Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2001 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-19-2001

A New Class of Non-Quest-Newton Methods and Their Global Convergence with Goldstein Line Search

Changhai Zhang

Kaoping Song

Yuxue Wang

A NEW CLASS OF NON-QUASI-NEWTON METHODS AND THEIR GLOBAL CONVERGENCE WITH GOLDSTEIN LINE SEARCH

Changhai Zhang
Address: Mathematical Science Department, Daqing Petroleum Institute, 151400 Anda, HeiLongJiang, China
Telephone: 86-459-4653702
Kaoping Song
Address: Petroleum Engineering Department, Daqing Petroleum Institute, 151400 Anda, HeiLongJiang, China
Telephone: 86-459-4654630
E-mail: skp@dqpi.edu.cn
Yuxue Wang
Address: Mathematical Science Department, Daqing Petroleum Institute, 151400 Anda, HeiLongJiang, China

ABSTRACT

In this paper, on the basis of the DFP method a class of non-quasi-Newton methods is presented. Under some condition the global convergence property of these methods with Goldstein line search on uniformly convex objective function is proved.

PROPOSAL OF NEW ALGORITHMS

For problems of unconstrained optimization

$$\min f(x), \quad x \in R^n \tag{1.1}$$

The quasi-Newton methods is one of the most well considered, and extensive methods and the DFP method which is one of quasi-Newton methods, given by Davidon[1], revised by Fletcher and Powell[2], was first derived.

Steps of DFP algorithm are as follows

Algorithm A

Step1. Given $x_1 \in R^n$, B_1 is $n \times n$ symmetric and positive definite matrix,

Step2. Calculate $g_k = \nabla f(x_k)$, if $g_k = 0$, then stop calculating, x_k can be obtained, otherwise turn to the next step.

Step3. $d_k = -B^{-1} g_k$.

Step4. Do line search to determine step length l_k.

Step5. $x_{k+1} = x_k + l_k d_k$.

Step6.

$$B_{k+1} = B_k - \frac{B_k s_k y_k^T + y_k s_k^T B_k}{s_k^T y_k} + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k y_k^T}{s_k^T y_k} \tag{1.2}$$

Where
$$s_k = x_{k+1} - x_k \tag{1.3}$$
$$y_k = g_{k+1} - g_k \tag{1.4}$$

Step7. $k := k+1$ and go to Step2.

Formula (1.2) is DFP update formula, the present paper revises (1.2), as follows

Step 6'

$$B^{k+1}(t,t) = B^k - \frac{B^k s^k y^{kT} + y^k s^{kT} B^k}{s_k^T y_k} + (Q^k(t,t) + s^{kT} B^k s^k) \frac{y^k y^{kT}}{(s_k^T y_k)^2} \tag{1.5}$$

Where

$$Q_k(t, t) = t y_k^T s + t\left(f_{k+1} - f_k - s_k^T g_k\right) \tag{1.6}$$

Here

$f_k = f(x_k)$, $t \in [0, T_1]$, $t \in [0, T_2]$, $t + t \geq t_0$, $0 < t_0 < 1$, T_1 and T_2 are constants, which are larger than 1. When $t = 1$, $t = 0$ formula (1.5), that is formula (1.2).

Make (1.5) replace (1.2), the rest procedures are the same to Algorithm **A**, the derived algorithms may be written as Algorithms. The characteristic of Algorithms may be analyzed as follows

From (1.5), we gain

$$B_{k+1}(t,t) s_k = \frac{Q_k(t,t)}{s_k^T y_k} y_k .$$

When $t \neq 1$ or $t \neq 1$, $\frac{Q_k(t,t)}{s_k^T y_k} \neq 1$, therefore Algorithms $B(t,t)$ are not the quasi-Newton methods. And since they contain DFP method, therefore Algorithms $B(t,t)$ are called a class of non-quasi-Newton methods.

To Algorithms $B(t,t)$, there are two methods to determine the steplength l_k. One is exact line search, the other is inexact line search. The paper examines Goldstein line search, that is to say l_k satisfy

$$f(x_k + l_k d_k) \leq f(x_k) + r\, l_k g_k^T d_k, \tag{1.8}$$

$$f(x_k + l_k d_k) \geq f(x_k) + s\, l_k g_k^T d_k, \tag{1.9}$$

Where r、s are constant, and $0 < r < s < 1$.

SEVERAL LEMMAS AND PROOFS [3][4][5]

In order to discuss the global convergence property of Algorithms $B(t,t)$ with Goldstein line search, we may assume objective function $f(x)$ to be as follows

(a) $f(x)$ is twice continuously differentiable;

(b) There exist positive constants m and M such that

$$m\|y\|^2 \leq y^2 \nabla^2 f(x) y \leq M\|y\|^2 , \tag{2.1}$$

For all $x \in R^n$ and all $y \in R^n$, where and hereinafter ‖ • ‖ stands for the Euclidean norm.

Lemma 1

$$m \le \frac{s_k^T y_k}{\|s_k\|^2} \le \frac{\|y_k\|^2}{s_k^T y_k} \le M. \qquad (2.2)$$

$$\sum_{k=1}^{\infty} l_k g_k^T H_k g_k = \sum_{k=1}^{\infty} -s_k^T g_k < +\infty. \qquad (2.3)$$

$$-\frac{2(1-s)M}{2M-m} s_k^T g_k \le s_k^T g_k - \frac{2(1-r)M}{m} s_k^T g_k \qquad (2.4)$$

Proof. From mean value theorem and condition (a), (b) we can get (2.2).

As for the proof of formula (2.3), see 〔6〕.

According to mean value theorem and condition (a), (b), we get

$$f(x_k + l_k d_k) - f(x_k) \ge s_k^T g_k + \frac{1}{2} m \| s_k \|^2 \qquad (2.5)$$

From (1.8) we have

$$f(x_k + l_k d_k) - f(x_k) \ge r s_k^T g_k \qquad (2.6)$$

And hence by (2.5) and (2.6) we get

$$m \| s_k \|^2 \le -2(1-r) s_k^T g_k \qquad (2.7)$$

Using (2.2)

$$\| s_k \|^2 \ge \frac{s_k^T y_k}{M} \qquad (2.8)$$

Which, with (2.7), implies that

$$s_k^T y_k \le -\frac{2(1-r)M}{m} s_k^T g_k \qquad (2.9)$$

From mean value theorem and (1.9) we obtain

$$-s_k^T g_{k+1} + \frac{1}{2} m \| s_k \|^2 \le -s g_k^T s_k, \qquad (2.10)$$

Therefore from (2.2) and (2.9) we get

$$s_k^T y_k \ge -(1-s) s_k^T g_k + \frac{m}{2M} s_k^T y_k, \qquad (2.11)$$

Which implies that

$$s_k^T y_k \ge -\frac{2(1-s)M}{2M-m} s_k^T g_k. \qquad (2.12)$$

Lemma 2

Let $a_k > 0, b_k > 0$ for all $k \ge 1$, and there exist positive constants b_1, b_2, such that

$$a_k \le b_1 + b_2 \sum_{j=1}^{k-1} l_j \qquad (2.13)$$

for all $k \ge 1$ and $\sum_{k=1}^{\infty} \frac{b_k}{a_k} < +\infty$, then $\sum_{k=1}^{\infty} b_k < +\infty$.

Lemma 3

There exist positive constants m_1 and M_1 such that

$$m_1 s_k^T y_k \le Q_k(t, t) \le M_1 s_k^T y_k \qquad (2.14)$$

For all positive integer k, where $Q_k(t, t)$ is from the definition of (1.6).

Proof. It follows from (1.6) and Taylor's formula that

$$\begin{aligned} Q_k(t,t) &= t s_k^T y_k + 2t(f(x_{k+1}) - f(x_k) - \nabla f(x_k)^T s_k) \\ &= t s_k^T y_k + t s_k^T \nabla^2 f(\overline{x_k}) s_k \end{aligned} \qquad (2.15)$$

where $\overline{x}_k$ is between x_k and x_{k+1}. From assumed condition (a) and (b), we obtain

$$m \| s_k \|^2 \le s_k^T \nabla^2 f(\overline{x_k}) s_k \le M \| s_k \|^2, \qquad (2.16)$$

And from (2.2) obtain

$$\frac{s_k^T y_k}{M} \le \| s_k \|^2 \le \frac{s_k^T y_k}{m} \qquad (2.17)$$

Thus (2.7)-(2.9) give

$$Q_k(t, t) \ge t s_k^T y_k + t \frac{m}{M} s_k^T y_k \ge t_0 \frac{m}{M} s_k^T y_k$$

And

$$Q_k(t, t) \le t s_k^T y_k + t \frac{M}{m} s_k^T y_k \ge (T_1 + T_2) \frac{M}{m} s_k^T y_k$$

Let $m_1 = t_0 \frac{m}{M}, M_1 = (T_1 + T_2) \frac{M}{m}$, then (2.14) hold.

Lemma 4

If B_k is symmetric and positive definite for $k \ge 1$, then $B_{k+1}(t, t)$ from the definition of (1.5) is symmetric and positive as well.

Proof. Formula (1.5) can be written as

$$B_{k+1}(t, t) = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{Q_k(t, t)}{(s_k^T y_k)^2} y_k y_k^T + v_k v_k^T \qquad (2.18)$$

Where

$$v_k = (s_k^T B_k s_k)^{\frac{1}{2}} \left(\frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k} \right).$$

Let

$$\tilde{B}_k = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{Q_k(t, t)}{(s_k^T y_k)^2} y_k y_k^T \qquad (2.19)$$

Then from the calculating formula of revised determinant of rank 2 and rank1, we obtain

$$\det(\tilde{B}_k) = \det(B_k) \frac{Q_k(t, t)}{(s_k^T B_k s_k)^2} \qquad (2.20)$$

$$\det(B_{k+1}(t, t)) = \det(B_k) \frac{y_k^T H_k y_k}{(s_k^T y_k)^2} Q_k(t, t) \qquad (2.21)$$

Where $H_k = B_k^{-1}$. From (2.2) and (2.14), we have $Q_k(t, t) > 0$ and hence $B_k + \frac{Q_k(t, t)}{(s_k^T y_k)^2} \times y_k y_k^T$ is positive definite. Thus from (2.19) and the interlocking eigenvalue theorem of rank1 revised matrix, we know that the least eigenvalue of $\widetilde{B}_k$ and det($\widetilde{B}_k$) have the same sign. But from (2.20), we know det($\widetilde{B}_k$)>0, therefore $\widetilde{B}_k$ is positive definite. Moreover, from (2.21), we know det($\widetilde{B}_{k+1}(t,t)$)>0, thus by using the interlocking eigenvalue theorem once again, we know that $B_{k+1}(t,t)$ is positive definite.

Lemma 5

For all $k \ge 1$, we have

$$\sum_{j=1}^{k} l_j \ge kC \tag{2.22}$$

Where C is a positive constant.

Proof. According to Lemmas 3, 4 and the proving process similar to [3], we can prove the lemma.

Lemma6

The following limit holds

$$\lim g_{kT} H_k g_k = 0 . \tag{2.23}$$

Proof. According to (1.5) and make use of matrix inversion formula, we get

$$H_{k+1(t}(t,t) = H_k - \frac{H_k y_k y_k^t H_k}{y_{kT} H_k y_k} + \frac{s_k s_k^T}{Q(t,t)_k} \tag{2.24}$$

From (2.24) and $g_{k+1} = g_k + y_k$, we get

$$g_{k+1}^T H_{k+1} g_{k+1} = y_l^T H_k g_k + 2 g_k^T H_{k+1} y_k + g_k^T H_{k+1} g_k$$
$$= \frac{s_k^T y_k}{Q_k(t,t)} + 2\frac{s_k^T y_k}{Q_k(t,t)} s_k^T g_k + g_k^T H_k g_k - \frac{(g_k^T H_k g_k)}{y_k^T H_k y_k} + \frac{(s_k^T g_k)^2}{Q_k(t,t)} \tag{2.25}$$

Then

$$\frac{(g_k^T H_k g_k)^2}{y_k^T H_k y_k} = g_k^T H_k g_k - g_{k+1}^T H_{k+1} g_{k+1} + \frac{s_k^T g_{k+1}}{Q_k(t,t)} \tag{2.26}$$

Replace index k of formula (2.26) with j, and extract the sum from 1 to k to j at the two ends, we get

$$\sum_{j=1}^{k} \frac{(g_j^T H_j y_j)^2}{y_j^T H_j y_j} = g_1^T H_1 g_1 - g_{k+1}^T H_{k+1} g_{k+1} + \sum_{j=1}^{k} \frac{(s_k^T g_{j+1})^2}{Q_j(t,t)} \tag{2.27}$$

But from (2.14), (2.3) and (2.4), we get

$$\sum_{j=1}^{k} \frac{(s_j^T g_{k+1})^2}{Q(t,t)} \le \frac{1}{m_1} \sum_{j=1}^{k} \left[\frac{s_j^T g_j}{s_j^T y_j} + 2 s_j^T g_j + s_j^T y_j \right]$$

$$\le \frac{1}{m_1} \frac{2(2s-1)Mm - m^2 + 4(1-r)(1-s)M^2}{2(1-s)Mm} \sum_{j=1}^{k} (-s_j^T g_j) \tag{2.28}$$

For $0 < r < s < 1$, we have

$$2(2s-1)Mm - m^2 + 4(1-r)(1-s)M^2$$
$$> 2[(1-s)^2 + (2s-1)]Mm + [2(1-s)^2 - 1]m^2$$
$$> (2s-1)^2 m^2 \ge 0$$

And hence by (2.27) and (2.28) we get

$$\sum_{j=1}^{\infty} \frac{(s_j^T g_{j+1})^2}{Q_j(t,t)} < +\infty \tag{2.29}$$

Therefore from (2.29), we get

$$\sum_{j=1}^{\infty} \frac{(g_j^T H_j y_j)^2}{y_j^T H_j y_j} < g_1^T H_1 g_1 + \sum_{j=1}^{\infty} \frac{(s_j^T g_{j+1})^2}{Q_j(t,t)} < +\infty \tag{2.30}$$

From (2.27)、(2.29) and (2.30), we know $\lim_{k\to\infty} g_k^T H_k g_k$ exists. If $\lim_{k\to\infty} g_k^T H_k g_k > 0$, then from (2.3) $\sum_{k=1}^{\infty} l_k < +\infty$ is known, that is in contradiction with (2.22), therefore (2.23) is correct.

Let $r_k = \frac{-s_k^T g_k}{s_k y_k}$, then from (2.4) and (2.23), we know $\lim_{k\to\infty} \frac{g_k^T H_k g_k}{r_k} = 0$, therefore there is a subsequence, which is monotone decreasing towards 0, in the sequence $\left\{ \frac{g_k^T H_k g_k}{r_k} \right\}$.

Lemma 7

Let $l = (m/M)^3$, then

$$\|g_j\|^2 \ge l \|g_{k+1}\|^2 \tag{2.31}$$

Holds for $j = k, k-1, \cdots, 1.$

Proof. see [4]

GLOBAL CONVERGENCE RESULTS[6][7]

The main results of this paper are introduced and proved as follows.

Theorem1. Assume conditions (a), (b) holds, and $\{x_k\}$ to be a sequence derived from Algorithms $B(t,t)$, if one of the following two conditions holds

(i) $\|g_{k+1}\|^2 - \|g_k\|^2 - \|y_k\|^2 = O\left(\frac{g_k^T H_k g_k}{r_k} \right)$,

(ii) when k is sufficiently large, $\left\{ \frac{g_k^T H_k g_k}{r_k} \right\}$ is monotone decreasing, and $s < \frac{m}{2M} + l\left(1 - \frac{m}{2M}\right)$, then

Algorithms $B(t,t)$ has global convergence property

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{3.1}$$

Proof. From (1.5), (2.3), (2.6) and $\frac{s_j^T y_j}{l_j} = \frac{y_j^T H_j g_j}{r_j}$, we get

$$tr(B_{k+1}) = tr(B_k) - 2\frac{y_k^T B_k s_k}{s_k^t y_k} + (Q_k(t,t) + s_k^T B_k s_k) \frac{\|y_k\|^2}{(y_k^T)^2}$$

$$=tr(B_1)-2\sum_{j=1}^{k}\frac{y_j^T B_j s_j}{s_j^T y_j}+\sum_{j=1}^{k}\left(Q_j(t,t)+s_j^T B_j s_j\right)\frac{\|y_k\|^2}{\left(s_j^T y_j\right)^2}$$

$$=tr(B_1)+\sum_{j=1}^{k}\frac{l_j\left(\|g_{j+1}\|^2-\|g_j\|^2-\|y_j\|^2\right)}{s_j^T y_j}$$

$$+\sum_{j=1}^{k}\frac{Q_j(t,t)+s_j^T B_j s_j}{s_j^T}\frac{\|y_j\|^2}{s_j^T y_j}$$

$$\leq tr(B_1)+\sum_{j=1}^{k}\frac{l_j\left(\|g_{j+1}\|^2-\|g_j\|^2-\|y_j\|^2\right)}{s_j^T g_j}+$$

$$\sum_{j=1}^{k}\frac{2M-m}{2(1-s)}l_j+MM_1k. \tag{3.2}$$

Assume condition (i) holds, and $\lim_{k\to\infty}\inf\|g_k\|>0$, then there is $e>0$ such that

$$\|g_k\|\geq e \tag{3.3}$$

For all $k\geq 1$. From condition (i), there exist positive constant L such that

$$\|g_{j+1}\|^2-\|g_j\|^2-\|y_j\|^2\leq L\frac{g_j^T H_j g_j}{r_j}$$

That is

$$\frac{r_j\left(\|g_{j+1}\|^2-\|g_j\|^2-\|y_j\|^2\right)}{g_j^T H_j g_j}\leq L \tag{3.4}$$

for all $j\geq 1$. From (3.2), (3.4) and (2.22), we get

$$tr(B_{k+1})\leq tr(B_1)+LR+\frac{2M-m}{2(1-s)}\sum_{j=1}^{k}l_j+M_1MR$$

$$\leq tr(B_1)+\left(\frac{L}{C}+MM_1+\frac{2M-m}{2(1-s)}\right)\sum_{j=1}^{k}l_j. \tag{3.5}$$

From (3.3)、(2.3) and, $tr(B_k)\geq\|B_k\|\geq\frac{\|g_k\|^2}{g_k^T H_j g_j}$, we know

$$\sum_{k=1}^{\infty}\frac{l_r}{tr(B_k)}\leq\sum_{k=1}^{\infty}\frac{g_k^T H_k g_k}{\|g_k\|^2}l_r\leq\frac{1}{e^2}\sum_{k=1}^{\infty}l_k g_k^T H_k g_k<+\infty \tag{3.6}$$

Thus, from Lemma 2, (3.5) and (3.6) we know $\sum_{k=1}^{\infty}l_k<+\infty$, this is in contradiction with (2.22).

Now assume condition(ii) to be correct, still suppose $\lim_{k\to\infty}\inf\|g_k\|>0$, then (3.4) holds. Let $m=\frac{(1-l)\left(1-\frac{m}{2M}\right)}{1-s}$, then

$$0<m<1. \tag{3.7}$$

From condition(ii), we know that there exists a positive integer k_0 such that when $k\geq k_0$,

$$\frac{g_{k_0}^T H_{k_0} g_{k_0}}{r_{k_0}}\geq\frac{g_{k_0+1}^T H_{k_0+1} g_{k_0+1}}{r_{k_0+1}}\geq\cdots\geq\frac{g_k^T H_k g_k}{r_k}\geq\cdots \tag{3.8}$$

We assume without loss of generality that $k_0=1$. From (3.8).(2.4) and (2.31), we obtain

$$\sum_{i=1}^{k}\frac{r_j\left(\|g_{j+1}\|^2-\|g_j\|^2-\|y_j\|^2\right)}{g_j^T H_j g_j}\leq\sum_{j=1}^{k}\left[\frac{r_{j+1}\|g_{j+1}\|^2}{g_{j+1}^T H_{j+1} g_{j+1}}-\frac{r_j\|g_j\|^2}{g_j^T H_j g_j}\right]+$$

$$\sum_{j=1}^{k}\left[\frac{r_j}{g_j^T H_j g_j}-\frac{r_{j+1}}{g_{j+1}^T H_{j+1} g_{j+1}}\right]\|g_{j+1}\|^2$$

$$\leq\frac{r_{k+1}\|g_{k+1}\|^2}{g_{k+1}^T H_{k+1} g_{k+1}}-\frac{r_1\|g_1\|^2}{g_1^T H_1 g_1}-\sum_{j=1}^{k}\left[\frac{r_{j+1}}{g_{j+1}^T H_{j+1} g_{j+1}}-\frac{r_j}{g_j^T H_j g_j}\right]\|g_{j+1}\|^2$$

$$=\frac{\|g_{j+1}\|^2}{g_{j+1}^T H_{j+1} g_{j+1}}r_{j+1}(1-l)-\frac{r_1}{g_1^T H_1 g_1}\left(\|g_1\|^2-l\|g_{j+1}\|^2\right)$$

$$\leq\frac{\left(1-\frac{m}{2M}\right)(1-l)}{1-s}\frac{\|g_{j+1}\|^2}{g_{j+1}^T H_{j+1} g_{j+1}}\leq m\|B_{j+1}\|\leq m\,tr(B_{j+1}) \tag{3.9}$$

Therefore from (3.2), (3.9), (3.7) and (2.22), we obtain

$$tr(B_{k+1})\leq\frac{tr(B_1)}{1-m}+\left[\frac{2M-m}{2(1-m)(1-s)}+\frac{MM_1}{C(1-m)}\right]\sum_{j=1}^{k}l_j \tag{3.10}$$

As the form of (3.10) is similar to that of (3.5), and hence (3.6) holds too. Therefore from (3.6),(3.10) and Lemma 2, we know $\sum_{j=1}^{\infty}l_j<+\infty$, which contradicts the (2.22), thus (3.1) must hold.

REFERENCES

[1] W.C. Davidon, Variable metric methods for minimization, Argonne National Laboratories Report ANL-5990,1959

[2] R.Fletcher, M.J.D.Powell, A rapid convergent descent method for minimization, The computer Journal, 6(1963),163-168

[3] R.Byrd, J.Nocedal, Y.Yuan, Global convergence of a class of quasi-Newton methods on convex problems, SIAM J. Numer. Anal. 24(1987),1171-1189

[4] Xu Da-chuan, Global convergence analysis of DFP method, Math. Numer. Sinica, 3(1997),287-292

[5] Ya-Xiang Yauan, Richard H.Byrd, Non-quasi-Newton Updates for Unconstrained Optimization, Journal of Computational Mathematics, 2(1995),95-107.

[6] Y.Yuan, Convergence of DFPalgorithm, Science in China (Series A), 38(1995),1281-1294.

[7] Chen Lanping, Jiao Baocong, Convergence properties of the preconvex part of non-quasi-Newton's family, Math. Numer. Sinica, 3(2000), 369-278.