Statistica Sinica **12**(2002),1015-1026

L_p-OPTIMALITY FOR REGRESSION DESIGNS UNDER CORRELATIONS

Kim-Hung Li and Nai N. Chan

The Chinese University of Hong Kong and University of Melbourne

Abstract: The input energy constraints in a linear dynamic system considered in this paper are of the form that the Euclidean norm of each column of its design matrix is bounded above by a constant. An exact L_p-optimal design is obtained in closed form which is easily computable. Interestingly, the L_p-optimal designs for the generalized and the ordinary least squares estimators coincide. An example is given to demonstrate how the results can be used to find a design that performs well under all L_p criteria.

Key words and phrases: CL vector, correlated error, dynamic systems, generalized least squares, L_p-optimal design, majorization, ordinary least squares.

1. Introduction

Consider the linear regression model

$$y = X\beta + e, \tag{1}$$

where y is an $n \times 1$ vector of observations, X is an $n \times r$ $(n \geq r)$ design matrix, β is an $r \times 1$ vector of unknown parameters, and e is an $n \times 1$ vector of random errors. The vector e has mean zero and covariance matrix $\sigma^2\Lambda$, where σ is an unknown parameter and Λ is a known $n \times n$ positive definite matrix. An exact optimal design for (1) is the design matrix which is "optimal" in a certain sense in a given experimental region.

The form of the experimental region depends on the meaning of the design variables which constitute the design matrix X. Most research in this area has been limited to situations where all elements in X can be independently chosen, or where restrictions are only allowed among elements in the same row. An example of the former case is weighing designs (see for example Banerjee (1975)), while typical examples of the latter include polynomial regression (see Gaffke and Krafft (1982), Constantine, Lim and Studden (1987), Kraft and Schaefer (1995), and Chang and Yeh (1998)), and the design of mixture experiments (see Chan (1995) for a review). However, experimental regions with restrictions on elements in the same column can arise naturally in real applications. For example, a column of

X may represent (a function of) the money or time to be spent in an experiment, so that control on the total amount of money or time specifies a constraint on the column. Optimal designs under column constraints are considered in Rao (1973, pp.235-236).

The experimental region considered in this paper is the set H of all $n \times r$ design matrices X (of rank r) of which the ith column has a Euclidean norm not exceeding c_i, $i = 1, \ldots, r$, where the c_i's are given positive numbers. Such a region has been considered in Dorogovcev (1971), Chan and Li (1989), and Li, Chan and Wong (1998). Without loss of generality, we assume that $c_1 \leq c_2 \leq \cdots \leq c_r$.

Restrictions on sum of squares are common in dynamic systems. In system theory, energy is defined as a sum of squares; it becomes an integral of a squared function in continuous time systems. A constraint on the energy of an input signal is a restriction on the sum of squared inputs (see for example Levadi (1966), and Mehra (1974)). If each column of X stores the values of an input signal, the experimental region H corresponds to putting an energy constraint on each of the signals.

For a given design matrix X of rank r, the best linear unbiased estimator of β based on the observation y is the generalized least squares estimator (GLSE) $(X'\Lambda^{-1}X)^{-1}X'\Lambda^{-1}y$. Further, the covariance matrix of the GLSE is $\sigma^2\Sigma$, where $\Sigma = (X'\Lambda^{-1}X)^{-1}$. Another unbiased estimator of the parameter β is the ordinary least squares estimator (OLSE) $(X'X)^{-1}X'y$.

Denote by $\mathrm{Var}\,(\hat{\beta})$ the $r \times r$ covariance matrix (depending on X) of an estimator $\hat{\beta}$ of β. We use the L_p-optimality criterion (see also the matrix mean ϕ_p-optimality criterion in Pukelsheim (1993, pp.140-143)) which is to find a design matrix $X \in H$ that minimizes $[tr\{\mathrm{Var}\,(\hat{\beta})^p\}]^{1/p}$. When p tends to zero, this reduces to D-optimality (which is to minimize the determinant of $\mathrm{Var}\,(\widehat{\beta})$); when p is one, it is simply A-optimality; and when p diverges to infinity, it becomes E-optimality (which is to minimize the largest eigenvalue of $\mathrm{Var}\,(\widehat{\beta})$).

As a single criterion, D-, A- and E-optimality are popular. However, there is also interest in multiple-objective optimal designs (Huang and Wong (1998)), which include compound optimal designs (Cook and Wong (1994)) and constrained optimal designs (Stigler (1971), Studden (1982), and Lee (1987, 1988)). A design is "good" if it performs well under a range of criteria. Robustness to the choice of criteria is desirable under certain circumstances. In this paper, we not only derive an L_p-optimal design for any specific p, but also provide sufficient insights to achieve a design that performs "well" under all L_p-criteria, as demonstrated by the example in Section 3.

To show how the regression setting arises in dynamic systems, consider a simple example from Levadi (1966) and Mehra (1974).

Example. Consider a continuous time system $dx(t)/dt = -x(t) + bu(t)$, where t denotes time in $[0, \tau]$, $x(t)$ is a scalar state variable, $u(t)$ is a scalar input, and b is an unknown parameter. The output variable $\psi(t)$ is the $x(t)$ observed with noise, i.e., $\psi(t) = x(t) + v(t)$, where $v(t)$ is a colored noise process with known autocovariance function. The initial state $x(0) = 0$. We observe $\psi(t)$ at discrete times, say $t_1, \ldots, t_m$. The input function, $u(t)$, is subject to an energy constraint $\int_0^\tau u^2(t)\, dt \le c^2$ and $\psi(t)$ can be expressed in terms of $u(t)$ as $\psi(t) = b \exp(-t) \int_0^t \exp(s) u(s) ds + v(t)$. Suppose that $u(t)$ belongs to a function space spanned by n orthonormal functions $g_i(t)$, $i = 1, \ldots, n$, $n \le m$, in $[0, \tau]$. (This assumption imposes modest restrictions on $u(t)$, as n can be any positive integer less than m and we are free to choose the $g_i(t)$'s.) Write $u(t) = \sum_{i=1}^n \alpha_i g_i(t)$. The energy constraint becomes $\sum_{i=1}^n \alpha_i^2 \le c^2$. The model can be expressed as $\psi(t) = b \sum_{i=1}^n \alpha_i h_i(t) + v(t)$, where $h_i(t) = \exp(-t) \int_0^t \exp(s) g_i(s) ds$ for $i = 1, \ldots, n$.

Let $A = [\alpha_1, \ldots, \alpha_n]'$, L be the $m \times n$ matrix with $h_i(t_j)$ as its (j, i)th element, $\Psi = [\psi(t_1), \ldots, \psi(t_m)]'$, and $V = [v(t_1), \ldots, v(t_m)]'$; then $\Psi = LAb + V$. Further, let Γ be the covariance matrix of V. The GLSE of b is then identical to that for model (1) with $X = A$, $y = (L'\Gamma^{-1}L)^{-1}L'\Gamma^{-1}\Psi$, and $\Lambda = (L'\Gamma^{-1}L)^{-1}$. The restriction on the design matrix X $(= A)$ is that the sum of squares of its elements is bounded above by c^2. Therefore, we arrive at a design setting for (1) in the desired experimental region.

For OLSE, the A-optimality problem under the experimental region H was considered by Chan (1982, 1987). A general and concise construction method of an A-optimal design had been suggested in Chan and Li (1989) and Li and Chan (1989), whilst Li, Chan and Wong (1998) provided an efficient algorithm for deriving an E-optimal design matrix. These results are special cases of the construction theorem in Section 2, where an exact L_p-optimal design matrix is given in closed form. It will be shown that the specified design matrix is L_p-optimal for both GLSE and OLSE. In Section 3, a linear systems example is used to demonstrate how the theorem can be used to derive a design which has high efficiency in the L_p-optimality family.

2. An Exact L_p-Optimal Design Matrix

Given two ordered vectors of dimension r, say $[a_i] \equiv [a_1, \ldots, a_r]$ and $[b_i] \equiv [b_1, \ldots, b_r]$ with $a_1 \le \cdots \le a_r$ and $b_1 \le \cdots \le b_r$, the vector $[a_i]$ is said to majorize $[b_i]$, written as $[a_i] \succ [b_i]$, if $\sum_{j=1}^i a_j \le \sum_{j=1}^i b_j$, $i = 1, \ldots, r-1$, and $\sum_{j=1}^r a_j = \sum_{j=1}^r b_j$; see, for example, Marshall and Olkin (1979, p.5). Let $D_r = \{[d_i] \equiv [d_1, \ldots, d_r] : 0 \le d_1 \le \cdots \le d_r\}$, and let D_{r+} be the subset of D_r with $d_1 > 0$.

In studying A-optimality in the experimental region H, Chan and Li (1989) defined a CL sequence, also referred to as a CL vector, and an algorithm was proposed for its construction. Li, Chan and Wong (1998) suggested an alternative and yet more efficient algorithm, which is used here to define the CL vector in a recursive manner.

Definition. Given $[a_i]$ and $[b_i]$ in D_{r+} , let $h = \max\{\sum_{j=1}^{i} a_j / \sum_{j=1}^{i} b_j : i = 1, \ldots, r\}$, and k $(1 \leq k \leq r)$ be the smallest integer such that $\sum_{j=1}^{k} a_j / \sum_{j=1}^{k} b_j = h$. We call $[d_1, \ldots, d_r]$ the CL vector of the pair $([a_i], [b_i])$ if $d_i = a_i/h,\ i = 1, \ldots, k$, and in the case $k < r$, $[d_{k+1}, \ldots, d_r]$ is the CL vector of the pair $([a_{k+1}, \ldots, a_r]$, $[b_{k+1}, \ldots, b_r])$ of vectors in $D_{(r-k)+}$.

When $r = 1$, the CL vector of $([a_1], [b_1])$ is $[b_1]$. For general r, the definition above either gives the CL vector directly (when $k = r$), or describes how it can be found based on CL vectors of smaller dimension. By induction, this defines the CL vector for all r. The definition can easily be used to find the CL vector of any pair of vectors in D_{r+} in a finite number of steps. For example, to find the CL vector $[d_1, \ldots, d_6]$ of $([1,\ 1,\ 3,\ 4,\ 6,\ 6], [1,\ 1,\ 2,\ 5,\ 5,\ 7])$, we have $h = 5/4$ and $k = 3$, and so $d_1 = a_1/h = 0.8,\ d_2 = a_2/h = 0.8,\ d_3 = a_3/h = 2.4$, and $[d_4, d_5, d_6]$ is the CL vector of $([4,\ 6,\ 6], [5,\ 5,\ 7])$. Then for the pair $([4,\ 6,\ 6], [5,\ 5,\ 7])$, we have by definition that $h = 1$, and $k = 2$, and so $d_4 = 4/h = 4$ and $d_5 = 6/h = 6$. As $[d_6]$ is the CL vector of $([6], [7])$, $d_6 = 7$. Thus the desired CL vector is $[d_i] = [0.8,\ 0.8,\ 2.4,\ 4,\ 6,\ 7]$. The pseudocode of an algorithm for finding the CL vector of a pair $[a_i]$ and $[b_i]$ in D_{r+} is given in Appendix 1.

As the CL vector is defined by a deterministic recursive algorithm, it must be unique. Furthermore, if $[d_i]$ is the CL vector of $([a_i], [b_i])$, then $[d_i] \succ [b_i]$. This important property can be easily proved by induction using the facts that (a) for the integer k in the definition, $[d_1, \ldots, d_k] \succ [b_1, \ldots, b_k]$, and (b) if $k < r$, $d_k \leq d_{k+1}$ $(= a_{k+1}/ \max\{\sum_{j=k+1}^{i} a_j / \sum_{j=k+1}^{i} b_j : i = k+1, \ldots, r\})$. Also, $[d_i] \in D_{r+}$, and k is the smallest integer j such that $\sum_{i=1}^{j} d_i = \sum_{i=1}^{j} b_i$.

Denote the $r \times r$ diagonal matrix with ith diagonal element $q_i, i = 1, \ldots, r$, by $\mathrm{diag}[q_1, \ldots, q_r]$. Let $0 < \lambda_1 \leq \cdots \leq \lambda_r \leq \cdots \leq \lambda_n$ be the eigenvalues of the matrix Λ arranged in ascending order of magnitude. For any $p > 0$, let $[d_{p,i}] \equiv [d_{p,1}, \ldots, d_{p,r}]$ be the CL vector of $([\lambda_1^{p/(p+1)}, \ldots, \lambda_r^{p/(p+1)}], [c_1^2, \ldots, c_r^2])$, where the c_i's are defined in Section 1.

Theorem. *In the regression setting with the experimental region H given in Section 1 and an $n \times r$ design matrix Z, the following are equivalent:*

(i) *Z is L_p-optimal for the GLSE;*

(ii) *Z is L_p-optimal for the OLSE;*

(iii) *The matrix Σ for the design matrix Z has eigenvalues $\lambda_i/d_{p,i},\ i = 1, \ldots, r$.*

(iv)
$$Z = P\,\mathrm{diag}[\sqrt{d_{p,1}}, \ldots, \sqrt{d_{p,r}}]Q', \tag{2}$$

where P is an $n \times r$ matrix of which the columns are orthonormal eigenvectors of Λ such that $\Lambda P = P \operatorname{diag}[\lambda_1, \ldots, \lambda_r]$, and Q is an $r \times r$ orthogonal matrix such that the ith diagonal element

$$(Q \operatorname{diag}[d_{p,1}, \ldots, d_{p,r}] Q')_{ii} = c_i^2, \quad i = 1, \ldots, r. \tag{3}$$

Statement (iv) of the Theorem provides a neat and efficient method for constructing an exact L_p-optimal design matrix. The existence of a matrix Q satisfying (3) follows from the fact that $[d_{p,i}] \succ [c_1^2, \ldots, c_r^2]$. A Fortran subroutine for the construction of Q in finite steps is given in Chan and Li (1983). As matrices P and Q in (2) are not necessarily unique, there may be more than one optimal design.

The theorem yields the A- and E-optimal designs suggested in Chan and Li (1989), and in Li, Chan and Wong (1998), by simply setting $p = 1$ and by letting p diverge to infinity, respectively. When p diverges to infinity, $[d_{p,i}]$ converges to the CL vector of $([\lambda_1, \ldots, \lambda_r], [c_1^2, \ldots, c_r^2])$.

There are two particular cases of interest. One is when the r smallest eigenvalues of Λ are identical, and the other is when all c_i's are equal. These two cases are considered in (a) and (b) of the following Corollary. Case (a) of the Corollary also applies when p tends to zero (corresponding to D-optimality) because $\lambda_i^{p/(p+1)}/c_i^2$ tends to $1/c_i^2$, non-increasing as the c_i's are arranged in non-descending order.

Corollary. *Under the same regression setting and notations of the theorem, we have*

(a) *If $\lambda_i^{p/(p+1)}/c_i^2$ is non-increasing in $i = 1, \ldots, r$ (in particular, if $\lambda_1 = \cdots = \lambda_r$), the exact L_p-optimal design is $Z = P \operatorname{diag}[c_1, \ldots, c_r]$.*

(b) *If $c_1 = \cdots = c_r = c$, and a Hadamard matrix G of order r exists, then the exact L_p-optimal design is*

$$Z = \left[c/(\sum_{i=1}^{r} \lambda_i^{p/(p+1)})^{1/2} \right] P \operatorname{diag}[\lambda_1^{p/[2(p+1)]}, \ldots, \lambda_r^{p/[2(p+1)]}] G.$$

3. An Example

As considered by Dorogovcev (1971) (see also Chang (1979), Chang and Wong (1981), and Li, Chan and Wong (1998)), suppose there is a continuous time output process $z(t)$ which depends on two user-supplied input functions $f_1(t)$ and $f_2(t)$ through the following equation:

$$z(t) = \beta_1 f_1(t) + \beta_2 f_2(t) + \xi(t). \tag{4}$$

Here $\xi(t)$ is a Gaussian process with $E(\xi(t)) = 0$, and $E(\xi(t_1)\xi(t_2)) = [\alpha + \min(t_1, t_2)]\sigma^2$ for a positive constant α, and any non-negative values t, t_1, and

t_2. We are going to observe the output $z(t)$ for a period of time, say from time 0 to 1. The input functions are subject to energy constraints $\int_0^1 f_1^2(t)dt \le c_1^2$ and $\int_0^1 f_2^2(t)dt \le c_2^2$, where $0 < c_1 \le c_2$. The problem is to choose appropriate input functions that produce the "best" estimators of β_1 and β_2 based on the observed output. If we impose no further restrictions, then $f_1(t)$ and $f_2(t)$ can be chosen to make our estimates of β_1 and β_2 as precise as we want. To see this, take orthogonal inputs $f_1(t) = \sqrt{2}c_1\cos(m\pi t)$ and $f_2(t) = \sqrt{2}c_2\cos((m+1)\pi t)$ for a positive integer m. From (4), $\int_0^1 z(t)f_1(t)dt = c_1^2\beta_1 + \int_0^1 \xi(t)f_1(t)dt$. As $\int_0^1 \xi(t)f_1(t)dt$ has mean zero and variance $\pi c_1^2\sigma^2/m^2$, $\hat{\beta}_1 \equiv \int_0^1 z(t)f_1(t)dt/c_1^2$ is an unbiased estimator of β_1 with standard error approaching zero as m increases. Similarly $\hat{\beta}_2 \equiv \int_0^1 z(t)f_2(t)dt/c_2^2$ converges to β_2 by increasing m.

Suppose we restrict both $f_1(t)$ and $f_2(t)$ to be quadratic functions of t. We can then express the input functions as linear combinations of the orthonormal functions $g_1(t) = 1$, $g_2(t) = \sqrt{12}(t-1/2)$ and $g_3(t) = \sqrt{180}(t^2-t+1/6)$ for $0 \le t \le 1$. For a given 3×2 matrix X, let $(f_1(t), f_2(t)) = (g_1(t), g_2(t), g_3(t))X$. The OLSE of $\beta = (\beta_1, \beta_2)'$ under (4) (Dorogovcev (1971)) is $T^{-1}\int_0^1 z(t)(f_1(t), f_2(t))'dt$, where $T = [T_{ij}]$ is a 2×2 matrix with $T_{ij} = \int_0^1 f_i(t)f_j(t)dt$. It is easy to see that this estimator is identical to the OLSE in (1) with X defined above,

$$y = \left[\int_0^1 z(t)g_1(t)dt,\ \int_0^1 z(t)g_2(t)dt,\ \int_0^1 z(t)g_3(t)dt\right]',$$
$$e = \left[\int_0^1 \xi(t)g_1(t)dt,\ \int_0^1 \xi(t)g_2(t)dt,\ \int_0^1 \xi(t)g_3(t)dt\right]'.$$

The energy constraints amount to restricting the first and second columns of the design matrix X to have Euclidean norms not exceeding c_1 and c_2 respectively. Now, consider the L_p-optimality criterion. From the autocovariance structure of $\xi(t)$, the matrix Λ for the linear model does not depend on the parameter α and is equal to

$$\Lambda = \begin{pmatrix} 1/3 & \sqrt{3}/12 & -\sqrt{5}/60 \\ \sqrt{3}/12 & 1/10 & 0 \\ -\sqrt{5}/60 & 0 & 1/42 \end{pmatrix}.$$

The two smallest eigenvalues of Λ are $\lambda_1 = 0.00916246$ and $\lambda_2 = 0.042751$. Their corresponding normalized eigenvectors form the matrix

$$P = \begin{pmatrix} 0.31625 & 0.29844 \\ -0.50251 & -0.75243 \\ 0.80466 & -0.58718 \end{pmatrix}. \tag{5}$$

By the definition of the CL vector, $[d_{p,1}, d_{p,2}]$ takes one of the following forms:

Case 1. When $(\lambda_1/\lambda_2)^{p/(p+1)} < (c_1/c_2)^2$, we have $d_{p,1} = \lambda_1^{p/(p+1)}(c_1^2+c_2^2)/(\lambda_1^{p/(p+1)} + \lambda_2^{p/(p+1)})$ and $d_{p,2} = \lambda_2^{p/(p+1)}(c_1^2 + c_2^2)/(\lambda_1^{p/(p+1)} + \lambda_2^{p/(p+1)})$. Optimal input functions, which depend on the value of p, can be found by first computing the matrix Z in (2) and then converting it back to the input functions.

Case 2. When $(\lambda_1/\lambda_2)^{p/(p+1)} \geq (c_1/c_2)^2$, we have $d_{p,1} = c_1^2$ and $d_{p,2} = c_2^2$ so that Q in the theorem is the identity matrix. The matrix Z in (2) becomes $Z = P\,\mathrm{diag}[c_1, c_2]$. For the matrix P in (5), the optimal input functions are

$$\begin{aligned} f_1(t) &= c_1[0.31625g_1(t) - 0.50251g_2(t) + 0.80466g_3(t)] \\ &= c_1(2.98589 - 12.53636t + 10.79562t^2), \end{aligned} \tag{6}$$

$$\begin{aligned} f_2(t) &= c_2[0.29844g_1(t) - 0.75243g_2(t) - 0.58718g_3(t)] \\ &= c_2(0.28911 + 5.26894t - 7.87543t^2). \end{aligned} \tag{7}$$

Given c_1, c_2, and p, we compute $(\lambda_1/\lambda_2)^{p/(p+1)}(= 0.21432^{p/(p+1)})$, and compare it to $(c_1/c_2)^2$ to determine whether Case 1 or Case 2 applies. L_p-optimal inputs can then be easily found.

If $(c_1/c_2)^2 \leq \lambda_1/\lambda_2$, Case 2 holds for all p, and the input functions in (6) and (7) are optimal for all L_p-criteria. When $(c_1/c_2)^2 > \lambda_1/\lambda_2$, the optimal inputs depend on the choice of p. If we have no definite p in mind, a natural approach is to choose inputs which are optimal for a certain L_p-criterion and at the same time perform well under other L_p-criteria. Let $\nu_{p,q}$ be the efficiency of the L_p-optimal inputs in the L_q-criterion. From (iii) of the Theorem, $\nu_{p,q} = \{[\sum_{i=1}^r (\lambda_i/d_{q,i})^q]/[\sum_{i=1}^r (\lambda_i/d_{p,i})^q]\}^{1/q}$. We wish to find p so as to maximize $\min_{q\geq 0} \nu_{p,q}$.

Let γ be such that $(\lambda_1/\lambda_2)^{\gamma/(\gamma+1)} = (c_1/c_2)^2$. As the matrix Z in (2) is invariant for all $p \leq \gamma$, we need only consider p larger than or equal to γ. Since $(1+x^w)^{1/w}/\sqrt{x}$ is non-decreasing in $x \geq 1$ for any positive w, it can be shown that for $p \geq \gamma$,

$$\begin{aligned} \min_{q\geq 0} \nu_{p,q} &= \min\{\nu_{p,0}, \nu_{p,\infty}\} \\ &= \min\{(c_1^2 + c_2^2)\lambda_1^{p/[2(p+1)]}\lambda_2^{p/[2(p+1)]}/[c_1c_2(\lambda_1^{p/(p+1)} + \lambda_2^{p/(p+1)})], \\ &\quad (\lambda_1 + \lambda_2)/[\lambda_2^{1/(p+1)}(\lambda_1^{p/(p+1)} + \lambda_2^{p/(p+1)})]\}. \end{aligned}$$

The first component in the bracket is decreasing with respect to p, while the second component is increasing as p increases. The maximum of $\min_{q\geq 0} \nu_{p,q}$ is attained when the two quantities in the bracket are equal. In other words, the optimal choice of p is the root of the equation $(\lambda_1/\lambda_2)^{p/[2(p+1)]} = (1 + \lambda_1/\lambda_2)c_1c_2/(c_1^2 + c_2^2)$. The minimum efficiency, $\min_{q\geq 0} \nu_{p,q}$, for this p is $\lambda_2(\lambda_1 + \lambda_2)(c_1^2 + c_2^2)^2/[\lambda_2^2(c_1^2 + c_2^2)^2 + (\lambda_1 + \lambda_2)^2 c_1^2 c_2^2]$.

As an example, suppose $c_1 = 1$ and $c_2 = 2$. The optimal choice of p is 15.0333. We have $d_{p,1} = 0.9545$ and $d_{p,2} = 4.0455$. By choosing

$$Q = \begin{pmatrix} 0.9926 & 0.1214 \\ 0.1214 & -0.9926 \end{pmatrix},$$

the matrix Z in (2) is

$$Z = \begin{pmatrix} 0.3796 & -0.5583 \\ -0.6710 & 1.4426 \\ 0.6369 & 1.2677 \end{pmatrix}.$$

The optimal choice of input functions is

$$\begin{aligned} f_1(t) &= 0.3796g_1(t) - 0.6710g_2(t) + 0.6369g_3(t) \\ &= 2.9660 - 10.8699t + 8.5454t^2, \end{aligned}$$

$$\begin{aligned} f_2(t) &= -0.5583g_1(t) + 1.4426g_2(t) + 1.2677g_3(t) \\ &= -0.2222 - 12.0111t + 17.0084t^2. \end{aligned}$$

For this choice of inputs, the minimum efficiency under all L_q-criteria is 0.9825.

Acknowledgements

The authors are grateful to the Associate Editor and the referees for their helpful comments and suggestions, and to Dr. Ken Sharpe for his help in improving the presentation of this paper.

Appendix 1. Pseudocode of an algorithm for finding the CL vector

Integer : t, r, i, k;
Real : $[a_i] \equiv [a_1, \ldots, a_r]$, $[b_i] \equiv [b_1, \ldots, b_r]$, $[d_i] \equiv [d_1, \ldots, d_r]$, h, u, v;
Input : r (the dimension), $[a_i]$ and $[b_i]$ in D_{r+};
Output : $[d_i]$ is the CL vector of $([a_i], [b_i])$;

Set $t = 1$;
Do while $t \leq r$,
  begin
    Set $h = 0$, $u = 0$, and $v = 0$;
    For $i = t, \ldots, r$, do
      begin
        Add a_i to u, and add b_i to v;
        If $u/v > h$ then begin $h = u/v$ and $k = i$ end;
      end;
    For $i = t, \ldots, k$, do $d_i = a_i/h$;
    Set $t = k + 1$;
  end;

Appendix 2. A proof of the theorem

To prove the theorem in Section 2, we need two lemmas. Lemma 1 can be easily verified by exchanging elements not in an ascending order. Lemma 2 is a generalization of Theorem 1 in Chan and Li (1989).

Lemma 1. *Given $p > 0$, $[s_i] \in D_{r+}$ and $q_i > 0$ for $i = 1, \ldots, r$, it holds that $\sum_{i=1}^{r}(s_i/q_i)^p \geq \sum_{i=1}^{r}(s_i/q_{(i)})^p$, where $q_{(i)}$, $i = 1, \ldots, r$, is the ith smallest value of $\{q_1, \ldots, q_r\}$.*

Lemma 2. *Given $p > 0$ and $[s_i], [b_i] \in D_{r+}$, the minimum of $\sum_{i=1}^{r}(s_i/q_i)^p$ in the set $\{[q_i] \in D_r : [q_i] \succ [b_i]\}$ is attained at $[q_i] = [t_i]$ if and only if $[t_i]$ is the CL vector of $([s_1^{p/(p+1)}, \ldots, s_r^{p/(p+1)}], [b_i])$.*

Proof. Write $\phi([q_i]) = \sum_{i=1}^{r}(s_i/q_i)^p$. As $\phi([q_i])$ is a positive continuous function of $[q_i]$ and the domain $\{[q_i] \in D_r : [q_i] \succ [b_i]\}$ is compact, there is at least one vector, $[t_i]$ say, in the domain at which the function attains its minimum. Obviously $[t_i] \in D_{r+}$. The sufficiency part is a consequence of the necessity part and the uniqueness of the CL vector. We need only to prove the necessity part.

We prove the lemma by induction on r. Clearly the result holds when $r = 1$. Suppose it holds when the dimension of the vectors is less than r for any fixed $r \geq 2$. Write $a_i = s_i^{p/(p+1)}$, $i = 1, \ldots, r$. Let v be the smallest integer such that $\sum_{j=1}^{v} t_j = \sum_{j=1}^{v} b_j$ (v must exist as the equality holds at least for $v = r$). We have two cases.

Case 1. $v = r$. As there is a total equality constraint on the q_i's, $\phi([q_i])$ can be considered as a function of $q_1, \ldots, q_{r-1}$. Let ζ, $\zeta > 0$, be the minimum $\{t_1, b_1 - t_1, \ldots, \sum_{i=1}^{r-1}(b_i - t_i)\}$. It can be shown that $\phi([q_i])$ attains its minimum at $[t_i]$ in the domain $\{[q_1, \ldots, q_{r-1}] : t_i - \zeta/(r-1) < q_i < t_i + \zeta/(r-1)$, for $i = 1, \ldots, r-1\}$ by Lemma 1 and the fact that after rearranging the q_i's in non-descending order, the resulting vector majorizes $[b_i]$. As $[t_i]$ is in the interior of the above domain, the derivative must vanish at $[t_i]$. It follows that $t_i = a_i \sum_{j=1}^{r} b_j / \sum_{j=1}^{r} a_j$, $i = 1, \ldots, r$. As $[t_i] \succ [b_i]$, the value k in the definition must be r. Thus $[t_i]$ is the CL vector of $([a_i], [b_i])$.

Case 2. $v < r$. For any vector $[u_i]$, define $[u_{\ell:j}] = [u_\ell, \ldots, u_j]$, where $\ell \leq j$. Furthermore, for any $[u_i] \in D_j$, define $K([u_i]) = \{[w_i] \in D_j : [w_i] \succ [u_i]\}$. Clearly $\phi([t_i]) \geq \min_{[q_{1:v}] \in K([b_{1:v}])} \sum_{i=1}^{v}(s_i/q_i)^p + \min_{[q_{v+1:r}] \in K([b_{v+1:r}])} \sum_{i=v+1}^{r}(s_i/q_i)^p$. Let $[q_{1:v}^*]$ and $[q_{v+1:r}^*]$ be the CL vectors of $([a_{1:v}], [b_{1:v}])$ and $([a_{v+1:r}], [b_{v+1:r}])$ respectively. Define w_i to be the ith smallest value of $\{q_1^*, \ldots, q_r^*\}$. Then, from the above inequality,

$$\phi([t_i]) \geq \phi([q_i^*]) \geq \phi([w_i]) \geq \phi([t_i]). \tag{A.1}$$

The first inequality in (A.1) follows from the inductive assumption, the second inequality follows from Lemma 1, and the last inequality is true because $[w_i] \succ [b_i]$. Therefore, the first inequality above must become equality, implying $t_i = q_i^*$ for $i = 1, \ldots, r$. From the definition of v, we can prove that the value k in the definition of the CL vector of $([a_i],[b_i])$ is v, implying that $[t_i]$ is the CL vector of $([a_i],[b_i])$.

Proof of the theorem. The proof proceeds by first showing that statements (i), (iii) and (iv) are equivalent, and then by showing that (iv) implies (ii), and (ii) implies (i).

To show (iv) implies (iii), we observe that the matrix Z in (2) is in H in view of (3). Moreover, $\Sigma = \{Q\,\mathrm{diag}[d_{p,1}/\lambda_1,\ldots,d_{p,r}/\lambda_r]Q'\}^{-1}$, showing that Σ has eigenvalues $\lambda_i/d_{p,i}$, $i = 1,\ldots,r$.

Next we prove that (iii) implies (i). By the singular value decomposition of a real matrix $X \in H$, we may write $X = AVB'$, where A is an $n \times r$ real matrix with orthonormal columns, $V = \mathrm{diag}[v_1,\ldots,v_r]$, with $0 < v_1 \le \cdots \le v_r$, and B is an $r \times r$ orthogonal matrix. It can be easily shown that for the matrix Σ in Section 1, the minimum of $tr(\Sigma^p)$ always occurs at an $X \in H$ such that the ith diagonal element $(X'X)_{ii}$ of $X'X$ is equal to c_i^2 , i.e., $(BV^2B')_{ii} = c_i^2$, $i = 1,\ldots,r$. Therefore, we can assume that the above equalities hold. This implies that $[v_1^2,\ldots,v_r^2] \succ [c_1^2,\ldots,c_r^2]$. As $tr(\Sigma^p) = tr[\{V^{-1}(A'\Lambda^{-1}A)^{-1}V^{-1}\}^p]$, we have by Theorem 6 (iv) and (v) in Wang and Gong (1993) that it is greater than or equal to

$$\sum_{i=1}^{r}[\text{the ith smallest eigenvalue of } (A'\Lambda^{-1}A)^{-1}]^p/v_i^{2p}. \tag{A.2}$$

By the Poincaré Separation Theorem (see, for example, Rao (1973, p.64)), the quantity in (A.2) is greater than or equal to

$$\sum_{i=1}^{r}(\lambda_i/v_i^2)^p. \tag{A.3}$$

As $[v_i^2] \succ [c_i^2]$, it follows from Lemma 2 that

$$tr(\Sigma^p) \ge \sum_{i=1}^{r}(\lambda_i/d_{p,i})^p, \tag{A.4}$$

providing a lower bound for $tr(\Sigma^p)$ in the case of the GLSE. This lower bound is attained by any matrix Z in (iii), showing that (iii) implies (i).

To show that (i) implies (iv), let X be an exact L_p-optimal design matrix for the GLSE, and decompose X by the singular value decomposition AVB'. As the lower bound in (A.4) is attainable at any matrix Z in (2), it must also be attainable at any L_p-optimal design matrix X. Therefore, the quantity in (A.2) should be equal to that in (A.3), implying that the matrix A must be one of the matrices P given in (iv). Also, the quantity in (A.3) should be equal to the greatest lower bound and hence, by Lemma 2, $[v_i^2] = [d_{p,i}]$. Clearly, $(X'X)_{ii} = c_i^2$ for all i, and therefore the corresponding B must be one of the matrices Q that satisfies (3). This completes the proof of (i) implying (iv).

For any matrix Z in (iv), Z is L_p-optimal for the GLSE. By Theorem 3.6 in Seber (1977, p.63), the GLSE and the OLSE of β for Z are identical. Therefore, by the optimality property of GLSE, we have (iv) implies (ii).

Finally, to show that (ii) implies (i), let X be an exact L_p-optimal design matrix in the OLSE case, and $\widehat{\beta}_0$ and $\widehat{\beta}_g$ be the OLSE and the GLSE, respectively, based on X. We have that

$$\sigma^{2p}\sum_{i=1}^{r}(\lambda_i/d_{p,i})^p = tr(\text{Var}\,(\widehat{\beta}_0)^p) \geq tr(\text{Var}\,(\widehat{\beta}_g)^p) \geq \sigma^{2p}\sum_{i=1}^{r}(\lambda_i/d_{p,i})^p,$$

where the first equality holds because (iv) implies (ii), and the last inequality follows from (A.4). Therefore, $tr(\text{Var}\,(\widehat{\beta}_g)^p) = \sigma^{2p}\sum_{i=1}^{r}(\lambda_i/d_{p,i})^p$, implying that X is also an exact L_p-optimal design matrix for the GLSE.

References

Banerjee, K. S. (1975). *Weighing Designs for Chemistry, Medicine, Economics, Operations Research, Statistics*. Marcel Dekker, New York.

Chan, L. Y. (1995). A review on optimal design for mixture models. In *Five Decades as a Mathematician and Educator – On the* 80th *Birthday of Professor Yung-Chow Wong* (Edited by K. Y. Chan and M. C. Liu), 45-88. World Scientific, Singapore.

Chan, N. N. (1982). A-optimality for regression designs. *J. Math. Anal. Appl.* **87**, 45-50.

Chan, N. N. (1987). Schur-convexity for A-optimal designs. *J. Math. Anal. Appl.* **122**, 1-6.

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. *Comput. 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. Multivariate Anal.* **11**, 85-101.

Chang, F.-C. and Yeh, Y.-R. (1998). Exact A-optimal designs for quadratic regression. *Statist. Sinica* **8**, 527-533.

Constantine, K. B., Lim, Y. B. and Studden, W. J. (1987). Admissible and optimal exact designs for polynomial regression. *J. Statist. Plann. Inference* **16**, 15-32.

Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrained and compound optimal designs. *J. Amer. Statist. Assoc.* **89**, 687-692.

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

Gaffke, N. and Krafft, O. (1982). Exact D-optimal designs for quadratic regression. *J. Roy. Statist. Soc. Ser. B* **44**, 394-397.

Huang, Y.-C. and Wong, W. K. (1998). Sequential construction of multiple-objective optimal designs. *Biometrics* **54**, 1388-1397.

Krafft, O. and Schaefer, M. (1995). Exact Elfving-minimax designs for quadratic regression. *Statist. Sinica* **5**, 475-484.

Lee, C. M.-S. (1987). Constrained optimal designs for regression models. *Commun. Statist. Theory Meth.* **16**, 765-783.

Lee, C. M.-S. (1988). Constrained optimal designs. *J. Statist. Plann. Inference* **18**, 377-389.

Levadi, V. S. (1966). Design of input systems for parameter estimation. *IEEE Trans. Automat. Control 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.

Li, K.-H., Chan, N. N. and Wong, K.-S. (1998). E-optimality for regression designs under correlations. *Statist. Sinica* **8**, 535-542.

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

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

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. Wiley, New York.

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

Stigler, S. M. (1971). Optimal experimental design for polynomial regression. *J. Amer. Statist. Assoc.* **66**, 311-318.

Studden, W. J. (1982). Some robust-type D-optimal designs in polynomial regression. *J. Amer. Statist. Assoc.* **77**, 916-921.

Wang, B.-Y. and Gong, M.-P. (1993). Some eigenvalue inequalities for positive semidefinite matrix power products. *Linear Algebra Appl.* **184**, 249-260.

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

E-mail: khli@cuhk.edu.hk

Department of Mathematics and Statistics, University of Melbourne, Parkville, Victoria 3052, Australia.

E-mail: nnchan@ms.unimelb.edu.au

(Received July 2000; accepted February 2002)