Some remarks on the paper of K. Varga and P. Fejes

By P. HUNYA

In practical applications of the iterative unconstrained optimization methods (for example in the gradient method discussed in [1]) a difficult problem is to find the initial estimation of the solution. Generally m points are randomly chosen from the region containing the optimum place and that which represents the minimum (maximum) value of function is considered as the initial estimation. Following this strategy the sequence generated by the iterative process converges with a probability generally less then 1. By the modified method proposed in [1] one step of the iteration is performed for all the m points before the selection and this may improve the probability of the convergency. However the verification of this property is heuristic, it is based on a number of experimental calculations with various type of functions. The upper and lower limits of the convergency are also given in [1].

In this paper a generalization of the modified method is described and the probability of the convergency is discussed in detail. (The problem of minimization is exemined, since all considerations are analogous in the case of maximization.)

Let us consider the continuous function of real values $F(x)$ defined on the complete metric space S. Let T be a subset of S with nonzero measure and let us suppose that F takes its minimum on T. Let $y=Mx$ be a mapping of T into T with the property

$$F(Mx) \leqq F(x). \tag{1}$$

There exists obviously a $T_{\text{conv}} \neq \emptyset$ subset of T for the elements x of which the sequence $x=M^0x, M^1x, M^2x, \ldots, M^nx, \ldots$ is convergent and

$$\lim_{n\to\infty} M^n x = \bar{x}, \tag{2}$$

where

$$F(\bar{x}) = \min_{x\in T} F(x);$$

that is, the iteration process generated by M converges to the solution of the optimization problem. To simlify the considerations we suppose the uniqueness of $\bar{x}$, however, this fact is not essential in the following proofs.

Suppose further that

$$\varrho(T_{\text{conv}}) \neq 0 \tag{3}$$

($\varrho(t)$ means the measure of the set $t \subset S$) and

$$\varrho(T_{\min}) \neq 0 \tag{4}$$

where $T_{\min}$ is a subset of T_{conv} defined by

$$T_{\min} = (x \mid F(x) < \inf_{y \in T - T_{\text{conv}}} F(y)). \tag{5}$$

Now let us consider the following procedure:

a) let the points $x_1, x_2, \ldots, x_m$ be chosen from T independently, with homogeneous distribution on T;

b) then form the sequence

$$M^n x_1, M^n x_2, \ldots, M^n x_m$$

for $n \geqq 0$ and

c) let a point $\bar{x}^n = M^n x_{k*}$ be chosen for which

$$F(M^n x_{k*}) = \min_{1 \leqq i \leqq m} F(M^n x_i).$$

We shall prove that by the previous conditions for $n \to \infty$ the probability $P = P(\bar{x}^n \in T_{\text{conv}})$ converges to the limit $\bar{P}$ while

$$P_n \leqq P \leqq \bar{P} \tag{6}$$

where P_n is a monotonic increasing sequence i.e.

$$P_n \leqq P_{n+1} \tag{7}$$

and

$$\lim_{n \to \infty} P_n = \bar{P}. \tag{7a}$$

In other words: by increasing the number of the iterative steps before selection the probability of the convergency approximates its upper limit with an arbitrary degree of accuracy. (The dependence on m is not considered in this paper.)

For the proof we define the following events:

1) A_n denotes the event that among the points $M^n x_1, \ldots, M^n x_m$ there exists element of T_{conv}

$$A_n = \{\exists i (M^n x_i \in T_{\text{conv}})\}. \qquad (n = 0, 1, 2, \ldots) \tag{8}$$

As a consequence of the definition of T_{conv} we have

$$A_n = A_k (= A), \tag{9}$$

since it is obvious that

$$M^k x \in T_{\text{conv}} \Leftrightarrow M^{k+1} x \in T_{\text{conv}}$$

for $k = 0, 1, 2, \ldots$.

2) B_n denotes that the selected point $\bar{x}^n$ is element of T_{conv}, that is, the iterational process converges:

$$B_n = \{\bar{x}^n \in T_{\text{conv}}\}. \tag{10}$$

3) C_n denotes the event that $\bar{x}^n$ is element of $T_{\min}$

$$C_n = \{\bar{x}^n \in T_{\min}\}. \tag{11}$$

4) And finally D_n denotes the event that at least one of the points $M^n x_1, M^n x_2, \ldots$ $\ldots, M^n x_m$ belongs to $T_{\min}$

$$D_n = \{\exists i(M^n x_i \in T_{\min})\}. \tag{12}$$

From the definitions 1)—4) immediately follows that for all n.

$$D_n \Leftrightarrow C_n \Rightarrow B_n \Rightarrow A_n \Leftrightarrow A, \tag{13}$$

consequently, the probabilities of A_n, B_n, C_n, D_n, A satisfy the relations

$$P(D_n) = P(C_n) \leqq P(B_n) \leqq P(A_n) = P(A). \tag{14}$$

For $n=0$ (14) contains as a special case one of the results of [1] for the values

$$P_0 = P(C_0) = 1-(1-\varrho_{\min})^m$$
$$\bar{P} = P(A_0) = 1-(1-\varrho_{\text{conv}})^m.$$

Using the notations $P_n = P(C_n) = P(D_n)$ and $\bar{P} = P(A) = P(A_n)$ (14) proves (6) also.

As a consequence of the condition (1) and the definition 4) we have

$$D_n \Rightarrow D_{n+1} \tag{15}$$

and this implies the inequality (7).

Let us consider now the sequence $D_0, D_1, \ldots, D_n, \ldots$. We shall prove that

$$\sum_{i=1}^{\infty} D_i = A. \tag{16}$$

From the continuity of the function $F(x)$ follows that for the elements $\underline{x}$ of T_{conv} we can find a natural number $n(x)$, such that $n > n(x)$ implies $M^n x \in T_{\min}$ Let T^i denote the subset of T_{conv} for the elements of which $n(x) = i$ ($i = 0, 1, 2, \ldots$) and define the event A^i as follows

$$A^i = \{\exists j(x_j \in T^i\}.$$

By the definition above $A_i \Rightarrow D_i$. On the other hand if there exists x_j ($1 \leqq j \leqq m$) such that $M^i x_j \in T_{\min}$ is true, then $x_j \in T_i$ also holds; that is, D_i implies A^i. Thus we have $A^i = D_i$. It is obvious that $A = \sum_{i=0}^{\infty} A^i$ so we get (16).

Because of (15) and (16) one of the basic limit theorems of the probability theory ([2], § 2.2) can be applied to the sequence $D_0, D_1, \ldots, D_n, \ldots$, therefore

$$\lim_{n\to\infty} P(D_n) = P(A)$$

so using the notations introduced previously we get (7a)

$$\lim_{n\to\infty} \underline{P}_n = \bar{P}.$$

Несколько замечаний к работе К. Варга и Р. Фейеш

В итерационных методах безусловной оптимизации при случайно выбранном начальном значении, получается последовательность сходящаяся к решению задачи, только с вероятностью P (обычно меньшей единицы). Вероятность сходимости может увеличиться, если из m случайно выбранных точек, считаем начальной ту точку, в которой достигается минимум (максимум). Дальнейшее улучшение получается при выпольнении n шагов итерационной процедуры перед выбором исходной точки.

Доказывается, что при довольно общих условиях, нижний предел вероятности P является монотонной функцией от n, и с ростом n, P сходится к своему верхнему пределу, обеспечивая этим увеличение вероятности конвергенции.

LABORATORY OF CYBERNETICS
JÓZSEF ATTILA UNIVERSITY
SZEGED, HUNGARY

References

[1] VARGA, K., P. FEJES, Combined application of drawing and steepest descent in generating initial estimates for subsequent optimization, *Acta Cybernet.*, v. 2, 1973, pp. 53—59.
[2] PRÉKOPA, A., *Valószínűségelmélet*, Budapest, 1962.

(Received June 28, 1972)