V. DUMITRU

Department of Economic Cybernetics
Academy of Economic Studies

On the Davidon-Fletcher-Powell Algorithm for Maximization

The Davidon-Fletcher-Powell algorithm, as described in [1] and [2], tries to find a local minimum of a function $f(x) \in C^2$, $x \in R^n$, through a number of Newton type approximations:

$$x^{k+1} = x^k - H^{-1}(x^k)\,\nabla f(x^k)$$

$$H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right) \tag{1}$$

where, however, matrix H^{-1} is not effectively calculated in every point x^k but is approximated only with the aid of the first order derivatives through

$$H_k^{-1} = H_{k-1}^{-1} - \frac{H_{k-1}^{-1} v_k v_k^T H_{k-1}^{-1}}{v_k^T H_{k-1}^{-1} v_k} + \frac{s_k s_k^T}{s_k^T v_k} \tag{2}$$

where

$$s_k = x^k - x^{k-1} \tag{3}$$

$$v_k = \bar{v} f(x^k) - \nabla f(x^{k-1}) \tag{4}$$

so that formula (1) becomes

$$x^{k+1} = x^k + \lambda_k d^k \tag{5}$$

where

$$d^k = -H_k^{-1} \nabla f(x^k) \tag{6}$$

and scaler λ_k provides the minimum on line d^k, that is

$$f(x^k + \lambda_k d^k) = \min_{\lambda > 0} f(x^k + \lambda d^k). \tag{7}$$

As H_0^{-1} one may take unit matrix I or any other symmetric and positive definite approximation. In this case, it is pointed out[2] that H_k^{-1} remains positive definite, a condition which is essential for the stability of the process.

Our computer programmes have used

$$H_0^{-1} = I. \tag{8}$$

If we have to solve a maximization problem for $f(x)$, obviously, we shall minimize $-f(x)$.

There are, however, situations in which things are not quite simple, from the practical point of view. For example, it is possible that the subroutine achieving the unidimensional optimization (7) is designed to provide only the maximum and thus, its modification for solving problems for minimum should require interventions which are either complicated — even impossible if it exists only as an object programme—or undesirable because our all system of optimization programmes is built for maximum.

Thus, the problem arises how to use in these conditions the D-F-P algorithm for the latter to solve a problem of maximum.

It might seem, and it in this way that the author has worked so far, that the only thing to change is direction d^k so that (6) becomes

$$d^k = H_k^{-1} \nabla f(x^k) \tag{6'}$$

and λ_k should determine the maximum on line d^k

$$f(x^k + \lambda_k d^k) = \max_{\lambda > 0} f(x^k + \lambda d^k) \tag{7'}$$

formulas (2), (5) and (8) remaining unchanged, because if d^k, calculated with (6) was a good direction for minimum, d^k given by (6') must be a good direction for maximum.

The reasoning is correct as long as H_k^{-1} remains, under these conditions, positive definite because in this case $(d^k)^T \nabla f(x^k) = \nabla^T f(x^k)\, H_k^{-1} \nabla f(x^k) > 0$, this is not understood by itself and experimental tests show that H_k^{-1} becomes negative definite, setting out from $H_0^{-1} = I$, even at the n-th iteration, in the process for searching a maximum. If it is so, when H_k^{-1} becomes negative definite λ_k should be allowed to become negative and in any case an attempt to localize the optimum on direction d^k, with a positive step, would be a failure which uselessly increases the number of estimations of the function.

The alternative we suggest is as in (5) λ_k be established in such a way as (7') be satisfied and formulas (2), (3), (4) and (6) remaining unchanged and

$$H_0^{-1} = -I \tag{8'}$$

provided that H_k^{-1} remains negative definite for any k, if approximations are calculated with formula (2). We have the following theorem:

If H_{k-1}^{-1} is symmetric and negative definite, thus H_k^{-1}, calculated through the Davidon-Fletcher-Powell formula (2) in conditions (3), (4), (5), (6) and (7') is symmetric and negative definite.

The proof we give follows the same scheme, known to the author, from the course delivered in the 1969—1970 academic year at Columbia University

by Philip Wolfe, as for the case in which it remained positive definite, when the algorithm for minimum was used.

The fact that H_k^{-1} is symmetric if H_{k-1}^{-1} is symmetric immediately results from formula (2). To show that H_k^{-1} is negative definite, we must show that whatever the vector $u \neq 0$, $u^T H_k^{-1} u < 0$.

It results from (2)

$$u^T H_k^{-1} u = u^T H_{k-1}^{-1} u - \frac{u^T H_{k-1}^{-1} v_k v_k^T H_{k-1}^{-1} u}{v_k^T H_{k-1}^{-1} v_k} + \frac{u^T s_k s_k^T u}{s_k^T v_k} \tag{9}$$

from where

$$u^T H_k^{-1} u = \frac{(u^T H_{k-1}^{-1} u)(v_k^T H_{k-1}^{-1} v_k) - (u^T H_{k-1}^{-1} v_k)^2}{v_k^T H_{k-1}^{-1} v_k} + \frac{(u^T s_k)^2}{s_k^T v_k} \tag{10}$$

As $v_k^T H_{k-1}^{-1} v_k < 0$, by hypothesis, and $(u^T s_k)^2$ is obviously positive, to show that $u^T H_k^{-1} u < 0$ it suffice to point out that:

$$(u^T H_{k-1}^{-1} u)(v_k^T H_{k-1}^{-1} v_k) - (u^T H_{k-1}^{-1} v_k)^2 \geqslant 0 \tag{11}$$

$$s_k^T v_k < 0. \tag{12}$$

Let us prove (11):

For this, we should first show that H_{k-1}^{-1} being negative definite, in this case

$$-H_{k-1}^{-1} = G \cdot G \tag{13}$$

with G a nonsingular symmetric matrix.

Indeed, $-H_{k-1}^{-1}$ being positively defined, there exist nonsingular matrix T so that

$$-T^{-1} H_{k-1}^{-1} T = \Lambda = \begin{bmatrix} \lambda_1 \cdot \ldots & 0 \\ & \ddots \\ 0 & \cdot \lambda_n \end{bmatrix} \tag{14}$$

Note with

$$\Lambda^{1/2} = \begin{bmatrix} \lambda_1^{1/2} \cdot \ldots & 0 \\ & \ddots \\ 0 & \cdot \lambda_n^{1/2} \end{bmatrix} \tag{15}$$

In this case

$$-H_{k-1}^{-1} = (T \Lambda^{1/2} T^{-1})^2 = G \cdot G \tag{16}$$

where

$$G = T \Lambda^{1/2} T^{-1}. \tag{17}$$

As, on the other hand, H_{k-1}^{-1} is by construction a symmetric matrix, if the initial approximation was symmetric, it results that G is also a symmetric matrix.

With this result obtained, (11) is a consequence of Schwarts's inequality:

$$(u^T H_{k-1}^{-1} u)(v_k^T H_{k-1}^{-1} v_k) - (u^T H_{k-1}^{-1} v_k)^2 = (u^T G G u)(v_k^T G G v_k) - (u^T G G v_k)^2 =$$

$$= (Gu)^T Gu \cdot (Gv_k)^T v_k - [(Gu)^T Gv]^2 =$$

$$= |Gu|^2 |Gv_k|^2 - [(Gu)^T Gv_k]^2 \geqslant 0 \tag{18}$$

the passing to the third member of relations (18) being possible due to the fact that G is a symmetric matrix and the inequality satisfied by the last term being a direct consequence of Schwartz's inequality

$$|Gu| \cdot |Gv_k| \geqslant |(Gu)^T (Gv_k)| \tag{19}$$

Let us now prove (12)
We have

$$s_k = x^k - x^{k-1}$$

from where, using (5) and (6)

$$s = -\lambda_{k-1} H_{k-1}^{-1} \nabla f(x^{k-1}). \tag{20}$$

So that

$$s_k^T v_k = -\lambda_{k-1} \nabla^T f(x^{k-1})\, H_{k-1}^{-1} (\nabla f(x^k) - \nabla f(x^{k-1})). \tag{21}$$

Supposing that optimization on direction

$$d^{k-1} = -H_{k-1}^{-1} \nabla f(x^{k-1})$$

was made with great accuracy

$$(d^{k-1})^T \nabla f(x^k) = -\nabla f(x^{k-1})\, H_{k-1}^{-1} \nabla f(x^k) = 0. \tag{22}$$

In this case (21) becomes

$$s_k^T v_k = \lambda_{k-1} \nabla^T f(x^{k-1})\, H_{k-1}^{-1} \nabla f(x^{k-1}) \tag{23}$$

which is a negative quantity if H_{k-1}^{-1} is negative definite and $\lambda_{k-1} > 0$ and thus relation (12) is proved.

With this, the Davidon-Fletcher-Powell algorithm for minimization becomes:

Considering x^0 and H_0^{-1} an approximation of $H^{-1}(x^0)$ symmetric and negative definite, calculate d^k by taking into account (5), (6) and (7'); H_k^{-1} being calculated with the aid of the recurrence formula(2).

In the subroutine programmed by us $H_0^{-1} = -I$.

The results of our tests, carried out in collaboration with S. Moga, show that this method reduces, in comparison with a method (2), (3), (4), (5), (6') (7') (8), the number of function evaluations and has a greater stability.

REFERENCES

1. DAVIDON W. C.: *Variable Metric Method for Minimization*, Research and Development Report ANL-5990 (Revised), Argonne National Laboratory, U.S. Atomic Energy Commissions, 1959
2. FLETCHER R., POWELL M. J. D.: *A Rapidly Convergent Descent Method for Minimization*, Computer J., 6 (1963)