Aplikace matematiky

Boris Gruber

Numerical determination of the relative minimum of a function of several variables by quadratic interpolation

Aplikace matematiky, Vol. 12 (1967), No. 2, 87–100

Persistent URL: http://dml.cz/dmlcz/103074

NUMERICAL DETERMINATION OF THE RELATIVE MINIMUM OF A FUNCTION OF SEVERAL VARIABLES BY QUADRATIC INTERPOLATION

BORIS GRUBER

(Received September 7, 1966.)

1. INTRODUCTION

Many physical and technical problems require to determine the relative minimum of a function of several variables. In this paper an algorithm is demonstrated which is suitable in the following cases:

a) The number of variables is so great (tens, or even hundreds) or the function is so complicated that the problem cannot be solved in the exact way known from classical analysis (putting the first partial derivatives equal to zero and inquiring into the corresponding quadratic form) and we have to modify the concept of the relative minimum in such a manner that accessible numerical methods are applicable. The uncertainty which arises will be smaller if not only one minimum of the function f is calculated but the whole sequence of minima of functions $f_1, \ldots, f_m$ depending upon a parameter. The probability that we have really found relative minima (in classical sense) increases with the "smoothness" of this sequence.

b) We are not interested in all minima of the function but only in certain ones which are of special importance for our problem. Besides from the (e.g. physical) character of the problem may be assumed that for each of these selected minima an "initial point" may be found which lies closer to this minimum than to the others. This situation occurs e.g. if looking for the equilibrium configuration of a system of mass points with central force acting beetween them. Doubtless a great number of these equilibrium configurations exist (the central force having a suitable form and the number of the mass points being great enough) but we are interested only in those of quite certain character, e.g. which correspond to an ideal crystal lattice or to the lattice with a vacancy, interstitial, dislocation, etc. We choose the initial configuration according to the case that is studied.

c) The algorithm is especially advantageous if the function has the form of a sum every member of which depends only upon one or a few variables. It is advantageous, too, if the calculation of the function values is much easier than the calculation of the values of the partial derivatives because these are not used.

The main effect of this method is that it enables us (even though only with a certain probability of the success) when looking for the relative minimum to change all co-ordinates of the given point simultaneously (not one by one) although only "very few" function values are calculated.

2. AUXILIARY CONCEPTS

Two points

$$X^0 = [x_1^0, \ldots, x_n^0], \tag{1}$$
$$X^* = [x_1^0 + \delta_1, \ldots, x_n^0 + \delta_n]$$

of the n-dimensional space ($n > 1$) and a positive real number δ are given. The points X°, X^* are said to be neighbouring[1]) (also X^* to be a neighbouring point of the point X° and *vice versa*) if such an integer i ($1 \leqq i \leqq n$) exists that

$$|\delta_i| = \delta\,, \quad \delta_j = 0 \quad \text{for} \quad j \neq i\,.$$

They are said to be adjoining[2]) (also X^* to be an adjoining point of the point X° and *vice versa*) if they are not identical and if

$$\text{either} \quad |\delta_i| = \delta \quad \text{or} \quad \delta_i = 0$$

holds for every i ($1 \leqq i \leqq n$). So two neighbouring points are also adjoining but the opposite is not true.

The symbol $\mathbf{K}(X^\circ, \delta, m)$ (m positive integer) denotes the set of points

$$X = [x_1, \ldots, x_n] \tag{2}$$

where every co-ordinate x_i assumes all the values

$$x_i^0 + k\delta/m \quad (k = 0, \pm 1, \ldots, \pm m)$$

independently on the other co-ordinates. Instead of $\mathbf{K}(X^\circ, \delta, 1)$ we write $\mathbf{K}(X^\circ, \delta)$. This set consists of the point X° and of all adjoining points of X° so that it contains exactly 3^n points. Further we denote by $\mathbf{R}(X^\circ, \delta)$ the set consisting of the point X° and of all neighbouring points of X°. This set has $2n + 1$ points and inclusion

$$\mathbf{R}(X^\circ, \delta) \subset \mathbf{K}(X^\circ, \delta) \tag{3}$$

holds (Fig. 1).

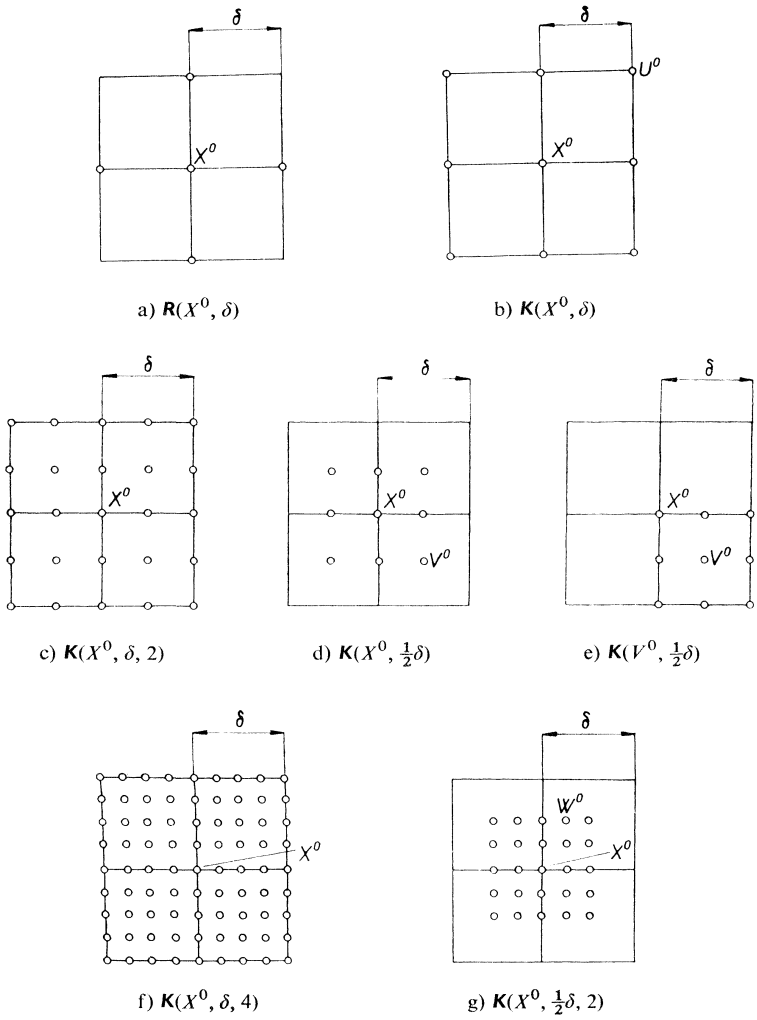
[1]) More precisely: δ-neighbouring.
[2]) More precisely: δ-adjoining.

A real-valued function f of n variables is termed to have a weak (or strong) relative minimum of the order δ at the point (1) if it is continuous in the interval

$$\langle x_1^0 - \delta, x_1^0 + \delta\rangle \times \ldots \times \langle x_n^0 - \delta, x_n^0 + \delta\rangle \tag{4}$$

and if

$$f(X^\circ) = \operatorname*{Min}_{X \in \mathbf{R}(X^0,\delta)} f(X) \quad (\text{or } f(X^\circ) = \operatorname*{Min}_{X \in \mathbf{K}(X^0,\delta)} f(X)) \tag{5}$$



a) $\mathbf{R}(X^0, \delta)$ b) $\mathbf{K}(X^0, \delta)$

c) $\mathbf{K}(X^0, \delta, 2)$ d) $\mathbf{K}(X^0, \frac{1}{2}\delta)$ e) $\mathbf{K}(V^0, \frac{1}{2}\delta)$

f) $\mathbf{K}(X^0, \delta, 4)$ g) $\mathbf{K}(X^0, \frac{1}{2}\delta, 2)$

Fig. 1 ($n = 2$).

is satisfied.[3]) To decide whether the function has a weak (or strong) relative minimum at a given point, we have to calculate $2n + 1$ (or 3^n) function values.[4])

3. QUADRATIC INTERPOLATION

Let be given the point (1), a positive real number δ and a real-valued function f of n variables continuous in the interval (4). Let us denote

$$
\begin{aligned}
(6) \qquad & y_0 = f(X^\circ)\,, \\
& \varepsilon_i = f(x_1^0, \ldots, x_{i-1}^0, x_i^0 + \delta, x_{i+1}^0, \ldots, x_n^0) - y_0\,, \\
& \bar{\varepsilon}_i = f(x_1^0, \ldots, x_{i-1}^0, x_i^0 - \delta, x_{i+1}^0, \ldots, x_n^0) - y_0 \quad (i = 1, \ldots, n)
\end{aligned}
$$

and construct the polynomial

$$
(7) \qquad P(x_1, \ldots, x_n) = y_0 + \sum_{i=1} P_i(x_i)
$$

where

$$
2\delta^2 P_i(x) = (\varepsilon_i + \bar{\varepsilon}_i)(x - x_i^0)^2 + \delta(\varepsilon_i - \bar{\varepsilon}_i)(x - x_i^0) \quad (i = 1, \ldots, n).
$$

Then equality

$$
P(X) = f(X)
$$

holds for $X \in \mathbf{R}(X^\circ, \delta)$.

In this paragraph first we are looking for the point

$$
(8) \qquad U^\circ = [u_1^0, \ldots, u_n^0] \in \mathbf{K}(X^\circ, \delta)
$$

satisfying

$$
(9) \qquad P(U^0) = \operatorname*{Min}_{X \in \mathbf{K}(X^0, \delta)} P(X)\,.\ ^5)
$$

Using denotation (2) we get

$$
(10) \qquad \operatorname*{Min}_{X \in \mathbf{K}(X^0, \delta)} P(X) = y_0 + \sum_{i=1}^{n} \operatorname*{Min}_{X \in \mathbf{K}(X^0, \delta)} P_i(x_i)
$$

having in mind that every of the functions P_i depends only upon one variable and that the set $\mathbf{K}(X^0, \delta)$ has the form of a Cartesian product. If $X \in \mathbf{K}(X^0, \delta)$ then $P_i(x_i)$ assumes one of the values $P_i(x_i^0)$, $P_i(x_i^0 + \delta)$, $P_i(x_i^0 - \delta)$, i.e. one of the values 0,

[3]) Of course, a function having a weak or a strong relative minimum at X^0 need not have a relative minimum at this point in the classical sense at all. E.g. $f(x, y) = (y - x^4)(y - x^2)$ has at the origin a strong relative minimum of the order δ for every $0 < \delta < 1$ but has not a relative minimum at this point because of $f(\delta, \delta^3) < 0$ for $0 < \delta < 1$.

[4]) More precisely, we must determine $f(X^0)$ and $2n$ (or $3^n - 1$) values $f(X) - f(X^0)$ the calculation of which may be eventually easier than the calculation of $f(X)$ (e.g. if f is a sum and every member depends only on a few variables).

[5]) Of course, more points with this property may exist. In respect of the decision of a computer the further procedure is formed in such a way that the point U^0 is determined uniquely.

ε_i, $\bar{\varepsilon}_i$. Therefore from (10) it follows

$$\text{(11)} \qquad \operatorname*{Min}_{X \in K(X^0, \delta)} P(X) = y_0 + \sum_{i=1}^{n} \operatorname{Min}(0, \varepsilon_i, \bar{\varepsilon}_i) .$$

It can be easily seen that

$$\text{(12)} \qquad \begin{aligned} &\text{for} \quad \varepsilon_i \geqq 0 , \quad \bar{\varepsilon}_i \geqq 0 \quad \text{is} \quad \operatorname{Min}(0, \varepsilon_i, \bar{\varepsilon}_i) = 0 = P_i(x_i^0) , \\ &\text{for} \quad \varepsilon_i < 0 , \quad \varepsilon_i < \bar{\varepsilon}_i \quad \text{is} \quad \operatorname{Min}(0, \varepsilon_i, \bar{\varepsilon}_i) = \varepsilon_i = P_i(x_i^0 + \delta) , \\ &\text{for} \quad \bar{\varepsilon}_i < 0 , \quad \bar{\varepsilon}_i \leqq \varepsilon_i \quad \text{is} \quad \operatorname{Min}(0, \varepsilon_i, \bar{\varepsilon}_i) = \bar{\varepsilon}_i = P_i(x_i^0 - \delta) \end{aligned}$$

(Fig. 2). So if we put

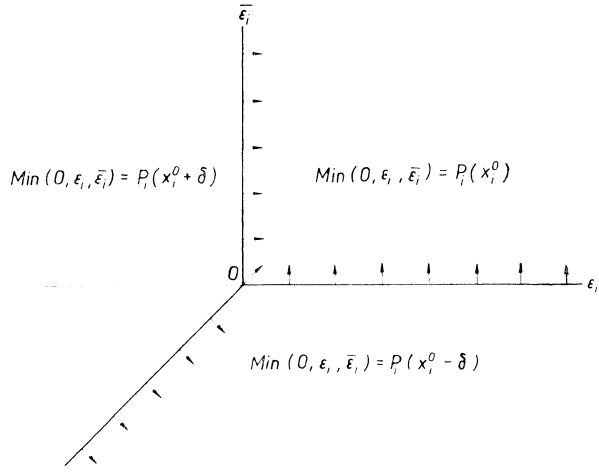


Fig. 2. (The arrows denote to what region the frontier belongs.)

$$\text{(13)} \qquad \begin{aligned} u_i^0 &= x_i^0 \qquad \text{for} \quad \varepsilon_i \geqq 0 , \quad \bar{\varepsilon}_i \geqq 0 , \\ u_i^0 &= x_i^0 + \delta \quad \text{for} \quad \varepsilon_i < 0 , \quad \varepsilon_i < \bar{\varepsilon}_i , \\ u_i^0 &= x_i^0 - \delta \quad \text{for} \quad \bar{\varepsilon}_i < 0 , \quad \bar{\varepsilon}_i \leqq \varepsilon_i \end{aligned}$$

the point U^0 is found.[6]) (Fig. 1b.) The points X^0, U^0 are adjoining if they are not identical. The last case occurs if and only if

$$\text{(14)} \qquad \varepsilon_i \geqq 0 , \quad \bar{\varepsilon}_i \geqq 0 \quad \text{for} \quad i = 1 , \ldots, n$$

[6]) For the sake of uniqueness (see footnote[5])) inequalities (12) are formed in such a manner that the regions defined by them in the plane $O\varepsilon_i\bar{\varepsilon}_i$ are disjoint. E.g. for $\varepsilon_i = 0$, $\bar{\varepsilon}_i \geqq 0$ is $\operatorname{Min}(0, \varepsilon_i, \bar{\varepsilon}_i) = 0 = \varepsilon_i = P_i(x_i^0) = P_i(x_i^0 + \delta)$ and it may be put with the same right $u_i^0 = x_i^0$ as well as $u_i^0 = x_i^0 + \delta$. In this case we prefer not to change the co-ordinate.

holds, i.e. if and only if the function f has a weak relative minimum of the order δ at the point X^0. Then we get according to (11)

$$\operatorname*{Min}_{X\in\mathbf{K}(X^0,\delta)} P(X) = y_0 = f(X^0) = P(X^0)$$

which means (see (5)) that the polynomial P has a strong relative minimum of the order δ at the point X^0.

Let us notice that at the point U^0 the polynomial P assumes its smallest value on a set containing 3^n points. But to determine the point U^0 it was necessary to know only $2n + 1$ values of the polynomial P.

Further it will be shown how to find the point

$$V^0 = [v_1^0, \ldots, v_n^0] \in \mathbf{K}(X^0, \delta, 2) \tag{15}$$

with the property

$$P(V^0) = \operatorname*{Min}_{X\in\mathbf{K}(X^0,\delta,2)} P(X)$$

assuming that f has a weak relative minimum of the order δ at the point X^0 [7]).

Similarly as in (10) we can write (using denotation (2))

$$\operatorname*{Min}_{X\in\mathbf{K}(X^0,\delta,2)} P(X) = y_0 + \sum_{i=1}^{n} \operatorname*{Min}_{X\in\mathbf{K}(X^0,\delta,2)} P_i(x_i)\,. \tag{16}$$

For $X \in \mathbf{K}(X^0, \delta, 2)$ the value $P_i(x_i)$ equals to one of the numbers

$$P_i(x_i^0),\ P_i(x_i^0 + \tfrac{1}{2}\delta)\,,\ P_i(x_i^0 - \tfrac{1}{2}\delta),\ P_i(x_i^0 + \delta)\,,\ P_i(x_i^0 - \delta)\,,$$

i.e. to one of the numbers

$$0,\ \tfrac{1}{8}(3\varepsilon_i - \bar{\varepsilon}_i),\ \tfrac{1}{8}(3\bar{\varepsilon}_i - \varepsilon_i),\ \varepsilon_i,\ \bar{\varepsilon}_i.$$

Denoting the smallest of them by a_i we get from (16)

$$\operatorname*{Min}_{X\in\mathbf{K}(X^0,\delta,2)} P(X) = y_0 + \sum_{i=1}^{n} a_i\,.$$

It can be easily recognized (remember that according to our assumption (14) holds) that

$$\begin{aligned}
&\text{for}\quad \tfrac{1}{3}\varepsilon_i \leqq \bar{\varepsilon}_i \leqq 3\varepsilon_i \quad \text{is}\quad a_i = P_i(x_i^0)\,,\\
&\text{for}\quad 3\varepsilon_i < \bar{\varepsilon}_i \qquad\quad \text{is}\quad a_i = P_i(x_i^0 + \tfrac{1}{2}\delta)\,,\\
&\text{for}\quad 3\bar{\varepsilon}_i < \varepsilon_i \qquad\quad \text{is}\quad a_i = P_i(x_i^0 - \tfrac{1}{2}\delta)
\end{aligned}$$

[7]) The procedure is formulated again in such a manner that V^0 is determined uniquely.

(Fig. 3). Therefore it is enough to put

$$
\begin{aligned}
(17) \qquad & v_i^0 = x_i^0 && \text{for} \quad \tfrac{1}{3}\varepsilon_i \leqq \bar{\varepsilon}_i \leqq 3\varepsilon_i\,, \\
& v_i^0 = x_i^0 + \tfrac{1}{2}\delta && \text{for} \quad 3\varepsilon_i < \bar{\varepsilon}_i\,, \\
& v_i^0 = x_i^0 - \tfrac{1}{2}\delta && \text{for} \quad 3\bar{\varepsilon}_i < \varepsilon_i\,.
\end{aligned}
$$

Because of $V^0 \in \boldsymbol{K}(X^0, \frac{1}{2}\delta)$ the polynomial P has a strong relative minimum of the order $\frac{1}{2}\delta$ at the point V^0 (Fig. 1c, d, e).

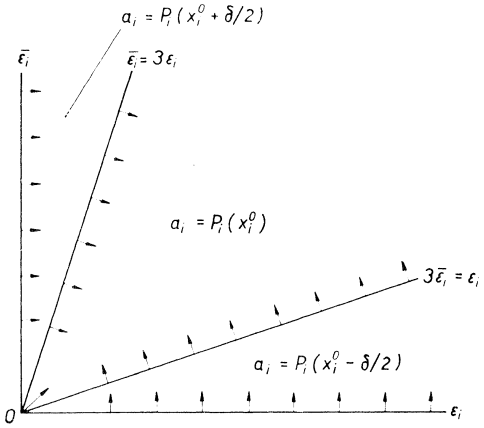


Fig. 3.

In a similar way the following may be stated: If

$$
\begin{aligned}
(18) \qquad & w_i^0 = x_i^0 && \text{for} \quad 9\varepsilon_i \leqq 15\bar{\varepsilon}_i \leqq 25\varepsilon_i\,, \\
& w_i^0 = x_i^0 + \tfrac{1}{4}\delta && \text{for} \quad 5\varepsilon_i < 3\bar{\varepsilon}_i < 21\varepsilon_i\,, \\
& w_i^0 = x_i^0 - \tfrac{1}{4}\delta && \text{for} \quad 5\bar{\varepsilon}_i < 3\varepsilon_i < 21\bar{\varepsilon}_i\,, \\
& w_i^0 = x_i^0 + \tfrac{1}{2}\delta && \text{for} \quad 7\varepsilon_i \leqq \bar{\varepsilon}_i\,, \quad \bar{\varepsilon}_i > 0\,, \\
& w_i^0 = x_i^0 - \tfrac{1}{2}\delta && \text{for} \quad 7\bar{\varepsilon}_i \leqq \varepsilon_i\,, \quad \varepsilon_i > 0\,,
\end{aligned}
$$

$$
(19) \qquad W^0 = [w_1^0, \ldots, w_n^0]
$$

is put then equality

$$
P(W^0) = \operatorname*{Min}_{X \in \boldsymbol{K}(X^0, \delta, 4)} P(X)
$$

holds assuming that f has a weak relative minimum of the order δ at X^0. For $W^0 \in \boldsymbol{K}(X^0, \frac{1}{2}\delta, 2)$, the polynominal P has at W^0 a strong relative minimum of the orders $\frac{1}{4}\delta$ and $\frac{1}{2}\delta$ (Fig. 1f, g).

Other cases of this kind are analogous.

4. SEARCH FOR THE RELATIVE MINIMUM

The initial point X^0, a positive real number δ and a real-valued function f of n variables are given. The function f is assumed to be continuous in a sufficiently great neighbourhood of the point X^0. Naturally we are led to one of these two procedures:

(a) We calculate the values $f(X)$ for $X \in \mathbf{R}(X^0, \delta)$ to know whether the function f has a weak relative minimum of the order δ at X^0. If this is not the case we find the point

$$Y^0 \in \mathbf{R}(X^0, \delta) \tag{20}$$

satisfying

$$f(Y^0) = \operatorname*{Min}_{X \in \mathbf{R}(X^0, \delta)} f(X)\ ^{8}) \tag{21}$$

and come back to the beginning having put Y^0 instead of X^0. If f has a weak relative minimum of the order δ at X^0, then we either finish the calculation when δ is for our purposes sufficiently small or return to the beginning having replaced the number δ by a suitable smaller positive number, e.g. $\frac{1}{2}\delta$.

(b) We proceed likewise as in the case (a), only the set $\mathbf{R}(X^0, \delta)$, the point Y^0 and the concept "weak relative minimum" are substituted by the set $\mathbf{K}(X^0, \delta)$, the point Z^0 and the concept "strong relative minimum", respectively.

From (3) it follows

$$f(Z^0) \leqq f(Y^0) \tag{22}$$

so that the procedure (b) in general stands for the shorter way to the relative minimum.[9]) Of course, to determine the point Z^0 it is necessary to calculate 3^n function values (in contrast to $2n + 1$ function values for determination of the point Y^0) and this may be at greater n (tens) even for a computer unfeasible.

Let us therefore approximate the function f by the polynomial P as it was demonstrated in the paragraph 3 and let us determine the point U^0 according to (13) and (8) (denotation (1) being assumed). For this $2n + 1$ values of the function f are enough. If this approximation is good we may expect that not only (9) but also

$$f(U^0) = \operatorname*{Min}_{X \in \mathbf{K}(X^0, \delta)} f(X) = f(Z^0)$$

holds, i.e.

$$f(U^0) \leqq f(Y^0) \tag{23}$$

[8]) Of course, more points with this property may exist.

[9]) The points X^0, Y^0 are neighbouring so that they differ only in one co-ordinate but X^0, Z^0 are adjoining and may differ in any number of co-ordinates.

is satisfied, too (see (22)). If here the sign $<$ occurs it means that we came closer to the value of the relative minimum than in the procedure (a) having done the same work[10]). That is the meaning of the approximation. If

$$f(U^0) < f(Y^0)$$

does not hold we have to content ourselves with the point Y^0. Let us mention that the point U^0 may differ from X^0 in any number of co-ordinates.

The approximation by means of the polynomial P may be also used in that case when we come to a point X^0 at which the function f has a weak relative minimum of the order δ this number being too great for the purpose of our calculation. Then we construct the point V^0 or W^0 as it was demonstrated in the paragraph 3.

So the search for the relative minimum by means of the quadratic interpolation may be done in the following way:

We calculate the $2n + 1$ values $f(X)$ for $X \in \boldsymbol{R}(X^0, \delta)$ and determine the numbers (6).

(α) If (14) does not hold the points Y^0 and U^0 satisfying (20), (21) and (13), (8) have to be found. When

$$f(U^0) < f(Y^0) \tag{24}$$

is true we return to the beginning having replaced the point X^0 by U^0; in the opposite case X^0 is replaced by Y^0.

(β) If (14) holds we either finish the calculation (when δ is sufficiently small) or construct the point V^0 or W^0 according to (17), (15) or (18), (19), respectively.[11]) When

$$f(V^0) < f(X^0) \quad \text{or} \quad f(W^0) < f(X^0) \tag{25}$$

is correct we come back to the beginning writing V^0, $\frac{1}{2}\delta$ or W^0, $\frac{1}{4}\delta$ instead of X^0, δ. If (25) does not hold we return to the beginning again having replaced δ by a smaller value.

At the end let us mention that the main idea of this method does not consist in the fact that P is a polynomial of the (at most) 2nd degree but that P has the form (7), where every of the functions P_i depends only upon the i-the variable and is uniquely determined by its function values at three points.

[10]) We have in mind the calculation of the function values of f.

[11]) We choose W^0 when having reasons for the assumption that the approximation of the function f by the polynomial P might be exceedingly good.

5. EXAMPLE

We will demonstrate the calculation of the equilibrium configuration and critical shear stress on the bilinear model of a crystal lattice. This lattice will be on the one hand "ideal", on the other hand with a "dislocation". The calculation was made by the computer LGP 30 in the Centrum of Numerical Mathematics of the Charles University.

The bilinear model consists of two parallel rows of "atoms". The distance between the rows is b and every row contains m atoms. Every atom may move only in the direction of its row. The whole configuration is characterized by the quantities

$$x_1, x_2, \ldots, x_{2m-1},$$

the meaning of which may be seen from the Fig. 4.

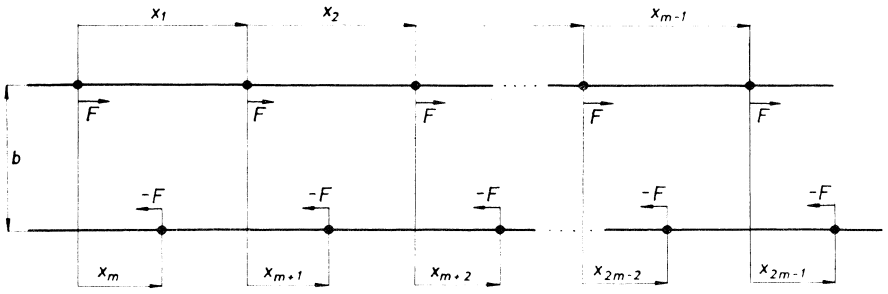


Fig. 4.

Central force is assumed to act between every couple of atoms. This force is determined by the Morse potential

$$V(r) = D(\mathrm{e}^{-2\alpha(r-r_0)} - 2\,\mathrm{e}^{-\alpha(r-r_0)});$$

here D and α are physical constants, r is the distance between the atoms and r_0 the equilibrium distance. Further we demand that a constant external force F acts in the positive direction on every atom of the upper row and in the negative direction on every atom of the lower row.

Introducing the dimensionless quantities

$$x = r/r_0\,, \quad \gamma = \alpha r_0\,, \quad f = Fr_0/D$$
$$v(x) = V(xr_0)/D = \mathrm{e}^{-2\gamma(x-1)} - 2\,\mathrm{e}^{-\gamma(x-1)}$$

the total dimensionless "potential energy" w of this system is given by the following

formula

$$w(x_1, \ldots, x_{2m-1}) = \sum_{1 \leqq i \leqq j \leqq m-1} \{v(x_i + \ldots + x_j) + v(-x_{i+m-1} + x_i + \ldots + x_j + x_{j+m}) + v(\sqrt{((-x_{i+m-1} + x_i + \ldots + x_j)^2 + b^2)}) + v(\sqrt{((x_i + \ldots + x_j + x_{j+m})^2 + b^2)})\} + \sum_{i=m}^{2m-1} \{fx_i + v(\sqrt{(x_i^2 + b^2)})\} .$$

The values of γ are between 3 and 4·5;[12]) we take $\gamma = 4$.

Respecting the capability of the computer that was used and the orientation character of the calculation we choose the simplest model with $m = 3$. Putting $b = \frac{1}{2}\sqrt{3}$ (the altitude of the equilateral triangle the side of which is equal to 1) it is natural to take the following values (if the external force does not act) as the initial configuration:

Ideal lattice	Lattice with dislocation
$x_1 = 1$	$x_1 = 1$
$x_2 = 1$	$x_2 = 1$
$x_3 = 0{\cdot}5$	$x_3 = -0{\cdot}5$
$x_4 = 0{\cdot}5$	$x_4 = 0{\cdot}5$
$x_5 = 0{\cdot}5$	$x_4 = 0{\cdot}5$

Calculating till 3 decimals we get this equilibrium configuration:

Ideal lattice	Lattice with dislocation
$x_1 = 0{\cdot}987$	$x_1 = 0{\cdot}960$
$x_2 = 0{\cdot}981$	$x_2 = 0{\cdot}975$
$x_3 = 0{\cdot}486$	$x_3 = -0{\cdot}386$
$x_4 = 0{\cdot}480$	$x_4 = 0{\cdot}431$
$x_5 = 0{\cdot}486$	$x_5 = 0{\cdot}454$

Now we let the force f gradually increase and look for the corresponding equilibrium configurations. The greatest value of f (using 3 decimals) for which in the model of ideal lattice an equilibrium configuration exists is $f = 1{\cdot}970$. The configuration reads

$$x_1 = 0{\cdot}965$$
$$x_2 = 0{\cdot}967$$
$$x_3 = 0{\cdot}232$$
$$x_4 = 0{\cdot}234$$
$$x_5 = 0{\cdot}232$$

[12]) See e.g. *L. A. Girifalco, V. G. Weizer*: Application of the Morse potential function to cubic metals, Phys. Rev. *114* (1959), 687.

For $f = 1{\cdot}971$ one doesn't succeed in finding the equilibrium configuration. Therefore we hold this value for the "critical shear stress" of our model because the "plastic deformation" just starts. For the model of the lattice with dislocation this critical shear stress is $f = 1{\cdot}222$. (If $f = 1{\cdot}221$ we get as the equilibrium configuration

$$\begin{aligned} x_1 &= 0{\cdot}958 \\ x_2 &= 0{\cdot}958 \\ x_3 &= -0{\cdot}573 \\ x_4 &= 0{\cdot}242 \\ x_5 &= 0{\cdot}271.) \end{aligned}$$

This value is only 1·61times smaller than the value for the ideal lattice in contradiction to the discrepances of about 4 orders obtained by the experiments on real materials. This is evidently caused by the "rigidity" of our model the atoms of which may move only in one direction.

I am indebted to Dr. J. Bílý from the Centrum of Numerical Mathematics of the Charles University for his valuable comments to my work. The example was calculated by the assistance of several workers of this Centrum. Special thanks are due to Mr. J. Kofroň and Mr. P. Doktor.

Výtah

NUMERICKÉ STANOVENÍ LOKÁLNÍHO MINIMA FUNKCE VÍCE PROMĚNNÝCH KVADRATICKOU INTERPOLACÍ

Boris Gruber

Je uveden algoritmus pro stanovení polohy lokálního minima funkce více proměnných, který spočívá na kvadratické interpolaci a je vhodný v těchto případech:

a) Počet proměnných je tak velký (desítky, popř. sta), event. funkce je tak složitá, že není možno řešit úlohu exaktně způsobem známým z klasické analýzy.

b) Nezajímáme se o všechna lokální minima dané funkce, nýbrž jen o jistá z nich, která mají pro nás zvláštní význam. Přitom z charakteru úlohy (např. fyzikálního) se dá předpokládat, že ke každému z těchto vybraných minim můžeme udat „výchozí bod", který leží k tomuto minimu blíže nežli k event. ostatním lokálním minimům.

c) Metoda je výhodná zvlášť tehdy (není to však podmínkou), má-li daná funkce tvar součtu, jehož každý člen závisí jen na jedné nebo několika málo proměnných. Rovněž se hodí v tom případě, když výpočet funkčních hodnot je značně jednodušší než výpočet hodnot parciálních derivací.

Algoritmus zní takto:

Je dán výchozí bod (1), kladné číslo δ a spojitá funkce f n proměnných splňující uvedené předpoklady. Označíme $\boldsymbol{R}(X^0, \delta)$ množinu skládající se z bodu X^0 a ze všech bodů tvaru

$$[x_1, \ldots, x_{i-1}, x_i \pm \delta, x_{i+1}, \ldots, x_n] \quad (i = 1, \ldots, n)$$

a vypočítáme čísla (6).

(α) Jestliže neplatí (14), určíme bod Y^0 splňující (20), (21) a bod U^0 podle (13), (8). Platí-li (24), vrátíme se na začátek nahradivše bod X^0 bodem U^0; neplatí-li (24), vrátíme se tam nahradivše X^0 bodem Y^0.

(β) Jestliže je splněno (14), pak buď výpočet skončíme (je-li pro naše účely číslo δ dostatečně malé), nebo sestrojíme bod V^0, resp. W^0 podle (17), (15), resp. (18), (19). (Pro bod W^0 se rozhodneme, máme-li důvody k předpokladu, že funkci f lze v okolí bodu X^0 zvlášť dobře aproximovat polynomem nejvýše druhého stupně.) Platí-li (25), vrátíme se na začátek píšíce V^0, $\frac{1}{2}\delta$, resp. W^0, $\frac{1}{4}\delta$ místo X^0, δ. Neplatí-li (25), vrátíme se na začátek nahradivše δ nějakou menší hodnotou.

Uvedená metoda je ilustrována příkladem výpočtu rovnovážné konfigurace bilineárního modelu atomové mřížky.

Резюме

ЧИСЛЕННОЕ УСТАНОВЛЕНИЕ ЛОКАЛЬНОГО МИНИМУМА ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ ПУТЕМ КВАДРАТИЧЕСКОЙ ИНТЕРПОЛЯЦИИ

БОРИС ГРУБЕР (Boris Gruber)

Приведен алгорифм для установления положения локального минимума функции многих переменных, основанный на квадратической интерполяции и пригодимый в следующих случаях:

a) Число переменных настолько велико (десятки, возможно сотни) или функция является настолько сложной, что задачу нельзя решить точно методами, известными из классического анализа.

b) Не интересуют нас все локальные минимумы данной функции, но только некоторые из них, имеющие особое значение. При этом по характеру задачи (напр. физическому) можно предполагать, что к каждому из этих выбранных минимумов можно определить „исходную точку", находящуюся ближе этого минимума чем других возм. локальных минимумов.

c) Метод выгоден особенно тогда (однако, это не обязательное условие), когда данная функция имеет вид суммы, каждый член которой зависит только от одной или от мало переменных. Также этот алгорифм пригоден в том случае, когда вычисление значений функции гораздо проще, чем вычисление значений частных производных.

Алгорифм состоит в следующем:

Задана исходная точка (1), положительное число δ и непрерывная функция f от n переменных удовлетворяющая предположениям. Обозначим посредством $\boldsymbol{R}(X^0, \delta)$ множество состоящее из точки X^0 и всех точек

$$[x_1, \ldots, x_{i-1}, x_i \pm \delta, x_{i+1}, \ldots, x_n] \quad (i = 1, \ldots, n)$$

и подсчитаем числа (6).

(α) Если не имеет место (14), находится точка Y^0 с помощью (20), (21) и точка U^0 с помощью (13), (8). Если имеет место (24), процесс повторятся с начала, но вместо точки X^0 берется U^0; если (24) не имеет место, вместо X^0 берется Y^0.

(β) Допустим (14) имеет место. Тогда или процесс может быть оборван (когда число δ достаточно мало для наших целей), или может быть построена точка V^0, соотв. W^0 следуя (17), (15), соотв. (18), (19). (Точка W^0 строится тогда, есть-ли у нас основания считать, что функцию f можно в окрестности X^0 особенно хорошо приблизить полиномом не более чем второй степени.) Если верно (25), процесс повторяется с начала, причем вместо X^0 и δ подставляется V^0, $\frac{1}{2}\delta$, соотв. W^0, $\frac{1}{4}\delta$. Если (25) не верно, повторяем процесс с начала, подставляя вместо δ некоторое меньшее число.

Применение метода иллюстрировано на примере вычислений положения равновесия билинейной модели атомной решетки.

Author's address: Dr. *Boris Gruber*, C.Sc., matematicko-fyzikální fakulta Karlovy university, Praha 2, Ke Karlovu 3.