The Annals of Statistics
1973, Vol. 1, No. 2, 342-352

SEQUENCES CONVERGING TO *D*-OPTIMAL DESIGNS OF EXPERIMENTS

BY CORWIN L. ATWOOD

Haile Sellassie I University

Fedorov (*Theory of Optimal Experiments* (1972)) gives a sequence of designs converging to a D-optimal design. Several modifications of that sequence are given to improve the speed of convergence. The analogous sequence for estimating some of the parameters is shown to converge to a D-optimal design, whether or not all the parameters are estimable under the limiting design. We prove the result $d(x, \xi)\xi(x) \leqq 1$, and several related results.

0. Introduction and summary. There have been several papers giving sequences of designs which converge to D-optimal designs. (For terminology, see the next section.) Wynn [5] considers only designs using a finite number of observations. Given the design ξ_n using n observations, he produces ξ_{n+1} by taking the next observation at a point x_0 where $d(x, \xi_n)$ is maximized. The resulting sequence $\det M(\xi_n)$ converges, though not monotonically, to the maximum possible value. Fedorov [2] deals with the continuous theory, in which the designs considered are probability measures. He adds that measure at x_0 which maximizes $\det M(\xi_{n+1})$, and obtains a sequence analogous to that of [5], except the convergence now is monotone. More recently Wynn [6] has given a similar result for exact designs when only some of the regression parameters are to be estimated. Fedorov mentions such a result for the continuous theory, but only in the case when the sequence of designs (which are intended for estimating only some of the parameters) converges to a design under which all the parameters are estimable.

Here, after proving the preliminary result that $d(x, \xi)\xi(x) \leqq 1$, we present some simple modifications of the sequence given by Fedorov for estimating all the parameters. In an example these modifications substantially increase the speed of convergence. We then prove that the analogous sequence for estimating some of the parameters converges to a D-optimal design, whether or not this limiting design allows all the parameters to be estimated. An interesting corollary of our methods is that an optimal design never assigns a point measure which is greater than the reciprocal of the number of parameters being estimated.

1. Terminology. We use the definitions and notation which have become standard. For fuller details, see almost any of the references listed. A controllable variable x can range over a space $\mathscr{X}$. On $\mathscr{X}$ are defined k regression functions $f_1, \cdots, f_k$, which we write as a column vector f. We can observe uncorrelated random variables Y_x, each with variance σ^2 and mean $f'(x)\theta$, for some unknown k-dimensional parameter θ. (Primes denote transposes.) A design

Received April 7, 1972.

ξ is a probability measure on $\mathscr{X}$. If ξ is supported on finitely many points and $\xi(x)$ is rational for all x, ξ is called exact and $\xi(x)$ is the proportion of observations to be taken at x. Otherwise we allocate observations to approximate the distribution given by ξ. Define $M(\xi)$ as the $k \times k$ matrix with ijth entry $\int f_i(x)f_j(x)\, d\xi(x)$.

Suppose we want to estimate θ. If ξ is exact then the covariance matrix of the best linear unbiased estimator of θ is proportional to $M^{-1}(\xi)$. Accordingly we call ξ D-optimal if it maximizes $\det M(\xi)$. The variance of the best linear unbiased estimator of $f'(x)\theta$ is proportional to $d(x, \xi) = f'(x)M^{-1}(\xi)f(x)$. Therefore we call ξ G-optimal if it minimizes $\bar{d}(\xi) = \max_x d(x, \xi)$, the maximum being taken over $\mathscr{X}$. Kiefer and Wolfowitz [4] showed that ξ is D-optimal if and only if it is G-optimal, and either condition is equivalent to $\max_x d(x, \xi) = k$. From the identity $\int d(x, \xi)\, d\xi(x) = k$, we conclude that the maximum (over the support of ξ) of $d(x, \xi)$ is $\geqq k$, and the minimum is $\leqq k$.

Now suppose instead that we want to estimate only some of the components of θ, say $\theta^{(1)}$ consisting of the first s components of θ. The covariance matrix of the best linear unbiased estimator of $\theta^{(1)}$ is proportional to $M^{(1)}$, the upper left $s \times s$ submatrix of $M^{-1}(\xi)$. (Use a limit if $M(\xi)$ is singular, as in [1] or [3].) Define $M^*(\xi) = [M^{(1)}(\xi)]^{-1}$. We call ξ D-optimal for estimating s out of k parameters if it maximizes $\det M^*(\xi)$. Let us define $r = k - s$ and define $M_r(\xi)$ as the lower right $r \times r$ submatrix of $M(\xi)$. Then it is not hard to show that

$$\det M(\xi) = \det M^*(\xi) \det M_r(\xi)\,. \tag{1.1}$$

If $M(\xi)$ is nonsingular, we can define

$$d_r(x, \xi) = f^{(2)\prime}(x)M_r^{-1}(\xi)f^{(2)}(x)\,,$$

where $f^{(2)}$ consists of the last r components of f. We then define

$$d_s(x, \xi) = d(x, \xi) - d_r(x, \xi)\,. \tag{1.2}$$

(The notation is compact but potentially misleading. The functions d_r and d_s correspond to estimating r out of r and s out of k parameters, respectively.) Kiefer [3] showed that if $M(\xi)$ is nonsingular, ξ is D-optimal if and only if $\max_x d_s(x, \xi)$ is minimized, and either condition is equivalent to $\max_x d_s(x, \xi) = s$. For any ξ the maximum (over the support of ξ) of $d_s(x, \xi)$ is $\geqq s$, and the minimum is $\leqq s$, paralleling the results for d.

For estimating θ or $\theta^{(1)}$ we will always assume that an optimal design exists. Sufficient conditions for this are that the regression functions are continuous and $\mathscr{X}$ is compact. By ξ^* we will always mean an optimal design.

2. Results. We first give a theorem which will be used frequently.

THEOREM 1. *If $M(\xi)$ is nonsingular, then for any x, $\xi(x)d(x, \xi) \leqq 1$.*

PROOF. The result is trivial if $\xi(x) = 0$. Fix x and suppose $\xi(x) > 0$. Let ε satisfy $0 < \varepsilon < \xi(x)$, and let $\alpha = \xi(x) - \varepsilon$. Define ξ_0 by $\xi_0(x) = \varepsilon/(1 - \alpha)$, and for $x' \neq x$, $\xi_0(x') = \xi(x')/(1-\alpha)$. Then ξ_0 is a probability measure with $M(\xi_0)$ nonsingular. Let ξ_x be concentrated at x. Then $\xi = (1-\alpha)\xi_0 + \alpha\xi_x$. By Theorem 2.6.1 of Fedorov [2],

$$d(x, \xi) = d(x, \xi_0)/[1 - \alpha + \alpha d(x, \xi_0)] . \tag{2.1}$$

Therefore

$$\begin{aligned}(\xi(x) - \varepsilon)d(x, \xi) &= \alpha d(x, \xi) \\ &= \alpha d(x, \xi_0)/[1 - \alpha + \alpha d(x, \xi_0)] \\ &< 1 .\end{aligned}$$

Taking the limit as $\varepsilon \to 0$, we obtain $\xi(x)d(x, \xi) \leqq 1$. □

There are two easy corollaries. The first is true because $d_s(x, \xi) \leqq d(x, \xi)$, the second because if ξ is optimal then $d(x, \xi) = k$ on points of support of ξ.

COROLLARY 1.1. *If $M(\xi)$ is nonsingular, $\xi(x)d_s(x, \xi) \leqq 1$.*

COROLLARY 1.2. *If ξ is optimal for θ, then $\xi(x) \leqq 1/k$ for all x.*

As Theorem 4 we will prove the analogue of Corollary 1.2 when estimating $\theta^{(1)}$. At present we can assert that $\xi(x) \leqq 1/s$ when ξ is optimal for $\theta^{(1)}$ only if we also assume that $M(\xi)$ is nonsingular.

We now summarize some results given by Fedorov [2], which we will also be using. Let ξ_n be any design, and let ξ_x be the design concentrated at one point x. Let $\xi_{n+1} = (1-\alpha)\xi_n + \alpha\xi_x$. This is a probability measure if it assigns nonnegative measure to every point, i.e. if $-\xi_n(x)/(1-\xi_n(x)) \leqq \alpha \leqq 1$. Note that if $\xi_n(x) > 0$ then we can make $\xi_{n+1}(x) < \xi_n(x)$ by making $\alpha < 0$. The details below will be slightly easier if we define $\beta = \alpha/(1-\alpha)$, so $-\xi_n(x) \leqq \beta \leqq \infty$. (It is only important to allow $\beta = \infty$, i.e. $\alpha = 1$, when $k = 1$. In that case the formulas below need minor rewriting whether we parametrize in terms of α or β. We leave this to the reader.) Then assuming that $M(\xi_n)$ is nonsingular we have

$$\det M(\xi_{n+1})/\det M(\xi_n) = (1+\beta)^{-k}(1 + \beta d(x, \xi_n)) . \tag{2.2}$$

The logarithm of (2.2) is defined if $\beta > -1$ and $\beta > -1/d(x, \xi_n)$. Therefore it is defined for β in $(-\xi_n(x), \infty)$, the interior of the domain of (2.2), using Theorem 1 to handle $\beta > -1/d(x, \xi_n)$. The derivative of the logarithm has a zero corresponding to a relative maximum of (2.2) at

$$\beta(x) = [d(x, \xi_n) - k][(k-1)d(x, \xi_n)]^{-1} . \tag{2.3}$$

This $\beta(x)$ is in the domain $[-\xi(x), \infty]$ if $d(x, \xi_n) \geqq k/(1 + (k-1)\xi_n(x))$. Otherwise the maximum is attained at $\beta = -\xi_n(x)$. The value of (2.2) attained at $\beta(x)$ is

$$[d(x, \xi_n)/k]^k[(k-1)/(d(x, \xi_n) - 1)]^{k-1} . \tag{2.4}$$

(When $k = 1$, (2.2) is maximized at $\beta = -\xi_n(x)$ or $\beta = \infty$, according as $d(x, \xi_n)$ is < 1 or > 1 respectively, and the maximum in the latter case is $d(x, \xi_n)$.)

Expression (2.4) is strictly increasing in $d(x, \xi_n)$ for $d(x, \xi_n) > k$, and strictly decreasing for $d(x, \xi_n) < k$. Thus we can make (2.4) large by maximizing $d(x, \xi_n)$. Fedorov therefore gives the following procedure. Let ξ_0 be any design with $M(\xi_0)$ nonsingular. At each step let x_0 be chosen to maximize $d(x, \xi_n)$, and let $\beta = \beta(x_0)$. Then Fedorov obtains the following theorem (his Theorem 2.5.3).

THEOREM 2 (Fedorov). *For the sequence of designs generated above*, $\det M(\xi_n) \to \det M(\xi^*)$ *monotonically, where* ξ^* *is optimal.*

Monotone convergence is not necessarily fast convergence. We now give a few refinements on the above procedure in order to improve the speed of convergence.

Since (2.4) is decreasing for $d(x, \xi_n) < k$, the expression may be maximized by choosing x_1 to minimize rather than maximize $d(x, \xi_n)$. The minimization is worthwhile over the support of ξ_n only, as will be seen below. In this case we have $\beta(x_1) \leqq 0$, and ξ_{n+1} is obtained by subtracting measure from ξ_n at x_1. So we use either x_1 and $\beta(x_1)$, assuming $\beta(x_1) \geqq -\xi_n(x_1)$, or x_0 and $\beta(x_0)$, whichever makes (2.2) larger. If $\beta(x_1) < -\xi_n(x_1)$, then we use either x_1 and $\beta = -\xi_n(x_1)$ or x_0 and $\beta(x_0)$, whichever makes (2.2) larger. In this last case we may conceivably do even better by considering x_2, the point of support of ξ_n where $d(x, \xi_n)$ is next smallest after x_1. The details are left to the reader. Allowing β to be negative can speed up convergence enormously, as will be discussed in connection with the example of Section 3.

A second refinement on the sequence is as follows. It may happen that $d(x, \xi_n)$ is maximized or minimized at several points simultaneously. This will typically happen if the model has some symmetry. In this case we can obtain several designs, say m such which we denote $\xi_{n+1,1}, \cdots, \xi_{n+1,m}$, such that $\det M(\xi_{n+1,i})$ is the same for all i. Since $-\log \det A$ is a convex function of A (mentioned for example in [3]), a convex combination of the $\xi_{n+1,i}$ will give a matrix M with $\det M \geqq \det M(\xi_{n+1,i})$ for any i. In fact it will be strictly greater unless the convex combination is of $\xi_{n+1,i}$ which all have exactly the same matrix $M(\xi_{n+1,i})$. Therefore rather than arbitrarily choosing one $\xi_{n+1,i}$, let $\xi_{n+1} = \sum_i m^{-1}\xi_{n+1,i}$.

Here is a third refinement, definitely less important than the other two. By Corollary 1.2, $\xi^*(x) \leqq 1/k$ for all x. So if we obtain $\xi_n(x) > 1/k$ for some x, we know that $\xi_n(x)$ is too large. Of course this can only happen if $k > 1$. Before investigating the extrema of $d(x, \xi_n)$ we can improve ξ_n by subtracting measure from ξ_n at x. Formally we choose $\alpha < 0$ such that $\xi_n'(x) = 1/k$, where $\xi_n' = (1 - \alpha)\xi_n + \alpha\xi_x$. Direct computation gives that the required α is $[1 - k\xi_n(x)]/[k - k\xi_n(x)]$. The corresponding β is

$$(1 - k\xi_n(x))/(k - 1)\,. \tag{2.5}$$

Clearly expression (2.5) is negative and $\geqq -\xi_n(x)$, and by Theorem 1 it is $\geqq \beta(x)$, where $\beta(x)$ is defined by (2.3). Inspection of the derivative of the logarithm of (2.2) shows that (2.2) is strictly decreasing in β for β both $\geqq \beta(x)$

and $\geqq -\xi_n(x)$. Also (2.2) equals 1 at $\beta = 0$. From this we conclude that if $\beta = (2.5)$ then (2.2) is > 1, so $\det M(\xi_n') > \det M(\xi_n)$. Thus we can replace ξ_n by ξ_n' without ever investigating the extrema of $d(x, \xi_n)$. This modification is probably not worth the trouble unless it is rather time consuming to look for the extrema of $d(x, \xi_n)$, since for most n $d(x, \xi_n)$ will probably be $\leqq 1/k$. Moreover even if we do use ξ_n' instead of ξ_n, we might get the same ξ_{n+1} in either case. In the example of Section 3, $\xi_2(A) > \frac{1}{3}$, but we get the same ξ_3 whether we base it on ξ_2 or ξ_2'.

The idea of the proof of Theorem 2 is that if for some $\varepsilon > 0$ and all n, $\det M(\xi_n) \leqq \det M(\xi^*) - \varepsilon$, then there is a $\delta > 0$ such that for all n,

$$\det M(\xi_{n+1})/\det M(\xi_n) > 1 + \delta ,$$

which is impossible. Each of the above improvements on the original sequence increases (possibly strictly) $\det M(\xi_{n+1})/\det M(\xi_n)$. Therefore the same proof works for our modified sequence. We state this formally.

COROLLARY 2.1. *For the sequence of designs generated above*, $\det M(\xi_n) \to \det M(\xi^*)$ *monotonically, where* ξ^* *is optimal.*

Let us now turn our attention to the case where we estimate s out of k parameters. Let ξ_n be any design for which $M(\xi_n)$ is nonsingular, and let x be any point. As before, let ξ_x be concentrated at x, let $\xi_{n+1} = (1 - \alpha)\xi_n + \alpha\xi_x$, and let $\beta = \alpha/(1 - \alpha)$. We want to choose α, or equivalently β, to maximize $\det M^*(\xi_{n+1})$. We summarize the relevant facts.

LEMMA. *Define* $\beta(x)$ *to be the larger solution, if solutions exist, of*

$$\beta^2 sdd_r + \beta(2sd_r + (s - 1)d_s) + s - d_s = 0 , \tag{2.6}$$

where d *means* $d(x, \xi_n)$, *etc. For* $\beta(x)$ *to exist it is sufficient, but not necessary, that* $d_s(x, \xi_n)$ *be* $\geqq s$, *and* $d_s(x, \xi_n) > s$ *if and only if* $\beta(x) > 0$. *If* $\beta(x)$ *exists and is* $\geqq -\xi_n(x)$, *then* $\det M^*(\xi_{n+1})$ *is maximized by* $\beta = \beta(x)$; *otherwise it is maximized by* $\beta = -\xi_n(x)$. *Finally, suppose* $\beta(x) \geqq -\xi_n(x)$. *Then* $d_s(x, \xi_{n+1}) = s$ *if and only if* $\beta = \beta(x)$.

PROOF. By (1.1) and (2.2)

$$\begin{aligned}
&\det M^*(\xi_{n+1})/\det M^*(\xi_n) \\
&\quad = [\det M(\xi_{n+1})/\det M_r(\xi_{n+1})]/[\det M(\xi_n)/\det M_r(\xi_n)] \\
&\quad = (1 + \beta)^{-k}(1 + \beta d(x, \xi_n))/[(1 + \beta)^{-r}(1 + \beta d_r(x, \xi_n))] \\
&\quad = (1 + \beta)^{-s}(1 + \beta d(x, \xi_n))/(1 + \beta d_r(x, \xi_n)) .
\end{aligned} \tag{2.7}$$

The logarithm of (2.7) is defined if $\beta > -1$, $\beta > -1/d(x, \xi_n)$ and $\beta > -1/d_r(x, \xi_n)$. Hence it is defined for $-\xi_n(x) < \beta < \infty$. The derivative of the logarithm is a negative quantity times the left side of (2.6). The left side of (2.6) is increasing at $\beta = 0$, and its sign changes as β goes from 0 to ∞ if and only if $d_s > s$. Thus (2.6) has a positive solution if and only if $d_s > s$, proving the first assertion. The next assertion is trivial if $d_r = 0$; assume that $d_r > 0$. If (2.6) has solutions then the smaller one is

$$-(2sd_r + (s-1)d_s)/(2sdd_r) - A$$

where A is nonnegative. But this is $\leqq -1/d$, hence $\leqq -\xi_n(x)$. So the only solution of (2.6) in $(-\xi_n(x), \infty)$ must be the larger root, $\beta(x)$. Since (2.7) approaches 0 as $\beta \to \infty$, the maximum must be at $\beta(x)$ if $\beta(x) > -\xi_n(x)$, and at $-\xi_n(x)$ otherwise. Finally, suppose $\beta(x) \geqq -\xi_n(x)$. Then $\beta = \beta(x)$ if and only if for our given x, det $M^*(\xi_{n+1})$ is maximized. This is equivalent to saying that if we use the same point x, and for some α' define

$$\xi_{n+2} = (1-\alpha')\xi_{n+1} + \alpha'\xi_x\,,$$

then det $M^*(\xi_{n+2})$ is maximized by $\alpha' = 0$, or equivalently $\beta' = 0$. This says that 0 must be a solution of (2.6), where now d means $d(x, \xi_{n+1})$, etc. But this is true if and only if $d_s(x, \xi_{n+1}) = s$. □

We are now ready to prove our main result. Let ξ_0 be any design such that $M(\xi_0)$ is nonsingular. Given any ξ_n, $n \geqq 0$, let x_n be chosen to maximize $d_s(x, \xi_n)$, and let $\beta = \beta(x_n)$. This determines ξ_{n+1}.

THEOREM 3. *For the sequence of designs given above*, det $M^*(\xi_n) \to$ det $M^*(\xi^*)$ *monotonically, where ξ^* is optimal for $\theta^{(1)}$.*

We could hitchhike on Wynn's proof [6]. By adding one point to an n-point design, he is taking $\beta = 1/n$ at the nth step. Thus we could argue that any convergence which can be obtained using $\beta = 1/n$ can be obtained faster using $\beta = \beta(x_n)$. This would work, but Wynn's problem has complications which are not present in our problem, and the proof we give below is therefore simpler than Wynn's proof.

PROOF. Since x_n maximizes $d_s(x, \xi_n)$ we have $d_s(x_n, \xi_n) \geqq s$. So when $\beta = \beta(x_n)$, det $M^*(\xi_{n+1})$/det $M^*(\xi_n)$ is maximized, hence it is $\geqq 1$. Therefore det $M^*(\xi_n)$ is a monotone increasing sequence.

A basic identity which we will use to prove convergence follows from (1.2) and (2.1). For any x, and any $\beta \geqq -\xi_n(x)$,

$$\begin{aligned} d_s(x, \xi_{n+1}) &= d(x, \xi_{n+1}) - d_r(x, \xi_{n+1}) \\ &= (1+\beta)d(x, \xi_n)/[1+\beta d(x, \xi_n)] \\ &\quad - (1+\beta)d_r(x, \xi_n)/[1+\beta d_r(x, \xi_n)] \\ &= (1+\beta)d_s(x, \xi_n)/[(1+\beta d(x, \xi_n))(1+\beta d_r(x, \xi_n))]\,. \end{aligned} \tag{2.8}$$

If we differentiate the logarithm of this last expression, we see that when $d(x, \xi_n) > 1$ the expression is monotone in β for $\beta \geqq 0$. Hence if $d_s(x, \xi_n) > s$, the expression is monotone in β for $\beta \geqq 0$. If $\beta = 0$, $d_s(x, \xi_{n+1}) = d_s(x, \xi_n)$, and by our lemma if $\beta = \beta(x) \geqq -\xi_n(x)$ then $d_s(x, \xi_{n+1}) = s$. (As an alternate proof of the last assertion of the lemma, we could observe that (2.8) reduces to (2.6) if and only if $d_s(x, \xi_{n+1}) = s$.)

Suppose now that $\det M^*(\xi_n)$ does not converge to $\det M^*(\xi^*)$, i.e. for some $\varepsilon > 0$ and all n, $\det M^*(\xi_n) \leqq \det M^*(\xi^*) - \varepsilon$. Then by Theorem 4.3 of [1] there is some $\delta > 0$ such that for all n, $\max_x d_s(x, \xi_n) \geqq s + \delta$. That is $d_s(x_n, \xi_n) \geqq s + \delta$.

For reasons which will become clear below we will choose $\delta > 0$ small enough that

$$(1 + \delta/s)^{\frac{1}{2}} \geqq 1 + \delta/4s . \tag{2.9}$$

Now define

$$\eta = \delta/(2s + \delta) . \tag{2.10}$$

Let $\beta' > 0$ be small enough so that if $0 \leqq \beta \leqq \beta'$ then

$$(1 - s\beta)(1 + s\beta(1 - \beta)(1 + \eta)) \geqq 1 + s\beta\eta/2 . \tag{2.11}$$

For each n, let β_n be $\beta(x_n)$ or β', whichever is smaller. Let $\alpha_n = \beta_n/(1 + \beta_n)$, and let $\xi' = (1 - \alpha_n)\xi_n + \alpha_n \xi_{x_n}$. Then

$$\det M^*(\xi_{n+1})/\det M^*(\xi_n) \geqq \det M^*(\xi')/\det M^*(\xi_n) .$$

By (2.8) and the remarks following it we have

$$\begin{aligned} s &\leqq d_s(x_n, \xi') \\ &= (1 + \beta_n)d_s(x_n, \xi_n)/[(1 + \beta_n d(x_n, \xi_n))(1 + \beta_n d_r(x_n, \xi_n))] \\ &\leqq (1 + \beta_n)d_s(x_n, \xi_n)/(1 + \beta_n d_r(x_n, \xi_n))^2 . \end{aligned}$$

Therefore

$$\begin{aligned} (1 + \beta_n d_r(x_n, \xi_n))^2 &\leqq (1 + \beta_n)d_s(x_n, \xi_n)/s \\ &= (1 + \beta_n)(1 + \delta_n/s) \end{aligned}$$

defining $\delta_n = d_s(x_n, \xi_n) - s$. Therefore

$$1 + \beta_n d_r(x_n, \xi_n) \leqq (1 + \beta_n)(1 + \delta_n/2s) .$$

From (2.7) we therefore have

$$\begin{aligned} &\det M^*(\xi_{n+1})/\det M^*(\xi_n) \\ &\quad \geqq \det M^*(\xi')/\det M^*(\xi_n) \\ &\quad = (1 + \beta_n)^{-s}(1 + \beta_n d(x_n, \xi_n))/[1 + \beta_n d_r(x_n, \xi_n)] \\ &\quad = (1 + \beta_n)^{-s}(1 + \beta_n d_s(x_n, \xi_n)/[1 + \beta_n d_r(x_n, \xi_n)]) \\ &\quad \geqq (1 - s\beta_n)(1 + \beta_n(s + \delta_n)/[(1 + \beta_n)(1 + \delta_n/2s)]) \\ &\quad \geqq (1 - s\beta_n)(1 + \beta_n(1 - \beta_n)(s + \delta_n)/(1 + \delta_n/2s)) \\ &\quad \geqq (1 - s\beta_n)(1 + \beta_n(1 - \beta_n)(s + \delta)/(1 + \delta/2s)) \end{aligned}$$

since by assumption $\delta_n \geqq \delta$. But by (2.10) this last expression equals

$$(1 - s\beta_n)(1 + s\beta_n(1 - \beta_n)(1 + \eta)) \geqq 1 + s\beta_n\eta/2$$

by (2.11) and our definition of β_n. Thus, defining $\eta' = s\eta/2$, we have

$$\det M^*(\xi_{n+1})/\det M^*(\xi_n) \geqq 1 + \eta'\beta_n .$$

Now there are two cases. Suppose first that $\{\bar{d}(\xi_n)\}$ is an unbounded sequence, where $\bar{d}(\xi_n)$ is defined as $\max_x d(x, \xi_n)$. From Theorem 2.6.1 of [2] we can conclude that

$$\begin{aligned}\bar{d}(\xi_{n+1}) = d(x_{n+1}, \xi_{n+1}) &\leqq (1 + \beta(x_n))d(x_{n+1}, \xi_n)\\ &\leqq (1 + \beta(x_n))\bar{d}(\xi_n) .\end{aligned}$$

Therefore

$$\bar{d}(\xi_{n+1}) \leqq \bar{d}(\xi_0) \textstyle\prod_0^n (1 + \beta(x_i))$$

or

$$\textstyle\prod_0^n (1 + \beta(x_i)) \geqq \bar{d}(\xi_{n+1})/\bar{d}(\xi_0) .$$

The right side can be made arbitrarily large by suitable choice of n. The left side is monotone in n, and therefore $\rightarrow \infty$. It is well known that for $\beta(x_i) \geqq 0$, the product $\prod (1 + \beta(x_i))$ converges if and only if $\sum \beta(x_i)$ converges, since $\sum \beta(x_i) \leqq \prod (1 + \beta(x_i)) \leqq \exp(\sum \beta(x_i))$. So in our case since $\prod (1 + \beta(x_i))$ diverges we conclude that $\sum \beta(x_i)$ diverges and therefore $\prod (1 + \eta'\beta(x_i))$ diverges. Now either $\beta_i = \beta'$ for infinitely many i, in which case $\prod (1 + \eta'\beta_i)$ diverges, or else $\beta_i = \beta(x_i)$ for all but finitely many i, in which case $\prod_0^n (1 + \eta'\beta_i)$ is a constant times $\prod_0^n (1 + \eta'\beta(x_i))$ for n large. Therefore $\prod (1 + \eta'\beta_i)$ must diverge.

Now suppose that $\bar{d}(\xi_n)$ is bounded by some constant A, for all n. By (2.8), for $x = x_n$ and $\beta = \beta(x_n)$ we have

$$\begin{aligned}s &= (1 + \beta(x_n))d_s(x_n, \xi_n)/[(1 + \beta(x_n)d(x_n, \xi_n))(1 + \beta(x_n)d_r(x_n, \xi_n))]\\ &\geqq d_s(x, \xi_n)/(1 + \beta(x_n)d(x_n, \xi_n))^2 .\end{aligned}$$

Therefore

$$\begin{aligned}(1 + \beta(x_n)d(x_n, \xi_n))^2 &\geqq d_s(x_n, \xi_n)/s\\ &\geqq 1 + \delta/s .\end{aligned}$$

Therefore by (2.9),

$$1 + \beta(x_n)d(x_n, \xi_n) \geqq 1 + \delta/4s .$$

Since $d(x_n, \xi_n) \leqq A$, we have

$$\beta(x_n) \geqq \delta/(4sA) \qquad \text{for all} \quad n .$$

We conclude that $\beta_n \geqq \min(\delta/(4sA), \beta') > 0$, for all n. Therefore in this case $\prod (1 + \eta'\beta_i)$ diverges.

Thus in either case

$$\det M^*(\xi_n) \geqq \det M^*(\xi_0) \textstyle\prod_0^{n-1} (1 + \eta'\beta_i) \rightarrow \infty .$$

But this is impossible, since $\det M^*(\xi_n) \leqq \det M^*(\xi^*)$ for all n. Therefore we must have $\det M^*(\xi_n) \rightarrow \det M^*(\xi^*)$. □

We modified the sequence of Theorem 2 in three ways to speed up the convergence. These modifications can also be applied to the sequence developed here for Theorem 3.

First, there is no reason why we should get the largest value of

$\det M^*(\xi_{n+1})/\det M^*(\xi_n)$ by choosing x_n to maximize $d_s(x, \xi_n)$. When estimating θ we had to choose between maximizing and minimizing $d(x, \xi_n)$. When estimating $\theta^{(1)}$ there seem to be many choices, since $\det M^*(\xi_{n+1})/\det M^*(\xi_n)$ depends in a complicated way on both $d_s(x, \xi_n)$ and $d_r(x, \xi_n)$. However we can certainly consider at least the points where $d_s(x, \xi_n)$ is maximized and minimized, and other points if it is not too inconvenient. For example, if we know that an optimal design is supported on a small given set of points, we might consider all these points. For each x considered the β to use is given by our lemma. Then we select that x from the points considered which gives the largest value of (2.7). A practical difficulty is that if we remove all the measure at x, this could conceivably make $M_r(\xi_{n+1})$ singular; this will happen if $1 + \beta d_r(x, \xi_n) = 0$. Therefore in this case the computer should be programmed to leave a tiny measure at x.

Second, if the model possesses some symmetry we may have several points $x_1, \cdots, x_m$ such that $d_s(x_i, \xi_n)$ is the same for all i, and $d_r(x_i, \xi_n)$ is the same for all i. If $\xi_{n+1,i} = (1-\alpha)\xi_n + \alpha\xi_{x_i}$, where α maximizes $\det M^*(\xi_{n+1,i})$, we would do better to let $\xi_{n+1} = \sum_i m^{-1}\xi_{n+1,i}$. This follows from the convexity of $-\log\det M^*(\xi)$ — see (2.12) of [3].

Third, suppose $\xi_n(x) > 1/s$. Let us choose $\alpha < 0$ such that $\xi_n'(x) = 1/s$, where $\xi_n' = (1-\alpha)\xi_n + \alpha\xi_x$. The corresponding β is

$$(1 - s\xi_n(x))/(s-1)\,. \tag{2.12}$$

Clearly this is $> -\xi_n(x)$. We want to show that $\det M^*(\xi_n') \geqq \det M^*(\xi_n)$. Since (2.6) has at most one solution $> -\xi_n(x)$, (2.7) must be monotone between 0 and $\beta(x)$. Because $\xi_n(x) > 1/s$, we have $d_s(x, \xi_n) < s$, by Corollary 1.1, so by our lemma, $\beta(x) < 0$. Since $\xi_n'(x) = 1/s$, we have $d_s(x, \xi_n') \leqq s$, so we would improve on ξ_n' by subtracting still more measure at x. Therefore the β given by (2.12) is $\geqq \beta(x)$, and of course < 0. By the monotonicity of (2.7) we conclude that $\det M^*(\xi_n') > \det M^*(\xi_n)$, so we shall replace ξ_n by ξ_n'.

If we use the sequence of designs given above, then the proof of Theorem 3 goes through word for word, except for trivial modification of the first paragraph of the proof. Therefore:

COROLLARY 3.1. *For the sequence of designs given above,* $\det M^*(\xi_n) \to \det M^*(\xi^*)$ *monotonically, where* ξ^* *is optimal for* $\theta^{(1)}$.

We are now in a position to give the analogue of Corollary 1.2.

THEOREM 4. *If* ξ *is optimal for* $\theta^{(1)}$, *then* $\xi(x) \leqq 1/s$ *for all* x.

As was mentioned after Corollary 1.2, the result is easy if $M(\xi)$ is nonsingular. We give here a proof which is valid even if $M(\xi)$ is singular.

PROOF. Suppose ξ is a design with $\xi(x) > 1/s$ for some x. Call $\xi(x) = (1+2h)/s$. Let $\xi' = (1-\alpha)\xi + \alpha\xi_x$, where $\alpha < 0$ is such that $\xi'(x) = (1+h)/s$. The idea of the proof is as follows. We will define designs ξ_ε and ξ_ε' which are close to ξ and ξ' respectively, show that ξ_ε' is better than ξ_ε, and conclude that ξ' is better than ξ.

Let ξ_0 be any design such that $\xi_0(x) = (1 + h)/s$ and $M(\xi_0)$ is nonsingular. For ε with $0 < \varepsilon < 1$, define $\xi_\varepsilon' = (1 - \varepsilon)\xi' + \varepsilon\xi_0$. Now $M(\xi_\varepsilon')$ is nonsingular and $\xi_\varepsilon'(x) = (1 + h)/s$, so by Corollary 1.1 $d_s(x, \xi_\varepsilon') \leqq s/(1 + h) < s$. Let us consider

$$\xi_\varepsilon = (1 - \alpha')\xi_\varepsilon' + \alpha'\xi_x .$$

At $\alpha' = 0$ the derivative of $\log[\det M^*(\xi_\varepsilon)/\det M^*(\xi_\varepsilon')]$ equals $d_s(x, \xi_\varepsilon') - s$. By the concavity of $\log\det M^*$ (see (2.12) of [3]), at any α'

$$\begin{aligned} \log\det M^*(\xi_\varepsilon) &\leqq \log\det M^*(\xi_\varepsilon') + \alpha'(d_s(x, \xi_\varepsilon') - s) \\ &= \log\det M^*(\xi_\varepsilon') - \alpha' hs/(1 + h) . \end{aligned} \tag{2.13}$$

Now choose $\alpha' = -\alpha/(1 - \alpha) > 0$, where α was defined above. Then

$$\xi_\varepsilon = (1 - \varepsilon)\xi + \varepsilon[(1 - \alpha')\xi_0 + \alpha'\xi_x] .$$

Take the limit of (2.13) as $\varepsilon \to 0$. Then we have

$$\begin{aligned} \log\det M^*(\xi) &\leqq \log\det M^*(\xi') - \alpha' hs/(1 + h) \\ &< \log\det M^*(\xi') . \end{aligned}$$

So ξ is not optimal. □

3. An example. Wynn [6] considers the following example:

$$f'(x)\theta = \theta_0 + \theta_1 x_1 + \theta_2 x_2 , \quad x = (x_1, x_2)$$

and $\mathscr{X}$ is the closed convex quadrilateral with vertices $A = (2, 2)$, $B = (-1, 1)$, $C = (1, -1)$ and $D = (-1, -1)$. He begins with one observation each at B, C and D. After successively adding 29 more points he obtains an optimal design for estimating θ: $\xi(A) = 10/32$, $\xi(B) = \xi(C) = 9/32$, $\xi(D) = 4/32$. For comparison we will consider the same example, and begin with the same design ξ_0, uniform on B, C and D. We use the sequence of Corollary 2.1.

Table 1 summarizes the results. Somewhere between steps 3 and 7 most people would decide, based on $\bar{d}$, that they are satisfied with the accuracy obtained. Although we naturally do not get the exact fractional form for $\xi^*(x)$ which

TABLE 1

n	$\xi_n(A)$	$\xi_n(B)$	$\xi_n(C)$	$\xi_n(D)$	$\det M(\xi_n)$	$\bar{d}(\xi_n)$	Point to be changed for ξ_{n+1}	β to get ξ_{n+1}
0	.0000	.3333	.3333	.3333	0.59259	25.5000	A	.4412
1	.3061	.2313	.2313	.2313	2.42516	3.2725	D	−.1110
2	.3443	.2602	.2602	.1353	2.51110	3.1756	A	−.0485
3	.3109	.2734	.2734	.1422	2.52838	3.0276	D	−.0183
4	.3167	.2785	.2785	.1262	2.53089	3.0216	A	−.0064
5	.3123	.2803	.2803	.1270	2.53120	3.0029	D	−.0022
6	.3130	.2809	.2809	.1251	2.53124	3.0024	A	−.0007
7	.3125	.2811	.2811	.1252	2.53124	3.0003	D	−.0002
∞	.3125	.28125	.28125	.1250	2.53125	3.0000		

Wynn obtains, if we should happen to have 32 observations our ξ_n rounds off to the correct fraction for $n \geqq 4$. From step 5 onwards any of our ξ_n is better than any design which Wynn obtains before getting the exact optimal design using 32 points.

If we instead use the sequence given by Theorem 2, with β always positive, the procedure soon settles down to adding minute measure alternately at A, B and C. By step 30 we have $\bar{d}(\xi_{30}) = 3.031$. Thus the sequence of Corollary 2.1 does better, in the $\bar{d}$ sense, in 3 steps than the sequence of Theorem 2 does in 30 steps.

This example shows dramatically the improvement which Corollary 2.1 can make over Theorem 2. Of course there is no guarantee that in every problem the sequence of Corollary 2.1 will arrive at a particular value of det $M(\xi)$ before the sequence of Theorem 2 does. At each step we have tried to choose a good ξ_{n+1}, given ξ_n. This does not necessarily improve the long term rate of convergence, although in most examples it probably does improve it.

Anyone who is constructing an optimal design using the procedures mentioned here should consider the recommendations and useful formulas given by Fedorov in Section 2.6 of [2].

REFERENCES

[1] ATWOOD, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist.* **40** 1570–1602.

[2] FEDOROV, V. V. (1972). *Theory of Optimal Experiments* (Tr. and ed. by E. M. Klimko and W. J. Studden). Academic Press, New York.

[3] KIEFER, J. (1961). Optimum designs in regression problems, II. *Ann. Math. Statist.* **32** 298–325.

[4] KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **14** 363–366.

[5] WYNN, H. P. (1970). The sequential generation of *D*-optimum experimental designs. *Ann. Math. Statist.* **41** 1655–1664.

[6] WYNN, H. P. (1972). Results in the theory and construction of *D*-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **34**.

MATHEMATICS DEPARTMENT
HAILE SELLASSIE I UNIVERSITY
P.O. BOX 1176
ADDIS ABABA, ETHIOPIA