MATHEMATICS

V. V. IVANOV

1962

SovietRxiv

Abstract

Full Text

MATHEMATICS

V. V. IVANOV

ON ALGORITHMS OF RAPID DESCENT

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

In the Soviet mathematical literature, works on methods of rapid descent are very few in number ([1]), although in automatic control these methods are used rather intensively, especially in the construction of so-called optimizers ([2]). In ([3]) the application of methods of rapid descent to the approximate solution of linear functional equations in arbitrary normed spaces is justified, in particular to the solution of problems of best linear approximation in an arbitrary metric and for an arbitrary set of elements.

The present note is devoted to the application of methods of rapid descent to the approximate solution of one of the basic problems of optimization theory —the determination of the least value of a function $Q(P)$ in the domain D : $H_j(P) \leqslant 0,\ j = 1, 2, \dots, m;\ P(x_1, x_2, \dots, x_n)$.

Suppose that Q and $H_j,\ j = 1, 2, \dots, m$, are everywhere continuously differentiable convex functions,* and that on the boundary of D $\operatorname{grad} H_j \neq 0,\ j = 1, 2, \dots, m$, and among the $\{H_j\}$ there are no two functions differing only in sign. Without loss of generality we shall assume D bounded, since one can always add to the functions $\{H_j\}$ the function $\rho(0, P) - R$ with sufficiently large R. To find the desired point P^0, at which the least value of Q in D is attained, we apply the rapid descent method Y, consisting of the following main parts.

A. **Finding a point lying inside D.** This can be done by the rapid descent method X (see ([3])), applied to the function

$$\Delta_1 = \max_{1 \leqslant j \leqslant m} H_j(P)$$

or to the function

$$\Delta_2 = \sum_1^m H_j(P)\, \delta_j(P),$$

where $\delta_j(P) = 0,\ H_j(P) < 0$, and $\delta_j(P) = 1,\ H_j(P) \geqslant 0$.

Moreover, problem A can be reduced to a finite sequence of problems of the type under consideration, but with initial points $\{P_j\}$ located beforehand inside the admissible domains. For example, let

$$Q_{P_1} = \sum_1^m H_j(P)\,\delta_j(P_1)$$

and $D_{P_1} : H_{j_1}(P) < 0$ for all $j_1,\ \delta_{j_1}(P_1) = 0$. Find a point $P_2 \in D_{P_1},\ Q_{P_1}(P_2) < 0$, and take

$$Q_{P_2} = \sum_1^m H_j(P)\,\delta_j(P_2)$$

and $D_{P_2} : H_{j_2}(P) < 0$ for all $j_2,\ \delta_{j_1}(P_2) = 0$. Then find $P_3 \in D_{P_2},\ Q_{P_2}(P_3) < 0$, and so on. It is clear that either such a construction is impossible, and then D has no interior points, or, for some $s,\ 1 \leqslant s \leqslant m,\ P_s \in D_s \subset D$.

B. **Finding** $\inf Q$ **in** D **in a given direction** ∇**.** Let the initial point be $P_1 \in D,\ \sigma > 0$. Form the sequence of points

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

As long as $H_j(T_r) \leqslant 0,\ j = 1, 2, \dots, m$, apply algorithm 2 (see ([3])). If r_0 is the first value for which, for some $j_0,\ 1 \leqslant j_0 \leqslant m,\ H_{j_0}(T_{r_0}) > 0$ and $H_j(T_{r_0-1}) \leqslant 0,\ j = 1, 2, \dots, m$, then take the point $R_1 = \frac{1}{2}(T_{r_0-1} + T_{r_0})$. If $H_j(R_1) \leqslant 0,\ j = 1, 2, \dots, m$, and $Q(R_1) - Q(T_{r_0-1}) \geqslant 0$, then the desired

---

* The convexity of any function $\varphi(P)$ means that $\varphi[\frac{1}{2}(P_1 + P_2)] \leqslant \frac{1}{2}[\varphi(P_1) + \varphi(P_2)]$.

the value of Q is attained on $[T_{r_0-2}, R_1]$, and it can be determined by algorithm 1 ([3]). If $Q(R_1) - Q(T_{r_0-1}) < 0$, or, for some $j_1,\ 1 \leq j_1 \leq m,\ H_{j_1}(R_1) > 0$, then we take, respectively, the point $R_2 = \frac{1}{2}(R_1 + T_{r_0})$ or $R_2' = \frac{1}{2}(R_1 + T_{r_0-1})$ and repeat with it the same operations as with R_1, etc.

In the case where the problem of solving the equations $H_j(P_1 + tV) = 0,\ j = 1, 2, \dots, m$, presents no particular difficulty (this will be so, for example, when $\{H_j\}$ is a sequence of linear or quadratic functions), it is more expedient to find the smallest positive root t' of the indicated equations and then determine $\inf Q$ on the segment $[P_1, P']$, $P' = P_1 + t'V$, by algorithm 1 ([3]).

C. Finding a direction of decrease of Q **in** D**.** If the initial point P_1 is located inside D, then as the desired direction we take $-\operatorname{grad} Q(P_1)$. Suppose now that P_1 has the property: $H_j(P_1) = 0,\ j = 1, 2, \dots, r$, and $H_j(P_1) < 0,\ j = r+1, r+2, \dots, m$. Let $\eta_j = -\operatorname{grad} H_j(P_1),\ j = 1, 2, \dots, r$, and $g = -\operatorname{grad} Q(P_1)$. The problem consists in finding a vector ξ which makes acute angles with all $\{\eta_j\}$ and with g.

In many problems, among the constraints $H_j \leq 0,\ j = 1, 2, \dots, m$, there are inequalities $-x_j \leq 0,\ j = 1, 2, \dots, n$. In that case, putting

$$\xi = x - \overline{OP}_1, \qquad x = \sum_1^n x_k e_k,$$

where $e_1(1, 0, \dots, 0)$, $e_2(0, 1, 0, \dots, 0), \dots, e_n(0, \dots, 0, 1)$, we reduce the problem to determining nonnegative solutions of the following system of linear inequalities:

$$(\xi, \eta_j) - \sigma = \sum_{k=1}^n (e_k, \eta_j) x_k - (\overline{OP}_1, \eta_j) - \sigma \geq 0, \qquad j = 1, 2, \dots, r;$$

$$(\xi, g) - \sigma = \sum_{k=1}^n (e_k, g) x_k - (\overline{OP}_1, g) - \sigma \geq 0 \tag{1}$$

under the additional condition

$$x_1 + x_2 + \dots + x_n \leq 1 + x_1^{(1)} + x_2^{(1)} + \dots + x_n^{(1)}. \tag{2}$$

From the theoretical point of view, it is more expedient to determine such a nonnegative solution of the first r inequalities (1) which realizes $\max(\xi, g)$ under condition (2). Thus, in the present case, the solution of problem C can be reduced to the use of standard subprograms of linear programming or to the search for nonnegative solutions of a system of linear equations ([4]).

In the general case we construct an equivalent initial orthonormal system of vectors $\tilde{\eta}_1, \tilde{\eta}_2, \dots, \tilde{\eta}_q$ and $\tilde{g}$ ([5]) so that

$$\tilde{\eta}_j = \sum_{k=1}^q \alpha_k^{(j)} \eta_k, \qquad j = 1, 2, \dots, q;$$

$$\eta_{q+i} = \sum_{k=1}^q \beta_k^{(i)} \eta_k, \quad i = 1, 2, \dots, p, \quad p + q = r, \quad 0 \leq p < r; \qquad g = \tilde{g} - \sum_{k=1}^q d_k \eta_k,$$

and put $\xi_j = (\xi, \eta_j),\ j = 1, 2, \dots, q$, and $\xi_0 = (\xi, g)$. Then the problem is reduced to determining a solution of the following system of inequalities:

$$\xi_j - \sigma \geq 0, \qquad j = 1, 2, \dots, q;$$

$$\sum_{k=1}^q \beta_k^{(i)} \xi_k - \sigma \geq 0, \qquad i = 1, 2, \dots, p; \tag{1'}$$

$$\xi_0 - \sigma \geq 0.$$

under the condition

$$\xi_0 + \xi_1 + \xi_2 + ... + \xi_q = 1. \tag{2'}$$

The desired vector is

$$\xi = \sum_1^q (\xi, \tilde{n}_j)\tilde{n}_j + (\xi, \tilde{g})\tilde{g} = \sum_{j=1}^q \left(\sum_{k=j}^q \alpha_j^{(k)} \tilde{n}_k + d_j \tilde{g} \right) \xi_j + \xi_0 \tilde{g}. \tag{3}$$

Here one can also pose the problem of determining $\max(\xi, g)$. If $\tilde{g} \neq 0$, then we obtain the following problem: to find a nonnegative solution of the first r inequalities $(1')$, for which

$$\xi_0 = 1 - (\xi_1 + \xi_2 + ... + \xi_q) \tag{4}$$

takes the greatest possible value. If $\tilde{g} = 0$, then one should determine a nonnegative solution of the first r inequalities $(1')$, satisfying the condition $\xi_1 + \xi_2 + ... + \xi_q = 1$, for which

$$\xi_0 = -\sum_1^q d_k \xi_k \tag{5}$$

takes the greatest possible value.

Let us note that in the case when all $H_j(P) = 0,\ j = 1, 2, \dots, r$, are essentially hyperplanes, one may set $\sigma = 0$ in the system $(1')$ and, for $\tilde{g} \neq 0$, set $\xi = \tilde{g}$. Then we arrive at the so-called method of gradient projection ([6]).

Starting from an arbitrary point $P_1 \in D$ and determining, by means of one of the algorithms indicated in C and B, a direction of decrease of Q in D and the least value of Q in D in this direction, we obtain a point $P_2 \in D$. Repeating the same operations again, we obtain $P_3 \in D$, then $P_4 \in D$, and so on.

Theorem. *The fast descent method Y, for sufficiently small σ, always converges in the sense that, whatever arbitrarily small number $\varepsilon > 0$ is given, either after a finite number of steps a point $P_s \in D$ is obtained for which none of the algorithms C determines a direction of decrease in D, and then*

$$Q(P_s) - \inf_{P \in D} Q(P) < \varepsilon, \tag{6}$$

or

$$\lim_{j\to\infty} Q(P_j) - \inf_{P\in D} Q(P) < \varepsilon. \tag{7}$$

Let us outline the proof of the theorem. Denote by D_ε the set of points of D for which $Q(P) - \inf_{P\in D} Q(P) \geq \varepsilon$, and assign to every point $P \in D$, as was done for P_1, the vectors $n_1, n_2, \dots, n_r$ and g (here the vectors themselves and r depend on P). Using the restrictions on $\{H_j\}$ and Q, it is not difficult to show that

$$C_1 = \inf_{P\in D_\varepsilon} \min\{\|n_1\|, \|n_2\|, \dots, \|n_r\|, \|g\|\} > 0$$

and that at every point $P \in D_\varepsilon$ one can draw a unit vector ξ which forms with the vectors $n_1, n_2, \dots, n_r$ and g acute angles not exceeding $\alpha < \pi/2$, where α does not depend on P. It is clear that, for $\sigma = \sigma_1 \leq C_1 \cos\alpha$, ξ will be a solution of (1). Represent ξ in the form $\xi = \xi' + \xi''$, where

$$\xi' = \sum_1^q (\xi, \tilde{n}_j)\tilde{n}_j + (\xi, \tilde{g})\tilde{g}.$$

By virtue of the orthogonality of ξ'' to ξ', the vector $\xi'/\|\xi'\|$ is a solution of the system $(1')$ for $\sigma = \sigma_1$. Hence the vector $\lambda\xi'/\|\xi'\|$, where $\lambda = (\xi_0 + \xi_1 + \xi_2 + \dots + \xi_q)^{-1}$, satisfies $(1')$ for $\sigma = \lambda\sigma_1$. It is easy to see that

$$C_2 = \inf_{P\in D_\varepsilon} \lambda > 0.$$

Take $\sigma = \sigma_2 \leq C_2 C_1 \cos\alpha$. Let the point P_s be such that no vector can be drawn from it which satisfies $(1')$ and $(2')$ for $\sigma = \sigma_2$. Then $P_s \in D_\varepsilon$, and, consequently, (6) is proved.

Let us now suppose that there is an infinite sequence of points $P_1, P_2, \dots, P_s, \dots$, at each of which one can draw a vector $\xi^{(s)}$ satisfying $(1')$ and $(2')$. Extract a convergent subsequence $\{P_{s_k}\}$, $\lim_{k\to\infty} P_{s_k} = P^0$. If $P^0 \in D_\varepsilon$, then there exists a direction of decrease ξ^0 which makes acute angles with $n_1^0, n_2^0, \dots, n_{r_0}^0$ and g^0, smaller than $a^0 < \pi/2$, where a^0 does not depend on σ. Therefore, beginning with some k, all $\xi^{(s_k)}$, for sufficiently small σ, make acute angles with $n_1, n_2, \dots, n_r$ and g not exceeding a^0. This means that the decrease of Q in passing from the point P_{s_k} to the point P_{s_k+1} cannot be less than some fixed number, i.e. $Q(P^0) = -\infty$. But this cannot be. Consequently, $P^0 \notin D_\varepsilon$, and (7) is proved.

Let us now consider the more general situation when Q is not necessarily convex and $\operatorname{grad} Q$ has a finite number of zeros. We shall find the direction of decrease as before, and $\inf Q$ in the given direction by algorithm B, but with the difference

that, in the case when the segment $[P_1, P']$ to the boundary of D is determined in advance, we shall again apply the first variant of algorithm B, taking the boundary surface to be $t-1 \leqslant 0$, where $Q(t) = Q[P_1 + t(P' - P_1)]$. In addition, when using algorithm 1 (3), in the case of equality of three values of Q, we shall continue the division of the segment with the center at the former point.

It is not difficult to verify that $Q'(t)$, for an arbitrary smooth curve $Q(t)$, has a finite number of zeros; the indicated algorithm for determining $\inf Q$ leads to a point P_2, which is either a point of $\min Q$, or the boundary point P'. In any case, the passage from P_1 to P_2, under the condition that P' is at a distance from P_1 not less than some fixed constant independent of the position of P_1, leads to a decrease of $Q(P_1) - Q(P_2)$ to zero only when $Q'_t(P_1) \to 0$. This means that either $\operatorname{grad} Q(P_1) \to 0$, or the angle between $\operatorname{grad} Q(P_1)$ and the vector $\overline{P_1 P'}$ tends to $\pi/2$. The same holds (as $\sigma \to 0$) if P_2 is found approximately in any finite number of steps, setting it equal to P' if it is sufficiently close to P'.

Thus, in the case under consideration, the method of fastest descent Y always leads to a point arbitrarily close to one of the stationary points of Q in D, or to a point of boundary minimum. The question of how to find the least value of Q in D remains, generally speaking, open.

Computing Center
of the Academy of Sciences of the Ukrainian SSR

Received
3 X 1961

REFERENCES

1. *Modern Mathematics for Engineers*, ed. by E. F. Beckenbach, IL, 1959.
2. A. A. Feldbaum, *Computing Devices in Automatic Systems*, 1959.
3. V. V. Ivanov, DAN, **143**, No. 3 (1962).
4. A. S. Barsov, *What Is Linear Programming*, Moscow, 1959.
5. V. V. Ivanov, E. A. Karagodova, E. K. Yadrenko, *Journal of Computational Mathematics and Mathematical Physics*, No. 3 (1961).
6. J. B. Rosen, *J. Soc. Ind. and Appl. Mat.*, 8, March, No. 1 (1960).

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