MINIMIZATION OF NONDIFFERENTIABLE FUNCTIONS IN THE PRESENCE OF NOISE

E. A. Nurminskii

UDC 62.50:007

The development of methods of minimization of nondifferentiable functions is of great interest for the solution of many problems such as minimax problems, two-stage mathematical-programming problems, etc. From the point of view of practical application it is of considerable interest to ascertain the performance of these methods in the presence of noise. The principal result of this paper consists in the fact that the specific character of nondifferentiable functions makes it possible in certain cases to get rid of the conventional conditions of stochastic approximation used for the suppression of random noise.

We shall consider nonconvex nondifferentiable functions whose properties are defined as follows [1, 2].

Definition. A continuous function $f(x)$ is said to be weakly convex if for any x there exists a set M(x) of vectors g such that for any y,

$$f(y) - f(x) \geqslant (g, y - x) + r(x, y) \tag{1}$$

where for $y \to x$ we have $r(x, y)|x-y|^{-1} \to 0$ uniformly in x in any closed bounded subset of E_n.

In [1], the vector g is called the quasigradient of a weakly convex function.

For solving the problem

$$\min_{x \in E_n} f(x) \tag{2}$$

we have proposed in [2] the quasigradient method

$$x^{s+1} = x^s - \rho_s g(x^s), \quad s = 0, 1, \ldots$$

and we proved its convergence under appropriate assumptions. In the present paper we shall ascertain the conditions of convergence of the algorithm

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

where

$$\xi^s(x^s, \omega) = g(x^s(\omega)) + \eta^s(\omega),$$

$\eta^s(\omega)$ being independent uniformly distributed random disturances, $E|\eta^s|^2 < \infty$. In the following we shall sharpen the requirements towards $\eta^s(\omega)$. We shall assume that problem (2) has the following property: If we denote by X* the set $\{x^*: 0 \in M(x^*)\}$, then there exists a positive δ such that for any $x \overline{\in} X^*$ we have $g \geqslant \delta$, where $g \in M(x)$.

This property is not very rigorous, and it holds for practically all linear minimax problems, i.e., for problems of form (2) in which $f(x) = \max_{i \in I}(c_i, x)$.

We have the following theorem.

THEOREM. Let $E|\eta^s| \leqslant \gamma$, where γ is sufficiently small,

$$\Sigma\rho_s = \infty, \ \rho_s/\rho_{s+1} \to 1, \ \rho_s \to +0,$$

$f(x)$ assumes on X* not more than a countable number of values, and the set $\{f(x) \leqslant f(x^0) + C\}$ is bounded for a positive C. Then the algorithm

$$x^{s+1}(\omega) = \begin{cases} x^s(\omega) - \rho_s \xi^s(x^s, \omega), & f(x^s) \leqslant f(x^0) + C, \\ x^0 \end{cases}$$

Translated from Kibernetika, No. 4, pp. 59-61, July-August, 1974. Original article submitted January 29, 1974.

will be convergent with probability 1, i.e., any limit point of the sequence $\{x^s(\omega)\}$ will belong to the set X^* with probability 1.

The conditions of the theorem contain an uncertainty related to the absence of an estimate for the quantity γ. This estimate will be obtained during the proof.

Suppose that the assertion of the theorem is not true and that for a set B, $P(B) > 0$, there exists for an $\omega \in B$ a subsequence $\{x^{n_k}(\omega)\}$ that is convergent to $x'(\omega) \in X^*$. Let us specify an $\omega \in B$ and omit in the following the dependence on ω.

Let us take a positive ϵ_0 such that for all vectors g we have

$$g \in \mathrm{co}\{\bigcup_{x \in U_{4\varepsilon_0}} M(x)\}, \quad |g| \geqslant \frac{1}{2}\delta, \tag{3}$$

where $U_{4\epsilon_0} = \{x: |x - x'| \leqslant 4\epsilon_0\}$.

For the time being we shall assume that $f(x') < f(x^0) + C$. Therefore, ϵ_0 can be taken in such a way that $f(x) \leqslant f(x^0) + C$ for $x \in U_{4\epsilon_0}(x')$. Later on we shall drop this assumption. By virtue of the semi-continuity from above of the set mapping M(x) [1], it is always possible to take ϵ_0 in such a way that (3) is satisfied.

Now let us assume that for a k' we have $|x^s - x^{n_{k'}}| \leqslant \epsilon_0$ for any $s \geqslant n_{k'}$. By setting $s = n_m$ and going over to the limit for $m \to \infty$, we obtain $|x' - x^{n_{k'}}| \leqslant \varepsilon_0$ or $|x' - x^s| \leqslant 2\varepsilon_0$ for $s \geqslant n_{k'}$. Let us consider the following quantities:

$$d_s = (g^{s+1}, x^{s+1} - x^{n_k}) = \left(g^{s+1}, \sum_{m=n_k}^{s} \rho_m \xi^m(x^m, \omega)\right) = \left(g^{s+1}, \sum_{m=n_k}^{s} \rho_m g^m\right) + \left(g^{s+1}, \sum_{m=n_k}^{s} \rho_m \eta^m(\omega)\right).$$

In [2] we have shown that for sufficiently large $s > n_k \geqslant n_{k'}$

$$\left(g^{s+1}, \sum_{m=n_k}^{s} \rho_m g^m\right) \geqslant \frac{\delta^2}{8} \sum_{m=n_k}^{s} \rho_m$$

or

$$d_s \geqslant \frac{\delta^2}{8} \sum_{m=n_k}^{s} \rho_m - \Delta \sum_{m=n_k}^{s} \rho_m |\eta^m(\omega)|.$$

It is possible to show that with probability 1,

$$\lim_{s \to \infty} \frac{\sum_{m=n_k}^{s} \rho_m |\eta^m(\omega)|}{\sum_{m=n_k}^{s} \rho_m} = E|\eta|.$$

Hence, for $E|\eta| \leqslant \gamma = \delta^2/32\Delta$ we have in the case of sufficiently large s,

$$d_s \geqslant \frac{\delta^2}{16} \sum_{m=n_k}^{s} \rho_m.$$

On the other hand,

$$f(x^{s+1}) - f(x^{n_k}) \leqslant -d_s + r(x^{s+1}, x^{n_k}).$$

Since by assumption we have $x^{s+1} \in U_{\varepsilon_0}(x^{n_k})$, it can be assumed that

$$|r(x^{s+1}, x^{n_k})| \leqslant \varepsilon' |x^{s+1} - x^{n_k}| \leqslant \varepsilon'' \sum_{m=n_k}^{s} \rho_m,$$

where ϵ'' is as small as desired. In particular, it can be assumed that $\epsilon'' < \delta^2/32$. Finally, we obtain

$$f(x^{s+1}) - f(x^{n_k}) \leqslant -\frac{\delta^2}{32} \sum_{m=n_k}^{s} \rho_m. \tag{4}$$

By going over in (4) to the limit for $s \to \infty$, we obtain a contradiction to the boundedness of the continuous function $f(x)$ on the closed bounded set $U_{4\epsilon_0}(x')$. Hence, the assumption made at the beginning of the proof is not true. Otherwise, we can say that

$$\min_{r>n_k} r : |x^r - x^{n_k}| > \varepsilon_0 = m_k < \infty. \qquad (5)$$

Since for sufficiently large k we have $x^{m_k} \in U_{4\varepsilon_0}(x')$, it follows that (4) holds also for $s = m_k - 1$. By denoting $-f(x) = W(x)$, we obtain

$$W(x^{m_k}) \geqslant W(x^{n_k}) + \frac{\delta^2}{32} \sum_{s=n_k}^{m_k-1} \rho_s.$$

Since $|x^{m_k} - x^{n_k}| > \varepsilon_0$, it follows that $\frac{\varepsilon_0}{c} \leqslant \sum_{s=n_k}^{m_k-1} \rho_s$, whence

$$W(x^{m_k}) - W(x^{n_k}) \geqslant \frac{\varepsilon_0 \delta^2}{32c}.$$

By going over in this inequality to the limit for $k \to \infty$, we obtain

$$\varliminf_{k\to\infty} W(x^{m_k}) > \lim_{k\to\infty} W(x^{n_k}). \qquad (6)$$

By virtue of formulas (5)–(6) proved above, the convergence of the algorithm follows from the general results of [3]. The subsequent analysis does not depend on the structure of the algorithm and can be carried out in the same way as in [2].

Let us make several remarks about the obtained result. First of all, let us note that in contrast to the conventional conditions of stochastic approximation (see, for example, [4]), it is not required that the step multipliers ρ_s satisfy the condition $\Sigma\rho_s^2 < \infty$. Moreover, the noise $\eta^s(\omega)$ can be biased, $E\eta^s \neq 0$; in particular, we can have $\eta^s(\omega) = \text{const}$, or, as can be easily seen from the proof, $\eta^s(\omega) \neq \text{const}$, but $|\eta^s| \leqslant \gamma$. This makes it possible to use finite-difference counterparts of the quasigradient method. In particular, if $f(x) = \max f(x, y)$, $y \in Y$, where $f(x, y)$ is differentiable with respect to x, it is possible to solve the $\min f(x)$ problem by the following method:

$$x^{s+1} = \begin{cases} x^s - \rho_s \tilde{g}(x^s), & f(x^s) \leqslant f(x^0) + C, \\ x^0, & \end{cases}$$

where $\tilde{g}(x^s) = [f(x^s + \Delta, y^s) - f(x^s, y^s)]/\Delta$, y^s is a solution of the problem $\max f(x^s, y)$, $y \in Y$, and Δ is a sufficiently small constant.

LITERATURE CITED

1. E. A. Nurminskii, "On the properties of a class of functions," in: Theory of Optimal Solutions, Izd. Inst. Kibernetika Akad. Nauk UkrSSR, Kiev (1972).
2. E. A. Nurminskii, "Quasigradient method of solution of nonlinear programming problem," Kibernetika, No. 1 (1973).
3. E. A. Nurminskii, "Conditions of convergence of nonlinear programming algorithms," Kibernetika, No. 6 (1972).
4. Yu. M. Ermol'ev, "Method of generalized stochastic gradients and of stochastic quasi-Fejer sequences," Kibernetika, No. 2 (1969).