The Annals of Statistics
1976, Vol. 4, No. 6, 1124–1138

CONVERGENT DESIGN SEQUENCES, FOR SUFFICIENTLY REGULAR OPTIMALITY CRITERIA

BY CORWIN L. ATWOOD

University of California at Davis

For an optimality criterion function Φ and a design ξ_n, approximate $\Phi(\mathbf{M}(\xi))$ near ξ_n by a quadratic Taylor expansion, let $\xi_n + \eta$ minimize this approximation, and let $\xi_{n+1} = \xi_n + \alpha\eta$, with α minimizing $\Phi(\mathbf{M}(\xi_{n+1}))$. If Φ satisfies regularity conditions, including strict convexity, possession of three continuous derivatives, and finiteness only for nonsingular $\mathbf{M}$, then $\mathbf{M}(\xi_n)$ converges to the optimal value for both the Fedorov steepest descent sequence and the above quadratic sequence, with the quadratic sequence having a faster asymptotic convergence rate. Methods are discussed for collapsing clusters of design points during the iterative process. In a simple example with D-optimality, the two methods are comparable. In a more complicated example the quadratic method is far superior.

1. Introduction and summary. For D-optimality and L-optimality, iterative algorithms have been given for finding optimal designs. The first were the steepest descent procedures of Fedorov [3] and Wynn [17], at each iteration adding mass to the design at a single point. Improvements have been proposed in [1] and [8], which are not essential departures from the steepest descent method. These procedures are also discussed and extended in [18], [13] and [14]. Fedorov and Malyutov [5], Whittle [16] and Kiefer [7] have pointed out that steepest descent methods can also be used for more general optimality criteria. Some essentially different iterative methods have been proposed for D-optimality by Sibson [9] and Silvey and Titterington [10], and for more general optimality by Gribik and Kortanek [6]. However these generally involve mathematical programming problems at each iteration, at least if the optimal support points are unknown, and their convergence speeds have not yet been considered in print. See [12] and [11] for further discussion.

A general failing of steepest descent methods is that they can converge very slowly. Quadratic methods, or "generalized Newton methods," in which the function to be minimized is approximated locally by a second degree polynomial, are commonly used as successful improvements to steepest descent methods, as discussed e.g., in [15].

In this paper a specific quadratic iterative method is proposed. For sufficiently regular optimality criteria, the Fedorov steepest descent sequence and the quadratic sequence are both shown to converge to an optimal design. The asymptotic convergence rates are found, and the improvement in a single iteration is seen

Received October 1975; revised May 1976.
AMS 1970 *subject classifications*. Primary 62K05; Secondary 65B99.
Key words and phrases. Optimal experimental designs, D-optimality, L-optimality, Φ-optimality, steepest descent, generalized Newton method.

to be asymptotically at least as great for the quadratic method as for the steepest descent method.

Specific comments are made concerning the application of these methods to D- and L-optimality. Methods are discussed for avoiding clusters of points, each point having small mass. Some slight extensions of Fedorov's lemmas on L-optimality are given. Finally we report the results of trying the steepest descent and quadratic methods in two examples. In the simpler example the execution times are comparable for the two methods. In the more complicated example the steepest descent sequence shows little hope of converging in any reasonable time, whereas the quadratic sequence converges fairly quickly.

2. Main results. We will assume the usual regression model, with $\mathbf{f} = (f_1, \cdots, f_k)^T$ a vector of real valued functions on a euclidean space $\mathscr{X}$. Let Ξ be the class of probability measures on $\mathscr{X}$, and for $\xi \in \Xi$ define the information matrix $\mathbf{M}(\xi) = \int \mathbf{f}(x)\mathbf{f}^T(x)\xi(dx)$. For an interpretation of these definitions, see [17] or [3].

We will also define $\mathscr{M} = \{\mathbf{M}(\xi) \mid \xi \in \Xi\}$, let $\mathscr{M}^+$ denote the set of nonsingular members of $\mathscr{M}$, and let $\mathscr{R}$ denote the set of all $k \times k$ matrices, with the euclidean topology. We will use $\|\ \|$ to denote an arbitrary but fixed norm on $\mathscr{R}$. Note that if η is any finite signed measure on $\mathscr{X}$, then $\mathbf{M}(\eta) = \int \mathbf{f}(x)\mathbf{f}^T(x)\eta(dx)$ is defined and in $\mathscr{R}$, although not necessarily in $\mathscr{M}$. As a convention, the letter ξ will denote an element of Ξ, and the letters η and ζ will denote signed measures. The letter $\mathbf{M}$ will denote an element of a neighborhood of $\mathscr{M}$, while letters such as $\mathbf{A}$ and $\mathbf{B}$ will denote arbitrary elements of $\mathscr{R}$.

Let Φ be a function which is real or $+\infty$ on $\mathscr{M}$. This function will be the optimality criterion function, and ξ^* will be called Φ-optimal if it minimizes $\Phi(\mathbf{M}(\xi))$. Two examples which satisfy the regularity assumptions to be given below are D-optimality:

$$\Phi(\mathbf{M}) = -\log |\mathbf{M}| ,$$

and L-optimality for nonsingular $\mathbf{C}$:

$$\Phi(\mathbf{M}) = \operatorname{tr} \mathbf{C}\mathbf{M}^{-1} ,$$

where $\mathbf{C}$ is given, symmetric and positive definite.

ASSUMPTIONS.

1. $\mathbf{f}$ is continuous and $\mathscr{X}$ is compact.
2. Φ is defined (possibly $+\infty$) and continuous on $\mathscr{M}$, and is defined and real-valued on $\mathscr{N}$, a neighborhood in $\mathscr{R}$ of $\mathscr{M}^+$.
3. The first three partial derivatives of Φ exist and are continuous in $\mathscr{N}$, with

$$\partial\Phi(\mathbf{M} + \alpha\mathbf{A})/\partial\alpha|_{\alpha=0} ,$$

$$\partial^2\Phi(\mathbf{M} + \alpha\mathbf{A} + \beta\mathbf{B})/\partial\alpha\, \partial\beta|_{\alpha=0,\beta=0}$$

and

$$\partial^3\Phi(\mathbf{M} + \alpha\mathbf{A} + \beta\mathbf{B} + \gamma\mathbf{C})/\partial\alpha\, \partial\beta\, \partial\gamma|_{\alpha=0,\beta=0,\gamma=0}$$

linear in $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, for $\mathbf{M} \in \mathscr{N}$, $\mathbf{A} \in \mathscr{R}$, $\mathbf{B} \in \mathscr{R}$, $\mathbf{C} \in \mathscr{R}$.

4. For all $\mathbf{M} \in \mathscr{M}^+$, and symmetric $\mathbf{A} \in \mathscr{R}$, $\mathbf{A} \neq \mathbf{0}$, we have

$$\partial^2\Phi(\mathbf{M} + \alpha\mathbf{A})/\partial\alpha^2|_{\alpha=0} > 0 .$$

5. $\Phi(\mathbf{M}) = \infty$ for singular $\mathbf{M} \in \mathscr{M}$.

Some comments on the assumptions are in order. Assumption 1 is standard. It implies that $\mathscr{M}$ is compact, so $\Phi(\mathbf{M})$ attains its minimum at some $\mathbf{M}^* = \mathbf{M}(\xi^*)$. Assumptions 2 and 3 are similar to those of [5], [16], [7] and [6]. Those papers only assume one derivative. However, most optimality criteria are differentiable infinitely often if they are differentiable at all, so in practice Assumption 3 is not as restrictive as it appears. All the results of the present paper can also be proved if, instead of 2 and 3, analogous assumptions are made on $\Phi^{[S]}$, defined as the restriction of Φ to symmetric matrices. The verification of this is routine.

Assumption 4 implies strict convexity of Φ on $\mathscr{M}^+$, but is slightly stronger. Strict convexity implies that the optimal $\mathbf{M}$ is unique. The stronger condition will be needed to define the quadratic sequence, and to get the asymptotic convergence rates of Theorem 2.2. Assumption 3 implies the existence of the derivative in Assumption 4. Note also that for any α' and any $\mathbf{A} \in \mathscr{R}$ and $\mathbf{M}' \in \mathscr{R}$ such that $\mathbf{M}' + \alpha'\mathbf{A} \in \mathscr{N}$, we have

$$\partial\Phi(\mathbf{M}' + \alpha\mathbf{A})/\partial\alpha|_{\alpha=\alpha'} = \sum a_{ij}\, \partial\Phi(\mathbf{M})/\partial m_{ij}|_{M=M'+\alpha' A}$$

and similarly the higher order derivatives exist and can be expressed as sums of products of elements of $\mathbf{A}$ and partial derivatives of Φ. In particular, the derivatives exist and are continuous in $\mathbf{M}$, $\mathbf{A}$ and α for $\mathbf{M} \in \mathscr{N}$, $\mathbf{A} \in \mathscr{R}$ and $\alpha = 0$.

Assumption 5 is a real restriction. Results when Assumption 5 is not satisfied will be given in a sequel.

We record here a fact mentioned by Kiefer [7, equation (6.5)]: If Φ is convex, as it is in our case, and ξ_x is supported at x, then for any $\mathbf{M} \in \mathscr{M}^+$

$$\min_x \partial\Phi(\mathbf{M} + \alpha[\mathbf{M}(\xi_x) - \mathbf{M}])/\partial\alpha|_{\alpha=0} \leqq \Phi(\mathbf{M}^*) - \Phi(\mathbf{M}) . \tag{2.1}$$

Here $\mathbf{M}^*$ is the optimal $\mathbf{M}$.

One more piece of machinery is useful. Suppose $\mathbf{M} \in \mathscr{N}$. For $\mathbf{M} + \alpha||\mathbf{A}||^{-1}\mathbf{A} \in \mathscr{N}$, define

$$K(\alpha; \mathbf{A}, \mathbf{M}) = \Phi(\mathbf{M} + \alpha||\mathbf{A}||^{-1}\mathbf{A}) - \Phi(\mathbf{M}) .$$

This is the increment in Φ obtained by moving a distance α in the direction $\mathbf{A}$. Here $\mathbf{A}$ has been standardized by dividing by its norm. If $\mathbf{A} \neq \mathbf{0}$, the nth derivative with respect to α is given at $\alpha = \alpha'$ by

$$K^{(n)}(\alpha'; \mathbf{A}, \mathbf{M}) = ||\mathbf{A}||^{-n}\partial^n\Phi(\mathbf{M} + \alpha\mathbf{A})/\partial\alpha^n|_{\alpha=\alpha'||A||^{-1}} . \tag{2.2}$$

This exists and is continuous for $n = 1, 2, 3$ and $\mathbf{M} + \alpha'||\mathbf{A}||^{-1}\mathbf{A} \in \mathscr{N}$, as mentioned in the discussion of Assumption 4. Therefore for $\mathbf{M} \in \mathscr{N}$ and $\mathbf{A} \in \mathscr{R}$, we have that $K^{(n)}(0; \mathbf{A}, \mathbf{M})$ exists and is continuous in $\mathbf{M}$ and $\mathbf{A}$. Moreover, since K is unchanged if $\mathbf{A}$ is multiplied by a constant, for $\mathbf{M} \in \mathscr{N}$ all possible values of $K^{(n)}(0; \mathbf{A}, \mathbf{M})$ are attained on $\{\mathbf{A} \in \mathscr{R} \mid ||\mathbf{A}|| = 1\}$.

LEMMA 2.1. *Let $\mathscr{M}_0$ be a subset of $\mathscr{M}^+$ which is closed in $\mathscr{M}$. Under Assumptions* 1–4, *there exist lower and upper bounds L_2, U_2, L_3 and U_3 such that for all* $\mathbf{M} \in \mathscr{M}_0$ *and symmetric* $\mathbf{A} \in \mathscr{R}$, $\mathbf{A} \neq \mathbf{0}$

$$0 < L_2 \leqq K''(0; \mathbf{A}, \mathbf{M}) \leqq U_2 ,$$
$$L_3 \leqq K'''(0; \mathbf{A}, \mathbf{M}) \leqq U_3 .$$

Without loss of generality U_2 and U_3 are positive and L_3 is negative.

PROOF. We have seen that $K^{(n)}(0; \mathbf{A}, \mathbf{M})$ is continuous in $\mathbf{M}$ and $\mathbf{A}$ and takes all its possible values when $||\mathbf{A}|| = 1$. Since $\mathscr{M}_0$ is compact and $\{\mathbf{A} \mid \mathbf{A}$ symmetric, $||\mathbf{A}|| = 1\}$ is compact, it follows that $K^{(n)}(0; \mathbf{A}, \mathbf{M})$ is bounded above and below. Since $K''(0; \mathbf{A}, \mathbf{M})$ attains its minimum but by Assumption 4 is never 0, the lower bound L_2 may be taken to be positive. □

We must now explicitly define the design sequences. They both begin with some $\mathbf{M}_0 = \mathbf{M}(\xi_0) \in \mathscr{M}^+$. Having obtained $\mathbf{M}_n = \mathbf{M}(\xi_n)$, let x_n minimize

$$\partial\Phi(\mathbf{M}_n + \alpha[\mathbf{M}(\xi_x) - \mathbf{M}_n])/\partial\alpha|_0$$

where ξ_x is concentrated at x. The steepest descent procedure sets $\xi_{n+1} = \xi_n + \alpha_n(\xi_{x_n} - \xi_n)$, with α_n chosen to minimize $\Phi(\mathbf{M}(\xi_{n+1}))$.

For the quadratic sequence, let ξ_0 be supported on a finite set. At the nth iteration let $\mathscr{X}_n$ be a finite set which contains x_n and the support of ξ_n. For ξ supported on $\mathscr{X}_n$, consider $\Phi(\mathbf{M}(\xi))$ as a function of m variables, namely the m values $\xi(y_i)$ for $y_i \in \mathscr{X}_n$. Write $\boldsymbol{\xi} = \boldsymbol{\xi}_n + \boldsymbol{\eta}$ where $\boldsymbol{\xi}$ is an m-dimensional column vector with components $\xi(y_i)$, and write $\Phi(\mathbf{M}(\xi)) = \psi(\boldsymbol{\xi})$. The second degree Taylor expansion of $\psi(\boldsymbol{\xi}_n + \boldsymbol{\eta})$ is

$$\psi(\boldsymbol{\xi}_n) + \boldsymbol{\eta}^T\mathbf{g} + \tfrac{1}{2}\boldsymbol{\eta}^T\mathbf{H}\boldsymbol{\eta} \tag{2.3}$$

where the gradient $\mathbf{g}$ is the m-dimensional column vector with ith element $\partial\psi(\boldsymbol{\xi}_n + \alpha\boldsymbol{\xi}_{y_i})/\partial\alpha|_{\alpha=0}$ and the Hessian $\mathbf{H}$ is the $m \times m$ symmetric matrix with ijth element

$$\partial^2\psi(\boldsymbol{\xi}_n + \alpha\boldsymbol{\xi}_{y_i} + \beta\boldsymbol{\xi}_{y_j})/\partial\alpha\,\partial\beta|_{\alpha=0,\beta=0}$$

for y_i and y_j in $\mathscr{X}_n$. As always, ξ_{y_i} is the design supported on y_i, and $\boldsymbol{\xi}_{y_i}$ is the corresponding m-vector.

Expression (2.3) is to be minimized with respect to $\boldsymbol{\eta}$, subject to the constraint that $\sum \eta(y_i) = 0$, i.e., $\boldsymbol{\eta}^T\mathbf{e} = 0$, where $\mathbf{e}$ is the column vector consisting of m 1's. To do this let λ be a Lagrange multiplier and set the usual derivatives equal to zero:

$$\begin{aligned} \mathbf{g} + \mathbf{H}\boldsymbol{\eta} - \lambda\mathbf{e} &= \mathbf{0} , \\ \boldsymbol{\eta}^T\mathbf{e} &= 0 . \end{aligned} \tag{2.4}$$

By Assumption 4, $\boldsymbol{\eta}^T\mathbf{H}\boldsymbol{\eta} > 0$ for all $\eta \neq 0$, i.e., $\mathbf{H}$ is invertible. Multiply (2.4) on the left by $\mathbf{e}^T\mathbf{H}^{-1}$, yielding

$$\mathbf{e}^T\mathbf{H}^{-1}\mathbf{g} + 0 - \lambda\mathbf{e}^T\mathbf{H}^{-1}\mathbf{e} = 0 .$$

Therefore the solution is

$$\boldsymbol{\eta} = \mathbf{H}^{-1}(-\mathbf{g} + \lambda\mathbf{e}), \quad \lambda = (\mathbf{e}^T\mathbf{H}^{-1}\mathbf{g})/(\mathbf{e}^T\mathbf{H}^{-1}\mathbf{e}). \tag{2.5}$$

Since expression (2.3) is convex, this stationary point gives the true minimum of (2.3). Since also $\boldsymbol{\eta}$ was chosen to minimize this strictly convex expression, it follows that $\boldsymbol{\eta}^T\mathbf{g}$, i.e.,

$$\partial\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\eta))/\partial\alpha|_0,$$

is strictly negative for nonoptimal $\mathbf{M}_n$.

Subject to qualifications to be given below, write η_n for the η of (2.5), and define $\xi_{n+1} = \xi_n + \alpha_n\eta_n$, with $\alpha_n > 0$ chosen to minimize $\Phi(\mathbf{M}(\xi_{n+1}))$.

The first qualification is that if for some y, $\eta_n(y) < 0$ while $\xi_n(y) = 0$, then $\xi_n + \alpha\eta_n$ is not a design for any $\alpha > 0$, since it assigns negative mass to y. Suppose that there is exactly one such y. We must go back and try to minimize (2.3) subject to the same constraints as before and the additional constraint $\eta(y) \geqq 0$. By the convexity of (2.3) and the fact that the original η_n minimizes (2.3), it is clear that the minimum subject to $\eta(y) \geqq 0$ occurs only when $\eta(y) = 0$. So the desired η can be found by deleting y from $\mathscr{X}_n$ and solving (2.4) as before, working now in $m - 1$ rather than m dimensions. If there are several such y then it is not obvious which y must be deleted, but ways to proceed are easy to devise.

The other qualification to the definition of η_n is a theoretical technicality. To prove Theorems 2.1 and 2.2, it is necessary that for n large, α_n should be in the interior of the range of possible α, i.e., the range of α for which $\xi_n(y) + \alpha\eta_n(y) \geqq 0$ for all y. To guarantee this, we require that there exist a constant $C > 0$ such that for all y and for n sufficiently large, η_n satisfies

$$||\mathbf{M}(\eta)||\xi_n(y) \geqq -C\eta(y). \tag{2.6}$$

This inequality is trivial if $\eta(y) \geqq 0$. In principle we must find η_n subject to this additional constraint, which we hereby incorporate into the definition of η_n. Observe that there exist possible η satisfying (2.6), and one such η is $\xi_x - \xi_n$ for arbitrary x. For $\mathbf{M}(\xi_x)$ is singular and $\mathbf{M}(\xi_n)$ is in

$$\mathscr{M}_0 = \{\mathbf{M} \in \mathscr{M} \mid \Phi(\mathbf{M}) \leqq \Phi(\mathbf{M}_0)\}. \tag{2.7}$$

So $||\mathbf{M}(\xi_x - \xi_n)||$ is at least the minimum distance from an element of $\mathscr{M}_0$ to a singular element of $\mathscr{M}$. If C is less than this number then $\xi_x - \xi_n = \eta$ satisfies (2.6).

In normal practice the constraint (2.6) can probably be ignored, since C can be set arbitrarily small, and since also we will recommend below that nearby support points be combined, so that $\xi_n(y)$ remains fairly large at all support points y. Condition (2.6) must be invoked explicitly only if the sequence is failing to converge satisfactorily because the best possible α_n is consistently the right end point of the range of possible α. I do not know if such a situation can actually arise. It does not arise in the examples of Section 4.

Note that if η satisfies (2.6) for all y, and $\eta(y) \geqq 0$ for y with $\xi_n(y) = 0$, then the same is true for $\alpha\eta$ with $\alpha \geqq 0$. That is, the two conditions refer to the direction rather than the magnitude of η. For any η satisfying these conditions, (2.3) is smallest at that $\eta' = \alpha\eta$ which minimizes

$$\alpha\boldsymbol{\eta}^T\mathbf{g} + \tfrac{1}{2}\alpha^2\boldsymbol{\eta}^T\mathbf{H}\boldsymbol{\eta} .$$

The minimum value is

$$(2.8) \qquad \begin{aligned} -[\boldsymbol{\eta}^T\mathbf{g}]^2/2\boldsymbol{\eta}^T\mathbf{H}\boldsymbol{\eta} &= -[\partial\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\eta))/\partial\alpha]^2/2[\partial^2\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\eta))/\partial\alpha^2] \\ &= -[K_n'(0; \mathbf{M}(\eta), \mathbf{M}_n)]^2/2K_n''(0; \mathbf{M}(\eta), \mathbf{M}_n) \end{aligned}$$

with the derivatives of Φ evaluated at $\alpha = 0$. Thus η_n has been found to minimize (2.8) among all η supported on $\mathscr{X}_n$ such that (2.6) is satisfied for all y and $\eta(y) \geqq 0$ when $\xi_n(y) = 0$. In particular, (2.8) is at least as small for $\eta = \eta_n$ as for $\eta = \xi_{x_n} - \xi_n$.

It will be shown below that (2.8) is the asymptotic change in Φ at the nth iteration. It follows that the improvement is asymptotically at least as good for the quadratic sequence as for the steepest descent sequence. This argument can easily be extended to include steepest descent sequences [1] which allow $\eta = \xi_n - \xi_x$ for x a support point of ξ_n.

We are now ready to prove the main results. For unified proofs, define a unified notation, setting $\xi_{n+1} = \xi_n + \alpha_n\zeta_n$; here $\zeta_n = \xi_{x_n} - \xi_n$ for the steepest descent sequence and $\zeta_n = \eta_n$ for the quadratic sequence. Write $\Delta = ||\mathbf{M}(\zeta_n)||$. By the remarks following (2.6), for both sequences there is a positive number C such that for all y and sufficiently large n

$$\Delta\xi_n(y) \geqq -C\zeta_n(y) .$$

From this, it follows easily that for all y and sufficiently large n

$$(2.9) \qquad 0 \leqq \alpha\Delta \leqq C \qquad \text{implies} \quad \xi_n(y) + \alpha\zeta_n(y) \geqq 0 .$$

THEOREM 2.1. *Under Assumptions* 1–5, $\Phi(\mathbf{M}_n) \to \inf_{\mathscr{M}} \Phi(\mathbf{M})$ *monotonically, for both the steepest descent and the quadratic sequence.*

PROOF. Since $\partial\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n))/\partial\alpha|_0 < 0$ in both cases, monotonicity is clear. Let $\mathscr{M}_0$ be defined by (2.7). Then $\mathscr{M}_0$ is contained in $\mathscr{M}^+$ and is closed in $\mathscr{M}$, so Lemma 2.1 applies. By monotonicity, $M_n \in \mathscr{M}_0$ for $n \geqq 0$.

Denote $K(\alpha; \mathbf{M}(\zeta_n), \mathbf{M}_n)$ by $K_n(\alpha)$, and as above denote $||\mathbf{M}(\zeta_n)||$ by Δ. Let A be the set of α such that $\xi_n + \alpha\zeta_n$ is a design with $\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n)) \leqq \Phi(\mathbf{M}_n)$. For any $\alpha \in A$ we have, by (2.2),

$$\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n)) - \Phi(\mathbf{M}_n) = \alpha\Delta K_n'(0) + \tfrac{1}{2}(\alpha\Delta)^2K_n''(\alpha^*\Delta)$$

for some α^* between 0 and α. But

$$K''(\alpha^*\Delta; \mathbf{M}(\zeta_n), \mathbf{M}_n) = K''(0; \mathbf{M}(\zeta_n), \mathbf{M}_n + \alpha^*\mathbf{M}(\zeta_n)) .$$

By the convexity of Φ and choice of α it follows that $\mathbf{M}_n + \alpha^*\mathbf{M}(\zeta_n) \in \mathscr{M}_0$, so by Lemma 2.1

$$\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n)) - \Phi(\mathbf{M}_n) \tag{2.10}$$

$$\leqq \alpha\Delta K_n'(0) + \tfrac{1}{2}(\alpha\Delta)^2 U_2 . \tag{2.11}$$

Suppose $K_n'(0)$ does not approach 0, i.e., for some $\delta > 0$ and infinitely many n, we have $K_n'(0) < -\delta$. Without loss of generality, δ may be chosen so that δ/U_2 is less than the number C in (2.6). Then

$$\begin{aligned} \Phi(\mathbf{M}_{n+1}) - \Phi(\mathbf{M}_n) &= \min_{\alpha\in A}[\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n)) - \Phi(\mathbf{M}_n)] \\ &\leqq \min_{\alpha\in A}[\alpha\Delta K_n'(0) + \tfrac{1}{2}(\alpha\Delta)^2 U_2] . \end{aligned} \tag{2.12}$$

Let α_0 be the right end point of A. If (2.10) is zero at $\alpha = \alpha_0$, then (2.11) is at least zero there, so by the convexity of (2.11) it is positive for $\alpha > \alpha_0$. If on the other hand (2.10) is negative at $\alpha = \alpha_0$, this means that $\xi_n + \alpha\zeta_n$ is not a design for $\alpha > \alpha_0$. In either case, the minimum in (2.12) is equal to the minimum over all α such that $\xi_n + \alpha\zeta_n$ is a design. From (2.9) and the choice of δ we have that $\xi_n + \alpha\zeta_n$ is a design for $\alpha\Delta = \delta/U_2$. Substitution of this value into (2.12) yields

$$\Phi(\mathbf{M}_{n+1}) - \Phi(\mathbf{M}_n) \leqq -\delta^2/2U_2 .$$

Summing over these infinitely many n, we obtain

$$\sum [\Phi(\mathbf{M}_{n+1}) - \Phi(\mathbf{M}_n)] = -\infty ,$$

which is impossible since $\Phi(\mathbf{M}(\xi))$ is bounded below. We conclude that $K_n'(0)\to 0$.

For the steepest descent sequence, let D be the minimum distance between an element of $\mathscr{M}_0$ and a singular element of $\mathscr{M}$. Then

$$\begin{aligned} K_n'(0) &= K'(0; \mathbf{M}(\xi_{x_n} - \xi_n), \mathbf{M}_n) \\ &= \|\mathbf{M}(\xi_{x_n}) - \mathbf{M}_n\|^{-1}\partial\Phi(\mathbf{M}_n + \alpha[\mathbf{M}(\xi_{x_n}) - \mathbf{M}_n])/\partial\alpha|_0 \\ &\leqq D^{-1}[\Phi(\mathbf{M}^*) - \Phi(\mathbf{M}_n)] \end{aligned} \tag{2.13}$$

by (2.1), so $\Phi(\mathbf{M}_n) \to \Phi(\mathbf{M}^*)$.

For the quadratic sequence

$$\begin{aligned} [K_n'(0)]^2/L_2 &\geqq [K'(0; \mathbf{M}(\eta_n), \mathbf{M}_n)]^2/K''(0; \mathbf{M}(\eta_n), \mathbf{M}_n) \\ &\geqq [K'(0; \mathbf{M}(\xi_{x_n} - \xi_n), \mathbf{M}_n)]^2/K''(0; \mathbf{M}(\xi_{x_n} - \xi_n), \mathbf{M}_n) \\ &\geqq [K'(0; \mathbf{M}(\xi_{x_n} - \xi_n), \mathbf{M}_n)]^2/U_2 \end{aligned}$$

by the remarks following (2.8). So $K_n'(0) \to 0$ implies, by (2.13), that $\Phi(\mathbf{M}_n) \to \Phi(\mathbf{M}^*)$. □

The assumption of third derivatives was not used in the above proof. A proof of the steepest descent portion of the theorem, under conditions similar to ours, is given in [4].

Observe that since $\Phi(\mathbf{M}_n) \to \Phi(\mathbf{M}^*)$, it follows that $\mathbf{M}_n \to \mathbf{M}^*$, by the compactness of $\mathscr{M}$ and the strict convexity of Φ. Since Ξ is compact, $\{\xi_n\}$ has an accumulation point ξ^*, and any such ξ^* must be optimal.

LEMMA 2.2. *Under Assumptions* 1–5,

$$\alpha_n \Delta \leqq -K_n'(0)/L_2 .$$

Here Δ and K_n remain as defined in the last proof.

PROOF. Since α_n minimizes $\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n))$ subject to the condition that $\xi_n(y) + \alpha\eta_n(y) \geqq 0$ for all y, it follows that

$$\begin{aligned} 0 &\geqq \partial\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n))/\partial\alpha|_{\alpha=\alpha_n} \\ &= \partial\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n))/\partial\alpha|_0 + \alpha_n \partial^2\Phi(\mathbf{M}_n + \alpha\mathbf{M}(\zeta_n))/\partial\alpha^2|_{\alpha=\alpha^*} \\ &= \Delta K_n'(0) + \alpha_n \Delta^2 K_n''(\alpha^*\Delta) \end{aligned}$$

for some α^* between 0 and α_n. As in the proof of Theorem 2.1 we conclude that

$$0 \geqq K_n'(0) + \alpha_n \Delta L_2$$

from which the assertion follows. □

THEOREM 2.2. *Under Assumptions* 1–5, *for both the steepest descent and the quadratic sequence, the improvement on the* nth *iteration is given by*

$$\Phi(\mathbf{M}_{n+1}) - \Phi(\mathbf{M}_n) = -(1 + \varepsilon_n)[K_n'(0)]^2/2K_n''(0)$$

where $\varepsilon_n \to 0$ *as* $n \to \infty$.

COROLLARY. *The improvement in a single iteration is asymptotically at least as good for the quadratic sequence as for the steepest descent sequence.*

The corollary follows from the remarks after (2.8).

PROOF OF THEOREM 2.2. The same reasoning as in the proof of Theorem 2.1 shows that

$$\Phi(\mathbf{M}_{n+1}) - \Phi(\mathbf{M}_n) \leqq \min_\alpha [\alpha\Delta K_n'(0) + \tfrac{1}{2}(\alpha\Delta)^2 K_n''(0) + (\alpha\Delta)^3 U_3/6] .$$

The minimum may be taken over the set of α for which $\xi_n + \alpha\zeta_n$ is a design and the expression in brackets is convex.

Consider $\alpha\Delta = -K_n'(0)/K_n''(0)$. The proof of Theorem 2.1 showed that $K_n'(0) \to 0$. So for n sufficiently large the $\alpha\Delta$ under consideration is less than the C of (2.9), so $\xi_n + \alpha\zeta_n$ is a design. Also for n large this $\alpha\Delta$ is less than $L_2/|U_3|$, so α is in the region of convexity of the expression in brackets above. Therefore for n large

$$\begin{aligned} \Phi(\mathbf{M}_{n+1}) - \Phi(\mathbf{M}_n) &\leqq -[K_n'(0)]^2/2K_n''(0) - [K_n'(0)/K_n''(0)]^3 U_3/6 \\ &\leqq -(1 + K_n'(0)U_3/3L_2^2)[K_n'(0)]^2/2K_n''(0) , \end{aligned}$$

and $K_n'(0)U_3/3L_2^2 \to 0$.

In the other direction,

$$\begin{aligned} \Phi(\mathbf{M}_{n+1}) - \Phi(\mathbf{M}_n) &\geqq \alpha_n \Delta K_n'(0) + \tfrac{1}{2}(\alpha_n \Delta)^2 K_n''(0) + (\alpha_n \Delta)^3 L_3/6 \\ &\geqq -[K_n'(0)]^2/2K_n''(0) + (\alpha_n \Delta)^3 L_3/6 \\ &\geqq -(1 + K_n'(0)U_2 L_3/3L_2^3)[K_n'(0)]^2/2K_n''(0) \end{aligned}$$

using Lemma 2.2 to get the last inequality. Since $K_n'(0) \to 0$, the theorem is proved. □

There are two respects in which the corollary to Theorem 2.2 falls short of a complete endorsement of the quadratic sequence over the steepest descent sequence. First, it is only asymptotic, and says nothing about how the sequences compare in the early iterations. Second, it does not take into account the effort needed to perform each iteration, which will be substantially greater for the quadratic sequence. The steepest descent method has the advantage that $\mathbf{M}^{-1}(\xi_{n+1})$ can be written as a simple function of $\mathbf{M}^{-1}(\xi_n)$, eliminating the need for matrix inversion. These simple computations are exploited for D- and L-optimality in [3]. If the quadratic method is actually to perform better, the decrease in Φ at each iteration must be substantial enough to justify the extra work. In Example 4.2 this is indeed the case.

We would expect the difference in methods to be greatest when the η to minimize (2.8) is selected from a large class, i.e., when there are many possible support points, not related to each other by any constraints such as symmetry. The two examples of Section 4 support this conjecture, and also suggest that the quadratic method seems to perform as well as the steepest descent method even in the early iterations.

3. Specific optimality criteria.

A. *D-optimality.* Here the optimality criterion function may be taken as

$$\Phi(\mathbf{M}) = -\log |\mathbf{M}|$$

if $|\mathbf{M}| > 0$, and $\Phi(\mathbf{M}) = \infty$ otherwise. Assumptions 1–5 are all easy to verify. In particular, 4 follows from the fact that the derivative under consideration equals $\operatorname{tr} \mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A} = \operatorname{tr} \mathbf{U}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}\mathbf{U}^T$, where $\mathbf{U}^T\mathbf{U} = \mathbf{M}^{-1}$. This is strictly positive if $\mathbf{M}$ is nonsingular and $\mathbf{A}$ is symmetric and nonzero. Using standard notation, we have $\partial\Phi(\mathbf{M}[\xi + \alpha(\xi_x - \xi)])/\partial\alpha|_0 = k - d(x, \xi)$, where $d(x, \xi) = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(x)$. The argument ξ will sometimes be suppressed below. For calculating (2.5), the ith element of $\mathbf{g}$ is $-d(y_i, \xi_n)$, and the ijth element of $\mathbf{H}$ is $d^2(y_i, y_j, \xi_n)$, where $d(x, y, \xi)$ is defined as $\mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(y)$.

Points of support. As mentioned by Fedorov [3] and Tsay [13, 14], steepest descent procedures tend to give designs with clusters of support points, with each point having very small mass. These can be collapsed into single points at the end of the iteration procedure, as suggested by Fedorov. Or they can be collapsed from time to time during the iterative process, for example after every tenth iteration; however, whenever this is done Fedorov's formulas [3, Theorem 2.6.1] are not applicable, and $\mathbf{M}^{-1}$ must be computed directly.

With the quadratic method there is no reason to defer combining nearby support points. On the contrary, the fewer support points there are, the easier will be the calculation of (2.5). The question then arises of the best way to combine them. We consider here the case of two nearby points on a line, and assume that **f** has three continuous derivatives in the local region of interest.

Suppose that $\mathscr{X}$ is one-dimensional, and that $d(x, \xi)$ is locally maximized at an interior point x_0, with $d(x_0, \xi) > k$ and the second derivative of $d(x, \xi)$ strictly negative at x_0. Suppose also that $x_1 = x_0 + \varepsilon_1$ is a support point of ξ, with ε_1 small. Consider adding measure γ to ξ at some $x = x_0 + \varepsilon$, and at the same time transferring all of the measure now at x_1 to x. What choice of x is best?

Formally, let

$$\xi' = (1 - \alpha - \beta)\xi + \alpha\xi_{x_1} + \beta\xi_x$$

where $\alpha = -\xi(x_1)$, and $\alpha + \beta = \gamma$, with γ presumably positive. Application twice of Theorem 2.5.1 of [3] yields

$$|\mathbf{M}(\xi')|/|\mathbf{M}(\xi)| = (1 - \gamma)^k\{[1 + ad(x_1, \xi)][1 + bd(x, \xi)] - abd^2(x, x_1, \xi)\} \tag{3.1}$$

where $a = \alpha/(1 - \alpha - \beta)$ and $b = \beta/(1 - \alpha - \beta)$. To avoid possible confusion about primes, we will here use one or two dots to denote first or second derivatives with respect to x, and continue to use T for transpose. By the assumptions on the behavior of $d(x)$ at x_0, we have

$$\begin{aligned} d(x) &\doteq d(x_0) + \tfrac{1}{2}\varepsilon^2\ddot{d}(x_0) \\ &= d(x_0) + \varepsilon^2[\mathbf{f}^T\mathbf{M}^{-1}\ddot{\mathbf{f}} + \dot{\mathbf{f}}^T\mathbf{M}^{-1}\dot{\mathbf{f}}]\,, \end{aligned} \tag{3.2}$$

where **f** and its derivatives are evaluated at x_0, **M** means $\mathbf{M}(\xi)$, and terms of order ε^3 are ignored. Similarly, we have

$$d(x, x_1) \doteq d(x_0) + \tfrac{1}{2}(\varepsilon^2 + \varepsilon_1^2)\mathbf{f}^T\mathbf{M}^{-1}\ddot{\mathbf{f}} + \varepsilon_1\varepsilon\dot{\mathbf{f}}^T\mathbf{M}^{-1}\dot{\mathbf{f}}\,. \tag{3.3}$$

Substitution of (3.2) and (3.3) into (3.1) yields

$$\begin{aligned} |\mathbf{M}(\xi')|/|\mathbf{M}(\xi)| \doteq (1 - \gamma)^k\{1 + ad(x_1) + b[d(x_0) + \tfrac{1}{2}\varepsilon^2\ddot{d}(x_0)] \\ + ab(\varepsilon - \varepsilon_1)^2d(x_0)\dot{\mathbf{f}}^T\mathbf{M}^{-1}\dot{\mathbf{f}}\}\,. \end{aligned} \tag{3.4}$$

For fixed $a < 0$ and $b \neq 0$, expression (3.4) is maximized when

$$\varepsilon = \varepsilon_1[ad(x_0)\dot{\mathbf{f}}^T\mathbf{M}^{-1}\dot{\mathbf{f}}]/[ad(x_0)\dot{\mathbf{f}}^T\mathbf{M}^{-1}\dot{\mathbf{f}} + \tfrac{1}{2}\ddot{d}(x_0)]\,. \tag{3.5}$$

Note that ε lies between ε_1 and 0, so x lies between x_1 and x_0.

So if it is decided that ξ' should have only one support point near x_0, then the best such point is approximated by $x = x_0 + \varepsilon$, with ε given by (3.5). To use (3.5) just as it is written, one must use γ, the additional new measure to be placed at x, and one must treat the local maxima one at a time, with $\mathbf{M}^{-1}$ and $d(x)$ being recomputed after each change. However, the quadratic sequence in Example 4.2 converged very satisfactorily when local adjustments were made as follows:

At any local maximum x_0 near a support point x_1 of ξ_n, the point $x = x_0 + \varepsilon$ was found from (3.5), with $a = -\xi_n(x_1)/(1 + \xi_n(x_1))$, i.e., with γ set to 0. Then $\mathscr{X}_n$ was defined to include such x, rather than the corresponding x_0 or x_1. The next design ξ_{n+1} was taken to be supported on $\mathscr{X}_n$. All of the calculations were based on the original $\mathbf{M}^{-1}(\xi_n)$ and $d(x, \xi_n)$.

When $\mathscr{X}$ has dimension greater than 1, the above method still can be used, by restricting attention to the line through x_0 and x_1. The derivatives are then understood to be derivatives along the line, and (3.5) gives the approximate location of the best support point, among points on the line.

B. *L-optimality*. It $\mathbf{M}$ is nonsingular, the criterion function is

$$\Phi(\mathbf{M}) = \operatorname{tr} \mathbf{CM}^{-1}$$

for some given symmetric positive definite $\mathbf{C}$. $\mathbf{C}$ must be nonsingular. If $\mathbf{M}$ is singular, define $\Phi(\mathbf{M}) = \infty$. All the assumptions are easy to check. Assumption 4 follows from the fact that the derivative in question equals $2 \operatorname{tr} \mathbf{CM}^{-1}\mathbf{AM}^{-1}\mathbf{AM}^{-1} = 2 \operatorname{tr} \mathbf{UAM}^{-1}\mathbf{CM}^{-1}\mathbf{AU}^T$, where $\mathbf{U}^T\mathbf{U} = \mathbf{M}^{-1}$. This is strictly positive if $\mathbf{M}$ is nonsingular and $\mathbf{A}$ is symmetric and nonzero. For calculating (2.5), the ith element of $\mathbf{g}$ is $-\varphi(y_i) = -\mathbf{f}^T(y_i)\mathbf{M}^{-1}\mathbf{CM}^{-1}\mathbf{f}(y_i)$, and the ijth element of $\mathbf{H}$ is $2\varphi(y_i, y_j)d(y_i, y_j)$, where $\varphi(x, y)$ is defined as $\mathbf{f}^T(x)\mathbf{M}^{-1}\mathbf{CM}^{-1}\mathbf{f}(y)$.

The analogue of (3.5) for L-optimality can be given in principle when $a = b$, but the algebra is very messy.

Extensions of Fedorov's results. In this subsection $\mathbf{C}$ is only required to be nonnegative definite, not necessarily positive definite. Write $L(\mathbf{M}^{-1})$ instead of $\Phi(\mathbf{M})$, and for any given x define the following notation:

$$\varphi = L[\mathbf{M}^{-1}(\xi)\mathbf{f}(x)\mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)] = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{CM}^{-1}(\xi)\mathbf{f}(x)$$
$$t = L(\mathbf{M}^{-1}(\xi)) = \operatorname{tr} \mathbf{CM}^{-1}(\xi)$$
$$d = d(x, \xi) = \mathbf{f}^T(x)\mathbf{M}^{-1}(\xi)\mathbf{f}(x) .$$

THEOREM 3.1. *Let* $\mathbf{M}(\xi)$ *be nonsingular. Then*:

(a) $td \geqq \varphi$, *with strict inequality if the rank of* $\mathbf{C}$ *is greater than* 1.

(b) $\partial L(\mathbf{M}^{-1}[(1-\alpha)\xi + \alpha\xi_x])/\partial\alpha = (1-\alpha)^{-2}(1+\alpha(d-1))^{-2}\{t - \varphi + 2\alpha t(d-1) + \alpha^2(d-1)[t(d-1) - \varphi]\}$.

Part (a) is stronger than Lemma 2.10.1 of Fedorov [3], which states that $\varphi(d-1) \geqq \varphi - t$. Part (b) enables one to calculate the steepest descent α for any x, by setting the expression in braces equal to zero. This last would seem to be of only academic interest to users of the quadratic sequence. However the steepest descent α can be substituted into (3.7) to give a numerical upper bound on $\Phi(\mathbf{M}^*) - \Phi(\mathbf{M})$. This complements the lower bound of (2.1) and is the analogue for L-optimality of (4.5) of [17].

PROOF. By Lemma 2.10.2 of [3] we have

$$\text{(3.7)} \qquad L(\mathbf{M}^{-1}[(1-\alpha)\xi + \alpha\xi_x]) = (1-\alpha)^{-1}(1+\alpha(d-1))^{-1}(t + \alpha[t(d-1) - \varphi]) .$$

If $td - \varphi = t + [t(d-1) - \varphi]$ is negative, then (3.7) approaches $-\infty$ as $\alpha \to 1$, which is impossible. If $td - \varphi = 0$, then

$$L(\mathbf{M}^{-1}[(1-\alpha)\xi + \alpha\xi_x]) = t(1 + \alpha(d-1))^{-1}$$

which is bounded as $\alpha \to 1$. This is impossible if rank $\mathbf{C} > 1$. This proves part (a). Part (b) is a routine exercise, based on (3.7). □

4. Numerical examples. Two examples with D-optimality were tried on a computer, using both the steepest descent method and the quadratic method. They are discussed in somewhat more detail in [2]. The steepest descent sequence incorporated suggestion 1 of Atwood [1], in which mass may be subtracted at a point with small $d(x, \xi)$ if this is better than adding mass at any point with large $d(x, \xi)$. The other modifications of [1] and St. John and Draper [8] were not used because then the time saving formulas of [3] for computing $\mathbf{M}^{-1}(\xi_{n+1})$ and $d(x, \xi_{n+1})$ could not have been used.

As shown in [1], if $\xi(x) > 1/k$ for some x, then ξ can be improved by reducing $\xi(x)$ to $1/k$ and distributing the excess mass proportionately among the other support points of ξ. In the quadratic sequence this was always done with any ξ_n obtained, before $\mathbf{M}(\xi_n)$ was calculated. For the steepest descent sequence in Example 4.2, clusters of design points were periodically collapsed into single points. At that time the resulting ξ was also adjusted to make $\xi(x) \leqq 1/k$ for all x.

In computing the quadratic sequence, $\mathbf{H}^{-1}$ need not be found explicitly, since only $\mathbf{H}^{-1}\mathbf{e}$ and $\mathbf{H}^{-1}\mathbf{g}$ are needed.

In the quadratic sequence, after finding η from (2.5) we were supposed to find α_n to minimize $\Phi(\mathbf{M}(\xi_n + \alpha\eta_n))$. It seemed satisfactory merely to approximate this α_n, using a few iterations of Newton's method. Of course, it is understood always that α_n must be such that $\xi_n + \alpha_n\eta_n$ does not assign negative mass to any point.

EXAMPLE 4.1. This is Wynn's example [17], in which

$$\mathbf{f}^T(x)\boldsymbol{\theta} = \theta_0 + \theta_1 x_1 + \theta_2 x_2, \qquad x = (x_1, x_2)$$

and $\mathscr{X}$ is the closed convex quadrilateral with vertices $A = (2, 2)$, $B = (-1, 1)$, $C = (1, -1)$ and $D = (-1, -1)$. A crucial simplifying feature of this problem is that an optimal design can only be supported on $\{A, B, C, D\}$. In each sequence, ξ_0 was taken to be uniform on $\{B, C, D\}$. Both sequences converged quickly. Iterations were stopped when $\max_x d(x, \xi_n)$ was less than 3.00005. The steepest descent sequence took 9 iterations to achieve this; the quadratic sequence took 3 iterations. The execution times were both approximately $\frac{1}{2}$ second on the UCD Burroughs 6700. Since exact execution times fluctuate with the time sharing mix in the machine, all that can be said is that the execution times were close to each other. The major portion of the computer time was for compilation rather than execution. The steepest descent program was considerably shorter to compile.

EXAMPLE 4.2. This example is given by Tsay [13, 14]

$$\mathbf{f}^T(x)\boldsymbol{\theta} = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x_+^{\,2} + \theta_4 (x - .3)_+^{\,2}$$

for $x \in [-1, 1]$, where

$$(x-a)_+^{\,m} = 0 \quad \text{if} \quad x < a$$
$$= (x-a)^m \quad \text{if} \quad x \geqq a\,.$$

As an initial design, ξ_0 was chosen to be uniform on $\{-1, -.5, 0, .5, 1\}$.

In the steepest descent sequence, clusters of points were collapsed into single points from time to time, following Fedorov [3, page 109]. Doing this fairly often seemed to help slightly more than doing it infrequently.

As $|\mathbf{M}(\xi_n)|$ increased monotonically, $\bar{d}(\xi_n) = \max_x d(x, \xi_n)$ oscillated, gradually tending downwards. It stayed below 5.5 from iteration 16 on, and below 5.05 from iteration 46 on. The procedure was terminated after 82 iterations, and 6.20 seconds. At that time $\bar{d}(\xi_n)$ was 5.028. Information for a few of the iterations is shown in Table 1. The values $\xi(x_i)$ are not shown in Tables 1 and 2, but all designs which are supported on only five points assign equal mass to each point.

Column 5 of Table 1 reflects the fact that the point .1281 was not ever moved after iteration 6, and other points were left untouched for long periods of time.

Tsay [13, 14] tries other steepest descent sequences with this example, using different choices of $\{\alpha_n\}$. His results are equally unsatisfying.

The quadratic sequence was much more successful, as can be seen in Table 2. It converged in 4 iterations, requiring 1.5 seconds to achieve $\bar{d}(\xi_n) = 5.00002$.

TABLE 1
Steepest descent sequence, example 4.2

Iteration	Time (sec.)	$10^7 \times \|\mathbf{M}\|$	$\bar{d}$	Points of Support After Collapsing Clusters
16	1.42	2.0792	5.47125	−1.0000
				−.5000
				.1281
				.5000
				.6177
				1.0000
46	3.72	2.1305	5.04634	−1.0000
				−.4875
				.1281
				.6125
				1.0000
82	6.20	2.1363	5.02770	−1.0000
				−.4804
				.1281
				.6125
				1.0000

TABLE 2
Quadratic sequence, example 4.2

Iteration	Time (sec.)	$10^7 \times \|M\|$	$\bar{d}$	Points of Support After Collapsing Clusters
0	.23	1.3613	7.58633	−1.0000
				−.5000
				.0000
				.5000
				1.0000
1	.65	1.9953	5.80774	−1.0000
				−.5000
				.1705
				.5000
				.6331
				1.0000
2	1.02	2.1219	5.27046	−1.0000
				−.5000
				−.4342
				.1252
				.5000
				.6141
				1.0000
3	1.25	2.1497	5.00155	−1.0000
				−.4493
				.1315
				.5986
				1.0000
4	1.50	2.1502	5.00002	−1.0000
				−.4551
				.1315
				.5996
				1.0000

For this sequence, $\mathscr{X}_n$ was taken to consist of the support of ξ_n, together with all x_i at which $d(x, \xi_n)$ was locally maximized and $> 1/k$. If such an x_i was close to a support point of ξ_n, the two points were combined according to the remarks following (3.5).

One lesson from the computing experience is that excessive caution in collapsing clusters slows down convergence. Points were considered close enough to be treated together if they were within some δ of each other. In the sequences summarized in the tables, δ was set equal to .06. Both sequences converge more slowly if $\delta = .02$, and more quickly if $\delta = .15$. The danger with too large a δ is that points may be combined which should have been left distinct. However for greatest speed in the early iterations it seems that δ should be as large as is felt to be safe.

In this example, $d(x, \xi)$ was a quartic polynomial in each of the three intervals $[-1, 0]$, $[0, .3]$ and $[.3, 1]$. This simple structure greatly eased the search for the local maxima. In a different example if the search for the local maxima were much harder, this would give a greater advantage to the sequence with fewer iterations, the quadratic sequence.

REFERENCES

[1] ATWOOD, C. L. (1973). Sequences converging to *D*-optimal designs of experiments. *Ann. Statist.* **1** 342–352.

[2] ATWOOD, C. L. (1976). Computational considerations for convergence to an optimal design. *Proc. 1976 Conf. on Information Sciences and Systems*, Dept. of Elec. Eng., Johns Hopkins Univ.

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

[4] FEDOROV, V. V. (1975). Numerical methods of constructing optimal designs for regression experiments. (In Russian) *Kibernetika* **11** 124–130.

[5] FEDOROV, V. V. and MALYUTOV, M. B. (1972). Optimal design in regression problems. *Math. Operationsforsch. Statist.* **3** 281–308.

[6] GRIBIK, P. R. and KORTANEK, K. O. (1976). Equivalence theorems and cutting plane algorithms for a class of experimental design problems. *SIAM J. Appl. Math.* To appear.

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

[8] ST. JOHN, R. and DRAPER, N. R. (1975). *D*-optimality and mixtures. Unpublished.

[9] SIBSON, R. (1972). Contribution to discussion of "Results in the theory and construction of *D*-optimum experimental designs" by H. P. Wynn. *J. Roy. Statist. Soc. Ser. B* **34** 181–183.

[10] SILVEY, S. D. and TITTERINGTON, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60** 21–32.

[11] SILVEY, S. D., TITTERINGTON, D. M. and TORSNEY, B. (1976). An algorithm for *D*-optimal designs on a finite design space. Unpublished.

[12] TITTERINGTON, D. M. (1975). Optimal design: some geometrical aspects of *D*-optimality. *Biometrika* **62** 313–320.

[13] TSAY, J.-Y. (1974). The iterative methods for calculating optimal experimental designs. Ph.D. thesis, Purdue Univ.

[14] TSAY, J.-Y. (1976). On the sequential construction of *D*-optimal designs. *J. Amer. Statist. Assoc.* To appear.

[15] WHITTLE, P. (1971). *Optimization Under Constraints.* Wiley-Interscience, New York.

[16] WHITTLE, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35** 123–130.

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

[18] WYNN, H. P. (1972). Results in the theory and construction of *D*-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **34** 133–147.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616