ON BAYESIAN METHODS FOR SEEKING THE EXTREMUM

J. Močkus

Institute of Physics and Mathematics
Academy of Sciences Lithuanian SSR
Vilnius, USSR

1. Introduction

Many well known methods for seeking the extremum had 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.

2. The definition of Bayesian methods

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

The probability distribution P on Ω is defined by the equalities

$$P\left\{\omega : f(x_i, \omega) < y_i, \, i = 1, \ldots, n, \, \omega \in \Omega\right\} = F_{x_1, \ldots, x_n}(y_1, \ldots, y_n), \, n = 1, 2, \ldots$$

where $F_{x_1, \ldots, x_n}(y_1, \ldots, y_n)$, $x_i \in A$, $i = 1, \ldots, n$ is the <u>a priori</u> probability distribution function.

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

$$Z_n = \left(f(x_i), x_i, \, i = 1, \ldots, n\right), \, n = 1, \ldots, N,$$

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

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

$$x_{n+1}(d) = d_n(z_n) = x_{n+1}(d, \omega), \, n = 1, \ldots, N.$$

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

$$E\left\{f\left(x_{N+1}(d^{o})\right)-f_{0}\right\}=\inf_{d} E\left\{f\left(x_{N+1}(d)\right)-f_{0}\right\}, \qquad (1)$$

where $f_{0}=\inf\limits_{x\in A} f(x,\omega).$

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

$$u_{N}(z_{N})=\inf_{x\in A} E\left\{f(x)|z_{N}\right\},$$

$$u_{n-1}(z_{n-1})=\inf_{x\in A} E\left\{u_{n}(z_{n-1},f(x),x)|z_{n-1}\right\}, n=N,...,2,$$

$$u_{0}=\inf_{x\in A} E\left\{u_{1}(f(x),x)\right\}, \qquad (2)$$

where $E\{f(x)|z_{N}\}$ is a conditional expectation of the stochastic variable $f(x)$ with respect to the stochastic vector z_{N}.

The equations (2) remain the same in the case of "noisy" observations, when at any fixed point x_{i} we observe the sum

$$\Psi(x_{i})=f(x_{i})+\eta_{i},$$

where stochastic variables $\eta_{i}, i=1,...,N$ are independent, the expectation of η_{i} is zero and variance is fixed. In such a case

$$z_{n}=\left(\Psi(x_{i}),x_{i},i=1,...,n\right), n=1,...,N.$$

3. An illustrative example

Suppose that

$$f(x,\omega)=(x-\omega)^{2}, A=\Omega=[-1,1], N=1,$$

and the a priori density function

$$p(\omega)=\frac{1}{2}, \omega\in\Omega,$$

then in accordance with the equations (2) the first observation

$$x_{1}(d^{o})=\pm 1,$$

and the final decision

$$X_2(d^o) = \begin{cases} 1 - \sqrt{f(1)}, & X_1 = 1, \\ -1 + \sqrt{f(-1)}, & X_1 = -1. \end{cases}$$

4. The convergence conditions

In accordance with the definition (1) the Bayesian method depends on the <u>a priori</u> probability distribution P . The conditions when the Bayesian method converges to the minimum of any continuous function are given by

Theorem 1. Assume:

1. the set A is compact, the functions $f(x,\omega)$ and the conditional expectations in (2) are continuous functions of x for all $\omega \in \Omega$, there exists the positive <u>a priori</u> density function and the finite expectation $E\{|f_o|\}$;

2. for all $x \in B(\omega), \delta > 0$ there exists $n_{\delta x}$ such that

$$\left| F_x(y|z_n) - F_\omega(y) \right| < \delta, \ n \geqslant n_{\delta x} ,$$

3.

$$F_{x_i}(y|z_n) = F_\omega(y), \ i = 1,...,n$$

4. for all $x \in A(\omega), y \in R$ there exists δ_{xy} independently of n such that

$$F_x(y|z_n) \geqslant \delta_{xy},$$

5. $\sup |U_N(z_N)|$ is uniformly integrable with respect to $F_x(y|z_N)$;

then

$$\lim_{N \to \infty} \left| f(x_{N+1}(d^o, \omega), \omega) - f_0 \right| = 0, \ \omega \in \Omega,$$

where $F_x(y|z_n)$ is the conditional probability distribution function of the stochastic variable $f(x)$ with respect to the stochastic vector z_n ,

$$F_\omega(y) = \begin{cases} 0, & y \leqslant y_\omega, \\ 1, & y > y_\omega, \end{cases}$$

$y_\omega = f(x_{i(x)}, \omega)$, the index $i(x)$ is defined by the equality

$$\tau_n = \|x - x_{i(x)}\| = \min_{1 \le i \le n} \|x - x_i\|,$$

and the sets $A(\omega), B(\omega)$ are defined by the conditions

$$A(\omega) = \left\{ x : \lim_{n \to \infty} \tau_n = \varepsilon, \ \varepsilon > 0, \ x \in A \right\},$$

$$B(\omega) = \left\{ x : \lim_{n \to \infty} \tau_n = 0, \ x \in A \right\},$$

the decision function d^0 corresponds to the solution of the equations (2) when conditional expectations are calculated using the usual formulas of the conditional probability density functions.

5. The one-stage method

One of the simplifications for the solution of the equations (2) is "one-stage" method [1] [3] when at each stage it is assumed that the following observation is the last one. In such a case the sequence of observations is defined by the equations

$$E\left\{ u(z_n, f(x_{n+1}), x_{n+1}) | z_n \right\} = \min_{x \in A} E\left\{ u(z_n, f(x), x) | z_n \right\}$$

where

$$u(z_{n+1}) = \min_{x \in A} E\left\{ f(x) | z_{n+1} \right\}, \ n = 0, \ldots, N.$$

The one-stage Bayesian method converges to the minimum of any continuous function under the conditions of theorem 1.

6. The restricted-memory case

The next simplification for the equations (2) is the restriction of "memory", when one can remember no more than $N_0 < N$ observations [1] [4]. The Bayesian method for seeking the extremum in a special case with $N_0 = 1$ happens to be similar to those of the well known methods of "random search" [5], when the next observation must be on the sphere with the centre at the point of the best observation.

In the case of the restricted memory the Bayesian methods do not necessarily converge.

7. The implementation

The one-stage method had been implemented in the case when the function to be minimized is considered as some realization of the stochastic Gaussian field with the expectation $\mathcal{M}$, standard devia-

tion 6 and the exponential correlation function

$$\varrho\left(x', x''\right) = exp\left(-\sqrt{\sum_{i=1}^{m} c_i^2 \left(x_i' - x_i''\right)^2}\right),$$

the parameters $\mathcal{M}, 6, c_i, i = 1,...,m$ are estimated by the maximum likelihood method on the basis of some additional observations.

The implementation of one-stage Bayesian method when $f(x)$ is the Viener process is considered in [6].

The development of some system of a priori distributions suitable for different classes of the functions f is probably the most important problem in the application of Bayesian approach to the global optimization.

The Bayesian methods had been applied to some practical problems [7] [8].

References

1. J. Mockus. On a Bayesian method for seeking an extremum. Automatika i vychislitelnaja tekhnika, N 3, 1972.
2. J. Mockus. On the Bayesian methods of optimization. In "Minimization Algorithms and Dynamic Systems" ed. by G. Szego and L. Dixon. North-Holland Publ., Amsterdam (in print).
3. V. Šaltenis. On a method of multiextremal optimization. Automatika i vychislitelnaja tekhnika, N 3, 1973.
4. A. Žilinskas and J. Mockus. On a Bayes method for seeking an extremum. Automatika i vychislitelnaja tekhnika, N 3, 1972.
5. L.A. Rastrigin. The stochastic methods of search. Nauka, Moscow, 1968.
6. A. Žilinskas. One-stage Bayesian method for unidimensional search. Preprint N 663-74. VINITI, Moscow, 1974.
7. L. Telksnys and V. Šaltenis. The application of multiextremal methods to the problem of estimation of parameters of random signals. Proc. 5th Conf. on extremal problems. Gorkij (in print).
8. A. Ališauskas, A. Lipskis, A. Mačiulis, V. Šaltenis. The application of multiextremal optimization to the planning of experiments for the development of high temperature - resistant polymeric compositions. Proc. 9th Seminar on the stochastic search. Kharkov (in print).