



Soviet-era science, translated into English

A. Yu. Levin

1965

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196501.07242>

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

Abstract

Full Text

A. Yu. Levin

On an Algorithm for Minimizing Convex Functions

(Presented by Academician A. N. Kolmogorov on 20 XI 1964)

Below a method is proposed for the approximate minimization of a convex function of several variables, based on a certain elementary-geometric idea. The proposed algorithm of centered sections (a.c.s.) also finds application in convex programming. The a.c.s. requires a more complicated program than gradient methods, but is more efficient than they are in terms of estimating the number of operations.

1. We consider the problem of ε -minimization of a convex function of m variables $f(X)$, $X = (x_1, \dots, x_m)$, on a convex polyhedron M_0 of the Euclidean space E_m , i.e., the problem of finding a point $X^0 \in M_0$ such that

$$f(X^0) - 2\varepsilon \leq \min f(X)$$

on M_0 (the error of the formula $\min f(X)$ on $M_0 \approx f(X^0) - \varepsilon$ does not exceed ε). It is assumed that f satisfies a Lipschitz condition:

$$|f(X^1) - f(X^2)| \leq c\rho(X^1, X^2)$$

for $X^1, X^2 \in M_0$ (the constant c can often be specified from a priori considerations). It is also assumed that the following basic operation (b.o.) of our algorithm can be carried out: for a given $X^* \in M_0$, the direction $\text{grad } f(X^*)$ is determined (the case of nondifferentiability of f at the point X^* may be disregarded, since, by convexity, f is differentiable almost everywhere). Naturally, it is desirable that the number of b.o.'s carried out in the process of ε -minimization be as small as possible.

For $m = 1$ ($M_0 = [a, b]$), the b.o. consists in determining $\text{sign } f'(x^*)$. ε -minimization is carried out by the obvious "bisection": computing $\text{sign } f'((a+b)/2)$ makes it possible to pass from $[a, b]$ to one of the intervals $[a, (a+b)/2]$, $[(a+b)/2, b]$, etc. The number of b.o.'s required for ε -minimization, as is easily calculated, satisfies the inequality

$$n \leq \max\{0, \log_2 [c(b-a)/2\varepsilon]\}.$$

We shall show that the idea of this algorithm, which is trivial for $m = 1$, can in a definite way be realized also in the multidimensional case. We shall consider in detail the case $m = 2$, which fully reveals the essence of the matter.

2. Carrying out the b.o. for an interior point X^* of the polygon M_0 makes it possible to cut M_0 into two parts by a straight line passing through X^* , and to discard one of them as not containing the minimum point X_{\min} ; indeed, if $\text{grad } f(X^*) \neq 0$, then, in view of the convexity of f ,

$$(X_{\min} - X^*, \text{grad } f(X^*)) < 0$$

(if $\text{grad } f(X^*) = 0$, then $X^* = X_{\min}$). The remaining part, which we shall denote by M_1 , is again a convex polygon; carrying out the b.o. for an interior point of M_1 , by the same principle we cut M_1 into two parts, one of which is discarded, and so on.

The section points (i.e., the points at which the b.o.'s are carried out) should, of course, be chosen in such a way that the leading polygons (l.p.) M_0, M_1, M_2, \dots "contract" sufficiently rapidly. It is very expedient to choose as the section point of the l.p. M_k its center of gravity ($k = 0, 1, \dots$). This recommendation is based on the well-known geometric fact ⁽²⁾: a straight line passing through the center of gravity of a convex polygon of area s divides it into parts the area of each of which is $\geq 4/9 s$. Thus, in our algorithm, independently of the directions of the sections, the area s_k of the l.p. M_k will not exceed

$$(5/9)^k s_0 \quad (k = 1, 2, \dots).$$

It may be expected that, along with the areas, the diameters d_k of the convex sets M_k will usually also decrease at a geometric* rate; however, generally speaking, this is not necessary, since the convex sets may take a strongly elongated form; the latter, for example, certainly occurs if the set of minimum points of f forms a segment. This complication can be overcome by using the fact that a convex figure of small area is well "approximated" by a segment. Namely, let us define the algorithm of ε -minimization as follows. We continue the centered sections of the convex sets until the width of the convex set becomes $\leq 2\varepsilon/c$; suppose that this requires n_1 oracle calls. Then we replace M_{n_1} by an "approximating" segment Δ , the distance from which to any point of M_{n_1} is $\leq \varepsilon/c$; evidently, Δ can be chosen of length $d'_{n_1} \leq d_{n_1}$. We now solve the one-dimensional problem of $\varepsilon/2$ -minimization of f on Δ . For this purpose the same oracle may be used, since the direction $\text{grad } f$ determines the sign of the derivative in any direction; let the number of oracle calls spent at this stage be n_2 . The midpoint X^0 of the last leading segment (of length $\leq 2\varepsilon/c$) gives the solution of the problem. Indeed, since $X_{\min} \in M_{n_1}$, there is an $x^1 \in \Delta$ such that $\rho(X_{\min}, X^1) \leq \varepsilon/c$, whence

$$f(X_{\min}) \geq f(X^1) - \varepsilon \geq \min f(X) - \varepsilon \geq f(X^0) - 2\varepsilon.$$

Let us now estimate the total number $n = n_1 + n_2$ of oracle calls expended. Let $n_1 n_2 \neq 0$. Denoting the width of M_i by h_i and taking into account the known inequality $hd \leq 2s$ (see (2)), we obtain the system of inequalities

$$h_{n_1-1} > 2\varepsilon/c, \quad d'_{n_1-1}/2^{n_2-1} > 2\varepsilon/c, \quad d'_{n_1} \leq d_{n_1} \leq d_{n_1-1}, \quad h_{n_1-1} d_{n_1-1} \leq 2s_{n_1-1} \leq 2s_0(5/9)^{n_1-1},$$

whence

$$n = n_1 + n_2 < 1 + \log_{1.8}(s_0 c^2 \varepsilon^{-2}).$$

If $n_1 \neq 0$, $n_2 = 0$, then, taking into account the inequalities $h^2 \leq s\sqrt{3}$ (see (2)), $h_{n_1-1} > 2\varepsilon/c$, we arrive at the better estimate

$$n = n_1 \leq 1 + \log_{1.8}(1/4\sqrt{3}s_0 c^2 \varepsilon^{-2}).$$

The case $n_1 = 0$ is also possible (if the width of M_0 is so small that the problem is in fact one-dimensional); in this case

$$n \leq \max\{0, \log_2(d'_0 c \varepsilon^{-1})\}.$$

Finally we obtain

$$n \leq \max\{0, \log_2(d'_0 c \varepsilon^{-1}), 1 + \log_{1.8}(s_0 c^2 \varepsilon^{-2})\}.$$

Thus, for the number of oracle calls, as also for $m = 1$, a geometric estimate of the form $n \leq \alpha + \beta |\ln \varepsilon|$ has been obtained; here α depends only on the dimensions of M_0 and on the Lipschitz constant, while β is constant.

3. The case $m > 2$ introduces no fundamentally new features. In exactly the same way, the execution of an oracle call for an interior point X^* of the convex set makes it possible, in accordance with the inequality

$$(X_{\min} - X^*, \text{grad } f(X^*)) < 0 \quad (\text{grad } f(X^*) \neq 0),$$

to cut the convex set into two parts by an $(m-1)$ -dimensional hyperplane passing through X^* , and to discard one of these parts. The choice of the centers of gravity of the convex sets as the cutting points again leads to a geometric estimate of the number of oracle calls, since the corresponding theorem on volumes also holds for $m > 2$: an $(m-1)$ -dimensional hyperplane passing through the center of gravity of a convex m -dimensional polyhedron of volume v divides it into parts, the volume of each of which is

$$\geq [1 - (1/(m+1))]^m v.$$

The proof of this proposition, due in its essential part to B. S. Mityagin, uses, in particular, the Brunn-Minkowski symmetrization principle (1). Thus, for the volumes v_k of the convex sets M_k , one can give an estimate independent of the dimension:

$$v_k \leq (1 - e^{-1})^k v_0 \quad (k = 1, 2, \dots).$$

The complications arising because of the possibility of "flattening" of the convex set are overcome by the same dimensionality-reduction device.

4. A geometric estimate of the number of oracle calls still does not mean that an analogous estimate holds for the total number of operations. Indeed, at each step of the descent algorithm, work of a twofold nature is performed: first, the oracle call, and second, work that we shall call auxiliary—the finding and storing of the elements of the new convex set (vertices, faces, center of gravity). Suppose that at the k -th step these stages require respectively l_k and l'_k operations—

* The terms “geometric,” “geometricity” here and below indicate analogy with the rate of decrease of a geometric progression.

If the numbers l_k may naturally be considered uniformly bounded, then the situation is different for l'_k , since the number r_k of $(m - 1)$ -dimensional faces of the auxiliary set M_k may increase without bound as k grows (at least theoretically). To avoid this increase of r_k , which is unpleasant in many respects, one can modify the algorithm of centers of gravity as follows.

For each m there exists a γ_m such that in E_m every convex polyhedron of volume v can be enclosed in an m -dimensional simplex of volume $\leq \gamma_m v$ (for $m = 2$ the unimprovable value $\gamma_2 = 2$ is known; see (3)). Choose an integer

$$q > -\ln \gamma_m / \ln(1 - e^{-1}) \approx 2.18 \ln \gamma_m$$

and supplement the algorithm of centers of gravity by the following rule: as soon as r_k becomes equal to $m + q$, M_k is enclosed in the corresponding m -dimensional simplex, which is then used as the auxiliary set. Since $r_{i+1} \leq r_i + 1$, the enclosing will be carried out no more often than every q steps; by the choice of q and the estimates given above, this ensures a geometric decrease of the volumes of the auxiliary sets. Since the number of operations needed for the enclosing depends only on m and q , we arrive at a geometric estimate for the number of operations in ε -minimization.

For $m = 2$, a more expedient way to avoid the growth of r_k is the following. It follows from Helly's theorem that in a convex heptagon there exists a point such that any line passing through it cuts the heptagon into parts each having no more than 6 sides. If, for $r_k = 7$, one chooses as the cutting point not the center of gravity of M_k , but the “Helly point” of M_k (the method for finding such points is obvious), then the possibility $r_k > 7$ is thereby excluded, while at the same time the geometric character of the method is preserved.

Another modification for $m = 2$: if the centers of gravity of the vertices of the auxiliary set are chosen as cutting points (their determination requires almost no computation), then the areas of the auxiliary sets will again decrease at a geometric rate. (The proof is based on the fact that, during the cutting process, r_k increases “no less intensively” than it decreases.)

Often the evaluation oracle makes it possible to determine not only the direction of grad $f(X^*)$, but also its magnitude, as well as the value $f(X^*)$. The algorithm of centers of gravity can be modified so as to use this additional information

for a more rapid reduction of the auxiliary set. We shall not dwell on this in greater detail here.

If M_0 is not given a priori, one can apply a trial method. Suppose, for example, that it is required to minimize a function in the plane. By two evaluations, the point X_{\min} (if it exists) is, generally speaking, enclosed in a certain angle ($< \pi$), and by several further evaluations, for more or less successfully chosen points of the angle, in a bounded convex polygon.

5. The practical applicability of the algorithm of centers of gravity seems expedient only for small m (say, for $m = 2, 3$), since the amount of auxiliary work apparently grows rapidly with m . At the same time it should be noted that the auxiliary work is essentially not connected with the form of the function being minimized, and therefore the corresponding standard subroutines can be used for a broad class of problems.

An important issue is the behavior of r_k , which requires experimental verification. For $m = 2$ several experiments were carried out; in them r_k did not increase, but fluctuated within the limits 3-5. As for gradient methods—the principal “competitors” of the algorithm of centers of gravity—apparently none of them provides a geometric estimate for the number of iterations; let us also note that the use of the algorithm of centers of gravity naturally removes the problem of the step size that arises in gradient methods. Such factors as the form of $f(X)$ are also important; roughly speaking, the “worse” $f(X)$ is, the more effective the algorithm of centers of gravity is in comparison with gradient methods. Indeed, the effectiveness of the latter, as is known, depends to a large extent on the properties of the function being minimized, whereas the algorithm of centers of gravity is little sensitive in this respect. Moreover, as the laboriousness of the evaluation oracle increases, the relative weight of the auxiliary work decreases.

6. Let us briefly consider one of the main applications of the a.d.s. As is well known, the problem of convex programming consists in minimizing $f(X)$ under the conditions $g_k(X) \leq 0$, $k = 1, 2, \dots, r$, where f, g_1, \dots, g_r are convex functions of n variables (n will be assumed sufficiently large). For a broad class of problems encountered in practice, the conditions are such that

$$f(X) = \sum_{i=1}^l f^i(x_1^i, \dots, x_{n_i}^i), \quad g_k(X) = \sum_{i=1}^l g_k^i(x_1^i, \dots, x_{n_i}^i), \quad k = 1, \dots, m,$$

where $\{x_1^i, x_2^i, \dots, x_{n_i}^i\}$, $i = 1, 2, \dots, l$ ($n_1 + n_2 + \dots + n_l = n$), are comparatively small groups of variables, and each of $g_{m+1}(X), \dots, g_r(X)$ depends only on the variables of some one group. Examples of this kind may be: the problem of block linear programming, the problem considered in ⁽⁵⁾, and many others. The difficulty of the problem under consideration is due to the “essential” constraints

$g_k(X) \leq 0$, $k = 1, 2, \dots, m$, since for $m = 0$ the problem would split into a number of low-dimensional problems which, in comparison with the original one, may be regarded as elementary. We shall show that even in the presence of a small number m of “essential” constraints the question can be reduced to the solution of a series of low-dimensional problems.

The set W in E_n , defined by the conditions $g_k(X) \leq 0$, $k = m + 1, \dots, r$, is obviously convex. We must minimize $f(X)$ on W under the conditions $g_k(X) \leq 0$, $k = 1, \dots, m$. As a generalization of the Kuhn-Tucker theorem shows (see ⁽⁴⁾), under small and natural additional assumptions this problem is equivalent to the problem of seeking in $W \times E'_m$ (E'_m is the nonnegative octant of E_m) a saddle point of the function $u(X, \Lambda) = f(X) + (\Lambda, G(X))$, where $\Lambda = (\lambda_1, \dots, \lambda_m)$, $G(X) = (g_1(X), \dots, g_m(X))$. Thus, it is required to find $\max h(\Lambda)$ on E'_m , where $h(\Lambda) = \min u(X, \Lambda)$, $X \in W$, is, evidently, a concave function. For fixed $\Lambda = \Lambda^*$, the problem of minimizing $u(X, \Lambda^*)$ on W splits into low-dimensional problems and is therefore not difficult; if $X(\Lambda^*)$ is a point of minimum and $h(\Lambda)$ is differentiable at the point Λ^* , then $\text{grad } h(\Lambda^*) = G[X(\Lambda^*)]$. We have thus arrived at the problem of minimizing on E'_m the convex function of m variables $h_1(\Lambda) = -h(\Lambda)$, and, for a given Λ^* , we are able to find $\text{grad } h_1(\Lambda^*)$, i.e. we are in the conditions of applicability of the a.d.s. (for small m). In view of the unboundedness of E'_m , one should resort (if it is impossible to estimate $\lambda_1, \dots, \lambda_m$ from above on the basis of a priori considerations) to the trial method. Let us also note that $h_1(\Lambda^*)$ may turn out to be undefined (because of the unboundedness below of $u(X, \Lambda^*)$ on W). This difficulty, however, is not essential: if, for some sequence X^1, X^2, \dots such that $u(X^k, \Lambda^*) \rightarrow -\infty$, one can effectively specify a vector G^* to which some subsequence of the vectors $G(X^k)$ converges in direction, then it is not hard to show that $(\Lambda_{\min} - \Lambda^*, G^*) \geq 0$, so that $-G^*$ may formally be used as $\text{grad } h_1(\Lambda^*)$. An analogous remark also applies to the case when $\inf u(X, \Lambda^*)$ is finite but is not attained on W .

The same method is also suitable for problems containing, among the constraints, linear equalities, with the difference that for the corresponding Lagrange multipliers the nonnegativity condition is removed (see ⁽⁴⁾).

7. A scheme of random search, which is in a certain sense a probabilistic analogue of the a.d.s., was considered in ⁽⁶⁾.

Voronezh State
University

Received
11 XI 1964

REFERENCES

1. G. Minkowski, *Uspekhi Mat. Nauk*, No. 2 (1936).

2. A. M. Yaglom, V. G. Boltyanskii, *Convex Figures*, Moscow, 1951.
3. A. M. Yaglom, I. M. Yaglom, *Nonelementary Problems in an Elementary Exposition*, Moscow, 1954.
4. K. J. Arrow, L. Hurwicz, H. Uzawa, *Studies in Linear and Nonlinear Programming*, IL, 1962.
5. I. A. Bakhtin, M. A. Krasnosel' skii, A. Yu. Levin, *Computational Mathematics and Mathematical Physics*, fasc. 3, No. 2 (1963).
6. A. Yu. Levin, A. S. Shvarts, *Proceedings of the Seminar on Functional Analysis*, Voronezh, vol. 7 (1963).

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

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