LIMIT EXTREMUM PROBLEMS

Yu.M. Ermol'ev, E.A. Nurminskiy
Institute of Cybernetics
Ukrainian Academy of Sciences
252627 Kiev 127 USSR

INTRODUCTION

In applications one has often to tackle extremum problems where the objective function happens to be not strictly fixed as in the general theory of nonlinear programming but changes versus some parameter (time, in particular). That is, instead of $F(x)$ there is a sequence of functions $F^N(x)$ in a certain sense approximating $F(x)$, on the basis of which the extremum of $F(x)$ is to be found. As a rule, one fails to execute the passage to the limit, to find $F(x)$ and then its extremum due to a number of circumstances of which the following might be emphasized:

1. The parameter N corresponds to the discrete time and $F^N(x)$ becomes known at the instant $t=N$ only. In this case the limit passage takes a whole time given for the problem solution.

2. The parameter N is an index of members of the sequence. It may be changed at one's discretion, "frozen", in particular, at some stages of the optimization process, however, the execution of limit passage is technically difficult. Such cases are particularly characteristic of problems of optimizing steady regime of controlled processes when averaged performance figures of the form

$$F(x) = \lim_{N\to\infty} F^N(x) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} g(k,x),$$

$$F(x) = \lim_{N\to\infty} F^N(x) = \lim_{N\to\infty} \sum_{k=1}^{N} \alpha^k g(k,x).$$

are dealt with.

3. The limit passage operation deteriorates some good properties of the function $F^N(x)$ which is characteristic of approximation problems when a function, being due to some reasons "bad", is approximated by a sequence of "good" ones and thus, instead of $F(x)$, it is of advantage to manipulate with the functions $F^N(x)$. An interesting and difficult problem arises here of optimizing the limit function $F(x)$ with the only use of information about members of the sequence $F^N(x)$ approximating $F(x)$. It is important to have in mind that if $F(x)$ is unknown then the examination of only one of the functions $F^N(x)$ in solving the problem approximately does not allow estimation of the accuracy of the obtained approximate solution. Such extremum searching procedures are treated here in which the search of the function $F^N(x)$ extremum is based on the analysis of a sequence of functions $F(x)$. The paper grounds on results of [3].

ALGORITHMS

Consider the extremum problem

$$\min F(x),\ x \in X, \tag{1}$$

where X is a convex closed bounded set, $F(x)$ - a convex but not necessarily continuously differentiable function determined as the limit of a functional sequence:

$$F(x) = \lim_{N \to \infty} F^N(x).$$

The following quite natural algorithm may be offered for the problem solution

$$x^{s+1} = \Pi_X(x^s - \rho_s \hat{F}_x^s(x^s)),\ s = 0, 1, \ldots, \tag{2}$$

where $\rho_s \geq 0$ are step-by-step factors, $\hat{F}_x^s(x^s)$ - a generalized gradient of the convex function $F(x)$, $\Pi_X(\cdot)$ - an operator of projection on the set X. The following theorem holds.

THEOREM 1. Let $F^s(x)$ be convex functions for each s, the sequence $F^s(x)$ uniformly converges on X, and $\rho_s \to 0, \sum \rho_s = \infty$. Then for each convergent subsequence $\{x^{s_k}\}$

$$\lim x^{s_k} = x^* \in X^*,$$
$$\lim F^s(x^s) = F^* = \min \{F(x), x \in X\},$$

where X^* is a set of solutions of problem (1). The requirement for the sequence $F^N(x)$ to converge uniformly is the most essential of the theorem assumptions. However, when functions $F^N(x)$ are convex the uniform convergence follows readily from the point one if some additional fairly weak assumptions are introduced. The rest of conditions do not differ from those of convergence of the known method of generalized gradients, formulated in [1], however, the study of convergence by method (2) with the direct application of the scheme in [1] under assumptions of the theorem is impossible. Theorem 1 is proved in [3] by an approach elaborated in [2]. Of great interest is also a convergence of the stochastic analogue of algorithm (2):

$$x^{s+1}(\omega) = \pi_X(x^s(\omega) - \rho_s \xi^s(x^s, \omega)), \tag{3}$$

where ξ^s is a random vector (a stochastic quasi-gradient) whose conditional expectation

$$E(\xi^s / x^0, \ldots, x^s) = \hat{F}^s_x(x^s). \tag{4}$$

THEOREM 2. Let assumptions of Theorem 1 be satisfied and $\sum \rho_s^2 < \infty$. Then algorithm (3) converges in the sense that for almost all ω the limit points of sequence $\{x^s(\omega)\}$ belong to the set X^* and with probability 1 $\lim F^s(x^s) = F^*$.

APPLICATIONS

On the method of penalty functions. With the use of the method of penalty functions the problem of minimization of $f^0(x)$ in the domain $\mathcal{D}$ determined by the constraints

$$f^i(x) \le 0, \; i = 1, \ldots, m, \quad x \in X$$

is approximated by the minimization of a function $F(x, c)$ in the domain X so that

$$\min_{x \in \mathcal{D}} f^0(x) = \lim_{c \to c^*} \min_{x \in X} F(x, c)$$

for some C^*, the C^* being often equal to 0, ∞, or must be a sufficiently large number. The C being fixed, the minimization of $F(x,c)$ in the domain X does not yield, generally speaking, the precise solution of the initial problem. However, if such arbitrary sequence $C^N, N=0,1,\ldots$ is chosen that $C^N \to C^*$ and if the limit extremum problem with the function $F^N(x) = F(x, c^N)$ is studied then under appropriate conditions the precise solution will be obtained by method (2). A question is interesting about choosing ways of the sequence that impacts the speed of convergence.

INTERCONNECTED EXTREMUM PROBLEMS

Sometimes there is a set of interconnected extremum problems in which solution of one problem prepares information for solving the others. In problems of vector optimization, for instance, the minimization problem

$$F(x) = \max_{\ell} \{ f^{\ell}(x) - \min_{x \in X} f^{\ell}(x) \}$$

is dealt with when choosing a compromise solution. In this case the problems of minimizing $f^{\ell}(x)$ prepare information for the basic problem of minimization of the function $F(x)$. Since the solution of each auxiliary problem necessitates an infinite number of iterations, the direct way to calculate $F(x)$ even in a separate point, to say nothing of its extremum search, is a nonconstructive one. In addition, if sequences of points $x^{\ell}(N)$ such that

$$f^{\ell}(x^{\ell}(N)) \to \min_{x \in X} f^{\ell}(x), \; N \to \infty,$$

are considered together with the limit extremum problem with the function

$$\overset{N}{F}(x) = \max_{\ell} \{ f^{\ell}(x)^{e} - f^{\ell}(x^{e}(N)) \},$$

then, if conditions of Theorem 1 are satisfied, procedure (2) helps us to find the precise minimum of $F(x)$.

OPTIMIZATION OF STEADY REGIMES

The results of Theorem 2 offer quite an effective way of solving function minimization problems of the form

$$F(x) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} g(k, x),$$

It is easily seen that with proper assumptions about the differentiability of functions $g(k,x)$ the stochastic gradient $\xi^{s}(x,\omega)$, satisfying relations (4), can be determined, for instance, as follows:

$$\xi^{s}(x, \omega) = g_x(\omega^{N}, x)$$

where ω^{N} is a random variable uniformly distributed on the set $\{1, 2, \ldots \omega\}$. Such a construction allows to calculate at each iteration a derivative of only one member of the increasing sum of terms $\sum_{k}^{N} g(k,x)$. A similar method is also applicable for minimizing

$$F(x) = \lim_{N\to\infty} \sum_{k=1}^{N} \alpha^{k} g(k,x)$$

ON THE RANDOM SEARCH METHOD

When the calculation of gradient of the objective function is complex the following method might prove helpful which also necessitates the solution of limit extremum problems. Instead of the objective function $F(x)$ we consider the function

$$F(x,\alpha) = EF(x - \eta(\alpha)) = \int_{E_n} F(x-y)P(y,\alpha)dy,$$

where the distribution $P(y,\alpha)$ of the random variable $\eta(\alpha)$ for $\alpha\to 0$ concentrates in 0, i.e. $F(x,\alpha)\to F(x)$, $\alpha\to 0$. Then subject to existence of corresponding integrals and $F(x-y)P(y,\alpha)\to 0$, $y\to 0$,

$$F_x(x,\alpha) = \int_{E_n} F_x(x-y)P(y,\alpha)\,dy = -\int_{E_n} F(x-y)\frac{P_y(y,\alpha)}{P(y,\alpha)}P(y,\alpha)dy.$$

Thus, the random variable

$$-F(x - \eta(\alpha))\frac{P_y(\eta(\alpha),\alpha)}{P(\eta(\alpha),\alpha)}$$

for the fixed x coincides on average with the gradient $F(x,\alpha)$. The examination of the sequence $\alpha^{s}\to 0$ and the limit extremum problem with functions $F^{s}(x) = F(x,\alpha^{s})$ completed, we obtain the possibility to organize, by procedure (3) and for

$$\xi^{s}(x,\omega) = -F(x - \eta(\alpha^{s}))\frac{P_y(\eta(\alpha^{s})\alpha^{s})}{P(\eta(\alpha^{s}),\alpha^{s})},$$

the iterative process where derivatives of $F(x)$ are not employed.

REFERENCES

1. Yu.M. Ermol'ev, Methods for Solving Nonlinear Extremum Problems, Kibernetika, No. 4, 1966.
2. E.A. Nurminskiy, Convergence Conditions of Nonlinear Programming Algorithms, Kibernetika, No.6, 1972.
3. Yu.M. Ermol'ev, E.A. Nurminskiy, Limit Extremum Problems, Kibernetika, No. 4, 1973.