

# STOCHASTIC APPROXIMATION FOR MULTIEXTREMAL PROBLEMS IN HILBERT SPACE

MATHEMATICS

1968

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-196801.05633>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

**Abstract**

**Full Text**

UDC 519.251.1

*MATHEMATICS*

E. M. VAISBORD, D. B. YUDIN

**STOCHASTIC APPROXIMATION FOR MULTIEXTREMAL PROBLEMS IN HILBERT SPACE**

*(Presented by Academician B. N. Petrov, 30 XI 1967)*

1°. Let us consider, in a separable Hilbert space  $H$ , a twice continuously differentiable functional  $\bar{f}(X)$ , bounded from below, satisfying the conditions

$$\|\nabla\bar{f}(X)\| < C < \infty; \quad \|d^2\bar{f}(X)/dX^2\| < K < \infty; \quad (1)$$

$$\lim_{\|X\| \rightarrow \infty} \|\nabla\bar{f}(X)\| > 0; \quad \lim_{\|X\| \rightarrow \infty} \bar{f}(X) > \inf_{X \in H} \bar{f}(X). \quad (2)$$

Let the functional  $\bar{f}(X)$ , generally speaking multi-extremal, attain its lower bound at some point  $X^* \in H$ . (Here we shall assume that  $\inf \bar{f}(X)$  is attained at a unique point. However, after obvious changes in the formulation of the main result of the present paper, its validity will be preserved for any finite number of such points.) Suppose also that for any  $\varepsilon > 0$

$$\inf_{H-S_\varepsilon(X^*)} \bar{f}(X) > \bar{f}(X^*), \quad (3)$$

where  $S_\varepsilon(X^*)$  is a neighborhood of the point  $X^*$ , and

$$\inf_{H-S_\varepsilon(G)} \|\nabla\bar{f}(X)\| > 0, \quad (4)$$

where  $G$  is the set of stationary points of the functional  $\bar{f}(X)$  (points  $X \in H$  for which  $\nabla\bar{f}(X) = 0$ ), and  $S_\varepsilon(G)$  is the  $\varepsilon$ -neighborhood of the set  $G$ .

In the finite-dimensional case, conditions (3) and (4) follow from condition (2) and the continuity of  $\bar{f}(X)$  and  $\nabla\bar{f}(X)$ . In the infinite-dimensional  $H$ , the sphere is not compact, and conditions (3), (4) do not follow from (2).

In the present note a process is proposed for constructing a sequence of elements  $X_n \in H$ ,  $n = 1, 2, \dots$ , which, for large  $n$ , falls with probability arbitrarily close to one into a prescribed neighborhood of the global minimum  $X^*$  of the functional  $\bar{f}(X)$ . It is assumed here that the functional  $\bar{f}(X)$  is determined with random error, i.e., for a value of the argument  $X$  we observe not  $\bar{f}(X)$ , but  $f(X) = \bar{f}(X) + y$ , where  $y$  is a random measurement error with zero mathematical expectation. The random variables  $y(X)$  are independent, identically distributed, with a distribution function possessing a continuous bounded density  $p_1(y)$  ( $-\infty < y < \infty$ ).

The proposed process for constructing the sequence  $X_n$  is a combination of one of the modifications of the stochastic approximation method with a jump-type random process and is an analogue of the procedure set forth in <sup>(1)</sup> for the finite-dimensional case.

2°. Consider the random process  $\xi_n = (X_n; \eta_n)$  with discrete time  $n = 1, 2, \dots$ . Here  $X_n \in H$ , and  $\eta_n$  is an integer from the interval  $[1, N]$ ,  $N$  being an integer. The transition from the state  $(X_n; \eta_n)$  at time  $n$  to the state  $(X_{n+1}; \eta_{n+1})$  at time  $n + 1$  occurs in accordance with the following rules. If  $\eta_n \neq 1$ ,

$$\eta_{n+1} = \begin{cases} \eta_n - 1 & \text{with probability } p(f(X_n)), \\ \min(\eta_n + 1, N) & \text{with probability } q(f(X_n)) = 1 - p(f(X_n)). \end{cases} \quad (5)$$

Here  $p(z)$  ( $-\infty < z < \infty$ ) is a monotonically increasing function satisfying the conditions

$$0 < p(z) < \delta_1 < 1/2.$$

If, however,  $\eta_n = 1$ , then  $\eta_{n+1} = i$ , where  $i$  is a random variable taking the values  $(2, 3, \dots, N)$  with an arbitrary distribution law.

$$X_{n+1} = \begin{cases} X_n - h_n \xi_n \operatorname{sign} [f(X_n + c_n \xi_n) - f(X_n)] & \text{if } \eta_n \neq 1, \\ Y \in H & \text{if } \eta_n = 1. \end{cases} \quad (6)$$

Here the following notation has been adopted:

$$h_n = h(n - \varphi(n))^{-(1/2+\delta)}, \quad c_n = c(n - \varphi(n))^{-r}, \quad (7)$$

$\varphi(n)$  is the largest integer satisfying the conditions

$$\varphi(n) < n, \quad \eta_{\varphi(n)} = 1; \quad (8)$$

$h$  and  $c$  are constants, and the numbers  $\delta$  and  $r$  satisfy the inequalities

$$\delta > 0, \quad r > 0, \quad \delta > r, \quad \delta + r < 1/2, \quad \delta + 2r > 1/2; \quad (9)$$

$\xi_n$  is a random element of the unit sphere of the Hilbert space  $H$ , whose distribution function  $Q(\xi_n)$  is such that the probability

$$P \left\{ \left| \left( \xi_n, \frac{\nabla \bar{f}(X_n)}{\|\nabla \bar{f}(X_n)\|} \right) \right| > \alpha \right\} > \beta, \quad (10)$$

where  $\alpha$  and  $\beta$  are fixed positive constants.

Note that in the finite-dimensional case one can guarantee fulfillment of condition (10) by choosing the distribution function  $Q(\xi_n)$  independent of the value of the argument  $X_n$ . Condition (10) is satisfied, for example, if  $Q(\xi_n)$  is the uniform distribution on the unit sphere. In the infinite-dimensional case, the distribution function  $Q(\xi_n)$  satisfying relation (10) must, generally speaking, depend on  $X_n$ .

$Y$  is a random point of the Hilbert space  $H$ , whose position is determined by some probability measure  $\mu$  on  $H$ . The only condition that the measure  $\mu$  must satisfy is that it be positive on every open set of the space  $H$ .

The process of searching for the global extremum  $\bar{f}(X)$  in  $H$ , defined by formulas (5)–(6), thus represents a combination of a certain variant of stochastic approximation when  $\eta_n \neq 1$  and random jumps when  $\eta_n = 1$ .

The main result of the paper can be formulated as the following theorem.

**Theorem.** For any  $\varepsilon > 0$  there exists a natural number  $N_0(\varepsilon)$ , and for all  $N > N_0$  one can specify a number  $n_0(N)$  such that, for  $n > n_0$ , the probability

$$P(\|X_n - X^*\| < \varepsilon) > 1 - \varepsilon.$$

3°. We give a brief outline of the proof of the stated assertion. From condition (1) it follows that the functional  $\bar{f}(X)$  satisfies the condition

$$\bar{f}(X + Y) \leq \bar{f}(X) + (\nabla \bar{f}(X), Y) + \frac{1}{2} K \|Y\|^2, \quad (11)$$

and from relations (6), (10), (11) the validity of the inequalities follows

$$M_{x_n}(\bar{f}(X_{n+1})) \leq \bar{f}(X_n) + a_n(M_{x_n}(Y_n), \nabla \bar{f}(X_n)) + a_n^2 C M_{x_n}(\|Y_n\|^2), \quad (12)$$

$$(M_{x_n}(Y_n), \nabla \bar{f}(X_n)) \leq -B_n^2 + b_n(k_2 + B_n), \quad (13)$$

$$\|M_{x_n}(Y_n)\|^2 \leq d_n + g_n B n^2, \quad (14)$$

$$M_{x_n}(\|Y_n - M_{x_n}(Y_n)\|^2) \leq d_n + g_n B n^2, \quad (15)$$

where  $M_{x_n}(z)$  is the conditional mathematical expectation of the random variable  $z$  under the condition that the set  $x_n$  of vectors  $X_k$ ,  $k \leq n$ , has taken the fixed value  $(X_1, \dots, X_n)$ ;

$$Y_n = \frac{1}{c_n} \xi_n, \quad d_n = h_n c_n, \quad C = \text{const} > 0, \quad k_2 = \text{const} > 0,$$

$$B_n^2 = 2\alpha^2 \beta \delta_2 \|\nabla \bar{f}(X_n)\|^2, \quad \delta_2 = \text{const} > 0,$$

$$b_n = k_3 c_n / \sqrt{2\alpha^2 \beta \delta_2}, \quad d_n = 4/c_n^2, \quad g_n = 0, \quad k_3 = \text{const}.$$

From conditions (7), (9) it follows that, for  $\varphi(n) = \text{const}$ ,

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n b_n < \infty, \quad \sum_{n=1}^{\infty} a_n^2 d_n < \infty, \quad \lim_{n \rightarrow \infty} b_n = \lim_{n \rightarrow \infty} a_n g_n = 0. \quad (16)$$

From relations (11)–(16) one can obtain that, in the absence of a jump (when  $\eta_n \neq 1$ ), for the process of variation of  $X_n$  defined by formula (6), all the conditions of theorem (5.2) of Fabian's paper [2] are fulfilled, except for the condition of finite dimensionality of the space  $H$ . An analysis of the proof of Fabian's theorem shows, however, that for the validity of its assertion the finite-dimensionality condition on the space  $H$  can be replaced by conditions (3), (4). From Fabian's theorem (5.2) and the first of inequalities (2) it follows that, in the absence of a jump (when  $\eta_n \neq 1$ ), the sequence  $X_n$ , defined by the recurrent relation (6), converges with probability equal to one to one of the level sets of the functional  $\bar{f}(X)$  passing through a stationary point of this functional. (The level set of  $\bar{f}(X)$  passing through  $X_0$  is the set of points  $X$  for which  $\bar{f}(X) = \bar{f}(X_0)$ .)

The random process  $(\zeta_n, n - \varphi(n)) = (X_n, \eta_n, n - \varphi(n))$  is a stationary Markov process with discrete time. It can be shown that this process satisfies the conditions ensuring the existence of a limiting probability for the residence of the point  $X_n$  in any Borel set of the space  $H$  and the validity of the ergodic theorem. From the ergodic theorem it follows that the ratio of the indicated limiting probabilities for two mutually complementary sets is equal to the ratio of the mathematical expectations of the times of continuous residence of the point  $X_n$  in these sets.

It can be shown that, when  $\eta_n \neq 1$  (in the absence of jumps), the sequence  $X_n$  converges with some positive probability to the point  $X^*$  of the global minimum of  $\bar{f}(X)$ . According to condition (3), the value of  $\bar{f}(X)$  in a sufficiently small neighborhood of  $X^*$  is less than the values of the functional in an  $\varepsilon$ -neighborhood of the level set passing through any other stationary point of  $\bar{f}(X)$ . Hence it follows that the mean value  $\bar{p}(X)$  of the probability  $p(\bar{f}(X_n))$  with which the point  $\eta_n$  passes to  $\eta_{n+1} = \eta_n - 1$  will, for points  $X \in S_\varepsilon(X^*)$ , be smaller than at points in a neighborhood of the level set passing through any other stationary point of  $\bar{f}(X)$ . The mathematical expectation of the time  $M(T)$  until the first jump of the point  $X_n$  belonging to some

set  $L$ , depends on the values of  $\bar{p}(X)$  on this set. The smaller the values assumed by  $\bar{p}(X)$  at the points of  $L$ , the larger the value  $M(T)$ . It can be shown that if the values of  $\bar{p}(X)$  at the points of the set  $L_1$  are strictly smaller than the values of this function at all points of the set  $L_2$ , then

$$\lim_{N \rightarrow \infty} \frac{M_{L_1}(T)}{M_{L_2}(T)} = \infty,$$

where  $M_{L_1}(T)$  and  $M_{L_2}(T)$  are the mathematical expectations of the times of continuous sojourn of the points  $X_n$  in the sets  $L_1$  and  $L_2$ , respectively. The development of the considerations presented constitutes the proof of the main assertion of the present note.

4°. The assertion of the theorem remains valid for substantially more general procedures of variation of  $\eta_n$ . In particular, the theorem is valid for an arbitrary process  $\eta_n$  satisfying the following property.

For  $p_1 < p_2$ ,

$$\lim_{N \rightarrow \infty} \frac{M(\tau \mid p(f(X_n)) < p_1)}{M(\tau \mid p(f(X_n)) > p_2)} = \infty,$$

where  $M(\tau \mid a > b)$  is the mathematical expectation of the time to reach the state  $\eta_n = 1$  under the condition  $a > b$ .

The theorem remains valid if one dispenses with the existence of a continuous bounded density of the distribution of the random variable  $y = f(X) - \bar{f}(X)$ . It is sufficient only to require that, for independent random variables  $y_1$  and  $y_2$  distributed identically with  $y$ , the condition

$$P(y_1 - y_2 < \sigma) > 1/2 + \alpha\sigma,$$

hold, where  $\sigma > 0$  is a sufficiently small quantity,  $\alpha = \text{const}$ .

The arbitrariness present in the organization of the process (5)–(6) can be used to accelerate the search. Acceleration can be achieved both by taking into

account a priori information about the function  $f(X)$ , and by using information accumulated in the course of the search.

Received 28 XI 1967

### CITED LITERATURE

<sup>1</sup> E. M. Vaisbord, D. B. Yudin, *Izv. AN SSSR, Tekhnicheskaya kibernetika*, No. 5 (1968). <sup>2</sup> V. Fabian, *Czechoslovak Mathematical Journal*, **10** (85), 123 (1960).

*Note: Figure translations are in progress. See original paper for figures.*

*Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.*