SUT Journal of Mathematics
Vol. 54, No. 1 (2018), 79–98

A hybrid method of three-term conjugate gradient method and memoryless quasi-Newton method for unconstrained optimization

Shummin Nakayama

(Received September 29, 2017; Revised May 21, 2018)

Abstract. Memoryless quasi-Newton methods are studied for solving large-scale unconstrained optimization problems. Nakayama et al. (2017) proposed a memoryless quasi-Newton method based on the spectral-scaling Broyden family and showed that the method satisfies the sufficient descent condition and converges globally. To relax the conditions on parameters in the method, we apply the modification technique by Kou and Dai (2015) to the method of Nakayama et al., and we give a hybrid method of the three-term conjugate gradient method and the memoryless quasi-Newton method based on the spectral-scaling Broyden family. We show that our method satisfies the sufficient descent condition, and we prove that the method converges globally. Furthermore, we give a concrete choice of parameters for our method. Finally, some numerical results are given.

AMS 2010 *Mathematics Subject Classification.* 90C30, 90C06.

Key words and phrases. Unconstrained optimization, memoryless quasi-Newton method, Broyden family, three-term conjugate gradient method, sufficient descent condition, global convergence.

§1. Introduction

Quasi-Newton methods are known as effective numerical methods for solving the following unconstrained optimization problem

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

where f is a smooth function. We denote its gradient ∇f by g. The quasi-Newton method is an iterative method of the form:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.1}$$

where $x_k \in \mathrm{R}^n$ is the k-th approximation to a solution, the step size $\alpha_k > 0$ is obtained by some line search, and the search direction d_k is given by

$$d_k = -H_k g_k. \tag{1.2}$$

Here, we denote $\nabla f(x_k)$ by g_k, and H_k is an approximation to the inverse Hessian $\nabla^2 f(x_k)^{-1}$. In this paper, we fix the initial direction by $d_0 = -g_0$. The matrix H_k is updated at each iteration such that the secant condition

$$H_k y_{k-1} = s_{k-1}$$

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. The well-known updating formulas are the BFGS, DFP and symmetric rank-one (SR1) formulas. This paper focuses on the Broyden family

$$H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} + \theta_{k-1} w_{k-1} w_{k-1}^T, \tag{1.3}$$

$$w_{k-1} = \sqrt{y_{k-1}^T H_{k-1} y_{k-1}} \left(\frac{s_{k-1}}{s_{k-1}^T y_{k-1}} - \frac{H_{k-1} y_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} \right), \tag{1.4}$$

where θ_{k-1} is a parameter. In particular, the formula (1.3) becomes the DFP formula when $\theta_{k-1} = 0$, and the BFGS formula when $\theta_{k-1} = 1$. Moreover, the Broyden family (1.3) is a convex combination of the DFP formula and the BFGS formula if $\theta_{k-1} \in [0, 1]$, and we say this interval convex class. The BFGS formula ($\theta_{k-1} = 1$) is known as one of the best choices in practice. On the other hand, Zhang and Tewarson [27] dealt with the preconvex class to find a better choice than the BFGS formula. The preconvex class means the interval $\theta_{k-1} > 1$. 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 (1.3) is also symmetric positive definite. This guarantees that the search direction satisfies the descent condition, namely, $g_k^T d_k = -g_k^T H_k g_k < 0$.

Since quasi-Newton methods need the storage of memories for matrices, it is difficult to apply quasi-Newton methods directly to large-scale unconstrained optimization problems. In order to remedy this difficulty, Shanno [22] proposed the memoryless quasi-Newton method for solving large-scale unconstrained optimization problems, and the method avoids the storage of memories for matrices. Specifically, Shanno substituted (1.3) and (1.4) with

$H_{k-1} = I$ into (1.2) and derived the following search direction:

$$\begin{aligned} d_k = &- g_k + \left(\theta_{k-1} \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(1 + \theta_{k-1} \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) d_{k-1} \\ &+ \left(\theta_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1 - \theta_{k-1}) \frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}} \right) y_{k-1}. \end{aligned} \tag{1.5}$$

The memoryless quasi-Newton method is closely related to the nonlinear conjugate gradient (CG) method. Under the exact line search, namely $g_k^T d_{k-1} = 0$, then the search direction (1.5) with $\theta_{k-1} = 1$ becomes

$$d_k = -g_k + \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} d_{k-1}, \tag{1.6}$$

which is identical to the nonlinear CG method with the Hestenes-Stiefel (HS) formula (see [17, 19], for example). Recently, three-term CG methods have been paid attention to (see [1, 18, 26], for example). By using the memoryless quasi-Newton method based on the BFGS formula, several three-term CG methods have been proposed (see [2, 24, 25], for example).

In this decade, several memoryless quasi-Newton methods have been studied. Kou and Dai [14] proposed the modified self-scaling memoryless BFGS method. Furthermore, several researchers have paid attention to the memoryless quasi-Newton methods based on other updating formulas instead of the BFGS formula. Nakayama et al. [15] proposed the memoryless quasi-Newton method based on the SR1 formula with the spectral-scaling secant condition [5]. The above methods always satisfy the sufficient descent condition which means that there exists a positive constant c such that

$$g_k^T d_k \leq -c \|g_k\|^2 \quad \text{for all } k, \tag{1.7}$$

where $\|\cdot\|$ is the ℓ_2 norm. Moreover, they showed the global convergence of the method for general objective functions. Nakayama et al. [16] also proposed the memoryless quasi-Newton method based on the spectral-scaling Broyden family [4] and gave a sufficient condition for the global convergence of the method. In their numerical experiments, they showed that the proposed method with the preconvex class performed better than the method with the convex class did.

This paper focuses on the modification of the method by Kou and Dai [14]. Considering their modification technique, we modify the memoryless quasi-Newton method based on the Broyden family [16] and propose a new method, which always satisfies the sufficient descent condition. We show that the method converges globally for general objective functions. Furthermore,

we give parameters for which our method can be regarded as a three-term CG method.

This paper is organized as follows. We recall some preliminaries and previous researches in Section 2. In Section 3, we propose a new method and show its global convergence. Finally, some numerical results are given.

§2. Preliminaries and previous researches

2.1. Preliminaries

In this subsection, we recall some preliminaries. We first make the following standard assumptions for the objective function.

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

$$\|x\| \leq \hat{a} \quad \text{for all } x \in \mathcal{L}. \tag{2.1}$$

(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

$$\|g(u) - g(v)\| \leq L\|u - v\| \quad \text{for all } u, v \in \mathcal{N}. \tag{2.2}$$

Note that, under Assumption 1, there exists a positive constant Γ such that

$$\|g(x)\| \leq \Gamma \quad \text{for all } x \in \mathcal{L}. \tag{2.3}$$

Throughout this paper, we assume that

$$g_k \neq 0 \quad \text{for all } k \geq 0,$$

otherwise a stationary point has been found.

In the line search procedure, we require the step size α_k in (1.1) to satisfy the Wolfe conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{2.4}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{2.5}$$

where $0 < \delta < \sigma < 1$. If the search direction satisfies the descent condition, then condition (2.5) yields

$$d_k^T y_k = g_{k+1}^T d_k - g_k^T d_k \geq -(1 - \sigma) g_k^T d_k > 0. \tag{2.6}$$

Since this paper deals with the Wolfe condition (2.4)–(2.5), we have

$$d_k^T y_k > 0 \quad (s_k^T y_k > 0) \quad \text{for all } k \geq 0, \tag{2.7}$$

which implies $\|s_k\| \neq 0$ and $\|y_k\| \neq 0$.

2.2. Modified self-scaling memoryless BFGS method

In this subsection, we review the memoryless quasi-Newton method by Kou and Dai [14]. Modifying the memoryless quasi-Newton method based on the self-scaling BFGS method, Kou and Dai [14] proposed the following search direction:

$$\begin{aligned} d_k = &- g_k + \left(\frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\tau_{k-1} + \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} \right) d_{k-1} \\ &+ \nu_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}, \end{aligned} \tag{2.8}$$

where $\nu_{k-1} \in [0, 1]$ is a parameter and τ_{k-1} is a scaling parameter satisfying

$$\tau_{k-1} \in \left[\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}, \ \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right],$$

which corresponds to the interval proposed by Oren and Luenberger [20, 21]. We note that since $s_{k-1}^T y_{k-1} > 0$,

$$\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}} \leq \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \tag{2.9}$$

holds by the Cauchy-Schwarz inequality. The search direction (2.8) with $\nu_{k-1} = 0$ can be regarded as a CG method of the form $d_k = -g_k + \beta_k d_{k-1}$, where β_k is defined by

$$\beta_k = \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\tau_{k-1} + \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}.$$

Furthermore, if $g_k^T d_{k-1} = 0$, then the search direction (2.8) becomes (1.6).

The following lemma was given by [14, Lemma 3.1].

Lemma 2.1. *If $\nu_{k-1} \in [0, 1)$ is a constant and* (2.7) *holds, then the search direction* (2.8) *satisfies the sufficient descent condition* (1.7) *for some positive constant c.*

In order to guarantee the global convergence, Kou and Dai modified the direction (2.8) as follows:

$$d_k = -g_k + \beta_k^{KD} d_{k-1} + \zeta_k^{KD} y_{k-1}, \tag{2.10}$$

$$\beta_k^{KD} = \max\left\{ \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\tau_{k-1} + \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}\right) \frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}, \rho \frac{g_k^T d_{k-1}}{\|d_{k-1}\|^2} \right\}$$

and

$$\zeta_k^{KD} = \begin{cases} 0 & \text{if } \beta_k^{KD} = \rho \frac{g_k^T d_{k-1}}{\|d_{k-1}\|^2} \\ \nu_{k-1} \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} & \text{otherwise,} \end{cases}$$

where $\rho \in (0,1)$ is a parameter. Note that, the above modified search direction always satisfies the sufficient descent condition (1.7). In the line search procedure, they used the improved Wolfe conditions [6]:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \min\left\{\epsilon |f(x_k)|, \delta \alpha_k g_k^T d_k + \eta_k\right\}, \tag{2.11}$$

and (2.5), where $0 < \delta < \sigma < 1$, $\epsilon > 0$ is a small number and $\eta_k = 1/k^2$.

The following theorem was given by [14, Theorem 4.2].

Theorem 2.2. *Suppose that Assumption* 1 *is satisfied. Consider the method* (1.1) *and* (2.10) *with* $\nu_{k-1} \in [0,1)$*, where the step size* α_k *satisfies the improved Wolfe conditions* (2.5) *and* (2.11)*. Then the method converges in the sense that*

$$\liminf_{k\to\infty} \|g_k\| = 0 \tag{2.12}$$

holds.

2.3. Memoryless quasi-Newton method based on the Broyden family

In this subsection, we recall the memoryless quasi-Newton method based on the Broyden family by Nakayama et al. [16]. They focused on the following spectral-scaling Broyden family [4]:

$$H_k = H_{k-1} - \frac{H_{k-1} y_{k-1} y_{k-1}^T H_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}} + \frac{1}{\gamma_{k-1}} \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} + \theta_{k-1} w_{k-1} w_{k-1}^T, \tag{2.13}$$

where w_{k-1} appears in (1.4) and $\gamma_{k-1} > 0$ is a scaling parameter. The formula (2.13) is the Broyden family based on the spectral-scaling secant condition [5]:

$$H_k y_{k-1} = \frac{1}{\gamma_{k-1}} s_{k-1},$$

where H_k is an approximation to $\left(\gamma_{k-1}\nabla^2 f(x_k)\right)^{-1}$. Nakayama et al. [16] gave the search direction of a memoryless quasi-Newton method based on (2.13), which is

$$\begin{aligned} d_k = &- g_k + \left(\theta_{k-1}\frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \theta_{k-1}\frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}\right)\frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}\right) d_{k-1} \\ &+ \left(\theta_{k-1}\frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1-\theta_{k-1})\,\frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}}\right) y_{k-1}, \end{aligned} \tag{2.14}$$

where $\hat{\gamma}_{k-1} = 1/\gamma_{k-1}$. Note that the search direction (2.14) with $\theta_{k-1} = 1$ corresponds to (2.8) with $\nu_{k-1} = 1$ and $\tau_{k-1} = \hat{\gamma}_{k-1}$. They gave the following proposition.

Proposition 2.3. *If conditions* (2.7)*,*

$$\hat{\gamma}_{k-1} \geq \theta_{k-1}\frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \tag{2.15}$$

and

$$0 < \theta_{min} \leq \theta_{k-1} \leq \theta_{max} < 2 \tag{2.16}$$

hold, then the search direction (2.14) *satisfies the sufficient descent condition* (1.7) *with* $c := \min\left\{\dfrac{\theta_{min}}{2},\ 1 - \dfrac{\theta_{max}}{2}\right\}$*, where* θ_{min} *and* θ_{max} *are constants satisfying* $0 < \theta_{min} \leq 1 \leq \theta_{max} < 2$.

In order to establish the global convergence of the method, Nakayama et al. [16] modified (2.14) as follows:

$$d_k = -g_k + \beta_k^{NNY} d_{k-1} + \zeta_k^{NNY} y_{k-1}, \tag{2.17}$$

$$\beta_k^{NNY} = \max\left\{\theta_{k-1}\frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \theta_{k-1}\frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}\right)\frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}, 0\right\}, \tag{2.18}$$

and

$$\zeta_k^{NNY} = \mathrm{sgn}(\beta_k^{NNY})\left(\theta_{k-1}\frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1-\theta_{k-1})\,\frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}}\right), \tag{2.19}$$

where $\mathrm{sgn}(\cdot)$ is defined by

$$\mathrm{sgn}(a) = \begin{cases} 1 & a > 0, \\ 0 & a = 0. \end{cases}$$

They proved the following convergence theorem [16, Theorem 3.6].

Theorem 2.4. *Suppose that Assumption* 1 *is satisfied. Consider the method* (1.1) *and* (2.17) *with*

$$\hat{\gamma}_{k-1} = \theta_{k-1}\frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \tag{2.20}$$

and (2.16)*, where the step size* α_k *satisfies the Wolfe conditions* (2.4)–(2.5)*. Then the method converges in the sense that* (2.12) *holds.*

§3. A hybrid method of three-term CG method and memoryless quasi-Newton method

In their numerical experiments, Nakayama et al. [16] showed that the method (1.1) and (2.17) with

$$\hat{\gamma}_{k-1} = \theta_{k-1}\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}. \tag{3.1}$$

performed better than the method with (2.20) did. However, (3.1) may not satisfy the condition (2.15), which dose not guarantee the global convergence of the method with (3.1). In order to relax the condition (2.15), we apply the modification technique by Kou and Dai [14] to the method of Nakayama et al. [16], and we give a new method. We show that the proposed method satisfies the sufficient descent condition without the condition (2.15), and the method converges globally. We note that the proposed method can adopt the parameter (3.1).

3.1. Proposed method

Based on the modification described in Section 2.2, we propose the following search direction

$$\begin{aligned} d_k = &- g_k + \left(\theta_{k-1}\frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \left(\hat{\gamma}_{k-1} + \theta_{k-1}\frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}\right)\frac{s_{k-1}^T g_k}{d_{k-1}^T y_{k-1}}\right) d_{k-1} \\ &+ \nu_{k-1}\left(\theta_{k-1}\frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + (1-\theta_{k-1})\frac{y_{k-1}^T g_k}{y_{k-1}^T y_{k-1}}\right) y_{k-1}, \end{aligned} \tag{3.2}$$

where ν_{k-1} is a parameter. Note that the search direction (3.2) with $\nu_{k-1} = 1$ is identical to (2.14), and that (3.2) with $\theta_{k-1} = 1$ corresponds to (2.8) with $\tau_{k-1} = \hat{\gamma}_{k-1}$. In the same way as (2.17), we modify (3.2) as follows:

$$d_k = -g_k + \beta_k^{NNY} d_{k-1} + \zeta_k^{new} y_{k-1} \tag{3.3}$$

and

$$\zeta_k^{new} = \nu_{k-1}\zeta_k^{NNY}, \tag{3.4}$$

where β_k^{NNY} and ζ_k^{NNY} are defined by (2.18) and (2.19), respectively. Note that, if $\beta_k^{NNY} > 0$ is satisfied, then the search direction (3.3) is identical to (3.2). Otherwise, the search direction (3.3) becomes the steepest descent direction ($d_k = -g_k$).

To establish the sufficient descent property of the proposed method, we impose the following condition on θ_{k-1}

$$0 \leq \theta_{k-1} \leq \bar{c}_1^2, \tag{3.5}$$

where $1 < \bar{c}_1 < 2$ is a constant. As a choice of ν_{k-1}, we consider the following:

$$\begin{cases} 0 \leq \nu_{k-1} \leq \bar{\nu}, & \text{if } 0 \leq \theta_{k-1} \leq 1, \\ 0 \leq \nu_{k-1} \leq \dfrac{\bar{c}_1}{\sqrt{\theta_{k-1}}} - 1, & \text{if } 1 < \theta_{k-1} \leq \bar{c}_1^2, \end{cases} \tag{3.6}$$

where $0 \leq \bar{\nu} < 1$. Then we give the sufficient condition for the search direction (3.3) to satisfy (1.7) as follows.

Proposition 3.1. *Assume that* (2.7) *holds. Then the search direction* (3.3) *with* (3.5)–(3.6) *satisfies the sufficient descent condition* (1.7) *for some positive constant* c.

Proof. By (2.18), we first note that $\beta_k^{NNY} \geq 0$ holds for all $k \geq 1$. For the case $\beta_k^{NNY} = 0$, the search direction (3.3) becomes the steepest descent direction which implies that the sufficient descent condition (1.7) with $c = 1$ holds. Otherwise, the search direction (3.3) is identical to (3.2). Thus, it is sufficient to show that (3.2) satisfies (1.7).

Using the relation $2u^T v \leq \|u\|^2 + \|v\|^2$ for any vectors u and v, we have

$$\begin{aligned} (1+\nu_{k-1})\frac{(y_{k-1}^T g_k)(d_{k-1}^T g_k)}{d_{k-1}^T y_{k-1}} &= \left(\frac{\sqrt{2}d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}\right)^T \left(\frac{(1+\nu_{k-1})}{\sqrt{2}} g_k\right) \\ &\leq \frac{1}{2}\left(\left\|\frac{\sqrt{2}d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}\right\|^2 + \left\|\frac{(1+\nu_{k-1})}{\sqrt{2}} g_k\right\|^2\right), \end{aligned}$$

and hence it follows from (2.7), (3.2) and $s_{k-1} = \alpha_{k-1}d_{k-1}$ that

$$\begin{aligned}
g_k^T d_k =& - \|g_k\|^2 + \theta_{k-1}(1+\nu_{k-1})\frac{(y_{k-1}^T g_k)(d_{k-1}^T g_k)}{d_{k-1}^T y_{k-1}} \\
& - \left(\hat{\gamma}_{k-1} + \theta_{k-1}\frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}\right)\frac{\alpha_{k-1}(d_{k-1}^T g_k)^2}{d_{k-1}^T y_{k-1}} + \nu_{k-1}(1-\theta_{k-1})\frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}} \\
\leq & - \|g_k\|^2 + \frac{\theta_{k-1}}{2}\left(\left\|\frac{\sqrt{2}d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}\right\|^2 + \left\|\frac{(1+\nu_{k-1})}{\sqrt{2}} g_k\right\|^2\right) \\
& - \left(\hat{\gamma}_{k-1} + \theta_{k-1}\frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}\right)\frac{\alpha_{k-1}(d_{k-1}^T g_k)^2}{d_{k-1}^T y_{k-1}} + \nu_{k-1}(1-\theta_{k-1})\frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}} \\
=& - \left(1 - \frac{\theta_{k-1}(1+\nu_{k-1})^2}{4}\right)\|g_k\|^2 - \hat{\gamma}_{k-1}\frac{\alpha_{k-1}(d_{k-1}^T g_k)^2}{d_{k-1}^T y_{k-1}} \\
& + \nu_{k-1}(1-\theta_{k-1})\frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}} \\
\leq & - \left(1 - \frac{\theta_{k-1}(1+\nu_{k-1})^2}{4}\right)\|g_k\|^2 + \nu_{k-1}(1-\theta_{k-1})\frac{(y_{k-1}^T g_k)^2}{y_{k-1}^T y_{k-1}}.
\end{aligned}$$

We consider the case $0 \leq \theta_{k-1} \leq 1$. Using $\nu_{k-1}(1-\theta_{k-1}) \geq 0$ and the Cauchy-Schwarz inequality,

$$\begin{aligned}
g_k^T d_k &\leq - \left(1 - \frac{\theta_{k-1}(1+\nu_{k-1})^2}{4}\right)\|g_k\|^2 + \nu_{k-1}(1-\theta_{k-1})\frac{\|y_{k-1}\|^2\|g_k\|^2}{\|y_{k-1}\|^2} \\
&= - \left((1-\nu_{k-1}) - \frac{\theta_{k-1}}{4}\left(1 - 2\nu_{k-1} + \nu_{k-1}^2\right)\right)\|g_k\|^2 \\
&= - \left((1-\nu_{k-1})\left(1 - \frac{\theta_{k-1}}{4}(1-\nu_{k-1})\right)\right)\|g_k\|^2
\end{aligned}$$

holds. Since it follows from $0 \leq \nu_{k-1}$ that

$$- \left(1 - \frac{\theta_{k-1}}{4}(1-\nu_{k-1})\right) \leq - \left(1 - \frac{1}{4} + 0\right) = -\frac{3}{4},$$

we obtain from $1 - \nu_{k-1} > 0$ and (3.6)

$$\begin{aligned}
g_k^T d_k &\leq -\frac{3(1-\nu_{k-1})}{4}\|g_k\|^2 \\
&\leq -\frac{3(1-\bar{\nu})}{4}\|g_k\|^2.
\end{aligned}$$

We next consider the case $1 < \theta_{k-1} \leq \bar{c}_1^2$. Since $\nu_{k-1}(1-\theta_{k-1}) \leq 0$ holds, we have from (3.6)

$$\begin{aligned} g_k^T d_k &\leq -\left(1 - \frac{\theta_{k-1}(1+\nu_{k-1})^2}{4}\right)\|g_k\|^2 \\ &\leq -\left(1 - \frac{\theta_{k-1}}{4}\frac{\bar{c}_1^2}{\theta_{k-1}}\right)\|g_k\|^2 \\ &= -\left(1 - \frac{\bar{c}_1^2}{4}\right)\|g_k\|^2. \end{aligned}$$

Therefore, the search direction (3.2) satisfies the sufficient descent condition (1.7) with $c = \min\left\{\frac{3(1-\bar{\nu})}{4},\ 1 - \frac{\bar{c}_1^2}{4}\right\}$. □

3.2. Global convergence

In this subsection, we prove the global convergence of the proposed method for general objective functions. We first introduce the following property.

Property 1. *Consider the method* (1.1) *and* (3.3). *Suppose that there exists a positive constant* ε *such that*

$$\varepsilon \leq \|g_k\| \quad \textit{for all } k \tag{3.7}$$

holds. Then we say that the method has Property 1 if there exists a positive constant $\bar{c}_2$ *such that*

$$\hat{\gamma}_{k-1} \leq \bar{c}_2\|d_{k-1}\| \quad \textit{for all } k.$$

The next proposition guarantees that the proposed method with some concrete choices of $\hat{\gamma}_{k-1}$ has *Property 1*.

Proposition 3.2. *Suppose that Assumption* 1 *is satisfied. Consider the method* (1.1) *and* (3.3), *where the step size* α_k *satisfies the Wolfe conditions* (2.4)–(2.5). *Then the following hold:*

1. *If we set*

$$\hat{\gamma}_{k-1} = \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}, \tag{3.8}$$

then the method has Property 1.

2. If we set

$$\hat{\gamma}_{k-1} = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}, \tag{3.9}$$

then the method has Property 1.

Proof. By (2.1) and (2.2), we obtain

$$\|y_{k-1}\| \leq L\|s_{k-1}\| \leq L(\|x_k\| + \|x_{k-1}\|) \leq 2L\hat{a}.$$

If (3.7) holds, then it follows from (1.7), (2.2), (2.6) and (2.9) that

$$\frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}} \leq \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \leq \frac{2L^2\hat{a}\|s_{k-1}\|}{\alpha_{k-1}c(1-\sigma)\varepsilon^2} = \frac{2L^2\hat{a}}{c(1-\sigma)\varepsilon^2}\|d_{k-1}\|.$$

Therefore, the methods have *Property 1.* □

We note that the proposed method with (2.20) has *Property 1.* Also, the method with (3.1) has *Property 1.* We emphasize that the proposed method can choose the parameter (3.1), which is a good choice in the previous research [16].

The following lemma corresponds to [16, Lemma 3.3]. Since the proof is almost same as that of [16, Lemma 3.3], we omit it. Note that the lemma imposes a different condition from that of [16, Lemma 3.3] on θ_{k-1}.

Lemma 3.3. *Consider the method* (1.1) *and* (3.3) *with* (3.5)*, where the step size α_k satisfies the Wolfe conditions* (2.4)–(2.5)*. Suppose that Assumption* 1 *is satisfied and there exists a positive constant ε such that* (3.7) *holds. If the method has Property 1, then there exist constants $b > 1$ and $\xi > 0$ such that for all k*

$$\beta_k^{NNY} \leq b \tag{3.10}$$

and

$$\|s_{k-1}\| \leq \xi \Longrightarrow \beta_k^{NNY} \leq \frac{1}{2b}. \tag{3.11}$$

The next lemma corresponds to [7, 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 (1.7) and (3.7), the vector u_k is well-defined. By defining

$$v_k = -\frac{1}{\|d_k\|}(g_k - \zeta_k^{new} y_{k-1}) \quad \text{and} \quad \eta_k = \beta_k^{NNY}\frac{\|d_{k-1}\|}{\|d_k\|},$$

equation (3.3) is written as

$$u_k = v_k + \eta_k u_{k-1}.$$

Then we get from the fact $\|u_k\| = \|u_{k-1}\| = 1$

$$\|v_k\| = \|u_k - \eta_k u_{k-1}\| = \|\eta_k u_k - u_{k-1}\|. \tag{3.12}$$

From the relations $\beta_k^{NNY} \geq 0$ and (3.12), we obtain

$$\begin{aligned} \|u_k - u_{k-1}\| &\leq (1+\eta_k)\|u_k - u_{k-1}\| \\ &= \|u_k - \eta_k u_{k-1} + \eta_k u_k - u_{k-1}\| \\ &\leq \|u_k - \eta_k u_{k-1}\| + \|\eta_k u_k - u_{k-1}\| \\ &= 2\|v_k\|. \end{aligned} \tag{3.13}$$

Since the search direction satisfies the descent condition, it follows from (2.6) that

$$g_k^T d_{k-1} \geq \sigma g_{k-1}^T d_{k-1} \geq \frac{-\sigma}{1-\sigma} d_{k-1}^T y_{k-1}$$

and

$$d_{k-1}^T y_{k-1} = g_k^T d_{k-1} - g_{k-1}^T d_{k-1} \geq g_k^T d_{k-1}.$$

Hence we obtain

$$\left|\frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}\right| \leq \max\left\{\frac{\sigma}{1-\sigma}, 1\right\}. \tag{3.14}$$

Using $\nu_{k-1} < 1$, $0 \leq \theta_{k-1} < 4$, (3.4) and (3.14), we get

$$\begin{aligned} \|\zeta_k^{new} y_{k-1}\| &\leq \nu_{k-1}\left(\left\|\theta_{k-1}\frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1}\right\| + |1-\theta_{k-1}|\frac{\|y_{k-1}\|\|g_k\|}{\|y_{k-1}\|^2}\|y_{k-1}\|\right) \\ &< 4\max\left\{\frac{\sigma}{1-\sigma},\ 1\right\}\|y_{k-1}\| + 3\|g_k\|. \end{aligned} \tag{3.15}$$

By (2.1), (2.2), (2.3), (3.13) and (3.15), for any positive integer m, we have

$$\begin{aligned}\sum_{k=0}^{m}\|u_k-u_{k-1}\|^2 &\le 4\sum_{k=0}^{m}\|v_k\|^2\\ &\le 4\sum_{k=0}^{m}\left(4\|g_k\|+4\max\left\{\frac{\sigma}{1-\sigma},\ 1\right\}\|y_{k-1}\|\right)^2\frac{1}{\|d_k\|^2}\\ &\le 64\left(\Gamma+2L\hat{a}\max\left\{\frac{\sigma}{1-\sigma},\ 1\right\}\right)^2\sum_{k=0}^{m}\frac{1}{\|d_k\|^2}.\end{aligned}$$

Since the search direction satisfies the sufficient descent condition, we obtain $\sum_{k=0}^{\infty}\frac{1}{\|d_k\|^2}<\infty$ by (3.7) (see [23, Lemma 3.1]). Therefore, we have $\sum_{k=0}^{\infty}\|u_k-u_{k-1}\|^2<\infty$. □

Let $\boldsymbol{N}$ denote the set of all positive integers. For $\lambda>0$ and a positive integer Δ, we define

$$\mathcal{K}^{\lambda}_{k,\Delta}:=\{i\in\boldsymbol{N}\mid\|s_{i-1}\|>\lambda,\ k\le i\le k+\Delta-1\}.$$

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 (3.10)–(3.11) hold, then a certain fraction of the steps cannot be too small. This lemma corresponds to [1, Lemma 3] and [9, Lemma 4.2]. Since the proof is almost same as in that of [1, Lemma 3] and [9, Lemma 4.2], we omit it.

Lemma 3.5. *Suppose that all assumptions of Lemma* 3.3 *hold. Then there exists $\lambda>0$ such that, for any $\Delta\in\boldsymbol{N}$ and any index k_0, there is an index $\hat{k}\ge k_0$ such that*

$$|\mathcal{K}^{\lambda}_{\hat{k},\Delta}|>\frac{\Delta}{2}.$$

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 [7, Theorem 3.6], we omit it.

Theorem 3.6. *Suppose that Assumption* 1 *is satisfied. Consider the method* (1.1) *and* (3.3) *with* (3.5)–(3.6). *Assume that the method has Property 1 and the step size α_k satisfies the Wolfe conditions* (2.4)–(2.5). *Then the method converges in the sense that*

$$\liminf_{k\to\infty}\|g_k\|=0 \tag{3.16}$$

holds.

Closing this section, we briefly consider a suitable choice of parameters. Nakayama et al. [16] showed that parameters

$$\theta_{k-1} = 1 + \frac{|g_k^T d_{k-1}|}{\|g_{k-1}\|\|d_{k-1}\|} \quad \text{and} \quad \theta_{k-1} = 1 + \left| \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \right|$$

were better choices than $\theta_{k-1} = 1$ in their numerical experiments. These results mean that the preconvex class is efficient, and hence we consider the following parameter:

$$\theta_{k-1} = 1 + t_k |g_k^T d_{k-1}|, \tag{3.17}$$

where t_k is a scalar parameter. Then, if we use the exact line search, the parameter θ_{k-1} becomes 1, and hence the search direction (3.2) with (3.17) is identical to the search direction of the CG method with HS formula (1.6). Therefore, we regard the proposed method with (3.17) as a hybrid method of the three-term CG method and the memoryless quasi-Newton method.

Now we give concrete choices of θ_{k-1} and ν_{k-1} that satisfy conditions (3.5)–(3.6). Let $\bar{c}_1 = 1.99$. If $1 < \theta_{k-1} \leq 1.2$, then

$$\frac{\bar{c}_1}{\sqrt{\theta_{k-1}}} - 1 \geq \frac{1.99}{\sqrt{1.2}} - 1 \approx 0.8166$$

holds. Using the above inequality and (3.17), we propose the following parameters:

$$\theta_{k-1} = 1 + \min\left\{ \left| \frac{d_{k-1}^T g_k}{g_{k-1}^T d_{k-1}} \right|,\ 0.2 \right\} \quad \text{and} \quad \nu_{k-1} = 0.8. \tag{3.18}$$

Note that (3.18) satisfies conditions (3.5)–(3.6).

§4. Numerical experiments

In this section, we report numerical experiments of the proposed method of the form (1.1) and (3.3). We tested 138 problems from the CUTEr library [3, 10]. The problems were used in [16], and their names and dimensions n are given in Table 1. All codes were written in C by modifying the software package CG-DESCENT Version 5.3 [11, 12, 13]. 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. The line search procedure was the default procedure of CG-DESCENT, which implies that we used the parameter values of $\sigma = 0.9$ and $\delta = 0.1$ in the Wolfe conditions (2.4)–(2.5).

To compare numerical performance among the tested methods, we adopt the performance profiles based on the CPU time by Dolan and Moré [8]. For n_s solvers and n_p problems, the performance profiles $P : \boldsymbol{R} \to [0, 1]$ 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 $P(\tau) = \frac{1}{n_p}|\{p \in \mathcal{P}|r_{p,s} \leq \tau\}|$, for all $\tau \geq 1$. 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 of performance profiles gives the percentage of the test problems for which a method is the best result. The top curve is the method that solves the most problems in a result that is within a factor τ of the best result. In order to prevent a measurement error, we set the minimum of the 0.1 seconds.

Table 2 presents the choices of parameters θ_{k-1}, $\hat{\gamma}_{k-1}$ and ν_{k-1} for the tested methods. ML1–3, KD and NEW are the method of the form (1.1) and (3.3). Note that ML1–3 correspond to memoryless quasi-Newton methods by Nakayama et al. [16], and KD corresponds to the method by Kou and Dai [14]. In our numerical experiments, for ML1–3 and KD, we chose the parameters recommended in their numerical experiments. NEW is the proposed method with (3.18). For NEW, we chose the parameters which performed better in our preliminary numerical experiments ([14, 16]). CGD is the well-known benchmark method based on the CG method by Hager and Zhang [12, 13], namely CG_DESCENT 5.3.

Figure 1 shows the performance profiles of the methods in Table 2. As mentioned at the beginning of Section 3, we see that ML3 performed better than ML2 did. This implies that the scaling parameter (3.1) is more efficient than (2.20). Since ML3 and NEW performed better than ML1 and KD did, respectively, we see that the preconvex class is significant for the Broyden family. We see that KD and NEW were superior to CGD and performed slightly better than ML3 did. Note that ML3 does not guarantee the global convergence. On the other hand, KD and NEW guarantee the global convergence. Thus, the proposed method is theoretically superior to the memoryless quasi-Newton method by Nakayama et al. [16].

Table 1: Test problems (names and dimensions) by CUTEr library

name	n	name	n	name	n	name	n
AKIVA	2	DIXMAANC	3000	HEART8LS	8	PENALTY1	1000
ALLINITU	4	DIXMAAND	3000	HELIX	3	PENALTY2	200
ARGLINA	200	DIXMAANE	3000	HIELOW	3	PENALTY3	200
ARGLINB	200	DIXMAANF	3000	HILBERTA	2	POWELLSG	5000
ARWHEAD	5000	DIXMAANG	3000	HILBERTB	10	POWER	10000
BARD	3	DIXMAANH	3000	HIMMELBB	2	QUARTC	5000
BDQRTIC	5000	DIXMAANI	3000	HIMMELBF	4	ROSENBR	2
BEALE	2	DIXMAANJ	3000	HIMMELBG	2	S308	2
BIGGS6	6	DIXMAANK	3000	HIMMELBH	2	SCHMVETT	5000
BOX3	3	DIXMAANL	3000	HUMPS	2	SENSORS	100
BOX	10000	DIXON3DQ	10000	JENSMP	2	SINEVAL	2
BRKMCC	2	DJTL	2	KOWOSB	4	SINQUAD	5000
BROWNAL	200	DQDRTIC	5000	LIARWHD	5000	SISSER	2
BROWNBS	2	DQRTIC	5000	LOGHAIRY	2	SNAIL	2
BROWNDEN	4	EDENSCH	2000	MANCINO	100	SPARSINE	5000
BROYDN7D	5000	EG2	1000	MARATOSB	2	SPARSQUR	10000
BRYBND	5000	ENGVAL1	5000	MEXHAT	2	SPMSRTLS	4999
CHAINWOO	4000	ENGVAL2	3	MOREBV	5000	SROSENBR	5000
CHNROSNB	50	ERRINROS	50	MSQRTALS	1024	STRATEC	10
CLIFF	2	EXPFIT	2	MSQRTBLS	1024	TESTQUAD	5000
COSINE	10000	EXTROSNB	1000	NONCVXU2	5000	TOINTGOR	50
CRAGGLVY	5000	FLETCBV2	5000	NONDIA	5000	TOINTGSS	5000
CUBE	2	FLETCHCR	1000	NONDQUAR	5000	TOINTPSP	50
CURLY10	10000	FMINSRF2	5625	OSBORNEA	5	TOINTQOR	50
CURLY20	10000	FMINSURF	5625	OSBORNEB	11	TQUARTIC	5000
CURLY30	10000	FREUROTH	5000	OSCIPATH	10	TRIDIA	5000
DECONVU	63	GENHUMPS	5000	PALMER1C	8	VARDIM	200
DENSCHNA	2	GENROSE	500	PALMER1D	7	VAREIGVL	50
DENSCHNB	2	GROWTHLS	3	PALMER2C	8	VIBRBEAM	8
DENSCHNC	2	GULF	3	PALMER3C	8	WATSON	12
DENSCHND	3	HAIRY	2	PALMER4C	8	WOODS	4000
DENSCHNE	3	HATFLDD	3	PALMER5C	6	YFITU	3
DENSCHNF	2	HATFLDE	3	PALMER6C	8	ZANGWIL2	2
DIXMAANA	3000	HATFLDFL	3	PALMER7C	8		
DIXMAANB	3000	HEART6LS	6	PALMER8C	8		

Table 2: Tested methods

Method name	θ_{k-1}	$\hat{\gamma}_{k-1}$	ν_{k-1}
ML1(BFGS)	1	(3.1)	1
ML2	$1+\min\left\{\dfrac{\lvert g_k^T d_{k-1}\rvert}{\lVert g_k\rVert\lVert d_{k-1}\rVert},\ 0.9\right\}$	(2.20)	1
ML3	$1+\min\left\{\dfrac{\lvert g_k^T d_{k-1}\rvert}{\lVert g_k\rVert\lVert d_{k-1}\rVert},\ 0.9\right\}$	(3.1)	1
KD	1	(3.9)	0.8
NEW	$1+\min\left\{\left\lvert\dfrac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}\right\rvert,\ 0.2\right\}$	(3.9)	0.8
CGD	CG_DESCENT Ver5.3		

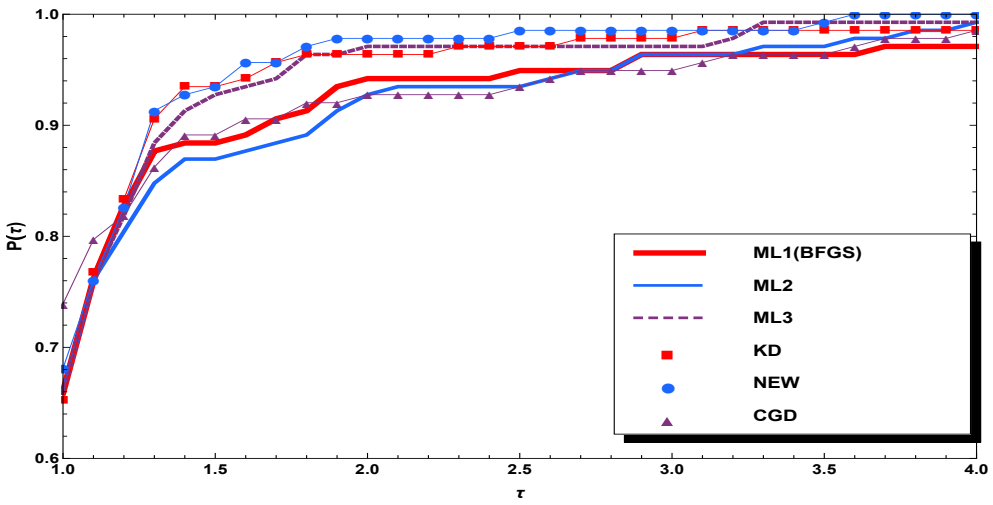

Figure 1: performance profiles based on CPU time

§5. Conclusions

In this paper, we have modified the memoryless quasi-Newton method based on the spectral-scaling secant condition [16] and proposed a new method. Furthermore, we have shown that the method always satisfies the sufficient descent condition and converges globally. In numerical experiments, we have shown that the proposed method performs better than existing memoryless quasi-Newton methods do, and we have found suitable parameters for the proposed method. A further study is to find more suitable choices for parameters θ_{k-1}, $\hat{\gamma}_{k-1}$ and ν_{k-1}.

Acknowledgment

I would like to thank Professor Hiroshi Yabe and Professor Yasushi Narushima for useful discussions and proofreading the manuscript. The author is grateful to the anonymous referee 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] N. Andrei, Accelerated adaptive Perry conjugate gradient algorithms based on the self-scaling memoryless BFGS update, *Journal of Computational and Applied Mathematics*, **325** (2017), 149–164.

[3] 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.

[4] 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.

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

[6] 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.

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

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

[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 : http://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, Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent, *ACM Transactions on Mathematical Software*, **32** (2006), 113–137.

[14] 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.

[15] 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.

[16] S. Nakayama, Y. Narushima and H. Yabe, Memoryless quasi-Newton methods based on spectral-scaling Broyden family for unconstrained optimization, *Journal of Industrial and Management Optimization*, to appear.

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

[18] 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.

[19] J. Nocedal and S.J. Wright, "Numerical optimization", 2^{nd} edition, Springer, 2006.

[20] S.S. Oren, Self-scaling variable metric (SSVM) algorithms, Part II: Implementation and experiments, *Management Science*, **20** (1974), 863–874.

[21] 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.

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

[23] 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.

[24] S. Yao and L. Ning, An adaptive three-term conjugate gradient method based on self-scaling memoryless BFGS matrix, *Journal of Computational and Applied Mathematics*, **332** (2018), 72–85.

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

[26] L. Zhang, W. Zhou and D.H. Li, Some descent three-term conjugate gradient methods and their global convergence, *Optimization Methods and Software*, **26** (2007), 697–711.

[27] 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.

Shummin Nakayama
Department of Applied Mathematics, Tokyo University of Science
1-3, Kagurazaka, Shinjuku, Tokyo 162-8601, Japan
E-mail: 1416702@ed.tus.ac.jp