DUAL DIRECTION METHODS FOR FUNCTION MINIMIZATION

Yu.M. Danilin
Institute of Cybernetics Ukrainan
Academy of Sciences, USSR

1°. Let

$$f(x) = \frac{1}{2}(Ax,x) + (b,x) + c,$$

$x \in E^n$, A be a symmetric matrix, and $(Ax,x)>0$, $\forall x \neq 0$, b - n - dimensional vector, c - scalar. Gradient of this function $f'(x) = A(x - x_*)$. Here and further x_* is a point of minimum.

Minimization of the assumed function is equivalent to the solution of the linear equation system

$$(f'(x_i), z_i) = (e_i, x_i - x_*), \qquad 0 \le i \le n-1 \qquad (1)$$

or the system of following equations

$$f(x_i) = f(x_*) + 1/2\,(f'(x_i), x_i - x_*), \qquad 0 \le i \le n \qquad (2)$$

where x_i are arbitrary points; $z_0, \dots, z_{n-1}$ - an arbitrary system of lineary independent vectors; $e_i = f'(x_i) - f'(x_i - z_i) = A z_i$.

The systems (1) and (2) may be written in the form

$$(e_i, x_*) = z^{i+1}, \qquad 0 \le i \le n-1 \qquad (3)$$

where

$$z^{i+1} = (e_i, x_i) - (f'(x_i), z_i) \qquad (4)$$

or (for the system (2))

$$z^{i+1} = (f'(x_i), x_i) - (f'(y_i), y_i) - 2[f(x_i) - f(y_i)] \qquad (5)$$

In the last case $z_i = x_i - y_i$; $e_i = f'(x_i) - f'(y_i)$; points y_i are chosen arbitrarily so as to provide the linear independence of the vectors $z_0, \dots, z_{n-1}$.

The finding of the point x_*, that satisfies the system (3), can be treated as some iterative process of the finding of points

$$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$$

satisfying the relations

$$(\bar{x}_{k+1}, e_i) = z^{i+1}, \qquad 0 \le i \le k \le n-1. \qquad (6)$$

From the comparison of formulas (4) and (6) it follows that $\bar{x}_n = x_*$.

The finding of the point $\bar{x}_{k+1}$ is realized with recurrent formulas. Let $x_0=\bar{x}_0$ be an arbitrary point, and

$$\bar{x}_{k+1} = \bar{x}_k + p_k , \qquad k = 0, 1, \dots, n-1. \tag{7}$$

So from the relations (6) it follows that vector p_k must satisfy such conditions:

$$(p_k, e_i) = 0, \qquad 0 \le i \le k-1 \tag{8}$$

$$(p_k, e_k) = z^{k+1} - (\bar{x}_k, e_k) \tag{9}$$

The last condition may be altered to the form:

$$(p_k, e_k) = -(f'(\bar{x}_k), \tau_k). \tag{10}$$

The choice of vector p_k, satisfying the conditions (8) and (9), permits a wide range of possibilites at $k \le n-2$. Thus, the formula (7) defines actually a wide class of the quadratic algorithms of minimization, in which second derivatives do not participate.

Here we dwell upon the study of the dual direction methods.

If the system of vectors $s_{k+1,i}, 0 \le i \le k$, is made dual (biortogonal) to the system $e_0, \dots, e_k$, it may be assumed that

$$p_k = \alpha_k s_{k+1,k} \tag{11}$$

where

$$\alpha_k = z^{k+1} - (\bar{x}_k, e_k) = -(f'(\bar{x}_k), \tau_k).$$

So, it appears that

$$\bar{x}_{k+1} = \bar{x}_k + [z^{k+1} - (\bar{x}_k, e_k)]\, s_{k+1,k} \tag{12}$$

Choosing p_k in form (11), the following equality $(f'(\bar{x}_k), \tau_k) = (f'(x_0), \tau_k)$ holds. Consequently

$$\bar{x}_n = x_* = x_0 - \sum_{i=0}^{n-1} (f'(x_0), \tau_i)\, s_{n,i} \tag{13}$$

2°. Let us dwell on the task of minimization of the non-quadratic functions. Let's assume $f(x)$ as a twice continously differentiable function, and $mJ \leq f''(x) \leq MJ$, $m>0$.

Now we consider the algorithms in the basis of which the formulas (12) and (13) lie.

The method, founded on the application of the formula (13), in its main facilities consists in the following steps. We make the sequence of points

$$x_{k+1} = x_k + \alpha_k \bar{p}_k \,, \qquad k = n, n+1, \dots \qquad (14)$$

wherein the vector

$$\bar{p}_k = -\sum_{i=0}^{n-1} \left(f'(x_k), z_{k-i}\right) s_{k+1,k-i} \qquad (15)$$

and as the multiplier α_k, defining the stepsize, we choose the greatest value of the parameter $0 \leq |\alpha| \leq 1$/ i.e. the value obtained by means of fracturing α/, that satisfies unequality

$$f(x_k+\alpha\bar{p}_k) - f(x_k) \leq \varepsilon\alpha\left(f'(x_k), \bar{p}_k\right), \qquad 0<\varepsilon<\frac{1}{2} \qquad (16)$$

At the initial stage of this process some iterations can be realized by the gradient method (that is $\bar{p}_k = -f'(x_k)$) - in the case, when the choice of $\bar{p}_k$ in the form of (15) has resulted in $\left(f'(x_k), \bar{p}_k\right) = 0$.

The vector z_k in the Eq.(15) is determined as $z_k = x_k - y_k$, where the point y_k is found arbitrarily so as to provide the carrying out of the following conditions: a) the vector system $z_k, \dots, z_{k-n+1}$ must be linearly independent at any k ; b) $\|z_k\| \to 0$ at $k \to \infty$.

The vectors $s_{k+1,k-i}$, $0 \leq i \leq n-1$, form the basis, dual to the basis $e_k, \dots, e_{k-n+1}$. The finding of the vectors is carried out with the recurrent formulas

$$s_{k+1,k} = \frac{s_{k,k-n}}{(s_{k,k-n}, e_k)}$$

$$s_{k+1,k-j} = s_{k,k-j} - \left(s_{k,k-j}, e_k\right) s_{k+1,k} \,, \qquad j = 1, \dots, n-1 .$$

If at some value of k it occurs that $\left(s_{k,k-n}, e_k\right) = 0$ / it can take place only at the initial stage of the process / we shall choose a new vector z_k and find the corresponding vector e_k. So we can always achieve the fulfilment of the condition $(S_{k,k-n}, e_k) \neq 0$.

Theorem 1. With above-mentioned assumptions for the sequence $\{x_k\}$, determined by formulas $\{(14),(15)\}$, the following statements $f(x_{k+1}) \le f(x_k)$, $\|x_k - x_*\| \to 0$ are true independently of the choice of initial point x_0; what's more, the speed of convergence is superlinear:

$$\|x_{N+\ell} - x_*\| \le C \lambda_N \lambda_{N+1} \dots \lambda_{N+\ell} \tag{17}$$

Here $\lambda_{N+\ell} < 1$ at any $\ell \ge 0$, $\lambda_i \to 0$ if $i \to \infty$.

If the matrix $f''(x)$ satisfies the Lip. condition, the estimate (17) is defined more precisely as follows:

$$\|x_{k+1} - x_*\| \le C \|x_{k-n+1} - x_*\| \, \|x_k - x_*\| .$$

The other dual direction algorithms may be found in the following way. The sequence $\{x_k\}$ is formed according to the formula (14), wherein vector $\bar{p}_k = \bar{x}_{k+1} - x_k$ and the point $\bar{x}_{k+1}$ is determined as shown in (12). The parameter α_k is chosen from the condition (16). The different algorithms correspond with the choice of the value of z^{k+1} by the formula (4) or (5). For the sake of brevity we name such algorithms as the methods $\{(14),(4)\}$, $\{(14),(5)\}$. In these algorithms the gradient steps are also possible at the initial stage of the process.

Theorem 2. If the function $f(x)$ and the vector system $z_k, \dots, z_{k-n+1}$ satisfy the above formulated requirements, the sequence $\{x_{\xi n}\}$, $\xi = 0, 1, \dots$, found by method $\{(14), (4)\}$, converges to the point x_* with the superlinear speed independently of the choice of the initial point x_0. And $f(x_{k+1}) \le f(x_k)$. The analogous properties are peculiar for the method $\{(14), (5)\}$, with the value of the vector z_k constrained as follows:

$$\|z_k\| \le \min \{ \|x_k - x_{k-1}\|, \|f'(x_k)\| \}$$

The properties of the method $\{(14),(15)\}$ (which was described in the form different from one given here) were studied in the papers [1 - 2] , where the above cited values of the speed of convergence and some other values were derived. The properties of the methods { (14),(4) } and {(14),(5)} were studied in detail in paper [3] . Note that the study of the algorithms alike in their meaning was accomplished in papers [4] and [5].

In conclusion we also notice that the methodology used in item 1° permits creation of conjugate direction methods as well. The theory of convergence of these methods was developed in the paper [6] .

References

1. Yu.M.Danilin, B.N.Pshenichniy, On the Methods of Minimization with the Accelerated Convergence(Russian). Journal "Vychisl. Mat. i Mat. Phys.", 10, N 6, 1970.
2. Yu.M.Danilin, B.N.Pshenichniy, The Estimates for the Speed of Convergence of a Certain Class of Minimization Algorithms (Russian). Journal "Dokl. Akad. Nauk SSSR", t. 213, N 2, 1973.
3. Yu.M.Danilin, On a Certain Class of the Minimization Algorithms with Superlinear Convergence (Russian). Journal "Vychisl. Mat. i Mat. Phys.", 14, N 3, 1974.
4. D.H.Jacobson, Oksman W., An algorithm that minimizes homogeneous functions on N variables in N+2 iterations and rapidly minimizes general functions. J.Math. Anal. Applik., v.38, 1972.
5. H.Y.Huang, Method of Dual matrices for function minimization. JOTA, v. 13, N 5, 1974.
6. Yu.M.Danilin, The Methods of Conjugate Directions for the Minimization Problems Solution (Russian). Journal "Kibernetika", N 5, 1971.