The Annals of Statistics
1978, Vol. 6, No. 6, 1273–1285

THE CONVERGENCE OF GENERAL STEP-LENGTH ALGORITHMS FOR REGULAR OPTIMUM DESIGN CRITERIA

BY CHIEN-FU WU[1] AND HENRY P. WYNN

University of California, Berkeley and Imperial College, London

For a regular optimality criterion function Φ, a sequence of design measures $\{\xi_n\}$ is generated using the iteration $\xi_{n+1} = (1 - \alpha_n)\xi_n + \alpha_n \bar{\xi}_n$, where $\bar{\xi}_n$ is chosen to minimize $\nabla \Phi(M(\xi_n), M(\xi))$ over all ξ and $\{\alpha_n\}$ is a prescribed sequence of numbers from (0, 1). This is called a general step-length algorithm for Φ. Typical conditions on $\{\alpha_n\}$ are $\alpha_n \to 0$ and $\Sigma_n \alpha_n = \infty$. In this paper, a dichotomous behavior of $\{\xi_n\}$ is proved under the above conditions on $\{\alpha_n\}$ for Φ satisfying some mild regularity conditions. Sufficient conditions for convergence to optimal designs are also established. This can be applied to show that the $\{\xi_n\}$ as constructed above do converge to an optimal design for most of the trace-related and determinant-related design criteria.

1. Introduction. Since the introduction of optimum design algorithms (Wynn (1970, 1972), Fedorov (1972)) there has been a successful search for algorithms for more general optimality criteria (Fedorov and Malyutov (1972), Gribik and Kortanek (1977), Whittle (1973), Atwood (1973, 1976 a, b)). This has been stimulated in part by the parallel development of the general equivalence theorems of Kiefer (1974). Some special algorithms have also been developed for the original D-optimality criterion (Silvey et al. (1978), Titterington (1976), Tsay (1976)).

More recently still the thesis of one of the present authors (Wu (1976)) draws the connection with the vast area of optimization theory and algorithms (see for example the books by Luenberger (1973), Polak (1971), Mangasarian (1969)). While this must inevitably lead to some simplification of the subject, the authors believe that it is useful to highlight the special features of the optimum design algorithms and even that optimization theory may gain a little in the process.

As is usual we consider a convex functional Φ on the space of $k \times k$ moment matrices $M(\xi) = [m_{ij}(\xi)]_{i,j}$ where

$$m_{ij}(\xi) = \int_{\mathcal{X}} f_i(x) f_j(x) \xi(dx)$$

and ξ is a probability measure on the compact set $\mathcal{X}$ (the design region) on which the $f_i(x)$'s, $i = 1, \cdots, k$, are continuous.

The purpose of optimum design theory is to find design measures which are Φ-optimal, that is to achieve

$$\inf \Phi(M(\xi))$$

Received April 1977; revised November 1977.

[1]Research supported by Office of Naval Research Contracts N00014-75-C-0444 and NR 042-036.

AMS 1970 *subject classifications*. Primary 62K05; Secondary 62L05.

Key words and phrases. General step-length algorithms, optimal design algorithms, convex programming, D-optimality, A-optimality, Φ_p-optimality, D_s-optimality, asymptotic convergence.

over all probability measures ξ on $\mathcal{X}$. Let $\mathfrak{M}$ be the set of all $M(\xi)$ which is a compact convex set in R^{k^2}. At this point we could abandon all reference to moment problem and design and treat the problem purely as a convex programming problem. We shall, rather, retain the design notation to draw out the special features. Three of these are:

(i) The measures with support at a single point play a special role in that they give the extreme points of $\mathfrak{M}$.

(ii) The functional Φ may be $+\infty$ at some points of $\mathfrak{M}$, typically for singular $M(\xi)$. For example for D-optimality $\Phi(M) = -\log \det(M) = \infty$ when M is singular.

(iii) Φ may have some special statistical meaning.

We shall use the directional derivative notation

$$\nabla \Phi(M_1, M_2) = \lim_{\alpha \to 0^+} \frac{\partial \Phi}{\partial \alpha}((1-\alpha)M_1 + \alpha M_2)$$

$$\nabla^2 \Phi(M_1, M_2) = \lim_{\alpha \to 0^+} \frac{\partial^2 \Phi}{\partial \alpha^2}((1-\alpha)M_1 + \alpha M_2)$$

when these quantities exist.

Necessary and sufficient conditions for Φ-optimality are discussed in Kiefer (1974) under various assumptions on Φ. Briefly they stem from the necessary and sufficient condition for local minimum:

$$\nabla \Phi(M(\xi^*), M(\xi)) \geqslant 0 \qquad \text{for all} \quad M(\xi) \text{ in } \mathfrak{M}.$$

The algorithms in this paper are of the kind discussed by Wynn (1970), Fedorov (1972) and Tsay (1976). A sequence of measures is generated using the iteration

$$\xi_{n+1} = (1-\alpha_n)\xi_n + \alpha_n \bar{\xi}_n \tag{1.1}$$

in which $\bar{\xi}_n$ is chosen to minimize $\nabla \Phi(M(\xi_n), M(\xi))$ over all ξ, and the general assumptions:

$$\alpha_n \to 0 \text{ and } \Sigma_0^\infty \alpha_n = \infty \qquad 0 \leqslant \alpha_n \leqslant 1. \tag{1.2}$$

When Φ is differentiable at M_1, i.e.,

$$\nabla \Phi(M_1) = \left[\frac{\partial \Phi(M_1)}{\partial m_{ij}} \right]_{i,j},$$

$\nabla \Phi(M_1, M_2)$ is linear in $M_2 - M_1$. That is,

$$\nabla \Phi(M_1, M_2) = \operatorname{tr}(\nabla \Phi(M_1)(M_2 - M_1)).$$

For this reason $\min_{M_2} \nabla \Phi(M_1, M_2)$ is achieved in such cases when M_2 is an extreme point of $\mathfrak{M}$ in which case

$$M_2 = f(\bar{x})f(\bar{x})^T$$

for some $\bar{x}$ in $\mathcal{X}$, i.e., $f(\bar{x})^T \nabla \Phi(M_1) f(\bar{x}) = \min_{x \in \mathcal{X}} f(x)^T \nabla \Phi(M_1) f(x)$.

Thus algorithm (1.1) has the following form:

$$\xi_{n+1} = (1 - \alpha_n)\xi_n + \alpha_n \xi_{x_n} \tag{1.3}$$

where ξ_{x_n} is concentrated at x_n with

$$f(x_n)^T \nabla \Phi(M(\xi_n))f(x) = \min_{x \in \mathfrak{X}} f(x)^T \nabla \Phi(M(\xi_n))f(x). \tag{1.4}$$

An important special case of (1.3) is the choice $\alpha_n = (n + 1)^{-1}$. This occurs in sequential design of experiments. Suppose we have already performed the experiments at $\{x_i\}_{i=1}^n$ up to the nth iteration (therefore, $M(\xi_n) = n^{-1}\Sigma_{i=1}^n x_i x_i^T$) and the next experiment is chosen to be performed at x_{n+1} with x_{n+1} chosen according to (1.4); then $M(\xi_{n+1}) = (1/(n + 1))\Sigma_{i=1}^{n+1} x_i x_i^T = (1 - (1/(n + 1)))M(\xi_n) + (1/(n + 1))x_{n+1}x_{n+1}^T$. General results proved in this paper will give an answer to the optimality of the above sequential procedure. Algorithms of this general kind are also useful in conjunction with a faster finite dimensional algorithm given by discretizing the design space.

Fedorov (1972, (2.6.21), (2.10.12)) gives the algorithm for D and L-optimality but without proof. The proof for the D-optimality case has been known to the second author for some years. Tsay (1976) gives it in detail following Wynn (1970) closely. The proof for L-optimality has not been given except in the special case $\alpha_n = 1/(n + 1)$ which requires further analysis (see Pázman (1974), Wynn (1975)).

The main purpose of this paper is to discuss the asymptotic behaviors of the algorithms mentioned above. A dichotomous theorem is proved in the next section. Sufficient conditions for convergence to optimal designs are discussed in Section 3. In the last section, we will show that most of the trace-related and determinant-related family of criteria satisfy the conditions in Section 3, including L-optimality and Kiefer's Φ_p-optimality. Under some additional conditions the result also holds for nonsingular D_s-optimality. Thus the convergence problem of algorithms (1.1) is settled for most of the commonly used optimality criteria. It will be seen that the main difficulty is the unboundedness mentioned in (ii) above. There seems to be little work on similar problems in the optimization literature. (For more general results in the framework of convex programming, see Wu (1976, 1978).)

2. A general result. For clarity we split the main result into two parts, first stating a theorem which emphasizes the possible unboundedness of the algorithm and then giving extra conditions in the next section which prevent this possibility.

Our initial regularity conditions are:

(A1) Φ is convex and bounded below on $\mathfrak{M}$ (but possibly $+\infty$) and achieves its minimum at $M(\xi^*)$ in $\mathfrak{M}$.

(A2) For any constant K ($> \min \Phi$), let

$$W(K) = \{M \in \mathfrak{M} : \Phi(M) \leqslant K\}.$$

(i) $\nabla^2\Phi(M_1, M_2)$ is continuous in M_1 (and M_2) for finite $\Phi(M_1)$ and (ii) sup $\{|\nabla^2\Phi(M_1, M_2)| : M_1 \in W(K), M_2 \in \mathfrak{M}\} = B(K) < \infty$.

These are straightforward regularity conditions given in terms of directional derivatives. Condition (A2) can be weakened, following similar conditions in optimization theory, to a Lipschitz condition on $\nabla\Phi$. This theory is developed by Wu (1976, 1978).

DICHOTOMOUS THEOREM. *Under* (A1) *and* (A2), *define a sequence* $\{M(\xi_n)\}_0^\infty$ *in* $\mathfrak{M}$ *satisfying*

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

for which

(i) $\bar{\xi}_n$ *achieves* $\inf_\xi \nabla\Phi(M(\xi_n), M(\xi))$,
(ii) $0 \leqslant \alpha_n \leqslant 1$, $\alpha_n \to 0$ *and* $\Sigma_0^\infty \alpha_n = \infty$.

Then either

(a) $\Phi(M(\xi_n)) \to \Phi(M(\xi^*))$

or

(b) *there is a subsequence* ξ_{n_r} *with*

$$\Phi(M(\xi_{n_r})) \to \infty.$$

PROOF. Suppose that (b) does not hold. Then there is a K such that $M(\xi_n) \in W(K)$ for all n. Expanding by a Taylor series in α_n:

$$\Phi(M(\xi_{n+1})) = \Phi(M(\xi_n)) + \alpha_n \nabla\Phi\big(M(\xi_n), M(\bar{\xi}_n)\big) + \alpha_n^2 \nabla^2\Phi\big((1-\bar{\alpha}_n)M(\xi_n) + \bar{\alpha}_n M(\bar{\xi}_n), M(\bar{\xi}_n)\big) \tag{2.1}$$

for some $\bar{\alpha}_n$ with $0 \leqslant \bar{\alpha}_n \leqslant \alpha_n$.

Now assume that

$$\Phi(M(\xi_n)) > \Phi(M(\xi^*)) + \varepsilon \qquad \text{for all} \quad n.$$

By convexity and the definition of $\bar{\xi}_n$

$$\begin{aligned} 0 < \varepsilon &< \Phi(M(\xi_n)) - \Phi(M(\xi^*)) \\ &\leqslant -\nabla\Phi(M(\xi_n), M(\xi^*)) \\ &\leqslant -\nabla\Phi\big(M(\xi_n), M(\bar{\xi}_n)\big). \end{aligned} \tag{2.2}$$

But the $\nabla^2\Phi$ term in (2.1) is bounded. Thus since $\alpha_n \to 0$ there is an n_0 such that for all $n > n_0$

$$\Phi(M(\xi_{n+1})) < \Phi(M(\xi_n)) - \alpha_n\frac{\varepsilon}{2}. \tag{2.3}$$

Summing (2.3) and using $\Sigma_0^\infty \alpha_n = \infty$ we see that $\Phi(M(\xi_n)) \to -\infty$, contradicting (A1). Thus there is an infinite subsequence $M(\xi_{n_s})$ with $\Phi(M(\xi_{n_s})) \to \Phi(M(\xi^*))$.

Now from (2.1) and the boundedness of $\nabla^2\Phi$ there is an n_1 such that for all $n > n_1$

$$\Phi(M(\xi_{n+1})) \leqslant \Phi(M(\xi_n)) + \varepsilon. \tag{2.4}$$

However, also using (2.1), (2.2) and the boundedness of $\nabla^2\Phi$, there is an n_2 such that for all $n > n_2$

$$\Phi(M(\xi_n)) > \Phi(M(\xi^*)) + \varepsilon \tag{2.5}$$

implies

$$\Phi(M(\xi_{n+1})) \leqslant \Phi(M(\xi_n)).$$

Now choose an $n_s > n_1, n_2$ such that

$$\Phi(M(\xi_{n_s})) < \Phi(M(\xi^*)) + \varepsilon$$

and (2.4) and (2.5) force

$$\Phi(M(\xi_n)) < \Phi(M(\xi^*)) + 2\varepsilon$$

for all $n > n_s$. The theorem is established. □

3. Conditions to eliminate unboundedness. Throughout this section, the conditions of the dichotomous theorem are assumed. To eliminate conclusion (b) in that theorem and thus give convergence to the optimum, additional conditions are needed. A few of them are listed below.

(C1) Φ is bounded above on $\mathfrak{M}$.

From the proof of the theorem this condition eliminates (b) trivially. However, it typically does not hold for many optimality criteria used in the optimal designs of experiments if convexity is to be preserved, as has already been mentioned.

(C2) $\Phi(M(\xi_n))$ is monotonically decreasing.

The condition clearly removes (b) since the whole sequence is bounded by the initial value. But it is also inappropriate since condition (ii) in the theorem does not in general give monotonicity. Monotone or optimum step length algorithms must be treated separately (see Fedorov (1972), Atwood (1973, 1976 a, b)). The next condition is more properly stated as a theorem.

(C3) *For any initial ξ_0 with finite $\Phi(M(\xi_0))$ there is an $\bar{\alpha}$, depending on ξ_0, such that if $\alpha_n \leqslant \bar{\alpha}$ for all n, (b) cannot hold.*

Since the proof is much more involved, it will be given in the appendix. A more general result for algorithms not necessarily of the form (1.1) can be given in the general framework of convex programming. The undesirable condition $\alpha_n \leqslant \bar{\alpha}$ in (C3) can also be removed via a theorem of Polyak (1967) if $M(\bar{\xi}_n) - M(\xi_n)$ is chosen, instead, to be a "support functional" to $W(\Phi(M(\xi_n)))$. (For details see Wu (1976, 1978).)

This condition is not very helpful in practice since it requires additional searches to get a suitable $\bar{\alpha}$. The following condition (C4), motivated from the original proofs for D-optimality, is more useful. It implies that there is an i_1 such that $\alpha_n \leqslant \bar{\alpha}(\xi_{i_1})$, the $\bar{\alpha}$ in (C3), for all $n \geqslant i_1$. An alternative proof that (C4) removes (b) can be given based on this connection with (C3).

(C4) There exists a $K < \infty$ such that for any $M(\xi) \notin W(K)$, the largest $\alpha^*(\xi)$ which achieves

$$\inf_{0 \leqslant \alpha \leqslant 1} \Phi\big((1-\alpha)M(\xi) + \alpha M(\bar{\xi})\big),$$

with $\bar{\xi}$ chosen to minimize $\nabla\,\Phi(M(\xi), M(\xi'))$ over all ξ', is greater than a positive constant η depending only on K.

This says that the optimum step is bounded away from zero when $\Phi(M(\xi))$ is bounded away from the infimum. From the convexity of Φ and $\nabla\,\Phi(M(\xi), M(\bar{\xi})) \leqslant 0$, $\Phi((1-\alpha)M(\xi) + \alpha M(\bar{\xi}))$ is monotonically decreasing in $0 \leqslant \alpha \leqslant \alpha^*(\xi)$.

Proof (that (C4) eliminates (b)). Take K in (C4) and K' such that $\infty > K' > K > \min \Phi$. Assume (b) holds. We can find a subsequence $\{M(\xi_{n_s})\}_{s=0}^{\infty}$ with (i) $M(\xi_{n_s}) \notin W(K)$ for all $s \geqslant 0$, (ii) $\alpha_n \leqslant \eta \leqslant \alpha^*(\xi_n)$ for all $n \geqslant n_0$, (iii) $\Phi(M(\xi_{n_s})) < \Phi(M(\xi_{n_{s+1}}))$ for all $s \geqslant 0$.

Now (ii) and the remark after (C4) imply that for $r \geqslant 0$

$$\Phi\big(M(\xi_{n_s+r})\big) \geqslant \Phi\big(M(\xi_{n_s+r+1})\big)$$

for all $M(\xi_{n_s+r}) \notin W(K)$. Therefore in order for (iii) to be true the sequence must return to $W(K)$ between n_s and n_{s+1}. That is, there exists a largest integer n_s' with $n_s < n_s' < n_{s+1}$ and $M(\xi_{n_s'}) \in W(K)$. For this n_s', we have

$$\Phi\big(M(\xi_{n_s'+1})\big) \geqslant \Phi\big(M(\xi_{n_{s+1}})\big) \quad \text{(by condition (C4))}.$$

Thus we have found an infinite subsequence $\{M(\xi_{n_s'})\}_{s=0}^{\infty}$ such that

$$M(\xi_{n_s'}) \in W(K) \qquad \text{for all} \quad s \geqslant 0$$

$$\text{and } \Phi\big(M(\xi_{n_s'+1})\big) \to \infty. \tag{3.1}$$

From $|M(\xi_{n_s'}) - M(\xi_{n_s'+1})| \leqslant \alpha_{n_s'} \cdot 2 \cdot \max\{|M| : M \in \mathfrak{M}\} \to 0$ and dist $(\partial W(K), \partial W(K')) \geqslant \delta > 0$ (from (5.1)), there exists an n_t such that

$$M(\xi_{n_s'+1}) \in W(K') \qquad \text{for all} \quad n_s' \geqslant n_t.$$

But then $\Phi(M(\xi_{n_s'+1}))$ is bounded in $W(K')$, contradicting (3.1). □

(C4) holds true for D-optimality. In this case

$$\alpha^*(\xi) = \frac{\bar{d}(\xi) - k}{k\big(\bar{d}(\xi) - 1\big)}$$

where $\bar{d}(\xi) = \sup_{x \in \mathfrak{X}} f(x)^T M(\xi)^{-1} f(x) > k + \delta$ for some $\delta > 0$ and all $M(\xi) \notin W(K)$ (see the inequality in Wynn (1970), (4.5) or Fedorov (1972), (2.5.13)).

The possible difficulty of verifying (C4) for many different criteria has led to a search for simpler conditions in terms of the first and second derivatives. What is needed is a condition which essentially controls the behavior of the second derivative for large Φ. We give in (C5) and (C6) two such conditions. In (C5), a decreasing function of Φ is well behaved and in the specialized version (C6) is exhibited a precise relationship between first and second derivatives. This has proved useful in practice (see Section 4), being satisfied by some well-known criteria. Conditions of this kind which control the behavior of unbounded functionals do not seem available in the optimization literature.

(C5) There exists a strictly decreasing continuous function h of $\Phi(M)$ and a constant $K < \infty$ such that

(i) $\psi(M) = h \circ \Phi(M)$ is uniformly continuous on $\mathfrak{M}$.
(ii) $\nabla^2\psi(M, \overline{M})$ exists and $\geqslant 0$ for $M \notin W(K)$ and $\overline{M}$ achieving $\inf_{M_1} \nabla \Phi(M, M_1)$.

PROOF (that (C5) eliminates (b)). We prove this by showing that (C5) $\Rightarrow$ (C4). With no loss of generality we will assume that $\sup \Phi(M) = \infty$ and $h(\infty) = 0$.

An equivalent condition to (C4) will be verified: there exists a positive η depending only on K such that $\Phi((1-\alpha)M + \alpha\overline{M})$ is $\downarrow$ in α for $0 \leqslant \alpha \leqslant \eta$ where $\overline{M}$ achieves $\inf_{M_1} \nabla \Phi(M, M_1)$ and $\Phi(M) \geqslant K'$ for some $K' > K$.

From (ii) and $h\downarrow$, $\psi((1-\alpha)M + \alpha\overline{M})$ is increasing in α provided it does not exceed $h(K)$ ($h(K) > 0$).

Let $D = 2\sup\{|M| : M \in \mathfrak{M}\}$ and $s(t)$, the reverse modulus of continuity of ψ, be $\inf\{|M_1 - M_2| : |\psi(M_1) - \psi(M_2)| \geqslant t,\ M_1, M_2 \in \mathfrak{M}\}$. From uniform continuity of ψ, $s(t) > 0$ for $t > 0$.

Thus if $\psi(M) \leqslant \frac{1}{2}h(K)$ and $\eta = s(h(K)/2)/D$, then $\psi((1-\alpha)M + \alpha\overline{M})$ is $\uparrow$ in α for $0 \leqslant \alpha \leqslant \eta$. This is equivalent to

$$\Phi\big((1-\alpha)M + \alpha\overline{M}\big) \text{ is } \downarrow \text{ in } \alpha \text{ for } 0 \leqslant \alpha \leqslant \eta$$

for

$$\Phi(M) \geqslant h^{-1}\left(\frac{h(K)}{2}\right) = K'. \qquad \square$$

The function h in (C5) should be chosen to make $h \circ \Phi(M)$ a nice function of M and also $h(x) \to 0$ fast as $x \to \infty$. We found $h(x) = e^{-dx}$ or x^{-d} for some large value of d particularly useful.

For $h(x) = e^{-dx}$, (C5) becomes a very simple condition, which therefore also eliminates (b).

(C6) There exist positive d and K such that

$$d\big(\nabla \Phi(M, \overline{M})\big)^2 \geqslant \nabla^2\Phi(M, \overline{M}) \qquad \text{for all} \quad M \notin W(K)$$

and $\overline{M}$ achieving $\inf_{M_1} \nabla \Phi(M, M_1)$.

4. Applications. From the results in the previous sections, in order to show that the $\{\xi_n\}$ constructed in (1.1) converge to an optimal design (conclusion (a) in the dichotomous theorem) we need to check the sufficient conditions discussed in Section 3 for various criteria, and the regularity conditions (A1) and (A2). For Examples 1 and 2, (A1) and (A2) certainly hold.

EXAMPLE 1. $\Phi(M) = \operatorname{tr}(M^{-1})$ (A-optimality criteria).

We need the slight assumption that there is at least one nonsingular M in $\mathfrak{M}$. The following fact will be needed in checking (C6).

Let $\lambda_1 \geqslant \cdots \geqslant \lambda_k$ be the ordered eigenvalues of M^{-1}. We can find constants $\underline{c}$ and $\bar{c}$ independent of M such that

$$\underline{c}\lambda_1^2 \leqslant \sup_x f(x)^T M^{-2} f(x) \leqslant \bar{c}\lambda_1^2. \tag{4.1}$$

The RHS inequality follows from the boundedness of $|f(x)|$ over $\mathfrak{X}$. The LHS inequality follows from the following simple argument:

Let $M^{-2} = \Sigma_{i=1}^k \lambda_i^2 z_i z_i^T$ be the spectral decomposition with $z_1, \cdots, z_k$ the normalized eigenvectors of M^{-2}. Then

$$\begin{aligned} \sup_{\mathfrak{X}} f(x)^T M^{-2} f(x) &= \sup_{\mathfrak{X}} \Sigma_{i=1}^k \lambda_i^2 \big(f(x)^T z_i\big)^2 \\ &\geqslant \sup_{\mathfrak{X}} \lambda_1^2 \big(f(x)^T z_1\big)^2 \\ &\geqslant \int_{\mathfrak{X}} \lambda_1^2 \big(f(x)^T z_1\big)^2 \xi(dx) \end{aligned}$$

where $M(\xi)$ is some *fixed* nonsingular matrix in $\mathfrak{M}$,

$$= \lambda_1^2 z_1^T M(\xi) z_1 \geqslant \lambda_1^2 \underline{c}$$

where $\underline{c}$ is the minimum eigenvalue of $M(\xi)$, since $|z_1| = 1$.

Now with $\Phi(M) = \operatorname{tr}(M^{-1})$,

$$\begin{aligned} \nabla\,\Phi(M, \overline{M}) &= \operatorname{tr}\big(M^{-2}(M - \overline{M})\big) \\ \nabla^2\Phi(M, \overline{M}) &= 2\operatorname{tr}\big(M^{-2}(M - \overline{M})M^{-1}(M - \overline{M})\big). \end{aligned}$$

Since $\overline{M}$ achieves $\inf_{M_1} \nabla\,\Phi(M, M_1)$, $\overline{M} = f(\bar{x})f(\bar{x})^T$ where $\bar{x}$ achieves $\sup_x f(x)^T M^{-2} f(x)$. Thus $(\nabla\,\Phi(M, \overline{M}))^2 = (\operatorname{tr} M^{-1} - f(\bar{x})^T M^{-2} f(\bar{x}))^2 \geqslant (\underline{c}\lambda_1^2 - \Sigma_1^k \lambda_i)^2$ for large λ_1 (from (4.1)). This shows that $(\nabla\,\Phi(M, \overline{M}))^2$ is of the order λ_1^4 as $\lambda_1 \to \infty$. It is also clear that $\nabla^2\Phi(M, \overline{M})$ is at most of the order λ_1^3 as $\lambda_1 \to \infty$. Thus $(\nabla\,\Phi(M, \overline{M}))^2$ dominates $\nabla^2\Phi(M, \overline{M})$ eventually as $\lambda_1 \to \infty$. Since $\operatorname{tr}(M^{-1}) \to \infty$ if and only if $\lambda_1 \to \infty$, condition (C6) is established.

EXAMPLE 2. $\Phi_p(M) = \operatorname{tr}(AM^{-p})$ for A positive definite and p positive integer (Kiefer's Φ_p-optimality criteria).

Convexity of $\Phi_p(M)$ was shown in Kiefer (1974). By a linear transformation on $\mathfrak{M}$, we can assume, with no loss of generality, that $A = I$.

$$\begin{aligned} \nabla\,\Phi_p(M, \overline{M}) &= p\operatorname{tr}\big(M^{-p-1}(M - \overline{M})\big) \\ \nabla^2\Phi_p(M, \overline{M}) &= p\Sigma_{r+s=p+2;\, r,s\geqslant 1} \operatorname{tr}\big[M^{-r}(M - \overline{M})M^{-s}(M - \overline{M})\big]. \end{aligned}$$

With the same technique developed in Example 1, it is easy to see that $(\nabla\Phi_p(M, \overline{M}))^2 \simeq \lambda_1^{2(p+1)}$ and $\nabla^2\Phi_p(M, \overline{M}) \simeq \lambda_1^{p+2}$ as $\lambda_1 \to \infty$. Condition (C6) is thus established. The result actually holds for any $p > 0$, but the calculation of $\nabla\,\Phi$ and $\nabla^2\Phi$ is much more tedious.

EXAMPLE 3. $\Phi(M) = -\log\det M^*$, M^* is the information matrix for estimating the first s ($s < k$) parameters of $\boldsymbol{\theta}$ (D_s-optimality criterion).

For nonsingular M, make the partitions

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}, \quad M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{12T} & M^{22} \end{bmatrix}$$

where $f_1(x)$ is s dimensional and M_{11} and M^{11} are $s \times s$. (Therefore $\det M^* = \det M/\det M_{22}$.) Define

$$d(x, M) = f(x)^T M^{-1} f(x), \quad d_2(x, M) = f_2(x)^T M_{22}^{-1} f_2(x),$$

$$\bar{d}^*(M) = \max_x (d(x, M) - d_2(x, M)) = d(\bar{x}, M) - d_2(\bar{x}, M) = d^*(\bar{x}, M).$$

The following conditions are sufficient for convergence of the algorithm in the dichotomous theorem.

(4.2) (1) $M(\xi_0)$ and $M(\xi^*)$ (the information matrix of an optimal ξ^*) are nonsingular.

(2) For any nonsingular M,

(i) $\sup\{d(\bar{x}, M) : \bar{d}^*(M) \leqslant K < \infty\} < \infty$

(ii) $\limsup \dfrac{d_2(\bar{x}, M)}{d(\bar{x}, M)} < 1$ as $\bar{d}^*(M) \to \infty$.

Since the proofs in Sections 2 and 3 only involve the convex combinations of $\{M(\xi_i)\}_0^\infty$ and $M(\xi^*)$, from assumption 1, we can restrict the following discussions to nonsingular M's.

$$\nabla \Phi(M, \overline{M}) = \min_{M_1} \nabla \Phi(M, M_1) = -\max_x (d(x, M) - d_2(x, M) - s)$$
$$= -\left(\bar{d}^*(M) - s\right),$$
$$\nabla^2 \Phi(M, \overline{M}) = \operatorname{tr} M^{-1}(\overline{M} - M) M^{-1}(\overline{M} - M)$$
$$- \operatorname{tr} M_{22}^{-1}\left(\overline{M}_{22} - M_{22}\right) M_{22}^{-1}\left(\overline{M}_{22} - M_{22}\right)$$
$$= \left(d(\bar{x}, M)^2 - 2d(\bar{x}, M) + k\right) - \left(d_2(\bar{x}, M)^2 - 2d_2(\bar{x}, M) + k - s\right)$$
$$= s + \bar{d}^*(M)(d(\bar{x}, M) + d_2(\bar{x}, M) - 2),$$

where $\overline{M} = f(\bar{x}) f(\bar{x})^T$.

According to assumption 2(i), $\nabla^2 \Phi(M, \overline{M})$ is bounded for $\bar{d}^*(M) \leqslant K < \infty$ (this is equivalent to $\Phi(M) \leqslant K' < \infty$). Assumption (A2) is thus satisfied while it is known that (A1) is true for D_s-criterion. The sufficient condition (C6) is equivalent to: there exists $c < \infty$ such that

$$c(d(\bar{x}, M) - d_2(\bar{x}, M)) \geqslant d(\bar{x}, M) + d_2(\bar{x}, M)$$

for all M with $\Phi(M) \geqslant K$, K a constant. This is obviously implied by assumption 2(ii). Thus under assumptions 1 and 2 we have convergence to optimality for this Φ.

Using the stronger results which are available for more specific sequences such as $\alpha_n = 1/(n+1)$ (see Pázman, 1974), the conditions 2(i) and (ii) in (4.2) may be avoided. However, in the absence of such results for the present fairly arbitrary sequences, conditions of this kind may be necessary to control the behavior of $\nabla^2\Phi$ for large and small Φ.

We now show that 2(i) and (ii) in (4.2) do indeed hold for certain one-dimensional polynomial regression models. Notice first that

$$\frac{d_2(\bar{x}, M)}{d^*(\bar{x}, M)} \leqslant c \tag{4.3}$$

for a finite constant c independent of M implies both 2(i) and 2(ii) of (4.2).

Consider polynomial regression on an interval $[-1, 1]$, so that

$$f_j(x) = x^{k-j} \qquad j = 1, \cdots, k.$$

The first s parameters are to be estimated. Let ξ be a probability measure on $[-1, 1]$ for which $M(\xi)$ is nonsingular. Let $p_0(x), \cdots, p_{k-1}(x)$ be the first k orthonormal polynomials with respect to ξ. Then

$$d(x, M) = \Sigma_{j=0}^{k-1} p_j^2(x)$$

$$d^*(x, M) = \Sigma_{j=k-s}^{k-1} p_j^2(x)$$

$$d_2(x, M) = \Sigma_{j=0}^{k-s-1} p_j^2(x).$$

Thus

$$\frac{d_2(\bar{x}, M)}{d^*(\bar{x}, M)} \leqslant \frac{\Sigma_{j=0}^{k-s-1} p_j^2(\bar{x}_j)}{\Sigma_{j=k-s}^{k-1} p_j^2(\bar{x}_{k-1})} \leqslant \frac{\Sigma_{j=0}^{k-s-1} p_j^2(\bar{x}_j)}{p_{k-1}^2(\bar{x}_{k-1})}, \tag{4.4}$$

where $\bar{x}_j$ achieves $\sup_{[-1,1]}|p_j(x)|$ for $j = 0, \cdots, k-1$. Thus (4.3) is implied by

$$\left|\frac{p_j(\bar{x}_j)}{p_{j+1}(\bar{x}_{j+1})}\right| \leqslant c_j \qquad j \geqslant 0$$

where c_j is a constant independent of ξ. Such a bound on the rate of increase of orthonormal polynomials seems unavailable in the literature so we prove it here. Write

$$p_j = k_j(x - \alpha_1) \cdots (x - \alpha_j)$$

$$p_{j+1} = k_{j+1}(x - \beta_1) \cdots (x - \beta_{j+1}),$$

so that

$$\left|\frac{p_j(\bar{x}_j)}{p_{j+1}(\bar{x}_{j+1})}\right| = \left|\frac{k_j}{k_{j+1}}\right| \frac{|(\bar{x}_j - \alpha_1) \cdots (\bar{x}_j - \alpha_j)|}{|(\bar{x}_{j+1} - \beta_1) \cdots (\bar{x}_{j+1} - \beta_{j+1})|}.$$

Now $|(\bar{x}_j - \alpha_1) \cdots (\bar{x}_j - \alpha_j)| \leqslant 2^j$ since the zeros of orthogonal polynomials on $[-1, 1]$ are contained in $[-1, 1]$. Furthermore

$$|\bar{x}_{j+1} - \beta_i| \geqslant \sup_{[-1,1]}|p_{j+1}(x)|/\sup_{[-1,1]}|p'_{j+1}(x)|$$

$$\geqslant (j+1)^{-2},$$

the last inequality following from a theorem of Markov for arbitrary polynomials on $[-1, 1]$ (see Todd (1963), (4.1)). It remains to prove that $|k_j/(k_{j+1})|$ is bounded. This can be shown using properties of Hankel determinants but the following is neater. Since $p_{j+1}(x) - (k_{j+1}/k_j)xp_j(x)$ is a polynomial of degree j, multiplying through by $p_{j+1}(x)$ and integrating gives

$$\frac{k_j}{k_{j+1}} = \int_{-1}^{1} xp_j(x)p_{j+1}(x)\xi(dx),$$

so that

$$\left|\frac{k_j}{k_{j+1}}\right| \leqslant \int_{-1}^{1} |p_j(x)|\,|p_{j+1}(x)|\xi(dx)$$
$$\leqslant 1,$$

by the Cauchy-Schwarz inequality.

By collecting together the last three inequalities, we have

$$\frac{p_j(\bar{x}_j)}{p_{j+1}(\bar{x}_{j+1})} \leqslant 2^j(j+1)^{2(j+1)}.$$

This establishes (4.4) and (4.3) and hence conditions 2(i) and (ii) in (4.2) for this example. The slight generalization to an arbitrary interval $[a, b]$ is straightforward.

APPENDIX

The following notation is needed in the proof:

$\partial W(K)$ = boundary of $W(K)$ in $\mathfrak{M}$,
$\text{dist}(A, B) = \min\{|M_1 - M_2| : M_1 \in A, M_2 \in B\}$ for $A, B \subseteq \mathfrak{M}$,
$D = 2\max\{|M| : M \in \mathfrak{M}\}$,
$P(K) = \sup\{|\nabla\Phi(M)| : M \in W(K)\}$ ($= \sup\{|\nabla\Phi(M)| : M \in \partial W(K)\}$ if Φ is convex) where l^2-norm is assumed throughout.

PROOF (that condition (C3) eliminates conclusion (b)). Choose K, K' such that $\infty > K' > K > \Phi(M_0)$ and $\bar{\alpha} = \min((K - \Phi^*)/B(K'), (K' - K)/D \cdot P(K'))$ where $\Phi^* = \inf\{\Phi(M) : M \in \mathfrak{M}\}$.

There are three stages to the proof. In the first stage we prove that $M(\xi_n) \in W(K)$ implies $M(\xi_{n+1}) \in W(K')$. In the second stage the stronger result that $M(\xi_n) \in W(K)$ for all n is established. The third stage is to appeal to the dichotomous theorem.

(i) $M(\xi_n) \in W(K) \Rightarrow M(\xi_{n+1}) \in W(K')$

$$|M(\xi_{n+1}) - M(\xi_n)| = \alpha_n|M(\bar{\xi}_n) - M(\xi_n)| \leqslant \alpha_n \cdot D \leqslant \bar{\alpha} \cdot D \leqslant (K' - K)/P(K').$$

It remains to prove that $(K' - K)/P(K') \leqslant \text{dist}(\partial W(K), \partial W(K'))$. This is geometrically quite obvious. An analytic proof is the following.

Choose $M \in \partial W(K)$, $M' \in \partial W(K')$ and let $g(u) = \Phi((1-u)M + uM')$.

$$\begin{aligned} 0 \leqslant g(1) - g(0) &= K' - K \\ &= \int_0^1 \langle \nabla\Phi(M + u(M' - M)), M' - M\rangle\, du \\ &\leqslant \max_{0 \leqslant u \leqslant 1} |\nabla\Phi(M + u(M' - M))|\,|M' - M| \\ &\leqslant P(K')|M' - M|. \end{aligned}$$

This implies that

$$\text{dist}(\partial W(K), \partial W(K')) \geqslant (K' - K)/P(K'). \tag{5.1}$$

(ii) $M(\xi_n) \in W(K)$ for all n.

Since $M(\xi_0) \in W(K)$, we need only to show that $M(\xi_n) \in W(K) \Rightarrow M(\xi_{n+1}) \in W(K)$ for any n. From (i), $M(\xi_{n+1}) \in W(K')$ and the upper bound $B(K')$ of $\nabla^2\Phi$ over $W(K')$ can be invoked. This upper bound plus (2.1) gives

(5.2)

$$\begin{aligned}\Phi(M(\xi_{n+1})) - \Phi(M(\xi_n)) &\leqslant \alpha_n\big(\nabla\,\Phi\big(M(\xi_n), M(\bar{\xi}_n)\big) + \alpha_n B(K')\big)\\ &\leqslant \alpha_n(-\Phi(M(\xi_n)) + \Phi^* + \alpha_n B(K')) \text{ from (2.2).}\end{aligned}$$

CASE 1. If $\Phi(M(\xi_n)) \geqslant \Phi^* + \alpha_n \cdot B(K')$, then

$$\Phi(M(\xi_{n+1})) \leqslant \Phi(M(\xi_n)) \leqslant K \text{ from (5.2).}$$

CASE 2. If $\Phi(M(\xi_n)) < \Phi^* + \alpha_n \cdot B(K')$, then

$$\begin{aligned}\Phi(M(\xi_{n+1})) &\leqslant (1 - \alpha_n)\Phi(M(\xi_n)) + \alpha_n\Phi^* + \alpha_n^2 B(K') \text{ from (5.2)}\\ &\leqslant (1 - \alpha_n)(\Phi^* + \alpha_n B(K')) + \alpha_n\Phi^* + \alpha_n^2 B(K')\\ &= \Phi^* + \alpha_n \cdot B(K') \leqslant \Phi^* + \bar{\alpha}\cdot B(K') \leqslant K.\end{aligned}$$

(iii) From (ii) $\Phi(M(\xi_n)) \leqslant K$ for all n and the rest of the proof is exactly the same as that of the dichotomous theorem. □

REFERENCES

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

[2] ATWOOD, C. L. (1976a). Convergent design sequences for sufficiently regular optimality criteria. *Ann. Statist.* **4** 1124–1138.

[3] ATWOOD, C. L. (1976b). Convergent design sequences for sufficiently regular optimality criteria—singular case. Preprint.

[4] FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

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

[6] GRIBIK, P. R. and KORTANEK, K. O. (1977). Equivalence theorems and cutting plane algorithms for a class of experimental design problems. *SIAM J. Appl. Math.* **32** 232–259.

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

[8] LUENBERGER, D. G. (1973). *Introduction to Linear and Nonlinear Programming*. Addison-Wesley, Reading, Mass.

[9] MANGASARIAN, O. L. (1969). *Nonlinear Programming*. McGraw-Hill, New York.

[10] PÁZMAN, A. (1974). A convergence theorem in the theory of D-optimum experimental designs. *Ann. Statist.* **2** 216–218.

[11] POLAK, E. (1971). *Computational Methods in Optimization—A Unified Approach*. Academic Press, New York.

[12] POLYAK, B. T. (1967). A general method of solving extremum problems. *Soviet Math. Dokl.* **8** 593–597.

[13] SILVEY, S. D., TITTERINGTON, D. M. and TORSNEY, B. (1978). An algorithm for optimal designs on a finite design space. *Commun. Statist. A*. To appear.

[14] TITTERINGTON, D. M. (1976). Algorithms for computing D-optimal designs on a finite design space. *Proc. 1976 Conf. on Information Sciences and Systems* 213–216. Johns Hopkins Univ.

[15] TODD, J. (1963). *Introduction to the Constructive Theory of Functions*. Birkhaüser Verlag, Berlin.

[16] TSAY, J. Y. (1976). On the sequential construction of D-optimal designs. *J. Amer. Statist. Assoc.* **71** 671–674.

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

[18] WU, C. F. (1976). Contributions to optimization theory with applications to optimal design of experiments. Ph.D. thesis, Univ. of California, Berkeley.

[19] WU, C. F. (1978). Some algorithmic aspects of the theory of optimal designs. *Ann. Statist.* **6** 1286–1301.

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

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

[22] WYNN, H. P. (1975). Simple conditions for optimal design algorithms. *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) 571–579. North-Holland, Amsterdam.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE
180 QUEEN'S GATE
LONDON SW7 5HH
ENGLAND