On Construction of Designs by Stepwise Method

Kôno, Kazumasa
College of General Education, Kyushu University

Sakaguchi, Koji
College of General Education, Kyushu University

https://doi.org/10.15017/1448935

出版情報：九州大学教養部数学雑誌. 7 (1), pp.1-12, 1969-09. College of General Education, Kyushu University
バージョン：
権利関係：

Math. Rep.
VII-1, 1969.

On Construction of Designs by Stepwise Method

by

K. KÔNO and K. SAKAGUCHI

[1969. 4. 1.]

§1. Introduction and preliminaries.

Recently, the properties of the optimal designs on regression problems have been studied by G. E. Elfving [2], S. Ehrenfeld [1], J. Kiefer [4],[5], J. Kiefer and J. Wolfowitz [6] and the other authors. Their results are mainly concerned with G-, or D-optimal criterion and their designs are characterized as the probability measure on some compact set. Their results are, therefore, not useful for designs which have a finite number of observations, except some special cases, and are usually approximative.

Our purpose in this paper is to construct exact designs for any size that take the largest information gain in some sence at each experimental stage.

Let $\mathfrak{X}$ be a p-dimensional compact set, $\theta'=(\theta_1, \cdots, \theta_k)$ be a k-dimensional parameter vector and $\boldsymbol{f}'(x)=(f_1(x),\cdots, f_k(x))$ be linearly independent and continuous functions defined on $\mathfrak{X}$. Each observations at x on $\mathfrak{X}$ yields independent random variable Y_x. Let

$$EY_x=\theta'\boldsymbol{f}(x), \qquad \mathrm{Var}\ Y_x=\sigma^2. \tag{1. 1}$$

Let ξ be a design which is the probability measure on $\mathfrak{X}$, then the information matrix of ξ in the sense of Fisher is

$$M(\xi)=(m_{ij}(\xi)), \qquad m_{ij}(\xi)=\int_{\mathfrak{X}} f_i(x)\ f_j(x)d\xi(x). \tag{1. 2}$$

If the size of design ξ is n and the matrix $M(\xi)$ is not singular, M.L E.'s of θ and $\theta'\boldsymbol{f}(x)$ have respectively,

$$\mathrm{Cov}(\hat{\theta})=\frac{1}{n}M^{-1}(\xi)\sigma^2, \tag{1. 3}$$

$$\mathrm{Var}(\hat{\theta}'\boldsymbol{f}(x))=\frac{1}{n}d(x,\ \xi)\sigma^2, \tag{1. 4}$$

where $\hat{\boldsymbol{\theta}}$ is M.L.E. of $\boldsymbol{\theta}$ and

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

The following criteria are reasonable.

Definition 1 (*D-optimality*).

ξ^* is D-optimal design if ξ^* satisfies $\max_{\xi} |M(\xi)|=|M(\xi^*)|$.

Definition 2 (*G-optimality*).

ξ^{**} is G-optimal design if ξ^{**} satisfies $\min_{\xi}\max_{x} d(x, \xi)=\max_{x} d(x, \xi^{**})$.

J. Kiefer and J. Wolfowitz [7] showed that the above two criteria are equivalent.

Now, let ξ take $x_1, x_2,\cdots,x_n$ (not necessarily distinct) with equal probability $\frac{1}{n}$ and let $X'=[\boldsymbol{f}(x_1),\cdots, \boldsymbol{f}(x_n)]$, then the information matrix

$$M(\xi)=\frac{1}{n}X'X.$$

M. Stone [8] discussed the construction of optimal designs based on the following information measure; let $P(\boldsymbol{\theta})$ be the a priori distribution of $\boldsymbol{\theta}$ and $P(\boldsymbol{\theta}/y_x)$ be the a posteriori distribution after observing y_x, then the information gain of observing y_x is defined by

(1. 6) $$\Delta I=I_{P(\theta)}-I_{P(\theta/y_x)},$$

where

(1. 7) $$I_{P(v)}=\int P(v)\ \log\ P(v)dv.$$

The fundamental theorem is as follows; if a priori distribution of $\boldsymbol{\theta}$ is $N(\mu, V_0)$ and the distribution of y_x is N.I.D. $(\boldsymbol{\theta}'\boldsymbol{f}(x), 1)$, the information gain ΔI of the design ξ with information matrix $M(\xi)$ is

(1. 8) $$\Delta I=\frac{1}{2}\log|I+V_0M(\xi)| =\frac{1}{2}\log|V_0||V_0^{-1}+M(\xi)|,$$

where V_0 is not singular.

M. Stone showed that the designs which maximized the information gain had almost the same properties of designs by above authors. While, this definition of information gain has the advantage that we can comparison the designs even if the information matrix $M(\xi)$ is singular.

We employ this information measure and attempt to construct the designs by the stepwise optimal procedure that allocates the next observa-

tion point to have the maximum information gain for any size of designs n and any a priori information matrix M which is singular or not singular.

§ 2. Stepwise optimal procedure and its properties.

Let $x_1, \cdots, x_n$ be a sequence of observation points on $\mathfrak{X}$, $M_i = M_0 + \sum_{j=1}^{i} \boldsymbol{f}(x_j)\boldsymbol{f}'(x_j)$, $(i=1, 2, \cdots)$, the i-th information matrix, and y_i be the observed value corresponding to x_i, $(i=1, 2, \cdots)$, then we have the increment of information by y_{i+1},

(2. 1) $$\Delta I_{i+1} = I_{P(\theta \mid y_1, \cdots, y_i)} - I_{P(\theta \mid y_1 \cdots y_i,\ y_{i+1})}$$
$$= \frac{1}{2} \log |M_i + \boldsymbol{f}(x_{i+1})\boldsymbol{f}'(x_{i+1})| \, |M_i|^{-1},$$

where M_i is the initial a priori dispersion matrix at the $(i+1)$ -st step.

We define the stepwise optimal rule ((rule S.O.)) as follows;

rule (S.O.): *Choose the observation point $x \in \mathfrak{X}$ that maximizes the information gain at each step.*

The information defined by (1. 5) and (1. 6) takes the average value on the observation point. Then our information is independent of the observed value y_x and therefore rule (S.O.) depends only on the observation points.

THEOREM 1. *Let M be $k \times k$ positive definite matrix and x be k-dimensional column vector, then*

(2. 2) $$|M + xx'| = |M| + x'M^*x = |M| \, [1 + x'M^{-1}x]$$

where $M^* = |M| \cdot M^{-1}$, i.e. M^* consists of the cofactors of the matrix M.

Proof. Since the rank of xx' is 1, there exists an orthogonal matrix O such that $Ox = (|x|, o, \cdots, o)'$.

(2. 3) $$|M + xx'| = |O| \, |M + xx'| \, |O'|$$
$$= |OMO'| + |x|^2 \times [(1,1) \text{ cofactor of the matrix } (OMO')].$$

$|OMO'| = |M|$ and

(2. 4) $$x'M^{-1}x = (Ox)'(OMO')^{-1}(Ox)$$
$$= (|x|, o, \cdots, o)(OMO')^{-1}(|x|, o, \cdots, o)'$$
$$= |x|^2 \times [(1,1) \text{ element of } (OMO')^{-1}]$$
$$= |x|^2 \times [(1,1) \text{ cofactor of } OMO']/|OMO'|.$$

We have the result of the theorem.

Remark 1. When M is singular, theorem 1 holds except the last equality. But we may restrict our interest to positive definite M.

Since M_i in (2. 1) is the inverse of a priori dispersion matrix of $\boldsymbol{\theta}$ at the $(i+1)$-st step, we restate by theorem 1 the rule (S.O.) as follows;

rule (S.O.)′ : *Choose* $x\in\mathfrak{X}$ *that maximize* $g_i(x)=\boldsymbol{f}(x)'M_i^{-1}\boldsymbol{f}(x)$ *for* $i=0, 1, 2, \cdots$.

Rule (S.O.) is to choose $x\in\mathfrak{X}$ that maximizes $|M_i+\boldsymbol{f}(x)\boldsymbol{f}'(x)|$ and it may be considered as conditional D-optimal criterion given $x_1, x_2, x_3, \cdots, x_i$ if we ignore M_0. Rule (S.O.)′ has the analogous form of $d(x, \xi)$ in the definition of G-optimal criterion.

THEOREM 2. *Let* $M_0^{(1)}$ *and* $M_0^{(2)}$ *be two a priori information matrix. If* $M_0^{(1)}-M_0^{(2)}$ *is non-negative difinite matrix, then*

$$\max_{x\in\mathfrak{X}}|M_0^{(1)}+xx'|\geqq\max_{x\in\mathfrak{X}}|M_0^{(2)}+xx'|. \tag{2. 5}$$

(We note the representation of (2. 5). In (2. 5), xx' is to be written in the form $\boldsymbol{f}(x)\boldsymbol{f}'(x)$, but if there are no confusions we will write as (2. 5) in the follows.)

Proof. Since $M_0^{(1)}$ and $M_0^{(2)}$ are symmetric matrix, there exists a regular transformation C such that

$$CM_0^{(1)}C'=\Lambda=\mathrm{diag}\ (\lambda_1, \cdots, \lambda_k),\quad CM_0^{(2)}C'=I. \tag{2. 6}$$

So $CC'=C'C=M^{(2)-1}$ and by the assumption $M_0^{(1)}$ and $M_0^{(2)}$, $\lambda_i\geqq 1$ $(i=1, 2, \cdots, k)$,

$$\begin{aligned}|M_0^{(i)}+xx'|&=|CC'|^{-1}\cdot|CM_0^{(i)}C'+Cxx'C'|\\&=|M_0^{(2)}|\cdot|CM_0^{(i)}C'+Cxx'C'|\quad(i=1, 2),\end{aligned} \tag{2. 7}$$

and if we set $Y=\{y|\ y=Cx,\ x\in\mathfrak{X}\}$ and apply theorem 1, then

$$\begin{aligned}\max_{x\in\mathfrak{X}}|M_0^{(1)}+xx'|&=\max_{x\in Y}|M_0^{(2)}||\Lambda+yy'|\\&=|M_0^{(2)}|[|\Lambda|+\max_{y\in Y}\ y'\Lambda^*y],\end{aligned} \tag{2. 8}$$

$$\begin{aligned}\max_{x\in\mathfrak{X}}|M_0^{(2)}+xx'|&=\max_{y\in Y}|M_0^{(2)}||I+yy'|\\&=|M_0^{(2)}|[1+\max_{y\in Y}\ y'y].\end{aligned} \tag{2. 9}$$

Therefore if we set $\sum_{i=1}^{k}y_i^{*2}=\max\sum_{i=1}^{k}y_i^2$, then

$$\begin{aligned}&\max_{x\in\mathfrak{X}}|M_0^{(1)}+xx'|-\max_{x\in\mathfrak{X}}|M_0^{(2)}+xx'|\\&=|M_0^{(2)}|[|\Lambda|+\max_{y\in Y}\sum(\prod_{j\neq i}\lambda_j)y_i^2-1-\max_{y\in Y}\sum y_i^2]\end{aligned} \tag{2. 10}$$

$$\geqq |M_0^{(2)}|\,[\prod_{i=1}^{k}\lambda_i - 1 + \sum(\prod_{j\neq i}\lambda_j - 1)y_i^{*2}] \geqq 0.$$

Remark 2. If we take $|M_0^{(1)}| \geqq |M_0^{(2)}|$ in place of nonnegative definiteness, the result of the theorem does not hold.

Example 1. $\mathfrak{X}=\{x \mid |x|\leq 1\}$, $M_0^{(1)}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, $M_0^{(2)}=\begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{3}{2}\end{pmatrix}$.

Then $|M_0^{(1)}|=1$, $|M_0^{(2)}|=\frac{3}{4}$ and $\max_{x\in\mathfrak{X}}|M^{(1)}+xx'|=2$, $\max_{x\in\mathfrak{X}}|M_0^{(2)}+xx'|=\frac{9}{4}$.

Remark 3. Above example shows that the inverse of the theorem 2 does not always hold.

THEOREM 3. *Under the same condition of theorem 2, the information gain* ΔI_1 *and* ΔI_2 *corresponding to* $M_0^{(1)}$ *and* $M_0^{(2)}$ *respectively of observing the next allocation point by the rule* (S.O.) *satisfy*

(2. 11) $$\Delta I_1 \geqq \Delta I_2.$$

Proof. ΔI_1 and ΔI_2 are respectively from (2. 1).

(2. 12) $$\Delta I_1 = \max_{x\in\mathfrak{X}} \frac{1}{2}\log |M_0^{(1)-1}|\,|M_0^{(1)}+F|$$

(2. 13) $$\Delta I_2 = \max_{x\in\mathfrak{X}} \frac{1}{2}\log |M_0^{(2)-1}|\,|M_0^{(2)}+F|$$

where $F=xx'$. Using the property (2. 7).

(2. 14) $$|M_0^{(1)-1}|\,|M_0^{(1)}+F| = \frac{|M_0^{(2)}|}{|M_0^{(1)}|}\,|\Lambda+Cxx'C'|$$

$$= \frac{|CM_0^{(2)}C'|}{|CM_0^{(1)}C'|}\,|\Lambda|\,[1+y'\Lambda^{-1}y]$$

$$= 1+y'\Lambda^{-1}y.$$

where $y=Cx$, $x\in\mathfrak{X}$. Similarly,

(2. 15) $$|M_0^{(2)-1}|\,|M_0^{(2)}+F| = 1+y'y.$$

(2. 16) $$\max_{y\in Y}\{1+y'\Lambda^{-1}y\} - \max_{y\in Y}\{1+y'y\}$$

$$= \max_{y\in Y}\sum_i \frac{y_i^2}{\lambda_i} - \max_{y\in Y}\sum_i y_i^2 \leqq 0,$$

because of $\lambda_i \geqq 1$, for $i=1, 2, \cdots, k$. Therefore we get the theorem.

Similarly to remark 2, the theorem 3 does not hold if we take $|M_0^{(1)}| \geqq |M_0^{(2)}|$ in place of non-negative definiteness.

For any non-zero vector $x\in\mathfrak{X}$ there exists some $\lambda(x)\geqq 1$ such that $\lambda(x)x\in\mathfrak{X}$ and $\lambda_1 x\notin\mathfrak{X}$ for all $\lambda_1>\lambda(x)$. Let

(2. 17) $$R(\mathfrak{X})=\{\lambda(x)x|\quad x\neq 0,\quad x\in\mathfrak{X}\},$$

THEOREM 4.

(2. 18) $$\max_{x\in\mathfrak{X}} x'M^{-1}x=\max_{x\in R(\mathfrak{X})} x'M^{-1}x$$

Proof. For each $x\in\mathfrak{X}$, there exists a point $\lambda(x)x\in R(\mathfrak{X})$ and $\lambda(x)\geqq 1$. Therefore, $x'M^{-1}x\leq\lambda^2(x)\cdot x'M^{-1}x=(\lambda(x)x)'M^{-1}(\lambda(x)x)$ for all $x\in\mathfrak{X}$.

Remark 4. If $\mathfrak{X}$ is convex set containing the origin, $R(\mathfrak{X})$ coincides the boundary set $\partial\mathfrak{X}$ of $\mathfrak{X}$.

THEOREM 5. *If $\mathfrak{X}$ is convex closure with a finite number of extremum points $\{b^{(1)},\cdots, b^{(m)}\}$. Then*

(2. 19) $$\max_{x\in\mathfrak{X}} x'M^{-1}x=\max_{x\in\{b^{(1)},\cdots,b^{(m)}\}} x'M^{-1}x=\max_{i=1,2,\cdots,m} b^{(i)\prime}M^{-1}b^{(i)}.$$

Proof. An artibitary point x of $\mathfrak{X}$ can be written in the form $\sum_{i=1}^{m}\lambda_i b^{(i)}$, where $\lambda_i\geqq 0$, $\sum_1^m\lambda_i=1$. Since $g(x)=x'M^{-1}x$ is a quadratic form of x and a convex function of x,

(2. 20) $$x'M^{-1}x\leqq\sum_1^m\lambda_i b^{(i)\prime}M^{-1}b^{(i)}\leqq\max_{i=1,\cdots,m} b^{(i)\prime}M^{-1}b^{(i)}.$$

Now, if the components of $f'(x)=(f_1(x),\cdots, f_k(x))$ are independent variables and $\mathfrak{X}$ is convex, then theorem 3 and 4 are useful. That is, for example, for linear regression problems, we may restrict our attention to boundary points of extremum points by rule (S.O.). But in general $f(\mathfrak{X})$ is not convex then there are the cases theorem 3 and 4 give us no information.

Example 2. $f'(x)=(1, x, x^2)$, $\mathfrak{X}=[-1, 1]$.

In this case $R(f(\mathfrak{X}))=f(\mathfrak{X})$.

In general case, an approach is to consider the geometric properties of the information matrix $M(\xi)$. (e.g. [3]).

§ 3. The construction of the designs for linear regression problems for the cases that $\mathfrak{X}$ is a sphere, simplex and cube.

(3. 1) $$EY_x=\theta'x,\qquad x'=(x^{(1)},\cdots, x^{(k)}).$$

We suppose that Y_x has independent and identical normal distribution.

(i) *The case that* $\mathfrak{X}$ *is a sphere*, e.g. $\mathfrak{X}=\{x\,|\,|x|\leqq 1\}$.
We may restrict $\mathfrak{X}$ to $\partial\mathfrak{X}=\{x\,|\,|x|=1\}$.

THEOREM 6. *If eigenvalues of* M_0 *are all equal, the system of allocation points* $x_{km+1},\cdots, x_{k(m+1)}(m=0, 1, \cdots)$ *given by the rule* (S.O.) *are mutually orthogonal and*

(3. 2) $$|M_n|=(\lambda+m+1)^r(\lambda+m)^{k-r},$$

where λ *is the equal eigenvalue of* M_0 *and* $n=km+r$, $0\leq r<k$.

Proof. M_i $(i=0, 1\cdots)$ being symmetric matrix, we can diagonalize M_i by some orthogonal transformation O.

(3. 3) $$g_i(x)=x'M_i^{-1}x=(Ox)'(OM_iO')^{-1}Ox=\sum_j\frac{y_{(j)}^2}{d_j^{(i)}}$$

where $D_i=OM_iO'=\text{diag}\ (d_1^{(i)},\cdots, d_k^{(i)})$ and $y=Ox$.

Therefore, we can maximize $g_i(x)$ if we set $y^{(j^*)}=1$ for j^* such that it satisfies $d_{j^*}^{(i)}=\min_{1\leq j\leq k} d_j^{(i)}$. For the sake of $\mathfrak{X}$ being sphere, the existence of $x\in\mathfrak{X}$ corresponding to $y=(0,\cdots, 1, \cdots, 0)'=Ox$ is guaranteed where 1 corresponds to the j^*-th coordinate. From the assumption, $D_0=\lambda I$ and so $d_j^*=\lambda$ for $j^*=1,\cdots, k$. If we set $y_1=Ox_1=(1, 0, \cdots 0)'$, then $D_1=\text{diag}\ (\lambda+1, \lambda,\cdots,\lambda)$ and $d_j^*=\lambda$ for $j^*=2, 3,\cdots, k$. If we set $y_2=(0, 1,\cdots, 0)'$ then we can maximize g_1 at the second step and $D_2=\text{diag}\ (\lambda+1, \lambda+1, \lambda,\cdots,\lambda)$. Similarily, we can take $y_3,\cdots, y_k$ as maximizing each $g_i(x)(i=2,\cdots, k-1)$ and being mutually orthogonal.

By the method of adapting $y_1,\cdots, y_k$, $D_k=(\lambda+1)I$. Therefore, after the $(k+1)$-st step the sequence of observation points consists of the replication of $x_1, x_2,\cdots, x_k$. In general,

$$D_n=\text{diag}\ (\underbrace{\lambda+m+1,\cdots, \lambda+m+1}_{r}, \underbrace{\lambda+m,\cdots, \lambda+m}_{k-r}),$$

$|D_n|=(\lambda+m+1)^r(\lambda+m)^{k-r}$, where $n=km+r$, $0\leqq r<k$.

$|D_n|=|M_n|$, because of $D_n=OM_nO'$.

THEOREM 7. *For any initial information matrix* M_0, *let* M_n *be the information matrix obtained by the rule* (S.O.), *then*

(3. 5) $$\lim_{n\to\infty}\frac{1}{n}M_n=\frac{1}{k}I.$$

Proof. We can diagonalize M_0 by some orthogonal transformation O. $OM_0O'=\Lambda=\text{diag}\ (\lambda_1^{(0)},\cdots, \lambda_k^{(0)})$. By employing successively the allocation point by the rule (S.O.) we have the first positive integer m such that

(3. 6) $$|\lambda_1^{(m)}-\lambda_i^{(m)}|<1 \quad (i=2, 3,\cdots, k),$$

where $OM_mO'=\mathrm{diag}\,(\lambda_1^{(m)},\cdots,\lambda_k^{(m)})$ and, without loss of generality, we can put $\lambda_1^{(m)}\leq\lambda_2^{(m)}\leq\cdots\leq\lambda_k^{(m)}$. For $n>m$, let $n-m=k\ell+r$, $0\leqq r<k$, then

$$\lambda_i^{(n)}=\begin{cases}\lambda_i^{(m)}+\ell+1 & 0\leqq i\leqq r\\ \lambda_i^{(m)}+\ell & r<i<k.\end{cases}\tag{3. 7}$$

Hence, $\frac{1}{n}\lambda_i^{(n)}$ converges to $\frac{1}{k}$ in each case and $\frac{1}{n}M_n\longrightarrow\frac{1}{k}I$ because $\frac{1}{n}D_n$ $=\frac{1}{n}OM_nO'\longrightarrow\frac{1}{k}I$.

Remark 5. It is easiby shown that we can make any k-consecutive allocation points to be mutuslly orthogonal for n larger than m and any M_0.

That is, after some steps, the design constructed by the rule (S.O.) consist of the system of unit length vectors which are mutually orthogonal eigenvectors of a priori dispersion matrix.

(ii) *The case that $\mathfrak{X}$ is a simplex*, e.g. $\mathfrak{X}=\{x=(x^{(1)},\cdots,x^{(k)})'\,|\,x^{(i)}\geqq 0,\ \Sigma x^{(i)}=1\}$.

From the theorem 3 we may restrict the set of extremum point $\varepsilon(\mathfrak{X})$ $=\{(1,0,\cdots,0),\cdots,(0,\cdots,0,1)\}$. Since vectors of $\varepsilon(\mathfrak{X})$ are mutually orthogonal, if we take $M_0=\lambda I$ then the situation is Similar to the sphere case. Therefore the rule (S.O.) takes sucessively points of $\varepsilon(\mathfrak{X})$. If we take M_0 arbitrarily, then the choice of $x\in\mathfrak{X}$ is to choose the point of $\varepsilon(\mathfrak{X})$ that gives the large value as possible to minimum eigenvalue as the proof of the theorem 3 and so the analogy of the remark 5 holds in this case. Obviously the analogy of theorem 5 holds.

(iii) *The case that $\mathfrak{X}$ is a cube*, e.g. $\mathfrak{X}=\{x\,|\,|x^{(i)}|\leqq 1\}$.

We may restrict $\mathfrak{X}$ to $\partial(\mathfrak{X})=\{x\,|\,|x^{(i)}|=1\}$.

Example 3. $k=2$, $M_0=\lambda I$.

The rule (S.O.) takes $\{(1,1),(1,-1)\}$ in any order, where the point (i,j) can be interchanged to $(-i,-j)$.

Example 4. $k=3$, $M_0=\lambda I$.

The rule (S.O.) takes $\{(1,1,1),(1,-1,1),(1,1,-1)\}$ in any order where (i,j,k) can be interchanged to $(-i,-j,-k)$.

COROLLARY OF THEOREM 3. *In the case $k=2^m$, there exists a system of 2^m-mutually orthogonal vectors in the set of extremum points $\varepsilon(\mathfrak{X})$. Therefore the rule* (S.O.) *takes those 2^m vectors in any order.*

In the case that $\mathfrak{X}$ is a cube, we do not get yet explicit designs by the rule (S.O.) for every k. But using the theorem 1, we can compute the maximum value of $x'M_i^{-1}x$ for every $x\in\partial(\mathfrak{X})$ which have 2^k-points by high speed computer.

We describe here the design precisely on three step. Here $M_0=\lambda I$. The first chosen point by the rule (S.O.) may be any point of $\partial(\mathfrak{X})$ and without loss of generality we assume $x_1=(1,\cdots,1)$. Then

$$M_1=\begin{pmatrix}\lambda+1 & & 1\\ & \ddots & \\ 1 & & \lambda+1\end{pmatrix},\quad M_1^{-1}=-\frac{1}{\lambda+k}\begin{pmatrix}m+k-1 & & -1\\ & \ddots & \\ -1 & & m+k-1\end{pmatrix}.$$

$$g_1(x)=x'M_1^{-1}x=\left[\sum_{i=1}^{k}\binom{m+k-1}{m}x^{(i)2}-\sum_{1\le i<j\le k}x^{(i)}x^{(j)}\right]\frac{1}{\lambda+k}. \tag{3.8}$$

We must take x_2 to maximize $g_1(x)$ in the second step. It is the point that minimizes $\sum_{1\le i<j\le k}x^{(i)}x^{(j)}$.

LEMMA. *Under* $|x^{(i)}|=1$, $i=1,\cdots,k$, *for minimizing* $\sum_{1\le i<j\le k}x^{(i)}x^{(j)}$ *we must change* $\left[\frac{k}{2}\right]$ *signs of* $x^{(i)}$*'s. If* k *is odd, the number of opposite sign may be* $\left[\frac{k}{2}\right]+1$.

Proof. Let ℓ be the number of $x^{(i)}$'s with positive sign and $(k-\ell)$ with negative sign. By direct computation

$$\sum_{1\le i<j\le k}x^{(i)}x^{(j)}=\frac{1}{2}\{(k-2\ell)^2-k\}. \tag{3.9}$$

To minimize $\sum x^{(i)}x^{(j)}$, we choose $\ell=\left[\frac{k}{2}\right]$ (or if k is odd, ℓ may be $\left[\frac{k}{2}\right]+1$).

Therefore at the second step, we must choose the point x such that a half of k components has positive and other half has negative sign (or $\left[\frac{k}{2}\right]+1$ if k is odd).

Without loss of generality we choose $x_2=(\underbrace{1,\cdots,1}_{[\frac{k}{2}]},\ \underbrace{-1,\cdots,-1}_{k-[\frac{k}{2}]})$. Then

(3. 10) $$M_2=\left(\begin{array}{ccc|ccc} \lambda+2 & & 2 & & & \\ & \ddots & & & 0 & \\ 2 & & \lambda+2 & & & \\ \hline & 0 & & \lambda+2 & & 2 \\ & & & & \ddots & \\ & & & 2 & & \lambda+2 \end{array}\right),$$

$$M_2^{-1}=\frac{1}{\lambda+2k}\left(\begin{array}{ccc|ccc} \lambda+2(k-1) & & -2 & & & \\ & \ddots & & & 0 & \\ -2 & & \lambda+2(k-1) & & & \\ \hline & 0 & & \lambda+2(k-1) & & -2 \\ & & & & \ddots & \\ & & & -2 & & \lambda+2(k-1) \end{array}\right).$$

x_3 which maximizes $g_2(x)=x'M_2^{-1}x$ minimizes

(3. 11) $$\sum_{1\le i<j\le\left[\frac{k}{2}\right]} x^{(i)}x^{(j)} + \sum_{\left[\frac{k}{2}\right]+1\le i<j\le k} x^{(i)}x^{(j)}.$$

Using lemma, $(x^{(1)},\cdots x^{([\frac{k}{2}])})$ and $(x^{([\frac{k}{2}]+1)},\cdots, x^{(k)})$ we change the sign of a half number of those components. That is, we choose as x_3, say,

$$(\underbrace{1,\cdots,1}_{[\frac{\alpha}{2}]}\ \underbrace{-1,\cdots,-1}_{\alpha-[\frac{\alpha}{2}]},\ \underbrace{1,\cdots,1}_{[\frac{k-\alpha}{2}]}\ \underbrace{-1,\cdots,-1}_{k-\alpha-[\frac{k-\alpha}{2}]}),$$

where $\alpha=\left[\frac{k}{2}\right]$.

§ 4. Quadratic regression with one variable.

$$EY_x=\theta_0+\theta_1x+\theta_2x^2,\quad \mathfrak{X}=[-1,\ 1].$$

In this case, $\boldsymbol{f}'(x)=(1,\ x,\ x^2)$, $\boldsymbol{\theta}'=(\theta_0,\ \theta_1,\ \theta_2)$, we restrict designs to symmetric ones. That is, ξ is symmetric then $\int_{\mathfrak{X}} x\,d\xi(x)=\int_{\mathfrak{X}} x^3\,d\xi(x)\equiv 0$. We may restrict an initial a priori information matrix M_0 to

(4. 1) $$M_0=\begin{pmatrix} m_1 & 0 & m_2 \\ 0 & m_2 & 0 \\ m_2 & 0 & m_3 \end{pmatrix}.$$

"0" elements correspond to elements of the first and the third order

in the $f(x)f'(x)$.

Especially, if we take $m_2=m_3$, then

$$m_1 \left\{\begin{matrix}\geqq\\<\end{matrix}\right\} \frac{3}{2}m_2 \longrightarrow x_1=\left\{\begin{matrix}1, & -1\\ 1, & 0, & -1\\ & 0\end{matrix}\right\}.$$

If we take

$$M_0=\lambda\begin{pmatrix}\frac{3}{2} & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 1\end{pmatrix} \tag{4. 2}$$

then the rule (S.O.) takes $\{-1, 0, 1\}$ in any order. M_0 in (4. 2) is considered, consulting the result, as follows. Let ξ^* be the design which assign equal probability $\frac{1}{3}$ on each point of $\{-1, 0, 1\}$. Let $\phi_0(x)=\frac{1}{\sqrt{3}}$, $\phi_1(x)=\frac{x}{\sqrt{2}}$ and $\phi_2(x)=\frac{1}{\sqrt{6}}(3x^2-2)$. Then $\phi'=(\phi_0, \phi_1, \phi_2)$ has the properties that

$$\int \phi_i(x)\phi_j(x)=\delta_{ij},\ (i, j=0, 1, 2).$$

$$EY_x=\theta'f(x)=\tau'\phi(x). \tag{4. 4}$$

$$\theta=\begin{pmatrix}\frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}}\\ 0 & \frac{1}{\sqrt{2}} & 0\\ 0 & 0 & \frac{3}{\sqrt{6}}\end{pmatrix}\tau=T\tau. \tag{4. 5}$$

We give uniform information to τ, then a priori information matrix M_0 on θ is

$$M_0=\lambda'(TT')^{-1}=\lambda\begin{pmatrix}\frac{3}{2} & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 1\end{pmatrix},\ \lambda=2\lambda'>0.$$

Under that information matrix M_0, our design obtained by the rule (S.O.) coincides the usual symmetric D-and G- optimal design.

References

[1] EHRENFELD, S. (1955) Complete class theorems in experimental designs. Proc. Third Berkeley Symp. Math. Stat. Prob. 1. 57-67. Univ. of California Press.

[2] ELFVING, G. (1952) Optimum allocation in linear regression theory. Ann. Math. Statist. 23. 225-262.

[3] FARELL, R. H., KIEFER, J. and WALBRAN, A. (1965) Optimum multivariate designs. Proc. Fifth Berkeley Symp. Math. Stat. Prob. 1. 113-138.

[4] KIEFER, J. (1959) Optimum experimental design. J. Roy. Statist. Soc., Ser. B. 21. 272-319.

[5] ——. (1961). Optimum designs in regression problem, II. Ann. Math. Statist. 32. 298-325.

[6] KIEFER, J. and WOLFOWITZ, J. (1959) Optimum design in regression problem. Ann. Math. Statist. 30. 271-294.

[7] —— and ——. (1960) The equivalence of two extremum problems. Canad. J. Math., Vol. 12. 363-366.

[8] STONE, M. (1968) Application of a measure of information to the design and comparison of regression experiments. Ann. Math. Statist. 29. 55-70.