



Soviet-era science, translated into English

ONE GENERAL METHOD FOR SOLVING EXTREMAL PROBLEMS

MATHEMATICS

1967

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196701.16899>

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

Abstract

Full Text

UDC 519.3

MATHEMATICS

B. T. POLYAK

ONE GENERAL METHOD FOR SOLVING EXTREMAL PROBLEMS

(Presented by Academician L. V. Kantorovich on 22 VI 1966)

The mathematical theory of extremal problems is at present developed unevenly. Thus, necessary conditions for an extremum ⁽¹⁻³⁾, existence theorems, and convergence conditions for minimizing sequences ⁽⁴⁾ have been obtained for a broad class of problems. At the same time, minimization methods have been studied only for the case of differentiable functionals and fairly simple constraints ⁽⁵⁾.

In this note a solution method is considered that is suitable for nonsmooth functionals and constraints of general form. Its idea was first stated by N. Z. Shor for the problem of an unconstrained extremum in a finite-dimensional space ⁽⁶⁾. It consists in choosing, instead of the antigradient, a direction defined by an arbitrary supporting functional. This idea is nontrivial, since the functional, generally speaking, does not decrease in the indicated direction. The transfer of this idea to the case of more general extremal problems, in combination with a proper choice of step length, constitutes the content of the article.

The problem under consideration consists in minimizing a functional $f(x)$ on a set Q of a Hilbert space E , where $Q = Q_1 \cap Q_2$, $Q_1 = \{x : g(x) \leq 0\}$, and $g(x)$ is a functional on E . We shall assume that Q_1 has interior points: $Q_1^0 \neq \phi$, and that we can project onto Q_2 , i.e., find $P_{Q_2}(x) \in Q_2$, $\|x - P_{Q_2}(x)\| = \inf_{y \in Q_2} \|x - y\|$.

The method consists in constructing the sequence

$$x^{n+1} = P_{Q_2}(x^n + \bar{x}^n), \quad (1)$$

where \bar{x}^n is an arbitrary supporting* vector at the point x^n to the set $\{x : f(x) \leq f(x^n)\}$, if $x^n \in Q_1$, and to the set $\{x : g(x) \leq g(x^n)\}$, if $x^n \notin Q_1$, while the step length $\lambda_n = \|\bar{x}^n\|$ satisfies the condition

$$\lim_{n \rightarrow \infty} \lambda_n = 0, \quad \sum_{n=0}^{\infty} \lambda_n = \infty. \quad (2)$$

In other words, we take a step along a supporting vector to the minimized functional if x^n satisfies the constraints, or along a supporting vector to the

constraints if they are violated, the step length not depending on the results of the computations.

Theorem 1. *Let $f(x), g(x)$ be quasiconvex ⁽⁴⁾ continuous functionals, let Q_2 be closed and convex, and let $Q_1^0 \cap Q_2 \neq \emptyset$. Then, for any x^0 , in the method (1), (2) there is a subsequence x^{n_k} for which*

$$x^{n_k} \in Q, \quad \lim_{k \rightarrow \infty} f(x^{n_k}) = f^* = \inf_{x \in Q} f(x).$$

Proof. First of all, the method (1), (2) is defined for any x^n that is not a solution. Next, choose an arbitrary $a > f^*$

* A linear functional $c \in E^*$ is called supporting to $R \subset E$ at a point $x \in R$, if $(c, x) \leq (c, y)$ for all $y \in R$.

and denote $S_\alpha = \{x : x \in Q_1, f(x) \leq \alpha\}$. There exists $\tilde{x} \in Q_2 \cap S_\alpha^0$ (by virtue of the assumption $Q_1^0 \cap Q_2 \neq \emptyset$ and $\alpha > f^*$). Choose $\rho > 0$ so that $x \in S_\alpha$ for all $\|x - \tilde{x}\| \leq \rho$. Suppose now that $x^n \notin S_\alpha$ for all n . Then, if $x^n \in Q_1$, then $S_\alpha = \{x : f(x) \leq f(x^n)\}$; if $x^n \notin Q_1$, then $S_\alpha \subset Q_1 \subset \{x : g(x) \leq g(x^n)\}$, so that in all cases $(\bar{x}^n, x^n) \leq (\bar{x}^n, x)$ for all $x \in S_\alpha$, in particular,

$$(\bar{x}^n, x^n) \leq (\bar{x}^n, \tilde{x} - \rho \bar{x}^n / \|\bar{x}^n\|).$$

Therefore

$$\begin{aligned} \|x^{n+1} - \tilde{x}\|^2 &= \|P_{Q_2}(x^n + \bar{x}^n) - \tilde{x}\|^2 \leq \|x^n + \bar{x}^n - \tilde{x}\|^2 \\ &= \|x^n - \tilde{x}\|^2 + \|\bar{x}^n\|^2 + 2(\bar{x}^n, x^n - \tilde{x} + \rho \bar{x}^n / \|\bar{x}^n\|) - 2\rho \|\bar{x}^n\| \\ &\leq \|x^n - \tilde{x}\|^2 + \lambda_n^2 - 2\rho \lambda_n. \end{aligned}$$

Choose N such that $\lambda_n \leq \rho$ for all $n \geq N$, and sum the inequalities obtained above from $n = N$ to $n = N + m$. We obtain

$$\begin{aligned} 0 \leq \|x^{N+m+1} - \tilde{x}\|^2 &\leq \|x^N - \tilde{x}\|^2 + \sum_{n=N}^{N+m} \lambda_n (\lambda_n - 2\rho) \leq \\ &\leq \|x^N - \tilde{x}\|^2 - \rho \sum_{n=N}^{N+m} \lambda_n, \end{aligned}$$

which is impossible by the divergence of $\sum \lambda_n$. Now choose $\alpha_k \rightarrow f^*$; by what has been proved, there will be $x^{n_k} \in S_{\alpha_k}$, and these x^{n_k} are the desired ones.

Under somewhat stronger assumptions one can prove a stronger result.

Theorem 2. *Suppose that, in addition to the conditions of Theorem 1, $S = \{x : f(x) = f^*, x \in Q\}$ is nonempty and from the conditions $f(x) \rightarrow f^*, x \in Q$, it follows that*

$$\rho(x, S) = \inf_{y \in S} \|x - y\| \rightarrow 0.$$

Then, in the method (1), (2), $\lim_{n \rightarrow \infty} \rho(x^n, S) = 0$.

The conditions of Theorem 2 are satisfied, for example, in the following cases: a) $f(x)$ is strongly convex ⁽⁴⁾ (then there exists a unique minimizer x^* and $\lim \|x^n - x^*\| = 0$); b) E is finite-dimensional, Q is bounded.

Let us now consider several special cases. If $Q = E$, then we obtain an unconstrained extremum problem. The method (1) takes the form

$$x^{n+1} = x^n + \bar{x}^n, \tag{1'}$$

where \bar{x}^n is a support vector to the set $\{x : f(x) \leq f(x^n)\}$. Such a method in the finite-dimensional case was proposed by N. Z. Shor ⁽⁶⁾. He, however, considered only the method with $\lambda_n \equiv \lambda$, which, generally speaking, does not converge (but allows one to find the minimum with accuracy up to $K\lambda$)^{*}. If $f(x)$ is a differentiable functional, then $\bar{x}^n = -\lambda_n f'(x^n) / \|f'(x^n)\|$, i.e. (1') becomes the gradient method. However, the rule for choosing the step length (2) differs from the usual variants of the gradient method ⁽⁷⁾, in which $f(x^n)$ decreases monotonically. A method close to (1'), (2) for differentiable functionals in a finite-dimensional space was considered by V. A. Vol'konskii ⁽⁸⁾. If $Q_1 = E$, then for a differentiable functional φ the method (1), (2) coincides with the gradient projection method ⁽⁵⁾, differing from its known variants in the rule for choosing the step length.

Let us note that the proposed method is applicable to approximately the same class of problems for which necessary and sufficient conditions for an extremum have been obtained ⁽¹⁻³⁾, and is formulated in the same terms of support functionals as these conditions.

Below we shall write out the concrete form of the method (1), (2) for a number of examples. In doing so, we use the well-developed ^(2, 3) technique for finding support functionals. As a rule, we shall not write out the entire set of support functionals, but shall confine ourselves only to its simplest representative.

** Note added in proof.* For this same case, a rule for choosing the step length (2) has been proposed in the article by Yu. M. Ermol'ev ⁽¹⁰⁾.

1. Solution of problems of mathematical programming and inequalities

Let Q_1 be defined by means of a finite number of inequalities:

$$Q_1 = \{x : g_i(x) \leq 0, i = 1, \dots, m\}.$$

Then

$$Q_1 = \{x : g(x) \leq 0\}, \quad g(x) = \max_{1 \leq i \leq m} g_i(x).$$

If $Q_1^0 \neq \emptyset$, and all $g_i(x)$ are quasiconvex, then a supporting vector to $\{x : g(x) \leq a\}$ has the form

$$c = \sum_{i=1}^m \beta_i c_i,$$

where $\beta_i \geq 0$, $\beta_i = 0$ if $g_i(x) < a$, and c_i are supporting vectors to $\{x : g_i(x) \leq a\}$. In particular, if $g_i(x)$ are differentiable, then

$$c = - \sum_{i=1}^m \beta_i g'_i(x);$$

in particular, one may take $c = -g'_j(x)$, where j is any of the indices for which $g_j(x) = a$. Thus, for the problem

$$\min f(x), \quad g_i(x) \leq 0, \quad i = 1, \dots, m,$$

where $f(x), g_1(x), \dots, g_m(x)$ are quasiconvex and differentiable, method (1), (2) takes the form

$$x^{n+1} = x^n + \bar{x}^n, \quad \bar{x}^n = \begin{cases} -\lambda_n f'(x^n) / \|f'(x^n)\|, & \text{if } g_i(x^n) \leq 0, \quad i = 1, \dots, m, \\ -\lambda_n g'_j(x^n) / \|g'_j(x^n)\|, & \text{if } g_j(x^n) = \max_{1 \leq i \leq m} g_i(x^n) > 0. \end{cases}$$

If $Q_1^0 \neq \emptyset$, for example, if there exists \tilde{x} for which $g_i(\tilde{x}) < 0$, $i = 1, \dots, m$, then Theorem 1 is applicable. If there are linear constraints of equality type or constraints of the type

$$\alpha_k \leq x_k \leq \beta_k$$

in the finite-dimensional case, then it is expedient to regard them as Q_2 . If the problem consists simply in solving the system of inequalities

$$g_i(x) \leq 0, \quad i = 1, \dots, m,$$

then the method

$$x^{n+1} = x^n - \lambda_n g'_j(x^n) / \|g'_j(x^n)\|, \quad g_j(x^n) = \max_{1 \leq i \leq m} g_i(x^n),$$

gives a solution in a finite number of steps when $Q_1^0 \neq \emptyset$.

2. Problems of best approximation

It is required to minimize

$$f(x) = \max_{t \in R} \left| a(t) + \sum_{i=1}^m x_i \varphi_i(t) \right|,$$

where $a(t), \varphi_1(t), \dots, \varphi_m(t)$ are given continuous functions on the compact set R , and $x = (x_1, \dots, x_m)$ is finite-dimensional. Method (1), (2) has the following form:

$$x_i^{n+1} = x_i^n + \bar{x}_i^n, \quad \bar{x}_i^n = -\frac{\lambda_n \varphi_i(t_n)}{(\sum_{i=1}^m \varphi_i^2(t_n))^{1/2}} \operatorname{sign} \left(a(t_n) - \sum_{i=1}^m x_i^n \varphi_i(t_n) \right),$$

where $t_n \in R$ is any of the points for which

$$\left| a(t_n) - \sum_{i=1}^m x_i^n \varphi_i(t_n) \right| = \max_{t \in R} \left| a(t) - \sum_{i=1}^m x_i^n \varphi_i(t) \right|.$$

If $\varphi_1(t), \dots, \varphi_m(t)$ are linearly independent, then Theorem 2 is applicable.

3. Optimal control problems

One minimizes

$$\int_0^T F(x, u, t) dt + \Phi(x(T)),$$

where

$$\frac{dx}{dt} = Ax + Bu,$$

$x(0)$ and T are fixed, under the constraints

$$u(t) \in M \subset E^r$$

for almost all $0 \leq t \leq T$, and

$$g(x(t)) \leq 0$$

for all $0 \leq t \leq T$. The problem is reduced to minimization of

$$f(u) = \int_0^T F(x, u, t) dt + \Phi(x(T)),$$

where $x = x(u)$ is found from the differential equation, under the constraints

$$u \in Q_1 \cap Q_2, \quad Q_1 = \{u : g(u) \leq 0\},$$

$$g(u) = \max_{0 \leq t \leq T} g(x; t), \quad Q_2 = \{u : u(t) \in M, 0 \leq t \leq T\},$$

in the space $L_2^r(0, T)$. Method (1), (2) takes the form

$$u^{n+1}(t) = P_M \left(u^n(t) + \lambda_n \frac{\bar{u}_n(t)}{\|\bar{u}_n\|_{L_2}} \right),$$

$$\bar{u}^n(t) = B^* \psi(t) - F_u(x^n, u^n, t), \quad \frac{d\psi(t)}{dt} = -A^* \psi(t) + F_x(x^n, u^n, t)$$

$$\psi(T) = -\Phi'(x^n(T)), \quad \text{if } q(x^n(t)) \leq 0, \quad 0 \leq t \leq T,$$

$$\bar{u}^n(t) = B^*\psi(t), \quad d\psi(t)/dt = -A^*\psi(t), \quad 0 \leq t \leq t_n,$$

$$\psi(t_n) = -q'(x^n(t_n)),$$

$$\psi(t) \equiv 0, \quad t_n < t \leq T, \quad \text{if } q(x^n(t_n)) = \max_{0 \leq t \leq T} q(x^n(t)) > 0.$$

Here P_M denotes projection (at each instant t) onto the finite-dimensional set M ; $F_u(x^n, u^n, t)$, $F_x(x^n, u^n, t)$, $\Phi'(x^n(T))$, $q'(x^n(t))$ denote supporting vectors to the finite-dimensional functions $F(x^n, u, t)$, $F(x, u^n, t)$, $\Phi(x)$, $q(x)$ at the corresponding points*, coinciding in the differentiable case with the derivatives. If $F(x, u, t)$ is continuous in $\{x, u, t\}$ and convex in $\{x, u\}$, $\Phi(x)$, $q(x)$ are continuous and convex, M is convex, closed, and bounded, and there exists $\tilde{u}(t) \in M$, $0 \leq t \leq T$, for which $q(\tilde{u}(t)) < 0$ for all $0 \leq t \leq T$, then Theorem 1 is applicable.

4. Variational problems with partial derivatives.

We shall confine ourselves to one concrete problem of this type ⁽⁹⁾.

One seeks $u(x, y) \in \dot{W}_2^1(S)$, minimizing

$$f(u) = \int_S \left(\frac{\mu}{2} |\nabla u|^2 + \tau |\nabla u| - cu \right) dx dy.$$

The method (1), (2) has the form

$$u^{n+1}(x, y) = u^n(x, y) + \lambda_n \frac{\bar{u}^n(x, y)}{\|\bar{u}^n\|_{W_2^1}}, \quad u^0(x, y) \in \dot{W}_2^1,$$

$$\Delta \bar{u}^n = \mu \Delta u^n + c + \psi, \quad \bar{u}_\Gamma^n = 0,$$

$$\psi(x, y) = \begin{cases} \tau \left(\frac{\partial}{\partial x} \frac{u_x^n}{|\nabla u^n|} + \frac{\partial}{\partial y} \frac{u_y^n}{|\nabla u^n|} \right), & \text{if } |\nabla u^n(x, y)| \neq 0, \\ 0, & \text{if } |\nabla u^n(x, y)| = 0. \end{cases}$$

Here Γ is the boundary of the domain S , $\mu, \tau > 0$. All the hypotheses of Theorem 2 are satisfied; the minimum u^* exists and is unique, so that $u^n \rightarrow u^*$ in \dot{W}_2^1 .

As is seen from the examples, the merit of the method (1), (2) from the computational point of view is its extraordinary simplicity and universality. The important question of the rate of convergence requires experimental study.

The author expresses gratitude to N. Z. Shor and E. S. Levitin for useful discussions.

Moscow State University
named after M. V. Lomonosov

Received
1 VI 1966

REFERENCES

1. L. V. Kantorovich, DAN, 28, No. 3 (1940).
2. A. Ya. Dubovitskii, A. A. Milyutin, Zhurn. vychislit. matem. i matem. fiz., 5, No. 3 (1965).
3. B. N. Pshenichnyi, Kibernetika, No. 5 (1965).
4. B. T. Polyak, DAN, 166, No. 2 (1966).
5. E. S. Levitin, B. T. Polyak, Zhurn. vychislit. matem. i matem. fiz., 6, No. 5 (1966).
6. N. Z. Shor, Candidate Dissertation, Kiev, 1964.
7. B. T. Polyak, Zhurn. vychislit. matem. i matem. fiz., 3, No. 4 (1963).
8. V. A. Volkonskii, Ekonomika i matematich. metody, 1, No. 2 (1965).
9. P. P. Mosolov, V. P. Myasnikov, PMM, 29, No. 3 (1965).
10. Yu. M. Ermoliev, Kibernetika, No. 4 (1966).

* A linear functional $c \in E^*$ is called a supporting functional to the functional $f(x)$ at the point x if $f(x + y) \geq f(x) + (c, y)$ for all $y \in E$.

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

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