^{-1}X'y$. Bhaumik (1995) gave reasons for using ordinary least squares rather than generalized least squares when the experimental errors in ϵ are correlated. In optimal design problems, different criteria are suggested in the choice of a design matrix in a prescribed experimental region $\mathbf{H}$, say. For a given estimator of β, a design matrix $Z \in \mathbf{H}$ is said to be E-optimal in $\mathbf{H}$ if the largest eigenvalue of the covariance matrix of the estimator with respect to all design matrices $X \in \mathbf{H}$ is minimized at $X = Z$. A design matrix Z is said to be A-optimal in $\mathbf{H}$ if the trace of the covariance matrix is minimized at $X = Z$.

The experimental region $\mathbf{H}$ considered in this paper is the set of all $n \times p$ real matrices of rank p whose ith column has a Euclidean norm not exceeding $c_i, i = 1, \ldots, p$, where c_i's are given positive numbers. Restrictions to column norms (rather than row norms) in a reproducing kernel inner product space formulation have long been considered in the time series literature (see Parzen (1961) for a comprehensive review). In the context of optimal regression designs, Rao (1965), p.193 considered a design region in terms of the column norms for an n-dimensional inner product (with kernel Λ^{-1}) space; here we prefer to use the region $\mathbf{H}$ which does not depend on the matrix Λ.

The choice of our experimental region $\mathbf{H}$ also has a close relation with the imposition of energy constraints on the input signals for identifying parameters in linear systems (Levadi (1966), Mehra (1974a, 1974b) and Kalaba and Spingarn (1982)). Pázman (1986), chapter 7 considered the applications of optimal regression designs to the physical fields such as the gravitational field, the electrostatic field, etc. The column norm restriction was also adopted in Chang and Wong (1981) for the optimal control of a regression experiment. Specifically we have the following:

Example. Dorogovcev (1971) (see also Chang (1979) and Chang and Wong (1981)) considered the model :

$$z(t) = \sum_{i=1}^{p} \beta_i f_i(t) + \xi(t), \tag{2}$$

where $z(t)$ is a continuous time output process, $f_i(t), i = 1, \ldots, p$, are p input functions, β_i's are unknown parameters, and $\xi(t)$ is a colored noise with zero mean. The output process $z(t)$ is to be observed from time a to b. The input functions, to be selected by the researcher, are subject to the energy constraints

$$\int_a^b f_i^2(t)dt \leq c_i^2, \quad i = 1, \ldots, p,$$

the left hand side being the energy content of the input signal $f_i(t)$. Suppose that each $f_i(t)$ is decomposable as

$$f_i(t) = \sum_{j=1}^{n} a_{ji} g_j(t), \tag{3}$$

where the a_{ji}'s are constants ($n \geq p$) to be specified, and $g_j(t), j = 1, \ldots, n$, are given continuous and orthonormal functions on the interval $[a, b]$. Write

$y_j = \int_a^b z(t)g_j(t)dt$, $j = 1, \ldots, n$, $y = [y_1, \ldots, y_n]', \beta = [\beta_1, \ldots, \beta_p]'$, and $A = [a_{ji}]$ which is an $n \times p$ matrix. Consider the model

$$y = A\beta + \zeta, \tag{4}$$

where $\zeta = [\zeta_1, \ldots, \zeta_n]'$ with $\zeta_j = \int_a^b \xi(t)g_j(t)dt$, $j = 1, \ldots, n$. It is easy to see that under the conditions in (3) the least squares estimator of β for the model in (2) is tantamount to that for the model in (4). The energy constraint is then equivalent to the ith column norm of the design matrix A in (4) not exceeding c_i, $i = 1, \ldots, p$.

A concise construction method of an A-optimal design in **H** for the ordinary least squares estimator has been suggested (Chan and Li (1989) and Li and Chan (1989)). A key part in the construction algorithm is the computation of a CL vector. Li and Chan (1989) introduced an algorithm that can produce a CL vector in a finite number of steps. In Section 2 we provide a simpler and more efficient algorithm for finding the CL vector. Based on the new algorithm, a method of constructing an E-optimal design matrix in **H** for the generalized least squares estimator is proposed in Section 3. The same design matrix is also E-optimal in **H** for the ordinary least squares estimator. The construction algorithm is illustrated by an example in Section 4.

2. CL Vector

In studying A-optimality for the ordinary least squares estimator in the design region **H** as given in Section 1, Chan and Li (1989) defined a CL sequence, which was also referred to as CL vector in Li and Chan (1989). Given two ordered vectors of dimension p, say $a_1 \leq \cdots \leq a_p$ and $b_1 \leq \cdots \leq b_p$, the vector $[a_i]$ is said to majorize $[b_i]$, written as $[a_i] \succ [b_i]$, if $\sum_{j=1}^i a_j \leq \sum_{j=1}^i b_j, i = 1, \ldots, p-1$, and $\sum_{j=1}^p a_j = \sum_{j=1}^p b_j$ (see, for example, Marshall and Olkin (1979), p.5). Denote

$$D = \{[d_i] : 0 \leq d_1 \leq \cdots \leq d_p\},$$

and D_+ to be the subset of D with $d_1 > 0$.

Definition. Given $[a_i], [b_i] \in D_+$, there exists a unique vector $[d_i]$ such that
(a) $[d_i] \in D_+$ and $[d_i] \succ [b_i]$;
(b) $d_i/a_i \leq d_{i+1}/a_{i+1}, i = 1, \ldots, p-1$;
(c) if $d_i/a_i < d_{i+1}/a_{i+1}$, then $\sum_{j=1}^i d_j = \sum_{j=1}^i b_j$.
This unique vector $[d_i]$ is called the CL vector of the pair $([a_i], [b_i])$.

The existence and uniqueness of a CL vector are proved in Chan and Li (1989). There it is also shown that

$$\min\Big\{\sum_{i=1}^p t_i^{-1}a_i^2 : [t_i] \in D, [t_i] \succ [b_i]\Big\} = \sum_{i=1}^p d_i^{-1}a_i^2.$$

Li and Chan (1989) proposed an algorithm for finding a CL vector. In what follows, we introduce an alternative approach which leads to a simpler and yet more efficient algorithm for constructing the CL vector.

Theorem 1. *Given* $[a_i]$ *and* $[b_i]$ *in* D_+, *let*

$$h = \max\Big\{\sum_{j=1}^{i} a_j / \sum_{j=1}^{i} b_j : i = 1, \ldots, p\Big\}.$$

Suppose k $(1 \le k \le p)$ *is an integer such that* $\sum_{j=1}^{k} a_j / \sum_{j=1}^{k} b_j = h$. *Write* $d_i = a_i/h$, $i = 1, \ldots, k$, *and if* $k < p$, *let* $[d_{k+1}, \ldots, d_p]$ *be the CL vector of* $([a_{k+1}, \ldots, a_p], [b_{k+1}, \ldots, b_p])$. *Then* $[d_1, \ldots, d_p]$ *is the CL vector of* $([a_i], [b_i])$.

Proof. The vector $[d_i]$ is shown to satisfy the conditions listed in the Definition above as follows :

(a) We show that

$$d_i/a_i \le d_{i+1}/a_{i+1}, \quad i = 1, \ldots, p-1. \tag{5}$$

As $d_i/a_i = 1/h$ for $i = 1, \ldots, k$, the inequalities in (5) hold for $i = 1, \ldots, k-1$. The proof of (5) is trivial if $k = p$. Suppose that $k < p$. The inequalities in (5) hold for $i > k$ as $[d_{k+1}, \ldots, d_p]$ is a CL vector. For $i = k$, let v be the smallest integer greater than k such that $\sum_{j=k+1}^{v} d_j = \sum_{j=k+1}^{v} b_j$. From the definition of a CL vector, $d_i/a_i = d_{i+1}/a_{i+1}$ for $k+1 \le i < v$. Therefore,

$$d_{k+1}/a_{k+1} = \sum_{j=k+1}^{v} d_j / \sum_{j=k+1}^{v} a_j = \sum_{j=k+1}^{v} b_j / \sum_{j=k+1}^{v} a_j;$$

from the definitions of h and k, the last quotient is not less than $1/h$, which is equal to d_k/a_k.

(b) We show that if $d_i/a_i < d_{i+1}/a_{i+1}$, then $\sum_{j=1}^{i} d_j = \sum_{j=1}^{i} b_j$. Clearly the statement holds for $i > k$. It also holds for $i \le k$ because for $i = 1, \ldots, k-1$, we always have $d_i/a_i = d_{i+1}/a_{i+1}$, and for $i = k$, we necessarily have $\sum_{j=1}^{k} d_j = \sum_{j=1}^{k} b_j$.

(c) The inequalities in (5) imply that $d_1 \le \cdots \le d_p$. It remains to show that $[d_i] \succ [b_i]$. Clearly $\sum_{j=1}^{p} d_j = \sum_{j=1}^{p} b_j$. From the definitions of h and k, $\sum_{j=1}^{i} d_j \le \sum_{j=1}^{i} b_j$ holds for $i \le k$. The inequality also holds for $i > k$ because of the fact that $\sum_{j=1}^{k} d_j = \sum_{j=1}^{k} b_j$, and the choice of $[d_{k+1}, \ldots, d_p]$.

3. An E-optimal Design Matrix

Write the $p \times p$ diagonal matrix with the ith diagonal element a_i, $i = 1, \ldots, p$, as $\mathrm{diag}[a_1, \ldots, a_p]$. Without loss of generality, assume that $c_1 \le \cdots \le c_p$. Let

$0 < \lambda_1 \le \cdots \le \lambda_p \le \cdots \le \lambda_n$ be the eigenvalues of the matrix Λ arranged in ascending order of magnitude, and P be an $n \times p$ real matrix whose columns are the orthonormal eigenvectors of Λ (implying that $P'P$ is the $p \times p$ identity matrix) with $\Lambda P = P \operatorname{diag}[\lambda_1, \ldots, \lambda_p]$. Let $[d_1, \ldots, d_p]$ be the CL vector of $([\lambda_1, \ldots, \lambda_p], [c_i^2])$. As $[d_i] \succ [c_i^2]$, we may construct a $p \times p$ orthogonal matrix Q such that the ith diagonal element

$$(Q \ \operatorname{diag}[d_1, \ldots, d_p] \ Q')_{ii} = c_i^2, \quad i = 1, \ldots, p, \tag{6}$$

using the algorithm given in Chan and Li (1983). Define

$$Z = P \ \operatorname{diag}[d_1^{1/2}, \ldots, d_p^{1/2}] \ Q'. \tag{7}$$

Theorem 2. *In the region* $\mathbf{H}$ *as given in Section* 1*, the design matrix* Z *in* (7) *is E-optimal for the generalized least squares estimator. The largest eigenvalue of* $(Z'\Lambda^{-1}Z)^{-1}$ *is simply*

$$\mu = \max\Big\{ \sum_{j=1}^{i} \lambda_j / \sum_{j=1}^{i} c_j^2 : i = 1, \ldots, p \Big\}.$$

The proof consists of two parts.

PART (a). The largest eigenvalue of the matrix Σ in (1) for any $X \in \mathbf{H}$ is greater than or equal to μ.

Proof. By the singular value decomposition of an $n \times p$ real matrix $X \in \mathbf{H}$ of rank p, we may write $X = ARB'$, where A is an $n \times p$ real matrix with orthonormal columns, $R = \operatorname{diag}[r_1, \ldots, r_p]$ with $0 < r_1 \le \cdots \le r_p$ (and so $[r_i] \in D_+$), and B is a $p \times p$ orthogonal matrix. Since $X'X = BR^2B'$ and the norm of the ith column of X does not exceed c_i, $i = 1, \ldots, p$, it can be shown that $[r_i^2]$ upper weakly majorizes $[c_i^2]$ in the sense that $r_1^2 + \cdots + r_i^2 \le c_1^2 + \cdots + c_i^2$, $i = 1, \ldots, p$. The matrix Σ in (1) becomes $(X'\Lambda^{-1}X)^{-1} = BR^{-1}(A'\Lambda^{-1}A)^{-1}R^{-1}B'$. Thus the eigenvalues of Σ are also the eigenvalues of $R^{-1}(A'\Lambda^{-1}A)^{-1}R^{-1}$. From Theorem 3 of Wang and Zhang (1992), the largest eigenvalue of $R^{-1}(A'\Lambda^{-1}A)^{-1}R^{-1}$ is greater than or equal to the ith smallest eigenvalue of $(A'\Lambda^{-1}A)^{-1}$ divided by r_i^2 for all $i = 1, \ldots, p$. By the Poincaré separation theorem, the ith smallest eigenvalue of $(A'\Lambda^{-1}A)^{-1}$, which is equal to the reciprocal of the ith largest eigenvalue of $A'\Lambda^{-1}A$, is greater than or equal to the ith smallest eigenvalue of Λ, which is λ_i. Therefore, for any $X \in \mathbf{H}$ with singular values r_i, $i = 1, \ldots, p$ (so that $[r_i^2]$ upper weakly majorizes $[c_i^2]$) we have

$$\begin{aligned} \text{largest eigenvalue of } \Sigma &\ge \max\{\lambda_i / r_i^2 : i = 1, \ldots, p\} \\ &\ge \max\Big\{ \sum_{j=1}^{i} \lambda_j / \sum_{j=1}^{i} r_j^2 : i = 1, \ldots, p \Big\} \end{aligned}$$

$$\geq \max\Big\{\sum_{j=1}^{i}\lambda_j/\sum_{j=1}^{i}c_j^2 : i=1,\ldots,p\Big\} = \mu;$$

the second inequality follows as

$$\sum_{j=1}^{i}\lambda_j/\sum_{j=1}^{i}r_j^2 \leq \max\{\lambda_j/r_j^2 \ : \ j=1,\ldots,i\}.$$

PART (b). The matrix Z constructed in (7) is a design matrix in $\mathbf{H}$ that attains the lower bound μ.

Proof. The matrix Z is in $\mathbf{H}$ in view of the equations in (6). Also,

$$\begin{aligned}
&\text{the largest eigenvalue of } (Z'\Lambda^{-1}Z)^{-1}\\
&= \text{the largest eigenvalue of } (\operatorname{diag}[d_1^{1/2},\ldots,d_p^{1/2}]P'\Lambda^{-1}P\operatorname{diag}[d_1^{1/2},\ldots,d_p^{1/2}])^{-1}\\
&= \max\{\lambda_i/d_i \ : \ i=1,\ldots,p\}.
\end{aligned}$$

As $[d_i]$ is the CL vector of $([\lambda_i],[c_i^2]), d_1/\lambda_1 \leq \cdots \leq d_p/\lambda_p$. Therefore, the largest eigenvalue of $(Z'\Lambda^{-1}Z)^{-1}$ is λ_1/d_1, which is equal to μ from Theorem 1 in Section 2.

It can be proved directly or through the use of Theorem 3.6 in Seber (1977), p.63 that the ordinary least squares and the generalized least squares estimators for the design matrix Z in (7) are identical. Therefore, Z in (7) is also E-optimal in $\mathbf{H}$ when the ordinary least squares estimator is used.

4. An Example

Suppose that an output discrete time process y_j is affected by two input signals x_{1j} and x_{2j} and an MA(1) disturbance ϵ_j such that

$$y_j = \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_j;$$

we aim at choosing an $n\times 2$ E-optimal design matrix $X=[x_{ji}]$ for the estimation of the parameters β_1 and β_2. The experimental region $\mathbf{H}$ is such that the norm of each column of X is less than or equal to $n^{1/2}$. In this case it can be shown that the inequality constraints may be replaced by the equalities that $\sum_{j=1}^{n} x_{ij}^2/n = 1$ for $i=1,2$. This restriction is similar to that used by Box and Jenkins (1970), pp. 416-420 in studying a case in which n tends to infinity and the stationary distribution of the input process has zero mean and unit variance.

Let the MA(1) process be $\epsilon_j = bu_{j-1} + u_j$, with u_j a Guassian white noise with constant variance σ^2. Then Λ is a tridiagonal matrix with diagonal elements equal to $(1+b^2)$, and sub-diagonal elements b. From Parlett (1980), p.130, the eigenvalues of Λ are $1+b^2+2b\cos(j\pi/(n+1))$, $j=1,\ldots,n$, with corresponding orthonormal eigenvectors

$$[2/(n+1)]^{1/2}[\sin(j\pi/(n+1)), \sin(2j\pi/(n+1)), \ldots, \sin(nj\pi/(n+1))].$$

Thus the two smallest eigenvalues of Λ are

$$\lambda_1 = 1+b^2+2b\cos\{[(n+1)I_{b\geq 0}-1]\pi/(n+1)\},$$

and

$$\lambda_2 = 1+b^2+2b\cos\{[(n+1)I_{b\geq 0}-2]\pi/(n+1)\},$$

where $I_{b\geq 0}$ is an indicator function taking the value 1 when $b \geq 0$ and 0 otherwise. Clearly the desired CL-vector as in (6) is $[d_1, d_2] = 2n[\lambda_1, \lambda_2]/(\lambda_1+\lambda_2)$, and we may choose

$$Q = 2^{-1/2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

The E-optimal design matrix in (7) is therefore

$$Z = \{2n/[(n+1)(\lambda_1+\lambda_2)]\}^{1/2}\ V\begin{bmatrix} \lambda_1^{1/2} & -\lambda_1^{1/2} \\ \lambda_2^{1/2} & \lambda_2^{1/2} \end{bmatrix},$$

where V is an $n\times 2$ matrix with the (j,i)th element equal to $\sin\{[(n+1)I_{b\geq 0} - i]j\pi/(n+1)\}$. This optimal design matrix depends on the MA(1) coefficient b. If b is not known except for its sign, and n is large, we may use, instead, the approximate optimal design matrix :

$$Z^* = [n/(n+1)]^{1/2}\ V\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

which has (componentwise) relative error of order $O(n^{-2})$.

Acknowledgement

The authors wish to thank the Chair-Editor and the referee for their helpful comments and suggestions.

References

Bhaumik, D. K. (1995). Majorization and D-optimality for complete block designs under correlations. *J. Roy. Statist. Soc. Ser. B* **57**, 139-143.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis Forecasting and Control.* Holden Day, San Francisco.

Chan, N. N. and Li, K.-H. (1983). Algorithm 115 : A Fortran subroutine for finding a real symmetric matrix with prescribed diagonal elements and eigenvalues. *Computer J.* **26**, 184-186.

Chan, N. N. and Li, K.-H. (1989). Majorization for A-optimal designs. *J. Math. Anal. Appl.* **142**, 101-107.

Chang, D.-S. (1979). Design of optimal control for a regression problem. *Ann. Statist.* **7**, 1078-1085.

Chang, D.-S. and Wong, C. S. (1981). A general approach to optimal control of a regression experiment. *J. Mult. Anal.* **11**, 85-101.

Dorogovcev, A. JA. (1971). Problems of optimal control of a regression experiment. *Selected Transl. Math. Statist. Probability* **10**, 35-41.

Kalaba, R. and Spingarn, K. (1982). *Control, Identification, and Input Optimization.* Plenum, New York.

Levadi, V. S. (1966). Design of input signals for parameter estimation. *IEEE Trans. Automat. Contr.* **AC-11**, 205-211.

Li, K.-H. and Chan, N. N. (1989). An algorithm for A-optimal regression designs. *Comput. Statist. Data Anal.* **8**, 49-57.

Marshall, A. W. and Olkin, I. (1979). *Inequalities : Theory of Majorization and Its Applications.* Academic Press, New York.

Mehra, R. K. (1974a). Optimal input signals for parameter estimation in dynamic systems - survey and new results. *IEEE Trans. Automat. Contr.* **AC-19**, 753-768.

Mehra, R. K. (1974b). Optimal inputs for linear system identification. *IEEE Trans. Automat. Contr.* **AC-19**, 192-200.

Parlett, B. N. (1980). *The Symmetric Eigenvalue Problem.* Prentice-Hall, Englewood Cliffs.

Parzen, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32**, 951-989.

Pázman, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

Seber, G. A. F. (1977). *Linear Regression Analysis.* Wiley, New York.

Wang, B. and Zhang, F. (1992). Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices. *Linear Algebra Appl.* **160**, 113-118.

Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

E-mail: khli@cuhk.edu.hk

E-mail: nnchan@cuhk.edu.hk

(Received August 1996; accepted June 1997)