A TYPE OF MODIFIED BFGS ALGORITHM WITH RANK DEFECTS AND ITS GLOBAL CONVERGENCE IN CONVEX MINIMIZATION

GE REN-DONG and XIA ZUN-QUAN

School of Science
Dalian Nationalities University
Dalian 116600
P. R. China
e-mail: bgrbgg@163.com

Department of Applied Mathematics
Dalian University of Technology
Dalian 116024
P. R. China

Abstract

A modified BFGS algorithm to solve the unconstrained optimization of a convex function is presented in this paper, whose Hessian matrix of the minimum point is of rank defects. The idea of the algorithm is to give a modified part of the convex function to obtain an equivalent model, then simplify the model to obtain the modified BFGS algorithm. The global convergence property of the algorithm is proved in this paper. And compared with the Tensor algorithm, it is shown that this method is more efficient for solving unconstrained optimization, whose object function is of rank defects as the latter’s restriction \( \text{rank}(\nabla^2 f(x^*)) \leq n - 2 \).

2010 Mathematics Subject Classification: 90C56.

Keywords and phrases: nonconvex function, unconstrained programme, BFGS algorithm, locally superlinear convergence, Quasi-Newton method.

Received July 20, 2009
The question studied in this article is to find \( x^* \in R^n \) such that \( f(x^*) \leq f(x), \forall x \in D, f : R^n \rightarrow R \). By now, BFGS algorithm is the best one for solving unconstrained optimization. But, if preserved the algorithm’s convergence, the Hessian matrix of objective function must be positive definite in neighborhood of optimal point. For the Hessian matrix being semi-positive definite, people seldom put to use BFGS method instead of making use of trust region methods and conjugate gradient methods for solving unconstrained optimization. Schnabel and other researchers [2, 3, 9] get some accomplishment in applying Tensor methods to solve unconstrained optimization under the condition of \( \text{rank}(\nabla^2 f(x^*)) \leq n - 2 \). Furthermore, there is no evident to show the convergent rate of Tensor methods faster than linear convergence under the condition of local convergence. The following is the process of Tensor methods. Let \( x_c \) is the current iterate, \( d \in R^n \), \( f'(x_c) \), and \( f''(x_c) \) are the first and the second derivatives at \( x_c \). Schnabel and Frank constructed the following model;

\[
M_T(x_c + d) = f(x_c) + f'(x_c)d + \frac{1}{2} f''(x_c)d^2 + \frac{1}{6} T_c d^3 + \frac{1}{24} V_c d^4,
\]

where the Tensor terms \( T_c \in R^{n \times n \times n \times n} \), \( V_c \in R^{n \times n \times n \times n} \) at \( x_c \) are symmetric. \( V_c \) is the solution satisfying the following constrained problem

\[
\begin{aligned}
\min_{V_c \in R^{n \times n \times n \times n}} & \| V_c \|_F, \\
\text{s.t.} & \quad V_c s^4 = \beta.
\end{aligned}
\]

\( T_c \) is the solution satisfying the following constrained problem

\[
\begin{aligned}
\min_{T_c \in R^{n \times n \times n}} & \| T_c \|_F, \\
\text{s.t.} & \quad T_c s^3 = \alpha,
\end{aligned}
\]
the next iterate \( x_c + d \) satisfies
\[
\min_{d \in \mathbb{R}^n} \| M_T(x_c + d) \|_F,
\]
where \( \alpha \) and \( \beta \) are determined by the past iteration information. With regard to modified BFGS methods, on the discussion of non-convex function, there are some jobs in recent years. Without considering \( \nabla^2 f(x^*) \) being rank deficient, there is only the such result: \( \lim_{k \to +\infty} \inf \| \nabla g(x_k) \| = 0 \) (see [4] 2000, 2001). Although, Li and Fukushima had proved the superlinear convergence in modified BFGS algorithm of the minor problems of non-convex, the \( \nabla^2 f(x^*) \) needs to be nonsingular.

The paper presented a new method in the aspect of \( f(x) \) being convex function on \( D \) and \( \text{rank}(\nabla^2 f(x^*)) = n - s \), \( (0 < s << n) \). In Section 2, we will discuss the method in detail. In Section 3, we will give out the proof of the global convergence.

2. A New Algorithm

It is necessary to make the following two basic assumptions for discussing the establishment of new algorithm.

**Assumption 1.** \( f(x) \in C^3_D \).

**Assumption 2.** \( \nabla^2 f(x) \) is positive definite, \( x \neq x^* \).

Firstly, let \( \text{Null}(\nabla^2 f(x^*)) = \{ \mu_1, \mu_2, \cdots, \mu_s \} \) and \( U = (\mu_1, \mu_2, \cdots, \mu_s) \). We select the matrix \( P \) with full column rank such that \( P^T U \) is nonsingular. It is easy to see that \( \nabla^2 f(x^*) + P P^T \) is nonsingular. When \( x \) is around \( x^* \), we define a new function as follow.

\[
\hat{f}(x) = f(x) + \lambda \mu(x)^T \nabla^2 f(x) \mu(x) = f(x) + \lambda h(x),
\]

(2.1)

where \( \mu(x) \) satisfies
\[
(\nabla^2 f(x) + P P^T) \mu(x) = P q,
\]

(2.2)
and \( q \in R^s, q \neq 0 \). When \( x = x^* \), it can be proved that \( \mu = U(P^T U)^{-1} q \in \text{Null}(\nabla^2 f(x^*)) \). Notice that, for \( \forall d \in R^n \), we have

\[
f^*(x^* + td) \cdot \mu^2 = f^*(x^*) \mu + tf^*(x^*) \cdot \mu^2 d + \frac{t^2}{2} f^{(4)}(x^*) \cdot \mu^2 d^2 + o(t^3),
\]

where the definition of derivatives of \( f(x) \) can be seen [7], as \( f^*(x^*) \cdot \mu = \mu^T \nabla^2 f(x^*) \mu = 0 \). \( f^*(x^* + td) \cdot \mu^2 \geq 0 \) (when \( t \) is small enough). Consequently, we have from the formulas given above

\[
\begin{align*}
 f''(x^*) \cdot \mu^2 d &= 0, \\
f^{(4)}(x^*) \cdot \mu^2 d^2 &\geq 0, \quad \forall d \in R^n.
\end{align*}
\]

Notice that for any \( \mu \in \text{Null}(\nabla^2 f(x^*)) \), (2.3) all hold. In the following research, we need the following assumption.

**Assumption 3.** There exists a \( q \neq 0 \), as setting \( \mu = U(P^T U)^{-1} q \), such that for any \( v \in \text{Null}(\nabla^2 f(x^*)) \), \( f^{(4)}(x^*) \cdot \mu^2 v^2 > 0 \).

For the sake of convenience, let

\[
f'(x) = (f_1(x), f_2(x), \ldots, f_n(x)), f_i(x) \in R, i = 1, 2, \ldots, n.
\]

We have \( \nabla^2 f(x) = (\nabla f_1(x), \nabla f_2(x), \ldots, \nabla f_n(x))^T \). Also denote that

\[
B(x) = \nabla^2 f(x) + PP^T.
\]

As

\[
(B(x)\mu(x))' = B'(x)\mu(x) + B(x)\mu'(x),
\]

where

\[
B'(x)\mu(x) = [\nabla^2 f_1(x)\mu(x), \nabla^2 f_2(x)\mu(x), \ldots, \nabla^2 f_n(x)\mu(x)]^T = \nabla^3 f(x)\mu(x).
\]

We have

\[
\mu'(x) = -B(x)^{-1} B'(x)\mu(x) = -B^{-1}(x)\nabla^3 f(x)\mu(x).
\]
According to (2.5) and (2.6), we have

\[ (\mu(x)^T \nabla^2 f(x)\mu(x))^\prime = \mu^T(x)\nabla^3 f(x)\mu(x) - 2\mu^T(x)\nabla^2 f(x)B^{-1}(x)\nabla^3 f(x)\mu(x). \]

Consequently,

\[ \hat{f}'(x) = f'(x) + \lambda h'(x) \]

\[ = f'(x) + \lambda \left[ (\nabla^3 f(x)\mu(x))^T \mu(x) - 2(\nabla^3 f(x)\mu(x))^T B^{-1}(x)\nabla^2 f(x)\mu(x) \right], \]

\[ \nabla \hat{f}(x) = \nabla f(x) + \lambda \left[ (\nabla^3 f(x)\mu(x))^T \mu(x) - 2(\nabla^3 f(x)\mu(x))^T B^{-1}(x)\nabla^2 f(x)\mu(x) \right], \]

because of the symmetry of \( \nabla^3 f(x)\mu(x) \), we have

\[ \nabla \hat{f}(x) = \nabla f(x) + \lambda \nabla h(x) \]

\[ = \nabla f(x) + \lambda \left[ (\nabla^3 f(x)\mu(x))^T \mu(x) - 2(\nabla^3 f(x)\mu(x))^T B^{-1}(x)\nabla^2 f(x)\mu(x) \right]. \quad (2.7) \]

From (2.3), it can be obtained easily \( \hat{f}'(x^*) = 0 \). Similarly, because

\[ (\nabla^3 f(x)\mu(x)\mu(x))^\prime = \begin{bmatrix} \mu^T(x)\nabla^2 f_1(x)\mu(x) \\ \mu^T(x)\nabla^2 f_2(x)\mu(x) \\ \vdots \\ \mu^T(x)\nabla^2 f_n(x)\mu(x) \end{bmatrix} \]

\[ = \nabla^4 f(x)\mu(x)\mu(x) + 2\nabla^3 f(x)\mu(x)\mu(x) \]

where \( \nabla^4 f(x) \) is defined by (2.6) and the way of definition of \( \nabla^3 f(x) \). Let

\[ t(x) = B^{-1}(x)\nabla^2 f(x)\mu(x), \]

then

\[ (\nabla^3 f(x)\mu(x)B^{-1}(x)\nabla^2 f(x)\mu(x))^\prime = \begin{bmatrix} \mu^T(x)\nabla^2 f_1(x)\mu(x) \\ \mu^T(x)\nabla^2 f_2(x)\mu(x) \\ \vdots \\ \mu^T(x)\nabla^2 f_n(x)\mu(x) \end{bmatrix} \]
Notice that, for any $v \in \text{Null}(\nabla^2 f(x^*))$, $v^T \nabla^3 f(x^*) \mu = 0$, and $t(x^*) = 0$, we have from (2.8)-(2.10),

$$v^T \nabla^2 h(x^*) v = v^T \nabla^4 f(x^*) \mu(x^*) \mu(x^*) v > 0.$$ 

According to the conclusions given above, when let $\lambda > 0$, the Hessian matrix of

$$\hat{f}(x) = f(x) + \lambda \mu(x) \nabla^2 f(x) \mu(x),$$

$\nabla^2 \hat{f}(x^*)$ is definite. For simplicity, we denote $x$ as $x_k$ and let $\mu_k = \mu(x_k)$, $t_k = t(x_k)$. When $x_k$ is an approximate value of $x^*$ and $h_k = O(\min(\|\nabla f(x_k)\|, \|x_{k+1} - x_k\|))$. Because of

$$\mu^T(x_k) \nabla^2 f(x_k) \mu(x_k) = \frac{f_k(x_k + h_k \mu_k) - 2f_k(x_k) + f_k(x_k - h_k \mu_k)}{h_k^2} + O(h_k^2),$$

(2.11)

and

$$t(x_k)^T \nabla^2 f(x_k) \mu(x_k) = t_k^T \left( \frac{\nabla^2 f_k(x_k + h_k \mu_k) - \nabla^2 f_k(x_k)}{h_k} \right) + O(h_k \|x_k - x^*\|),$$

(2.12)

we have from (2.7),

$$\nabla h(x_k) = [\nabla^3 f(x_k) \mu(x_k)] \mu(x_k) - 2\nabla^3 f(x_k) \mu(x_k) B^{-1}(x_k) \nabla^2 f(x_k) \mu(x_k)$$

$$= \nabla^3 f(x_k) \mu(x_k) \mu(x_k) - 2\nabla^3 f(x_k) \mu(x_k) t(x_k)$$
\[
\begin{align*}
\mu^T x_k \nabla^2 f_1(x_k) \mu(x_k) & =\mu^T x_k \nabla^2 f_2(x_k) \mu(x_k) \quad \cdots \quad \mu^T x_k \nabla^2 f_n(x_k) \mu(x_k) \\
\mu^T x_k \nabla^2 f_1(x_k) \mu(x_k) & =\mu^T x_k \nabla^2 f_2(x_k) \mu(x_k) \quad \cdots \quad \mu^T x_k \nabla^2 f_n(x_k) \mu(x_k)
\end{align*}
\]

\[
\begin{align*}
\begin{bmatrix}
f_1(x_k + h_k \mu_k) - 2f_1(x_k) + f_1(x_k - h_k \mu_k) \\
f_n(x_k + h_k \mu_k) - 2f_n(x_k) + f_n(x_k - h_k \mu_k)
\end{bmatrix}
& = \\
\begin{bmatrix}
h_k^2 \\
h_k^2
\end{bmatrix}
\begin{bmatrix}
\frac{\nabla f_1(x_k + h_k \mu_k) - \nabla f_1(x_k)}{h_k} \\
\frac{\nabla f_n(x_k + h_k \mu_k) - \nabla f_n(x_k)}{h_k}
\end{bmatrix}
\end{align*}
\]

\[
\begin{align*}
\begin{bmatrix}
O(h_k^2) \\
O(h_k^2)
\end{bmatrix}
& = \\
\begin{bmatrix}
O(h_k \|x_k - x^*\|) \\
O(h_k \|x_k - x^*\|)
\end{bmatrix}
\end{align*}
\]

Let
\[
g_k = \nabla f(x_k) + \lambda \left\{ \frac{\nabla f(x_k + h_k \mu_k) - 2\nabla f(x_k) + \nabla f(x_k - h_k \mu_k)}{h_k^2} \\
- \frac{2 \nabla^2 f(x_k + h_k \mu_k) - \nabla^2 f(x_k)}{h_k} \right\}. 
\]

(2.13)

It can be proved easily that
\[
\nabla f(x_k) - g_k = \lambda (O(h_k^2) + O(h_k \|x_k - x^*\|)). 
\]

(2.14)

Meanwhile, we can obtain from Assumption 1 and (2.11)-(2.13) that for any \( x_k \in D \), there exits a positive number \( e \) such that
\[
\|\nabla f(x_k) - g_k\| \leq \lambda e \|\nabla f(x_k)\| (\|\nabla f(x_k)\| + \|x_k - x^*\|). 
\]

(2.15)
A modified BFGS algorithm for solving unconstrained optimization is defined by the following steps.

**Algorithm 2.1 (GMBFGS).** Choose initial point $x_0$ and initial positive-definite matrix $B_0$. Choose positive numbers $\epsilon, \delta, \sigma_1, \sigma_2$ such that $0 < \sigma_1 < \sigma_2 < 1$. Compute $\hat{g}_0 = \nabla f(x_0)$, and let $k = 0$.

**Step 1.** Solving the following linear equation to get $p_k$:

$$B_k p_k + \hat{g}_k = 0.$$  

**Step 2.** Find a stepsize $\alpha_k > 0$ satisfying the Wolfe-type linear search conditions:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T p_k,$$

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq \sigma_2 \nabla f(x_k)^T p_k,$$

with $0 < \sigma_1 < \sigma_2 < 1$. Moreover, as long as $\alpha_k = 1$ satisfies the sufficient decrease condition of Wolfe-type linear search conditions, we take $\alpha_k = 1$.

**Step 3.** Compute $x_{k+1} = x_k + \alpha_k p_k$ and $\|\nabla f(x_{k+1})\|$. If $\|\nabla f(x_{k+1})\| \leq \varepsilon$, stop.

**Step 4.** Compute $\gamma_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ and $\eta_k = \frac{\gamma_k^T s_k}{s_k^T s_k} + \|\nabla f(x_k)\|$. If $\eta_k > \delta$, then (let $y_k = \gamma_k + \|\nabla f(x_k)\| s_k$ and $\hat{g}_k = \nabla f(x_{k+1})$), go to **Step 6**.

**Step 5.** Computing $\mu_{k+1} = \mu(x_{k+1})$ given from (2.2), $t_{k+1} = t(x_{k+1})$ given from (2.9), and $g_{k+1}, y_k = g_{k+1} - g_k$ given from (2.13), let $\hat{g}_k = g_{k+1}$. 

$$\lim_{k \to 0} g_k = \nabla f(x_k). \quad (2.16)$$
Step 6. Update

\[ B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}. \]

Step 7. \( k = k + 1 \), go to Step 1.

In general, numerical experiments show that Step 5 of the algorithm is carried out only a few times before the accuracy has been met.

3. Convergence Analysis

In the following, \( \| \cdot \| \) denotes Euclidean norm,

Before the proof of convergence, we restate the basic assumptions:

Assumption H1: \( f(x): \mathbb{R}^n \to \mathbb{R}^1 \) is at last fourth continuously differentiable in the open convex \( D \subseteq \mathbb{R}^n \) and is bounded below in \( D \) containing the bounded level set

\[ \mathbb{L} = \{ x \mid f(x) \leq f(x_0) \}. \]

Assumption H2: \( \text{rank}(\nabla^2 f(x^*)) = n - s \), \( 0 < s << n \).

Assumptions H3: There exists an \( x^* \in D \) such that \( \nabla f(x^*) = 0 \), and there exists a \( q \in \mathbb{R}^s \) such that for any \( v \in \text{Null}(\nabla^2 f(x^*)) \), \( \nabla^4 f(x^*) \cdot \mu^2 \cdot v^2 > 0 \), where \( \mu = U(P^T U)^{-1} q \).

Assumption H4: \( \nabla f(x^*), \nabla^2 \hat{f}(x^*) \) satisfies Lipschitz condition in \( D \), i.e., there exists a \( \gamma \) such that for \( x, \bar{x} \in D \),

\[ \| \nabla f(x) - \nabla f(\bar{x}) \| \leq \gamma \| x - \bar{x} \|; \]

\[ \| \nabla^2 \hat{f}(x) - \nabla^2 \hat{f}(\bar{x}) \| \leq \gamma \| x - \bar{x} \|. \]

Assumption H5: There is a positive \( B \) such that for \( B_k \)

\( (k = 1, 2, \cdots, ) \)
Assumption H₆: There is a positive B such that for αₖ (k = 1, 2, ...,)
\[ \min_k |\alpha_k| \geq \alpha^* > 0. \]

Assumption H₇: There is a positive ρ such that for \( L \in x(0) \)
\[ \|\nabla f(x_k)\| \geq \rho \|\nabla h(x_k)\|. \]

In order to discuss the general global convergence properties of GMBFGS algorithm, at first, we want to introduce some necessary lemmas.

**Lemma 3.1.** When \( x, x + d, y, z \in D \), by basic assumptions, then the following two inequality is hold.

1. \( \|\nabla f(x + d) - \nabla f(x) - \nabla^2 f(x)d\| \leq \frac{\gamma}{2} \|d\|^2. \)
2. \( \|\nabla f(x) - \nabla f(y) - \nabla^2 f(z)(x - y)\| \leq \gamma \sigma(x, y)\|x - y\|, \) where \( \sigma(x, y) = \max\{\|x - z\|, \|y - z\|\}. \)

**Lemma 3.2.** When \( A, B \in \mathbb{R}^{n,n} \), A is inversible, \( \|A^{-1}\| \leq \alpha, \|A - B\| \leq \beta \), and \( \alpha \cdot \beta < 1 \), then B is also inversible. Moreover, we have

\[ \|B^{-1}\| \leq \frac{\alpha}{1 - \alpha \beta}. \] (3.1)

Let \( \tilde{G}(x_k) = \int_0^1 \nabla^2 \hat{f}(x_k + \delta(x^* - x_k))d\delta \), when \( \delta \) is small enough and
\[ \|\tilde{G}(x_k) - \nabla^2 \hat{f}(x^*)\| \leq \delta. \]

By Lemma 3.2, it is easy to see that \( \tilde{G}(x_k) \) is inversible and the inverse of \( \tilde{G}(x_k) \) is bounded.

**Lemma 3.3.** There exists a \( \varepsilon_2 > 0 \) and two constant; \( L_1, L_2 > 0 \) such that for \( \|x_k - x^*\| \leq \varepsilon_2 \), one has
A TYPE OF MODIFIED BFGS ALGORITHM …

(1) \[ \| \mu(x_k) - \mu(x^*) \| \leq L_1 \| x_k - x^* \|. \]

(2) \[ \| \nabla f(x_k) - g_k \| \leq L_2 \| x_k - x^* \|^2, \quad (3.2) \]

where \( h_k = O(\| \nabla f(x_k) \|) \).

**Proof.** Let \( \hat{\beta} = \| (\nabla^2 f(x^*) + PP^T)^{-1} \| \). By basic Assumption \( H_3 \) and Lemma 3.2, if \( \varepsilon_2 > 0 \) is small enough such that \( \hat{\beta} \gamma \varepsilon_2 < 1 \), then for \( \| x_k - x^* \| < \varepsilon_2 \), one has

\[ \| (\nabla^2 f(x_k) + PP^T)^{-1} \| \leq \frac{\hat{\beta}}{1 - \hat{\beta} \gamma \varepsilon_2}. \]

Also by (2.2), let \( L_1 = \frac{\hat{\beta}^2 \gamma}{1 - \hat{\beta} \gamma \varepsilon_2} \), we have

\[ \| \mu(x_k) - \mu(x^*) \| = \| (\nabla^2 f(x_k) + PP^T)^{-1} Pq - (\nabla^2 f(x^*) + PP^T)^{-1} Pq \| \]

\[ \leq \| (\nabla^2 f(x) + PP^T)^{-1} - (\nabla^2 f(x^*) + PP^T)^{-1} \| \| Pq \| \]

\[ \leq \hat{\beta} \| (\nabla^2 f(x_k) + PP^T)^{-1} \| \| \nabla^2 f(x_k) - \nabla^2 f(x^*) \| \]

\[ \leq \frac{\hat{\beta}^2 \gamma}{1 - \hat{\beta} \gamma \varepsilon_2} \| x_k - x^* \| = L_1 \| x_k - x^* \|. \]

Further, by (2.13), (2.14), and conclusion (1) of the lemma, it can be proved easily that the conclusion (2) is hold.

In the following, we discuss the general global convergence properties. We know from the discussion of the above section that, under basic Assumptions \( H_1 \) and \( H_3 \), \( \nabla^2 \hat{f}(x^*) \) is positive definite. By the basic Assumption \( H_4 \), there is a neighborhood \( U_0(x^*) = \{ x \| x - x^* \| \leq \varepsilon_0 \} \) of
\( x^* \) such that \( \nabla^2 \hat{f}(x) \) is uniformly positive definite for all \( x \in U_0(x^*) \).

Therefore, there is constants \( M > 0, m > 0 \) such that for all \( x \in U(x^*) \)

\[
\|\nabla \hat{f}(x)\| = \|\nabla \hat{f}(x) - \nabla \hat{f}(x^*)\| \geq m\|x - x^*\|.
\] (3.3)

\[
M\|d\|^2 \geq d^T \nabla^2 \hat{f}(x)d \geq m\|d\|^2.
\] (3.4)

**Lemma 3.4.** There exists two positive constants \( \delta, R_3 \) such that for \( \|\nabla f(x_k)\| \leq \delta \), one has

\[
\|x_k - x^*\| \leq R_3\|x_{k+1} - x_k\|.
\] (3.5)

**Proof.** At first, from the strick convex property of \( f(x) \), we have that when \( \|f(x_k)\| \to 0, x_k \to x^* \). Then, when \( \delta \) is small enough and \( \|\nabla f(x_k)\| \leq \delta \),

\[
\|x_k - x^*\| \leq \epsilon_2.
\]

Also by the Algorithm 2.1, Assumption \( H_6 \), and Lemma 3.3, we have

\[
x_{k+1} - x_k = -\alpha_k B_k^{-1} \hat{g}_k = \alpha_k B_k^{-1}(\nabla \hat{f}(x_k) - \hat{g}_k) - \alpha_k B_k^{-1}(\nabla \hat{f}(x_k) - \nabla \hat{f}(x^*))
\]

\[
= [-\alpha_k B_k^{-1} \tilde{G}(x_k) + O(\|x_k - x^*\|)](x_k - x^*).
\]

Next, when \( \delta \) is small enough, then \( \|\tilde{G}(x_k) - A\| \leq \delta' \) with \( \alpha \delta' < 1 \). By Lemma 3.2 and Assumption \( H_5 \), \( B_k^{-1} \) and \( \tilde{G}(x_k) \) are invertible. Hence, 

\[
- B_k^{-1} \tilde{G}(x_k) + O(\|x_k - x^*\|) \]

is invertible and their inverse are bounded.

The bound is denoted by \( R_3 \). Hence, one has

\[
\|x_k - x^*\| \leq R_3\|x_{k+1} - x_k\|.
\] (3.6)

**Lemma 3.5.** There exists a real number \( 0 < r_3 < 1 \) such that, when \( \|\nabla f(x_k)\| < \delta \) and \( s_k \neq 0, y_k \neq 0 \), then
A TYPE OF MODIFIED BFGS ALGORITHM …

(1) \( \|x_{k+1} - x^*\| \leq r_3 \|x_k - x^*\| \).

(2) \( 2s_k^T \tilde{G}(x_k)s_k \geq s_k^T y_k \geq 1 / 2s_k^T \tilde{G}(x_k)s_k \).

(3.7)

**Proof.** At first, when \( \|\nabla f(x_k)\| \leq \delta \),

\[
x_{k+1} - x^* = x_k - x^* - \alpha_k B_k^{-1} \tilde{g}_k
\]

\[
= x_k - x^* + \alpha_k B_k^{-1} (\nabla \hat{f}(x_k) - \tilde{g}_k) - \alpha_k B_k^{-1} (\nabla \hat{f}(x_k) - \nabla \hat{f}(x^*))
\]

Next, we set

\[
r_3 = 1 + B \ell_2 \varepsilon_0 + B r.
\]

And let \( \delta \) in Lemma 3.4 is small enough such that when \( \|\nabla f(x_k)\| < \delta \),

\[
\|x_k - x^*\| \leq \varepsilon_2 / r_2,
\]

then according to Lemma 3.2, one has

\[
\|x_{k+1} - x^*\| \leq \|x_k - x^*\| + B \|\nabla \hat{f}(x_k) - \tilde{g}_k\| + B \|\nabla \hat{f}(x_k) - \nabla \hat{f}(x^*)\|
\]

\[
\leq \|x_k - x^*\| + B L_2 \|x_k - x^*\|^2 + B \|\nabla \hat{f}(x_k) - \nabla \hat{f}(x^*)\|
\]

\[
\leq r_3 \|x_k - x^*\| \leq \varepsilon_2.
\]

(3.8)

Consequently, by Lemma 3.3,

\[
\|\tilde{g}_{k+1} - \nabla \hat{f}(x_{k+1})\| \leq L_2 \cdot \|x_{k+1} - x^*\|^2 \leq L_2 \cdot r_3 \cdot \|x_k - x^*\|^2.
\]

Furthermore, when \( \varepsilon_4 \) is small enough and \( \|x_k - x^*\| \leq \varepsilon_4 \), by Lemma 3.4,

\[
s_k^T \cdot y_k = s_k^T (\tilde{g}_{k+1} - \tilde{g}_k)
\]

\[
= s_k^T (\nabla \hat{f}(x_{k+1}) - \nabla \hat{f}(x_k)) - s_k^T (\nabla \hat{f}(x_{k+1}) - \tilde{g}_{k+1}) + s_k^T (\nabla \hat{f}(x_k) - \tilde{g}_k)
\]

\[
= s_k^T \cdot \int_0^1 \nabla^2 \hat{f}(x_k + \theta(x_{k+1} - x_k))d\theta \cdot s_k + O(\|x_k - x^*\|^3)
\]

\[
= s_k^T \cdot \int_0^1 \nabla^2 \hat{f}(x_k + \theta(x_{k+1} - x_k))d\theta \cdot s_k + O(\|x_{k+1} - x_k\|^3),
\]
By Assumption H6, it is easy to know that there exists a positive number $\mu$ such that for any $\nabla f(x_k)$, then
\[
\nabla f(x_k)^T B_k^{-1} \nabla f(x_k) \geq \mu \|\nabla f(x_k)\|^2.
\]
And by Assumption H7,
\[
\nabla f(x_k)^T B_k^{-1} \nabla h(x_k) \leq B \|\nabla f(x_k)\| \|\nabla h(x_k)\| \leq B / p \|\nabla f(x_k)\|^2.
\]
When $\|\nabla f(x_k)\| \leq \delta$ and we set $\lambda \leq \frac{\mu \rho}{2B(1 + \rho\delta(\delta + \epsilon))}$, then from (2.15), and the proof of Lemma 3.4,
\[
\nabla f(x_k)^T p_k = -\nabla f(x_k) B_k^{-1} \hat{g}_k
\]
\[
= -\nabla f(x_k) B_k^{-1} \nabla \hat{f}(x_k) + \nabla f(x_k) B_k^{-1} (\nabla \hat{f}(x_k) - \hat{g}_k)
\]
\[
\leq -(\mu - \lambda B / \rho - \lambda B \epsilon(\delta + \epsilon + 2)) \|\nabla f(x_k)\|^2 < 0. \quad (3.9)
\]
Hence, $p_k$ is a descent direction. By the Step 4 of Algorithm 2.1, (3.4), and Lemma 3.4, there exists a constant $c > 0$ such that for any $k \geq 1$,
\[
\frac{y_k^T s_k}{s_k^T s_k} \geq c. \quad (3.10)
\]
By the Wolfe-type condition, for any $x_k \in L$, $\|\nabla f(x_k)\|$, $\|\nabla^2 f(x_k + ts_k)\|$, $\|\nabla^2 \hat{f}(x_k + ts_k)\|$ are all bounded. Moreover, there exists a constant $C > 0$ such that, from (3.4),
\[
\frac{y_k^T y_k}{y_k^T s_k} \leq C. \quad (3.11)
\]

**Lemma 3.6 (Zoutendijk's Theorem).** Consider any iteration of form $x_{k+1} = x_k + \alpha_k p_k$, where $p_k$ is a descent direction and $\alpha_k$ satisfies the
Wolfe conditions. Suppose that $f$ is bounded below in $\mathbb{R}^n$ and $f$ is continuously differentiable in an open set $N$ containing the level set $L = \{x | f(x) \leq f(x_0)\}$, where $x_0$ is the starting point of the iteration. Assume also that the gradient $\nabla f$ is Lipschitz continuous on $N$, then

$$\sum_{k \geq 0} \cos^2 \theta_k \|\nabla f_k\| < \infty,$$

where

$$\cos \theta_k = -\frac{\nabla f_k^T p_k}{\|\nabla f_k\| \|p_k\|}.$$

**Theorem 3.1** (Global Convergence Theorem). Let $B_0$ is any symmetric positive-definite initial matrix, and let $x_0$ be a starting point for which Assumptions $H_0 \cdot H_6$ is satisfied. Then, the sequence $\{x_k\}$ generated by Algorithm 2.1 converges to the minimizer $x^*$ of $f$.

**Proof.** Let us define

$$c_k = \frac{y_k^T s_k}{s_k^T s_k}, \quad C_k = \frac{y_k^T y_k}{y_k^T s_k},$$

and notice from (3.10) and (3.11),

$$c_k \geq c, \quad C_k \leq C.$$

By computing the trace of the BFGS approximation, we obtain that

$$\text{trace}(B_{k+1}) = \text{trace}(B_k) - \frac{\|B_k s_k\|^2}{s_k B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k}.$$  \hfill (3.14)

We can also show from the determinant property of BFGS, update of $B_k$ that

$$\det(B_{k+1}) = \det(B_k) \frac{y_k^T s_k}{s_k^T B_k s_k}.$$  \hfill (3.15)
Let us define
\[
\cos \hat{\theta}_k = \frac{s_k^T B_k s_k}{\|s_k\| \|B_k s_k\|}, \quad q_k = \frac{s_k^T B_k s_k}{s_k^T s_k},
\] (3.16)
so that \( \theta_k \) is the angle between \( s_k \) and \( B_k s_k \). We then obtain that
\[
\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} = q_k \cos^2 \hat{\theta}_k.
\] (3.17)

In addition, we have from (3.12) that
\[
\det(B_{k+1}) = \det(B_k) \frac{y_k^T s_k}{s_k^T s_k} \frac{s_k^T s_k}{s_k^T B_k s_k} = \det(B_k) \frac{m_k}{q_k}.
\] (3.18)

We now combine the trace and determinant by introducing the following function of a positive-definite matrix \( B \):
\[
\psi(B) = \text{trace}(B) - \ln(\det(B)),
\] (3.19)
where \( \ln(\cdot) \) denotes the natural algorithm. It is not difficult to show that \( \psi(B) > 0 \). By using (3.12) and (3.14)-(3.19), we have that
\[
\psi(B_{k+1}) = \psi(B_k) + C_k - \frac{q_k}{\cos^2 \hat{\theta}_k} - \ln c_k + \ln q_k
\]
\[
= \psi(B_k) + (C_k - \ln c_k - 1) + \left[ 1 - \frac{q_k}{\cos^2 \hat{\theta}_k} + \ln \frac{q_k}{\cos^2 \hat{\theta}_k} \right] + \ln \cos^2 q_k.
\] (3.20)

Now, since the function \( h(t) = 1 - t + \ln t \leq 0 \) for all \( t > 0 \), the term inside the square brackets is non-positive, and thus from (3.13) and (3.20), we have
\[
0 < \psi(B_{k+1}) \leq \psi(B_1) + c k + \sum_{j=1}^{k} \ln \cos^2 \hat{\theta}_j,
\] (3.21)
where we can assume that constant \( e = C - \ln c - 1 \) to be positive, without loss of generally.
Let us proceed by contradiction and assume that \( \cos \theta_j \to 0 \). Then, there exists \( J_1 > 0 \) such that for all \( j > J_1 \), we have
\[
\ln \cos^2 \hat{\theta}_j < -2e.
\]
Using this inequality in (3.21), we find the following relation to \( e \) true for all \( k > J_1 \):
\[
0 < \psi(B_{k+1}) \leq \psi(B_1) + \sum_{j=1}^{J_1} \ln \cos^2 \hat{\theta}_j + 2eL_1 - ek.
\]
However, the right-hand-side is negative for large \( k \), giving a contradiction. Then, there exists a subsequence of indices \( \{j_k\} \) such that \( \{ \cos \hat{\theta}_{j_k} \} \geq \zeta > 0 \). Furthermore, we have from (3.16), Algorithm 2.1, and (2.16),
\[
\lim_{\lambda \to \infty} \cos \hat{\theta} = \lim_{\lambda \to \infty} -\frac{p_k^T \hat{g}_k}{\|p_k\| \|\hat{g}_k\|} = \lim_{\lambda \to \infty} -\frac{p_k^T \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} = \cos \theta \geq \zeta > 0.
\]
By Assumptions H_1 – H_6 and Zoutendijk’s theorem, this limit implies that
\[
\lim \inf \|\nabla f(x_k)\| \to 0.
\]
Since, the property of strict convex of \( f \), the latter limit is enough to prove that
\[
x_k \to x^*.
\]

4. Numerical Examples

\[
f(x_1, x_2) = x_1^4 + x_1^2 + x_2^4,
\]
we have
\[
\nabla f(x) = [4x_1^3 + 2x_1, 4x_2^3]^T, \quad \nabla^2 f(x) = \begin{bmatrix} 12x_1^2 + 2 & 0 \\ 0 & 12x_2^2 \end{bmatrix}.
\]
In the process of this algorithm realization, we set
\[
\lambda = 0.0001, \ \delta = 10^{-5}, \ \sigma_1 = 0.2, \ \sigma_2 = 0.9, \ \varepsilon = 10^{-12}.
\]
The initial iteration value
\[
x = (1, 1)^T.
\]
The random matrix and random vector obtained by the Matlab command \texttt{rand( )} are
\[
P = \begin{bmatrix}
0.41027020699095 & 0.05789130478427 \\
0.89364953091353 & 0.35286813221700
\end{bmatrix},
\]
\[
q = [0.81316649730376 \ 0.00986130066092]^T,
\]
when carrying out the 38-step, the result becomes
\[
\|g(x_k)\| = 9.419483961805627e - 013,
\]
\[
x = 1.0e - 004 \cdot [0.00000000001502 \ 0.61752560551708]^T,
\]
at the same time, Step 5 of the algorithm is only carried out 2 times. This shows that the new algorithm is very effective.

References


