Journal of the Operations Research Society of Japan
Vol. 61, No. 1, January 2018, pp. 53–70


# A MEMORYLESS SYMMETRIC RANK-ONE METHOD WITH SUFFICIENT DESCENT PROPERTY FOR UNCONSTRAINED OPTIMIZATION

Shummin Nakayama
*Tokyo University of Science*

Yasushi Narushima
*Yokohama National University*

Hiroshi Yabe
*Tokyo University of Science*

(Received July 30, 2016; Revised October 11, 2017)

*Abstract* Quasi-Newton methods are widely used for solving unconstrained optimization problems. However, it is difficult to apply quasi-Newton methods directly to large-scale unconstrained optimization problems, because they need the storage of memories for matrices. In order to overcome this difficulty, memoryless quasi-Newton methods were proposed. Shanno (1978) derived the memoryless BFGS method. Recently, several researchers studied the memoryless quasi-Newton method based on the symmetric rank-one formula. However existing memoryless symmetric rank-one methods do not necessarily satisfy the sufficient descent condition. In this paper, we focus on the symmetric rank-one formula based on the spectral scaling secant condition and derive a memoryless quasi-Newton method based on the formula. Moreover we show that the method always satisfies the sufficient descent condition and converges globally for general objective functions. Finally, preliminary numerical results are shown.

**Keywords**: Nonlinear programming, unconstrained optimization, memoryless quasi-Newton method, symmetric rank-one formula, sufficient descent condition

## 1. Introduction

In this paper, we consider the following unconstrained optimization problem:

$$\min \ f(x), \tag{1.1}$$

where $f : \boldsymbol{R}^n \to \boldsymbol{R}$ is a sufficiently smooth function. We denote its gradient $\nabla f$ by $g$. Usually, iterative methods are used for solving problem (1.1), and they are of the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

where $x_k \in \boldsymbol{R}^n$ is the $k$th approximation to a solution, $\alpha_k > 0$ is a step size, and $d_k$ is a search direction.

Quasi-Newton methods are known as effective methods for solving problem (1.1). The search direction of quasi-Newton methods is given by

$$d_k = -H_k g_k, \tag{1.3}$$

where $H_k$ is an approximation to the inverse Hessian $\nabla^2 f(x_k)^{-1}$. The matrix $H_k$ is chosen so that the secant condition

$$H_k y_{k-1} = s_{k-1} \tag{1.4}$$

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 Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula is well-known as an effective updating formula and is given by

$$H_k = H_{k-1} - \frac{H_{k-1}y_{k-1}s_{k-1}^T + s_{k-1}y_{k-1}^T H_{k-1}}{s_{k-1}^T y_{k-1}} + \left(1 + \frac{y_{k-1}^T H_{k-1} y_{k-1}}{s_{k-1}^T y_{k-1}}\right) \frac{s_{k-1}s_{k-1}^T}{s_{k-1}^T y_{k-1}}. \quad (1.5)$$

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, numerical methods which do not need to store any matrix have been studied. As such methods, the limited memory BFGS (L-BFGS) method [19, 25], nonlinear conjugate gradient methods (see [9, 14, 16, 24], for example) and memoryless quasi-Newton methods [30] are well-known. In this paper, we focus on memoryless quasi-Newton methods.

Shanno [30] proposed the memoryless quasi-Newton methods based on the BFGS formula, namely, the method (1.3) and (1.5) with $H_{k-1} = I$, where $I$ denotes the identity matrix. Then, the search direction is given by

$$d_k = -g_k + \left(\frac{y_{k-1}^T g_k}{s_{k-1}^T y_{k-1}} - \left(1 + \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}\right) \frac{g_k^T s_{k-1}}{s_{k-1}^T y_{k-1}}\right) s_{k-1} + \frac{g_k^T s_{k-1}}{s_{k-1}^T y_{k-1}} y_{k-1}. \quad (1.6)$$

We call a memoryless quasi-Newton method with the BFGS formula the memoryless BFGS method, and we use the same manner for the other methods. Recently, several researchers dealt with the memoryless BFGS method [17, 18]. If the exact line search is used, then the search direction (1.6) becomes

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

because the exact line search implies $g_k^T s_{k-1} = 0$. This direction is a search direction of the nonlinear conjugate gradient method with the HS formula [16]. Since the memoryless BFGS method deeply relates to the conjugate gradient method, the method has been paid attention to. Several three-term conjugate gradient methods were proposed based on the memoryless BFGS method or its variants [1, 23, 33].

The symmetric rank-one (SR1) formula is also known as an effective updating formula and is given by

$$H_k = H_{k-1} + \frac{(s_{k-1} - H_{k-1}y_{k-1})(s_{k-1} - H_{k-1}y_{k-1})^T}{(s_{k-1} - H_{k-1}y_{k-1})^T y_{k-1}}. \quad (1.7)$$

In comparison with the BFGS formula, this formula has interesting properties as follows (see [5, 26, 29]):

- The formula has a self-dual relation.
- For the strictly convex quadratic objective function, the BFGS method with the exact line search terminates at $n$ steps and $H_n$ is the inverse Hessian. On the other hand, the SR1 method without a line search terminates at $n+1$ steps and $H_n$ is the inverse Hessian for the same objective function.
- For the SR1 method, $H_k$ approaches the inverse Hessian at an optimal solution under certain assumptions.

Therefore, many researchers have studied the SR1 method (see [3, 5, 22], for example). However, the SR1 formula does not necessarily retain a positive definiteness, and hence the search direction may not satisfy the descent condition (namely, $g_k^T d_k = -g_k^T H_k g_k < 0$). In order to overcome this difficulty, the sized SR1 formula

$$H_k = \theta_{k-1} H_{k-1} + \frac{(s_{k-1} - \theta_{k-1} H_{k-1} y_{k-1})(s_{k-1} - \theta_{k-1} H_{k-1} y_{k-1})^T}{(s_{k-1} - \theta_{k-1} H_{k-1} y_{k-1})^T y_{k-1}} \tag{1.8}$$

was considered (see [27, 31]). Under the assumptions that $H_{k-1}$ is positive definite, $\theta_{k-1} > 0$ and $s_{k-1}^T y_{k-1} > 0$, the matrix $H_k$ is positive definite if and only if the parameter $\theta_{k-1}$ satisfies

$$\theta_{k-1} \notin \left[ \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}}, \frac{s_{k-1}^T H_{k-1}^{-1} s_{k-1}}{s_{k-1}^T y_{k-1}} \right]. \tag{1.9}$$

As a choice of $\theta_{k-1}$, some researchers chose the following parameters [27]:

$$\theta_{k-1} = \frac{s_{k-1}^T H_{k-1}^{-1} s_{k-1}}{s_{k-1}^T y_{k-1}} + \sqrt{\left( \frac{s_{k-1}^T H_{k-1}^{-1} s_{k-1}}{s_{k-1}^T y_{k-1}} \right)^2 - \frac{s_{k-1}^T H_{k-1}^{-1} s_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}}} \tag{1.10}$$

and

$$\theta_{k-1} = \frac{s_{k-1}^T H_{k-1}^{-1} s_{k-1}}{s_{k-1}^T y_{k-1}} - \sqrt{\left( \frac{s_{k-1}^T H_{k-1}^{-1} s_{k-1}}{s_{k-1}^T y_{k-1}} \right)^2 - \frac{s_{k-1}^T H_{k-1}^{-1} s_{k-1}}{y_{k-1}^T H_{k-1} y_{k-1}}}. \tag{1.11}$$

These parameters are solutions of the following problem

$$\min \ \kappa \left( H_{k-1}^{-\frac{1}{2}} H_k H_{k-1}^{-\frac{1}{2}} \right),$$

where $\kappa(A)$ denotes the condition number of a matrix $A$. Parameters (1.10)–(1.11) satisfy condition (1.9) if $s_{k-1}$ and $H_{k-1} y_{k-1}$ are linearly independent.

This paper considers a memoryless SR1 method. Several researchers studied memoryless SR1 methods. For example, Moyi and Leong [21] gave the following search direction

$$d_k^{ML} = -\theta_{k-1} g_k - \frac{(s_{k-1} - \theta_{k-1} y_{k-1})^T g_k}{(s_{k-1} - \theta_{k-1} y_{k-1})^T y_{k-1}} (s_{k-1} - \theta_{k-1} y_{k-1}). \tag{1.12}$$

This corresponds to the memoryless quasi-Newton method based on (1.8). They chose the parameter $\theta_{k-1}$ by

$$\theta_{k-1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} - \sqrt{\left( \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \right)^2 - \frac{s_{k-1}^T s_{k-1}}{y_{k-1}^T y_{k-1}}}, \tag{1.13}$$

which is (1.11) with $H_{k-1} = I$. For uniformly convex objective functions, they showed that the method satisfies the sufficient descent condition and converges globally. Here, the sufficient descent condition means that there exists a positive constant $c$ such that

$$g_k^T d_k \leq -c \|g_k\|^2 \quad \text{for all } k, \tag{1.14}$$

where $\| \cdot \|$ is the $\ell_2$ norm. Modarres et al. [20] proposed the memoryless quasi-Newton method with a variant of the sized SR1 formula (1.8) based on the modified secant condition [32]. Their method satisfies the descent condition. In addition, they established the global convergence of the method for uniformly convex objective functions. However, the global convergence of the above methods were not shown for general objective functions.

To our knowledge, memoryless SR1 (or sized SR1) methods which satisfy the sufficient descent condition and converge globally for general objective functions have not been studied. The sufficient descent condition plays an important role in establishing the global convergence of the general iterative methods for general objective functions, and hence we consider a memoryless SR1 method which always satisfies the sufficient descent condition (1.14). For this purpose, we introduce the spectral scaling secant condition. This condition was proposed by Cheng and Li [4] in order to improve the performance of the quasi-Newton method based on the secant condition (1.4). In this paper, we derive an SR1 formula by using the spectral scaling secant condition and propose a new memoryless quasi-Newton method based on this formula. Furthermore, we show that the method always satisfies the sufficient descent condition, and we prove its global convergence property for general objective functions.

This paper is organized as follows. In Section 2, we propose a new memoryless SR1 method which always generates the sufficient descent direction. In Section 3, we prove the global convergence properties of our method for uniformly convex objective functions and general objective functions, respectively. Finally, some numerical results are presented in Section 4.

## 2. Our Method

We recall the spectral scaling secant condition proposed by Cheng and Li [4]. They scaled the objective function for numerical stability and considered the following approximate relation

$$\gamma_{k-1} f(x) \approx \gamma_{k-1} \left( f(x_k) + g_k^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k) \right), \tag{2.1}$$

where $\gamma_{k-1}$ is a scaling parameter. Differentiating (2.1) and substituting $x_{k-1}$ into $x$, we get

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

which yields the spectral scaling secant condition:

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

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

$$H_k y_{k-1} = \frac{1}{\gamma_{k-1}} s_{k-1}. \tag{2.2}$$

Cheng and Li [4] claimed that the spectral scaling secant condition has a preconditioned property. In fact, they showed that the BFGS method based on the spectral scaling secant condition performed better than the method based on the standard secant condition did in numerical experiments.

In this section, we present a new memoryless SR1 method based on the spectral scaling secant condition (2.2). The SR1 formula based on (2.2) is given by

$$H_k = H_{k-1} + \frac{\left( \frac{1}{\gamma_{k-1}} s_{k-1} - H_{k-1} y_{k-1} \right) \left( \frac{1}{\gamma_{k-1}} s_{k-1} - H_{k-1} y_{k-1} \right)^T}{\left( \frac{1}{\gamma_{k-1}} s_{k-1} - H_{k-1} y_{k-1} \right)^T y_{k-1}}. \tag{2.3}$$

We call the formula (2.3) the spectral scaling SR1 (SS-SR1) formula.

Now, considering a memoryless quasi-Newton method based on (2.3), and putting

$$p_{k-1} = s_{k-1} - \gamma_{k-1} y_{k-1}, \tag{2.4}$$

we obtain

$$d_k = -g_k - \frac{p_{k-1}^T g_k}{\gamma_{k-1} p_{k-1}^T y_{k-1}} p_{k-1}. \tag{2.5}$$

Then we have the following theorem.

**Theorem 2.1.** *Under the assumptions that* $\gamma_{k-1} > 0$ *and* $s_{k-1}^T y_{k-1} > 0$*, the search direction* (2.5) *satisfies the descent condition if and only if the parameter* $\gamma_{k-1}$ *satisfies*

$$\gamma_{k-1} \notin \left[ \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}, \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \right]. \tag{2.6}$$

*Moreover, if the parameter* $\gamma_{k-1}$ *satisfies*

$$0 < \gamma_{k-1} < \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}, \tag{2.7}$$

*then the search direction satisfies the sufficient descent condition* (1.14) *with* $c = 1$*, namely,*

$$g_k^T d_k \le -\|g_k\|^2 \tag{2.8}$$

*holds.*

*Proof.* By taking into account the relation between (1.8) and (2.3) with $H_{k-1} = I$, the condition (1.9) corresponds to (2.6), and hence the matrix $H_k$ updated by (2.3) with $H_{k-1} = I$ is positive definite if and only if the parameter $\gamma_{k-1}$ satisfies (2.6). Therefore, the search direction (2.5) is the descent direction if and only if the parameter $\gamma_{k-1}$ satisfies (2.6).

We next consider the case that $\gamma_{k-1}$ satisfies (2.7), which guarantees the following inequality from (2.4)

$$p_{k-1}^T y_{k-1} = (s_{k-1} - \gamma_{k-1} y_{k-1})^T y_{k-1} > 0. \tag{2.9}$$

Therefore, we have

$$g_k^T d_k = -\|g_k\|^2 - \frac{\left(p_{k-1}^T g_k\right)^2}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \le -\|g_k\|^2.$$

□

Theorem 2.1 guarantees that the search direction (2.5) with (2.7) satisfies the sufficient descent condition. Therefore, throughout this paper, we adapt the condition (2.7) for $\gamma_{k-1}$. Furthermore, to establish the global convergence of the method, we consider a restart strategy. To guarantee the well-definedness of the updating formula, the SR1 formula (1.7) is usually used only if

$$|(s_{k-1} - H_{k-1} y_{k-1})^T y_{k-1}| \ge \mu \|s_{k-1} - H_{k-1} y_{k-1}\| \|y_{k-1}\|, \quad \mu \in (0,1) \tag{2.10}$$

holds. If (2.10) does not hold, then the next matrix defined as $H_k = H_{k-1}$ (see [5]). Following this scheme, when

$$p_{k-1}^T y_{k-1} \ge \mu \|p_{k-1}\| \|y_{k-1}\|, \quad \mu \in (0,1) \tag{2.11}$$

does not hold, we set $H_k = H_{k-1} = I$, because $H_{k-1}$ is set to be the identity matrix in the memoryless quasi-Newton method. In this case, we have the steepest descent direction $d_k = -H_k g_k = -g_k$.

Summarizing the above arguments, we present the search direction of the memoryless SS-SR1 method as follows:

$$d_k = \begin{cases} -g_k & \text{if } k=0 \text{ or } p_{k-1}^T y_{k-1} < \mu \|p_{k-1}\| \|y_{k-1}\|, \\ -g_k + \beta_k^N p_{k-1} & \text{otherwise,} \end{cases} \tag{2.12}$$

where

$$\beta_k^N = \frac{-p_{k-1}^T g_k}{\gamma_{k-1} p_{k-1}^T y_{k-1}}, \tag{2.13}$$

and $\gamma_{k-1}$ is a parameter satisfying (2.7). Note that $p_{k-1}^T y_{k-1} > 0$ holds by (2.9). Since $d_k = -g_k$ implies $g_k^T d_k = -\|g_k\|^2$, we obtain by (2.8) and (2.12) that

$$g_k^T d_k \leq -\|g_k\|^2 \quad \text{for all } k \geq 0. \tag{2.14}$$

Therefore, our method always satisfies the sufficient descent condition (1.14) with $c = 1$. We note that the method (2.12) can be regarded as a three-term conjugate gradient like method.

In the line search, we require a step size $\alpha_k$ to satisfy the Wolfe conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k, \tag{2.15}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{2.16}$$

where $0 < \delta < \sigma < 1$. Since $d_k$ satisfies the sufficient descent condition (2.14) and the curvature condition (2.16) in the Wolfe conditions, we have

$$d_k^T y_k \geq (1-\sigma)\|g_k\|^2. \tag{2.17}$$

Therefore, $s_k^T y_k > 0$ always holds under the Wolfe conditions.

## 3. Global Convergence of Our Method

In this section, we show the global convergence property of our method for uniformly convex objective functions in Section 3.1 and general objective functions in Section 3.2, respectively. Throughout this section, we assume that $g_k \neq 0$ for all $k$, otherwise a stationary point has been found.

In order to establish the global convergence, we restrict the interval (2.7) to

$$\gamma_{k-1} \in \left( \rho \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}, \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} \right), \tag{3.1}$$

where $\rho \in (0,1)$ is arbitrarily chosen, and make the following standard assumptions for the objective function.

**Assumption 3.1.** *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 $\nu$ such that*

$$\|x\| \leq \nu \quad \textit{for all } x \in \mathcal{L}. \tag{3.2}$$

**Assumption 3.2.** *The objective function $f$ is continuously differentiable on an open convex neighborhood $\mathcal{N}$ of $\mathcal{L}$, and its gradient $g$ is Lipschitz continuous in $\mathcal{N}$, namely, there exists a positive constant $L$ such that*

$$\|g(u) - g(v)\| \le L\|u - v\| \quad \text{for all } u, v \in \mathcal{N}. \tag{3.3}$$

The above assumptions imply that there exists a positive constant $\hat{\nu}$ such that

$$\|g(x)\| \le \hat{\nu} \quad \text{for all } x \in \mathcal{L}. \tag{3.4}$$

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

**Lemma 3.1.** *Suppose that Assumptions 3.1–3.2 are satisfied. Consider any method in the form (1.2), where $d_k$ and $\alpha_k$ satisfy the sufficient descent condition (1.14) and the Wolfe conditions (2.15) and (2.16), respectively. If*

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty,$$

*then the method converges globally in the sense that*

$$\liminf_{k\to\infty} \|g_k\| = 0 \tag{3.5}$$

*holds.*

### 3.1. Global convergence for uniformly convex objective functions

In this subsection, we prove the global convergence of our method for uniformly convex objective functions. Since $f$ is a uniformly convex function, there exists a positive constant $m$ such that

$$(g(u) - g(v))^T(u - v) \ge m\|u - v\|^2 \quad \text{for all } u, v \in \boldsymbol{R}^n. \tag{3.6}$$

Then we obtain the following convergence theorem.

**Theorem 3.1.** *Suppose that the objective function $f$ is a uniformly convex function and that Assumption 3.2 holds. Consider the method (1.2) and (2.12) with (3.1). Assume that $\alpha_k$ satisfies the Wolfe conditions (2.15) and (2.16). Then*

$$\lim_{k\to\infty} \|g_k\| = 0 \tag{3.7}$$

*holds. Therefore, the generated sequence $\{x_k\}$ converges to the global minimizer.*

*Proof.* By (3.6), we get

$$s_{k-1}^T y_{k-1} \ge m\|s_{k-1}\|^2. \tag{3.8}$$

It follows from (2.11), (2.13), (3.1), (3.3) and (3.8) that

$$|\beta_k^N| = \left|\frac{p_{k-1}^T g_k}{\gamma_{k-1} p_{k-1}^T y_{k-1}}\right| < \frac{y_{k-1}^T y_{k-1}}{\rho s_{k-1}^T y_{k-1}} \frac{\|p_{k-1}\|\|g_k\|}{\mu\|p_{k-1}\|\|y_{k-1}\|} = \frac{\|y_{k-1}\|\|g_k\|}{\rho\mu s_{k-1}^T y_{k-1}} \le \frac{L\|g_k\|}{\rho\mu m\|s_{k-1}\|} \tag{3.9}$$

and

$$\gamma_{k-1} < \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} \le \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \le \frac{1}{m}. \tag{3.10}$$

We now consider the search direction (2.12). We first note that (2.14) is always satisfied. Next we estimate the norm of the search direction (2.12). If $d_k = -g_k$, then inequality (3.4) yields

$$\|d_k\| = \|g_k\| \leq \hat{\nu}. \tag{3.11}$$

Otherwise, using (2.4), (3.3), (3.4), (3.9) and (3.10), we obtain

$$\begin{aligned}
\|d_k\| &= \| -g_k + \beta_k^N p_{k-1}\| \\
&\leq \|g_k\| + |\beta_k^N| \|p_{k-1}\| \\
&\leq \|g_k\| + \frac{L\|g_k\|}{\rho\mu m \|s_{k-1}\|}\|p_{k-1}\| \\
&= \|g_k\| + \frac{L\|g_k\|}{\rho\mu m \|s_{k-1}\|}\|s_{k-1} - \gamma_{k-1} y_{k-1}\| \\
&\leq \|g_k\| + \frac{L\|g_k\|}{\rho\mu m \|s_{k-1}\|}(1 + L\gamma_{k-1})\|s_{k-1}\| \\
&\leq \|g_k\| + \frac{L\|g_k\|}{\rho\mu m \|s_{k-1}\|}\left(1 + \frac{L}{m}\right)\|s_{k-1}\| \\
&= \left(1 + \frac{L}{\rho\mu m}\left(1 + \frac{L}{m}\right)\right)\|g_k\| \\
&\leq \left(1 + \frac{L}{\rho\mu m}\left(1 + \frac{L}{m}\right)\right)\hat{\nu}.
\end{aligned} \tag{3.12}$$

Since the search direction is bounded, we get

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty.$$

By Lemma 3.1, we have (3.5). Since the objective function $f$ is uniformly convex, (3.5) yields (3.7), and the generated sequence $\{x_k\}$ converges to the global minimizer. □

Since the search direction is bounded by (3.11)–(3.12) and satisfies the sufficient descent condition (2.14), we can prove the $R$-linear convergence of the memoryless SS-SR1 method for uniformly convex objective functions under the assumption that the step size is bounded in a similar way to the proof of [6, Theorem 3.1].

### 3.2. Global convergence for general objective functions

In this subsection, we prove the global convergence of our method for general objective functions. To establish the global convergence, we modify $\beta_k^N$ as follows:

$$\beta_k^{N+} = \max\left\{0, \beta_k^N\right\}. \tag{3.13}$$

If condition $p_{k-1}^T y_{k-1} < \mu\|p_{k-1}\|\|y_{k-1}\|$ holds infinitely many times, the search direction (2.12) becomes the steepest descent direction infinitely many times, which guarantees (3.5) (see [26, Section 3.2]). Accordingly, we hereafter assume without loss of generality that $p_{k-1}^T y_{k-1} \geq \mu\|p_{k-1}\|\|y_{k-1}\|$ holds for all $k \geq 1$. Then we have $d_k = -g_k + \beta_k^{N+} p_{k-1}$.

We now estimate the norm of the search direction of our method. From (2.12), (2.13) and (2.4), we have the following relations

$$
\begin{aligned}
\|d_k\|^2 &= \left\| -g_k - \frac{p_{k-1}^T g_k}{\gamma_{k-1} p_{k-1}^T y_{k-1}} p_{k-1} \right\|^2 \\
&= \|g_k\|^2 + \frac{2\left(p_{k-1}^T g_k\right)^2}{\gamma_{k-1} p_{k-1}^T y_{k-1}} + \left( \frac{p_{k-1}^T g_k}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|p_{k-1}\|^2 \\
&\le \|g_k\|^2 + \left( \frac{2\|g_k\|^2}{\gamma_{k-1} p_{k-1}^T y_{k-1}} + \left( \frac{\|p_{k-1}\|\|g_k\|}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \right) \|p_{k-1}\|^2 \\
&\le \|g_k\|^2 + \left( \frac{\|g_k\|^2}{\|p_{k-1}\|^2} + \frac{2\|g_k\|^2}{\gamma_{k-1} p_{k-1}^T y_{k-1}} + \left( \frac{\|p_{k-1}\|\|g_k\|}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \right) \|p_{k-1}\|^2 \\
&= \|g_k\|^2 + \left( \frac{\|g_k\|}{\|p_{k-1}\|} + \frac{\|p_{k-1}\|\|g_k\|}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|p_{k-1}\|^2 \\
&= \|g_k\|^2 + \left( \frac{\gamma_{k-1} p_{k-1}^T y_{k-1} + \|p_{k-1}\|^2}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|g_k\|^2 \\
&= \|g_k\|^2 + \left( \frac{p_{k-1}^T (\gamma_{k-1} y_{k-1} + p_{k-1})}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|g_k\|^2 \\
&= \|g_k\|^2 + \left( \frac{p_{k-1}^T (\gamma_{k-1} y_{k-1} + s_{k-1} - \gamma_{k-1} y_{k-1})}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|g_k\|^2 \\
&= \|g_k\|^2 + \left( \frac{p_{k-1}^T s_{k-1}}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|g_k\|^2 \\
&\le \|g_k\|^2 + \left( \frac{\|s_{k-1}\|\|p_{k-1}\|}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|g_k\|^2 \\
&\le \|g_k\|^2 + \left( \frac{\alpha_{k-1}\|p_{k-1}\|\|g_k\|}{\gamma_{k-1} p_{k-1}^T y_{k-1}} \right)^2 \|d_{k-1}\|^2. \qquad (3.14)
\end{aligned}
$$

Therefore, by defining

$$\Psi_k = \frac{\alpha_{k-1}\|p_{k-1}\|\|g_k\|}{\gamma_{k-1} p_{k-1}^T y_{k-1}}, \qquad (3.15)$$

the relation (3.14) yields

$$\|d_k\|^2 \le \|g_k\|^2 + \Psi_k^2 \|d_{k-1}\|^2 \quad \text{for all } k \ge 1.$$

It follows from (2.7), (2.9), (3.15) and $\alpha_{k-1} > 0$ that

$$\Psi_k > 0 \quad \text{for all } k \ge 1.$$

The following lemma implies that $\Psi_k$ will be small when the step $s_{k-1}$ is too small. This lemma corresponds to *Property* $(*)$ of conjugate gradient methods originally given by Gilbert and Nocedal [10].

**Lemma 3.2.** *Suppose that Assumptions* 3.1–3.2 *are satisfied. Consider the method* (1.2) *and* (2.12) *with* (3.1) *and* (3.13), *where* $\alpha_k$ *satisfies the Wolfe conditions* (2.15) *and* (2.16). *Assume that there exits a positive constant* $\varepsilon$ *such that*

$$\varepsilon \le \|g_k\| \quad \textit{for all } k. \qquad (3.16)$$

*Then, there exist constants $b > 1$ and $\xi > 0$ such that*

$$\Psi_k \le b \tag{3.17}$$

*and*

$$\|s_{k-1}\| \le \xi \Longrightarrow \Psi_k \le \frac{1}{b}. \tag{3.18}$$

*Proof.* It follows from (2.11), (2.17), (3.1), (3.2), (3.3), (3.15) and (3.16) that

$$\begin{aligned}
\Psi_k &= \frac{\alpha_{k-1}\|p_{k-1}\|\|g_k\|}{\gamma_{k-1}p_{k-1}^T y_{k-1}} \\
&< \frac{y_{k-1}^T y_{k-1}}{\rho s_{k-1}^T y_{k-1}} \frac{\alpha_{k-1}\|p_{k-1}\|\|g_k\|}{\mu\|p_{k-1}\|\|y_{k-1}\|} \\
&\le \frac{\alpha_{k-1}\|y_{k-1}\|\|g_k\|}{\rho\mu s_{k-1}^T y_{k-1}} \\
&\le \frac{\alpha_{k-1}L\|s_{k-1}\|\|g_k\|}{\rho\mu\alpha_{k-1}(1-\sigma)\|g_k\|^2} \\
&\le \frac{L\|s_{k-1}\|}{\rho\mu(1-\sigma)\|g_k\|} \\
&\le \frac{2L\nu}{\rho\mu(1-\sigma)\varepsilon} \\
&:= \bar{b}.
\end{aligned}$$

We define $b = 1 + \bar{b}$ and

$$\xi = \frac{\rho\mu(1-\sigma)\varepsilon}{Lb}.$$

If $\|s_{k-1}\| \le \xi$, then we obtain

$$\Psi_k \le \frac{L\|s_{k-1}\|}{\rho\mu(1-\sigma)\varepsilon} \le \frac{1}{b}.$$

Therefore, the proof is complete. □

The following lemma corresponds to [7, Lemma 3.4] and [23, Lemma 2.3].

**Lemma 3.3.** *Suppose that all assumptions of Lemma 3.2 are satisfied. Then, $d_k \ne 0$ and*

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

*hold, where $u_k = d_k/\|d_k\|$.*

*Proof.* Since $d_k \ne 0$ follows from (2.14) and $\varepsilon \le \|g_k\|$, the vector $u_k$ is well-defined. By defining

$$v_k = -\frac{1}{\|d_k\|}(g_k + \beta_k^{N+}\gamma_{k-1}y_{k-1}) \quad \text{and} \quad \eta_k = \beta_k^{N+}\frac{\|s_{k-1}\|}{\|d_k\|},$$

it follows from (2.12) that

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

This form and the identity $\|u_k\| = \|u_{k-1}\| = 1$ yield

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

Using this relation and $\beta_k^{N+} \geq 0$, we obtain

$$\begin{aligned} \|u_k - u_{k-1}\| &\leq (1+\eta_k)\|u_k - u_{k-1}\| \\ &= \|u_k - \eta_k u_{k-1} + \eta_k u_k - u_{k-1}\| \\ &\leq \|u_k - \eta_k u_{k-1}\| + \|\eta_k u_k - u_{k-1}\| \\ &= 2\|v_k\|. \end{aligned} \tag{3.19}$$

From (2.11), (2.13) and (3.13),

$$\beta_k^{N+} < \frac{\|p_{k-1}\|\|g_k\|}{\mu\gamma_{k-1}\|p_{k-1}\|\|y_{k-1}\|} \leq \frac{\|g_k\|}{\mu\gamma_{k-1}\|y_{k-1}\|} \tag{3.20}$$

holds. Using Lemma 3.1 and $\varepsilon \leq \|g_k\|$, we have

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} < \infty.$$

Therefore, (3.4), (3.19) and (3.20) yield

$$\begin{aligned} \sum_{k=0}^{\infty} \|u_k - u_{k-1}\|^2 &\leq 4\sum_{k=0}^{\infty} \|v_k\|^2 \\ &\leq 4\sum_{k=0}^{\infty} (\|g_k\| + \beta_k^{N+}\gamma_{k-1}\|y_{k-1}\|)^2 \frac{1}{\|d_k\|^2} \\ &\leq 4\left(\hat{\nu} + \frac{\hat{\nu}}{\mu}\right)\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} \\ &< \infty. \end{aligned}$$

Hence, the proof is complete. □

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

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

Let $|\mathcal{K}_{k,\Delta}^{\lambda}|$ denote the number of elements in $\mathcal{K}_{k,\Delta}^{\lambda}$. The following lemma shows that if the magnitude of the gradient is bounded away from zero and (3.17)–(3.18) hold, then a certain fraction of the steps cannot be too small. This lemma can be proved in the same way as [7, Lemma 3.5] and [10, Lemma 4.2]. Thus we omit the proof.

**Lemma 3.4.** *Suppose that all assumptions of Lemma* 3.2 *hold. Then there exists* $\lambda > 0$ *such that, for any* $\Delta \in \boldsymbol{N}$ *and any index* $k_0$*, there is an index* $\hat{k} \geq k_0$ *such that*

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

Now we obtain the global convergence result of our method. Since the proof of the theorem is exactly same as [7, Theorem 3.6], we omit it.

**Theorem 3.2.** *Suppose that Assumptions* 3.1–3.2 *are satisfied. Consider the method* (1.2) *and* (2.12) *with* (3.1) *and* (3.13)*. Assume that* $\alpha_k$ *satisfies the Wolfe conditions* (2.15) *and* (2.16)*. Then the method converges in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

As a concrete choice of $\gamma_{k-1}$ in (2.12), we choose the following parameter:

$$\gamma_{k-1} = \Gamma_k \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} \quad \text{with} \quad \Gamma_k \in (\Gamma_{\min}, \Gamma_{max}), \tag{3.21}$$

where $0 < \Gamma_{min} < \Gamma_{max} < 1$. Obviously, parameter (3.21) satisfies (3.1) with $\rho = \Gamma_{min}$. Therefore, Theorem 3.2 guarantees that the method (1.2) and (2.12) with (3.21) and (3.13) converges globally. Moreover, we obtain the following convergence result of our method. The proof of this corollary can be found in Appendix.

**Corollary 3.1.** *Suppose that Assumptions* 3.1–3.2 *are satisfied. Consider the method* (1.2) *and* (2.12) *with*

$$\gamma_{k-1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} - \sqrt{\left(\frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}\right)^2 - \frac{s_{k-1}^T s_{k-1}}{y_{k-1}^T y_{k-1}}} \tag{3.22}$$

*and* (3.13). *Assume that* $\alpha_k$ *satisfies the Wolfe conditions* (2.15) *and* (2.16). *Then the method converges in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

We note that the parameter (3.22) corresponds to (1.13).

## 4. Numerical Results

In this section, we report numerical results to compare the memoryless SS-SR1 method with other memoryless quasi-Newton methods. We investigate numerical performance of the tested methods on 134 problems from the CUTEr library [2, 11]. The names of test problems and their dimensions $n$ are given in Table 1. The problems were listed in Hager [12]. Although Hager [12] 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 and different local solutions were obtained when different solvers were applied to those omitted problems.

All codes were written in C by modifying the software package CG-DESCENT Version 5.3 [12, 13, 15]. They were run on a PC with 2.66GHz Intel Core i7, 8.0 GB RAM memory and Linux OS Ubuntu 14. The stopping rule was $\|g_k\|_\infty \le 10^{-6}$. We stopped the algorithm if CPU time exceeded 600 seconds or if a numerical overflow occurred. The line search procedure was the default procedure of CG-DESCENT. We used the parameter values of $\delta = 0.01$ and $\sigma = 0.1$ in the Wolfe conditions (2.15)–(2.16), and $\mu = 10^{-6}$ in the restart rule (2.11). In our experiments, the restart seldom occurred. Table 2 presents the methods used in our experiments. As the value of $\Gamma_k$ in (3.21) approached 1, the performance of the proposed method became poor. Therefore, we chose small values for $\Gamma_k$.

In order to compare numerical performance among the tested methods, we adopt the performance profiles of Dolan and Moré [8]. For $n_s$ solvers and $n_p = 134$ problems, the performance profiles $P : \boldsymbol{R} \to [0, 1]$ is defined as follows: Let $P$ and $S$ be the set of problems and the set of solvers, respectively. For each problem $p \in P$ and for each solver $s \in S$, 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 $P(\tau) = \frac{1}{n_p}\text{size}\{p \in P | r_{p,s} \le \tau\}$ for all $\tau > 0$, where sizeA, for any set A, stands for number of the elements in that set. Note that $P(\tau)$ is the probability for solver $s \in S$ such that a performance ratio $r_{p,s}$ is within a factor $\tau > 0$ 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 solved the most problems in a result that is within a factor $\tau$ of the best result.

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

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

Table 2: Tested methods

| Method name | | Algorithm |
|---|---|---|
| ML | : | method (1.12) and (1.13) by Moyi and Leong |
| mlSS-SR1(∗) | : | our method (2.12), (3.13) and (3.22) |
| mlSS-SR1(0.1) | : | our method (2.12), (3.13) and (3.21) with $\Gamma_k = 0.1$ |
| mlSS-SR1(0.01) | : | our method (2.12), (3.13) and (3.21) with $\Gamma_k = 0.01$ |
| mlSS-SR1(0.001) | : | our method (2.12), (3.13) and (3.21) with $\Gamma_k = 0.001$ |
| mlBFGS | : | memoryless BFGS method (1.6) |

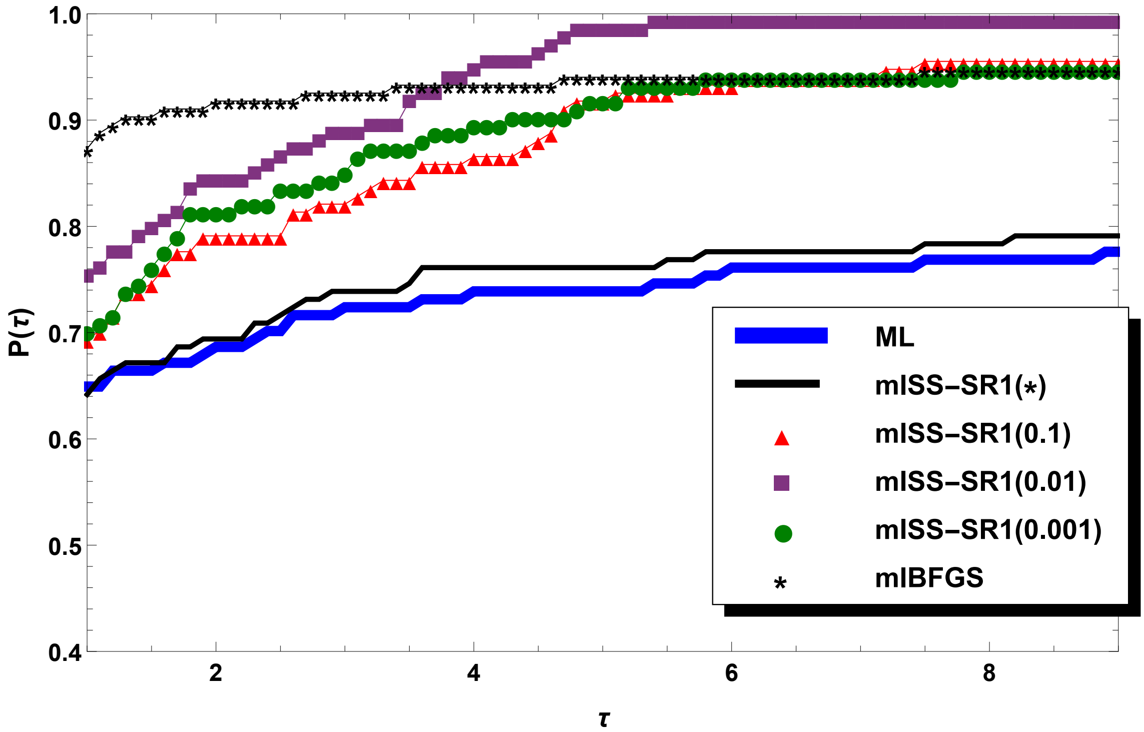


Figure 1: Performance profiles of memoryless quasi-Newton methods based on CPU time

In Figure 1, we give the performance profiles based on the CPU time. In order to prevent a measurement error, we set the minimum of the measurement 0.1 seconds. We observe from Figure 1 that mlSS-SR1(0.1), (0.01) and (0.001) are better than mlSS-SR1($*$) and ML. In our methods, parameter (3.21) with $\Gamma_k = 0.1,\ 0.01,\ 0.001$ are better than (3.22). Note that mlSS-SR1(0.01) performed better than mlSS-SR1(0.1) did, but mlSS-SR1(0.001) performed poorer than mlSS-SR1(0.01) did. As mentioned above, the performance of our method became poor as the value of $\Gamma_k$ approached 1. However, we cannot have any special tendency as the value of $\Gamma_k$ becomes small. Comparing mlSS-SR1(0.01) with mlBFGS, we observe that the performance profile of mlBFGS is over that of mlSS-SR1(0.01) in the interval $\tau < 4$, and the performance profile of mlSS-SR1(0.01) is over that of mlBFGS when $\tau \geq 4$. Therefore mlBFGS is superior to mlSS-SR1(0.01) from the viewpoint of the time efficiency. On the other hand, mlSS-SR1(0.01) is superior to mlBFGS from the viewpoint of the robustness.

## 5. Conclusion

In this paper, we have dealt with the spectral scaling secant condition, and we have derived the SS-SR1 formula (2.3). Based on the formula with the restart strategy, we have proposed the memoryless SS-SR1 method which always generates the sufficient descent direction and converges globally for general objective functions under standard assumptions and the Wolfe conditions (2.15) and (2.16). Finally, we have presented preliminary numerical results to investigate numerical performance of our method. Our further research is to find a suitable choice of $\gamma_{k-1}$ in (2.12).

## Acknowledgment

This research is supported in part by JSPS KAKENHI (grant number 17K00039).

## References


[1] N. Andrei: A scaled nonlinear conjugate gradient algorithm for unconstrained optimization. *Optimization*, **57** (2008), 549–570.

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

[3] H. Byrd, F. Khalfan, and B. Schnabel: Analysis of a symmetric rank-one trust region method. *SIAM Journal on Optimization*, **6** (1996), 1025–1039.

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

[5] A.R. Conn, N.I.M. Gould, and Ph.L. Toint:Convergence of quasi-Newton matrices generated by the symmetric rank one update. *Mathematical Programming*, **50** (1991), 177–195.

[6] Y.H. Dai: On the nonmonotone line search. *Journal of Optimization Theory and Applications*, **112** (2002), 312–330.

[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] R. Fletcher and C.M. Reeves: Function minimization by conjugate gradients. *The Computer Journal*, **7** (1964), 149–154.

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

[11] 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.

[12] W.W. Hager: Hager's web page: *http://people.clas.ufl.edu/hager/*, the last access date was May 2, 2017.

[13] 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.

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

[15] 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.

[16] 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), 409–436.

[17] S.B. Kafaki: On optimality of the parameters of self-scaling memoryless quasi-Newton updating formulae. *Journal of Optimization Theory and Applications*, **167** (2015), 91–101.

[18] 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.

[19] D.C. Liu and J. Nocedal: On the limited memory method for large-scale optimization. *Mathematical Programming*, **45** (1989), 503–528.

[20] 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.

[21] 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.

[22] Y. Narushima: Symmetric rank-one method based on some modified secant conditions for unconstrained optimization. *SUT Journal of Mathematics*, **47** (2011), 25–43.

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

[25] J. Nocedal: Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, **35** (1980), 773–782.

[26] J. Nocedal and S.J. Wright: *Numerical Optimization (Second edition), Springer Series in Operations Research* (Springer Verlag, New York, 2006).

[27] M.R. Osborne and L.P. Sun: A new approach to the symmetric rank one updating algorithm, *IMA Journal of Numerical Analysis*, **19** (1999), 497–508.

[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] W. Sun and Y. Yuan: *Optimization Theory and Methods: Nonlinear Programming, Springer Optimization and Its Application* (Springer Science+Business Media, New York, 2006).

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

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

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

[33] 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.

## Appendix

### Proof of Corollary 3.1

By letting $a = y_{k-1}^T y_{k-1}$, $b = s_{k-1}^T y_{k-1}$ and $c = s_{k-1}^T s_{k-1}$, parameter (3.22) is rewritten by

$$\gamma_{k-1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} - \sqrt{\left(\frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}\right)^2 - \frac{s_{k-1}^T s_{k-1}}{y_{k-1}^T y_{k-1}}} = \frac{c}{b} - \sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}}.$$

If condition $p_{k-1}^T y_{k-1} < \mu \|p_{k-1}\| \|y_{k-1}\|$ holds infinitely many times, the search direction (2.12) becomes the steepest descent direction infinitely many times, which guarantees (3.5).

Therefore, we consider the case that (2.11) is satisfied. Then we have $(s_{k-1}^T y_{k-1})^2 = b^2 < ac = s_{k-1}^T s_{k-1} y_{k-1}^T y_{k-1}$. In fact, if $b^2 = ac$ holds, then

$$\begin{aligned}(s_{k-1} - \gamma_{k-1} y_{k-1})^T y_{k-1} &= b - \frac{ac}{b} + a\sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}} \\ &= \frac{b^2 - ac}{b} + a\sqrt{\frac{c(ac - b^2)}{ab^2}} \\ &= 0,\end{aligned}$$

which implies that (2.11) does not hold. Since

$$\left(\frac{b}{a}\right)^2 - \frac{c}{a} = \frac{b^2 - ca}{a^2} < 0,$$

we obtain

$$\begin{aligned}\frac{b}{a} - \gamma_{k-1} &= \frac{b}{a} - \left(\frac{c}{b} - \sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}}\right) \\ &= \sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}} - \left(\frac{c}{b} - \frac{b}{a}\right) \\ &= \sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}} - \sqrt{\left(\frac{c}{b} - \frac{b}{a}\right)^2} \\ &= \sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}} - \sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a} - \frac{c}{a} + \left(\frac{b}{a}\right)^2} \\ &> 0.\end{aligned}$$

Hence we have

$$\gamma_{k-1} < \frac{b}{a} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}. \tag{A.1}$$

Next, we get

$$\frac{c}{b} - \frac{b}{2a} = \frac{2ac - b^2}{2ab} > 0$$

and

$$\begin{aligned}\left(\frac{c}{b} - \frac{b}{2a}\right)^2 &= \left(\frac{c}{b}\right)^2 - \frac{c}{a} + \left(\frac{b}{2a}\right)^2 \\ &> \left(\frac{c}{b}\right)^2 - \frac{c}{a} \\ &= \frac{c(ac - b^2)}{ab^2} \\ &> 0.\end{aligned}$$

From the above relations, we obtain

$$\sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}} < \frac{c}{b} - \frac{b}{2a}.$$

Parameter (3.22) yields

$$\gamma_{k-1} = \frac{c}{b} - \sqrt{\left(\frac{c}{b}\right)^2 - \frac{c}{a}} > \frac{b}{2a} = \frac{1}{2}\frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}. \tag{A.2}$$

Therefore, it follows from (A.1) and (A.2) that

$$\frac{1}{2}\frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}} < \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} - \sqrt{\left(\frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}\right)^2 - \frac{s_{k-1}^T s_{k-1}}{y_{k-1}^T y_{k-1}}} < \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}},$$

which implies that (3.1) holds with $\rho = 1/2$. Therefore, the result follows directly from Theorem 3.2. □

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