Global Convergence of Perry-Shanno Memoryless Quasi-Newton-type Method

Yigui Ou *, Jun Zhang and Qian Zhou
Department of Mathematics, Hainan University, Haikou 570228, P.R.China
(Received 11 August 2008, accepted 10 September 2010)

Abstract: In this paper, we analyze the global convergence properties of Perry-Shanno memoryless quasi-
Newton type method associating with a new line search model. preliminary numerical results are also re-
ported.

Keywords: unconstrained optimization; memoryless quasi-Newton method; line search; global convergence

1 Introduction

In this paper, we consider the following unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable function.

Quasi-Newton method is a well-known and useful method for solving unconstrained convex programming and the
BFGS method is the most effective quasi-Newton type methods for solving unconstrained optimization problems from
the computation point of view. For the current iterate $x_k \in \mathbb{R}^n$ and symmetric positive definite matrix $B_k \in \mathbb{R}^{n \times n}$, the
next iterate is obtained by

$$x_{k+1} = x_k + a_k d_k$$

where $a_k > 0$ is a step-size obtained by a one-dimensional line search, and

$$d_k = -B_k^{-1} \nabla f(x_k)$$

is a descent direction with $B_k^{-1}$ being available and approximating the inverse of the Hessian matrix of $f$ at $x_k$. Throughout
this paper, we use $\| \|$ to denote the Euclidean vector or matrix norm and denote $\nabla f(x_k)$ by $g_k$.

Although the quasi-Newton type method is known to be remarkably robust in practice, one will not be able to establish
truly convergence results for general nonlinear objective functions, that is, one cannot prove that the iterates generated
by this method approach a stationary point of the problem from any starting point and any (suitable) initial Hessian
approximation. Therefore, there has been an ever-expanding interest in quasi-Newton type methods since the first quasi-
Newton method was suggested by Davidon (1959) and improved by Fletcher and Powell (1963) (hence the name “DFP”
formula), and there is a vast literature on the problem of convergence properties of the quasi-Newton type method for
solving problem (1). For details, see Refs[1], [2], [3] and references therein.

In 1978, Shanno [4] presented a memoryless quasi-Newton-type method, in which the search direction $d_k$ is defined by

$$\begin{cases}
    d_k = -B_k^{-1} g_k & (k \geq 1), \\
    d_0 = -g_0
\end{cases}$$

and $B_k$ is updated by the following Perry and Shanno formula

$$B_{k+1} = \frac{\| g_k \|^2}{y_k^T s_k} I + \frac{y_k y_k^T}{y_k^T s_k} - \frac{\| g_k \|^2}{s_k^T y_k s_k} s_k s_k^T$$

*Corresponding author. E-mail address: ouyigu@tom.com

Copyright © World Academic Press, World Academic Union
IUNS.2011.04.15/455
with \( s_k = x_{k+1} - x_k, y_k = g_{k+1} - g_k \).

If we use reverse form \( H_k \) of \( B_k \), then we can obtain the formula of \( H_k \)
\[
H_{k+1} = \frac{y_k^T s_k}{\|y_k\|^2} I + \frac{s_k^T s_k}{y_k^T s_k} - \frac{1}{\|y_k\|^2} \left( y_k s_k^T + s_k y_k^T \right)
\]  
(5)
and the next search direction can be written as
\[
d_{k+1} = -B_{k+1}^{-1} g_{k+1} = -H_{k+1} g_{k+1}
\]  
(6)
Moreover, Shanno proved the global convergence properties of this method in association with Wolf step-size selection rules. Subsequently, Werner [5] generalized Shanno’s conclusion with a step-size \( a_k \) satisfying
\[
f(x_k + a_k d_k) \leq f(x_k) - \min \{ \mu_1(-g_k^T d_k), \mu_2(-g_k^T d_k)^2 \}
\]  
(7)
where \( \mu_1 \) and \( \mu_2 \) are two positive constants. Recently, Liu and Jing [6] analyzed the convergence of the method with nonmonotone Wolfe line search.

As mentioned above, the problem of convergence properties of the quasi-Newton type method has been studied extensively and a great deal of achievements have been made. However, the proofs of the convergence properties only have relation to concrete step-size selection rules. Recently, Han, Liu and Guo [7, 8, 9] have proven the global convergence properties of several BFGS-type methods for solving problem (1) associated with general line search model. Motivated by the general line search model and (7), we propose a new line search technique, and consider the global convergence properties of Perry-Shanno memoryless quasi-Newton type method associating with this new technique. Compared with [6]-[9], we obtain the strong convergence property: \( \lim_{k \to +\infty} \| g_k \| = 0 \) instead of the weak property: \( \lim_{k \to +\infty} \inf \| g_k \| = 0 \).

This paper is organized as follows: In section 2, we describe the new line search model and related quasi-Newton type method. In section 3, we analyze the convergence property of the method with this new model. In section 4, we present numerical experiments. Some conclusions are summarized in section 5.

2 New line search model

In this section, a new method with a new line search model is given. To this end, we first give some definitions as in Ref[7], which is useful subsequently.

**Definition 1**  
The function \( \sigma : [0, +\infty) \to [0, \infty) \) is called forcing function if \( \lim_{i \to +\infty} \sigma(t_i) = 0 \) for any \( t_i \geq 0, i = 1, 2, \cdots \), then one has \( \lim_{i \to +\infty} t_i = 0 \).

Using the definition of forcing function, the following definition can be given.

**Definition 2**  
The general form of step-size selection rules: Let \( a > 0, \beta \in [0, 1], \sigma_1, \sigma_2, \sigma_3 \) and \( \psi \) are forcing functions, choose the step length \( a_k \) to satisfy one of the following three situations

\( \text{(S1)} \) \( f(x_k + a_k d_k) \leq f(x_k) - \sigma_1(r_k), \) where \( r_k = \frac{-g_k^T d_k}{\| d_k \|^2} \).

\( \text{(S2)} \) \( f(x_k + a_k d_k) \leq f(x_k) - \sigma_2(-a_k g_k^T d_k), \) and if \( a_k \| d_k \| \leq \min \{ a, \psi(r_k) \} \), then there exists \( b_k \) satisfying

\[
\nabla f(x_k + b_k a_k d_k)^T d_k \geq \beta g_k^T d_k
\]  
(8)
and \( \{ b_k \}, \{ x_k + b_k a_k d_k \} \) are bounded.

\( \text{(S3)} \) \( f(x_k + a_k d_k) \leq f(x_k) - \sigma_3(r_k) \| d_k \| \).

Inspired by the above definition 2 and (7), we introduce a new line search model, in which the step-size \( a_k \) is defined as
\[
f(x_k + a_k d_k) \leq f(x_k) - \min \{ \sigma_1(r_k \| d_k \|), \sigma_2(r_k), \sigma_3(r_k) \| d_k \| \}
\]  
(9)
where \( \sigma_1, \sigma_2 \) and \( \sigma_3 \) are forcing functions, and \( r_k = \frac{-g_k^T d_k}{\| d_k \|^2} \).

**Remark 1**  
Similarly to the proof of Theorem 2 in Ref[6], we can immediately deduce that the most popular step-size rules are special cases of the line search model (9), such as the Goldstein step-size rule, the Wolfe step-size rule and the Armijo step-size rule. Moreover, if we choose \( \sigma_1(t) = \mu_1 t, \sigma_2(t) = \mu_2 t^2 \) and \( \sigma_3(t) = \sigma_1(t) \), then (9) is transformed into (7).
Remark 2 Obviously, if \( a_k \) denotes the step-size computed by the general line search model (S1) or (S3), then \( a_k \) must satisfy the new line search rule (9). That is to say, the step-size \( a_k \) generated by (9) is easier to find from the theorem point of view.

Associated with the new line search model (9), we give a modified Perry-Shanno memoryless quasi-Newton type method (MPSMQN) as follows.

The MPSMQN method

Given an initial point \( x_0 \in \mathbb{R}^n \) and a symmetric positive definite matrix \( B_0 \in \mathbb{R}^{n \times n} \).

\[
\begin{align*}
    k := 0, \\
    \text{while } g_k = 0 \\
    \{ \\
        d_k = -B_k^{-1}g_k = H_k g_k, \\
        x_{k+1} = x_k + a_k d_k
    \} \quad \text{(10)}
\end{align*}
\]

Remark 3 In the above method, \( a_k \) is the step-size defined by (9), and \( B_k \) or \( H_k \) is updated by the formula (4) and (5), respectively.

3 Convergence analysis

In order to analyze the global convergence of our proposed method. We need the following assumptions.

Assumptions

H1. The level set \( L_0 = \{ x \in \mathbb{R}^n | f(x) \leq f(x_0) \} \) is contained in a bounded convex set \( D \).

H2. The function \( f(x) \) is twice continuously differentiable.

H3. There exist positive constants \( c_1 \) and \( c_2 \) such that

\[
    c_1 \| z \|^2 \leq z^T \nabla^2 f(x) z \leq c_2 \| z \|^2, \quad \forall x, z \in \mathbb{R}^n \quad \text{(11)}
\]

Lemma 4 Under Assumption H2 and H3, there exists a constant \( c_3 > 0 \) such that

\[
    \frac{\| y_k \|^2}{y_k^T s_k} \leq c_3 \quad \text{(12)}
\]

\[
    \frac{y_k^T s_k}{\| y_k \|^2} \leq \frac{1}{c_1} \quad \text{(13)}
\]

Proof. For the proof of inequality (12), see Ref[7]. Here we only prove the inequality (13). By using Assumption H2 and H3, we have

\[
    y_k^T s_k = \int_0^1 s_k^T \nabla^2 f(x_k + ts_k) s_k dt \geq c_1 \| s_k \|^2 \quad \text{(14)}
\]

From (12) and the Cauchy-Schwartz inequality, it follows that

\[
    \frac{y_k^T s_k}{\| y_k \|^2} \leq \frac{\| s_k \|^2}{y_k^T s_k} \leq \frac{1}{c_1} \quad \text{(15)}
\]

This proof is completed.

Remark 5 From Lemma 1, it follows

\[
    c_1 \leq \frac{\| y_k \|^2}{y_k^T s_k} \leq c_3 \quad \text{(16)}
\]

Lemma 6 Let the matrix \( B_k \) be defined by the formula (4). Then \( B_k \) is symmetric and positive definite, so is \( H_k \) defined by the formula (5).

Proof. We will prove

\[
    x^T B_k x > 0, \quad \forall x \neq 0 \quad \text{(17)}
\]

by induction.
Obviously, $B_0$ is symmetric and positive definite by the proposed method. We now suppose that (17) holds for some $k \geq 1$. Then by the update formula (4) we have

$$x^T B_{k+1} x = \frac{\|y_k\|^2}{y_k^T s_k} (x^T x - \frac{(s_k^T x)^2}{s_k^T s_k}) + \frac{(y_k^T x)^2}{y_k^T s_k}$$

(18)

From the Cauchy-Schwartz inequality, it immediately follows that

$$x^T x - \frac{(s_k^T x)^2}{s_k^T s_k} \geq 0$$

(19)

In addition, the second term in (18) is also nonnegative because of $s_k^T y_k > 0$ (see (14)). Therefore we obtain that

$$x^T B_{k+1} x \geq 0$$

(20)

In what follows, we must prove that at least one term in (18) is strictly larger than zero. Since $x/\|x\| = 0$, the inequality (19) becomes an equality if and only if $s_k$ is parallel to $x$, i.e., there exists a constant $\gamma/\|x\| = 0$ such that $x = \gamma s_k$. Thus

$$\frac{(y_k^T x)^2}{y_k^T s_k} = \gamma^2 y_k^T s_k > 0$$

(21)

which indicates that if the first term in (18) equals zero, the second term must be strictly larger than zero. Thus, for any $x/\|x\| = 0$, we always have $x^T B_{k+1} z > 0$.

In addition, the positive definiteness of $H_k$ is obvious, since it is the reverse form of $B_k$. This proof is completed. \qed

**Theorem 3** Suppose that Assumptions H1, H2 and H3 hold. Let $\{x_k\}$ be an infinitely sequence generated by the memoryless Perry-Shanno quasi-Newton type method with the rule (9). Then

$$\lim_{k \to +\infty} \|g_k\| = 0$$

(22)

**Proof.** Suppose on the contrary that there exist $\epsilon > 0$ and an infinite subset of $K$ of $\{0, 1, 2, \cdots\}$ such that

$$\|g_k\| \geq \epsilon, \forall k \in K$$

(23)

By using the update formula (4) and (5), we have

$$trace(B_k) = n \frac{\|y_{k-1}\|^2}{y_{k-1}^T s_{k-1}}$$

(24)

$$trace(H_k) = (n - 2) \frac{y_{k-1}^T s_{k-1} - 2 \|s_{k-1}\|^2}{y_{k-1}^T s_{k-1}}$$

(25)

Combining (13) with Lemma 1 yields

$$trace(B_k) \leq n c_3$$

(26)

$$trace(H_k) \leq \frac{n}{c_1}$$

(27)

Since

$$\frac{\|d_k\|^2}{g_k^T H_k g_k} = \frac{\|H_k g_k\|^2}{g_k^T H_k g_k} \leq trace(H_k) \leq \frac{n}{c_1}$$

(28)

which implies

$$\frac{\|d_k\|^2}{-g_k^T d_k} = \frac{\|d_k\|^2}{\|g_k\| \|d_k\| \cos \theta_k} = \frac{\|d_k\|}{\|g_k\| \cos \theta_k} \leq \frac{n}{c_1}$$

(29)

where

$$\cos \theta_k = \frac{-g_k^T d_k}{\|g_k\| \|d_k\|}$$
Note that
\[ \|d_k\| = \|H_k g_k\| \geq \|g_k\| \frac{\|g_k\|}{\text{tr}(B_k)} \geq \epsilon \frac{\epsilon}{n c_3}, \quad \forall k \in K \] (30)

From (29) and (30), we have
\[ r_k = -g_k^T d_k = \|g_k\| \cos \theta_k \geq \frac{c_1 \|d_k\|}{n} \geq \epsilon \frac{c_1}{n^2 c_3}, \quad \forall k \in K \] (31)
and
\[ -g_k^T d_k = r_k \|d_k\| \geq \epsilon \frac{c_1 \epsilon}{n^2 c_3} = \epsilon \frac{c_1^2}{n^2 c_3}, \quad \forall k \in K \] (32)

Since \( \sigma_1, \sigma_2 \) and \( \sigma_3 \) are forcing functions, then there exist \( \epsilon_1 > 0, \epsilon_2 > 0 \) and \( \epsilon_3 > 0 \) such that
\[ \sigma_1(-g_k^T d_k) \geq \epsilon_1 \]
\[ \sigma_2(r_k) \geq \epsilon_2 \]
\[ \sigma_3(r_k) \|d_k\| \geq \epsilon_3 \]
for any \( k \in K \). Thus
\[ \sum_{k \in K} \min\{\sigma_1(-g_k^T d_k), \sigma_2(r_k), \sigma_3(r_k) \|d_k\|\} \geq \sum_{k \in K} \min\{\epsilon_1, \epsilon_2, \epsilon_3\} = +\infty \] (33)

On the other hand, it follows from (9) that \( f(x_{k+1}) \leq f(x_k) \) for all \( k \), which, together with Assumption H1, implies that \( \lim_{k \to \infty} f(x_k) \) exists. Hence from (9) we can deduce that
\[ \sum_{k=0}^{+\infty} \min\{\sigma_1(-g_k^T d_k), \sigma_2(r_k), \sigma_3(r_k) \|d_k\|\} \leq f(x_0) - \lim_{k \to \infty} f(x_k) < +\infty \]
which implies
\[ \sum_{k \in K} \min\{\sigma_1(-g_k^T d_k), \sigma_2(r_k), \sigma_3(r_k) \|d_k\|\} < \infty \]
This is in contradiction with (33), which implies that the conclusion of Theorem 3 is true. This completes the proof. \( \blacksquare \)

4 Numerical experiments

To illustrate the feasibility of the proposed method, Algorithm MPSMQN was then implemented in Matlab 7.1 and run on a computer for four standard test functions as follows (see Ref [10]).

Test 1. Rosenbrock function.
\[ f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2. \]

Test 2. Wood function.
\[ f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3^2 - x_4)^2 + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1) \]

Test 3. Powell singular function.
\[ f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4. \]

\[ f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2. \]

Throughout the computational experiments, we choose the following forcing functions:
\[ \sigma_1(t) = \mu_1 t, \quad \sigma_2(t) = \mu_2 t^2 \text{ and } \sigma_3(t) = \sigma_1(t) \]
where the parameters used are \( \mu_1 = 10^{-4}, \mu_2 = 0.1 \). The stopping condition is \( \|g_k\| \leq 10^{-6} \).

To validate the MPSMQN method from a computational point of view, we compare it with the traditional Perry-Shanno memoryless quasi-Newton-type method (abbreviated as the PS method) with the following Wolf linesearch, which is one of the most popular line search rules.

\[
\begin{align*}
\{ f(x_k + a_k d_k) &\leq f(x_k) + \rho a_k g_k^T d_k \\
g(x_k + a_k d_k) &\geq \sigma g_k^T d_k
\end{align*}
\]

LINS homepage: http://www.nonlinearscience.org.uk/
Table 1: Numerical results for testing problems

<table>
<thead>
<tr>
<th>Test problem</th>
<th>n</th>
<th>MPSMQN</th>
<th>PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1.</td>
<td>2</td>
<td>55/131/111</td>
<td>61/205/123</td>
</tr>
<tr>
<td>Test 2.</td>
<td>4</td>
<td>161/361/323</td>
<td>202/480/405</td>
</tr>
<tr>
<td>Test 3.</td>
<td>4</td>
<td>99/223/199</td>
<td>152/345/305</td>
</tr>
<tr>
<td>Test 4.</td>
<td>2</td>
<td>39/106/79</td>
<td>45/122/91</td>
</tr>
</tbody>
</table>

where $0 < \rho < 0.5$, $\sigma \in (\rho, 1)$. The PS method is implemented in the same way.

Table 1 lists the numerical results which are given in the form of $k/k_f/k_g$, where $k$, $k_f$ and $k_g$ denote the number of iterations, function evaluations and gradient evaluations, respectively. Note that we don’t give the initial points, since they are the same as that in Ref [10].

Of course, we cannot draw some general conclusions from the rather limited numerical tests. Comparing the results given by our method with the PS method, however, our method is comparable to that in computational effort, which indicates that the proposed algorithm is effective in some sense. Further improvement is expected from more sophisticated implementation.

5 Conclusion

In this paper, we propose a new line search model for unconstrained optimization problem. By using this new model, we establish the strong global convergence properties of Perry-Shanno memoryless quasi-Newton type method. Numerical results show its feasibility.

Acknowledgements

This work is partially supported by the NSF of Hainan University (No. HD09XM87).

References