THE QUASIGRADIENT METHOD FOR THE SOLVING OF THE NONLINEAR PROGRAMMING PROBLEMS

E. A. Nurminskii

UDC 512.25/26+219.3

In many applied extremal problems with nondifferentiable functions the requirement of concavity is too burdensome. At the same time, these functions possess a series of special properties which allow us to hope for the construction of some general approach.

In the present paper we introduce the class of weakly concave (convex) functions, which, in particular, includes the differentiable as well as the concave (convex) functions and is closed with respect to the operation of taking the minimum (maximum).

For the weakly concave functions we introduce the concept of the quasigradient, which coincides in the case of differentiable functions with the usual gradient and with the generalized gradient for concave functions. We prove the convergence of the quasigradient method for the solving of the maximization problem.

Definition 1. A continuous function $f : E_n \to E_1$ is said to be weakly concave if for every $x \in E_n$ there exists a nonempty set M(x) of vectors g such that for all $y \in E_n$

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

where for $y \to x$, $\frac{r(x, y)}{|x - y|} \to 0$ uniformly with respect to x in each compact subset of E_n.

From the definition it follows easily that M(x) is convex and closed. Let us prove that the set M(x) is bounded. Indeed, if we assume the opposite, then for some x^0 there exists a sequence $g^n \in M(x^0)$ such that $|g^n| \to \infty$. Without loss of generality we assume that $|g^n| \geq 1$. We consider the corresponding sequence $\{y^n\}$, where $y^n = -\frac{g^n}{|g^n|^{3/2}} + x^0$. We note that $|y^n - x^0| \to 0$ for $n \to \infty$. Then, by the definition of a weakly concave function

$$f(y^n) - f(x^0) \leqslant -|g^n|^{1/2} + r(x^0, y^n) \to -\infty$$

for $n \to \infty$, which contradicts the boundedness of $f(x)$ on the compact set $\{x : |x - x^0| \leq 1\}$.

Definition 2. A vector g satisfying inequality (1) will be called a quasigradient of the weakly concave function $f(x)$.

It is easy to see that in the case of a differentiable function, the quasigradient coincides with the usual gradient while for concave functions, with the generalized gradient.

We mention some properties of weakly concave functions.

1. If $f_i(x)$ are weakly concave for $i = 1, 2, \ldots, m$, then $f(x) = \min_{1 \leqslant i \leqslant m} f_i(x)$ is a weakly concave function.

2. If $f_\alpha(x)$ is weakly concave for each $\alpha \in A$, where A is a compact topological space and, moreover,

---

Translated from Kibernetika, No. 1, pp. 122-125, January-February, 1973. Original article submitted January 24, 1972.

for $y \to x \frac{r_\alpha(x, y)}{|x-y|} \to 0$ uniformly with respect to α, then $f(x) = \min_{\alpha \in A} f_\alpha(x)$ is a weakly concave function.

Let us prove the second property. Let $A(x) = \{\alpha : f_\alpha(x) = f(x)\}$ and $\alpha \in A(x)$. By the definition of a weakly concave function

$$f_\alpha(y) - f_\alpha(x) \leqslant (g_\alpha, y - x) + r_\alpha(x, y) \leqslant (g_\alpha, y - x) + \max_{\alpha \in A} r_\alpha(x, y).$$

But for $\alpha \in A(x)$

$$f_\alpha(x) = f(x), \quad f_\alpha(y) \geqslant f(y).$$

Then we have

$$f(y) - f(x) \leqslant (g_\alpha, y - x) + \max_{\alpha \in A} r_\alpha(x, y) = (g_\alpha, y - x) + r(x, y),$$

where $r(x, y)$ satisfies the necessary condition in the definition of an asymptotically concave function.

We consider the problem

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

where $f(x)$ is a weakly concave function.

At the formulation of the maximization problem of a weakly concave function there arises the question of the necessary conditions for an extremum. Here we will consider the maximum problem in the absence of restrictions; therefore the necessary condition for an extremum is $0 \in M(x)$.

We impose the following conditions on $f(x)$:

$$X^* = \{x : 0 \in M(x)\} \text{ is compact,} \tag{3}$$

$$G(a) = \{x : f(x) \geqslant a\} \text{ is compact for every } a. \tag{4}$$

In order to solve problem (2) we consider the sequence $\{x^n\}$, generated by the following relation:

$$x^{s+1} = \begin{cases} x^s + \rho_s g(x^s) & \text{for } x^s + \rho_s g(x^s) \in S, \\ y \in A & \text{for } x^s + \rho_s g(x^s) \bar{\in} S; \end{cases} \tag{5}$$

x^0 is the initial approximation, S and A ($A \subset S$) are some compact subsets of E_n which will be defined explicitly later, and ρ_s is a numerical sequence such that

$$\rho_s > 0, \ \rho_s \to 0, \ \frac{\rho_{s+1}}{\rho_s} \to 1, \ \sum_{s=0}^{\infty} \rho_s = \infty.$$

Our proof of the convergence of the algorithm (5) will be based on the approach described in [1], where it has been proved that in order that any convergent subsequence of the sequence $\{x^n\}$, generated by the relation (5), should converge to the set of the solutions X^*, it is sufficient that the following conditions should hold:

1) $\lim_{n \to \infty} |x^{n+1} - x^n| = 0,$

where n is not a "jump" moment, i.e., $x^n + \rho_n g(x^n) \in S$.

2) $x^n \in S$.

3) For any subsequence $\{x^{n_k}\}$ such that $\lim_{k \to \infty} x^{n_k} = x \bar{\in} X^*$, for all k and for sufficiently small $\varepsilon > 0$

there exist indices $m_k < \infty$, defined by the relations

$$m_k = \min_{r > n_k} r : |x^r - x^{n_k}| > \varepsilon.$$

4) There exist a continuous function W(x) such that $\underline{\lim}_{k\to\infty} W(x^{m_k}) > \lim_{k\to\infty} W(x^{n_k}) = W(x')$ for arbitrary subsequences $\{x^{n_k}\}$, $\{x^{m_k}\}$, connected by the condition 3.

5) W(x) have a finite number of values on X^*.

6) $\min_{x\in A} W(x) > \max_{x\in\partial S} W(x)$.

For the sequence $\{x^n\}$, generated by the relation (5), we set $A = \{x^0\}$, $S = G(f(x^0) - \delta)$, where $\delta > 0$ is some constant. Obviously, the set A is compact, while the set S is compact by virtue of (4).

As W(x) we will use $f(x)$. Conditions 1, 2, 6 hold by virtue of the assumptions; condition 5 is also assumed to be true. In order to prove that the conditions 3, 4 hold, we need some preliminary result.

LEMMA. Let D be a convex compact set, not containing the origin; let $\{y^k\}$ be some sequence of vectors from D, and let $\{z^k\}$ be the sequence formed from the vectors y^k in the following manner:

$$z^1 = y^1, \qquad (6)$$
$$z^{k+1} = (1 - \sigma_k) z^k + \sigma_k y^{k+1},$$

where

$$1 \geqslant \sigma_k \geqslant 0, \ \sigma_k \to 0, \ \sum_{k=1}^{\infty} \sigma_k = \infty. \qquad (7)$$

Then there exists $\gamma > 0$ and $\tilde{N} < \infty$ such that for an arbitrary sequence $\{y^k\}$, at least for one $k \leq \tilde{N}$

$$(z^k, y^{k+1}) > \gamma.$$

Proof. Since D does not contain the origin, for all m we have $0 < \delta \leq |y^m| \leq \Delta < \infty$, where

$$\delta = \min_{y\in D} |y|, \quad \Delta = \max_{y\in D} |y|.$$

We set $\gamma = \frac{1}{2}\delta^2$ and we assume that for all k

$$(z^k, y^{k+1}) \leqslant \gamma. \qquad (8)$$

Then from (6) we have

$$|z^{s+1}|^2 = |z^s|^2 + 2\sigma_s((z^s, y^{s+1}) - |z^s|^2) + \sigma_s^2 |y^{s+1} - z^s|^2.$$

Since $z^s \in D$ and (8) holds, for sufficiently large s we have

$$\delta^2 \leqslant |z^{s+1}|^2 \leqslant |z^s|^2 - \sigma_s \delta^2 + 2\Delta^2\sigma_s^2 \leqslant |z^s|^2 - \gamma\sigma_s.$$

Summing this inequality with respect to s, and taking into account (7), we obtain a contradiction. Consequently, there exists $\tilde{s}$ such that

$$(z^{\tilde{s}}, y^{\tilde{s}+1}) > \gamma.$$

From the proof it is clear that $\tilde{s} < \tilde{N}$, where $\tilde{N} < \infty$ does not depend on the choice of the sequence $\{y^k\}$. The lemma is proved.

Remark. The lemma remains true also when D is some not necessarily convex subset of a convex compact set not containing the origin.

We prove now that the conditions 3, 4 hold for the sequence $\{x^n\}$, generated by the relation (5). We assume that there exists a subsequence $\{x^{n_k}\} \to x' \overline{\in} X^*$ and that at first $x' \in \text{int } S$. Since $0 \overline{\in} M(x')$, one can show that there exists $\varepsilon > 0$ such that for some direction e, all $x \in U_{4\varepsilon}(x') = \{x : |x - x'| \leq 4\varepsilon\}$ and all $g \in M(x)$ satisfy the following inequalities:

$$(g, e) \geqslant \delta > 0, \qquad (9)$$

$$|g| \leqslant \Delta, \tag{10}$$

where δ and Δ are some constants, $U_{4\varepsilon}(x') \subset S$.

The convex compact set of the vectors g which satisfy the inequalities (9), (10) will be denoted by $\tilde{D}$. According to the construction of ε, we have

$$\tilde{D} \supset \bigcup_{x \in U_{4\varepsilon}(x')} M(x).$$

We assume now that condition 3 does not hold, i.e., for all $s > n_{k'}$ $|x^s - x^{n_{k'}}| \leqslant \varepsilon$. Putting in this inequality $s = n_k$ and taking the limit for $k \to \infty$, we have $|x' - x^{n_{k'}}| \leq \varepsilon$, from where it follows, taking into account the previous inequality, that $|x' - x^s| \leq 2\varepsilon$ for $s > n_{k'}$. We note that here we can consider k' arbitrarily large.

In order to apply the lemma, we note that

$$x^s - x^{n_k} = \sum_{r=n_k}^{s-1} \rho_r g(x^r) = \left(\sum_{r=n_k}^{s-1} \rho_r \right) z^{(k)}_{s-n_k}, \tag{11}$$

where $z^{(k)}_{s-n_k}$ can be obtained from a relation similar to (6):

$$z_1^{(k)} = g(x^{n_k}),$$

$$z^{(k)}_{m+1} = (1 - \sigma^{(k)}_m) z^{(k)}_m + \sigma^{(k)}_m g(x^{n_k+m})$$

for $\sigma^{(k)}_{s-n_k} = \dfrac{\rho_s}{\sum\limits_{r=n_k}^{s} \rho_r}$.

It can be easily checked that $\lim\limits_{m\to\infty} \sigma^{(k)}_m = 0$, $\sum\limits_{m=1}^{\infty} \sigma^{(k)}_m = \infty$, since the conditions (7) hold for every fixed k.

In addition we note that

$$\lim_{k\to\infty} \sigma^{(k)}_m = \frac{1}{m+1}.$$

By assumption $g(x^{n_k+m}) \in \tilde{D}$. Then, as it follows from the lemma, there exists $\tilde{N}_k$ such that $(z^{(k)}_{\tilde{s}_k}, g(x^{n_k+\tilde{s}_k})) > \frac{1}{2}\delta^2$ at least for one $\tilde{s}_k \leq \tilde{N}_k$. By virtue of (12) [sic] $\tilde{N}_k$ is uniformly bounded: $\tilde{N}_k \leq N < \infty$. Substituting (11), we obtain

$$(x^{n_k+\tilde{s}_k} - x^{n_k}, \; g(x^{n_k+\tilde{s}_k})) > \frac{1}{2}\delta^2 \sum_{r=n_k}^{n_k+\tilde{s}_k-1} \rho_r. \tag{13}$$

We note that inequality (1) can be rewritten in the form

$$f(y) - f(x) \geqslant (g, y - x) - r(y, x),$$

where

$$g = g(y) \in M(y).$$

Considering this inequality for $y = x^{n_k+\tilde{s}_k}$, $x = x^{n_k}$, taking into account (13), we obtain

$$f(x^{n_k+\tilde{s}_k}) - f(x^{n_k}) \geqslant \frac{1}{2}\delta^2 \sum_{r=n_k}^{n_k+\tilde{s}_k-1} \rho_r - r(x^{n_k+\tilde{s}_k}, x^{n_k}),$$

where $\tilde{s}_k \leq N < \infty$.

We note that $\|x^{n_k+\tilde{s}_k} - x^{n_k}\| \to 0$ for $k \to \infty$.

By virtue of this, for sufficiently large k we have

$$\frac{|r(x^{n_k+\tilde{s}_k}, x^{n_k})|}{\|x^{n_k+\tilde{s}_k} - x^{n_k}\|} \leqslant \frac{\delta^2}{4\Delta}.$$

Since

$$\|x^{n_k+\tilde{s}_k} - x^{n_k}\| \leqslant \Delta \sum_{r=n_k}^{n_k+\tilde{s}_k-1} \rho_k,$$

we have

$$f(x^{n_k+\tilde{s}_k}) - f(x^{n_k}) \geqslant \frac{1}{4}\delta^2 \sum_{r=n_k}^{n_k+\tilde{s}_k-1} \rho_r. \tag{14}$$

We recall that by assumption $x^s \in U_{2\varepsilon}(x') \subset U_{4\varepsilon}(x')$ for $s > n_k$. Repeating the arguments given at the proof of the inequality (14), for an arbitrary index $s > n_k$, we obtain

$$f(x^s) - f(x^{n_k}) \geqslant \frac{1}{4}\delta^2 \sum_{r=n_k}^{s-1} \rho_r - \frac{1}{4}\delta^2 \sum_{r=s-r_k}^{s-1} \rho_r + f(x^s) - f(x^{s-r_k}), \tag{15}$$

where $r_k \leq N < \infty$.

$$\sum_{r=s-r_k}^{s-1} \rho_r \leqslant N \sup_{r>s-r_k} \rho_r \to 0 \quad \text{for} \quad s \to \infty.$$

Similarly

$$f(x^s) - f(x^{s-r_k}) \to 0 \quad \text{for} \quad s \to \infty,$$

since $\|x^s - x^{s-r_k}\| \to 0$.

Therefore, taking the limit in (15) for $s \to \infty$, by virtue of (6) we obtain a contradiction with the boundedness of the continuous function $f(x)$ on the compact set $U_{2\varepsilon}(x')$. The obtained contradiction proves that condition 3 holds. Let

$$m_k = \min_{r>n_k} r : \|x^r - x^{n_k}\| > \varepsilon.$$

By definition $x^{m_k} \bar{\in} U_\varepsilon(x^{n_k})$, but for sufficiently large k, $x^{m_k} \in U_{4\varepsilon}(x')$. Therefore the inequality (15) remains true also for $s = m_k$, i.e.,

$$f(x^{m_k}) - f(x^{n_k}) \geqslant \frac{1}{4}\delta^2 \sum_{r=n_k}^{m_k-1} \rho_r - \frac{1}{4}\delta^2 \sum_{r=m_k-r_k}^{m_k-1} \rho_r + f(x^{m_k}) - f(x^{m_k-r_k}).$$

But

$$\varepsilon < \|x^{m_k} - x^{n_k}\| \leqslant \Delta \sum_{r=n_k}^{m_k-1} \rho_r.$$

Therefore

$$f(x^{m_k}) - f(x^{n_k}) \geqslant \frac{\varepsilon\delta^2}{4\Delta} - \frac{\delta^2}{4} \sum_{r=m_k-r_k}^{m_k-1} \rho_r + f(x^{m_k}) - f(x^{m_k-r_k}), \tag{16}$$

where, as mentioned before, $r_k \leq N < \infty$.

Because of this,

$$\lim_{k\to\infty} |f(x^{m_k-r_k}) - f(x^{m_k})| = 0,$$

$$\lim_{k\to\infty} \sum_{r=m_k-r_k}^{m_k-1} \rho_r = 0.$$

Therefore, taking in (16) the limit for $k \to \infty$, we obtain

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

Thus, we have proved that the conditions 3, 4 hold at the interior points x' of the set S. As asserted in the remark to the proof of Theorem 2 of [1], this guarantees the finiteness of the number of "jumps" in the set A; i.e., there exist only a finite number of indices n such that $x^n + \rho_n g(x^n) \in S$. Therefore, one can prove in a similar way that the conditions 3, 4 hold at the points x' which are not interior points of the set S.

The convergence of the algorithm (5) is a consequence of the fact that conditions 1-6 hold.

LITERATURE CITED

1. E. A. Nurminskii, "Conditions for the convergence of the algorithms of nonlinear programming," Kibernetika, No. 6 (1972).