The Annals of Statistics
1977, Vol. 5, No. 4, 790-794

A CONVERGENCE THEOREM IN L-OPTIMAL DESIGN THEORY[1]

BY JIA-YEONG TSAY

University of Cincinnati Medical Center

It is shown that the Fedorov procedure for finding L-optimal designs always converges to an L-optimal design, even when the limiting design is singular. Two lemmas are given first. An example illustrates the case of singular limiting design.

1. Introduction. The theory of L-optimal (or linear optimal) designs was first introduced by Fedorov (1972), although some special cases had been considered before by many other authors, such as Elfving (1952), Kiefer and Wolfowitz (1959), Karlin and Studden (1966), etc. For the detail, see Tsay (1976). Fedorov suggests an iterative method to calculate L-optimal designs and shows that under some conditions the constructed sequence converges to an L-optimal design (see Theorem 2.10.1, Fedorov (1972)). One of his conditions is that the limiting design be nonsingular. The purpose of this paper is to prove that the Fedorov procedure always converges to an L-optimal design, even if the limiting design is singular.

Let $f' = (f_1, \cdots, f_k)$ be a vector of linearly independent continuous functions on a compact set X and let $\theta' = (\theta_1, \cdots, \theta_k)$ be a vector of unknown parameters (primes denote transposes). Consider the regression model with mean $\theta' f(x)$ and variance σ^2. An experimental design ξ is a probability measure on X. The information matrix of a design ξ is denoted by $M(\xi) = \int f(x)f'(x)\, d\xi(x)$. Let L be a nonnegative linear functional on the set of nonnegative definite symmetric matrices. A design ξ^* is L-optimal if $L(M^{-1}(\xi^*)) = \min_\xi L(M^{-1}(\xi))$. For a singular design ξ, $L(M^{-1}(\xi))$ is defined as $\lim_{\varepsilon\to 0} L\{(M(\xi) + \varepsilon I)^{-1}\}$. In the next section it will be shown that the Fedorov procedure always converges to an L-optimal design. An example is given in Section 3 for the case of singular limiting design.

2. Results. Let $\phi(x, \xi) = L\{M^{-1}(\xi)f(x)f'(x)M^{-1}(\xi)\}$ and $\bar{\phi}(\xi) = \max_x \phi(x, \xi)$. Given a nonsingular design ξ_0, define

$$\xi_{n+1} = (1 - \alpha_n)\xi_n + \alpha_n \xi(x_n)$$

where x_n is a point maximizing $\phi(x, \xi_n)$, $\xi(x_n)$ a design with mass one at x_n and

$$\alpha_n = \{\phi(x_n, \xi_n) - L(M^{-1}(\xi_n))\}/\gamma\phi(x_n, \xi_n)(d(x_n, \xi_n) - 1) \qquad (2.1)$$

with some constant $\gamma \geqq 1$ (Fedorov has only $\gamma > 1$). Here $d(x, \xi)$ is defined as

Received January 1976; revised December 1976.

[1] This research was supported in part by NIH Postdoctoral Fellowship 1 F32 ES05041-01.

AMS 1970 *subject classification.* Primary 62K05.

Key words and phrases. Optimal experimental designs, iterative procedure, regression model, L-optimal design, information matrix.

$f'(x)M^{-1}(\xi)f(x)$. We have the following theorem (compare with Theorem 2.10.1, Fedorov (1972)).

THEOREM. *For the sequence of designs* $\{\xi_n\}$ *generated above*, $L(M^{-1}(\xi_n)) \to L(M^{-1}(\xi^*))$, *where* ξ^* *is an L-optimal design.*

Before proving the theorem we will prove two lemmas.

LEMMA 1. *For the sequence of designs* $\{\xi_n\}$ *generated above*, $\sum_n (d(x_n, \xi_n))^{-1} = \infty$.

PROOF. Suppose $\sum_n (d(x_n, \xi_n))^{-1} < \infty$, then by the definition of α_n

$$\begin{aligned} \alpha_n &= \{\phi(x_n, \xi_n) - L(M^{-1}(\xi_n))\}/\gamma\phi(x_n, \xi_n)(d(x_n, \xi_n) - 1) \\ &< \{(1+\gamma)\phi(x_n, \xi_n) - L(M^{-1}(\xi_n))\}/\gamma\phi(x_n, \xi_n)\, d(x_n, \xi_n) \\ &< (1+\gamma)\phi(x_n, \xi_n)/\gamma\phi(x_n, \xi_n)\, d(x_n, \xi_n) \\ &= (1+\gamma)/\gamma\, d(x_n, \xi_n) \end{aligned} \tag{2.2}$$

where the first inequality is simply that for any $a > 0$, $b > 0$, $c > 0$ if $0 < a/b < 1$, then $a/b < (a+c)/(b+c)$. Using (2.2) we have $\gamma\alpha_n/(1+\gamma) < (d(x_n, \xi_n))^{-1}$, hence

$$\gamma/(1+\gamma) \cdot \textstyle\sum_n \alpha_n < \sum_n (d(x_n, \xi_n))^{-1} < \infty\,,$$

from which it follows that $\prod_n (1 - \alpha_n) > 0$.

By the definition of ξ_n, it is known that

$$\begin{aligned} \xi_1 &= (1-\alpha_0)\xi_0 + \alpha_0\xi(x_0) > (1-\alpha_0)\cdot\xi_0 \\ \xi_2 &= (1-\alpha_1)\xi_1 + \alpha_1\xi(x_1) > (1-\alpha_1)(1-\alpha_0)\xi_0 \\ &\cdots \\ \xi_n &> (1-\alpha_{n-1})\cdots(1-\alpha_0)\xi_0 = \textstyle\prod_{i=0}^{n-1}(1-\alpha_i)\cdot\xi_0\,. \end{aligned} \tag{2.3}$$

Let n go to infinity, we get from (2.3)

$$\lim \xi_n \geqq \textstyle\prod_n (1-\alpha_n)\cdot\xi_0\,.$$

But $\prod_n (1 - \alpha_n) > 0$ and ξ_0 is nonsingular, it follows that $\lim \xi_n$ is nonsingular, hence $\limsup d(x_n, \xi_n) < \infty$. This implies that $\sum_n (d(x_n, \xi_n))^{-1} = \infty$ which leads to a contradiction. Consequently, we have $\sum_n (d(x_n, \xi_n))^{-1} = \infty$.

LEMMA 2. *Let* ξ^* *be an L-optimal design. For any nonsingular design* ξ *the following inequality holds*:

$$L(M^{-1}(\xi)) - L(M^{-1}(\xi^*)) \leqq \bar{\phi}(\xi) - L(M^{-1}(\xi))\,.$$

(This is a special case of (6.5) of Kiefer (1974).)

PROOF. Define $g(\alpha) = L\{((1-\alpha)M(\xi) + \alpha M(\xi^*))^{-1}\}$ for $0 \leqq \alpha \leqq 1$. Then $g(\alpha)$ is convex in α, thus $g(1) - g(0) \geqq \dot{g}(0)$. But the derivative $\dot{g}(0)$ has the relation

$$\dot{g}(0) = L(M^{-1}(\xi)) - \textstyle\int \phi(x, \xi)\, d\xi^*(x) \geqq L(M^{-1}(\xi)) - \bar{\phi}(\xi)\,.$$

It follows that

$$\bar{\phi}(\xi) - L(M^{-1}(\xi)) \geqq -\dot{g}(0) \geqq g(0) - g(1) = L(M^{-1}(\xi)) - L(M^{-1}(\xi^*)) .$$

This proves the lemma.

PROOF OF THEOREM. Denote $L(M^{-1}(\xi_n))$, $L(M^{-1}(\xi^*))$, $\max_x \phi(x, \xi_n)$, $d(x_n, \xi_n)$ by L_n, L^*, $\bar{\phi}_n$, d_n respectively.

From Lemma 2.10.2(2) of Fedorov (1972) we have

$$L_0 \geqq L_1 \geqq L_2 \geqq \cdots \geqq L^* .$$

Now since $\{L_n\}$ is a bounded monotone decreasing sequence, it must converge, say, to $\underline{L}$. We will show that $\underline{L} = L^*$.

Suppose $\underline{L} > L^*$. Then there exists a number $\delta > 0$ such that for all n

$$L_n > L^* + \delta . \tag{2.4}$$

Combining Lemma 2 and (2.4) we get

$$\delta < \bar{\phi}_n - L_n \qquad \text{for all} \quad n . \tag{2.5}$$

From (2.10.6) of Fedorov, (2.1) and (2.5) it follows that

$$\begin{aligned} L_n - L_{n+1} &= \alpha_n/(1-\alpha_n) \cdot \{\bar{\phi}_n/(1+\alpha_n(d_n - 1)) - L_n\} \\ &= \frac{\bar{\phi}_n - L_n}{\gamma\bar{\phi}_n(d_n - 1) - (\bar{\phi}_n - L_n)} \left\{ \frac{\bar{\phi}_n}{1 + (\bar{\phi}_n - L_n)/\gamma\bar{\phi}_n} - L_n \right\} \\ &= \frac{\bar{\phi}_n - L_n}{\gamma\bar{\phi}_n(d_n - 1) - (\bar{\phi}_n - L_n)} \cdot \frac{(\gamma\bar{\phi}_n - L_n)(\bar{\phi}_n - L_n)}{(1+\gamma)\bar{\phi}_n - L_n} \\ &> \frac{\bar{\phi}_n - L_n}{\gamma\bar{\phi}_n d_n} \cdot \frac{(\gamma\bar{\phi}_n - \gamma L_n)\delta}{(1+\gamma)\bar{\phi}_n - (1+\gamma)L_n + \gamma L_n} \\ &= \frac{1}{\gamma\, d_n \bar{\phi}_n/(\bar{\phi}_n - L_n)} \cdot \frac{\gamma\delta}{(1+\gamma) + \gamma L_n/(\bar{\phi}_n - L_n)} \\ &= \frac{1}{d_n(1 + L_n/(\bar{\phi}_n - L_n))} \cdot \frac{\delta}{1 + \gamma + \gamma L_n/(\bar{\phi}_n - L_n)} \\ &> \frac{1}{d_n(1 + L_0/\delta)} \cdot \frac{\delta}{1 + \gamma + \gamma L_0/\delta} \\ &= c/d_n \end{aligned} \tag{2.6}$$

where $c = \delta/(1 + L_0/\delta)(1 + \gamma + \gamma L_0/\delta)$ is a constant.

By Lemma 1 and (2.6) we obtain

$$\begin{aligned} L_0 - L_N &= \textstyle\sum_{n=0}^{N-1} (L_n - L_{n+1}) \\ &> c \cdot \textstyle\sum_{n=0}^{N-1} d_n^{-1} \to \infty \qquad \text{as} \quad N \to \infty . \end{aligned} \tag{2.7}$$

From (2.7) it follows that $L_N \to -\infty$ as $N \to \infty$ which is impossible. Therefore, $\underline{L} = L^*$. This proves the theorem.

REMARK. It seems that the theorem may still hold if the α_n in (2.1) is relaxed to the form

$$\alpha_n = \{\phi(x_n, \xi_n) - L(M^{-1}(\xi_n))\}/\gamma L(M^{-1}(\xi_n))(d(x_n, \xi_n) - 1) \tag{2.8}$$

with a constant $\gamma > 1$. Theorem 3.1 (a) of Atwood (1976) assures the α_n in (2.8) to be less than 1, hence the sequence $\{\xi_n\}$ so constructed is a sequence of probability measures. The monotonicity of $\{L_n\}$ follows from (2.8) and the first equality of (2.6). However, a different technique is needed to prove Lemma 1. The following corollary does generalize the theorem in this direction, although not quite as far as may be possible.

COROLLARY. *Let λ be a positive constant. Define*

$$\alpha_n = (\bar{\phi}_n - L_n)/\gamma_n \bar{\phi}_n (d_n - 1)$$

with γ_n satisfying both

$$\gamma_n \geqq \lambda$$

and

$$\gamma_n \bar{\phi}_n > L_n . \tag{2.9}$$

Then for the sequence of designs $\{\xi_n\}$ generated with this α_n, $L(M^{-1}(\xi_n)) \to L(M^{-1}(\xi^*))$, where ξ^* is an L-optimal design. (*Note that the strict inequality in* (2.9) *is used to avoid the occurrence of singular designs in* $\{\xi_n\}$.)

PROOF. Lemma 1 can be proved as before, since $\gamma_n \geqq \lambda$ gives

$$\alpha_n < (1 + \gamma_n)/\gamma_n \, d(x_n, \xi_n) \leqq (1 + \lambda)/\lambda \, d(x_n, \xi_n)$$

as the form of the inequality (2.2). The monotonicity of $\{L_n\}$ can be obtained from the first line of (2.6) and (2.9). Therefore the corollary can be proved just as the theorem was with minor algebraic modification.

3. Example. We will give an example for the case of a singular limiting design. Let us consider the model $\theta' f(x) = \theta_0 + \theta_1 x$ on $[0, 1]$. Define $L(D) = D_{11}$, where $D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$ is a covariance matrix. Let $\xi_0 = \left\{ \begin{smallmatrix} 0, & 1 \\ \frac{1}{2}, & \frac{1}{2} \end{smallmatrix} \right\}$, and $\gamma = 1$ in (2.1). Then it is easy to see that $M^{-1}(\xi_0) = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$, $L(M^{-1}(\xi_0)) = 2$ and $\phi(x, \xi_0) = 4x^2 - 8x + 4$. Since $x_0 = 0$ maximizes $\phi(x, \xi_0)$, using (2.1) we have $\alpha_0 = \frac{1}{2}$, $\xi_1 = \left\{ \begin{smallmatrix} 0, & 1 \\ \frac{3}{4}, & \frac{1}{4} \end{smallmatrix} \right\}$. Similarly, $M^{-1}(\xi_1) = \begin{bmatrix} \frac{4}{3} & \frac{4}{3} \\ \frac{4}{3} & \frac{16}{3} \end{bmatrix}$, $L(M^{-1}(\xi_1)) = \frac{4}{3}$, $\phi(x, \xi_1) = \frac{16}{9}x^2 - \frac{32}{9}x + \frac{16}{9}$ and $\sup_x \phi(x, \xi_1) = \phi(0, \xi_1) = \frac{16}{9}$. Hence $x_1 = 0$, $\alpha_1 = \frac{3}{4}$ and $\xi_2 = \left\{ \begin{smallmatrix} 0, & 1 \\ \frac{15}{16}, & \frac{1}{16} \end{smallmatrix} \right\}$.

Let $a_n = (\frac{1}{2})^{2^n}$. Then using the same procedure we can inductively find that

$$\xi_n = \begin{Bmatrix} 0 & , & 1 \\ 1 - a_n, & & a_n \end{Bmatrix}, \qquad M^{-1}(\xi_n) = \begin{bmatrix} \dfrac{1}{1 - a_n} & \dfrac{1}{1 - a_n} \\ \dfrac{1}{1 - a_n} & \dfrac{1}{a_n(1 - a_n)} \end{bmatrix},$$

$$L(M^{-1}(\xi_n)) = 1/(1 - a_n), \qquad \phi(x, \xi_n) = (1 - x)^2/(1 - a_n)^2, \qquad x_n = 0$$

and $\alpha_n = 1 - a_n$.

Thus we have

$$\lim_{n\to\infty} L(M^{-1}(\xi_n)) = 1 = \lim_{n\to\infty} \phi(x_n, \xi_n) \qquad \text{and} \qquad \lim_{n\to\infty} \xi_n = \xi(0)$$

where $\xi(0)$ is a design concentrating all mass at the point 0. It is obvious that $\xi(0)$ is singular and L-optimal.

It is interesting to note that $\xi(0)$ is the only L-optimal design for this problem. Furthermore, if the procedure in the corollary is used, we can get a design as close to the optimal one as we want in one step. For instance, γ_0 could be chosen very close to but bigger than $\frac{1}{2}$, then α_0 would become very close to but smaller than 1. This makes ξ_1 very close to $\xi^* = \xi(0)$. In this example, the problem was to find a design which would provide an estimator with maximum efficiency for the intercept of the regression line $\theta_0 + \theta_1 x$. The design $\xi(0)$ is to take all observations at the point 0 and hence this design is disconnected for θ_0 and θ_1. For the theory of connected designs we refer the interested reader to Eccleston and Hedayat (1974). If we are interested in estimating the slope instead, the one and only one optimal design is to take half of the observations at 0 and the remaining half at 1. This design certainly is nonsingular.

Acknowledgments. The author wishes to thank Professor William J. Studden for his useful suggestions. The constructive comments of the referee (particularly the generalization of the theorem) and the Associate Editor are greatly appreciated.

REFERENCES

[1] Atwood, C. R. (1976). Convergent design sequences, for sufficiently regular optimality criteria. *Ann. Statist.* **4** 1124–1138.

[2] Eccleston, J. A. and Hedayat, A. (1974). On the theory of connected designs: characterization and optimality. *Ann. Statist.* **2** 1238–1255.

[3] Elfving, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 252–262.

[4] Fedorov, V. V. (1972). *Theory of Optimal Experiments* (translated by W. J. Studden and E. M. Klimko, eds.). Academic Press, New York.

[5] Karlin, S. and Studden, W. J. (1966). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

[6] Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879.

[7] Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30** 271–294.

[8] Tsay, J-Y. (1976). Linear optimal experimental designs. *Proc. 1976 Conf. on Information Sciences and Systems*, 222–226. Dept. of Elec. Engrg., Johns Hopkins Univ.

Department of Environmental Health
Division of Biostatistics
University of Cincinnati Medical College
Cincinnati, Ohio 45267