SIAM J. NUMER. ANAL.
Vol. 9, No. 3, September 1972

A SEQUENTIAL METHOD SEEKING THE GLOBAL MAXIMUM OF A FUNCTION*

BRUNO O. SHUBERT†

Abstract. In this paper a sequential search method for finding the global maximum of an objective function is proposed. The method is applicable to an objective function of a single variable defined on a closed interval and such that some bound on its rate of change is available. The method is shown to be minimax. Computational aspects of the method are also discussed.

1. Introduction. Maximum seeking methods deal with the problem of estimating and locating the maximum of a real function f by successively evaluating or sampling the function for various values of its argument (sampling points). Such a method must prescribe a rule to choose the sampling points (to be called the *sampling rule*), and a function (to be called the *estimate*) depending, in general, on all samples obtained and approximating the unknown value of the maximum. The method is called sequential if the sampling rule utilizes the values of previous samples to determine the next sampling point.

Following Spang [10], we can distinguish three basic types of these methods: (i) gradient methods; (ii) sequential minimax search procedures; and (iii) random and grid search methods. The main drawback of methods of the first two types is that they require the function f to be unimodal, i.e., to have a single maximum only. This is obvious because these methods are all based on the "hill-climbing" principle of moving the next sampling point in the direction in which the function increases. If the function f is not unimodal, these methods will successfully reach a local maximum only, which may be of little value in most applications.

Random methods do not generally suffer this drawback. However, the maximum is found only with some probability as long as the number of samples is finite. Furthermore, being nonsequential, random methods as well as grid search require a very large number of samples to estimate and locate the maximum with reasonable confidence level and residual uncertainty (see [10]).

Several attempts have been made to combine the "hill-climbing" principle with nonsequential global search [4], [5], [7], [8], [12], et al. Roughly speaking, either a finite random or deterministic global search is used first to locate a favorable starting point for a gradient method to follow (e.g., [4], [5]), or random excursions are made occasionally from the region where the gradient search is currently being performed (e.g., [7], [12]). In the first case, there is still no guarantee that the global maximum will actually be found, while in the second case the disadvantages of random methods are still present and some rather strong regularity conditions may be required to assure convergence.

In this paper, we propose a deterministic sequential method seeking the global maximum of a function of a single real variable defined on a closed interval $\langle a, b\rangle$. Although in principle the method could be extended to functions of several variables, such an extension does not seem to be computationally feasible, unless

* Received by the editors December 9, 1969, and in revised form December 28, 1970.
† Department of Operations Analysis, Naval Postgraduate School, Monterey, California 93940.

some discrete approximation is used in the domain [9]. The same is generally true for methods not based on the gradient such as those in [1], [2], [3] and [6]. Restrictive as this may appear, there are still many important applications (see, for instance, [13] or [2, § 6]) to which we might add the problem of finding the maximum likelihood or minimum chi-square estimators for a single-parameter family of distributions.

The only assumption we make about the function f is that it is globally Lipschitzian. That is, we assume that there is a constant C, the value of which is known, and such that $|f(x) - f(x')| \leqq C|x - x'|$ for any x, x' from the domain $\langle a, b\rangle$. It is usually not too difficult to compute such a constant. For a differentiable function f this amounts to finding some upper bound on its derivative. Even if the function f is given empirically, the constant C can often be obtained from the physical nature of the function, e.g., maximum possible acceleration if f is a velocity. In any case, if it is desired to estimate the value or location of the maximum of a multimodal function with predetermined accuracy, knowledge of C or some equivalent information is necessary to determine a stopping rule no matter what method is used.

The proposed sampling rule is formally defined in the next section, where also the convergence of the method is proved. In addition, an upper bound to the estimation error is obtained. In § 3, it is shown that the sampling rule is in fact a minimax rule with respect to the class of all functions satisfying the Lipschitz condition with the same constant. Finally, in the last section we discuss the computational aspects of the method, describe a suitable algorithm and present a numerical example.

2. Convergence of the method. Let $f(x)$ be a real function defined on a closed interval $\langle a, b\rangle$ and such that there is a constant C such that for every $x \in \langle a, b\rangle$, $x' \in \langle a, b\rangle$,

$$|f(x) - f(x')| \leqq C|x - x'|. \tag{1}$$

Let

$$\phi = \max_{x \in \langle a,b\rangle} f(x)$$

be the global maximum of the function f, and let

$$\Phi = \{x \in \langle a, b\rangle : f(x) = \phi\}$$

be the set of all x for which this maximum is attained.

We define the sampling sequence $x_0, x_1, x_2, \cdots$ of points from $\langle a, b\rangle$ recursively as follows:

$$\begin{aligned} &x_0 \in \langle a, b\rangle, \\ &x_{n+1} \text{ such that } F_n(x_{n+1}) = M_n, \qquad n = 0, 1, 2, \cdots, \end{aligned} \tag{2}$$

otherwise arbitrary, where

$$M_n = \max_{x \in \langle a,b\rangle} F_n(x), \tag{3}$$

$$F_n(x) = \min_{k=0,\cdots,n} \{f(x_k) + C|x - x_k|\}. \tag{4}$$

We now prove the following theorem.

THEOREM. *As* $n \to \infty$,

$$f(x_n) \to \phi, \tag{5}$$

$$M_n \downarrow \phi, \tag{6}$$

and

$$\inf_{x \in \Phi} |x - x_n| \to 0. \tag{7}$$

Proof. Let $x_0, x_1, x_2, \cdots$ be the sampling sequence defined by (2) and let us denote by X the set of all values of this sequence:

$$X = \{x \in \langle a, b \rangle : x = x_n \text{ for some } n = 0, 1, 2, \cdots\}.$$

First, let us consider the case when all the sampling points are different, that is,

$$x_m \neq x_n \quad \text{for all } m \neq n. \tag{8}$$

The set X is then infinite and therefore has at least one limit point in $\langle a, b \rangle$. Let $\varepsilon > 0$ and let z be a limit point of X such that

$$f(z) \leqq M - \varepsilon, \tag{9}$$

where $M = \lim_{n \to \infty} M_n$ is the limit of the nonincreasing sequence $M_0, M_1, M_2, \cdots$ defined by (3). Further, let $x_{n_0}, x_{n_1}, x_{n_2}, \cdots$ be a sequence of points from X converging to the limit point z, and let $k(\varepsilon)$ be such that

$$k \geqq k(\varepsilon) \Rightarrow |x_{n_k} - z| < \varepsilon/2C, \tag{10}$$

where C is the constant from (1).

By (1), (9) and (10), for all $k \geqq k(\varepsilon)$,

$$f(x_{n_k}) \leqq C|x_{n_k} - z| + f(z) < \varepsilon/2 + f(z) < M - \varepsilon/2,$$

and by (4),

$$n \geqq n_{k(\varepsilon)} \Rightarrow F_n(x) \leqq F_{n_k}(x) \leqq C|x - x_{n_k}| + f(x_{n_k})$$

for all $x \in \langle a, b \rangle$. Hence, for $k \geqq k(\varepsilon)$,

$$n \geqq n_k \quad \text{and} \quad |x - x_{n_k}| \leqq \varepsilon/2C \quad \text{imply } F_n(x) < M,$$

and since $M_n \geqq M$ for all n, it follows from (2) and (3) that no $x_n \in X$, $n > n_k$, can lie in the interval $\langle x_{n_k} - \varepsilon/2C, x_{n_k} + \varepsilon/2C \rangle$. However, by (10), the point z belongs to the interior of this interval, which contradicts z being a limit point of X. Therefore, all limit points z of X must satisfy the inequality $f(z) > M - \varepsilon$ for all

$\varepsilon > 0$, and since by (3) and (4) for all $x \in \langle a, b \rangle$ and all n,

$$f(x) \leqq M_n, \quad \text{and hence also} \quad f(x) \leqq M, \tag{11}$$

we must have

$$f(z) = M \tag{12}$$

for all limit points z of the sampling sequence. Finally, by condition (1), the function f must be continuous on $\langle a, b \rangle$ so that all three statements of the theorem follow from (12) and from the obvious fact that $f(x_n) \leqq \phi \leqq M$.

It remains to consider the case when assumption (8) is not satisfied, that is, when

$$x_m = x_n \quad \text{for } m > n. \tag{13}$$

First notice that by (4), for all $i = 0, 1, \cdots, j; j = 0, 1, 2, \cdots$,

$$F_j(x_i) \leqq f(x_i). \tag{14}$$

(In fact, $F_j(x_i) = f(x_i)$ since $f(x) \leqq F_j(x)$ for all $x \in \langle a, b \rangle$.) By the construction of the sampling sequence and by (13), (14) and (11), we have

$$M_{m-1} = F_{m-1}(x_m) = F_{m-1}(x_n) \leqq f(x_n) \leqq M,$$

and since $M_n \downarrow M$, this implies

$$M_{m-1+k} = M \quad \text{for all } k = 0, 1, 2, \cdots. \tag{15}$$

Now let the integers j and k be such that $0 \leqq j \leqq k - 1$. Then by (15), (14) and (11),

$$M = M_{m-1+k} = F_{m-1+k}(x_{m+j}) \leqq f(x_{m+j}) \leqq M$$

so that

$$f(x_{m+j}) = M \quad \text{for all } j = 0, 1, 2, \cdots,$$

and the theorem follows by the same argument as in the previous case. The proof is complete.

Let us now turn our attention to the choice of an estimate ϕ_n of the global maximum ϕ. From what we have just proved it follows that we could use either the last sample $f(x_n)$ itself or the maximum M_n defined by (3). It seems more appropriate, however, to use the largest sample obtained so far:

$$\phi_n = \max \{f(x_0), \cdots, f(x_n)\}, \qquad n = 0, 1, 2, \cdots,$$

which by (6) also converges to ϕ. Furthermore, since obviously $M_n \geqq \phi \geqq \phi_n$, the difference

$$\Delta_n = M_n - \phi_n,$$

which by (5) and (6) converges to zero from above, gives a convenient upper bound to the error $\phi - \phi_n$.

It is important to estimate the rate with which this error approaches zero. To that purpose let us consider the sequence Δ_n^*, $n = 0, 1, 2, \cdots$, where

$$\Delta_n^* = \sup_{f \in \mathscr{L}_{\langle a,b\rangle}(C)} \Delta_n. \tag{16}$$

Here, $\mathscr{L}_{\langle a,b\rangle}(C)$ denotes the class of all functions f defined on the interval $\langle a, b\rangle$ and satisfying (1) with the same constant C. Again let $x_0, x_1, x_2, \cdots$ be the sampling sequence defined by (2), and let us denote for awhile the respective sample values by $y_0, y_1, y_2, \cdots$. Since Δ_n can depend only on the first $n + 1$ samples and is clearly a constant with respect to the first sample y_0, we can write for $n = 1, 2, \cdots$,

$$\Delta_n^* = \sup_{y_1 \in Y_1} \cdots \sup_{y_n \in Y_n} \Delta_n(y_1, \cdots, y_n),$$

where for $k = 1, \cdots, n$,

$$Y_k = \{f(x_k) : f \in \mathscr{L}_{\langle a,b\rangle}(C) \text{ and } f(x_j) = y_j \text{ for } j = 0, \cdots, k-1\}.$$

We are now going to show that

$$\Delta_n^* = \Delta_n(y_0, y_0, \cdots, y_0). \tag{17}$$

Let k be an integer, $0 \leqq k < n$, and let us denote

$$\hat{y}_{n-k} = \max\{y_0, \cdots, y_{n-k}\}$$

and

$$M_{n,k} = \max_{x \in \langle a,b\rangle} F_{n,k}(x),$$

where $F_{n,k}$ is defined by (4) with $y_{n-k+1}, \cdots, y_n$ all replaced by $\hat{y}_{n-k}$. Then we have

$$\Delta_n(y_1, \cdots, y_{n-k}, \hat{y}_{n-k}, \cdots, \hat{y}_{n-k}) = M_{n,k} - \hat{y}_{n-k}. \tag{18}$$

Further, let $G_{n,k}$ be defined as $F_{n,k}$ with the $(n-k)$th term $y_{n-k} + C|x - x_{n-k}|$ left out from the set so that

$$M_{n,k} = \max_{x \in \langle a,b\rangle} \min\{y_{n-k} + C|x - x_{n-k}|, G_{n,k}\},$$

and let $g = \max_{x \in \langle a,b\rangle} G_{n,k}(x)$. Then either $g > G_{n,k}(x_{n-k})$, in which case $M_{n,k} = g$ or $g = G_{n,k}(x_{n-k})$, in which case $M_{n,k} = \max\{\frac{1}{2}(G_{n,k}(x_{n-k}) + y_{n-k}), g'\}$, where $g' \leqq g$ is the second maximum of $G_{n,k}$, more precisely its largest local maximum at a point different from x_{n-k}. Thus $M_{n,k}$ as a function of y_{n-k} on Y_{n-k} is of the form

$$M_{n,k}(y_{n-k}) = \begin{cases} q & \text{for } y_{n-k} \leqq t, \\ q + \frac{1}{2}y_{n-k} & \text{for } y_{n-k} > t, \end{cases}$$

where q and t are some constants. On the other hand,

$$\hat{y}_{n-k}(y_{n-k}) = \begin{cases} \hat{y}_{n-k-1} & \text{for } y_{n-k} \leqq \hat{y}_{n-k-1}, \\ y_{n-k} & \text{for } y_{n-k} > \hat{y}_{n-k-1}, \end{cases}$$

and since always $\hat{y}_{n-k-1} \in Y_{n-k}$, the right-hand side of (18) is maximized for $y_{n-k} = \hat{y}_{n-k-1}$. Hence

$$\max_{y_{n-k} \in Y_{n-k}} \Delta_n(y_1, \cdots, y_{n-k}, \hat{y}_{n-k}, \cdots, \hat{y}_{n-k}) = \Delta_n(y_1, \cdots, y_{n-k-1}, \hat{y}_{n-k-1}, \cdots, \hat{y}_{n-k-1})$$

for all $k = 0, \cdots, n-1$, and (17) follows by induction. Thus, for any $n = 0, 1, 2, \cdots$, the supremum in (16) is attained if the function f is a constant. The case $f = \text{const.}$ is easy to analyze for an arbitrary choice of $x_0 \in \langle a, b \rangle$; if $(x_0 - a)/(b - x_0) = 2^m$ for some $m = -\infty, \cdots, -1, 0, 1, \cdots, +\infty$, then whenever $M_n = \max F_n$ decreases it decreases to $1/2$ of its previous value while the number of new maxima so arisen is twice the number of maxima created by the previous decrease of M_n. If x_0 is not chosen as above, then the process is, up to some n, the same as that just described, while from this n on M_n decreases to $1/2$ of its pre-previous value and the number of new maxima is twice that at the pre-previous decrease. Thus, the length of intervals during which M_n does not change increases with powers of two, while the values of M_n decrease with the same rate. Therefore, if $f = y_0 = \text{const.}$, $M_n - y_0 = O(1/n)$.[1] In particular, if $x_0 = \frac{1}{2}(a + b)$, $M_n - y_0 \leqq C(b - a)/n$. Hence, from (14), we conclude that Δ_n as well as the estimation error $\phi - \phi_n$ decreases to zero uniformly in $f \in \mathscr{L}_{\langle a,b \rangle}(C)$ with the rate $O(1/n)$.

At this point it may be argued that if a random or systematic grid search were used, the estimation error would decrease with the same rate. Indeed, for $f \in \mathscr{L}_{\langle a,b \rangle}(C)$ we would have $\phi - \phi_n \leqq C\delta_n/2$, where δ_n is the largest distance between two neighboring sampling points, and $\delta_n = O(1/n)$, in particular, $\delta_n = (b - a)/n$ for grid search [10, p. 362]. However, since these methods are non-sequential, the rate would remain the same for all functions $f \in \mathscr{L}_{\langle a,b \rangle}(C)$. With our method, on the other hand, the rate $O(1/n)$ actually occurs only in the most unfavorable case $f = \text{const.}$, while for other functions, the error will decrease the faster the more pronounced is the maximum.

Besides estimating the value, we may also want to locate the global maximum, that is, to determine the set Φ of all $x \in \langle a, b \rangle$ at which the maximum is attained. For $n = 0, 1, 2, \cdots$, let

$$\Phi_n = \{x \in \langle a, b \rangle : F_n(x) \geqq \phi_n\}, \tag{19}$$

where F_n is the function (4). Since $\phi_n \leqq \phi_{n+1} \leqq \phi$ and $f(x) \leqq F_{n+1}(x) \leqq F_n(x)$ for all $x \in \langle a, b \rangle$, we have

$$\Phi \subset \Phi_{n+1} \subset \Phi_n, \qquad n = 0, 1, 2, \cdots.$$

Thus, the set (19) may be called a *set of uncertainty* in the location of the maximum after n samples. Furthermore, if the set-theoretical limit $\Phi_\infty = \bigcap_{n=0}^{\infty} \Phi_n$ contains an open interval I, then $F_n(x) \geqq \phi$ for all $x \in I$ and all $n = 0, 1, 2, \cdots$, and hence the sampling sequence $x_0, x_1, x_2, \cdots$ must have a limit point in I. In view of (7), this implies that the set-theoretical difference $\Phi_\infty - \Phi$ consists of at most a countable number of isolated points. Consequently, as $n \to \infty$, the Lebesgue measure of this difference converges to zero. This convergence, however, is not uniform in $f \in \mathscr{L}_{\langle a,b \rangle}(C)$; take, for instance, $f(x) = px$, $x \in \langle 0, 1 \rangle$, $0 < p \leqq 1$.

[1] If $\{a_n\}$ is a sequence of numbers, the symbol $a_n = O(1/n)$ means that there is a constant $K > 0$ such that $|a_n| \leqq K/n$ for all n.

3. Optimality of the sampling rule. We shall now study the sampling rule (2) as a member of the class $S_{\langle a,b\rangle}$ of all sequential sampling rules. Any such sampling rule s is determined by a sequence of functions $s = (\sigma_0, \sigma_1, \cdots)$, where $\sigma_0 = \xi_0 \in \langle a, b\rangle$, $\sigma_1(\xi_0, f(\xi_0)) = \xi_1 \in \langle a, b\rangle, \cdots,$

$$\sigma_n(\xi_0, \cdots, \xi_{n-1}; f(\xi_0), \cdots, f(\xi_{n-1})) = \xi_n \in \langle a, b\rangle, \cdots$$

is the sampling sequence. With any such sampling rule $s \in S_{\langle a,b\rangle}$ and any function $f \in \mathscr{L}_{\langle a,b\rangle}(C)$, we associate the error

$$e_n(f, s) = \phi(f) - \phi_n(f, s), \qquad n = 0, 1, 2, \cdots,$$

where $\phi(f)$ is the global maximum of f and $\phi_n(f, s)$ is the estimate

$$\phi_n(f, s) = \max\{f(\xi_0), \cdots, f(\xi_n)\}. \tag{20}$$

Let us try to find the minimax sampling rule s^*, that is, an $s^* \in S_{\langle a,b\rangle}$ such that

$$\sup_{f\in\mathscr{L}_{\langle a,b\rangle}(C)} e_n(f, s^*) = \inf_{s\in S_{\langle a,b\rangle}} \sup_{f\in\mathscr{L}_{\langle a,b\rangle}(C)} e_n(f, s). \tag{21}$$

Let $f_0 \in \mathscr{L}_{\langle a,b\rangle}(C)$, and for $k = 0, 1, 2, \cdots$, let $f_{k+1} \in \mathscr{L}_{\langle a,b\rangle}(C)$ be such that $f_{k+1}(\xi_j) = f_k(\xi_j)$ for $j = 0, \cdots, k$. Then the right-hand side of (21) can be written as

$$\inf_{\sigma_0} \sup_{f_0} \cdots \inf_{\sigma_n} \sup_{f_n} e_n(f, s).$$

Further, for $k = 0, 1, 2, \cdots$, let

$$F_k(x) = \min_{j=0,\cdots,k} \{f_k(\xi_j) + C|x - \xi_j|\}, \tag{22}$$

$$M_k = \max_{x\in\langle a,b\rangle} F_k(x),$$

$$\hat{y}_k = \max\{f_k(\xi_0), \cdots, f_k(\xi_k)\},$$

and let (α_k, β_k) be an interval in $\langle a, b\rangle$ such that

$$x \in (\alpha_k, \beta_k) \Rightarrow F_k(x) > \hat{y}_k,$$

and such that M_k is attained for some $x_{k+1} \in (\alpha_k, \beta_k)$. Considering now $\sup_{f_n} e_n \cdot (f_n, s)$ as a function of $\xi_n = \sigma_n(\xi_0, \cdots, \xi_{n-1}; f_n(\xi_0), \cdots, f_n(\xi_{n-1}))$, we have

$$\sup_{f_n} e_n(f_n, s) = \begin{cases} M_{n-1} - \hat{y}_{n-1} & \text{if } \xi_n \notin \langle \alpha_{n-1}, \beta_{n-1}\rangle, \\ \max\{C/2|\xi_n - \alpha_{n-1}|, C/2|\xi_n - \beta_{n-1}|, M'_{n-1} - \hat{y}_{n-1}\} & \\ \qquad \text{if } \xi_n \in \langle \alpha_{n-1}, \beta_{n-1}\rangle, & \end{cases}$$

where $M'_{n-1} = \max_{x\in\langle a,b\rangle - \langle \alpha_{n-1}, \beta_{n-1}\rangle} F_{n-1}(x)$ is the second maximum of F_{n-1}. Since

$$\min_{\xi_n\in\langle \alpha_{n-1}, \beta_{n-1}\rangle} \max\{C/2|\xi_n - \alpha_{n-1}|, C/2|\xi_n - \beta_{n-1}|\} = \tfrac{1}{2}(M_{n-1} - \hat{y}_{n-1})$$
$$\leqq M_{n-1} - \hat{y}_{n-1},$$

it follows that

$$\inf_{\sigma_n} \sup_{f_n} e_n(f_n, s) = \max\{\tfrac{1}{2}(M_{n-1} - \hat{y}_{n-1}), M'_{n-1} - \hat{y}_{n-1}\}, \tag{23}$$

and this is achieved for ξ_n satisfying the inequality

$$C|\xi_n - x_n| \leqq \max\{2(M'_{n-1} - \hat{y}_{n-1}) - (M_{n-1} - \hat{y}_{n-1}), 0\}. \tag{24}$$

Next, it is easily seen that (23) is maximized by choosing f_{n-1} such that $f_{n-1}(\xi_{n-1}) = \hat{y}_{n-2}$, and that the resulting expression is then minimized by choosing ξ_{n-1} such that

$$\tfrac{1}{2}(M_{n-1} - \hat{y}_{n-2}) = M'_{n-1} - \hat{y}_{n-2}.$$

This happens for $\xi_{n-1} = x_{n-1}$ and at the same time the right-hand side of (24) becomes zero so that $\xi_n = x_n$. Continuing in this manner we conclude that the right-hand side of (21) is achieved for

$$\xi_0 = \tfrac{1}{2}(a + b), \quad \xi_1 = x_1, \cdots, \xi_n = x_n. \tag{25}$$

Finally, noticing that (25) is true for every $n = 0, 1, 2, \cdots$ and that the points $x_1, x_2, \cdots$ coincide with the sampling sequence (2) of § 2 (with $x_0 = \frac{1}{2}(a + b)$), we resolve that the sampling rule (2) proposed in this paper is in fact a minimax rule s^* as defined by (21).

Remark. It can be shown by essentially the same reasoning that the sampling rule (2) is also a minimax rule for the location uncertainty. More precisely, (2) also satisfies (21) with $e_n(f, s)$ replaced by $\lambda_n(f, s)$—the Lebesgue measure of $\Phi_n(f, s) - \Phi(f)$, where $\Phi_n(f, s)$ is the set (19) with ϕ_n and F_n defined by (20) and (22) respectively and $\Phi(f) = \{x \in \langle a, b\rangle : f(x) = \phi(f)\}$.

4. The algorithm. The sampling rule defined in § 2 is not difficult to program for a digital computer. This has been done in [11], where the computation aspects of the method are investigated in detail. The typical iterative step of the corresponding algorithm can be described as follows:

For every $n = 1, 2, \cdots$ the function F_n defined by (4) must be stored in the computer's memory. This is most conveniently done by storing the ordered set of data

$$D_n = \{(t_1, z_1), \cdots, (t_{H_n}, z_{H_n})\},$$

where $z_1 \leqq z_2 \leqq \cdots \leqq z_{H_n}$, and the vectors (t_i, z_i), $i = 1, \cdots, H_n$, are coordinates of the maxima of F_n. Hence, and by (3), $z_{H_n} = M_n$, and according to (2) the next sampling point $x_{n+1} = t_{H_n}$.

The function f is now evaluated or sampled at x_{n+1},

$$f(x_{n+1}) = y_{n+1},$$

and the memory content D_n is updated by dropping (t_{H_n}, z_{H_n}) and adding two new vectors (t_l, z_l), (t_r, z_r), where

$$z_l = z_r = \tfrac{1}{2}(z_{H_n} + y_{n+1}),$$

$$t_l = t_{H_n} - \frac{1}{2C}(z_{H_n} - y_{n+1}), \qquad t_r = t_{H_n} + \frac{1}{2C}(z_{H_n} - y_{n+1}).$$

The new set of data D_{n+1} is then obtained by rearranging the vectors $(t_1, z_1), \cdots, (t_{H_{n-1}}, z_{H_{n-1}}), (t_l, z_l), (t_r, z_r)$ in the order of nondecreasing second component. It is

easy to see that D_{n+1} is again the set of coordinates of all the maxima of F_{n+1} and the iteration may follow.

The initial set of data depends of course on the choice of the first sampling point x_0. For instance, if $x_0 = \frac{1}{2}(a + b)$, we could start with

$$D_1 = \left\{\left(a, y_0 + \frac{C}{2}(b - a)\right), \quad \left(b, y_0 + \frac{C}{2}(b - a)\right)\right\}$$

adding only (t_r, z_r) and (t_l, z_l) in the next two iterations, respectively, and then continuing as described above.

The iterations continue until $z_{H_n} - \phi_n \leqq \varepsilon$, where $\varepsilon > 0$ is the desired accuracy in estimating the maximum ϕ and ϕ_n is the estimate updated at each iteration by $\phi_{n+1} = \max\{\phi_n, y_{n+1}\}$.

The uncertainty set Φ_n defined by (16) is easily computed from the final set of data D_n as the union of the intervals

$$\left\langle t_i - \frac{1}{C}(z_i - \phi_n), \quad t_i + \frac{1}{C}(z_i - \phi_n)\right\rangle, \qquad i = 1, \cdots, H_n,$$

for which $z_i \geqq \phi_n$. It may be further simplified by merging together the intervals less apart than some predetermined distance (see [11]).

The main computational disadvantage of the algorithm as described so far is that the memory content increases by one vector (t_i, z_i) at each iteration. This can be remedied to some extent by dropping at each iteration all vectors $(t_i, z_i) \in D_n$ such that $z_i < \phi_{n+1}$. Although the function F_n is no longer being stored completely, it is easy to see that the sequence of sampling points generated by the algorithm remains the same as before. Experiments seem to indicate that the memory content then increases far more slowly, depending, of course, on the function f. (If $f = \text{const.}$, it will still increase by one vector per iteration.)

The algorithm has been tested for various functions f using the IBM 360/67 computer. For instance, we tried the trigonometric polynomial

$$f(x) = \sum_{k=1}^{5} k \sin((k + 1)x + k)$$

on the interval $\langle -10, 10\rangle$. This function has the global maximum $\phi = 12.0313 \cdots$ at two different points $x = -6.7745 \cdots$ and $x = 5.7918 \cdots$, and 19 other local maxima no less than 1.2 apart. The constant $C = 70$ was obtained as a simple upper bound to the derivative

$$\left|\frac{df}{dx}\right| \leqq \sum_{k=1}^{5} k(k + 1) = 70.$$

For the desired accuracy $\varepsilon = 0.01$, the algorithm required 444 iterations (samples) giving the estimate $\phi_n = 12.0313$. The residual uncertainty in the location was reduced to three intervals: $\langle -6.7907, -6.7595\rangle$, $\langle -0.5129, -0.4261\rangle$, $\langle 5.7749, 5.8061\rangle$. (There is a local maximum $f(x) = 12.0312 \cdots$ at $x = -0.4914 \cdots$.) The largest number of vectors (t_i, z_i) stored was less than 250; the total computer time needed (including that for computing the values $f(x_n)$) was 13 seconds.

All this compares very favorably with nonsequential methods. For example, the grid search would require, for the same accuracy $\varepsilon = 0.01$ and with the same constant $C = 70$, placing the sampling points no farther than $\delta = 0.02/70$ apart, which would in turn require as many as $20/\delta = 7.10^4$ samples. A random method would require an even larger number of samples for a sufficient confidence level (see [10]). This indicates that the sequential method is preferable, especially in cases when the function f is difficult or time consuming to evaluate or when the sampling of f is costly.

REFERENCES

[1] J. H. Beamer and D. J. Wilde, *Minimax optimization of unimodal function by variable block search*, Management Sci., 16 (1970), pp. 529–541.

[2] G. Berman, *Minimization by successive approximation*, this Journal, 3 (1966), pp. 123–133.

[3] ———, *Lattice approximation to the minima of functions of several variables*, J. Assoc. Comput. Mach., 16 (1969), pp. 286–294.

[4] H. O. Hartley and R. C. Pfaffenberger, *Statistical control for non-linear optimization*, Tech. Rep. 18, Institute of Statistics, Texas A & M University, College Station, 1969.

[5] J. D. Hill, *A search technique for multimodal surfaces*, IEEE Trans. System Science and Cybernetics, SSC-5 (1969), pp. 2–8.

[6] J. Kiefer, *Optimum sequential search and approximation methods under minimum regularity assumptions*, J. Soc. Indust. Appl. Math., 5 (1957), pp. 105–136.

[7] J. Matyas, *Random optimization*, Automat. Remote Control, 26 (1965), pp. 244–251.

[8] S. A. Pijavskii, *An algorithm for finding the absolute minimum of functions*, Theory of Optimal Solutions, No. 2, Akad. Nauk Ukrain, SSR, Kiev, 1967.

[9] B. O. Shubert, *Sequential optimization of multimodal discrete function with bounded rate of change*, Tech. Rep. NPS55SY70101A, Naval Postgraduate School, Monterey, Calif., 1970.

[10] H. A. Spang, III, *A review of minimization technique for nonlinear functions*, SIAM Rev., 4 (1962), pp. 343–365.

[11] R. L. Springfield, *A computerized algorithm for sequential search of the global maximum*. M.S. thesis, Dept. of Operations Analysis, Naval Postgraduate School, Monterey, Calif., 1970.

[12] E. M. Vaysbord and D. B. Yudin, *Multiextremal stochastic approximation*, Engrg. Cybernetics, 5 (1968), pp. 1–10.

[13] D. J. Wilde, *Optimum Seeking Methods*, Prentice-Hall, Englewood Cliffs, N.J., 1964.