MATHEMATICS

1962

SovietRxiv

Abstract

Full Text

MATHEMATICS

V. V. IVANOV

ON A GENERAL APPROXIMATE METHOD FOR SOLVING LINEAR PROBLEMS

(Presented by Academician I. N. Vekua on 3 November 1961)

Let E_1 and E_2 be linear normed spaces, and let a linear operator A be given with domain $G_1 \subseteq E_1$ and range $G_2 \subseteq E_2$. It is required to construct a sequence $\{x_n\}$ for which the relation

$$\lim_{n\to\infty} \|y - Ax_n\| = \inf_{x\in G_1} \|y - Ax\|, \qquad y \in E_2. \tag{1}$$

is satisfied.

Suppose that in G_1 there exists an A-complete system $e_1, e_2, \dots, e_n$ (^{1}), and we shall seek x_n in the form

$$x_n = \sum_1^n \alpha_k e_k, \tag{2}$$

(α_k may always be assumed real) from the condition that the sought values of the parameters α_k^0 minimize

$$\Delta(P) = \left\| y - \sum_1^n \alpha_k y_k \right\|, \qquad y_k = Ae_k \neq 0, \qquad P(\alpha_1, \alpha_2, \dots, \alpha_n). \tag{3}$$

It is not difficult to show that

$$\lim_{n\to\infty} \left\| y - \sum_1^n \alpha_k^0 y_k \right\| = \inf_{x\in G_1} \|y - Ax\|. \tag{4}$$

Further, it is easy to see that

$$\Delta\left[\tfrac{1}{2}(P_1 + P_2)\right] \leq \tfrac{1}{2}[\Delta(P_1) + \Delta(P_2)],$$

i.e. Δ is a convex function. Therefore it is natural, for finding α_k^0, to apply various methods of steepest descent.

We consider an approximate method X, consisting of three principal parts.

1. **Determination of $\inf \Delta$ on a given interval $[P', P'']$.**
   Put $\Delta(t) = \Delta(P)$, $P = P' + t(P'' - P')$. Having computed Δ at three values $t = t^{(1)} = \frac{1}{2}$, $t^{(1)} \pm \frac{1}{2}$, choose $t = t^{(2)}$ for which Δ is smallest. If all three values are equal, then they coincide with the required one. Otherwise, repeating the indicated operation for the values $t = t^{(2)}$, $t^{(2)} \pm {\frac{1}{2}}^2$, we find $t = t^{(3)}$, for which Δ is smallest, and so on.*

2. **Determination of $\inf \Delta$ in a given direction $-\nabla$.**
   Let P_1 be the initial point. Form the sequence of points

   $$T_r = P_1 + 2^{r-2}\sigma\nabla, \qquad r = 2, 3, \dots \quad (T_0 = T_1 = P_1),$$

   where σ is a given positive number, and compute

   $$\gamma_r = \Delta(T_r) - \Delta(T_{r-1}), \qquad r = 2, 3, \dots .$$

   Let r_0, $r_0 \geq 2$, be the first value for which $\gamma_{r_0} \geq 0$. Then the required value of Δ is attained on the interval $[T_{r_0-2}, T_{r_0}]$, and it can be found by algorithm 1.

Remark 1. For functions Δ of the form (3), Problem 2 reduces to determining the least value of the function $\varphi(a) = \|u - av\|$, where $v \neq \theta$ and u are given elements. It is not difficult to conclude that in this case the desired value is attained on the segment $[-2\|u\|/\|v\|,\ 2\|u\|/\|v\|]$, and Algorithm 1 determines it in the number of steps $k = \log_2 4\|u\|/h$ with an accuracy not exceeding h.

3. Determination of the direction of decrease of Δ. Suppose that there is an algorithm for searching through points which determines $\inf \Delta$ along the boundary Γ_n of an n-dimensional cube with center at the given point P', with faces parallel to the coordinate planes and with edge length 2δ. As soon as, in the course of solving this problem, a point $P'' \in \Gamma_n$ is obtained for which $\Delta(P') - \Delta(P'') > \sigma$, we shall regard the ray

$$P = P' + t(P'' - P'), \qquad t \geq 0,$$

as the desired direction.

Noting that for any points P_1 and P_2

$$|\Delta(P_1) - \Delta(P_2)| \leq \rho(P_1, P_2)\left(\sum_1^n \|y_k\|^2\right)^{1/2} \leq C_1\rho(P_1, P_2),$$

and denoting by P'^0 the point nearest to P' at which $\Delta(P'^0) = \inf \Delta$ over all $\{\alpha_k\}$, let us join P' with P'^0 by a straight-line segment. Then, if

$$\Delta(P') - \Delta(P'^0) > C_1\delta\sqrt{n},$$

* If $t^{(j)} > 1$ or $t^{(j)} < 0$, then the required value is attained, respectively, at $t = 1$ or $t = 0$.

then along this segment Δ will decrease, and, by the convexity of Δ,

$$\Delta(P') - \Delta(P''') \geq \frac{\delta}{\rho(P', P'^0)} [\Delta(P') - \Delta(P'^0)], \tag{5}$$

where P''' is the point of intersection of the segment with Γ_n.

Let $P_1 = P'$ and $P_2 = P' + t^0(P'' - P')$, where

$$\Delta[P' + t^0(P'' - P')] = \inf_t \Delta[P' + t(P'' - P')].$$

Taking P_2 as the initial point instead of P_1, in an analogous way we obtain P_3, then P_4, etc. Using (5) and the fact that

$$\Delta(P) = \Delta(P^*) = \left\| y - \sum_1^m a_i y_{k_i} \right\|, \tag{6}$$

where $\{y_{k_i}\}$ is a linearly independent system equivalent to $\{y_k\}$, it is not difficult to prove the following theorem:

Theorem 1. *Method X always converges in the sense that, whatever $\varepsilon > 0$, for $\delta < \varepsilon / C_1 \sqrt{n}$ and for sufficiently small σ*

$$\Delta(P_j) - \inf_{\{\alpha_k\}} \Delta(P) < \varepsilon \tag{7}$$

for all j, starting from some one.

Remark 2. In some problems (for example, in solving a system of inequalities) the corresponding function Δ, while possessing the property of convexity, does not, generally speaking, possess all the properties of a function of the form (3). Method X can also be extended to these problems, with the same convergence theorem for the method, under the condition that the range of values of all those P for which $\Delta(P) \leq \text{const} = \Delta(P_1)$ is bounded.

The most laborious task is that of determining $\inf \Delta$ on Γ_n. To solve it we choose a sequence of numbers $\sigma \ll \delta_{n-1} \ll \delta_{n-2} \ll \cdots \ll \delta_1 = \delta$ and determine the least values of $2n$ convex functions of $n-1$ variables:

$$\Delta(\alpha_1 \pm \delta, \alpha_2, \dots, \alpha_n), \quad \Delta(\alpha_1, \alpha_2 \pm \delta, \alpha_3, \dots, \alpha_n), \dots,$$

$$\dots, \Delta(\alpha_1, \alpha_2, \dots, \alpha_{n-1}, \alpha_n \pm \delta).$$

We do this by Method X, taking the corresponding $(n-1)$-dimensional cubes with edge length $2\delta_2 \ll 2\delta_1$. In this way, by induction, the whole matter is reduced to determining least values in the given directions. As is clear, to implement such an algorithm one needs a computer with a large word length.

For comparatively large n, it will in practice be more advantageous to solve this problem by computing Δ at the nodes of uniform cubic grids with gradually decreasing step size, superimposed on Γ_n. In all those cases when P'^0 lies outside Γ_n, the largest step h^* for which points P'' having the property $\Delta(P') - \Delta(P'') > \sigma$ are found on the grids may naturally be called a δ, σ-measure of the conditioning of the original problem. The smaller h^*, the worse conditioned the problem is and the more difficult it is to solve.

As an example, consider the problem of determining the optimal parameters of a linear differential equation with constant coefficients of order not higher than the 3rd from a given transition function $h(t)$. Introduce the function

$$\Delta(P) = \max_t \left| \frac{d^3 h}{dt^3} + a_1 \frac{d^2 h}{dt^2} + a_2 \frac{dh}{dt} + a_3 h + a_4 \right|, \qquad P(a_1, a_2, a_3, a_4). \tag{8}$$

The problem is to find P^0 for which $\Delta(P)$ attains the smallest value. The function Δ is convex, and by algorithm X the problem can be solved effectively, since the number of undetermined parameters is small. The known algorithms of best approximation (^{2}) are, generally speaking, not applicable in this case, since the system of functions d^2h/dt^2, dh/dt, h, 1 need not be a Chebyshev system.

Let $\Delta(P)$ be an arbitrary everywhere continuously differentiable convex function of n real variables. In seeking $\inf \Delta$ in this case, algorithm 3 can be substantially simplified by taking as the sought directions $-\operatorname{grad} \Delta$ (the method of steepest descent) or, successively, the vectors $e_1(1, 0, \dots, 0)$, $e_2(0, 1, 0, \dots, 0), \dots, e_n(0, \dots, 0, 1)$ (the method of coordinate descent).

Theorem 2. *The method of steepest descent always converges in the sense that*

$$\lim_{j \to \infty} \Delta(P_j) = \inf_{\{P\}} \Delta(P), \tag{9}$$

where

$$P_j = P_{j-1} + t_j \operatorname{grad} \Delta(P_{j-1}), \qquad \Delta(P_j) = \inf_{\{t\}} \Delta[P_{j-1} + t \operatorname{grad} \Delta(P_{j-1})]. \tag{10}$$

Proof. Assuming that $\operatorname{grad} \Delta(P_j) \neq 0$, $j = 1, 2, \dots$, select from $\{P_j\}$ a convergent subsequence $\{P_{j_k}\}$, $\lim_{k \to \infty} P_{j_k} = \widetilde{P}$.

If $\operatorname{grad} \Delta(\widetilde{P}) \neq 0$, then

$$\Delta(\widetilde{P}) \leq \inf_{\{i_k\}} \Delta[P_{j_k} + t_{j_k+1} \operatorname{grad} \Delta(P_{j_k})] \leq$$

$$\leq \lim_{k \to \infty} \Delta[P_{j_k} + t^0 \operatorname{grad} \Delta(P_{j_k})] = \inf_{\{t\}} \Delta[\widetilde{P} + t \operatorname{grad} \Delta(\widetilde{P})] < \Delta(\widetilde{P}).$$

Consequently, $\operatorname{grad} \Delta(\widetilde{P}) = 0$, and the theorem may be regarded as proved.

Suppose that at every point, except possibly at points where $\Delta(P)$ assumes its smallest value, the minimum of Δ in each of the directions e_s, $s = 1, 2, \dots, n$, is attained only for a single value of the argument. Then the following is true:

Theorem 3. *The method of coordinate descent always converges in the sense that*

$$\lim_{j \to \infty} \Delta(P_j^s) = \inf_{\{P\}} \Delta(P), \qquad s = 1, 2, \dots, n, \tag{11}$$

where

$$P_j^{s+1} = P_j^s + t_j^s e_{s+1}, \qquad \Delta(P_j^{s+1}) = \inf_{\{t\}} \Delta(P_j^s + t e_{s+1}), \qquad P_j^{n+1} = P_{j+1}^1,\ e_{n+1} = e_1. \tag{12}$$

Proof. Let $\lim\limits_{k \to \infty} P_{jk}^s = \widetilde{P}^s,\ 1 \le s \le n$, and let

$$\gamma_i = \inf_{\{t\}} \Delta(\widetilde{P}^s + t e_{s+i}) = \Delta(\widetilde{P}^s), \quad i = 1, 2, \dots, r; \quad \gamma_{r+1} < \Delta(\widetilde{P}^s).$$

If $\{t_{jk}^{s+i}\}$ do not converge to zero, then, putting $\lim\limits_{l \to \infty} t_{k_l}^{s+i} = t^{s+i} \neq 0$, we find

$$\lim_{l \to \infty} \Delta(P_{jk_l}^s + t_{jk_l}^{s+i} e_{s+i}) = \Delta(\widetilde{P}^s + t^{s+i} e_{s+i}) = \Delta(\widetilde{P}^s),$$

which is impossible. Consequently,

$$\lim_{k \to \infty} t_{jk}^{s+i} = 0, \quad i = 1, 2, \dots, r.$$

We now have the estimate

$$d = \Delta(\widetilde{P}^s) - \gamma_{r+1} \le \Delta(P_{jk}^s + t_{jk}^{s+1} e_{s+1} + t_{jk}^{s+2} e_{s+2} + \dots + t_{jk}^{s+r} e_{s+r} + t^0 e_{s+r+1}) -$$

$$-\Delta(\widetilde{P}^s + t^0 e_{s+r+1}) \le \text{const} \left[\rho(P_{jk}^s, \widetilde{P}^s) + \sum_{i=1}^{r} \left| t_{jk}^{s+i} \right| \right] \to 0,$$

which is impossible. Hence, at the point $\widetilde{P}^s$ there is no decrease in any of the coordinate directions. This means that the tangent plane to the surface Δ at the point $(\widetilde{\Delta}, \widetilde{P}^s)$, $\widetilde{\Delta} = \Delta(\widetilde{P}^s)$, has the equation $\Delta = \widetilde{\Delta}$. But, by virtue of the convexity of Δ, the tangent plane is simultaneously a supporting plane, so that all points of Δ, except $(\widetilde{\Delta}, \widetilde{P}^s)$, are situated above the plane $\Delta = \widetilde{\Delta}$. This is equivalent to the assertion of the theorem.

In the case when E_1 and E_2 are Hilbert spaces and $\Delta(P)$ is a function of the form (3), one can prove more precise theorems than Theorems 2 and 3. Namely, it turns out that the corresponding points P_j and P_j^s always converge to quite definite points, depending, generally speaking, on the initial point (see (^{3})).

Computing Center
Academy of Sciences of the USSR

Received
30 X 1961

REFERENCES

1. S. G. Mikhlin, *Direct Methods in Mathematical Physics*, 1959.

2. E. Ya. Remez, *General Computational Methods of Chebyshev Approximation. Problems with Linearly Entering Real Parameters*, Kiev, 1957.

3. V. V. Ivanov, *Zhurnal vychislitel' noi matematiki i matematicheskoi fiziki*, **1**, No. 6 (1961).

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

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