Journal of Mathematical Research & Exposition
May, 2009, Vol. 29, No. 3, pp. 423–433
DOI:10.3770/j.issn:1000-341X.2009.03.006
Http://jmre.dlut.edu.cn

Derivation and Global Convergence for Memoryless Non-quasi-Newton Method

JIAO Bao Cong[1], *YU Jing Jing*[1,2], *CHEN Lan Ping*[1]
(1. School of Mathematical Sciences, Capital Normal University, Beijing 100037, China;
2. Department of Electrical Engineering, Qingdao Harbor Vocational Technology College, Shandong 266404, China)
(E-mail: jiaobc3093@126.com; yujingjing9204@163.com; chenlanp@mail.cnu.edu)

Abstract In this paper, a new class of memoryless non-quasi-Newton method for solving unconstrained optimization problems is proposed, and the global convergence of this method with inexact line search is proved. Furthermore, we propose a hybrid method that mixes both the memoryless non-quasi-Newton method and the memoryless Perry-Shanno quasi-Newton method. The global convergence of this hybrid memoryless method is proved under mild assumptions. The initial results show that these new methods are efficient for the given test problems. Especially the memoryless non-quasi-Newton method requires little storage and computation, so it is able to efficiently solve large scale optimization problems.

Keywords memoryless non-quasi-Newton method; Wolfe line search; global convergence.

Document code A
MR(2000) Subject Classification 65K10; 90C30
Chinese Library Classification O174.13; O221.2

1. Introduction

The problem we consider is unconstrained optimization calculation:

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

where $f: R^n \to R, f \in C^2$.

It is known that the memoryless quasi-Newton methods have been successfully used for solving problem (1.1). The memoryless quasi-Newton methods were originated with the work of Perry[1] and Shanno([2,3]) in the 1970s, and have been developed and studied then by many authors: Perry[1], Powell[4], and Dai[9] et al.. However, it still remains unanswered for the general objective functions[2,5,6]. At the same time Huang(1970) proposed a class of update formulas, where the updated matrix need not satisfy quasi-Newton equation, only need the generated search directions are conjugacy when the method is used for the convex quadratic functions. This implies that the method possesses property of quadratic termination. In 1991, Yuan[7] proposed a quasi Newton method. Zhao and Duan[8] established a non-quasi-Newton

Received date: 2007-05-26; **Accepted date**: 2008-07-06
Foundation item: the National Natural Science Foundation of China (No. 60472071); the Science Foundation of Beijing Municipal Commission of Education (No. KM200710028001).

equation and proposed a pseudo-Newton method in 1996. Chen and Jiao[10] proposed a new non-quasi-Newton method in 1997. In this paper we derive a new class of memoryless formula from non-quasi-Newton equation, in which the update matrices are symmetric positive definite, and prove that the method with inexact line search converges globally. Numerical experiments indicate that it is able to efficently solve larger scale optimization problems.

2. Derivation of memoryless non-quasi-Newton method

Suppose that the objective function $f(x) \in C^2$, $g(x) = \nabla f(x)$, $g_k = g(x_k)$, $f_k = f(x_k)$, $\gamma_k = g_{k+1} - g_k$ and $\delta_k = x_{k+1} - x_k$, when $\|\delta_k\|$ is sufficiently small ($\|\cdot\|$ denotes its Euclidean norm). Hesse matrix G_k for x_k possesses the property:

$$R_k \approx \frac{1}{2}\delta_k^{\mathrm{T}} G_k \delta_k, \tag{2.1}$$

where $R_k = f_{k+1} - f_k - g_k \mathrm{T} \delta_k$. Specially, the above formula equality holds true strictly for quadratic functions. Consider Hestenes-Stiefel conjugate gradient method iteration formula:

$$x_{k+1} = x_k + \lambda_k d_k, \tag{2.2}$$

$$d_1 = -g_1,$$

$$d_{k+1} = -g_{k+1} + \frac{\gamma_k^{\mathrm{T}} g_{k+1}}{\gamma_k^{\mathrm{T}} d_k} d_k, \quad k \geq 2, \tag{2.3}$$

where λ_k is a search steplength. Notice that (2.3) can be written as the following form,

$$d_{k+1} = -(I - \frac{d_k \gamma_k^{\mathrm{T}}}{\gamma_k^{\mathrm{T}} d_k}) g_{k+1},$$

where I is a unit matrix. Since $\delta_k = x_{k+1} - x_k = \lambda_k d_k$, we have

$$\frac{d_k \gamma_k^{\mathrm{T}}}{\gamma_k^{\mathrm{T}} d_k} = \frac{\delta_k \gamma_k^{\mathrm{T}}}{\gamma_k^{\mathrm{T}} \delta_k},$$

$$d_{k+1} = -(I - \frac{\delta_k \gamma_k^{\mathrm{T}}}{\gamma_k^{\mathrm{T}} \delta_k}) g_{k+1}. \tag{2.4}$$

Denote $H_{k+1} = I - \frac{\delta_k \gamma_k^{\mathrm{T}}}{\gamma_k^{\mathrm{T}} \delta_k}$. Then

$$d_{k+1} = -H_{k+1} g_{k+1}.$$

To make H_{k+1} inherit the positive definiteness of H_k, we update H_{k+1} by adding an update term. For convenience, the updated matrix is still denoted by H_{k+1}. Hence

$$H_{k+1} = I - \frac{\gamma_k \delta_k^{\mathrm{T}} + \delta_k \gamma_k^{\mathrm{T}}}{\delta_k^{\mathrm{T}} \gamma_k} + \frac{\|\gamma_k\|^2}{(\delta_k^{\mathrm{T}} \gamma_k)^2} \delta_k \delta_k^{\mathrm{T}} + \frac{\delta_k \delta_k^{\mathrm{T}}}{2R_k}. \tag{2.5}$$

We denote the inverse of H_{k+1} by $B_{k+1} = {H_{k+1}}^{-1}$. Then

$$B_{k+1} = I - \frac{\delta_k \delta_k^{\mathrm{T}}}{\|\delta_k\|^2} + \frac{2R_k}{(\delta_k^{\mathrm{T}} \gamma_k)^2} \gamma_k \gamma_k^{\mathrm{T}}. \tag{2.6}$$

We require B_{k+1} to satisfy the non-quasi-Newton equation[8]

$$\delta_k^{\mathrm{T}} B_{k+1} \delta_k = 2R_k. \tag{2.7}$$

In order to decrease the iterative error, multiplying H_{k+1} by a coefficient t_k, we obtain

$$H_{k+1} = t_k(I - \frac{\gamma_k \delta_k^{\mathrm{T}} + \delta_k \gamma_k^{\mathrm{T}}}{\delta_k^{\mathrm{T}} \gamma_k} + \frac{\|\gamma_k\|^2}{(\delta_k^{\mathrm{T}} \gamma_k)^2} \delta_k \delta_k^{\mathrm{T}}) + \frac{\delta_k \delta_k^{\mathrm{T}}}{2R_k},$$

where $t_k = \frac{\delta_k^{\mathrm{T}} \gamma_k}{\|\gamma_k\|^2}$. Then

$$H_{k+1} = \frac{\delta_k^{\mathrm{T}} \gamma_k}{\|\gamma_k\|^2} I + \frac{\delta_k \delta_k^{\mathrm{T}}}{\delta_k{}^{\mathrm{T}} \gamma_k} - \frac{\gamma_k \delta_k^{\mathrm{T}} + \delta_k \gamma_k^{\mathrm{T}}}{\|\gamma_k\|^2} + \frac{\delta_k \delta_k^{\mathrm{T}}}{2R_k} \tag{2.8}$$

and

$$B_{k+1} = \frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}} \delta_k} I - \frac{\|\gamma_k\|^2}{\|\delta_k\|^2 \gamma_k^{\mathrm{T}} \delta_k} \delta_k \delta_k^{\mathrm{T}} + \frac{2R_k}{(\delta_k^{\mathrm{T}} \gamma_k)^2} \gamma_k \gamma_k^{\mathrm{T}}. \tag{2.9}$$

In fact, (2.9) satisfies non-quasi-Newton equation(2.7).

We call (2.8) and (2.9) the memoryless non-quasi-Newton update formulas. For quadratic functions, (2.8) and (2.9) are the Perry-Shanno memoryless quasi-Newton formulas. For non-quadratic functions, (2.8), (2.9) are not the same as the Perry-Shanno memoryless quasi-Newton formulas.

The following theorem shows the B_k is a positive definite matrix in update Formula (2.9).

Theorem 2.1 *If $R_k > 0$ and $\delta_k^{\mathrm{T}} \gamma_k > 0$ for all $k \geq 1$, then B_{k+1} is a positive definite matrix.*

Proof For all $z \in R^n, z \neq 0$, we have

$$z^{\mathrm{T}} \frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}} \delta_k} (I - \frac{\delta_k \delta_k^{\mathrm{T}}}{\|\delta_k\|^2}) z = \frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}} \delta_k} (\|z\|^2 - \frac{(z^{\mathrm{T}} \delta_k)^2}{\|\delta_k\|^2}) \geq 0,$$

$$z^{\mathrm{T}} (\frac{2R_k}{(\delta_k^{\mathrm{T}} \gamma_k)^2} \gamma_k \gamma_k^{\mathrm{T}}) z = \frac{2R_k}{(\delta_k^{\mathrm{T}} \gamma_k)^2} (z^{\mathrm{T}} \gamma_k)^2 \geq 0,$$

and it is impossible that the equality in the above two inequalities holds simultaneously. Therefore, $z^{\mathrm{T}} B_{k+1} z > 0$; that is, B_{k+1} is a positive definite matrix.

3. The global convergent of memoryless non-quasi-Newton method

We make the following assumptions:

H1 (i) $f \in C^2$, and the level set $D_1 = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded, that is, there exists

$$L_1 > 0 \quad \text{such that} \quad \|x\| \leq \frac{1}{2} L_1, \quad \forall x \in D_1.$$

(ii) There exist positive numbers $M > m > 0$, for all $x \in D_1, u \in R^n$, such that

$$m\|u\|^2 \leq u^{\mathrm{T}} G(x) u \leq M\|u\|^2, \tag{3.1}$$

where $G(x)$ is Hesse matrix of $f(x)$.

Consider the following iteration

$$x_{k+1} = x_k + \lambda_k d_k, \tag{3.2}$$

$$d_1 = -g_1, \tag{3.3}$$

$$\begin{aligned} d_{k+1} &= -H_{k+1}g_{k+1} = -B_{k+1}^{-1}g_{k+1} \\ &= (\frac{\delta_k^{\mathrm{T}} g_{k+1}}{\|\gamma_k\|^2} - \frac{\delta_k^{\mathrm{T}} g_{k+1}}{\delta_k{}^{\mathrm{T}}\gamma_k} - \frac{\delta_k^{\mathrm{T}} g_{k+1}}{2R_k})\delta_k + \frac{\delta_k^{\mathrm{T}} g_{k+1}}{\|\gamma_k\|^2}\gamma_k - \frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2}g_{k+1}, \quad k \geq 2, \end{aligned} \tag{3.4}$$

where

$$B_{k+1} = \frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}}\delta_k}I - \frac{\|\gamma_k\|^2}{\|\delta_k\|^2\gamma_k^{\mathrm{T}}\delta_k}\delta_k\delta_k^{\mathrm{T}} + \frac{2R_k}{(\delta_k^{\mathrm{T}}\gamma_k)^2}\gamma_k\gamma_k^{\mathrm{T}}, \tag{3.5}$$

$$H_{k+1} = B_{k+1}^{-1} = \frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2}I + \frac{\delta_k\delta_k^{\mathrm{T}}}{\delta_k{}^{\mathrm{T}}\gamma_k} - \frac{\gamma_k\delta_k^{\mathrm{T}} + \delta_k\gamma_k^{\mathrm{T}}}{\|\gamma_k\|^2} + \frac{\delta_k\delta_k^{\mathrm{T}}}{2R_k}. \tag{3.6}$$

In this paper, we require the steplength λ_k to satisfy Wolfe linesearch, that is,

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \beta\lambda_k g_k^{\mathrm{T}} d_k, \tag{3.7}$$

$$g(x_k + \lambda_k d_k)^{\mathrm{T}} d_k \geq \sigma g_k^{\mathrm{T}} d_k, \tag{3.8}$$

where $\beta \in (0, \frac{1}{2}), \sigma \in (\beta, 1)$. The sketch of the memoryless non-quasi-Newton method is as follows:

Step 1. Choose starting point $x_1, d_1 = -g_1 = - \bigtriangledown f(x_1), \varepsilon > 0, k := 1$.

Step 2. If $\|g_k\| \leq \varepsilon$, then stop; otherwise, determine a steplength λ_k by using Wolfe linesearch. Set $x_{k+1} := x_k + \lambda_k d_k$.

Step 3. Determine $\delta_k = x_{k+1} - x_k, g_{k+1} = \bigtriangledown f(x_{k+1}), \gamma_k = g_{k+1} - g_k$. If $-g_k^{\mathrm{T}} d_k \leq \varepsilon$, then $d_{k+1} = -g_{k+1}$, set $k := k + 1$, go to Step 2; otherwise, go to Step 4.

Step 4. Compute d_{k+1} by using (3.4), set $k := k + 1$, go to Step 2.

In the following, we will write the non-memoryless quasi-Newton method as Method A.

Lemma 3.1 *Assume that Assumption H1 holds. Then Method A satisfies*

$$\frac{m}{M} \leq \frac{\gamma_k^{\mathrm{T}}\delta_k}{2R_k} \leq \frac{M}{m}, \tag{3.9}$$

$$\frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}}\delta_k} \leq \frac{M^2}{m}. \tag{3.10}$$

Proof Since

$$\gamma_k = \int_0^1 G(x_k + t\delta_k)\delta_k \mathrm{d}t,$$

from which and Assumption H1, we have

$$\|\gamma_k\| \leq M\|\delta_k\|,$$

$$m\|\delta_k\|^2 \leq \gamma_k^{\mathrm{T}}\delta_k = \delta_k^{\mathrm{T}}\int_0^1 G(x_k + t\delta_k)\mathrm{d}t\delta_k \leq M\|\delta_k\|^2,$$

$$\frac{m}{2}\|\delta_k\|^2 \leq R_k = \int_0^1\int_0^{\mathrm{T}} \delta_k^{\mathrm{T}} G(x_k + \alpha\delta_k)\delta_k \mathrm{d}\alpha\mathrm{d}t \leq \frac{M}{2}\|\delta_k\|^2,$$

$$\frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}}\delta_k} \leq \frac{M\|\gamma_k\|\|\delta_k\|}{m\|\delta_k\|^2} \leq \frac{M^2}{m}.$$

Thus

$$\frac{m}{M} \leq \frac{\gamma_k^{\mathrm{T}}\delta_k}{2R_k} \leq \frac{M}{m}.$$

Theorem 3.2 *If for any $k \geq 1$, steplengthes $\lambda_1, \lambda_2, \dots, \lambda_k$ are generated by Method A. Then there exists a positive constant $c_1 > 0$ such that*

$$\prod_{j=1}^{k} \lambda_j \geq c_1^k. \tag{3.11}$$

Proof By (3.6), we have

$$\mathrm{tr}(H_{k+1}) = (n-2)\frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2} + (1 + \frac{\delta_k^{\mathrm{T}}\gamma_k}{2R_k})\frac{\|\delta_k\|^2}{\delta_k^{\mathrm{T}}\gamma_k}. \tag{3.12}$$

Since $(\delta_k^{\mathrm{T}}\gamma_k)^2 \leq \|\delta_k\|^2\|\gamma_k\|^2$, we have

$$\mathrm{tr}(H_{k+1}) \leq (n - 1 + \frac{M}{m})\frac{\|\delta_k\|^2}{\delta_k^{\mathrm{T}}\gamma_k} = c_2\frac{\|\delta_k\|^2}{\delta_k^{\mathrm{T}}\gamma_k}, \tag{3.13}$$

$$\mathrm{tr}(H_{k+1}) \geq (n - 1 + \frac{m}{M})\frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2} = c_3\frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2}, \tag{3.14}$$

where $c_2 = n - 1 + \frac{M}{m}, c_3 = n - 1 + \frac{m}{M}$.

By $\delta_k = \lambda_k d_k = \lambda_k(-H_k g_k)$ and (3.8), we have

$$\mathrm{tr}(H_{k+1}) \leq \frac{c_2\lambda_k}{1-\sigma}\frac{\|H_k g_k\|^2}{g_k^{\mathrm{T}}H_k g_k}. \tag{3.15}$$

Because H_k is a positive definite matrix, we have

$$\frac{\|H_k g_k\|^2}{g_k^{\mathrm{T}}H_k g_k} \leq \mathrm{tr}(H_k).$$

Since $H_1 = I$, we deduce the relation

$$\mathrm{tr}(H_{k+1}) \leq \frac{c_2\lambda_k}{1-\sigma}\mathrm{tr}(H_k) \leq \ldots \leq (\frac{c_2}{1-\sigma})^k(\prod_{j=1}^{k}\lambda_j)\mathrm{tr}(H_1) = (\frac{c_2}{1-\sigma})^k(\prod_{j=1}^{k}\lambda_j)n. \tag{3.16}$$

By Lemma 3.1 and (3.14), we have

$$\mathrm{tr}(H_{k+1}) \geq c_3\frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2} \geq \frac{c_3 m}{M^2}. \tag{3.17}$$

From (3.16) and (3.17), it follows that

$$\frac{c_3 m}{M^2} \leq (\frac{c_2}{1-\sigma})^k(\prod_{j=1}^{k}\lambda_j)n.$$

By the above arguments we have

$$\prod_{j=1}^{k}\lambda_j \geq \frac{\frac{c_3 m}{M^2}}{(\frac{c_2}{1-\sigma})^k n} \geq c_1^k,$$

where $c_1 = \frac{1-\sigma}{c_2}\min\{1, \frac{c_3 m}{M^2 n}\}$.

Theorem 3.3 *Assume that the Assumption H1 holds. Let $\{x_k\}$ be generated by Method A. Then,*

$$\lim_{k\to+\infty}\inf\|g_k\| = 0. \tag{3.18}$$

Proof The proof is by contradiction. Assume that the opposite of (3.18) holds, that is, there is a positive constant $\mu > 0$, such that $\|g_k\| \geq \mu$, $\forall k$. Due to (3.5) and Lemma 3.1, we know

$$\mathrm{tr}(B_{k+1}) = (n - 1 + \frac{2R_k}{\delta_k^{\mathrm{T}} \gamma_k}) \frac{\|\gamma_k\|^2}{\delta_k^{\mathrm{T}} \gamma_k} \leq (n - 1 + \frac{M}{m}) \frac{M^2}{m} = c_4, \tag{3.19}$$

since B_k is a positive definite matrix, we have

$$\frac{\|g_k\|^2}{g_k^{\mathrm{T}} H_k g_k} \leq \mathrm{tr}(B_k).$$

Therefore,

$$\frac{\mu^2}{g_k^{\mathrm{T}} H_k g_k} \leq \frac{\|g_k\|^2}{g_k^{\mathrm{T}} H_k g_k} \leq \mathrm{tr}(B_k) \leq c_4,$$

$$g_k^{\mathrm{T}} H_k g_k \geq \frac{\mu^2}{c_4} = c_5, \tag{3.20}$$

$$\prod_{j=1}^{k} g_j^{\mathrm{T}} H_j g_j \geq c_5^k. \tag{3.21}$$

Multiplying (3.11) with (3.21) and using the algebraic-geometric mean inequality, we have

$$(\frac{\sum_{j=1}^{k} \lambda_j g_j^{\mathrm{T}} H_j g_j}{k})^k \geq \prod_{j=1}^{k} \lambda_j g_j^{\mathrm{T}} H_j g_j \geq (c_1 c_5)^k = c_6^k. \tag{3.22}$$

It implies

$$\sum_{j=1}^{k} \lambda_j g_j^{\mathrm{T}} H_j g_j \geq k c_6.$$

Let $k \to +\infty$. We have

$$\sum_{j=1}^{+\infty} \lambda_j g_j^{\mathrm{T}} H_j g_j \geq +\infty.$$

By $d_j = -H_j g_j$ and (3.7), we obtain

$$+\infty \leq \sum_{j=1}^{+\infty} \lambda_j g_j^{\mathrm{T}} H_j g_j = \sum_{j=1}^{+\infty} (-\lambda_j g_j^{\mathrm{T}} d_j) \leq \frac{1}{\beta} \sum_{j=1}^{+\infty} (f_j - f_{j+1}),$$

which contradicts Assumption H1(i). Therefore,

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

4. Hybrid memoryless non-quasi-Newton method

In this section, we will derive a hybrid method for obtaining a globally convergent iteration under the weaker condtions based on the memoryless non-quasi-Newton method in Section 3. First, we let

$$\bar{K} = \{k | \frac{R_k}{\|\delta_k\|^2} > \frac{\varepsilon_1 \|g_k\|^\alpha}{\gamma_k^{\mathrm{T}} \delta_k}\}, \alpha > 0, \varepsilon_1 > 0.$$

Denote the memoryless non-quasi-Newton formula by H_{k+1}^{NEW} and Perry-Shanno memoryless quasi-Newton formula by H_{k+1}^{PS}, i.e.,

$$H_{k+1}^{\mathrm{NEW}} = \frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2}I + \frac{\delta_k\delta_k^{\mathrm{T}}}{\delta_k{}^{\mathrm{T}}\gamma_k} - \frac{\gamma_k\delta_k^{\mathrm{T}} + \delta_k\gamma_k^{\mathrm{T}}}{\|\gamma_k\|^2} + \frac{\delta_k\delta_k^{\mathrm{T}}}{2R_k},$$

$$H_{k+1}^{PS} = \frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2}I + 2\frac{\delta_k\delta_k{}^{\mathrm{T}}}{\delta_k^{\mathrm{T}}\gamma_k} - \frac{1}{\|\gamma_k\|^2}(\gamma_k\delta_k^{\mathrm{T}} + \delta_k\gamma_k^{\mathrm{T}}).$$

Let

$$H_{k+1} = \begin{cases} H_{k+1}^{\mathrm{NEW}}, & \text{if } k \in \bar{K}, \\ H_{k+1}^{\mathrm{PS}}, & \text{otherwise.} \end{cases} \tag{4.1}$$

We make the following assumptions

H2 (i) $f(x) \in C^2$, and is a convex function in D_2, where the level set $D_2 = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded; that is, there exists a positive constant $L_2 > 0$ such that $\|x\| \leq \frac{1}{2}L_2, \forall x \in D_2$.

(ii) $g(x)$ satisfies Lipschitz condition: $\exists L_3 > 0, \|g(x) - g(y)\| \leq L_3\|x - y\|, \forall x, y \in D_2$.

Replace H_{k+1} with (4.1) in Method A, then we obtain the hybrid memoryless non-quasi-Newton method and name it as Method B.

Theorem 4.1 *Assume that Assumption H_2 holds. Let the sequence $\{x_k\}$ be generated by Method B. Suppose $\exists\mu > 0$, such that $\|g_k\| \geq \mu$ hold for all k. Then there exist $M_1, M_2, M_3 > 0$, such that*

$$M_1 \leq \frac{\gamma_k^{\mathrm{T}}\delta_k}{2R_k} \leq M_2, \tag{4.2}$$

$$\frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}}\delta_k} \leq M_3. \tag{4.3}$$

Proof Let D_2^* denote the convex closure of the level set D_2, that is, D_2^* is the smallest closed convex set containing D_2. Since D_2 is bounded, D_2^* is bounded, too. By the continuity of Hesse matrix $G(x)$, there exists $M' > 0$, such that $\|G(x)\| \leq M'$ for all $x \in D_2^*$. Since H_{k+1}^{PS} is a positive definite matrix, Method B is a descented method. The minimization sequence $\{x_k\} \subset D_2$, then $x_k + \theta\delta_k = (1-\theta)x_k + \theta x_{k+1} \in D_2^*$ for all $0 \leq \theta \leq 1$. Therefore,

$$\begin{aligned} R_k = f_{k+1} - f_k - g_k^{\mathrm{T}}\delta_k &= \int_0^1\int_0^\theta \delta_k^{\mathrm{T}}G(x_k + \theta\delta_k)\mathrm{d}\alpha\mathrm{d}\theta \\ &\leq \int_0^1\int_0^\theta \|\delta_k\|\|G(x_k + \theta\delta_k)\delta_k\|\mathrm{d}\alpha\mathrm{d}\theta \\ &\leq \frac{M'}{2}\|\delta_k\|^2. \end{aligned}$$

If $k \in \bar{K}$, by (3.8), we know

$$\gamma_k^{\mathrm{T}}\delta_k > \varepsilon_1\frac{\|\delta_k\|^2\|g_k\|^\alpha}{R_k} \geq \frac{2\varepsilon_1\mu^\alpha}{M'}. \tag{4.4}$$

From (i) in Assumption H2, we have $\|\delta_k\|^2 \leq L_2^2$, then

$$\frac{\|\delta_k\|^2}{\gamma_k^{\mathrm{T}}\delta_k} \leq \frac{M'L_2^2}{2\varepsilon_1\mu^\alpha}. \tag{4.5}$$

Since

$$\gamma_k^{\mathrm{T}}\delta_k \le \|\gamma_k\|\|\delta_k\| \le L_3\|\delta_k\|^2, \tag{4.6}$$

and (4.5), (4.6), we have

$$\frac{2\varepsilon_1\mu^{\alpha}}{M'L_2^2} \le \frac{\gamma_k^{\mathrm{T}}\delta_k}{\|\delta_k\|^2} \le L_3. \tag{4.7}$$

By (4.7) and the definition of $\bar{K}$, we obtain

$$\frac{2R_k}{\|\delta_k\|^2} \ge \frac{2\varepsilon_1\|g_k\|^{\alpha}}{\gamma_k^{\mathrm{T}}\delta_k} \ge \frac{2\varepsilon_1\mu^{\alpha}}{L_2^2L_3}. \tag{4.8}$$

From (4.7) and (4.8), it implies that

$$\frac{4\varepsilon_1\mu^{\alpha}}{(M'L_2)^2} \le \frac{\gamma_k^{\mathrm{T}}\delta_k}{2R_k} = \frac{\gamma_k^{\mathrm{T}}\delta_k}{\|\delta_k\|^2}\frac{\|\delta_k\|^2}{2R_k} \le \frac{(L_2L_3)^2}{2\varepsilon_1\mu^{\alpha}}.$$

Let $M_1 = \frac{4\varepsilon_1\mu^{\alpha}}{(M'L_2)^2}$, $M_2 = \frac{(L_2L_3)^2}{2\varepsilon_1\mu^{\alpha}}$, then (4.2) holds. By (4.5), we have

$$\frac{\|\gamma_k\|^2}{\gamma_k^{\mathrm{T}}\delta_k} \le \frac{M'L_3^2\|\delta_k\|^2}{2\varepsilon_1\mu_2^{\alpha}} = M'M_2. \tag{4.9}$$

If $k \notin \bar{K}$, then there exists $\overline{G}$, such that $\int_0^1 G(x_k + t\delta_k)\mathrm{d}t\delta_k = \overline{G}\delta_k, t \in (0,1)$; that is, $\gamma_k = g_{k+1} - g_k = \overline{G}\delta_k$. By Assumption H2(i) we know, $\overline{G}$ is a semi-positive definite matrix. Thus we let $z_k = \overline{G}^{\frac{1}{2}}\delta_k$, where $\overline{G}^{\frac{1}{2}}\overline{G}^{\frac{1}{2}} = \overline{G}$. By Assumption H2, there exist $M' > 0$, $\|\overline{G}(x_k)\| \le M', \forall x_k \in D_2$. Then

$$\frac{\gamma_k^{\mathrm{T}}\gamma_k}{\gamma_k^{\mathrm{T}}\delta_k} = \frac{\delta_k^{\mathrm{T}}\overline{G}^2\delta_k}{\delta_k^{\mathrm{T}}\overline{G}\delta_k} = \frac{z_k^{\mathrm{T}}\overline{G}z_k}{z_k^{\mathrm{T}}z_k} \le M'.$$

We take $M_3 = \max\{M'M_2, M'\}$, then (4.3) holds.

Theorem 4.2 *If problem (1.1) satisfies Assumption H2, and the sequence $\{x_k\}$ is generated by Method B, then we have*

$$\lim_{k\to+\infty}\inf\|g_k\| = 0. \tag{4.10}$$

Proof First we prove that $\exists c_7 > 0$, $\prod_{j=1}^{k}\lambda_j \ge c_7^k$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are generated by Method B. By (4.2) and (3.14), we know

$$\mathrm{tr}(H_{k+1}^{\mathrm{NEW}}) \ge (n - 1 + M_1)\frac{\delta_k^{\mathrm{T}}\gamma_k}{\|\gamma_k\|^2} \ge \frac{c_3'm}{M_3},$$

$c_3' = n - 1 + M_1$. Since

$$\mathrm{tr}(H_{k+1}^{PS}) = (n-2)\frac{s_k^{\mathrm{T}}y_k}{\|y_k\|^2} + 2\frac{\|s_k\|^2}{s_k^{\mathrm{T}}y_k} \ge n\frac{\gamma_k^{\mathrm{T}}\delta_k}{\|\gamma_k\|^2} \ge \frac{n}{M'},$$

let $t_1 = \min\{\frac{c_3'm}{M_3}, \frac{n}{M'}\}$. Then

$$\mathrm{tr}(H_{k+1}) \ge t_1. \tag{4.11}$$

Since H_{k+1} is a positive definite matrix, by (3.15) and using the definition of H_{k+1}^{PS}, we have

$$\mathrm{tr}(H_{k+1}^{\mathrm{NEW}}) \le \frac{c_2\lambda_k}{1-\sigma}\frac{\|H_kg_k\|^2}{g_k^{\mathrm{T}}H_kg_k} \le \frac{c_2\lambda_k}{1-\sigma}\mathrm{tr}(H_k),$$

$$\mathrm{tr}(H_{k+1}^{\mathrm{PS}}) \le \frac{n\lambda_k}{1-\sigma}\frac{\|H_k g_k\|^2}{g_k^{\mathrm{T}} H_k g_k} \le \frac{n\lambda_k}{1-\sigma}\mathrm{tr}(H_k).$$

By $c_2 = n - 1 + \frac{n}{M} \ge n$, we have $\mathrm{tr}(H_{k+1}) \le \frac{c_2\lambda_k}{1-\sigma}\mathrm{tr}(H_k)$. Thus

$$\mathrm{tr}(H_{k+1}) \le \frac{c_2\lambda_k}{1-\sigma}\mathrm{tr}(H_k) \le \ldots \le (\frac{c_2}{1-\sigma})^k(\prod_{j=1}^{k}\lambda_j)\mathrm{tr}(H_1) = (\frac{c_2}{1-\sigma})^k(\prod_{j=1}^{k}\lambda_j)n.$$

From (4.11), we have

$$\prod_{j=1}^{k}\lambda_j \ge \frac{t_1}{n(\frac{c_2}{1-\sigma})^k} \ge (c_7)^k,$$

where $c_7 = \frac{1-\sigma}{c_2}\min\{1, \frac{t_1}{n}\}$.

Assume there exists $\mu > 0$, such that $\|g_k\| \ge \mu$. By (4.3) and (3.19), we have

$$\mathrm{tr}(B_{k+1}^{\mathrm{NEW}}) \le c_4, \quad \mathrm{tr}(B_{k+1}^{\mathrm{PS}}) = n\frac{\|\gamma_k\|^2}{\delta_k^{\mathrm{T}}\gamma_k} \le \frac{n}{M'},$$

where $c_4 = (n - 1 + \frac{M}{m})\frac{M^2}{m}$. Set $t_2 = \min\{c_4, \frac{n}{M'}\}$. Then

$$\mathrm{tr}(B_{k+1}) \le t_2.$$

By a similar way to the proof of Theorem 3.3, we can obtain

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

5. Numerical experiments

In order to test the given methods in this paper, we performed some numerical experiments, and compared DY conjugate gradient method[12] with both Wolfe linesearch and BFGS method. All of the test functions are from [11].

Test function	Function name
BADSCD	powell badly scaled functing
FROTH	freudenstein and roth
BOX	box three-dimensional function
SING	powell singular function
KOWOSB	kowalik and osborne function
SINGX	extended powell singular function
WOOD	wood function
PENALTY I	penalty function I
VARDIM	variable dimensioned function
ROSEX	extended rosenbrock function
TRID	broyden tridiagonal function

Table 5.1 List of test problems

Table 5.2 shows the number of iterations and Table 5.3 shows the CPU carry-out time (seconds). All of the methods terminate with $\|g(x_k)\| \leq 10^{-6}$. We take $\sigma = 0.01, \beta = 0.1, \varepsilon = 10^{-6}$ in Wolfe linesearch. In Method B we take α as follows: if $\|g_k\| \geq 1$, then $\alpha = 0.01$; if $\|g_k\| < 1$, then $\alpha = 1$. "F" denotes the number of iterations over 5000 or the CPU carry-out time over 600 seconds.

Problem	Dimension	Method A Ite.	Method B Ite.	BFGS Ite.
BADSCD	2	101	96	F
FROTH	2	11	11	8
BOX	3	58	16	15
SING	4	263	189	21
KOWOSB	4	378	174	15
SINGX	5000	169	/209	F
WOOD	1000/5000	188/263	188/266	F/F
PENALTY I	1000/5000	39/41	39/41	56/50
VARDIM	50/100/200/1000	11/13/14/16	11/13/14/16	13/10/F/17
ROSEX	50/100/1000/5000	19/20/19/18	19/20/19/18	42/42/21/F
TRID	100/200/1000/5000	34/38/35/35	34/38/35/35	74/83/78/75

Table 5.2 Test results

Problem	Dimension	Method A CPU(Sec.)	Method B CPU(Sec.)	DY CPU(Sec.)
BADSCD	2	3	3	1
FROTH	2	2	2	65
BOX	3	3	2	3
SING	4	7	6	54
KOWOSB	4	5	5	7
SINGX	5000	512	581	F
WOOD	1000/5000	223/1112	217/1424	F/F
PENALTY I	1000/5000	57/259	45/227	30/112
VARDIM	50/100/200/1000	2/3/5/23	1/2/4/20	F/3/F/F
ROSEX	50/100/1000/5000	3/4/27/118	1/3/24/64	3/4/35/195
TRID	100/200/1000/5000	6/12/44/212	4/9/41/149	5/29/57/298

Table 5.3 Test results (CPU)

From the numerical results in Tables 5.2 and 5.3, we see that the non-quasi Newton method and the hybrid non-memoryless method are sometimes more efficient. Especially the memoryless non-quasi-Newton method requires little storage and computation. So it is able to efficiently solve large scale optimization problems.

References

[1] PERRY J M. *A class of conjugate gradient algorithms with a two step variable metric memory* [J]. Discussion paper 269. Center for Mathematical Studies in Economic and Management Science,Northwestern University, 1977.

[2] SHANNO D F. *On the convergence of a new conjugate gradient algorithm* [J]. SIAM J. Numer. Anal., 1978, **15**(6): 1247–1257.

[3] SHANNO D F. *Conjugate gradient methods with inexact searches* [J]. Math. Oper. Res., 1978, **3**(3): 244–256.

[4] POWELL M J D. *Restart procedures for the conjugate gradient method* [J]. Math. Programming, 1977, **12**(2): 241–254.

[5] TOINT PH L. *Global convergence of the patitioned BFGS algorithms for convex partially separable optimization* [J]. Math. Programming, 1997, **77**: 69–94.

[6] XI Shaolin. *Nonliner Optimization Methods* [M]. Beijing: Higher Education Press, 1992.

[7] YUAN Yaxiang. *A modified BFGS algorithm for unconstrained optimization* [J]. IMA J. Numer. Anal., 1991, **11**(3): 325–332.

[8] ZHAO Yunbin, DUAN Yurong. Derivation and Property of Pesudo-Newton-N Class. Derivation and Property of Pesudo-Newton-N Class, Communication On Applied Mathematics and Computation, 1996,10(1), pp.82-90.

[9] DAI Yuhong. *Convergence properties of a memoryless quasi-Newton method* [J]. J. Numer. Methods Comput. Appl., 2000, **21**(1): 28–32

[10] CHEN Lanping, JIAO Baocong. *A Class of Non-Quasi-Newton methods and its convergence* [J]. Comm Appl. Computer, 1997, **11**(2): 9–17.

[11] MORÉ J J, GARBOW B S, HILLSTROM K E. *Testing unconstrained optimization software* [J]. ACM Trans. Math. Software, 1981, **7**(1): 17–41.

[12] DAI Yuhong, YUAN Yaxiang. *A Nonlinear conjugate gradient method with a strong Global convergence propertie* [J]. SIAM J Optimization, 1999, **10**: 177–182.