MEMORYLESS QUASI-NEWTON METHODS BASED ON SPECTRAL-SCALING BROYDEN FAMILY FOR UNCONSTRAINED OPTIMIZATION

SHUMMIN NAKAYAMA*
Department of Applied Mathematics, Graduate School of Science
Tokyo University of Science, 1-3, Kagurazaka
Shinjuku-ku, Tokyo 162-8601, Japan

YASUSHI NARUSHIMA
Faculty of International Social Sciences
Yokohama National University, 79-4 Tokiwadai
Hodogaya-ku, Yokohama 240-8501, Japan

HIROSHI YABE
Department of Applied Mathematics
Tokyo University of Science, 1-3, Kagurazaka
Shinjuku-ku, Tokyo 162-8601, Japan

(Communicated by Fabian Bastin)

Abstract. Memoryless quasi-Newton methods are studied for solving large-scale unconstrained optimization problems. Recently, memoryless quasi-Newton methods based on several kinds of updating formulas were proposed. Since the methods closely related to the conjugate gradient method, the methods are promising. In this paper, we propose a memoryless quasi-Newton method based on the Broyden family with the spectral-scaling secant condition. We focus on the convex and preconvex classes of the Broyden family, and we show that the proposed method satisfies the sufficient descent condition and converges globally. Finally, some numerical experiments are given.

1. Introduction. We are concerned with the unconstrained optimization problem

$$\min f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. We denote its gradient $\nabla f$ by $g$. Quasi-Newton methods are known as effective numerical methods for solving problem (1), and they are iterative methods of the form

$$x_{k+1} = x_k + \alpha_k d_k,$$

where $x_k \in \mathbb{R}^n$ is the $k$-th approximation to a solution, $\alpha_k$ is a step size, and $d_k$ is a search direction obtained by solving the linear system of equations

$$B_k d_k = -g_k$$

2010 Mathematics Subject Classification. Primary: 90C30, 90C06; Secondary: 65K05.
Key words and phrases. Unconstrained optimization, memoryless quasi-Newton method, Broyden family, sufficient descent condition, global convergence.

* Corresponding author: Shummin Nakayama.

1773
or directly from
\[ d_k = -H_k g_k. \]  
Here \( g_k \) denotes \( g(x_k) \), \( B_k \) is a nonsingular approximation matrix to the Hessian \( \nabla^2 f(x_k) \) and \( H_k = B_k^{-1} \). Throughout this paper, we choose the identity matrix \( I \) as the initial matrix \( B_0 \) or \( H_0 \), namely, \( d_0 = -g_0 \). The matrix \( B_k \) or \( H_k \) is updated at each iteration such that the secant condition
\[ B_k s_{k-1} = y_{k-1} \quad \text{or} \quad H_k y_{k-1} = s_{k-1} \]  
is satisfied, where \( s_{k-1} \) and \( y_{k-1} \) are defined by
\[ s_{k-1} = x_k - x_{k-1} = \alpha_{k-1} d_{k-1} \quad \text{and} \quad y_{k-1} = g_k - g_{k-1}, \]
respectively. There are several kinds of updating formulas. The well-known updating formulas are the BFGS, DFP and symmetric rank-one (SR1) formulas. In this paper, we focus on the Broyden family
\[ B_k = B_{k-1} - \frac{B_{k-1} s_{k-1} y_{k-1}^T}{s_{k-1}^T B_{k-1} s_{k-1}} + \frac{y_{k-1} y_{k-1}^T}{s_{k-1}^T y_{k-1}} + \phi_{k-1} v_{k-1} v_{k-1}^T, \]  
where \( \phi_{k-1} \) is a parameter. In particular, by setting \( \phi_{k-1} = 0, \phi_{k-1} = 1, \)
\[ \phi_{k-1} = \frac{s_{k-1}^T y_{k-1}}{(y_{k-1} - B_{k-1} s_{k-1})^T s_{k-1}} \quad \text{and} \quad \phi_{k-1} = \frac{s_{k-1}^T y_{k-1}}{(y_{k-1} + B_{k-1} s_{k-1})^T s_{k-1}}, \]
the formula \( (5) \) becomes the BFGS, DFP, SR1 and Hoshino formula \([16]\), respectively. If \( \phi_{k-1} \in [0, 1] \), the family \( (5) \) is a convex class, which is a convex combination of the BFGS and DFP formulas. Note that the BFGS, DFP and Hoshino formulas belong to the convex class. The BFGS and DFP formulas have a dual relation each other, while the SR1 and Hoshino formulas are self-dual, respectively (see \([24, 30]\), for example). In general, \( \phi_{k-1} = 0 \) is suggested as the best choice in the convex class. However, to find a better choice than \( \phi_{k-1} = 0 \), Zhang and Tewarson \([34]\) dealt with the preconvex class of the Broyden family \( (5) \), which means the interval \( (\phi_{k-1}^*, 0) \) in the Broyden family, where \( \phi_{k-1}^* \) is defined by
\[ \phi_{k-1}^* = \frac{1}{1 - \mu_{k-1}} \quad \text{with} \quad \mu_{k-1} = \frac{(s_{k-1}^T B_{k-1} s_{k-1}) (y_{k-1}^T H_{k-1} y_{k-1})}{(s_{k-1}^T y_{k-1})^2}. \]
If \( B_{k-1} \) is symmetric positive definite and \( s_{k-1}^T y_{k-1} > 0 \) holds, then \( B_k \) updated by \( (5) \) with \( \phi_{k-1} > \phi_{k-1}^* \) is also symmetric positive definite (see \([30]\), for example). This guarantees that the search direction satisfies the descent condition:
\[ g_k^T d_k < 0 \quad \text{for all} \quad k, \]  
because the positive definiteness of \( B_k \) implies
\[ g_k^T d_k = -g_k^T B_k^{-1} g_k < 0. \]
Zhang and Tewarson \([34]\) claimed that the quasi-Newton method based on the Broyden family with the preconvex class is efficient when the parameter \( \phi_{k-1} \) was chosen appropriately.

By considering the inverse of \( (5) \), the Broyden family of \( H_k \) is represented by
\[ H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} + \theta_{k-1} w_{k-1} w_{k-1}^T, \]  
where \( \theta_{k-1} \) is a parameter.
exists a positive constant $c$ such that

\[
\phi_{k-1} - 1 = \frac{\phi_{k-1} - 1}{\phi_{k-1}(1 - \mu_{k-1}) - 1}.
\]

Note that $\phi_{k-1} = 0$ corresponds to $\theta_{k-1} = 1$ and $\phi_{k-1} = 1$ corresponds to $\theta_{k-1} = 0$. We emphasize that the convex class $\phi_{k-1} \in [0, 1]$ is mapped into $\theta_{k-1} \in [0, 1]$ and the preconvex class $\phi_{k-1} \in [\phi_{k-1}^*, 0)$ is mapped into $\theta_{k-1} \in (1, \infty)$. Moreover, if $s_{k-1}^T y_{k-1} > 0$, $\theta_{k-1} \geq 0$ and $H_{k-1}$ is symmetric positive definite, then the matrix $H_k$ updated by the Broyden family \((7)\) is also symmetric positive definite (see [30], for example).

For solving large-scale unconstrained optimization problems, Shanno [27] proposed the memoryless quasi-Newton method to avoid the storage of memories for matrices. Specifically, Shanno substituted \((7)\) with $H_{k-1} = I$ into \((3)\) and derived the following search direction:

\[
d_k = -g_k + \left( \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) d_{k-1} - \left( 1 + \frac{y_{k-1}^T g_k}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} d_{k-1} + \left( \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) y_{k-1}.
\]

For the case $\theta_{k-1} = 1$, the search direction \((9)\) yields the memoryless quasi-Newton method based on the BFGS formula (the memoryless BFGS method):

\[
d_k = -g_k + \left( \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) d_{k-1} - \left( 1 + \frac{y_{k-1}^T g_k}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} d_{k-1} + \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}.
\]

If the exact line search is used, namely $g_k^T d_{k-1} = 0$, then the memoryless BFGS method is identical to the nonlinear conjugate gradient (CG) method with the Hestenes-Stiefel formula (see [13, 15, 22], for example). Thus, memoryless quasi-Newton methods closely relate to CG methods and three-term CG methods [1, 23]. In this decade, several memoryless quasi-Newton methods have been studied. For example, Kou and Dai [17] proposed the modified self-scaling memoryless BFGS method which converges globally for general objective functions. Furthermore, several researchers have paid attention to the memoryless quasi-Newton method based on the SR1 formula instead of the BFGS formula. Modarres et al. [19] presented the memoryless quasi-Newton method based on the self-scaling SR1 formula [29] with the modified secant condition, and proved the global convergence property of the proposed method for uniformly convex objective functions. Nakayama et al. [21] also proposed the memoryless quasi-Newton method based on the SR1 formula with the spectral-scaling secant condition [4]. Some researchers proposed CG methods based on the memoryless quasi-Newton method (see [5, 20, 33], for example).

In this paper, we consider the memoryless quasi-Newton methods based on the Broyden family with the spectral-scaling secant condition which always satisfies the sufficient descent condition. Here, the sufficient descent condition means that there exists a positive constant $c$ such that

\[
g_k^T d_k \leq -c\|g_k\|^2 \quad \text{for all } k,
\]
where $\| \cdot \|$ is the Euclidean norm. This condition is important in establishing the global convergence of general iterative methods. Although, $\theta_{k-1} = 1$ in (7) (namely the BFGS formula) is known as one of the best choices, we also focus on the convex class $\theta_{k-1} \in [0, 1]$ and the preconvex class $\theta_{k-1} \in (1, \infty)$ to find a better choice than $\theta_{k-1} = 1$. In particular, the following are our research subjects.

- We give a condition for the parameters which guarantees the global convergence of the method.
- We investigate how a choice of the parameter $\theta_{k-1}$ affects numerical performance. Specifically, we will compare the convex class and the preconvex class in the numerical experiments.

In addition, based on Zhang and Tewarson’s claim [34], we expect that good choices of the parameter can be found from the preconvex class.

This paper is organized as follows. In Section 2, we propose a memoryless quasi-Newton method based on the Broyden family with the spectral-scaling secant condition [4] which satisfies the sufficient descent condition. In Section 3, we prove the global convergence properties of the proposed methods for uniformly convex and general objective functions, respectively. Finally, some numerical results are shown in Section 4.

2. Memoryless quasi-Newton method based on the Broyden family. In this section, we present a new memoryless quasi-Newton method based on the Broyden family which always satisfies the sufficient descent condition (10).

In order to improve the performance of the quasi-Newton method, Cheng and Li [4] proposed the spectral-scaling secant condition. They scaled the objective function for numerical stability and considered the following approximate relation

$$
\gamma_{k-1} f(x) \approx \gamma_{k-1} \left( f(x_k) + g_k^T (x - x_k) + \frac{1}{2} (x - x_k)^T \nabla^2 f(x_k)(x - x_k) \right), \tag{11}
$$

where $\gamma_{k-1} > 0$ is a scaling parameter. Differentiating (11) and substituting $x_{k-1}$ into $x$, we have

$$
\gamma_{k-1} \nabla^2 f(x_k) s_{k-1} \approx \gamma_{k-1} y_{k-1},
$$

which yields the spectral-scaling secant condition:

$$
B_k s_{k-1} = \gamma_{k-1} y_{k-1}, \tag{12}
$$

where $B_k$ is an approximation to $\gamma_{k-1} \nabla^2 f(x_k)$. Setting $H_k = B_k^{-1}$ yields the relation

$$
H_k y_{k-1} = \frac{1}{\gamma_{k-1}} s_{k-1}. \tag{13}
$$

Chen and Cheng [3] proposed the Broyden family based on the spectral-scaling secant condition (13) as follows

$$
H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} + \frac{1}{\gamma_{k-1} s_{k-1}^T s_{k-1}} + \theta_{k-1} w_{k-1} w_{k-1}^T, \tag{14}
$$

where $w_{k-1}$ appears in (8). As shown in (7), if $s_{k-1}^T y_{k-1} > 0$ and $H_{k-1}$ is symmetric positive definite, then the matrix $H_k$ updated by the Broyden family (14) with $\theta_{k-1} \geq 0$ is also symmetric positive definite.

We now consider the search direction of a memoryless quasi-Newton method based on (14). From (3) and (14) with $H_{k-1} = I$, we have the following search
direction
\[
d_k = -g_k + \left( \theta_{k-1} \frac{y^T_{k-1}g_k}{d^T_{k-1}y_{k-1}} - \left( \gamma_{k-1} + \theta_{k-1} \frac{y^T_{k-1}y_{k-1}}{s^T_{k-1}y_{k-1}} \right) \frac{s^T_{k-1}g_k}{d^T_{k-1}y_{k-1}} \right) d_{k-1} \\
+ \left( \theta_{k-1} \frac{d^T_{k-1}g_k}{d^T_{k-1}y_{k-1}} + (1 - \theta_{k-1}) \frac{y^T_{k-1}g_k}{y^T_{k-1}y_{k-1}} \right) y_{k-1},
\]
(15)

where \( \gamma_{k-1} = 1/\gamma_{k-1} \). The following proposition gives the sufficient condition for (15) to satisfy (10).

**Proposition 2.1.** If
\[
d^T_{k-1}y_{k-1} > 0,
\]
(16)

\[
\gamma_{k-1} \geq \frac{y^T_{k-1}y_{k-1}}{s^T_{k-1}y_{k-1}}
\]
(17)

and
\[
0 < \theta_{\min} \leq \theta_{k-1} \leq \theta_{\max} < 2
\]
(18)

hold, then the search direction (15) satisfies the sufficient descent condition (10) with \( c := \min \left\{ \frac{\theta_{\min}}{2}, 1 - \frac{\theta_{\max}}{2} \right\} \), where \( \theta_{\min} \) and \( \theta_{\max} \) are constants satisfying
\[
0 < \theta_{\min} \leq 1 \leq \theta_{\max} < 2.
\]

**Proof.** It follows from relations \( 2u^Tv \leq \|u\|^2 + \|v\|^2 \) for any vectors \( u \) and \( v \), (15), (16) and (17) that
\[
y^T_k d_k = -\|g_k\|^2 + 2\theta_{k-1} \left( \frac{y^T_{k-1}g_k}{d^T_{k-1}y_{k-1}} \right) \frac{\left( d^T_{k-1}g_k \right)}{d^T_{k-1}y_{k-1}} + 2\theta_{k-1} \left( \frac{y^T_{k-1}g_k}{s^T_{k-1}y_{k-1}} \right) \frac{\left( s^T_{k-1}g_k \right)}{s^T_{k-1}y_{k-1}} + (1 - \theta_{k-1}) \left( \frac{y^T_{k-1}g_k}{y^T_{k-1}y_{k-1}} \right) \frac{\left( y^T_{k-1}g_k \right)}{y^T_{k-1}y_{k-1}}
\]
\[
\leq -\|g_k\|^2 + 2\theta_{k-1} \left( \|\sqrt{2}d^T_{k-1}g_k\|_{\sqrt{2}y_{k-1}} \right) \left( \|\sqrt{2}g_k\|_{\sqrt{2}y_{k-1}} \right) + (1 - \theta_{k-1}) \left( \frac{y^T_{k-1}g_k}{y^T_{k-1}y_{k-1}} \right) \frac{\left( y^T_{k-1}g_k \right)}{y^T_{k-1}y_{k-1}}
\]
\[
= -\left( \frac{y^T_{k-1}g_k}{y^T_{k-1}y_{k-1}} \right) \frac{\left( y^T_{k-1}g_k \right)}{y^T_{k-1}y_{k-1}}
\]
(19)
If $0 < \theta_{k-1} \leq 1$, then from $1 - \theta_{k-1} \geq 0$, (18) and (19), we have
\[ g_k^T d_k \leq -\left(1 - \frac{\theta_{k-1}}{2}\right) \|g_k\|^2 + (1 - \theta_{k-1}) \frac{\|y_{k-1}\|^2}{y_{k-1}^Ty_{k-1}} \leq -\frac{\theta_{\min}}{2} \|g_k\|^2. \]
Otherwise $1 < \theta_{k-1} < 2$ holds and hence (18) and (19) yield
\[ g_k^T d_k \leq -\left(1 - \frac{\theta_{k-1}}{2}\right) \|g_k\|^2 \leq -\left(1 - \frac{\theta_{\max}}{2}\right) \|g_k\|^2. \]
Therefore, the search direction (15) satisfies the sufficient descent condition (10) with $c := \min \left\{ \frac{\theta_{\min}}{2}, 1 - \frac{\theta_{\max}}{2} \right\}$. \hfill $\square$

We emphasize that (15) is an extension of the memoryless BFGS method based on (13) to the memoryless quasi-Newton method based on the Broyden family. On the other hand, by modifying the memoryless BFGS method based on (4), Kou and Dai [17] proposed the following search direction
\[ d_k = -g_k + \left(\frac{y_{k-1}^Ty_k}{d_{k-1}^Ty_{k-1}} - \left(\frac{y_{k-1}^Ty_{k-1}y_{k-1}^T}{d_{k-1}^Ty_{k-1}} - \frac{s_{k-1}^Ty_{k-1}y_{k-1}^T}{d_{k-1}^Ty_{k-1}}\right)\right) d_{k-1} + \xi_{k-1} g_{k-1} d_{k-1}^T y_{k-1}, \]
where $\tau_{k-1} > 0$ and $\xi_{k-1} \in [0, 1]$ are parameters. If $\tau_{k-1}$ is set to be $\hat{\tau}_{k-1}$, then (20) with $\xi_{k-1} = 1$ corresponds to (15) with $\theta_{k-1} = 1$.

3. Global convergence. In this section, we prove the global convergence of our method for uniformly convex and general objective functions, respectively. In order to show global convergence properties, we make the following standard assumptions for objective functions.

**Assumption 1.**
(i) The level set $\mathcal{L} = \{ x \mid f(x) \leq f(x_0) \}$ at the initial point $x_0$ is bounded, namely, there exists a positive constant $\hat{a}$ such that
\[ \|x\| \leq \hat{a} \quad \text{for all } x \in \mathcal{L}. \]  
(ii) The objective function $f$ is continuously differentiable on an open convex neighborhood $\mathcal{N}$ of $\mathcal{L}$, and its gradient $g$ is Lipschitz continuous on $\mathcal{N}$, namely, there exists a positive constant $L$ such that
\[ \|g(u) - g(v)\| \leq L \|u - v\| \quad \text{for all } u, v \in \mathcal{N}. \]

Under the above assumptions, there exists a positive constant $\Gamma$ such that
\[ \|g(x)\| \leq \Gamma \quad \text{for all } x \in \mathcal{L}. \]

Throughout this paper, we assume that $g_k \neq 0$ for all $k \geq 0$, otherwise a stationary point has been found.

In the line search procedure, we require the step size $\alpha_k$ in (2) to satisfy the Wolfe conditions:
\[ f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k, \]
\[ g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \]
where $0 < \delta < \sigma < 1$. If the search direction satisfies the curvature condition (25) and the descent condition (6), then we have
\[ d_{k-1}^T y_{k-1} = g_{k-1}^T d_{k-1} - g_{k-1}^T d_{k-1} \geq -(1 - \sigma) g_{k-1}^T d_{k-1} > 0, \]
which yields (16). In addition, we have
\[ g_k^T d_{k-1} \geq \sigma g_{k-1}^T d_{k-1} \geq \frac{-\sigma}{1 - \sigma} d_{k-1}^Ty_{k-1}. \] (27)
On the other hand, it follows from (6) that
\[ d_{k-1}^Ty_{k-1} = g_k^T d_{k-1} - g_{k-1}^T d_{k-1} \geq g_k^T d_{k-1}. \] (28)
Therefore, from (27) and (28), we obtain
\[ \left| \frac{g_k^T d_{k-1}}{d_{k-1}^Ty_{k-1}} \right| \leq \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\}. \] (29)

The following lemma [28, Lemma 3.1] is useful in showing the global convergence of our method.

**Lemma 3.1.** Suppose that Assumption 1 is satisfied. Consider any iterative method in the form (2), where \( d_k \) and \( \alpha_k \) satisfy the sufficient descent condition (10) and the Wolfe conditions (24)–(25), respectively. If
\[ \sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty, \]
then the method converges globally in the sense that
\[ \liminf_{k \to \infty} \|g_k\| = 0 \] (30)
holds.

We first show the global convergence of our method for uniformly convex objective functions. We note that Assumption 1 (i) is always satisfied for uniformly convex objective functions.

**Theorem 3.2.** Suppose that the objective function \( f \) is a uniformly convex function and Assumption 1 is satisfied. Consider the method (2) and (15) with (17)–(18). Assume the step size \( \alpha_k \) satisfies the Wolfe conditions (24)–(25) and \( \gamma_{k-1} \) is bounded, namely, there exists a positive constant \( \Gamma \) such that
\[ \gamma_{k-1} \leq \Gamma \quad \text{for all} \ k \geq 1. \] (31)
Then
\[ \lim_{k \to \infty} \|g_k\| = 0 \] (32)
holds. Furthermore, the generated sequence \( \{x_k\} \) converges to the global minimizer.

**Proof.** Since the objective function is uniformly convex, there exists a positive constant \( m \) such that
\[ (g(u) - g(v))^T (u - v) \geq m\|u - v\|^2 \quad \text{for all} \ u, v \in \mathbb{R}^n, \]
which implies that
\[ s_{k-1}^T y_{k-1} \geq m\|s_{k-1}\|^2. \] (33)
We have from (15), (18), (22), (23), (31), (33) and the fact \( s_{k-1} = \alpha_{k-1} d_{k-1} \)

\[
|d_k| = \| -g_k + \left( \theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left( \frac{\gamma_{k-1} + \theta_{k-1} y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{s_{k-1}^T y_{k-1}} \right) s_{k-1} \\
+ \left( \theta_{k-1} \frac{s_{k-1}^T g_k}{s_{k-1}^T y_{k-1}} \right) (1 - \theta_{k-1}) \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}} \right) y_{k-1} \| \\
\leq \| g_k \| + \left( 2 \frac{L}{m} \| s_{k-1} \| \| g_k \| + \| \hat{\gamma} + 2 \frac{L^2}{m} \| s_{k-1} \| \| g_k \| \right) \| s_{k-1} \| \\
+ 2L \frac{m}{\| s_{k-1} \|^2} \| g_k \| \\
= \left( 2 + \frac{1}{m} \left( 4L + \hat{\gamma} + 2 \frac{L^2}{m} \right) \right) \| g_k \| \\
\leq \left( 2 + \frac{1}{m} \left( 4L + \hat{\gamma} + 2 \frac{L^2}{m} \right) \right) \Gamma.
\]

Hence we obtain \( \sum_{k=0}^{\infty} \frac{1}{\| d_k \|^2} = \infty \). By Proposition 2.1 and Lemma 3.1, we have (30). Since the objective function is uniformly convex, (30) yields (32), and the generated sequence \( \{ x_k \} \) converges to the global minimizer.

We next show the global convergence of the proposed method for general objective functions. In order to establish the global convergence of the proposed method, we modify the search direction (15) as follows:

\[
d_k = -g_k + \beta_k d_{k-1} + \zeta_k y_{k-1} \quad k \geq 1,
\]

\[
\beta_k = \max \left\{ \theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left( \frac{\gamma_{k-1} + \theta_{k-1} y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}, 0 \right\},
\]

\[
\zeta_k = \text{sgn}(\beta_k) \left( \theta_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1 - \theta_{k-1}) \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}} \right),
\]

where \( \text{sgn}(\cdot) \) is defined by

\[
\text{sgn}(a) = \begin{cases} 
1 & a > 0, \\
0 & a = 0.
\end{cases}
\]

Equations (35) and (36) mean a safeguard in the sense that the search direction becomes the steepest descent direction \( (d_k = -g_k) \) when the coefficient of \( d_{k-1} \) in (15) is nonpositive. Note that such a modification is widely used in showing the global convergence of three-term CG methods and memoryless quasi-Newton methods (see [13, 17, 21, 22], for example).

If \( \beta_k > 0 \) is satisfied, then Proposition 2.1 holds. Otherwise, the search direction (34) becomes the steepest descent direction. Therefore, the search direction (34) always satisfies the sufficient descent condition (10).

In order to establish the global convergence of the method, we now introduce the following property.

Property 1. Consider the method (2) and (34)-(36) with (17)-(18). Suppose that there exists a positive constant \( \varepsilon \) such that

\[
\varepsilon \leq \| g_k \| \quad \text{for all } k.
\]
Then we say that the method has Property 1 if there exists a positive constant \( \bar{c} \) such that

\[
\gamma_{k-1} \leq \bar{c} \| s_{k-1} \| \quad \text{for all } k.
\]

If the method has Property 1, the following lemma holds.

**Lemma 3.3.** Consider the method (2) and (34)–(36) with (17)–(18), where the step size \( \alpha_k \) satisfies the Wolfe conditions (24)–(25). Suppose that Assumption 1 is satisfied and there exists a positive constant \( \varepsilon > 0 \) such that (37) holds. If the method has Property 1, then there exist constants \( b > 1 \) and \( \xi > 0 \) such that for all \( k \)

\[
\beta_k \leq b
\]

and

\[
\| s_{k-1} \| \leq \xi \quad \Rightarrow \quad \beta_k \leq \frac{1}{2b}.
\]

**Proof.** It follows from (10), (26) and (37) that

\[
d_{k-1}^T y_{k-1} \geq -(1 - \sigma) y_{k-1}^T d_{k-1} \geq c(1 - \sigma) || y_{k-1} ||^2 \geq c(1 - \varepsilon^2).
\]

From (18), (21), (22), (23), (29), (35), (38) and (41), we have

\[
\beta_k \leq \left| \theta_{k-1} - \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right| + \left( \theta_{k-1} \gamma_k - 1 + \theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) \left| \theta_{k-1} \frac{d_{k-1}^T y_{k-1}}{d_{k-1}^T y_{k-1}} \right|
\]

\[
\leq 2 \left| \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right| + 2 \left( \bar{c} \| s_{k-1} \| + 2 \left| \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right| \right) \left| \frac{d_{k-1}^T y_{k-1}}{d_{k-1}^T y_{k-1}} \right|
\]

\[
\leq 2 \left( \frac{L^T}{c(1 - \sigma) \varepsilon^2} + \left( \bar{c} + \frac{4 L^2 \hat{\alpha}}{c(1 - \sigma) \varepsilon^2} \right) \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\} \right) \| s_{k-1} \|
\]

\[
\leq 4 \left( \frac{L^T}{c(1 - \sigma) \varepsilon^2} + \left( \bar{c} + \frac{4 L^2 \hat{\alpha}}{c(1 - \sigma) \varepsilon^2} \right) \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\} \right) \hat{a}
\]

\[
:= \bar{b}.
\]

We now define \( b = 1 + \bar{b} \) and

\[
\xi = \frac{1}{4 \left( \frac{L^T}{c(1 - \sigma) \varepsilon^2} + \left( \bar{c} + \frac{4 L^2 \hat{\alpha}}{c(1 - \sigma) \varepsilon^2} \right) \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\} \right) b}.
\]

If \( \| s_{k-1} \| \leq \xi \), then by (42) we obtain

\[
\beta_k \leq 2 \left( \frac{L^T}{c(1 - \sigma) \varepsilon^2} + \left( \bar{c} + \frac{4 L^2 \hat{\alpha}}{c(1 - \sigma) \varepsilon^2} \right) \max \left\{ \frac{\sigma}{1 - \sigma}, 1 \right\} \right) \xi = \frac{1}{2b}.
\]

Therefore, this lemma is proved.

This lemma means that \( \beta_k \) will be small when \( \| s_{k-1} \| \) is small. This characterization is originally proposed for the CG methods by Gilbert and Nocedal [9], and it is called Property (⋆) (see also [6]). The next lemma corresponds to [6, Lemma 3.4] and [9, Lemma 4.1].

**Lemma 3.4.** If all assumptions of Lemma 3.3 are satisfied, then \( d_k \neq 0 \) and

\[
\sum_{k=0}^{\infty} \| u_k - u_{k-1} \|^2 < \infty
\]

holds, where \( u_k = d_k / \| d_k \| \).
Proof. Since \(d_k \neq 0\) follows from (10) and (37), the vector \(u_k\) is well-defined. By defining
\[
v_k = -(g_k - \zeta_k y_{k-1}) \frac{1}{\|d_k\|} \quad \text{and} \quad \eta_k = \beta_k \frac{\|d_{k-1}\|}{\|d_k\|},
\]
equation (34) is written as
\[
u_k = v_k + \eta_k u_{k-1}.
\]
Then we get from the fact
\[
\|u_k\| = \|u_{k-1}\| = 1
\]
\[
\|v_k\| = \|u_k - \eta_k u_{k-1}\| = \|\eta_k u_k - u_{k-1}\|.
\]
From the relations \(\beta_k \geq 0\) and (43), we obtain
\[
\|u_k - u_{k-1}\| \leq (1 + \eta_k)\|u_k - u_{k-1}\|
\]
\[
= \|u_k - \eta_k u_{k-1} + \eta_k u_k - u_{k-1}\|
\]
\[
\leq \|u_k - \eta_k u_{k-1}\| + \|\eta_k u_k - u_{k-1}\|
\]
\[
= 2\|v_k\|.
\]
(44)

It follows from (18), (29) and (36) that
\[
\|\zeta_k y_{k-1}\| \leq \left| \theta_{k-1} \frac{\theta_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right| + |1 - \theta_{k-1}| \frac{\|y_{k-1}\| \|g_k\|}{\|y_{k-1}\|^2} \|y_{k-1}\|
\]
\[
\leq 2 \max \left\{ \frac{\sigma}{1 - \sigma}, \ 1 \right\} \|y_{k-1}\| + \|g_k\|.
\]
(45)

Hence, by (21), (22), (23), (44) and (45), for any positive integer \(m\), we have
\[
\sum_{k=0}^{m} \|u_k - u_{k-1}\|^2 \leq 4 \sum_{k=0}^{m} \|v_k\|^2
\]
\[
\leq 4 \sum_{k=0}^{m} \left( 2\|g_k\| + 2 \max \left\{ \frac{\sigma}{1 - \sigma}, \ 1 \right\} \|y_{k-1}\| \right)^2 \frac{1}{\|d_k\|^2}
\]
\[
\leq 16 \left( \Gamma + 2L\tilde{a} \max \left\{ \frac{\sigma}{1 - \sigma}, \ 1 \right\} \right)^2 \sum_{k=0}^{m} \frac{1}{\|d_k\|^2}.
\]

Therefore, Lemma 3.1 implies
\[
\sum_{k=0}^{\infty} \|u_k - u_{k-1}\|^2 < \infty.
\]
\[
\square
\]

Let \(\mathcal{N}\) denote the set of all positive integers. For \(\lambda > 0\) and a positive integer \(\Delta\), we define
\[
\mathcal{K}^\lambda_{k, \Delta} := \{ i \in \mathcal{N} \mid \|s_{i-1}\| > \lambda, \ k \leq i \leq k + \Delta - 1 \}.
\]
Let \(|\mathcal{K}^\lambda_{k, \Delta}|\) denote the number of elements in \(\mathcal{K}^\lambda_{k, \Delta}\). The following lemma shows that if the magnitude of the gradient is bounded away from zero and (39)–(40) hold, then a certain fraction of the steps cannot be too small. This lemma corresponds to [1, Lemma 3] and [9, Lemma 4.2], and the proof is similar to that of [1, Lemma 3] and [9, Lemma 4.2], but we write it for the readability.
Lemma 3.5. Suppose that all assumptions of Lemma 3.3 hold. Then there exists \( \lambda > 0 \) such that, for any \( \Delta \in \mathbb{N} \) and any index \( k_0 \), there is an index \( k \geq k_0 \) such that
\[
|K^\lambda_{k,\Delta}| > \frac{\Delta}{2}.
\]

Proof. We proceed by contradiction. Suppose that for any \( \lambda > 0 \), there exist \( \Delta \in \mathbb{N} \) and \( k_0 \) such that
\[
|K^\lambda_{k,\Delta}| \leq \frac{\Delta}{2} \quad \text{for all } k \geq k_0.
\]
Let \( b > 1 \) and \( \xi > 0 \) be given in Lemma 3.3. We set \( \lambda = \xi \). Then it follows from (39), (40) and (46) that
\[
k_{0} + (i+1)\Delta \quad \prod_{k \in K_{k',\Delta}} \beta_k = \prod_{k \in K_{k',\Delta}} \beta_k \prod_{k \notin K_{k',\Delta}} \beta_k \leq b^{\Delta/2} \left( \frac{1}{2b} \right)^{\Delta/2} = \left( \frac{1}{2} \right)^{\Delta/2}
\]
for any \( i \geq 0 \), where \( k' = k_0 + i\Delta + 1 \). The relation (47) yields
\[
\prod_{j=k_0+1}^{k_0+(i+1)\Delta} 2\beta_j^2 = \prod_{j=k_0+1}^{k_0+i\Delta} 2\beta_j^2 \prod_{j=k_0+(i+1)\Delta+1}^{k_0+i\Delta} 2\beta_j^2 \leq 2^{i\Delta} \left( \frac{1}{2} \right)^{i\Delta} = 1.
\]
For all indices \( i \geq 1 \) and \( k_0 \leq l \leq k_0+i\Delta \), there exists an index \( i' \) such that \( k_0 + i'\Delta \leq l \leq k_0 + (i'+1)\Delta \), and we have
\[
\prod_{j=l}^{k_0+i\Delta} 2\beta_j^2 = \prod_{j=l}^{k_0+(i'+1)\Delta} 2\beta_j^2 \prod_{j=k_0+(i+1)\Delta+1}^{k_0+i\Delta} 2\beta_j^2 \leq \prod_{j=l}^{k_0+(i'+1)\Delta} 2\beta_j^2.
\]
Therefore, by (39), (49) and \( b \geq 1 \), we have
\[
\prod_{j=l}^{k_0+i\Delta} 2\beta_j^2 \leq (2b)^\Delta := t_1.
\]
We get from (34) that
\[
||dk_i||^2 \leq (|| - g_k + \zeta_k y_{k-1} || + \beta_k ||dk_{k-1}||^2) \leq 2 || - g_k + \zeta_k y_{k-1} || + 2\beta_k^2 ||dk_{k-1}||^2
\]
for all \( k \geq 1 \). Using (21), (22), (23) and (45), we obtain
\[
|| - g_k + \zeta_k y_{k-1} || \leq ||g_k|| + ||\zeta_k y_{k-1}||
\]
\[
\leq 2 ||g_k|| + 2 \max \left\{ \frac{\sigma}{1-\sigma}, 1 \right\} ||y_{k-1}||
\]
\[
\leq 2\Gamma + 4L\hat{\alpha} \max \left\{ \frac{\sigma}{1-\sigma}, 1 \right\}
\]
\[
:= t_2.
\]
Relations (48), (50), (51) and (52) yield
\[
||dk_{k_0+i\Delta}||^2 \leq 2 || - g_{k_0+i\Delta} + \zeta_{k_0+i\Delta} y_{k_0+i\Delta-1} || + 2\beta_{k_0+i\Delta}^2 ||dk_{k_0+i\Delta-1}||^2
\]
\[
\leq 2t_2^2 + 2\beta_{k_0+i\Delta}^2 (2t_2^2 + 2\beta_{k_0+i\Delta}^2 ||dk_{k_0+i\Delta-2}||^2)
\]
\[
\leq 2t_2^2 + 2t_2^2 \sum_{l=k_0+2}^{k_0+i\Delta} \prod_{j=l}^{k_0+i\Delta} 2\beta_j^2 + ||dk_l|| \prod_{j=k_0+1}^{k_0+i\Delta} 2\beta_j^2
\]
\[
\leq 2t_2^2 + 2(i\Delta - 1)t_1 t_2^2 + t_3,
\]
then it follows from (10), (18), (21), (22) and (26) that 

\( (i) \) Obviously, \( \hat{\theta} \) and 

Proof. From Theorem 3.6, to prove this corollary, it is sufficient to show that \( \hat{\theta} \) satisfies the Wolfe conditions (24)–(25). Then the method converges in the sense that (30) holds.

By using Theorem 3.6, the global convergence of our method with some concrete choices of parameters \( \hat{\gamma}_{k-1} \) and \( \theta_{k-1} \) can be easily obtained.

Corollary 3.7. Suppose that Assumption 1 is satisfied. Consider the method (2) and (34)–(36) with (17)–(18). Assume that the method has Property 1 and the step size \( \alpha_k \) satisfies the Wolfe conditions (24)–(25). Then the method converges in the sense that (30) holds.

Using Lemmas 3.4 and 3.5, we obtain the following global convergence theorem of our method. Since the proof of the theorem is exactly same as [6, Theorem 3.6], we omit it.

**Theorem 3.6.** Suppose that Assumption 1 is satisfied. Consider the method (2) and (34)–(36) with (17)–(18). Assume that the method has Property 1 and the step size \( \alpha_k \) satisfies the Wolfe conditions (24)–(25). Then the method converges in the sense that (30) holds.

(i) Let \( \theta_{k-1} \) be a constant satisfying \( 0 < \theta_{k-1} < 2 \) and set \( \hat{\gamma}_{k-1} = \theta_{k-1} \frac{y_k^T y_{k-1}}{y_{k-1}^T y_{k-1}} \).

If the step size \( \alpha_k \) satisfies the Wolfe conditions (24)–(25), then the method converges in the sense that (30) holds.

(ii) Let

\[
\theta_{k-1} = \frac{s_{k-1}^T y_{k-1}}{(s_{k-1}^T \hat{\gamma}_{k-1} y_{k-1}) y_{k-1}} \quad \text{and} \quad \hat{\gamma}_{k-1} = \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}},
\]

where \( \rho \geq 1 \) is a constant. If the step size \( \alpha_k \) satisfies the Wolfe conditions (24)–(25), then the method converges in the sense that (30) holds.

(iii) Let \( \theta_{k-1} = 1 + \frac{q_{k-1}^T d_{k-1}}{q_{k-1}^T y_{k-1}} \) and \( \hat{\gamma}_{k-1} = \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \). If the step size \( \alpha_k \) satisfies the strong Wolfe conditions (24) and

\[
|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k,
\]

then the method converges in the sense that (30) holds.

Proof. From Theorem 3.6, to prove this corollary, it is sufficient to show that \( \hat{\gamma}_{k-1} \) and \( \theta_{k-1} \) satisfy (17) and (18), respectively, and that the method has Property 1.

(i) Obviously, \( \hat{\gamma}_{k-1} \) and \( \theta_{k-1} \) satisfy (17) and (18), respectively. If (37) is satisfied, then it follows from (10), (18), (21), (22) and (26) that

\[
\hat{\gamma}_{k-1} = \frac{\theta_{k-1} y_{k-1}^T y_{k-1}}{\alpha_{k-1} q_{k-1}^T d_{k-1} y_{k-1}} \leq \frac{4L^2 \alpha_{k-1}^2}{\alpha_{k-1} c(1 - \sigma) \varepsilon^2} = \frac{4L^2 \alpha_{k-1}^2}{c(1 - \sigma) \varepsilon^2} ||d_{k-1}||,
\]

which means that the method has Property 1 with \( \tilde{c} = \frac{4L^2 \alpha_{k-1}^2}{c(1 - \sigma) \varepsilon^2} \).

(ii) Since \( \theta_{k-1} \) is rewritten as \( \theta_{k-1} = \frac{\theta_{k-1}^T}{1+\theta_{k-1}} \), \( \theta_{k-1} \) satisfies (18) because of \( \frac{1}{2} \leq \theta_{k-1} < 1 \). In a similar way to (i), we have from (10), (21), (22), (26) and (37)

\[
\hat{\gamma}_{k-1} = \frac{\theta_{k-1}^T y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \leq \frac{4L^2 \alpha_{k-1}^2}{c(1 - \sigma) \varepsilon^2} ||d_{k-1}||.
\]
Therefore, the method has Property 1 with \( \tilde{c} = \frac{2\rho L^2 \hat{\gamma}}{c(1-\sigma)\varepsilon^2} \).

(iii) From (53) and the fact \( g_k^T d_{k-1} < 0 \), we have

\[
1 \leq \theta_{k-1} = 1 + \left| \frac{g_k^T d_{k-1}}{g_k^T_1 d_{k-1}} \right| \leq 1 + \sigma < 2,
\]

which guarantees (18) with \( \theta_{\min} = 1 \) and \( \theta_{\max} = 1 + \sigma \). In a similar way to (i), we obtain from (10), (21), (22), (26) and (37)

\[
\hat{\gamma}_{k-1} = \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \leq \frac{2(1 + \sigma) L^2 \hat{\gamma}}{c(1-\sigma)\varepsilon^2} \|d_{k-1}\|.
\]

Therefore, the method has Property 1 with \( \tilde{c} = \frac{2(1 + \sigma) L^2 \hat{\gamma}}{c(1-\sigma)\varepsilon^2} \).

Note that the case (ii) in Corollary 3.7 corresponds to the memoryless quasi-Newton method based on the Hoshino formula with \( \tilde{\gamma}_{k-1} = \rho \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \), and that \( \theta_{k-1} \) belongs to the convex class. Moreover, if we use the choice

\[
\theta_{k-1} = 1 + \left| \frac{g_k^T d_{k-1}}{g_k^T_1 d_{k-1}} \right|,
\]

then we have the CG method with the Hestenes-Stiefel formula under the exact line search. This is a reason why we consider the case (iii) in Corollary 3.7.

4. Numerical experiments. In this section, we report numerical experiments to investigate numerical performance of the method of the form (2) and (34)–(36). We tested 138 problems from the CUTEr library [2, 10]. The problems were listed in Hager [11]. Although Hager [11] considered 145 tests, we did not consider the remaining test here due to the fact that the memory of our PC was insufficient for some of them. The names of test problems and their dimensions were given in Table 1. All codes were written in C by modifying the software package CG-DESCENT Version 5.3 [11, 12, 14]. They were run on a PC with 3.5GHz Intel Core i5, 32.0 GB RAM memory and Linux OS Ubuntu 16. We stopped the algorithm if \( \|g_k\|_\infty \leq 10^{-6} \) held or if CPU time exceeded 600 seconds. The line search procedure was the default procedure of CG-DESCENT, which means that the Wolfe conditions (24)–(25) with \( \sigma = 0.9 \) and \( \delta = 0.1 \) were used.

To compare numerical performance between the tested methods, we adopt the performance profiles based on the CPU time by Dolan and Moré [7]. For \( n_s \) solvers and \( n_p \) problems, the performance profiles \( P : \mathbb{R} \to [0, 1] \) is defined as follows: Let \( \mathcal{P} \) and \( \mathcal{S} \) be the set of problems and the set of solvers, respectively. For each problem \( p \in \mathcal{P} \) and for each solver \( s \in \mathcal{S} \), we define \( t_{p,s} \) CPU time required to solve problem \( p \) by solver \( s \). The performance ratio is given by \( r_{p,s} = t_{p,s} / \min_{s} t_{p,s} \). Then, the performance profile is defined by \( P(\tau) = \frac{1}{n_p} \sum_{p \in \mathcal{P}} \frac{1}{n_s} \sum_{s \in \mathcal{S}} I(r_{p,s} \leq \tau) \), for all \( \tau \geq 1 \), where sizeA, for any set \( A \), stands for number of the elements in that set. Note that \( P(\tau) \) is the probability for solver \( s \in \mathcal{S} \) such that a performance ratio \( r_{p,s} \) is within a factor \( \tau \geq 1 \) of the best result. The left side of the figure gives the percentage of the test problems for which a method is the best result, and the right side gives the percentage of the test problems that are successfully solved by each of the methods. The top curve is the method that solve the most problems in a result that is within a factor \( \tau \) of the best result. In order to prevent a measurement error, we set the minimum of the 0.1 seconds.
Table 1. Test problems (names and dimensions) by CUTEr library

| name    | n  | name    | n  | name    | n  |
|---------|----|---------|----|---------|----|
| AKIVA   | 2  | DIXMAANC3000 | HEART6LS | 8  | PENALTY1 | 1000 |
| ALLINTU | 4  | DIXMAAND3000 | HELIX | 3  | PENALTY2 | 200  |
| ARGLINA | 200 | DIXMAANE3000 | HIELOW | 3  | PENALTY3 | 200  |
| ARGLINB | 200 | DIXMAANF3000 | HILBERTA | 2  | POWELL6G | 5000 |
| ARWHEAD | 5000 | DIXMAANG3000 | HILBERTB | 10 | POWER | 10000 |
| BARD    | 3  | DIXMAANH3000 | HimmelBB| 2  | QUARTC | 5000 |
| BDQRTIC | 5000 | DIXMAANI3000 | HimmelBF| 4  | ROSENBR | 2   |
| BEALE   | 2  | DIXMAANJ3000 | HimmelBG| 2  | S308  | 2    |
| BIGGS6  | 6  | DIXMAANK3000 | HimmelBH| 2  | SCHMIVETT | 5000 |
| BOX3    | 3  | DIXMAANL3000 | HUMPS | 2  | SENSORS | 100  |
| BOX     | 10000 | DIXON3DQ10000 | JENSM | 2  | SINEVAL | 2    |
| BRKMCC  | 2  | DJTL | 2 | KOWOSB | 4  | SINEQUAD | 5000 |
| BROWNAL | 200 | DQDRTIC5000 | LIARWHID| 5000 | SISSE | 2    |
| BROWNs2 | 2  | DQRTIC5000 | LOGHAIRY| 2  | SNAIL | 2    |
| BROWNd4 | 4  | EDENSCH | 2000 | MANCINO | 100 | SPARSINE | 5000 |
| BROYD7D | 5000 | EG2  | 1000 | MARATOSB | 2  | SPARSQR | 10000 |
| BRYBND  | 5000 | ENGVAlI5000 | MEXHAT | 2  | SPRMRTS| 4999 |
| CHAINWOO | 4000 | ENGVAl2 | 3 | MOREBV | 5000 | SROSENBR | 5000 |
| CHNROSNB | 50 | ERRINROS50 | MSQRTALS | 1024 | STRATEC | 10   |
| CLIFF2  | 2  | EXPIT | 2 | MSQRTBLS | 1024 | TESTQUAD | 5000 |
| COSINE  | 10000 | EXTROSNB10000 | NONCVXU2 | 5000 | TOINTGOR | 50 |
| CRAAGLYV | 5000 | FLETCBV2 | 5000 | NONDIA | 5000 | TOINTGSS | 5000 |
| CUBE    | 2  | FLETCR | 1000 | NONDQR | 5000 | TOINTPS | 50 |
| CURLY10 | 10000 | FMINSRF2 | 5625 | OSORNEA | 5  | TOINTQR | 50 |
| CURLY20 | 10000 | FMINSURF | 5625 | OSORNEB | 11 | TQUARTIC | 5000 |
| CURLY30 | 10000 | FREUROTH | 5000 | OSCIPATH | 10 | TRIDIA | 5000 |
| DECONVU | 63 | GENHUMPS | 5000 | PALMER1C | 8  | VARDIM | 200 |
| DENSCHNA | 2  | GENROSE | 500 | PALMERID | 7  | VAREJGVL | 50 |
| DENSCHNB | 2  | GROWTHLS | 3 | PALMER2C | 8  | VIBREBAM | 8 |
| DENSCHNC | 2  | GULF | 3 | PALMER3C | 8  | WATSON | 12 |
| DENSCHNd | 3  | HAIRY | 2 | PALMER4C | 8  | WOODS | 4000 |
| DENSCHNE | 3  | HATFLDD | 3 | PALMER5C | 6  | YITU | 3 |
| DENSCHNF | 2  | HATFLDE | 3 | PALMER6C | 8  | ZANGWIL2 | 2 |
| DIXMAANA | 3000 | HATFLDFL | 3 | PALMER7C | 8  | |
| DIXMAANB | 3000 | HEART6LS | 6  | PALMER8C | 8  | |

4.1. Numerical comparison with convex and preconvex classes. In this subsection, we compare the convex class with the preconvex class of the Broyden family. Specifically, we investigate both the cases that $\theta_{k-1}$ is constant and $\theta_{k-1}$ is variable.

4.1.1. Numerical comparison of constant parameter $\theta_{k-1}$. We set $\theta_{k-1}$ to be constant numbers, and we use the following parameter

$$
\hat{\gamma}_{k-1} = \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}},
$$

which guarantees that the method has Property 1 from Corollary 3.7. Tables 2–3 present the tested methods for constant parameters $\theta_{k-1}$. In Method 1 (DFP), $\theta_{k-1} = 0$ yields $\hat{\gamma}_{k-1} = 0$ by (54). Thus we set $\hat{\gamma}_{k-1} = 0.01 \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}$ in Method 1 (DFP) instead of (54). Note that the global convergence properties of Methods 1 (DFP) and 9 are not guaranteed.

In Figure 1, we compare the methods in Table 2. Method 5 (BFGS) is superior to the other methods. However, since we observe that Methods 6–9 performed better
than Methods 1 (DFP) and 2–4 did, we can see that the preconvex class is a better choice than the convex class excluding $\theta_{k-1} = 1$. In each class, the performance of the method becomes better as $\theta_{k-1}$ approaches 1. Figure 2 gives the performance profiles of the methods in Table 3 in which the parameter $\theta_{k-1}$ is chosen close to 1. Then, we have a similar tendency to Figure 1.

Summarizing the above observations, the preconvex class is better than the convex class excluding $\theta_{k-1} = 1$. However, $\theta_{k-1} = 1$ is superior to the other parameters when we use constant numbers as $\theta_{k-1}$.

4.1.2. Numerical comparison of variable parameter $\theta_{k-1}$. We next consider the case where the value for the parameter $\theta_{k-1}$ is a variable number. In order to guarantee the condition (18), we modify $\theta_{k-1}$ as follows

$$\theta_{k-1} = \min \left\{ \theta_{\text{max}}, \max \left\{ \theta_{\text{min}}, \hat{\theta}_{k-1} \right\} \right\},$$

where $\theta_{\text{min}} = 0.1$ and $\theta_{\text{max}} = 1.9$. Table 4 gives the choices of the parameter $\hat{\theta}_{k-1}$. For Methods 5 (BFGS) and 19–21, we choose the scaling parameter $\hat{\gamma}_{k-1}$ in (54), and for Method 18 (Hoshino), we choose the scaling parameter $\hat{\gamma}_{k-1} = \frac{y_{k-1}^T y_k - y_{k-1}^T y_{k-1}}{s_{k-1}^T y_k - s_{k-1}^T y_{k-1}}$. Method 18 (Hoshino) corresponds to the proposed method based on the Hoshino formula and its global convergence is established by (ii) of Corollary 3.7. It follows from Corollary 3.7 that the global convergence properties of all methods in Table 4 are guaranteed. In Methods 19–21, we choose the parameters $\hat{\theta}_{k-1}$ so that they become 1 under the exact line search. Note that Methods 5 (BFGS) and 18 (Hoshino) belong

---

### Table 2. Tested methods (Methods 1–9)

| Method number | $\theta_{k-1}$ | Class   | Global convergence |
|---------------|----------------|---------|--------------------|
| 1 (DFP)       | 0              | convex  | not established    |
| 2             | 0.25           | convex  | established        |
| 3             | 0.5            | convex  | established        |
| 4             | 0.75           | convex  | established        |
| 5 (BFGS)      | 1              | convex  | established        |
| 6             | 1.25           | preconvex | established  |
| 7             | 1.5            | preconvex | established  |
| 8             | 1.75           | preconvex | established  |
| 9             | 2              | preconvex | not established |

### Table 3. Tested methods (Methods 5 and 10–17)

| Method number | $\theta_{k-1}$ | Class   | Global convergence |
|---------------|----------------|---------|--------------------|
| 5 (BFGS)      | 1              | convex  | established        |
| 10            | 0.8            | convex  | established        |
| 11            | 0.85           | convex  | established        |
| 12            | 0.9            | convex  | established        |
| 13            | 0.95           | convex  | established        |
| 14            | 1.05           | preconvex | established  |
| 15            | 1.1            | preconvex | established  |
| 16            | 1.15           | preconvex | established  |
| 17            | 1.2            | preconvex | established  |
to the convex class, and that Methods 19 and 21 belong to the preconvex class. On the other hand, since the sign of $d_{k-1}^T g_k$ is unknown, it is not clear whether Method 20 belongs to the convex/preconvex class or not.

In Figure 3, we observe that Method 19 is comparable with Method 5 (BFGS), and Method 21 performed slightly better than Method 5 (BFGS) did. Method
Table 4. Tested methods (Methods 5 and 18–21)

| Method number | $\theta_{k-1}$ | Class | Global convergence |
|---------------|----------------|-------|---------------------|
| 5 (BFGS)      | $1$            | convex| established         |
| 18 (Hoshino)  | $\frac{s_k^T y_{k-1}}{(s_{k-1} + \gamma_{k-1} y_{k-1})^T y_{k-1}}$ | convex| established         |
| 19            | $1 + \frac{d_{k-1}^T g_k}{d_{k-1}^T g_{k-1}}$ | preconvex| established |
| 20            | $1 + \frac{\|d_{k-1}\|\|g_k\|}{d_{k-1}^T g_k}$ | unknown| established |
| 21            | $1 + \frac{\|d_{k-1}\|\|g_k\|}{\|d_{k-1}\|\|g_k\|}$ | preconvex| established |

20 did not outperform Methods 5 (BFGS), 19 and 21, and Method 18 (Hoshino) performed poorly. Therefore, we see that the preconvex class is a better choice than the convex class, at least in these results. Moreover, although the preconvex class is inferior to the choice $\theta_{k-1} = 1$ when the parameter is constant, we find a better choice of parameters from the preconvex class if we choose $\theta_{k-1}$ adoptively. Thus, we expect that more efficient parameters can be found from the preconvex class and it is our further study.

Figure 3. Performance profiles of methods in Table 4

4.2. Numerical comparison of scaling parameter $\gamma_{k-1}$. Finally, we investigate how a choice of the scaling parameter $\gamma_{k-1}$ in (35) affects numerical performance. As a choice of common parameters for the self-scaling BFGS formula, Oren
Table 5. Tested methods (Methods 5, 19 and 21–25)

| Method number | $\hat{\gamma}_{k-1}$ | $\hat{\theta}_{k-1}$ | Global convergence |
|---------------|-----------------|-----------------|-------------------|
| 5 (BFGS)      | (54)            | 1               | established       |
| 19            | (54)            | $1 + \left| \frac{d^T_{k-1}g_k}{d^T_{k-1}g_{k-1}} \right|$ | established       |
| 21            | (54)            | $1 + \|d_{k-1}\|\|g_k\|$ | established       |
| 22 (BFGS)     | (56)            | 1               | not established   |
| 23            | (56)            | $1 + \left| \frac{d^T_{k-1}g_k}{d^T_{k-1}g_{k-1}} \right|$ | not established   |
| 24            | (56)            | $1 + \|d_{k-1}\|\|g_k\|$ | not established   |
| 25 (CG_DESCENT) [11, 12, 14] | established |

[25] and Oren and Luebenberg [26] suggested the following parameters:

$$\frac{s^T_{k-1}y_{k-1}}{s^T_{k-1}B_{k-1}s_{k-1}} \quad \text{and} \quad \frac{y^T_{k-1}H_{k-1}y_{k-1}}{s^T_{k-1}y_{k-1}},$$

These parameters correspond to

$$\hat{\gamma}_{k-1} = \theta_{k-1} \frac{s^T_{k-1}y_{k-1}}{s^T_{k-1}s_{k-1}} \quad \text{(56)}$$

and (54) in our method, respectively. In this subsection, we deal with (56) in addition to (54). As a choice of $\hat{\theta}_{k-1}$, we choose

$$\hat{\theta}_{k-1} = 1 + \left| \frac{d^T_{k-1}g_k}{d^T_{k-1}g_{k-1}} \right| \quad \text{and} \quad \hat{\theta}_{k-1} = 1 + \left| \frac{d^T_{k-1}g_k}{\|d_{k-1}\|\|g_k\|} \right|,$$

because these parameters give better numerical results in Figure 3. Then we give $\hat{\gamma}_{k-1}$ and $\hat{\theta}_{k-1}$ in Table 5. We also modify $\hat{\theta}_{k-1}$ as in (55). Method 25 is CG_DESCENT Version 5.3 which is the well-known benchmark software package based on the CG method by Hager and Zhang [11, 12, 14]. Note that the global convergence properties of Methods 5 (BFGS), 19 and 21 are guaranteed. On the other hand, since

$$\frac{s^T_{k-1}y_{k-1}}{s^T_{k-1}s_{k-1}} \leq \frac{y^T_{k-1}y_{k-1}}{s^T_{k-1}y_{k-1}},$$

holds, the choice (56) may not satisfy the condition (17). Thus the global convergence properties of Methods 22 (BFGS), 23 and 24 may not be guaranteed.

Figure 4 shows the performance profiles of the methods in Table 5. We observe that Methods 23 and 24 performed better than Method 25 (CG_DESCENT) did and Method 22 (BFGS) is almost comparable with Method 25 (CG_DESCENT). Also, Methods 22 (BFGS), 23, and 24 are superior to Methods 5 (BFGS), 19 and 21. Thus, we see that (56) is more efficient than (54). However, the global convergence of the method with (56) has not been established, and hence it is one of the further studies. Furthermore, since we observe that Methods 23 and 24 performed better than Method 22 (BFGS) did, the preconvex class is significant for the proposed method as in Section 4.1.2.
5. **Concluding remarks.** In this paper, we have proposed the memoryless quasi-Newton methods based on the Broyden family with the spectral-scaling secant condition, which always satisfy the sufficient descent condition. In addition, we have given the conditions for the parameters $\theta_{k-1}$ and $\hat{\gamma}_{k-1}$ in (34)–(36) that guarantee the global convergence of the method. In numerical experiments, although $\theta_{k-1} = 1$ (namely the BFGS formula) is a better choice, we have found good parameters in the preconvex class. Moreover, if we choose the parameters $\theta_{k-1}$ and $\hat{\gamma}_{k-1}$ suitably, then our methods are superior to or at least comparable with CG DESCENT Version 5.3. Therefore, we suggest that it is important to consider the preconvex class. Our further works are to find more suitable choices for $\theta_{k-1}$ and $\hat{\gamma}_{k-1}$, and to relax conditions on parameters $\theta_{k-1}$ and $\hat{\gamma}_{k-1}$ for the global convergence property.

In this paper, we have dealt with the spectral-scaling secant condition (12) by [4]. This is based on the secant condition (4). To extend the proposed method, we can consider the spectral-scaling secant condition based on other secant conditions (see [8, 18, 31, 32]). For example, the spectral-scaling secant condition based on the MBFGS secant condition [18] is given by

$$B_{k}s_{k-1} = \gamma_{k-1}z_{k-1}, \quad z_{k-1} = y_{k-1} + p\|g_{k-1}\|^2s_{k-1},$$

where $p$ and $q$ are positive constants. By replacing the vector $y_{k-1}$ in (34)–(36) with $z_{k-1}$, we can present the method based on the above secant condition and prove its global convergence property in a similar way to the proofs in this paper. Also, the other secant conditions can be considered similarly.

**Acknowledgment.** This research is supported in part by JSPS KAKENHI (grant number 17K00039). The authors are grateful to the anonymous referees whose comments helped to improve the paper.
REFERENCES

[1] M. Al-Baali, Y. Narushima and H. Yabe, A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization, Computational Optimization and Applications, 60 (2015), 89–110.

[2] I. Bongart, A. R. Conn, N. I. M. Gould and P. L. Toint, CUTE: constrained and unconstrained testing environment, ACM Transactions on Mathematical Software, 21 (1995), 123–160.

[3] Z. Chen and W. Cheng, Spectral-scaling quasi-Newton methods with updates from the one parameter of the Broyden family, Journal of Computational and Applied Mathematics, 248 (2013), 88–98.

[4] W. Y. Cheng and D. H. Li, Spectral scaling BFGS method, Journal of Optimization Theory and Applications, 146 (2010), 305–319.

[5] Y. H. Dai and C. X. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search, SIAM Journal on Optimization, 23 (2013), 296–320.

[6] Y. H. Dai and L. Z. Liao New conjugacy conditions and related nonlinear conjugate gradient methods, Applied Mathematics and Optimization, 43 (2001), 87–101.

[7] E. D. Dolan and J. J. Moré, Benchmarking optimization software with performance profiles, Mathematical Programming, 91 (2002), 201–213.

[8] J. A. Ford and I. A. Moghrabi, Multi-step quasi-Newton methods for optimization, Journal of Computational and Applied Mathematics, 248 (2013), 88–98.

[9] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM Journal on Optimization, 2 (1992), 21–42.

[10] N.I.M. Gould, D. Orban, and P.L. Toint, CUTEr and SifDec: A constrained and unconstrained testing environment, revisited, ACM Transactions on Mathematical Software, 29 (2003), 373–394.

[11] W. W. Hager, Hager’s web page: https://people.clas.ufl.edu/hager/.

[12] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal on Optimization, 16 (2005), 170–192.

[13] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient method, Pacific Journal of Optimization, 2 (2006), 35–58.

[14] W. W. Hager and H. Zhang, Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent, ACM Transactions on Mathematical Software, 32 (2006), 113–137.

[15] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards, 49 (1952), 499–506.

[16] S. Hoshino, A formulation of variable metric methods, IMA Journal of Applied Mathematics, 10 (1972), 394–403.

[17] C. X. Kou and Y. H. Dai, A modified self-scaling memoryless Broyden-Fletcher-Goldfarb-Shanno method for unconstrained optimization, Journal of Optimization Theory and Applications, 165 (2015), 209–224.

[18] D. H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, Journal of Computational and Applied Mathematics, 129 (2001), 15–35.

[19] F. Modarres, M. A. Hassan and W. J. Leong, Memoryless modified symmetric rank-one method for large-scale unconstrained optimization, American Journal of Applied Sciences, 6 (2009), 2054–2059.

[20] A.U. Moyi and W.J. Leong, A sufficient descent three-term conjugate gradient method via symmetric rank-one update for large-scale optimization, Optimization, 65 (2016), 121–143.

[21] S. Nakayama, Y. Narushima and H. Yabe, A memoryless symmetric rank-one method with sufficient descent property for unconstrained optimization, Journal of the Operations Research Society of Japan, 61 (2018), 53–70.

[22] Y. Narushima and H. Yabe, A survey of sufficient descent conjugate gradient methods for unconstrained optimization, SUT Journal of Mathematics, 50 (2014), 167–203.

[23] Y. Narushima, H. Yabe and J. A. Ford, A three-term conjugate gradient method with sufficient descent property for unconstrained optimization, SIAM Journal on Optimization, 21 (2011), 212–230.

[24] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd edition, Springer, 2006.

[25] S. S. Oren, Self-scaling variable metric (SSVM) algorithms, Part II: Implementation and experiments, Management Science, 20 (1974), 863–874.
[26] S. S. Oren and D. G. Luenberger, Self-scaling variable metric (SSVM) algorithms, Part I: Criteria and sufficient conditions for scaling a class of algorithms, Management Science, 20 (1974), 845–862.

[27] D. F. Shanno, Conjugate gradient methods with inexact searches, Mathematics of Operations Research, 3 (1978), 244–256.

[28] K. Sugiki, Y. Narushima and H. Yabe, Globally convergent three-term conjugate gradient methods that use secant condition and generate descent search directions for unconstrained optimization, Journal of Optimization Theory and Applications, 153 (2012), 733–757.

[29] L. Sun, An approach to scaling symmetric rank-one update, Pacific Journal of Optimization, 2 (2006), 105–118.

[30] W. Sun and Y. Yuan, Optimization Theory and Methods: Nonlinear Programming, Springer, 2006.

[31] Z. Wei, G. Li and L. Qi, New quasi-Newton methods for unconstrained optimization problems, Applied Mathematics and Computation, 175 (2006), 1156–1188.

[32] J. Z. Zhang, N. Y. Deng and L. H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, Journal of Optimization Theory and Applications, 102 (1999), 147–167.

[33] L. Zhang, W. Zhou and D. H. Li, A descent modified Polak-Ribi`ere-Polyak conjugate gradient method and its global convergence, IMA Journal of Numerical Analysis, 26 (2006), 629–640.

[34] Y. Zhang and R. P. Tewarson, Quasi-Newton algorithms with updates from the preconvex part of Broyden’s family, IMA Journal of Numerical Analysis, 8 (1988), 487–509.

Received July 2017; revised February 2018.

E-mail address: 1416702@ed.tus.ac.jp
E-mail address: narushima@ynu.ac.jp
E-mail address: yabe@rs.kagu.tus.ac.jp