Superquadraticly convergent methods for minimization functions

Stanisław M. Grzegórski[*], Edyta Łukasik

*Department of Computer Science, Lublin University of Technology,
Nadbystrzycka 36b, 20-618 Lublin, Poland*

Abstract

In the paper locally superquadraticly convergent methods for minimization functions are considered. Threefold symmetric approximations to partial derivatives of the third order are constructed.

1. Introduction

Let $f : D \subset R^n \to R$, $f \in C^3(D)$, D - open set. We want to find $x^* \in D$ such that $\nabla f(x^*) = 0$. For a given $x_0 \in D$ the Newton method defines the sequence $\{x_k\}$ in the following way

$$\nabla^2 f(x_k)s_k = -\nabla f(x_k), \ x_{k+1} = x_k + s_k, \ k = 0,1,2,\cdots . \tag{1}$$

If the matrix $\nabla^2 f(x^*)$ is nonsingular then Newton method is locally quadraticly convergent to x^*, i.e. there exist $c > 0$ and $\varepsilon > 0$ such that, if $\left\| x^* - x_0 \right\| < \varepsilon$, then

$$\left\| x_{k+1} - x^* \right\| \le c \left\| x_k - x^* \right\|^2 . \tag{2}$$

To assure global convergence of the method one should consider a sequence

$$x_{k+1} = x_k + t_k s_k, \ t_k \in R, \ k = 0,1,2,\cdots \tag{3}$$

and the parameter t_k should satisfy the global convergence conditions. If the matrix $\nabla^2 f(x^*)$ is singular, then the Newton method is divergent or at most linearly convergent to x^*. To assure a great speed of convergence for singular problems one applies the method of the third rate of convergence: for a given $x_0 \in D$ the sequence $\{x_k\}$ is defined as

$$x_{k+1} = x_k + s_k, \ k = 0,1,2,\cdots \tag{4}$$

---

[*] Corresponding author: *e-mail address*: s.grzegorski@pollub.pl

where s_k is the solution of the system of quadratic equations

$$\nabla f(x_k) + \nabla^2 f(x_k)s_k + \frac{1}{2}(\nabla^3 f(x_k)s_k, s_k) = 0. \tag{5}$$

When the calculation of the operator $\nabla^3 f(x_k)$ is too expensive or is not attainable for computation then we propose a new class of the methods which are locally superquadraticly convergent to x^*, i.e.

$$\lim_{k \to \infty} \frac{\left\| x_{k+1} - x^* \right\|}{\left\| x_k - x^* \right\|^2} = 0. \tag{6}$$

Let $B_k = (B_k^1, B_k^2, \cdots, B_k^n)$, $B_k^i \in R^{n \times n}$, $B_k^i = (B_k^i)^T$, $i = 1, 2, \cdots, n$. The sequence $\{x_k\}$ is defined by (1.4) and

$$\nabla f(x_k) + \nabla^2 f(x_k)s_k + \frac{1}{2}(B_k s_k, s_k) = 0. \tag{7}$$

If the problem $\min_{x \in D} f(x)$ is regularly singular at x^*, i.e.

$$\det(\nabla^2 f(x^*)) = 0, \text{ and } \left\| \nabla f(x) \right\| \geq c \left\| x - x^* \right\|^2, \ c > 0, \ x \in D, \tag{8}$$

then the sequence $\{x_k\}$ defined by (1.4) and (1.7) is locally superlinearly convergent to x^*, if the operators B_k are constructed in an adequate way. In this paper such algorithms are given.

2. The BFGS method

The DFP (Davidon [1], Fletcher and Powell [2]) method is very well known as the method of aproximation to the Hessian $\nabla^2 f(x_k)$. This formula has the form

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \ y_k = \nabla f(x_{k+1}) - \nabla f(x_k). \tag{9}$$

The DFP formula, for nonsingular problems, guarantees local superlinear convergence of the method

$$x_{k+1} = x_k + s_k, \ B_k s_k = -\nabla f(x_k), \ k = 0, 1, 2, \cdots \tag{10}$$

We may use the DFP formula to approximate the operator $\nabla^3 f(x_k)$. Namely, let

$$B_k = (B_k^1, B_k^2, \cdots, B_k^n), \ B_k^i \in R^{n \times n}, \ B_k^i = (B_k^i)^T, \ i = 1, 2, \cdots, n \tag{11}$$

and let $\nabla_i^2 f(x_k)$ denote i-th column of the matrix $\nabla^2 f(x_k)$, $y_k^i = \nabla_i^2 f(x_{k+1}) - \nabla_i^2 f(x_k)$. Then

$$B_{k+1}^i = B_k^i - \frac{B_k^i s_k s_k^T B_k^i}{s_k^T B_k^i s_k} + \frac{y_k^i (y_k^i)^T}{(y_k^i)^T s_k}, \ i = 1, 2, \cdots, n. \tag{12}$$

Note that the operators B_{k+1} satisfy the equation

$$B_{k+1}s_k = \nabla^2 f(x_{k+1}) - \nabla^2 f(x_k), \ k = 0,1,2,\cdots \qquad (13)$$

Now, from the general theory for the systems of nonlinear equations [3], [4] local superquadratic convergence of the method results (4), (7) with the update (12). On the other hand, the DFP updates do not satisfy the following properties of the partial derivatives

$$\frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_l} = \frac{\partial^3 f(x)}{\partial x_i \partial x_l \partial x_j} = \cdots = \frac{\partial^3 f(x)}{\partial x_l \partial x_j \partial x_i}, \ i,j,l = 1,2,\cdots,n \ . \qquad (14)$$

In this case, we say the operator $\nabla^3 f(x)$ is threefold symmetric (T-symmetric).

It is worth remarking that the operator $\nabla^3 f(x)$ has only $P(n) = \frac{1}{6}n(n+1)(n+2)$

different elements and the DFP aproximations have $Q(n) = \frac{1}{2}n^2(n+1)$ different

elements, which means that the BFGS formula is not adequate for approximation to $\nabla^3 f(x)$. In the next Section we give a new formula for the update of B_k and B_k will be T-symmetric.

3. New approximation to $\nabla^3 f(x)$

The approximation B_k to $\nabla^3 f(x)$ satisfies secant equation (13) and operators B_k should be threefold symmetric. If we take

$$B_{k+1} = B_k + E \ , \qquad (15)$$

then

$$Es_k = \nabla^2 f(x_{k+1}) - \nabla^2 f(x_k) - B_k s_k = Y \ . \qquad (16)$$

In that case we have to solve the problem

$$\min \|E\|^2 , \ \|E\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} E_{ijk}^2 \qquad (17)$$

under constraints

$$Es = Y, \ s \in R^n, \ Y \in R^{n \times n}, \ Y = Y^T \qquad (18)$$

and

$$E_{ijk} = E_{ikj} = E_{jik} = E_{jki} = E_{kij} = E_{kji}, \ i,j,k = 1,2,\cdots,n \ . \qquad (19)$$

Remark. If we take another norm of the operator E, then we get another formula for the update B_k.

Let $\Lambda \in R^{n \times n}$. In our case the lagrangian has the form

$$L(E,\Lambda) = \frac{1}{2}\|E\|^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}\Lambda_{ij}(\sum_{k=1}^{n} E_{ijk}s_k - Y_{ij}) \ . \qquad (20)$$

From this we have

$$\frac{\partial L(E,\Lambda)}{\partial E_{pqr}} = E_{pqr} + \Lambda_{pq}s_r = 0 . \tag{21}$$

The fact $E_{pqr} = E_{qpr}$ implies $\Lambda = \Lambda^T$. Since the operator E is threefold symmetric then equation (21) may be written as

$$E_{pqr} = -\frac{1}{3}(\Lambda_{pq}s_r + \Lambda_{pr}s_q + \Lambda_{qr}s_p) \quad \text{for } 1 \le p \le q \le r \le n . \tag{22}$$

Now, the equation $Es = Y$ has the form

$$\sum_{l=1}^{n}(\Lambda_{ij}s_l + \Lambda_{il}s_j + \Lambda_{jl}s_i)s_l = -3Y_{ij} \quad 1 \le i \le j \le n \tag{23}$$

or is in the matrix form

$$\Lambda\|s\|^2 + \Lambda ss^T + ss^T\Lambda = -3Y . \tag{24}$$

Therefore

$$s^T\Lambda s = -\frac{1}{\|s\|^2}s^TYs, \; u = \Lambda s = \frac{1}{2\|s\|^2}(-3Ys + \frac{s^TYs}{\|s\|^2}s) . \tag{25}$$

Finally

$$\Lambda = -\frac{1}{\|s\|^2}(3Y + us^T + su^T) . \tag{26}$$

To calculate the new threefold symmetric update $B_{k+1} = B_k + E$ we use the formulae (22), (25) and (26).

4. Remarks on the local superquadratic convergence of the method

At first we describe the proposed algorithm:

a) Let $x_0 \in R^n$ and $B_0 = (B_0^1, B_0^2, \cdots, B_0^n)$ - threefold operator be given. Let $k = 0$,

b) Solve, using for example the Newton method, the system of quadratic equations

$$\nabla f(x_k) + \nabla^2 f(x_k)s_k + \frac{1}{2}(B_k s_k, s_k) = 0 ,$$

c) Calculate $x_{k+1} = x_k + s_k$, $\nabla f(x_{k+1})$, $\nabla^2 f(x_{k+1})$,

d) Update the operator B_k using the formulae from Section 3,

e) If a stop criterion is not satisfied, then $k := k + 1$ and return to point b.

To explain a character of convergence of the method we introduce some notations. Let

$$H_k = \int_0^1 \nabla^3 f(x_k + t(x_{k+1} - x_k))dt , \qquad (27)$$

$$L_k = \{X \in R^{n \, x \, n \, x \, n} : Xs_k = \nabla^2 f(x_{k+1}) - \nabla^2 f(x_k)\}, \qquad (28)$$

$$Q = \{X \in R^{n \, x \, n \, x \, n} : X \text{ is threefold symmetric operator}\}. \qquad (29)$$

The set Q is linear subspace in $R^{n \, x \, n \, x \, n}$ and $H_k \in Q$. Applying Theorem 3.2.7 [5] we have

$$H_k s_k = \nabla^2 f(x_{k+1}) - \nabla^2 f(x_k), \qquad (30)$$

which means that $Q \cap L_k$ is a nonempty linear set. The proposed norm is generated by inner product, so the operator B_{k+1} is defined as the orthogonal projection of the operator B_k onto the set $Q \cap L_k$, and from Pitagoras Theorem (see [3]) we get

$$\|B_{k+1} - H_k\|^2 + \|B_k - B_{k+1}\|^2 = \|B_k - H_k\|^2 \quad k = 0,1,2,\cdots \qquad (31)$$

The inequality $\|B_{k+1} - H_k\| \le \|B_k - H_k\|$ implies local linear convergence of the sequence $\{x_k\}$. From equations (31) it results additionally

$$\sum_{k=0}^{\infty} \|B_{k+1} - B_k\|^2 < \infty . \qquad (32)$$

The last inequality and the secant equation (13) assure local superquadratic convergence of the proposed algorithm [4].

References

[1] Davidon W.C.,*Variable metric method for minimization*, Report ANL-5990 Rev (1959), Argonne National Laboratories, Ill.

[2] Fletcher R., Powell M.J.D., *A rapidly convergent descent method for minimization*, Comput. J., 6 (1963) 163.

[3] Grzegórski S.M., *Orthogonal projections on convex sets for Newton-like methods*, SIAM J. on Numer. Anal., 22 (1985) 1208.

[4] Grzegórski S.M., Łukasik E., *Theory of convergence for 2-rank iterative methods*, prepared for publication.

[5] Ortega J.M., Rheinboldt W.C., *Iterative Solution of Nonlinear Solution in Several Variables*, Academic Press, New York, London, (1970).