The application of Bayesian methods for seeking the extremum

Chapter · September 2014

CITATIONS
607

READS
10,956

3 authors, including:

Vytautas Tiesis
Vilnius University
15 PUBLICATIONS **636** CITATIONS
SEE PROFILE

Antanas Zilinskas
Vilnius University
190 PUBLICATIONS **2,917** CITATIONS
SEE PROFILE

9. Conclusions.

In the above paragraphs we have described a class of global methods, within the framework of stochastic automata in time-varying environment.

Some particular implementations have been suggested with the aid of results from statistics of extrema.

The methods share the troubles common to all the space-covering methods, i.e. inapplicability in high dimensions, due to the high number of cells needed for a precise description.

Tests have been performed on a number of test functions, showing the efficiency of the methods in finding global optima.

References

[1] Chichinadze, V.K.; "The Ψ-transform for solving linear and nonlinear programming problems". Automatica (1969) vol. 5, pp.347-355.

[2] Hartman, J.K.; Some experiments in global optimisation. Naval Postgraduate School - Monterey - Cal. - Tech. Rep., May 1972.

[3] Dal Castello, G.; Mereau, P.; Pacquet, J.G.; "Methode numerique pour la recherche d'un optimum global". INFOR, June 1975, pp. 185-196.

[4] Clough, D.J.; "An asymptotic extreme value sampling theory for estimation of a global maximum". CORS J., July 1969, pp.102-115.

[5] McMurtry, G.J.; Fu, K.S.; "Variable structure automaton used as a multimodal searching technique". IEEE Autom. Contr., AC-11, n. 3, July 1966, pp. 379-387.

TOWARDS GLOBAL OPTIMISATION 2
L.C.W. Dixon and G.P. Szegö (eds.)

THE APPLICATION OF BAYESIAN METHODS FOR SEEKING THE EXTREMUM

J. Mockus, V. Tiešis, A. Žilinskas
Institute of Mathematics and Cybernetics
Academy of Sciences of the Lithuanian SSR
Vilnius

The purpose of this paper is to describe how the Bayesian approach can be applied to the global optimization of multiextremal functions. The function to be minimized is considered as a realization of some stochastic function. The optimization technique based upon the minimization of the expected deviation from the extremum is called Bayesian. The implementation of Bayesian methods is considered.

The results of the application to the minimization of some standard test functions are given.

INTRODUCTION

Many well known methods for seeking the extremum have been developed on the basis of quadratic approximation. In some problems of global optimization the function to be minimized can be considered as a realization of some stochastic function. The optimization technique based upon the minimization of the expected deviation from the extremum is called Bayesian.

The description of such methods is given in [1, 2, 3]. However, to make this paper reasonably complete a brief definition of the Bayesian methods will be given.

DEFINITION OF BAYESIAN METHODS

Assume the function to be minimized is a realization $f(x,\omega)$ of some stochastic function $f(x)$, where $x \in A \subset R^n$ and $\omega \in \Omega$ is some fixed but unknown index.

The probability distribution P on Ω is defined by the equalities:

$$F_{x_1,\dots,x_m}(y_1,\dots,y_m) = P\left\{\omega : f(x_1,\omega) < y_1,\dots, f(x_m,\omega) < y_m\right\} \quad (1)$$

where P is a priori probability of an event:

$$\left\{\omega : f(x_1,\omega) < y_1,\dots, f(x_m,\omega) < y_m\right\} \quad (2)$$

The observation is the evaluation of function f at some fixed point x_i. The vector

$$z_m = (f(x_1),\dots, f(x_m), x_1,\dots, x_m)$$

contains the information gained in all the observations from 1 to m.

A decision function is the measurable vector function $d = (d_0, \ldots, d_N)$, which expresses the dependence between the point of the next observation and the results of the previous observations

$$x_{m+1} = d_m(z_m), \quad m = 0,1,\ldots,N. \tag{3}$$

The decision function d^0 is called the Bayesian method for seeking the minimum, if it minimizes the expected deviation from the extremum

$$\min_d E\left\{ f(x_{N+1}) - f_0 \right\} \tag{4}$$

where E defines the expectation and $f(x_{N+1})$ is the value of function f at the point of final decision x_{N+1}.

The criterion (1) is satisfied under some conditions [2 - 3] by the solution of the following recurrent equations

$$u_N(z_N) = \min_{x \in A} E\left\{ f(x) / z_N \right\}$$

$$u_{m-1}(z_{m-1}) = \min_{x \in A} E\left\{ u_m(z_{m-1}, f(x), x) / z_{m-1} \right\} \quad m = N,\ldots,2 \tag{5}$$

$$u_0 = \min_{x \in A} E\left\{ u_1(f(x), x) \right\}$$

where $E\left\{ f(x) / z_N \right\}$ is a conditional expectation of the random variable $f(x)$ with respect to the random vector z_m.

In accordance with the definition (4) the Bayesian method depends on a priori probability distribution P. The conditions under which the Bayesian methods converge to the minimum of any continuous function are given in [4]. For example, those conditions are satisfied for a Markov process.

THE ONE-STAGE METHOD

One of the simplifications for the solution of the equations (2) is "one-stage" method [1] when at each stage it is assumed that the next observation is the least one. In such a case the sequence of observations is defined by the equations

$$E\left\{ u(z_m, f(x_{m+1}), x_{m+1}) / z_m \right\} = \min_{x \in A} E\left\{ u(z_m, f(x), x) / z_m \right\} \tag{6}$$

where

$$u(z_{m+1}) = \min_{x \in A} E\left\{ f(x) / z_{m+1} \right\}, \quad m = 0,1,\ldots,N. \tag{7}$$

The one-stage Bayesian method converges to the minimum of any continuous function under the same conditions as eq. (5).

SELECTION OF AN A PRIORI DISTRIBUTION FUNCTION

The selection of a priori distribution function (1) is the only way to adjust the Bayesian method to a given class of optimization problems.

For simplicity it will be supposed that the set A is an n-dimensional cube

$[-1,1]^n$. One of the weakest conditions to satisfy the continuity of realizations of a stochastic function is the independence of n-th differences [6], which are a discrete approximation of an n-th derivatives. It is known [6] that the continuity of realizations and independence of the n-th differences implies the Gaussian probability distribution. Furthermore it is very reasonable to suppose that the stochastic model is homogeneous on A. All those conditions (continuity, independence of n-th differences and homogeneouity on A) are satisfied in a case of the Gaussian stochastic function of n variables with the expectation $\mathcal{M}$ and the covariance [7]

$$K_{x_j x_k} = \sigma^2 \prod_{i=1}^{n} \left(1 - \frac{|x_j^i - x_k^i|}{2}\right) \tag{8}$$

where

$$x_j^i, \ x_k^i \in [-1,1].$$

The covariance (8) can be expressed in the following form

$$K_{x_j x_k} = \frac{1}{2^n} \sum_{l=1}^{2^n} K_{x_j x_k}(x_l)$$

where x_l, $l = 1,\ldots,2^n$ are the vertices of an n-dimensional cube $-1,1^n$

$$K_{x_j x_k}(x_l) = \sigma^2 \prod_{i=1}^{n} |v_i|, \text{ and } v_i = \begin{cases} \min(x_j^i - x_l^i, \ x_k^i - x_l^i), & \text{if } x_j^i \geqslant x_l^i, \ x_k^i \geqslant x_l^i \\ \min(x_l^i - x_j^i, x_l^i - x_k^i), & \text{if } x_j^i < x_l^i, \ x_k^i < x_l^i. \end{cases}$$

Consequently, when $\mathcal{M} = 0$ the stochastic function with covariance (8) is the sum of 2^n of Wiener fields [6] with the origins on the vertices of the cube $[-1,1]^n$.

When n=1 it is the sum of two independent Wiener processes running in the opposite directions: the origin of the first is at -1, the origin of the second is at +1.

In such a way one can define a priori distribution functions correct up to some unknown parameters: $\mathcal{M}$ and σ^2. The estimation of those parameters can be realized on the basis of some number M of additional observations. It is convenient to distribute those observations uniformly with the equal probability on A.

The unbiased maximum likelihood estimations $\bar{\mathcal{M}}$ and $\bar{\sigma^2}$ of the parameters $\mathcal{M}$ and σ^2 are correspondingly:

$$\bar{\mathcal{M}} = \frac{\sum_{i,k=1}^{M} \gamma^{-1}_{x_i x_k} f(x_k)}{\sum_{i,k=1}^{M} \gamma^{-1}_{x_i x_k}}$$

and

$$\bar{\sigma}^2 = \frac{1}{M-1} \sum_{i,k=1}^{M} \rho^{-1}_{x_i x_k} \bar{z}_i \bar{z}_k$$

where $\bar{z}_i = f(x_i) - \bar{\mathcal{M}}$ and $\rho^{-1}_{x_i x_k}$ is an element of the inverse correlation matrix with the elements $\rho_{x_i x_k} = K_{x_i x_k} / \sigma^2$.

The maximum likelihood estimation $\bar{\sigma}^2$ of the only parameter σ^2 in a special case of a Wiener process on the unit interval where the distance between the additional observations are equal is very simple:

$$\bar{\sigma}^2 = \sum_{i=1}^{M} (f(i/M) - f((i-1)/M)^2$$

The computer simulation had demonstrated [8] that the effectiveness of the one-step Bayesian method (6) can be improved if the estimation of the variance σ^2 is increased by the factor $\alpha > 1$. In the case of the Wiener process the recommended [8] value of $\alpha = 7$. Apparently, it is because the increasing of the estimation of σ^2 makes a search of the extremum more "global" and so in some way approximates the influence of the consequent steps which are neglected in the case of the one-step method.

The Gaussian field with the expectation $\mathcal{M}$ and covariance

$$K_{x_i x_k} = \sigma^2 \ell^{-\sqrt{\sum_{i=1}^{M} c_i^2 (x_j^i - x_k^i)^2}} \qquad (9)$$

also were investigated [5, 9]. This function is interesting because in a special case when $n=1$ it corresponds to the stationary Markov process.

THE RELATION OF EFFECTIVENESS OF BAYESIAN SEARCH ON A PRIORI DISTRIBUTION

The arbitrariness of an a priori distribution makes it necessary to investigate the behaviour when a Bayesian method designed to minimize the realizations of one stochastic function actually is minimizing the realization of another quite different stochastic function. In [10] the one-step Bayesian method intended for the minimization of realizations of the Wiener process is applied to the minimization of realizations of the stationary Markov process with zero expectation, unit variance and exponential correlation.

The results (when the scale factor $c=1$) are given in the first row of Table 1. In the second row of this table the results are given for the one-stage Bayesian method designed for the minimization of the Markov process. Similarly in the third row the results of the Monte-Carlo method with random uniformly distributed observations are given.

Table 1
The dependence of the results of the Bayesian method on the a priori distribution

m	15	30	45	60	75
1	-0.929	-0.984	-1.009	-1.020	-1.026
2	-0.927	-0.989	-1.0011	-1.0022	-1.1027
3	-0.848	-0.907	-0.933	-0.951	-0.956

As one can see that in spite of considerable difference between the a priori distributions corresponding to row 1 and row 2 (the Wiener process is not even a stationary one), the observed results of the Bayesian methods designed on the basis of true and false stochastic models are very close. So it is possible to suppose, that the efficiency of Bayesian methods does not depend very much on the a priori distribution.

ONE-DIMENSIONAL SEARCH

For the solution of some multidimensional problems of optimization it is convenient to use the coordinate optimization technique, when along each coordinate a one-dimensional global search is carried out [11]. The one-dimensional search also can be applied to multidimensional optimization using the condition

$$\min_{x_1} \min_{x_2} \dots \min_{x_n} f(x_1, \dots, x_n) \qquad (10)$$

The condition (10) is convenient in the two-dimensional case. Therefore, methods of one-dimensional search are of special interest.

The one-dimensional one-step Bayesian method is relatively simple [8]. In the case when the a priori distribution corresponds to the Wiener process on the interval $[0,1]$ the coordinates $m+1$-th observation are

$$x_{m+1} = \arg \max_{0 \leqslant x \leqslant 1} w_{m+1}(x)$$

where

$$w_{m+1}(x) = \sigma_m(x) \int_{-\infty}^{\frac{f(x_{0m}) - \mathcal{M}_m(x)}{\sigma_m(x)}} \Pi(z)\,dz, \quad m=0,1,\dots,N$$

$$\mathcal{M}_0(x) = 0, \quad \sigma_0^2(x) = \sigma^2 x, \quad f(x_{0m}) = \min_{1 \leqslant i \leqslant m} f(x_i), \quad f(x_{00}) = 0$$

and $\Pi(z)$ is a probability integral. Let $x_{i_1}, \dots, x_{i_m}$ be the coordinates of observations ordered by the increasing values. Then

$$\mathcal{M}_m(x) = \frac{f(x_{i_{k+1}})(x - x_{i_k}) + f(x_{i_k})(x_{i_{k+1}} - x)}{x_{i_{k+1}} - x_{i_k}}$$

and

$$\sigma^2_m(x) = \frac{(x - x_{i_k})(x_{i_{k+1}} - x)}{(x_{i_{k+1}} - x_{i_k})}$$

if $x_{i_k} < x < x_{i_{k+1}}$. If $x = x_{i_k}$, $k = 1, \ldots, n-1$, then $\mathcal{M}_n(x_{i_k}) = f(x_{i_k})$ and $\sigma^2_m(x_{i_k}) = 0$.

The function $w_m(x)$ is unimodal on the interval $x_{i_k} \leqslant x \leqslant x_{i_{k+1}}$ and so it can be minimized using usual methods of one-dimensional optimization.

COMBINATION OF THE BAYESIAN METHODS AND THE METHODS OF LOCAL OPTIMIZATION

Such properties of stochastic functions as the independence of n-th differences usually do not contradict a priori notions about the "global" behaviour of real functions. Unfortunately, some important local characteristics of the realizations are very different. For example, the realizations of the Wiener process are not differentiable at almost every x. This makes the adequacy of the stochastic models (8) or (9) very doubtful when short distances are considered. The local inadequacy of the stochastic models (8) and (9) is one reason which can explain the decline in efficiency which occurs near the termination of the Bayesian procedures of optimization when the distances between the subsequent observations are usually small. Therefore, it would be very useful to detect in advance such regions where the general stochastic model no longer describes the behaviour of the function to be minimized sufficiently well because the distances between the nearest observations have become too small. In such regions it is quite reasonable to use the quadratic approximation approach which is widely used in some very efficient methods of local optimization.

In the case when the Wiener process was used as a general stochastic model [12], the following rule for the detection of intervals of local inadequacy was used. If at the k-th observation for some l_k, $4 \leqslant l_k \leqslant k-3$ the condition

$$f(x_{i_k - 1}) > f(x_{i_k}), \quad i_k = l_k - 2, \ldots, l_k \tag{11}$$

and

$$f(x_{i_k}) < f(x_{i_k + 1}), \quad i_k = l_k, \ldots, l_{k+2}$$

is satisfied, then the interval $(x_{l_k + 2}), (x_{l_k - 2})$ is considered as the interval of local inadequacy.

The probability of inequality (11) is less than 0.016 in the case of the Wiener process when $x_{i_k}, x_{i_k} = l_k - 2, \ldots, l_k$ are fixed. This inequality is quite natural, if the function $f(x)$ is unimodal at the interval $(x_{l_{k-2}}, x_{l_{k+2}})$. Therefore when such an interval is detected the corresponding local minimum is calculated correct up to $\varepsilon > 0$ using the algorithm of the parabolic approximation.

In the multidimensional case the direct detection of the regions where the function is unimodal with a given probability is more complicated. So far in such case the simplified procedure is used. The procedure is based upon the assumption, that local minima usually are in the neighbourhood of such points x_l where observed values of $f(x_l)$ are significantly lower than conditional expectation. Such points can be detected using the following inequality

$$\delta_l \leqslant \delta \tag{12}$$

where

$$\delta_l = \frac{f(x_l) - E\left\{f(x_l) / z_{l-1}\right\}}{\sigma\left\{f(x_l) / z_{l-1}\right\}},$$

$\sigma\left\{f(x_l) / z_{l-1}\right\}$ is the conditional standard deviation and δ is some fixed number which depends on a significance level. In some situations it is more convenient to fix not some significance level but the number L, of local minima. In such cases the local minimization is carried out from the L starting points with lowest values of δ_l.

NUMERICAL EXAMPLES

For the minimization of seven standard test functions, three algorithms were considered. The first algorithm was Bayesian of the type (6), (7), (8), (12) with the fixed numbers of global observations N and local minimizations L. For the local minimization with simple constraints the modification of a variable metric method [13], [14] was used. The local minimization was terminated if the norm of the gradient was less than $5 \cdot 10^{-4}$ or if the value of the function was decreasing less than $5 \cdot 10^{-4}$ in k iterations, where $k = 2$ if $n = 2$ and $k = 4$ if $n > 2$. The number of observations performed for the local minimization will be denoted N_1 and the point, that was found $x_p = (x^1_p, \ldots, x^n_p)$.

The second algorithm was a combination of the Monte Carlo method with local minimization. The first N_2 observations were random and uniformly distributed. From the best of those observations the local minimization was carried out by the same method as in the case of the first algorithm. The procedure was replated K times.

The third algorithm was of the type (10) using the Bayesian algorithm (12) for one-dimensional minimization. The same accuracy level of 0.001 was fixed for both the coordinates and the function.

The first test function considered was Branin's function, which has three equal local minima:

$$f(x^1,x^2)=a(x^2-b(x^1)^2+(x^1-d)+l(1-f)\cos x^1+l \qquad (13)$$

$$a=1,\quad b=5.1/4\pi^2,\quad c=5/\pi,\quad d=6,\quad l=10,\quad f=1/8\pi,$$

$$-5\leqslant x^1\leqslant 10,\qquad 0\leqslant x^2\leqslant 15.$$

The results of the minimization of function (13) by the Bayesian algorithm when $N=70$, $L=4$ are given in Table 2.

Table 2
The results of the optimization of Branin's function by the Bayesian algorithm

l	δ_l	x_l^1	x_l^2	x_p^1	x_p^2	N_1
19	-2.30	9.457	2.379	9.425	2.475	22
70	-2.09	-3.034	15.000	-3.142	12.275	39
39	-1.78	3.330	1.358	3.142	2.275	32
55	-1.48	3.087	2.520	3.142	2.275	26

In the case of function (13) and the Monte-Carlo algorithm when $N_2=1$, $K=300$ the average number of observations was 45. The point of the first local minimum (-3.142, 12.275) was found in 35,5 % of the cases, the point of the second local minimum (3.142, 2.275) - in 39,5 % of the cases, the point of the third local minimum (9.425, 2.475) - in 25 % cases.

The third algorithm was terminated after 1021 observations of function (13). The results are given in Table 3.

Table 3
The local minima of Branin's function which were found by the third algorithm

x^1	-3.14159	3.14159	9.42478
x^2	12.27500	2.27500	2.47500
f	0.39879	0.39879	0.39879

The second test function was that of Goldstein and Price with four local minima:

$$f(x^1,x^2)=\left[1+(x^1+x^2+1)^2(19-14x^1+3(x^1)^2-14x^2+6x^1x^2+3(x^2)^2)\right]\times$$

$$\times\left[30+(2x^1-3x^2)(18-32x^1+12(x^1)^2+48x^2-36x^1x^2+27(x^2)^2)\right] \qquad (14)$$

$$-2\leqslant x^1,\ x^2\leqslant 2.$$

The results of the minimization of the function (14) by the Bayesian algorithm when $N=70$, $L=4$ are given in Table 4.

Table 4
The results of the minimization of the Goldstein and Price function by the Bayesian algorithm

l	δ_l	x_l^1	x_l^2	x_p^1	x_p^2	N_1
59	-6.02	-0.025	-0.944	0.000	-1.000	55
26	-2.06	-0.455	-0.579	-0.600	-0.400	84
44	-1.81	1.098	-1.367	0.000	-1.000	83
45	-1.42	-0.608	-0.493	-0.600	-0.400	70

The results of the minimization of function (14) by the Monte-Carlo algorithm with $K=200$ are given in Tables 5 and 6, where β_1 is the percentage of the cases when the global minimum was found, N_3 is the average number of observations, β_2 is the percentage of the cases when the result of the optimization was in the region of a singular point (x_s^1, x_s^2).

Table 5
The results of the optimization of the Goldstein and Price function by the Monte-Carlo algorithm

N_2	1	5	10	30	50	70
β_1	50	64.5	68	81.5	89	89
N_3	92	82	79	94	112	131

Table 6
The percentage of the cases when the results of optimization by Monte-Carlo method was in the region of the singular point of the Goldstein and Price function (x_s^1, x_s^2)

x_s^1	0.000	-0.600	1.800	1.200	-0.398
x_s^2	-1.000	-0.400	0.200	0.800	-0.602
β_2	50	25.5	15.5	8	1

The third algorithm was terminated after 520 observations of function (14). One local minimum (0.00001, -1.00000) where $f=3$ was calculated with given accuracy.

The third, fourth and fifth test functions from Shekel's family were given by the formula

$$f(x)=-\sum_{i=1}^{k}\frac{1}{(x-a_i)^T(x-a_i)+c_i} \qquad (15)$$

$$x=(x^1,\ldots,x^n)^T,\quad a_i=(a_i^1,\ldots,a_i^n)^T,\quad c_i>0$$

$$0\leqslant x^j\leqslant 10,\quad j=1,\ldots,n.$$

The values of the parameters used are those specified in the introduction to this volume.

The results of the minimization of function (15) by the Bayesian algorithm N = 70, L = 5 and Monte-Carlo algorithm (K = 200) are given in Tables 7 and 8, correspondingly.

Table 7
The results of the minimization of Shekel's function by the Bayesian algorithm

k = 5			k = 7			k = 10		
δ_l / N_1	x_l	x_p	δ_l / N_1	x_l	x_p	δ_l / N_1	x_l	x_p
51	0.713	1.000	61	4.398	4.994	27	5.009	4.001
-7.51	0.815	1.000	-11.6	4.397	4.995	-4.08	2.798	4.001
117	0.570	1.000	171	3.275	3.006	216	4.831	4.000
	1.487	1.000		2.586	3.007		3.034	4.000
14	5.938	5.999	29	4.053	4.994	32	5.285	4.995
-3,68	6.803	6.000	-6.56	4.566	4.995	-2.41	6.148	4.994
137	5.398	5.999	189	2.907	3.006	243	3.654	3.008
	5.734	6.000		3.146	3.007		3.213	3.007
15	4.378	4.000	12	6.393	5.998	57	6.196	6.006
-2.98	3.370	4.000	-2.15	5.404	6.000	-2.01	1.351	2.010
227	2.892	4.000	164	6.253	5.997	261	8.077	6.004
	4.434	4.000		4.654	5.999		1.481	2.009
70	4.057	3.002	63	2.730	4.001	40	1.402	3.001
-1.99	7.739	6.998	-2.15	5.186	4.001	-1.45	6.696	7.000
155	3.358	3.002	359	6.153	3.999	189	3.909	3.001
	7.499	6.998					6.890	7.000
67	5.700	3.001	42	0.836	3.001	39	2.799	3.001
-0.45	3.037	6.998	-0.97	9.595	7.001	-0.67	8.785	7.000
468	8.856	3.001	326	0.501	3.000	230	0.617	3.001
	2.394	6.998		9.807	7.000		6.477	7.000

Table 8
The results of the minimization of Shekel's function by the Monte-Carlo algorithm where N_3 is the average number of observations, β_1 - percentage of the cases when the global minimum was found

	N_2	1	5	10	30	50	70
k = 5	β_1	38.5	38	32.5	32.5	24.5	28.5
	N_3	353	306	288	265	263	263
k = 7	β_1	44	39.5	32	33	22	32.5
	N_3	369	308	288	272	266	285
k = 10	β_1	38	34	34	30.5	21	27.5
	N_3	343	295	280	267	265	275

The sixth and seventh test functions from Hartman's family were given by the formula

$$f(x) = -\sum_{i=1}^{K} c_i \exp\left(-\sum_{j=1}^{n} \alpha_{ij}(x^j - P_{ij})^2\right) \qquad (16)$$

$$x = (x^1, \ldots, x^n), \quad \alpha_i = (\alpha_{i1}, \ldots, \alpha_{in}), \quad P_i = (P_{i1}, \ldots, P_{in})$$

The standard parameters values were used.

The results of the minimization of function (16) by the Bayesian algorithm (N = 70, L = 5) and Monte-Carlo algorithm (K = 200) are given in Tables 9, 10, correspondingly, where N_3 is the average number of observations and β_1 is the percentage of cases when the global minimum was found.

Table 9
The results of the minimization of Hartman's function by the Monte-Carlo algorithm

	N_2	1	5	10	30	50	70
n = 3	β_1	60	71	89.5	92.5	95.5	99.5
	N_3	103	97	97	118	126	155
n = 6	β_1	64	71	74	64	51	57
	N_3	321	268	259	270	278	303

Table 10
The results of the minimization of Hartman's functions by the Bayesian algorithm

n=3				n=6			
l δ_l	x_l	x_p	N_1	l δ_l	x_l	x_p	N_1
37	0.0000	0.1146	87	65	0.0692	0.2017	246
-4.84	0.5668	0.5556		-3.01	0.4275	0.1500	
	0.8297	0.8525			0.5566	0.4768	
					0.3245	0.2753	
					0.2232	0.3117	
					0.8545	0.6573	
28	0.8398	0.1146	110	5	0.5467	0.2017	257
-4.45	0.4609	0.5556		-2.00	0.3877	0.1500	
	0.8904	0.8525			0.8051	0.4768	
					0.4761	0.2753	
					0.4542	0.3117	
					0.8605	0.6573	
48	0.9573	0.1146	91	15	0.6265	0.4047	221
-2.95	0.5447	0.5556		-1.66	0.9441	0.8824	
	0.9278	0.8525			0.9181	0.8462	
					0.7578	0.5740	
					0.0802	0.1389	
					0.2041	0.0385	
24	0.4527	0.3687	78	58	0.2167	0.4077	202
-1.38	0.0000	0.1176		-1.18	0.8614	0.8824	
	0.1854	0.2675			1.0000	0.8462	
					0.9419	0.5740	
					0.9640	0.1387	
					0.0000	0.0385	
34	0.5003	0.1093	77	46	0.4026	0.4047	236
-1.38	0.8212	0.1093		-0.68	1.0000	0.8824	
	0.5331	0.5641			0.0000	0.8461	
					0.9010	0.5740	
					0.9800	0.1388	
					0.3206	0.0385	

CONCLUSIONS

The Bayesian algorithms require significantly less observations (function evaluations) than the Monte-Carlo algorithms under the conditions which were considered in this paper. However, applying the Bayesian algorithms it is necessary to perform a supplementary calculations to find the best point at the next observation. It is not so important when the observations are complicated or expensive. But when the function evaluations are simple it is reasonable to simplify the Bayesian algorithms using some approximation of the equations (6) (7).

REFERENCES

1. Mockus J.B. (1972). On Bayesian Methods for Seeking the Extremum. Automatics and Computers, No. 3, 53-62 (Russian).

2. Mockus J.B. (1974). On Bayesian Methods for Seeking the Extremal Point, Kybernetes, Vol. 3, 103-108.
3. Mockus J.B. (1975). On Bayesian Methods of Optimization, Towards Global Optimization, North-Holland, Amsterdam, 166-181.
4. Mockus J.B. (1975). On Bayesian Methods for Seeking the Extremum, Lecture Notes in Computer Science, 27, Springer Verlag, Berlin, Heidelberg, New York, 400-404.
5. Žilinskas A.G., Mockus J.B. (1970). On a Bayesian Method for Seeking the Minimum, Automatics and Computers, No. 3, 42-44 (Russian).
6. Katkauskaitė A.J. (1972). The Random Fields with Independent Increments, Lithuanian Mathematical Transaction, No. 4, 75-85 (Russian).
7. Mockus J.B. (1977). On Bayesian Methods for Seeking the Extremum and Their Application, Information Processing 77, B. Gilchrist (ed.), North-Holland, Amsterdam, 195-200.
8. Žilinskas A.G. (1975). The One-step Bayesian Method for Seeking the Extremum of the Function at One Variable, Cybernetics, No. 1, 139-144 (Russian).
9. Šaltenis V.R. (1971). On the Methods for the Multiextremal Optimization, Automatics and Computers, No. 3, 34-38 (Russian).
10. Žilinskas A.G., Timofeev L.L. (1976). On the Relation of the Efficiency of Bayesian Methods for Seeking the Extremum and the a Priori Distribution, Numerical Methods of Nonlinear Programming, Naukova Dumka, Kiev, 186-189.
11. Mockus J.B. (1976). Multiextremal Problems in Design, Nauka, Moscow (Russian).
12. Žilinskas A.G. (1974). On a Simplified Bayesian Algorithm for Seeking the Minimum, The Computer-aided Design of Engineering Units and Technological Processes, Gorky State University, Gorky, 54-57(Russian).
13. Goldfarb B.D. (1969). Extension of Davidson's Variable Metric Method to Minimization Under Linear Unequality and Equality Unstrain't, SIAM Journal Applied Mathematics, Vol. 17, No. 4, 733-764.
14. Biggs M.C. (1971). Minimization Algorithms Making use of Nonquadratic Properties of the Objective Function, Journal of the Inst. of Mathematics and Applications, Vol. 8, No. 3, 315-327.