SINGLE-STEP BAYESIAN SEARCH METHOD FOR AN EXTREMUM OF FUNCTIONS OF A SINGLE VARIABLE

A. G. Zhilinskas

UDC 62–50:519.83

Search methods for an extremum with the least average error are called Bayesian if the averaging is carried out by using an *a priori* probability distribution within the class of optimized functions [1]. As it was noted in [1], it is very difficult to implement Bayesian methods in the form of digital-computer programs; therefore related methods such as Bayesian search methods with finite memory [1-3] or single-step Bayesian methods [1, 4, 5] are of interest.

In the present article it is proposed that realizations of the Wiener process be minimized. An optimal search method in the sense of mean improvement per single step is described, that is, a single-step Bayesian method (SSBM). It is shown that irrespective of the adopted SSBM for searching for the parameter of a Wiener process the SSBM converges for any continuous function. Using digital-computer simulation the convergence characteristics of an SSBM are determined in the case of minimizing a Wiener process realization. An example is given of optimizing a multiextremal (piecewise-quadratic) function as well as an easy exposition of the search strategy.

DESCRIPTION OF THE SEARCH METHOD

Let us suppose that a real function $f(x)$ $(0 \leqslant x \leqslant 1)$ is to be minimized. To construct a search method it is necessary to determine the following:

a) the point of the k-th test x_k $(0 \leqslant x_k \leqslant 1)$ as a function of $Z_{k-1} = ((f(x_1), x_1), \ldots, (f(x_{k-1}), x_{k-1}))$ $k = 1, 2, \ldots, Z_0 = 0$;

b) x_{0k} as a function of Z_k where x_{0k} denotes the position of the minimum after the k-th step, that is, the k-th approximation of the extremum points.

The search method is now constructed based on the concept of single-step Bayesian optimality [1, 4, 5]. Let $f(x)$ be a portion of a realization of the Wiener process with the parameter σ and $f(0) = 0$ [6].

At the (k + 1)-th step of the procedure the coordinates of the preceding tests and the associated values are already known. They are ordered according to the coordinate x_i and are denoted by $x_i^k (i = \overline{1, k})$:

$$x_0^k = 0 \leqslant x_1^k \leqslant \ldots \leqslant x_k^k \leqslant 1, \ \{x_i^k, i = \overline{1, k}\} = \{x_i, i = \overline{1, k}\}, \tag{1}$$

the vector $(x_1^k, \ldots, x_k^k)$ being regarded as constant. One adopts as the solution x_{0k} the minimum point of the conditional mathematical expectation of $f(x)$ relative to $\widetilde{Z}_k = ((f(x_1^k), x_1^k), \ldots, (f(x_k^k), x_k^k))$:

$$x_{0k} = \arg\min_{0 \leqslant x \leqslant 1} M\{f(x) \mid \widetilde{Z}_k\}. \tag{2}$$

Since $M\{f(x) \mid \widetilde{Z}_k\}$ is a piecewise-linear function of x [5, 7] therefore $x_{0k} = \arg\min_{x_i^k} f(x_i^k)$. The coordinate x_{k+1} of the (k + 1)-th test is selected in such a way that the mean improvement on the (k + 1)-th step $\varphi_{k+1}(x)$ is maximized:

$$x_{k+1} = \arg\max_{0 \leqslant x \leqslant 1} \varphi_{k+1}(x) = \arg\max_{0 \leqslant x \leqslant 1} (f(x_{0k}) - M\{f(x_{0k+1}) \mid \widetilde{Z}_k\}) = \arg\max_{0 \leqslant x \leqslant 1} (f(x_{0k}) - M\{\min(f(x_{0k}), f(x)) \mid \widetilde{Z}_k\}).$$

It is easily shown that

Translated from Kibernetika, No. 1, pp. 139-144, January-February, 1975. Original article submitted January 17, 1973.

$$\varphi_{k+1}(x) = \sigma_k(x) \int_{-\infty}^{\frac{f(x_{0k}) - m_k(x)}{\sigma_k(x)}} \Pi(z)\,dz, \quad k = 0, 1, \ldots, \tag{3}$$

where $m_k(x) = M\{f(x) \mid \tilde{Z}_k\}$, $\sigma_k^2(x) = D\{f(x) \mid \tilde{Z}_k\}$, $k = 1, 2, \ldots$, that is, the conditional mathematical expectation and the conditional variance of $f(x)$ are given, respectively, by

$$m_0(x) = 0, \; \sigma_0^2(x) = \sigma^2 \cdot x, \; f(x_{00}) = 0, \tag{4}$$

$$\Pi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

It follows from (2)-(4) that $x_1 = 1$; by using (1) it is therefore found that $x_k^k = 1$, $k = 1, 2, \ldots$. Then [5, 7] for $x_{i-1}^k < x < x_i^k$, $i = \overline{1, k}$, one has

$$m_k(x) = \frac{f(x_i^k)(x - x_{i-1}^k) + f(x_{i-1}^k)(x_i^k - x)}{x_i^k - x_{i-1}^k},$$

$$\sigma_k(x) = \sigma \sqrt{\frac{(x - x_{i-1}^k)(x_i^k - x)}{x_i^k - x_{i-1}^k}};$$

$$m_k(x_i^k) = f(x_i^k), \; \sigma_k(x_i^k) = 0. \tag{5}$$

To determine the coordinate of the (k + 1)-th test it is necessary to find the maximum of $\varphi_{k+1}(x)$ in the interval [0, 1], that is, to find the maxima of $\varphi_{k+1}(x)$ in the subintervals $[x_{i-1}^k, x_i^k]$, $i = \overline{1, k}$, and then compare them. Since the Wiener process (5) is Markovian in the case of $f(x_{0k}) = f(x_{0k-1})$, therefore it is sufficient to find the maxima of $\varphi_{k+1}(x)$ in the intervals $[x_{l-1}^k, x_l^k]$, $[x_l^k, x_{l+1}^k]$, at the (k + 1)-th step where

$$l = \arg \max_{1 \leqslant i \leqslant k-1} \left(\max_{x_{i-1}^{k-1} \leqslant x \leqslant x_i^{k-1}} \varphi_k(x) \right), \; x_l^k = x_k,$$

since

$$\max_{x_{i-1}^k \leqslant x \leqslant x_i^k} \varphi_{k+1}(x) = \max_{x_{i-1}^{k-1} \leqslant x \leqslant x_i^{k-1}} \varphi_k(x), \; 1 \leqslant i \leqslant l-1,$$

$$\max_{x_i^k \leqslant x \leqslant x_{i+1}^k} \varphi_{k+1}(x) = \max_{x_{i-1}^{k-1} \leqslant x \leqslant x^{k-1}} \varphi_k(x), \; l+1 \leqslant i \leqslant k-1.$$

THEOREM. For $x_{i-1}^k \leqslant x \leqslant x_i^k$ the function $\varphi_{k+1}(x)$, k = 1, 2, ..., is unimodal and convex upwards in a neighborhood of the maximum.

Proof. The following notation is introduced: if $f(x_{i-1}^k) < f(x_i^k)$, then

$$t_{ki} = \frac{x_i^k - x}{x_i^k - x_{i-1}^k},$$

$$z_{ki-1} = f(x_{i-1}^k) - f(x_{0k}), \; z_{ki} = f(x_i^k) - f(x_{0k}),$$

otherwise $t_{ki} = \dfrac{x - x_{i-1}^k}{x_i^k - x_{i-1}^k}$,

$$z_{ki-1} = f(x_i^k) - f(x_{0k}), \; z_{ki} = f(x_{i-1}^k) - f(x_{0k}).$$

Then for $x_{i-1}^k \leqslant x \leqslant x_i^k$, $0 \leqslant t_{ki} \leqslant 1$ one has

$$\varphi_{k+1}(x) = \tilde{\varphi}_{k+1}(t_{ki}) = c_{ki} \sqrt{t_{ki}(1 - t_{ki})} \int_{-\infty}^{-\frac{z_{ki-1} t_{ki} + z_{ki}(1 - t_{ki})}{c_{ki} \sqrt{t_{ki}(1 - t_{ki})}}} \Pi(z)\,dz,$$

where $c_{ki} = \sigma \sqrt{x_i^k - x_{i-1}^k}$.

The first derivative of the function $\tilde{\varphi}_{k+1}(t_{ki})$ shows that $\tilde{\varphi}_{k+1}(t_{ki})$ increases for $t_{ki} = 0$ and decreases for $t_{ki} = 1$. Consequently, the maximum point is an interior point of the interval (0, 1). The second derivative of $\tilde{\varphi}_{k+1}(t_{ki})$ shows that $\tilde{\varphi}_{k+1}(t_{ki})$ is convex upwards for $a_{ki} \leqslant t_{ki} \leqslant b_{ki}$ and convex downwards for $0 \leqslant t_{ki} < a_{ki}$, $b_{ki} < t_{ki} \leqslant 1$, where

$$a_{ki} = \frac{1}{2} + \frac{z_{ki}^2 - z_{ki-1}^2 - \sqrt{c_{ki}^4 + 4c_{ki}^2 z_{ki} z_{ki-1}}}{2(z_{ki-1} + z_{ki})^2 + c_{ki}^2},$$

$$b_{ki} = \frac{1}{2} + \frac{z_{ki}^2 - z_{ki-1}^2 + \sqrt{c_{ki}^4 + 4c_{ki}^2 z_{ki} z_{ki-1}}}{2(z_{ki-1} + z_{ki})^2 + c_{ki}^2}.$$

Consequently, $\tilde{\varphi}_{k+1}(t_{ki})$ $(0 \leqslant t_{ki} \leqslant 1)$ is unimodal and the maximum point t_{ki}^* satisfies the inequality

$$a_{ki} \leqslant t_{ki}^* \leqslant b_{ki}. \tag{6}$$

COROLLARY. It follows from the expression for the function $\tilde{\varphi}_{k+1}(t_{ki})$ that the point t_{ki}^* lies between the maxima of the function $t_k(1 - t_{ki})$ and $-\frac{z_{ki-1} + z_{ki}(1 - t_{ki})}{\sqrt{t_{ki}(1 - t_k)}}$. By using (6) one obtains

$$\max\left(\frac{1}{2}, a_{ki}\right) \leqslant t_{ki}^* \leqslant \min\left(\frac{z_{ki}}{z_{ki-1} + z_{ki}}, b_{ki}\right).$$

Since the point x_{0k} is determined in an elementary way the implementation of the algorithm reduces to the determination of the point x_k. In accordance with the theorem, the point t_{ki}^*, and hence also x_k, can be calculated by the Fibonacci method with a high degree of accuracy or [since the derivative of the function $\tilde{\varphi}_{k+1}(t_{ki})$ is known] by the method of interval-halving.

In [5] a simple algorithm was described which is close to the SSBM.

PROOF OF CONVERGENCE

The SSBM was constructed as procedure optimal in some sense for a Wiener process characterized by the parameter σ. However, it can also be analyzed as not related to a Wiener process, namely, as a search method dependent on the parameter σ. The convergence of the method in a class of continuous functions is, therefore, of interest. In the investigation of this problem the properties of the maximum of the function $\tilde{\varphi}_{k+1}(t_{ki})$ play an essential part.

LEMMA 1. $\tilde{\varphi}_{k+1}(t_{ki}^*)$ is a not increasing function of z_{ki-1}, z_{ki} and a not decreasing function of c_{ki}.

Proof. It is obvious that $\tilde{\varphi}_{k+1}(t_{ki})$ is a decreasing function of z_{ki-1}, z_{ki} and an increasing function of c_{ki} with t_{ki} kept constant. Consequently, $\tilde{\varphi}_{k+1}(t_{ki})$ is a not increasing function of z_{ki-1}, z_{ki} and a not decreasing function of c_{ki}.

LEMMA 2. t_{ki}^* is an increasing function of z_{ki} and a decreasing function of z_{ki-1}, c_{ki}.

Proof. t_{ki}^* in the form of an implicit function of z_{ki-1}, z_{ki}, c_{ki} is given by the equation

$$\left.\frac{\partial \tilde{\varphi}_{k+1}(t_{ki})}{\partial t_{ki}}\right|_{t_{ki}=t_{ki}^*} = 0.$$

It is not difficult to show that

$$\left.\frac{\partial^2 \tilde{\varphi}_{k+1}(t_{ki})}{\partial t_k^2}\right|_{t_{ki}=t_{ki}^*} < 0, \quad \left.\frac{\partial^2 \tilde{\varphi}_{k+1}(t_{ki})}{\partial t_{ki}\partial z_{ki-1}}\right|_{t_{ki}=t_{k}^*} < 0,$$
$$\left.\frac{\partial^2 \tilde{\varphi}_{k+1}(t_{ki})}{\partial t_{ki}\partial z_{ki}}\right|_{t_{ki}=t_{ki}^*} > 0, \quad \left.\frac{\partial^2 \tilde{\varphi}_{k+1}(t_{ki})}{\partial t_{ki}\partial c_k}\right|_{t_{ki}=t_{ki}^*} < 0. \tag{7}$$

In accordance with the rules for differentiating an implicit function the inequalities (7) imply the validity of the lemma.

THEOREM. For any continuous function $f(x)$ such that $f(0) = 0$ and for any constant $\sigma < 0$ one has

$$\lim_{k\to\infty} f(x_{0k}) = \min_{0\leqslant x\leqslant 1} f(x).$$

Proof. It will be shown that $\lim_{k\to\infty}(\max_{1\leqslant i\leqslant k}(x_i^k - x_{i-1}^k)) = 0$ since then the assertion of the theorem follows directly.

Let us assume the opposite; let us assume that there exists a $\Delta > 0$ such that

$$x_i^k - x_{i-1}^k = \max_{1 \leqslant i \leqslant k} (x_i^k - x_{i-1}^k) > \Delta$$

for all k = 1, 2, ...
The function $f(x)$ attains its maximum and its minimum on the interval [0, 1]; the notation is introduced

$M = \max_{0 \leqslant x \leqslant 1} f(x)$, $m = \min_{0 \leqslant x \leqslant 1} f(x)$. Since $z_{kl} \leqslant M - m$, $z_{kl} \leqslant M - m$, $x_i^k - x_{i-1}^k > 0$, therefore it follows from Lemma 1 that

$$\max_{x_{i-1}^k \leqslant x \leqslant x_i^k} \varphi_{k+1}(x) \geqslant \frac{1}{2} \sigma \sqrt{\Delta} \int_{-\infty}^{\frac{2(m-M)}{\delta \sqrt{\Delta}}} \Pi(z)\, dz.$$

But for any interval $[x_{i-1}^k, x_i^k]$ one has

$$\max_{x_{i-1}^k \leqslant x \leqslant x_i^k} \varphi_{k+1}(x) \leqslant \frac{1}{2} \sigma \sqrt{x_i^k - x_{i-1}^k} \int_{-\infty}^{0} \Pi(z)\, dz;$$

therefore in accordance with the search procedure the calculations in the intervals whose length is smaller than $g \cdot \Delta$,

$$g = \left(\int_{-\infty}^{\frac{2(m-M)}{\sigma \sqrt{\Delta}}} \Pi(z)\, dz \Big/ \int_{-\infty}^{0} \Pi(z)\, dz \right)^2,$$

will not be carried out ($0 < g < 1$).

It then follows from Lemma 2 that for all k one has

$$\min_{1 \leqslant i \leqslant k} (x_i^k - x_{i-1}^k) \geqslant h \cdot g \cdot \Delta, \tag{8}$$

where $0 < h = 1 - t^* < 1$,

$$t^* = \arg \max_{0 \leqslant t \leqslant 1} \sigma \sqrt{g \cdot \Delta \cdot t(1-t)} \int_{-\infty}^{\frac{m-M}{\sigma} \sqrt{\frac{1-t}{t \cdot g \cdot \Delta}}} \Pi(z)\, dz.$$

Now it follows from (8) that for $k > 1/(h \times g \cdot \Delta)$ one has $\sum_{i=1}^{k} (x_i^k - x_{i-1}^k) > 1$, which is not possible. The inconsistency arrived at proves the theorem.

Remark. The constraint $f(0) = 0$ is not essential since the origin can always be chosen at $f(0)$.

EXPERIMENTAL INVESTIGATION OF CONVERGENCE

The most important characteristic of the convergence method in the case of realizations of the random process $f(x)$ is its mathematical expectation $f(x_{0k})$ as a function of k. It follows from the results of the preceding section that the SSBM with the parameter σ converges for almost all realizations of the Wiener process with the parameter $\sigma_0 = 1$. It will be of some interest now to find out how the convergence depends on σ.

This problem has been investigated using a digital computer. The realizations of the Wiener process $f(x)$ have been replaced by the realizations of the process $g_N(x)$ ($0 \leqslant x \leqslant 1$):

$$g_N(x) = h(\bar{x}_{i-1}) + (h(\bar{x}_i) - h(\bar{x}_{i-1}))(x - \bar{x}_{i-1}) \cdot N,$$

$$\bar{x}_{i-1} \leqslant x \leqslant \bar{x}_i, \quad \bar{x}_i = i/N, \qquad i = \overline{1, N},$$

where $h(0) = 0$, $h(\bar{x}_i) = h(\bar{x}_{i-1}) + \xi/\sqrt{N}$, ξ is the normally distributed normalized random quantity.

The fact that for $N \to \infty$ the process $g_N(x)$ converges in some specified sense to the Wiener process with a parameter equal to unity forms the basis of our approximation, and one also has $M\{\min_{0 \leqslant x \leqslant 1} g_N(x)\} \to M\{\min_{0 \leqslant x \leqslant 1} f(x)\}$ [6]. It was assumed in the calculations that N was equal to 1000; the averaging was carried out

TABLE 1

σ \ k	15	30	45	60	75
1	—0,7212	—0,7252	—0,7255	—0,7258	—0.7259
3	—0,7515	—0,7789	—0,7799	—0.7811	—0.7813
5	—0,7430	—0,7772	—0,7834	—0.7861	—0.7867
7	—0,7247	—0.7680	—0,7811	—0.7850	—0,7869
9	—0,7131	—0,7553	—0,7734	—0,7822	—0,7852
UDE	—0,6268	—0.6927	—0.7151	—0.7294	—0.7342
SSBMO	—0,7315	—0.7716	—0.7814	—0,7855	—0,7870

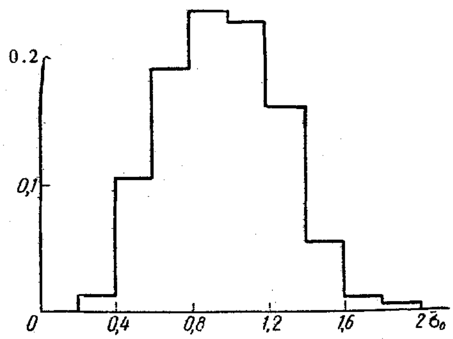

Fig. 1. Histogram of parameter estimates for a Wiener process.

over 200 realizations. The arithmetic mean of the experimentally obtained values of $\min_{0\leqslant x\leqslant 1} g_N(x)$ was equal to -0.7890. It is also noted that $M\{\min_{0\leqslant x\leqslant 1} f(x)\} = -0.7979$.

The experimental results showing the dependence of the averaged value of $g_N(x_{0k})$ on k and σ are given in Table 1. For comparison the averaged results are also given of optimization by the search method with uniform distribution of experiments (UDE) over the interval [0, 1].

In the case of minimizing realizations of the Wiener process with parameter $\sigma_0 = 1$ one can see from the table that for $k \leqslant 75$ one can select the optimal value of σ. Since the SSBM is optimal in the sense of mean improvement on one step, therefore for $k > 1$ the optimal value of σ is not unity as one could have expected.

The convergence was investigated for $\sigma_0 = 1$; the latter, however, does not limit the generality since any Wiener process can be converted into the investigated one by an appropriate change of the scale.

ESTIMATE OF THE PARAMETER

Let a realization of the Wiener process with an unknown parameter σ_0 be subject to minimization. It follows from the preceding sections that the SSBM converges for almost all realizations for any parameter σ; the rate of convergence however, depends on the choice of σ. Consequently, the necessity arises to find an estimate of σ_0; the following two facts are important:

a) only a fairly small number of experiments can be used to estimate a parameter;

b) it can be seen from the table that by varying σ in the interval $(3\sigma_0, 9\sigma_0)$ in which the optimal value of σ for $k \leqslant 75$ is found the average error varies only slightly.

Let us consider the case in which σ_0 is estimated independently for each realization by employing the maximum likelihood method using m uniformly distributed experiments:

$$\bar{\sigma}_0 = \sqrt{\sum_{i=1}^{m} (f(i/m) - f((i-1)/m))^2}. \qquad (8)^*$$

For m = 5 the estimate (8) was obtained for 200 realizations of the process $g_N(x)$. The average estimate was s = 0.9684. Its histogram is shown in Fig. 1.

Since for $30 \leqslant k \leqslant 75$ the value 7s is close to the optimal value of the SSBM parameter and since the relative frequency of the event $(3\sigma_0 < 7\bar{\sigma}_0 < 9\sigma_0)$ exceeds 0.9, therefore it is advantageous to select the SSBM parameter equal to $7\bar{\sigma}_0$. Averaged results of the search by a single-step Bayesian method with an estimate of σ_0 and with the parameter $\sigma = 7\bar{\sigma}_0$ (SSBMO) are shown in Table 1 (the experiments for estimating the parameter are counted as search experiments).

In the case of minimizing the functions which are not realizations of a Wiener process the recommended value for the parameter of the search method is $7\bar{\sigma}_0$ where $\bar{\sigma}_0$ is calculated by using the formula (8).

STOPPING RULE OF THE SEARCH

The optimal stopping rule of the search is of theoretical interest. To construct a rule of this kind one needs "the cost" of a single evaluation of the minimized function [1]. In practice, however, it is rarely possible to measure "the cost" of one evaluation in units of the minimized function; moreover, it is difficult to implement such a role. Therefore a simpler rule will be analyzed.

The specifying in advance of the number of evaluations of the minimized function is one such stopping rule. It seems to be suitable in the case of limited machine time, that is, when a single evaluation of the minimized function is very costly. The ALGOL program for SSBMO with a fixed number of evaluations is

*Numbered as in Russian original – Publisher.

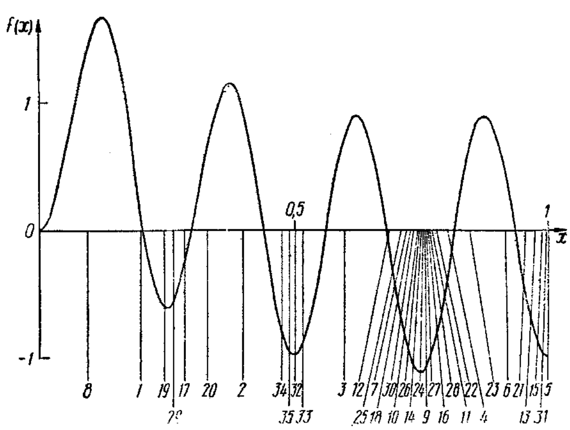

Fig. 2. Diagram of the search strategy using the SSBMO method; the points at which computations have been carried out are noted.

given in [8].

In some problems it is essential that the search be stopped when the prescribed accuracy has been reached. Since in an SSBMO algorithm one computes the mean improvement φ_k of the k-th step and regarding the latter as an accuracy estimate, the search should be stopped at the k-th step if

$$\varphi_k < \varepsilon, \quad (9)$$

where ϵ is the required accuracy in function values.

The stopping rule (9) for the search was investigated experimentally in the case of minimizing realizations of a Wiener process.

In all 200 realizations were minimized, the search being stopped in accordance with the rule (9) with ϵ = 0.05. In 180 cases the minimum was found with an accuracy ϵ the average number of calculations carried out up to the stop being 37.3.

EXAMPLES

The SSBMO was used to minimize the function

$$f(x) = \min_{1 \leqslant i \leqslant 30} \{(x - \xi_i)^2 \cdot 200 + \eta_i\}, \qquad 0 \leqslant x \leqslant 1,$$

where ξ_i are independent random quantities with uniform distribution in the interval [0, 1], η_i are independent random Gaussian quantities with zero average and standard deviation of 1.5.

The averaging was taken over 50 realizations; the average of the minima of $f(x)$ was -3.0223. The search results were as follows: $\overline{f(x_{030})} = -3.0206$, $\overline{f(x_{040})} = -3.0216$, $\overline{f(x_{050})} = -3.0219$ where bars denote minimization over 50 realizations. It is noted that after 100 steps the average of the values reached by the UDE method was -3.0208.

In Fig. 2 the search strategy for the SSBMO method is clearly shown. The minimized function was $f(x) = 2(x - 0.75)^2 + \sin(8\pi x - \pi/2) - 0.125$ $(0 \leqslant x \leqslant 1)$. The minimum of the function is at the point $x_0 = 0.75$ and then $f(x_0) = -1.125$. The search stops when the specified number N = 35 of evaluations of the minimized function have been carried out; $x_{0N} = 0.7511$, $f(x_{0N}) = -1.1246$; an improved result was achieved in the 24th experiment.

To illustrate the search with the stopping rule (9) the maximization is considered of the function

$$f(x) = \sum_{i=1}^{5} i \times \sin((i+1)x + i),$$
$$-10 \leqslant x \leqslant 10;$$

the global maximum of $f(x)$ is 12.0313 being reached at the points -6.7745 and 5.7919 [9]. The minimax method required 444 evaluations to reach the guaranteed error ϵ = 0.01 [9]. The same function was maximized by using the SSBMO algorithm with the stopping rule (9) for ϵ = 0.01. The search stopped after N = 255 evaluations of the maximized function, the best point being $x_{0N} = 5.7945$, $f(x_{0N}) = 12.0301$; other local maxima were also found with a high accuracy, namely, $x' = -6.7680$, $f(x') = 12.0245$, $x'' = -0.5014$, $f(x'') = 12.0156$.

LITERATURE CITED

1. I. B. Motskus, "Bayesian search methods of extremum," Avtomat. Vychisl. Tekh., No. 3 (1972).
2. A. G. Zhilinskas and I. B. Motskus, "A Bayesian search method for an extremum," Avtomat. Vychisl. Tekh., No. 4 (1972).
3. A. G. Zhilinskas, I. B. Motskus, and L. L. Timofeev, "A Bayesian search method for an extremum with limited memory," Avtomat. Vychisl. Tekh., No. 6 (1972).
4. V. R. Shaltyanis, "A method of multiextremal optimization," Avtomat. Vychisl. Tekh., No. 3 (1971).
5. A. G. Zhilinskas, "Optimal planning of experiments in search for an extremum of a function of a single variable," Inst. Fiz. Mat. Akad. Nauk. Litovsk. SSR, Dep. VINITI, No. 2130-70.

6. I. I. Gikhman and A. V. Skorokhod, Introduction to the Theory of Random Processes [in Russian], Nauka, Moscow (1965).
7. H. Kushner, "A versatile stochastic model of a function of unknown and time-varying form," J. Math. Anal. Appl., 5, 150-167.
8. A. G. Zhilinskas, "Single-step Bayesian method of one-dimensional search," Inst. Fiz. Mat. Akad. Nauk Litovsk. SSR, Dep. VINITI, No. 663-74.
9. B. O. Shubert, "A sequential method seeking the global maximum of a function," SIAM J. Num. Anal., 9, 379-388.