American Journal of Computational Mathematics, 2020, 10, 1-22
https://www.scirp.org/journal/ajcm
ISSN Online: 2161-1211
ISSN Print: 2161-1203

A Scaled Conjugate Gradient Method Based on New BFGS Secant Equation with Modified Nonmonotone Line Search

Tsegay Giday Woldu[1], Haibin Zhang[1], Yemane Hailu Fissuh[2]

[1]Department of Applied Mathematics, Beijing University of Technology, Beijing, China
[2]Department of Statistics, Aksum University, Aksum, Ethiopia
Email: tsegikuhil@gmail.com, zhanghaibn@bjut.edu.cn, yemaoroba@gmail.com

How to cite this paper: Woldu, T.G., Zhang, H.B. and Fissuh, Y.H. (2020) A Scaled Conjugate Gradient Method Based on New BFGS Secant Equation with Modified Nonmonotone Line Search. *American Journal of Computational Mathematics*, **10**, 1-22.
https://doi.org/10.4236/ajcm.2020.101001

Received: September 24, 2019
Accepted: January 5, 2020
Published: January 8, 2020

Abstract

In this paper, we provide and analyze a new scaled conjugate gradient method and its performance, based on the modified secant equation of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method and on a new modified nonmonotone line search technique. The method incorporates the modified BFGS secant equation in an effort to include the second order information of the objective function. The new secant equation has both gradient and function value information, and its update formula inherits the positive definiteness of Hessian approximation for general convex function. In order to improve the likelihood of finding a global optimal solution, we introduce a new modified nonmonotone line search technique. It is shown that, for nonsmooth convex problems, the proposed algorithm is globally convergent. Numerical results show that this new scaled conjugate gradient algorithm is promising and efficient for solving not only convex but also some large scale nonsmooth nonconvex problems in the sense of the Dolan-Moré performance profiles.

Keywords

Conjugate Gradient Method, BFGS Method, Modified Secant Equation, Nonmonotone Line Search, Nonsmooth Optimization

1. Introduction

The conjugate gradient method (CG) and Quasi-Newton method are two major popular iterative methods for solving smooth unconstrained optimization problems, and Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is one of the most efficient quasi-Newton methods for solving small and medium sized unconstrained optimization problems [1] [2] [3] [4]. The iterative method is computed by

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

where α_k is a step size and d_k is a search direction. For continuously differentiable function $h: R^n \to R$, the minimization problem:

$$\min_{x \in R^n} h(x) \tag{2}$$

has been well studied for several decades. Conjugate gradient method is among the preferable methods for solving problem (2) with search direction d_k given by

$$d_k = \begin{cases} -\nabla h_k + \beta_k d_{k-1} & \text{if } k \geq 1, \\ -\nabla h_k & \text{if } k = 0, \end{cases} \tag{3}$$

where ∇h_k is the gradient of an objective function $h(x)$ at k iterate and β_k is a scalar describing the attributes of the CG methods.

Some well-known formulas for the scalar β_k are the Hestenes-Stiefel (HS) [5], Fletcher-Reeves (FR) [6], Polak-Ribière and Polak (PRP) [7], and Dai-Yuan (DY) [8] given by

$$\beta_k^{HS} = \frac{\nabla h_k^{\mathrm{T}} y_k}{d_{k-1}^{\mathrm{T}} y_k}, \beta_k^{PRP} = \frac{\nabla h_k^{\mathrm{T}} y_k}{\|\nabla h_{k-1}\|^2},$$

$$\beta_k^{FR} = \frac{\|\nabla h_k\|^2}{\|\nabla h_{k-1}\|^2}, \beta_k^{DY} = \frac{\|\nabla h_k\|^2}{d_{k-1}^{\mathrm{T}} y_k},$$

where $y_k = \nabla h(x_k) - \nabla h(x_{k-1})$ and $\|\cdot\|$ denotes the Euclidean norm. Due to their simplicity and low memory requirement, CG methods are more effective and desirable for large scale unconstrained smooth problems [9] [10]. The global convergence properties of nonlinear CG methods have been analyzed under the weak Wolfe line search

$$\begin{cases} h(x_k + \alpha_k d_k) \leq h(x_k) + \varsigma \alpha_k \nabla h_k^{\mathrm{T}} d_k, \\ \nabla h(x_k + \alpha_k d_k)^{\mathrm{T}} d_k \geq \rho \nabla h_k^{\mathrm{T}} d_k, \end{cases} \tag{4}$$

and the strong Wolfe line search:

$$\begin{cases} h(x_k + \alpha_k d_k) \leq h(x_k) + \varsigma \alpha_k \nabla h_k^{\mathrm{T}} d_k, \\ \left|\nabla h(x_k + \alpha_k d_k)^{\mathrm{T}} d_k\right| \leq \rho \left|\nabla h_k^{\mathrm{T}} d_k\right|, \end{cases} \tag{5}$$

where $0 < \varsigma < \rho < 1$. CG methods use relatively little memory for large scale problems and require no numerical linear algebra, so each step is quite fast. However, they do not have second order information of the objective function, and typically converge much more slowly than Newton or quasi-Newton methods.

The quasi-Newton method is an iterative method with second order information of the objective function, and BFGS is the effective quasi-Newton method with the search direction

$$d_k = -B_k \nabla h_k, \tag{6}$$

where B_k is an approximation of the Hessian matrix of h at x_k. The update formula for B_k is defined by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} B_k s_k} + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}, \tag{7}$$

where s_k is defined as $s_k = x_{k+1} - x_k$, and the Hessian approximation B_{k+1} of (7) satisfies the standard secant equation

$$B_{k+1} s_k = y_k, \tag{8}$$

if $y_k^{\mathrm{T}} s_k > 0$, which is known as the curvature condition. The BFGS method has very interesting properties and remains one of the most respectable quasi-Newton methods for unconstrained optimization [11]. The theory of BFGS method and its global convergence have been established by many researchers (see [12]). For convex objective function, using some special inexact line search, it has been proved that the BFGS method is globally convergent (see [13] [14] [15]). However, when the objective function is nonconvex, the BFGS method under exact line search may fail to converge [16]. Moreover, Dai [17] proved that the BFGS method may fail for nonconvex functions with Wolfe line search techniques given in (4) and (5) [18]. Wolfe line search technique is the most common monotone line search technique, and it may leads to small steps without making significant progress to the minimum when the contours of the objective functions are a family of curves with large curvature (see [19] [20] [21] [22]). In order to overcome this drawback, the first nonmonotone line search technique was proposed by Grippo *et al.* [19] for Newton's method. With this initiative, many nonmonotone line search techniques have been proposed in recent years [23]. Yuan *et al.* [24] developed a modified limited memory BFGS method with the update formula that has a higher order approximation to exact Hessian, and its convergence property is analyzed under the nonmonotone line search type. However, the method converges for only uniformly convex functions. Li *et al.* [25] proposed a new BFGS algorithm with modified secant equation which achieves both global and superlinear convergence for generally convex functions under the nonmonotone line search of [19]. Su and Rong [26] introduced and established a new spectral CG method and its implementation under a modified nonmonotone line search technique. They introduced a new spectral conjugate gradient direction

$$d_k = \begin{cases} -\theta_k \nabla h_k + \beta_k d_{k-1} & \text{if } k \ge 1, \\ -\nabla h_0 & \text{if } k = 0, \end{cases} \tag{9}$$

where

$$\theta_k = 1 + \frac{\beta_k d_k^{\mathrm{T}} \nabla h_k}{\|\nabla h_k\|^2}, \tag{10}$$

$$\beta_k = \frac{\nabla h_k^{\mathrm{T}} y_{k-1}}{(1-\tau)\|\nabla h_{k-1}\|^2 + \tau d_{k-1}^{\mathrm{T}} y_{k-1}}, \tag{11}$$

and $\tau \in [0,1]$. It is not difficult to notice that the denominator of (11) is the convex combination of the denominator of the conjugate parameters in HS and PRP conjugate gradient methods. The choice of spectral parameter given (10) ensures the sufficient descent property of the search direction without dependence of line search. The convergence property of their method analyzed under a new modified nonmonotone line search with some mild conditions. However, this spectral CG method has only first order information, and excludes second order information. When the number of dimension is large, the CG methods are more effective compared to the BFGS methods in term of the CPU-time but in term of the number of iterations and the number of function evaluations, the BFGS methods are better. In order to incorporate the remarkable properties of the CG and BFGS methods and to overcome their drawbacks, many hybrid of CG and BFGS methods are introduced for unconstrained smooth optimization [27] [28] [29]. However, the usage of these methods is mainly restricted to solve smooth optimization problems. Recently, Yuan *et al.* [30] [31] [32] [33] introduced some CG approaches to solve nonsmooth convex large scale problems using the smoothing regularization, and under some assumptions, the global convergence properties of these approaches are analyzed. Yuan and Wei [34] proposed the Barzilai and Borwein (BB) gradient method with nonmonotone line search to solve nonsmooth convex optimization problems. Some implementable quasi-Newton methods are also introduced for solving the same problem (see [35] [36] [37] [38]). More recently, Ou and Zhou [39] introduced a modified scaled BFGS preconditioned CG algorithm, and under appropriate assumptions, the method is proven to possess global convergence for nonsmooth convex functions.

Motivated by the work of Ou and Zhou [39], in this paper, we propose a hybrid approach of the a scaled CG method and a modified BFGS method to combine the simplicity of CG method and the Hessian approximation of BFGS method. Our work is mainly focused in developing the scaled conjugate search direction that includes the second order information of the objective function by incorporating the modified secant equation of BFGS method. Opposing the work of Ou and Zhou [39], our method has both the function and gradient value information of the objective function. Moreover, our method leads to better descent direction than the CG methods proposed so far. To the best of our knowledge, this is the first work to incorporate the scaled CG algorithm with the BFGS secant equation which contains both the function and gradient value information of the objective function for solving large scale nonsmooth optimization. Under the new modified nonmonotone line search technique, the global convergence of the algorithm is analyzed for nonsmooth convex problems.

The paper is organized as follows. In the next section, we consider a nonsmooth convex problem and review their basic results. In Section 3, we propose a new scaled CG algorithm that incorporates the BFGS secant equation which has both function value and gradient information of the objective function via the smoothing regularization. Using the new modified nonmonotone line search technique, we prove the global convergence of our new algorithm for non-smooth convex problems. Numerical results and related comparisons are reported in Section 4. Finally, Section 5 concludes our work.

2. Nonsmooth Convex Problems and Their Basic Results

In this section, we consider the unconstrained optimization problem

$$\min_{x \in R^n} f(x), \tag{12}$$

where $f: R^n \to R$ is a possibly nonsmooth convex function. This problem is equivalent to the following problem

$$\min_{x \in R^n} F(x), \tag{13}$$

where $F: R^n \to R$ is the Moreau-Yosida regularization of f [40], which is defined by

$$F(x) = \min_{z \in R^n} \left\{ f(z) + \frac{1}{2\lambda} \|z - x\|^2 \right\}, \tag{14}$$

where λ is a positive parameter. The function F is a finite-valued, continuously differentiable convex function even though the function f is nondifferentiable (see [40]). Let $p(x)$ be the unique solution of (14). In what follows, we can express $F(x)$ as

$$F(x) = f(p(x)) + \frac{1}{2\lambda} \|p(x) - x\|^2. \tag{15}$$

Moreover, the gradient of F is globally Lipschitz continuous, *i.e.*,

$$\|g(x) - g(y)\| \le \frac{1}{\lambda} \|x - y\|, \ \forall x, y \in R^n, \tag{16}$$

where

$$g(x) = \nabla F(x) = \frac{x - p(x)}{\lambda}. \tag{17}$$

The point $x \in R^n$ is an optimal solution to (12) if and only if $g(x) = 0$ (see [40]). Furthermore, under reasonable conditions the gradient of F is semismooth and some of its remarkable properties are given in [41] [42].

Several methods have been proposed to solve (13) by incorporating bundle methods and quasi-Newton methods ideas [43] [44] [45], but it is burdensome to evaluate the exact value of $p(x)$ at any given point x [46]. Luckily, for each $x \in R^n$ and any $\varepsilon > 0$, we can have $p^\alpha(x, \varepsilon) \in R^n$ such that

$$f(p^\alpha(x, \varepsilon)) + \frac{1}{2\lambda} \|p^\alpha(x, \varepsilon) - x\|^2 \le F(x) + \varepsilon. \tag{18}$$

Therefore, we can approximate $F(x)$ and $g(x)$ by

$$F^{\alpha}(x,\varepsilon)=f\left(p^{\alpha}(x,\varepsilon)\right)+\frac{1}{2\lambda}\left\|p^{\alpha}(x,\varepsilon)-x\right\|^{2}, \tag{19}$$

and

$$g^{\alpha}(x,\varepsilon)=\frac{x-p^{\alpha}(x,\varepsilon)}{\lambda}, \tag{20}$$

respectively. Implementable algorithms to define such a $p^{\alpha}(x,\varepsilon)$ for non-smooth convex model can be seen in [47]. The noticeable attributes of $F^{\alpha}(x,\varepsilon)$ and $g^{\alpha}(x,\varepsilon)$ by the following proposition [48].

Proposition 1. Let $p^{\alpha}(x,\varepsilon)$ be a vector that satisfies (18), and let $F^{\alpha}(x,\varepsilon)$ and $g^{\alpha}(x,\varepsilon)$ be defined by (19) and (20), respectively. Then we obtain

$$F(x)\le F^{\alpha}(x,\varepsilon)\le F(x)+\varepsilon, \tag{21}$$

$$\left\|p^{\alpha}(x,\varepsilon)-p(x)\right\|\le\sqrt{2\lambda\varepsilon}, \tag{22}$$

and

$$\left\|g^{\alpha}(x,\varepsilon)-g(x)\right\|\le\sqrt{2\varepsilon/\lambda}. \tag{23}$$

Proposition 1 shows that the approximations of $F^{\alpha}(x,\varepsilon)$ and $g^{\alpha}(x,\varepsilon)$ can be made arbitrarily close to the exact values of $F(x)$ and $g(x)$ respectively.

3. A Scaled CG Method Based on New BFGS Secant Equation

In this section, we introduce the new scaled CG search direction that incorporates the modified BFGS secant equation, and then describe the new algorithm for solving nonsmooth problems. We make use of a modified nonmonotone line search technique introduced by [23] to compute a step size. Based on the above approximations, we redefine the search direction of CG method (3) to solve problem (13) as follows:

$$d_{k+1}=\begin{cases}-g^{\alpha}(x_{k+1},\varepsilon_{k+1})+\beta_{k+1}d_k & \text{if } k\ge 1,\\ -g^{\alpha}(x_{k+1},\varepsilon_{k+1}) & \text{if } k=0,\end{cases} \tag{24}$$

where ε is an appropriately chosen positive number. Ou and Zhou [39] provided a search direction defined by

$$d_{k+1}=\begin{cases}-\tilde{Q}_{k+1}g^{\alpha}(x_{k+1},\varepsilon_{k+1}) & \text{if } k\ge 1,\\ -g^{\alpha}(x_{k+1},\varepsilon_{k+1}) & \text{if } k=0,\end{cases} \tag{25}$$

where $\tilde{Q}_{k+1}\in R^{n\times n}$ is defined

$$\tilde{Q}_{k+1}=\tilde{\theta}_{k+1}I-\tilde{\theta}_{k+1}\frac{w_k s_k^{\mathrm{T}}+s_k w_k^{\mathrm{T}}}{w_k^{\mathrm{T}}s_k}+\left[1+\tilde{\theta}_{k+1}\frac{w_k^{\mathrm{T}}w_k}{w_k^{\mathrm{T}}s_k}\right]\frac{s_k s_k^{\mathrm{T}}}{w_k^{\mathrm{T}}s_k}, \tag{26}$$

with

$$\tilde{\theta}_{k+1}=\frac{s_k^{\mathrm{T}}s_k}{w_k^{\mathrm{T}}s_k},$$

where

$$w_k = y_k^* + t_k s_k. \tag{27}$$

The vector y_k^* and t_k in (27) are defined as

$$y_k^* = g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right) - g^\alpha\left(x_k, \varepsilon_k\right) \tag{28}$$

and

$$t_k = t + \max\left\{\frac{s_k^{\mathrm{T}} y_k}{\left\|s_k\right\|^2}, 0\right\}\left(t > 0\right). \tag{29}$$

It is easy to observe that the (27) has only gradient value information. In order to have both gradient and function value information, we replace (27) and (29) by

$$w_k^* = y_k^* + \max\left\{t_k^*, 0\right\} s_k, \tag{30}$$

and

$$t_k^* = \frac{6\left[F\left(x_k\right) - F\left(x_k + \alpha_k d_k\right)\right] + 3\left(g^\alpha\left(x_k + \alpha_k d_k, \varepsilon_{k+1}\right) + g^\alpha\left(x_k, \varepsilon_k\right)\right)^{\mathrm{T}} s_k}{\left\|s_k\right\|^2}, \tag{31}$$

respectively. Thus, the BFGS method with the secant equation

$$B_{k+1} s_k = w_k^*, \tag{32}$$

and the update formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} B_k s_k} + \frac{w_k^* w_k^{*\mathrm{T}}}{w_k^{*\mathrm{T}} s_k}, \tag{33}$$

has both gradient and function value information, and the matrix B_{k+1} inherits the positive definiteness of B_k for generally convex functions. Using the secant Equation (32), we propose the new search direction is defined by

$$d_{k+1} = \begin{cases} -\ddot{\theta}_{k+1} g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right) + \beta_{k+1} d_k - \vartheta_{k+1} w_k^*, & \text{if } k \geq 1, \\ -g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right), & \text{if } k = 0, \end{cases} \tag{34}$$

where

$$\ddot{\theta}_{k+1} = 2 - \frac{d_k^{\mathrm{T}} g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right)}{\left\|g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right)\right\|^2}\left(\frac{g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right)^{\mathrm{T}} w_k^*}{\left\|d_k\right\|\left\|w_k^*\right\|}\right), \tag{35}$$

$$\beta_{k+1} = \frac{g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right)^{\mathrm{T}} w_k^*}{\left\|d_k\right\|\left\|w_k^*\right\| + \left|d_k^{\mathrm{T}} y_k^*\right|}, \tag{36}$$

and

$$\vartheta_{k+1} = \frac{d_k^{\mathrm{T}} g^\alpha\left(x_{k+1}, \varepsilon_{k+1}\right)}{\left\|d_k\right\|\left\|w_k^*\right\|}. \tag{37}$$

Now, based on the above search direction, we describe our new scaled CG algorithm with a modified nonmonotone line search for solving problem (13) as follows.

Algorithm 1

Step 0. Given $\epsilon > 0, \beta \in (0,1), \varepsilon \in (0,1), \sigma \in (0,1)$, and a point $x_0 \in R^n$. Set $d_0 = -g^\alpha(x_0, \varepsilon_0)$ and $k := 0$.

Step 1. If $\|g^\alpha(x_k, \varepsilon_k)\| < \epsilon$, then stop, else go to the next step.

Step 2. Compute the search direction d_k by using (34)-(37).

Step 3. Set trial step size $\alpha_k = 1$.

Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$ and choose a scalar ε_{k+1} such that $0 < \varepsilon_{k+1} < \varepsilon_k$.

Step 5. Let $\mu \in (0,1]$, $M \geq 1$ is a positive integer, define $m(k) = \min\{k+1, M\}$, and choose

$$\mu_{ki} \geq \mu,\ i = 0, 1, 2, \cdots, m(k)-1,\ \sum_{i=0}^{m(k)-1} \mu_{ki} = 1.$$

Let $\alpha_k \geq 0$ be bounded above and satisfy:

$$\begin{aligned} &F^\alpha(x_k + \alpha_k d_k, \varepsilon_{k+1}) \\ &\leq \max\left[F^\alpha(x_k, \varepsilon_k), \sum_{i=0}^{m(k)-1} \mu_{ki} F^\alpha(x_{k-i}, \varepsilon_{k-i})\right] \sigma \alpha_k g^\alpha(x_k, \varepsilon_k)^{\mathrm{T}} d_k. \end{aligned} \tag{38}$$

If (38) does not holds, define $\alpha_k = \beta \alpha_k$ and go to step 5.

Step 6. Set $K := k + 1$ and go to step 1.

It can be observed that the line search technique in step 5 of Algorithm 1 is a nonmonotone line search technique with some modifications.

Convergence Analysis

In this subsection, we establish the global convergence of our method for non-smooth convex problem (12). To prove the global convergence of Algorithm 1, the following Lemmas are needed.

Lemma 1. *Assume that the search direction* d_k *is generated by Algorithm* 1, *then for all* $k \geq 0$, *we have*

$$g^\alpha(x_{k+1}, \varepsilon_{k+1})^{\mathrm{T}} d_{k+1} \leq -\|g^\alpha(x_{k+1}, \varepsilon_{k+1})\|^2, \tag{39}$$

and

$$\|d_{k+1}\| \leq 5\|g^\alpha(x_{k+1}, \varepsilon_{k+1})\|. \tag{40}$$

Proof. If $k = 0$, then

$$g^\alpha(x_0, \varepsilon_0)^{\mathrm{T}} d_0 = -g^\alpha(x_0, \varepsilon_0)^{\mathrm{T}} g^\alpha(x_0, \varepsilon_0) = -\|g^\alpha(x_0, \varepsilon_0)\|^2,$$

and

$$\|d_0\| = \|-g^\alpha(x_0, \varepsilon_0)\| \leq \|g^\alpha(x_0, \varepsilon_0)\| \leq 5\|g^\alpha(x_0, \varepsilon_0)\|.$$

Let $k \geq 1$, then from (34) we have

$$
\begin{aligned}
& g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} d_{k+1} \\
& =-\ddot{\theta}_{k+1}\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2}+\beta_{k+1} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} d_{k}-\vartheta_{k+1} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*} \\
& \leq-\ddot{\theta}_{k+1}\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2}+\frac{g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*}}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|+\left|d_{k}^{\mathrm{T}} y_{k}^{*}\right|} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} d_{k} \\
& \quad-\frac{d_{k}^{\mathrm{T}} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*} \\
& =-2\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2}+\frac{d_{k}^{\mathrm{T}} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*} \\
& \quad+\frac{g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*}}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|+\left|d_{k}^{\mathrm{T}} y_{k}^{*}\right|} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} d_{k} \\
& \quad-\frac{d_{k}^{\mathrm{T}} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*} \\
& =-2\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2}+\frac{g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*}}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|+\left|d_{k}^{\mathrm{T}} y_{k}^{*}\right|} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} d_{k} \\
& \leq-2\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2}+\left(\frac{\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2}}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|}\right)\left\|d_{k}\right\|\left\|w_{k}^{*}\right\| \\
& =-\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2} .
\end{aligned}
$$

Once more, (34) yields that

$$
\begin{aligned}
\left\|d_{k+1}\right\| & =\left\|-\ddot{\theta}_{k+1} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)+\beta_{k+1} d_{k}-\vartheta_{k+1} w_{k}^{*}\right\| \\
& =\left\|-\ddot{\theta}_{k+1} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)+\frac{g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)^{\mathrm{T}} w_{k}^{*}}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|+\left|d_{k}^{\mathrm{T}} y_{k}^{*}\right|} d_{k}-\frac{d_{k}^{\mathrm{T}} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|} w_{k}^{*}\right\| \\
& \leq\left\|\ddot{\theta}_{k+1} g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|+2\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\| \\
& \leq 4\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|+\left(\frac{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|}{\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{2}}\right) \frac{\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\|^{3}}{\left\|d_{k}\right\|\left\|w_{k}^{*}\right\|} \\
& \leq 5\left\|g^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\right\| .
\end{aligned}
$$

Thus, the proof is completed.

Lemma 1 shows that the search direction d_k developed in (34)-(37) leads to the most sufficiently descent direction and it belongs to a trust region.

Lemma 2. *Let the step size* α_k *satisfy* (38), *then there exist* $\beta > 0$ *satisfy a*

$$
\alpha_{k} \geq \min \left\{1, \frac{(1-\sigma) \beta}{L} \frac{\left|g^{\alpha}\left(x_{k},\varepsilon_{k}\right)^{\mathrm{T}} d_{k}\right|}{\left\|d_{k}\right\|^{2}}\right\}. \tag{41}
$$

Proof. If $\alpha_k = 1$ satisfies the formula (38), then the proof is completed. Otherwise, there exist β such that

$$
\begin{aligned}
&F^{\alpha}\left(x_k+\frac{\alpha_k}{\beta}d_k,\varepsilon_{k+1}\right)\\
&>\max\left\{F^{\alpha}\left(x_k,\varepsilon_k\right),\sum_{i=0}^{m(k)-1}\mu_{ki}F^{\alpha}\left(x_{k-i},\varepsilon_{k-i}\right)\right\}+\sigma\frac{\alpha_k}{\beta}g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\
&>F^{\alpha}\left(x_k,\varepsilon_k\right)+\sigma\frac{\alpha_k}{\beta}g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k.
\end{aligned}
$$

Thus,

$$
F^{\alpha}\left(x_k+\frac{\alpha_k}{\beta}d_k,\varepsilon_{k+1}\right)-F^{\alpha}\left(x_k,\varepsilon_k\right)>\sigma\frac{\alpha_k}{\beta}g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k. \tag{42}
$$

Using mean value theorem, we have

$$
\begin{aligned}
&F^{\alpha}\left(x_k+\alpha d_k,\varepsilon_{k+1}\right)-F^{\alpha}\left(x_k,\varepsilon_k\right)\\
&=\int_0^{\alpha}\left(g^{\alpha}\left(x_k+td_k,\varepsilon_{k+1}\right)-g^{\alpha}\left(x_k,\varepsilon_k\right)\right)^{\mathrm{T}}d_k\mathrm{d}t+\alpha g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\
&\le\frac{1}{2}L\alpha^2\left\|d_k\right\|^2+\alpha g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k.
\end{aligned}
$$

Combining the above inequality with (42), we have

$$
\alpha_k\ge\min\left\{1,\frac{(1-\sigma)\beta}{L}\frac{\left|g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\right|}{\left\|d_k\right\|^2}\right\}.
$$

Thus, the proof is completed.

Lemma 3. *Assume that the sequence* $\{x_k\}$ *is generated by Algorithm* 1, *then we have*

$$
\begin{aligned}
&F^{\alpha}\left(x_k,\varepsilon_k\right)\\
&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k-2}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i+\sigma\alpha_{k-1}g^{\alpha}\left(x_{k-1},\varepsilon_{k-1}\right)^{\mathrm{T}}d_{k-1}\\
&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k-1}\alpha_r g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i.
\end{aligned}
$$

Proof. We prove this lemma by induction. For $k=1$, by (38) and $\mu\le1$, we have

$$
\begin{aligned}
F^{\alpha}\left(x_1,\varepsilon_1\right)&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\sigma\alpha_0 g^{\alpha}\left(x_0,\varepsilon_0\right)^{\mathrm{T}}d_0\\
&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\alpha_0 g^{\alpha}\left(x_0,\varepsilon_0\right)d_0
\end{aligned}
$$

Assume the equation holds for $1,2,\cdots,k$, and we need to show for $k+1$. To show the condition, we have considered two cases.

Case 1:

$$
\max\left[F^{\alpha}\left(x_k,\varepsilon_k\right),\sum_{i=0}^{m(k)-1}\mu_{ki}F^{\alpha}\left(x_{k-i},\varepsilon_{k-i}\right)\right]=F^{\alpha}\left(x_k,\varepsilon_k\right).
$$

Then, from (38), we have

$$
\begin{aligned}
F^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)&=F^{\alpha}\left(x_k+\alpha_k d_k,\varepsilon_{k+1}\right)\\
&\le F^{\alpha}\left(x_k,\varepsilon_k\right)+\sigma\alpha_k g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\
&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k-1}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i+\sigma\alpha_k g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\
&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i.
\end{aligned}
$$

Case 2:

$$\max\left[F^{\alpha}\left(x_k,\varepsilon_k\right),\ \sum_{i=0}^{m(k)-1}\mu_{ki}F^{\alpha}\left(x_{k-i},\varepsilon_{k-i}\right)\right]=\sum_{i=0}^{m(k)-1}\mu_{ki}F^{\alpha}\left(x_{k-i},\varepsilon_{k-i}\right),$$

let $n=\min\left[k,m-1\right]$. Then, again from (38),

$$\begin{aligned}F^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)&=F^{\alpha}\left(x_k+\alpha_k d_k,\varepsilon_{k+1}\right)\\&\le\sum_{j=0}^{n}\mu_{kj}F^{\alpha}\left(x_{k-j},\varepsilon_{k-j}\right)+\sigma\alpha_k g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\&\le\sum_{j=0}^{n}\mu_{kj}\left[F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k-j-2}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i\right.\\&\quad\left.+\sigma\alpha_{k-j-1}g^{\alpha}\left(x_{k-j-1},\varepsilon_{k-j-1}\right)^{\mathrm{T}}d_{k-j-1}\right]+\sigma\alpha_k g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k.\end{aligned}$$

Thus, by imposing

$$\left(1,2,\cdots,n\right)\times\left(1,2,\cdots,k-n-2\right)\subset\left\{\left(j,i\right):0\le j\le n,0\le i\le k-n-2\right\},$$

and

$$\sum_{j=0}^{n}\mu_{kj}=1,\ \mu_{kj}\ge\mu,$$

we have

$$\begin{aligned}&F^{\alpha}\left(x_{k+1},\varepsilon_{k+1}\right)\\&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sum_{i=0}^{k-n-2}\left(\sum_{j=0}^{n}\mu_{kj}\right)\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i\\&\quad+\sigma\sum_{j=0}^{n}\mu_{kj}\alpha_{k-j-1}g^{\alpha}\left(x_{k-j-1},\varepsilon_{k-j-1}\right)^{\mathrm{T}}d_{k-j-1}+\sigma\alpha_k g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k-n-2}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i\\&\quad+\mu\sigma\sum_{i=k-j-1}^{k-1}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i+\sigma\alpha_k g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\&=F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k-1}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i+\sigma\alpha_k g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k\\&\le F^{\alpha}\left(x_0,\varepsilon_0\right)+\mu\sigma\sum_{i=0}^{k}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}}d_i.\end{aligned}$$

Theorem 1. *Assume that the sequences* $\left\{x_k\right\}$ *and* $\left\{d_k\right\}$ *are generated by Algorithm 1. Let F is bounded below on the level set* $\mathcal{L}_0=\left\{x\in R^n\mid F\left(x\right)\le F\left(x_0\right)\right\}$ *and*

$$\lim_{k\to\infty}\varepsilon_k=0.$$

Then

$$\lim_{k\to\infty}g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}}d_k=0. \tag{43}$$

Proof. Suppose that (43) is not true. Then there exist constants $\gamma>0$ and k_0 such that

$$g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}} d_k \le -\gamma, \quad \forall k > k_0. \tag{44}$$

From Lemma 3, we have

$$F^{\alpha}\left(x_0,\varepsilon_0\right) - F^{\alpha}\left(x_k,\varepsilon_k\right) \ge -\mu\sigma\sum_{i=0}^{k-1}\alpha_r g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}} d_i. \tag{45}$$

By (40), (41) and (44), we have

$$\begin{aligned}
&F^{\alpha}\left(x_0,\varepsilon_0\right) - F^{\alpha}\left(x_k,\varepsilon_k\right)\\
&\ge -\mu\sigma\sum_{i=0}^{k-1}\alpha_i g^{\alpha}\left(x_i,\varepsilon_i\right)^{\mathrm{T}} d_i \ge \mu\sigma\gamma\sum_{i=0}^{k-1}\alpha_i\\
&\ge \mu\sigma\gamma\sum_{i=0}^{k-1}\min\left\{1, \frac{\left(1-\sigma\right)\beta}{L}\frac{\left|g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}} d_k\right|}{\left\|d_k\right\|^2}\right\}\\
&\ge \mu\sigma\gamma\sum_{i=0}^{k-1}\min\left\{1, \frac{\left(1-\sigma\right)\beta}{25L}\right\}.
\end{aligned}$$

Letting $k \to \infty$, we have

$$\mu\sigma\gamma\sum_{k=0}^{\infty}\min\left\{1, \frac{\left(1-\sigma\right)\beta}{25L}\right\} \le \sum_{k=0}^{\infty} F^{\alpha}\left(x_0,\varepsilon_0\right) - F^{\alpha}\left(x_k,\varepsilon_k\right),$$

and this contradicts our assumption on *F*. Hence the theorem is proved.

Theorem 2. *Let the conditions in Lemma 1 and Theorem 1 hold, then Algorithm 1 converges for nonsmooth problem* (12).

Proof. From Lemma 1 and Theorem 1, we have

$$0 \ge \lim_{k\to\infty}\left(-\left\|g^{\alpha}\left(x_k,\varepsilon_k\right)\right\|^2\right) \ge \lim_{k\to\infty} g^{\alpha}\left(x_k,\varepsilon_k\right)^{\mathrm{T}} d_k = 0.$$

Then,

$$\lim_{k\to\infty}\left\|g^{\alpha}\left(x_k,\varepsilon_k\right)\right\| = 0. \tag{46}$$

Thus, (23) and convergence of sequence $\left\{\varepsilon_k\right\}$ yield

$$0 \le \lim_{k\to\infty}\left\|g^{\alpha}\left(x_k,\varepsilon_k\right) - g\left(x_k\right)\right\| \le \lim_{k\to\infty}\sqrt{2\varepsilon/\lambda} = 0.$$

Hence,

$$\lim_{k\to\infty}\left\|g\left(x_k\right)\right\| = 0. \tag{47}$$

Let x^* be an accumulation point of $\left\{x_k\right\}$. Then there exists a subsequence $\left\{x_k\right\}_K$ satisfying

$$\lim_{k\in K, k\to\infty} x_k = x^*. \tag{48}$$

Thus, (17), (43) and (47) yield $x^* = p\left(x^*\right)$. Therefore x^* is an optimal solution of nonsmooth problem (12).

4. Numerical Experiments for Large Scale Nonsmooth Problems

In this section, we present some numerical experiments to examine the efficiency of Algorithm 1 for some large scale nonsmooth academic test problems which are introduced in [49]. The details of these large scale nonsmooth academic test problems with their initial points $x_i^{(1)}$ and the minimum values $f(x^*)$ are listed as follows:

Problem 1

$$f(x) = \max_{1 \le i \le n} x_i^2$$

$x_i^{(1)} = i$ for $i = 1, \cdots, n/2$ and
$x_i^{(1)} = -i$ for $i = n/2 + 1, \cdots, n$

$$f(x^*) = 0.$$

Problem 2

$$f(x) = \max_{1 \le i \le n} \left| \sum_{j=1}^{n} \frac{x_j}{i + j - 1} \right|$$

$x_i^{(1)} = i$ for $i = 1, \cdots, n$.

$$f(x^*) = 0.$$

Problem 3

$$f(x) = \sum_{i=1}^{n-1} \max\left\{ -x_i - x_{i+1}, -x_i - x_{i+1} + \left(x_i^2 + x_{i+1}^2 - 1 \right) \right\}$$

$x_i^{(1)} = -0.5$ for $i = 1, \cdots, n$;
$f(x^*) = -\sqrt{2(n-1)}$.

Problem 4

$$f(x) = \sum_{i=1}^{n-1} \max\left\{ x_i^4 + x_{i+1}^2, (2 - x_i)^2 + (2 - x_{i+1})^2, 2e^{-x_i + x_{i+1}} \right\}$$

$x_i^{(1)} = 2$ for $i = 1, \cdots, n$;
$f(x^*) = 2(n-1)$.

Problem 5

$$f(x) = \max\left\{ \sum_{i=1}^{n-1} \left(x_i^4 + x_{i+1}^2 \right), \sum_{i=1}^{n-1} \left((2 - x_i)^2 + (2 - x_{i+1})^2 \right), \sum_{i=1}^{n-1} \left(2e^{-x_i + x_{i+1}} \right) \right\}$$

$x_i^{(1)} = 2$ for $i = 1, \cdots, n$;
$f(x^*) = 2(n-1)$.

Problem 6

$$f(x) = \max_{1 \le i \le n} \left\{ g\left(-\sum_{i=1}^{n} x_i \right), g(x_i) \right\},$$

where $g(y) = \ln(|y| + 1)$;
$x_i^{(1)} = 1$ for $i = 1, \cdots, n$;
$f(x^*) = 0$.

Problem 7

$$f(x) = \sum_{i=1}^{n-1} \left(|x_i|^{x_{i+1}^2 + 1} + |x_{i+1}|^{x_i^2 + 1} \right)$$

$x_i^{(1)} = -1$ when $\mathrm{mod}(i,2) = 1, (i = 1, \cdots, n)$ and
$x_i^{(1)} = 1$ when $\mathrm{mod}(i,2) = 0, (i = 1, \cdots, n)$;
$f(x^*) = 0.$

Problem 8

$$f(x) = \sum_{i=1}^{n-1} \left(-x_i + 2\left(x_i^2 + x_{i+1}^2 - 1\right) + 1.75\left|x_i^2 + x_{i+1}^2 - 1\right| \right)$$

$x_i^{(1)} = -1$ for $i = 1, \cdots, n$;
$f(x^*) = varies.$

Problem 9

$$f(x) = \max\left\{ \sum_{i=1}^{n-1} \left(x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1 \right), \sum_{i=1}^{n-1} \left(-x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1 \right) \right\}$$

$x_i^{(1)} = -1.5$ when $\mathrm{mod}(i,2) = 1, (i = 1, \cdots, n)$ and
$x_i^{(1)} = 2.0$ when $\mathrm{mod}(i,2) = 0, (i = 1, \cdots, n)$;
$f(x^*) = 0.$

Problem 10

$$f(x) = \sum_{i=1}^{n-1} \max\left\{ x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1, -x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1 \right\}$$

$x_i^{(1)} = -1.5$ when $\mathrm{mod}(i,2) = 1, (i = 1, \cdots, n)$ and
$x_i^{(1)} = 2.0$ when $\mathrm{mod}(i,2) = 0, (i = 1, \cdots, n)$;
$f(x^*) = 0.$

The problems 1 - 5 are convex functions, and the others are nonconvex functions. We test the above problems with the dimension of $n = 1000$, $n = 3000$, $n = 5000$, $n = 6000$, $n = 10000$, $n = 12000$, $n = 20000$, $n = 50000$, $n = 60000$ and $n = 100000$. For convenience sake, we denote Algorithm 1 by scaled conjugate gradient method based on modified secant equation of BFGS method (**SCG-MBFGS**), and in order to demonstrate validity of our algorithm, we also list the results of other three algorithms **MPRP** in [30], **MHS** in [31] and **MSBFGS-CG** in [39]. All algorithms were implemented in Fortran 90 and run on a PC with an intel(R) Core(TM)i3-3110M CPU at 2.40 GHz, 4.00 GB of RAM, and the Windows 7 operating system. We stopped the iteration when the condition $\left\| g^\alpha (x_k, \varepsilon_k) \right\| \le 10^{-10}$ was satisfied. The parameters for **SCG-MBFGS** were chosen as $M = 10 \beta = 0.6, \sigma = 0.85 \lambda = \mu = 1$. All parameters for other three methods are chosen as in [30] [31] and [39] respectively. **Table 1** shows the numerical results of **SCG-MBFGS**, **MPRP**, **MHS** and **MSBFGS-CG** on the given test problems. The columns in **Table 1** have the following meanings:

Dim: the dimensions of problem.

NI: the total number of iterations.

NF: the number of function evaluations.

TIME: the CPU time in seconds.

$f(x)$: the value of $f(x)$ at the final iteration.

From the numerical results in **Table 1**, it is not difficult to see that

Table 1. Numerical results for 10 problems with given initial points and dimensions.

No.	Dim	Algorithm 3.1	MHS	MPRP	MSBFGS−CG
		NI/NF/$f(x)$/TIME	NI/NF/$f(x)$/TIME	NI/NF/$f(x)$/TIME	NI/NF/$f(x)$/TIME
1	1000	213/4351/6.9117E−8/0.3034E+0	225/4710/6.9354E−8/0.4358E+0	225/4710/6.7102E−6/0.6947E+0	186/1601/2.6568E−9/0.5977E+0
	3000	219/4509/1.9754E−8/0.2356E+0	238/5273/6.6915E−8/1.1157E+0	228/5196/6.5327E−6/1.6902E+0	219/2310/2.2089E−9/1.3871E+0
	5000	224/4597/5.7735E−7/1.6248E+0	246/5284/8.1495E−7/1.2650E+0	250/5253/9.0184E−7/1.9379E+0	242/2725/1.2183E−9/1.6962E+0
	6000	227/4608/6.81679E−8/1.1756E+0	253/5298/6.7261E−8/1.2323E+0	252/5308/5.8563E−6/2.0417E+0	248/2849/1.8972E−8/2.0233E+0
	10,000	230/4719/3.0882E−6/1.9051E+0	264/5387/7.3580E−6/1.8361E+0	261/5466/3.4089E−8/2.2684E+0	253/2996/3.4045E−9/2.1527E+0
	12,000	234/4841/8.5011E−8/2.1084E+0	264/5529/6.5042E−8/2.4321E+0	266/5573/9.0925E−7/2.3537E+0	261/3328/4.9085E−9/2.2886E+0
	20,000	243/5150/7.7681E−8/2.43061E+0	275/5830/6.9187E−8/2.6893E+0	270/5769/5.68982E−8/2.8049E+0	275/4573/6.1823E−10/2.6981E+0
	50,000	268/5539/1.0736E−8/2.9909E+0	289/6002/7.6945E−7/2.8937E+0	286/5991/8.0015E−6/3.0539E+0	283/5597/7.0687E−8/2.9298E+0
	60,000	277/5913/5.9244E−8/3.1618E+0	294/6125/6.4520E−8/2.9931E+0	295/6110/9.6412E−7/3.2291E+0	288/6018/1.7851E−8/3.1037E+0
	100,000	309/6884/8.6471E−8/4.1954E+0	324/7018/6.7620E−8/3.0068E+0	311/6983/5.1947E−7/4.7095E+0	311/6836/6.6109E−8/3.6814E+0
2	1000	58/1507/8.0315E−8/5.3728E+1	96/1605/8.0248E−9/5.4181E+1	91/1482/6.1664E−9/2.9639E+1	94/1301/4.0582E−9/4.1905E+1
	3000	66/1682/6.7354E−9/5.3848E+2	110/1917/6.9902E−9/5.8991E+2	105/11760/8.8715E−6/1.7785E+2	103/1420/3.0085E−9/2.5608E+2
	5000	77/1920/5.9738E−7/6.8072E+2	115/1963/7.1280E−5/8.8693E+2	111/1938/5.0135E−7/7.8101E+2	119/1499/2.6731E−9/9.6649E+2
	6000	84/2007/8.3699E−9/2.3108E+3	118/2103/8.45247E−9/2.5384E+3	112/1966/6.3507E−7/1.5338E+3	129/1573/1.9629E−9/1.9867E+3
	10,000	93/2235/5.6492E−6/2.9875E+3	126/2302/6.0153E−7/5.5084E+3	120/2127/4.8937E−8/3.3185E+3	141/1901/4.8003E−9/4.7503E+3
	12,000	96/2276/9.3619E−9/1.0518E+4	127/2310/9.5284E−9/1.1372E+4	123/2162/7.0430E−5/1.0882E+4	147/2243/6.1285E−8/1.2026E+4
	20,000	108/2419/8.9409E−9/1.6564E+4	165/2676/7.9950E−9/1.7582E+4	130/2296/2.1027E−8/1.0947E+4	155/2489/3.0318E−8/1.3732E+4
	50,000	119/2693/9.0057E−6/1.9714E+4	185/2709/5.8276E−8/1.8972E+4	141/2604/4.8826E−5/1.4608E+4	168/2707/5.6590E−7/1.5809E+4
	60,000	121/2701/9.9354E−9/2.0018E+4	215/2786/6.9338E−9/2.0358E+4	148/2651/3.0798E−5/1.5101E+4	173/2842/3.7206E−7/1.8104E+4
	100,000	127/2820/8.9262E−9/2.1883E+4	240/3008/6.9393E−9/2.2960E+4	195/2907/8.7203E−8/3.1048E+4	189/2803/6.6907E−8/2.1061E+4
3	1000	16/79/6.2899E−9/5.0301E−2	37/114/7.2687E−9/4.8344E−2	37/114/7.2687E−9/0.3916E−1	37/110/2.3278E−9/0.2861E−1
	3000	23/82/5.6856E−7/3.1106E−2	37/123/5.3594E−9/4.4758E−2	37/116/8.5786E−9/0.56813E−1	37/113/4.0082E−9/0.4193E−1
	5000	28/89/8.3625E−8/4.4800E−2	39/123/6.0082E−6/6.1701E−2	39/120/9.0931E−9/0.6191E−1	39/116/5.9987E−9/0.5518E−1
	6000	31/90/6.8937E−9/5.4075E−2	40/123/5.4561E−9/7.6563E−2	39/121/8.4838E−9/0.8434E−1	39/117/6.1096E−9/0.6991E−1
	10,000	39/110/5.9901E−7/1.107E−1	40/125/4.9830E−8/1.1071E−1	40/123/9.0941E−9/0.5089E+0	40/121/1.6943E−9/0.4805E+0
	12,000	40/113/1.5721E−9/1.125E−1	40/125/1.6369E−9/1.3819E−1	40/125/8.6969E−9/0.7394E+0	40/123/1.8803E−9/0.6097E+0
	20,000	44/135/6.0317E−9/2.011E−1	50/157/6.5211E−9/1.5450E−1	48/151/7.1062E−9/0.6928E+0	41/149/3.9088E−9/0.6681E+0
	50,000	47/158/7.0052E−9/8.062E−1	50/160/5.9735E−6/6.1943E−1	55/153/6.1343E−9/0.7906E+0	49/161/6.1874E−9/0.6714E+0
	60,000	49/161/6.0976E−9/8.3069E−1	50/165/5.1158E−9/7.4703E−1	59/158/6.1999E−9/0.8005E+0	53/161/6.607E−9/0.6968E+0
	100,000	55/180/6.1437E−9/1.169E+0	60/174/6.0371E−9/1.0152E+0	60/166/5.1093E−9/1.5108E+0	62/173/7.0773E−9/1.3005E+0
4	1000	5/74/2.8725E+3/0.0416E+0	7/89/2.5699E+3/0.3053E−1	7/89/1.9980E+3/0.1489E−1	7/89/1.8835E+3/0.1069E−1
	3000	5/74/1.9435E+4/0.0522E+0	7/89/2.3654E+4/0.4348E−1	7/89/1.7105E+3/0.1863E−1	7/89/1.7098E+3/0.1306E−1
	5000	5/74/1.5184E+5/0.0637E+0	7/89/1.6941E+4/0.4463E−1	7/89/1.7329E+4/0.2814E+0	7/89/1.8026E+3/0.2804E+0
	6000	5/85/1.5352E+4/0.0696E+0	7/89/1.5433E+4/0.48906E−1	7/89/1.1998E+4/0.3819E+0	7/89/1.4620E+4/0.2997E+0
	10,000	6/89/1.2674E+4/0.1009E+0	7/89/2.0593E+4/0.8813E−1	7/89/1.8801E+5/0.4097E+0	7/89/1.3806E+4/0.3991E+0

Continued

	12,000	6/89/2.99787E+4/0.1047E+0	7/89/3.0868E+4/0.9556E−1	7/89/2.3998E+4/0.5523E+0	7/89/2.2855E+4/0.4895E+0
	20,000	6/89/2.9304E+4/0.4095E+0	7/89/8.3594E+4/0.3845E+0	7/89/1.6858E+5/0.9945E+0	7/89/1.5008E+4/0.8028E+0
	50,000	7/89/1.9998E+5/0.4569E+0	7/89/5.0973E+4/0.4611E+0	7/89/2.0017E+4/1.2931E+0	7/89/1.8801E+4/1.1561E+0
	60,000	7/89/2.0953E+5/0.9996E+0	7/89/1.5435E+5/0.6831E+0	7/89/1.1999E+5/1.3883E+0	7/89/1.4096E+5/1.3468E+0
	100,000	7/89/1.9468E+5/3.5798E+0	7/89/2.04359E+5/3.4358E+0	7/89/2.0071E+5/6.1947E+0	7/89/2.1188E+6/6.1609E+0
5	1000	5/22/1.9899E+3/0.6409E−2	7/86/2.5699E+3/0.7187E−3	7/85/1.9980E+3/0.6281E−1	7/86/1.1063E+4/0.5437E−1
	3000	7/29/2.3548E+3/0.1138E−1	17/95/1.9353E+4/0.2854E−2	16/95/2.7126E+3/0.7897E−1	16/94/2.0019E+4/0.7669E−1
	5000	14/42/3.0083E+3/0.1307E−1	21/114/2.0891E+4/0.6341E−2	21/114/1.8462E+4/0.8633E−1	21/114/1.1192E+4/0.6675E−1
	6000	15/58/1.6253E+4/0.15625E−1	21/114/1.5433E+4/0.6912E−2	21/114/1.1999E+4/0.9487E−1	21/114/1.0930E+4/0.9036E−1
	10,000	15/87/2.5984E+3/0.1171E+0	21/114/4.8327E+5/0.1088E+0	21/114/1.5827E+4/0.1984E+0	21/114/2.3072E+4/0.1772E+0
	12,000	16/92/3.0879E+4/0.1299E+0	21/114/3.0868E+4/0.1108E+0	21/114/2.3998E+4/0.2789E+0	21/114/2.1984E+4/0.2683E+0
	20,000	19/96/1.9348E+5/0.1515E+0	21/114/2.9306E+4/0.1935E+0	21/114/1.7102E+5/0.3195E+0	21/114/1.5120E+5/0.2956E+0
	50,000	22/101/2.1184E+4/0.4091E+0	21/114/1.8256E+5/0.3638E+0	21/114/1.5293E+5/0.4093E+0	21/114/2.0968E+5/0.3885E+0
	60,000	22/115/1.9356E+5/0.6657E+0	21/114/1.5434E+5/0.52975E+0	21/114/1.1999E+5/0.5383E+0	21/114/2.3001E+5/0.4896E+0
	100,000	25/117/2.0009E+5/1.3158E+0	21/114/1.6972E+5/1.4358E+0	21/114/1.7582E+5/4.9145E+0	21/114/2.2293E+5/3.6984E+0
6	1000	59/1047/6.1866E−9/2.8017E−1	78/1047/6.3785E−9/2.0109E−1	77/1026/6.8037E−9/1.0741E−1	71/891/3.6735E−9/0.1009E+0
	3000	63/1216/6.2814E−9/2.9565E−1	85/1168/5.6373E−9/2.4364E−1	81/1153/7.0961E−9/5.9417E−1	77/921/6.8083E−9/0.5067E+0
	5000	66/1290/8.5604E−9/1.1307E+0	91/1308/7.9629E−8/1.0254E+0	90/1281/7.8404E−9/1.3245E+0	82/937/7.9864E−9/1.2838E+0
	6000	68/1374/6.9799E−9/1.1864E+0	92/1323/8.2692E−9/1.0613E+0	91/1316/5.5483E−9/1.5115E+0	86/998/7.8405E−9/1.3644E+0
	10,000	88/1457/8.0915E−8/2.1735E+0	94/1430/6.6018E−7/2.1506E+0	96/1401/9.9366E−9/2.9348E+0	88/1208/5.7163E−9/2.8861E+0
	12,000	95/1492/6.8749E−9/2.2673E+0	99/1452/5.6143E−9/2.2001E+0	97/1440/8.5456E−9/3.1554E+0	93/1417/9.9355E−9/3.1047E+0
	20,000	106/1504/4.5481E−9/3.5061E+0	107/1671/6.9926E−9/3.3804E+0	104/1597/6.7138E−9/5.1947E+0	105/1531/6.8017E−9/5.0093E+0
	50,000	110/1691/5.2964E−7/1.2993E+1	110/1688/8.0153E−6/1.2054E+1	110/1665/6.1343E−9/1.1832E+1	112/1682/7.7084E−9/1.1570E+1
	60,000	112/1702/6.7549E−9/1.3608E+1	112/1707/6.4697E−9/1.2916E+1	112/1698/7.5022E−9/1.3527E+1	112/1704/5.9938E−9/1.2643E+1
	100,000	117/1808/6.9354E−8/8.968E+1	119/1793/7.0642E−9/1.0054E+2	116/1731/6.7139E−9/9.4368E+1	120/1750/8.8802E−9/9.1146E+1
7	1000	30/81/6.7682E−9/1.4821E−1	38/117/7.2687E−9/1.2581E−1	38/117/7.2687E−9/0.7074E+0	35/114/1.5021E−9/0.6621E+0
	3000	36/85/5.8186E−9/1.0918E−1	39/121/6.5165E−9/1.7375E−1	39/121/9.2894E−9/0.8198E+0	37/116/3.8037E−9/0.6997E+0
	5000	38/94/7.1437E−8/2.2432E−1	41/123/6.2047E−7/5.3942E−1	40/123/9.0932E−9/1.8426E+0	39/120/5.2352E−9/1.6801E+0
	6000	39/102/5.9864E−9/2.7029E−1	41/126/5.4561E−9/7.1584E−1	40/123/9.8411E−8/4.6316E+0	40/123/7.0942E−9/4.3308E+0
	10,000	41/117/9.9318E−5/1.9885E+0	46/129/6.1533E−8/1.0827E+0	41/125/1.8188E−8/5.9017E+0	41/126/2.1136E−9/5.8892E+0
	12,000	44/122/8.2994E−9/2.0108E+0	48/131/8.1848E−9/1.3739E+0	46/130/9.1246E−9/8.4288E+0	46/129/1.8277E−9/7.8591E+0
	20,000	46/142/6.9354E−8/2.1563E+0	54/162/6.9354E−8/5.5384E+0	49/142/7.8906E−9/9.9472E+0	47/138/5.0472E−9/9.0988E+0
	50,000	51/163/6.8141E−8/9.9182E+0	67/185/7.0255E−7/8.2977E+0	55/153/9.0948E−8/1.5013E+1	58/160/6.5509E−9/1.6395E+0
	60,000	51/169/6.9354E−8/1.0056E+1	72/191/6.9354E−8/9.7955E+0	57/160/1.7923E−8/2.1545E+1	60/164/1.4301E−9/1.9896E+1
	100,000	64/1905/6.9354E−8/1.4891E+1	87/2012/6.9354E−8/1.4358E+1	66/1936/1.8862E−8/7.4791E+1	71/1942/2.0446E−9/7.0104E+1
8	1000	31/77/−2.9735E+2/1.5502E−2	36/114/−2.4974E+2/3.2875E−2	37/114/−2.4975E+4/3.9808E−2	38/116/−5.3979E+3/3.5074E−2
	3000	34/84/−2.1068E+2/2.9607E−2	39/118/−1.8852E+2/3.7518E−2	38/115/−2.0965E+4/7.17093E−2	39/119/−2.4796E+4/4.6310E−2
	5000	37/88/−1.7664E+3/3.3509E−2	39/122/−1.5326E+3/6.3006E−2	39/120/−1.24995E+5/1.9137E−1	41/123/−3.2153E+4/1.7892E−1

Continued

	6000	37/94/−1.4998E+3/3.8201E−2	39/123/−1.4998E+3/7.9009E−2	39/120/−1.2488E+5/2.0213E−1	41/126/−1.9248E+4/1.8068E−1
	10,000	39/101/−1.5683E+3/1.0021E−1	40/126/−2.7053E+3/1.1906E−1	40/123/−2.4997E+5/3.0749E−1	43/129/−2.0127E+4/2.9976E−1
	12,000	40/107/−2.9998E+3/1.0131E−1	40/126/−2.9998E+3/1.5488E−1	40/123/−1.4887E+5/4.7906E−1	43/131/−2.4998E+4/3.7175E−1
	20,000	40/113/−2.9231E+3/4.1561E−1	41/138/−2.9409E+3/3.4008E−1	41/126/−2.5827E+6/9.0468E−1	46/137/−2.5037E+5/6.5936E−1
	50,000	40/120/−2.0583E+3/7.0931E−1	42/154/−1.8099E+3/5.9141E−1	43/157/−1.24997E+6/9.3709E−1	47/165/−1.7036E+5/7.9384E−1
	60,000	41/137/−1.9357E+3/8.5164E−1	42/160/−1.4999E+3/6.8471E−1	57/161/−1.24995E+6/9.9741E−1	54/174/−1.4881E+5/8.8056E−1
	100,000	41/152/−1.5094E+3/1.0171E+0	46/197/−2.9354E+3/9.8405E−1	71/196/−1.499E+5/9.4011E−1	66/205/−1.6068E+5/9.0985E−1
9	1000	36/47/9.0397E−9/4.5592E−2	39/174/9.04688E−9/3.265E−2	37/114/5.4897E−9/4.6481E−2	34/101/7.0463E−9/3.9807E−2
	3000	39/55/6.5101E−9/4.9901E−2	39/176/6.8510E−9/3.375E−2	38/117/5.4993E−9/6.1903E−2	36/106/5.5439E−9/4.364E−2
	5000	39/61/8.2853E−9/8.8256E−2	41/181/7.0387E−9/7.9083E−2	39/120/6.8294E−9/9.9665E−2	36/113/3.8145E−9/5.1750E−2
	6000	40/67/5.9491E−9/9.6753E−2	42/183/6.7514E−9/9.4875E−2	39/121/6.9655E−9/1.6048E−1	39/114/5.6824E−9/1.2194E−1
	10,000	40/88/6.5687E−8/1.7263E−1	43/185/8.6588E−8/1.6537E−1	40/123/6.8253E−9/1.6047E−1	39/120/4.3654E−9/1.4481E−1
	12,000	42/96/6.8993E−9/1.806E−1	43/186/6.7466E−9/1.5463E−1	42/139/6.7574E−9/1.8889E−1	42/127/5.8553E−9/1.6617E−1
	20,000	46/112/6.9506E−9/2.9899E−1	53/195/6.9307E−9/2.03582−1	47/145/6.7908E−9/4.7286E−1	45/139/6.1085E−9/3.6828E−1
	50,000	53/137/9.0249E−8/5.5128E−1	58/204/6.1284E−9/6.1639E−1	56/157/8.5275E−9/8.0077E−1	56/151/7.0083E−9/7.6994E−1
	60,000	55/146/7.9419E−9/8.9897E−1	59/220/8.4274E−9/8.5912E−1	57/168/9.6814E−9/8.3438E−1	56/166/8.8087E−9/8.5238E−1
	100,000	66/203/8.5929E−9/1.0168E+0	74/242/7.9908E−9/1.0008E+0	69/214/8.7106E−9/1.047E+0	73/211/7.8997E−9/1.0183E+0
10	1000	31/84/6.7887E−9/5.3011E−2	39/120/6.8185E−9/3.2375E−2	39/120/6.8184E−9/4.9791E−2	37/112/6.0424E−9/3.3661E−2
	3000	32/91/6.8185E−9/4.0807E−2	41/124/7.1982E−9/3.7025E−2	40/123/6.9058E−9/8.1673E−2	38/117/6.0885E−9/6.0898E−2
	5000	34/98/8.4897E−9/6.7185E−2	42/127/6.4897E−9/5.0688E−2	41/126/8.5258E−9/3.2518E−1	39/121/1.8205E−9/2.8856E−1
	6000	34/102/5.1906E−9/8.8093E−2	42/129/5.1156E−9/6.3125E−2	41/126/8.6835E−9/4.5537E−1	40/123/1.5825E−9/3.8677E−1
	10,000	37/118/8.6826E−9/1.4077E−1	43/130/5.0362E−9/1.0055E−1	42/129/8.5262E−9/5.0619E−1	42/125/2.6473E−9/4.6409E−1
	12,000	40/120/6.0162E−9/1.7186E−1	43/132/5.1157E−9/1.3838E−1	42/130/6.2721E−9/6.1254E−1	42/126/3.5262E−9/5.2537E−1
	20,000	46/139/5.9398E−9/4.0125E−1	51/164/6.8235E−9/2.7099E−1	50/149/6.7894E−9/9.4011E−1	53/139/5.8502E−9/8.8816E−1
	50,000	74/188/5.2798E−9/9.0155E−1	84/216/6.8873E−9/7.9928E−1	87/222/5.3291E−9/1.0004E+0	78/197/6.0074E−9/8.9038E−1
	60,000	76/197/7.0307E−9/9.9906E−1	87/220/6.3948E−9/8.8975E−1	87/226/5.3563E−9/1.1664E+0	88/204/5.5493E−9/1.1107E+0
	100,000	84/225/6.3793E−9/1.0626E+0	97/244/6.4077E−9/1.0037E+0	99/255/5.0979E−9/1.4198E+0	94/222/4.9906E−9/1.2684E+0

SCG-MBFGS is superior or competitive to the other three methods in solving the given problems in terms of number of iteration, number of function evaluations and CPU time. Furthermore, to directly illustrate the performances of our method, we employed the tool provided by Dolan and Moré [50] to analyze and compare the efficiency of the method in terms of the number of iterations, number of function evaluations and CPU time. **Figures 1-3** represent the computational performance profiles of the above algorithms regarding the number of iterations, number of function evaluations and CPU time respectively. From the 3 figures, we can observe that for the given test problems, **SCG-MBFGS** is competitive or superior to other three methods in terms of number of iteration, function evaluations and CPU time respectively.

From **Figure 1** and **Figure 2**, we also notice that **SCG-MBFGS** performs better than the other methods do in terms of the numbers of iterations and function evaluations. **Figure 3** indicates that **MHS** is comparable to **SCG-MBFGS** in terms of CPU time, and since the search direction of **MHS** is developed with only first order information while **SCG-MBFGS**, **MPRP** and **MSBFGS-CG** are with second order information, it is reasonable to need less CPU time for **MHS**.

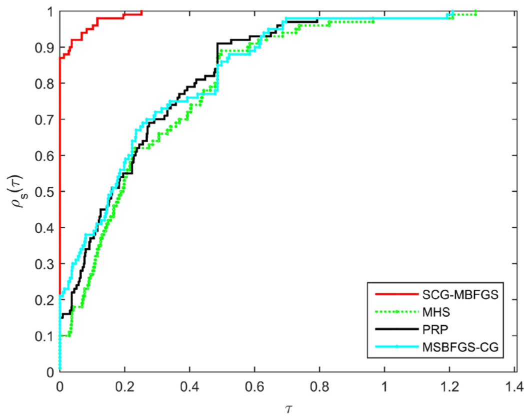

Figure 1. Performance profiles of these three methods based on NI.

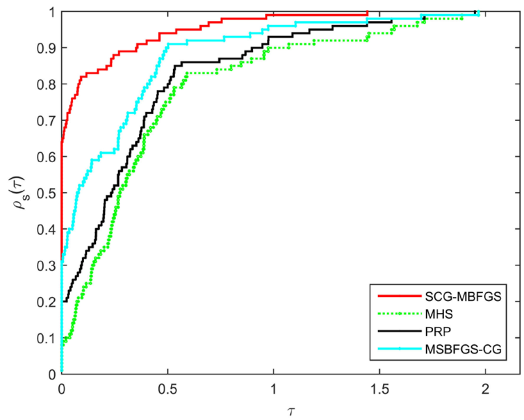

Figure 2. Performance profiles of these three methods based on NF.

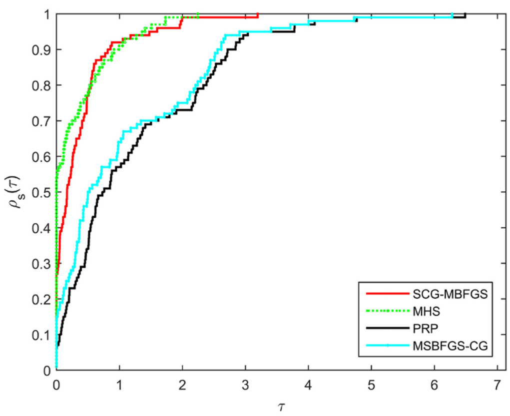

Figure 3. Performance profiles of these three methods based on CPUTIME.

5. Conclusion

In this paper, we propose a new scaled conjugate gradient method which incorporates a modified secant equation of BFGS method. This modified secant equation contains both function value and gradient information of the objective function, and its Hessian approximation update generates positive definite matrix. Under a modified nonmonotone line search and some mild conditions, the strong global convergence of the proposed method is analyzed for nonsmooth convex problems. The search direction of our new method generates sufficiently descent condition and belongs to a trust region. Compared with existing nonsmooth CG methods, the search direction of our approach is more descent direction. Numerical results and related comparisons show that the proposed method is effective for solving large scale nonsmooth optimization problems.

Acknowledgements

The authors would like to thank the reviewers and editor for their valuable comments which greatly improve our paper. This work is supported by the National Natural Science Foundation of China [Grant No. 11771003].

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Broyden, C.G. (1970) The Convergence of a Class of Double-Rank Minimization Algorithms: I. General Considerations. *Journal of the Institute of Mathematics and Its Applications*, **6**, 76-90. https://doi.org/10.1093/imamat/6.1.76

[2] Fletcher, R. (1970) A New Approach to Variable Metric Algorithms. *The Computer Journal*, **13**, 317-322. https://doi.org/10.1093/comjnl/13.3.317

[3] Goldfarb, D. (1970) A Family of Variable Metric Methods Derived by Variation Mean. *Mathematics of Computation*, **23**, 23-26. https://doi.org/10.1090/S0025-5718-1970-0258249-6

[4] Shanno, D.F. (1970) Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, **24**, 647-656. https://doi.org/10.1090/S0025-5718-1970-0274029-X

[5] Hestenes, M.R. and Stiefel, E. (1952) Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, **49**, 409-436. https://doi.org/10.6028/jres.049.044

[6] Fletcher, R. and Reeves, C.M. (1964) Function Minimization by Conjugate Gradients. *The Computer Journal*, **7**, 149-154. https://doi.org/10.1093/comjnl/7.2.149

[7] Polak, H. (1969) The Conjugate Gradient Method in Extreme Problems. *USSR Computational Mathematics and Mathematical Physics*, **9**, 94-112. https://doi.org/10.1016/0041-5553(69)90035-4

[8] Dai, Y.H. and Yuan, Y. (1999) A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property. *SIAM Journal on Optimization*, **10**, 177-182. https://doi.org/10.1137/S1052623497318992

[9] Hager, W.W. and Zhang, H. (2006) A Survey of Nonlinear Conjugate Gradient Methods. *Pacific Journal of Optimization*, **2**, 35-58.

[10] Sun, W.Y. and Yuan, Y.X. (2006) Optimization Theory and Methods: Nonlinear Programming. Springer, New York.

[11] Nocedal, J. (1992) Theory of Algorithms for Unconstrained Optimization. *Acta Numerica*, **1**, 199-242. https://doi.org/10.1017/S0962492900002270

[12] Dennis, J.E. and Moré, J.J. (1974) A Characterization of Superlinear Convergence and Its Application to Quasi-Newton Methods. *Mathematics of Computation*, **28**, 549-560. https://doi.org/10.1090/S0025-5718-1974-0343581-1

[13] Byrd, R., Nocedal, J. and Yuan, Y. (1987) Global Convergence of a Class of Quasi-Newton Methods on Convex Problems. *SIAM Journal on Numerical Analysis*, **24**, 1171-1189. https://doi.org/10.1137/0724077

[14] Byrd, R. and Nocedal, J. (1989) A Tool for the Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization. *SIAM Journal on Numerical Analysis*, **26**, 727-739. https://doi.org/10.1137/0726042

[15] Griewank, A. (1991) The Global Convergence of Partitioned BFGS on Problems with Convex Decompositions and Lipschitzian Gradients. *Mathematical Programming*, **50**, 141-175. https://doi.org/10.1007/BF01594933

[16] Mascarenhas, W.F. (2004) The BFGS Method with Exact Line Searches Fails for Non-Convex Objective Functions. *Mathematical Programming*, **99**, 49-61. https://doi.org/10.1007/s10107-003-0421-7

[17] Dai, Y.H. (2002) Convergence Properties of the BFGS Algorithm. *SIAM Journal on Optimization*, **13**, 693-701. https://doi.org/10.1137/S1052623401383455

[18] Wolfe, P. (1969) Convergence Conditions for Ascent Methods. *SIAM Review*, **11**, 226-235. https://doi.org/10.1137/1011036

[19] Grippo, L., Lampariello, F. and Lucidi, S. (1986) A Nonmonotone Line Search Technique for Newton's Method. *SIAM Journal on Scientific Computing*, **23**, 707-716. https://doi.org/10.1137/0723046

[20] Toint, P.L. (1996) An Assessment of Nonmonotone Line Search Technique for Unconstrained Optimization. *SIAM Journal on Scientific Computing*, **17**, 725-739. https://doi.org/10.1137/S106482759427021X

[21] Toint, P.L. (1997) A Nonmonotone Trust Region Algorithm for Nonlinear Programming Subject to Convex Constraints. *Mathematical Programming*, **77**, 69-94. https://doi.org/10.1007/BF02614518

[22] Panier, E.R. and Tits, A.L. (1991) Avoiding Maratos Effect by Means of Nonmonotone Line Search Constrained Problems. *SIAM Journal on Numerical Analysis*, **28**, 1183-1190. https://doi.org/10.1137/0728063

[23] Yu, Z. and Pu, D. (2008) A New Nonmonotone Line Search Technique for Unconstrained Optimization. *Journal of Computational and Applied Mathematics*, **219**, 134-144. https://doi.org/10.1016/j.cam.2007.07.008

[24] Yuan, G., Wei, Z. and Wu, Y. (2010) Modified Limited Memory BFGS Method with Nonmonotone Line Search for Unconstrained Optimization Problems. *Journal of the Korean Mathematical Society*, **47**, 767-788. https://doi.org/10.4134/JKMS.2010.47.4.767

[25] Li, X., Wang, B. and Hu, W. (2017) A Modified Nonmonotone BFGS Algorithm for Unconstrained Optimization. *Journal of Inequalities and Applications*, **183**, 1-18. https://doi.org/10.1186/s13660-017-1453-5

[26] Su, K. and Rong, Z. (2015) A Spectral Conjugate Gradient Method under Modified

Nonmonotone Line Search Technique. *Mathematica Aeterna*, **5**, 537-549.

[27] Andrei, A. (2007) Scaled Conjugate Gradient Algorithms for Unconstrained Optimization. *Computational Optimization and Applications*, **38**, 401-416. https://doi.org/10.1007/s10589-007-9055-7

[28] Babaie-Kafaki, S. (2013) A Modified Scaled Memoryless BFGS Preconditioned Conjugate Gradient Method for Unconstrained Optimization. *4OR*, **11**, 361-374. https://doi.org/10.1007/s10288-013-0233-4

[29] Babaie-Kafaki, S. and Chanbari, R. (2017) A Class of Descent Four-Term Extension of the Dai-Liao Conjugate Gradient Method Based on the Scaled Memoryless BFGS Update. *Journal of Industrial & Management Optimization*, **13**, 649-658. https://doi.org/10.3934/jimo.2016038

[30] Yuan, G., Wei, Z. and Li, G. (2014) A Modified Polak-Ribière-Polyak Conjugate Gradient Algorithm for Nonsmooth Convex Programs. *Journal of Computational and Applied Mathematics*, **255**, 86-96. https://doi.org/10.1016/j.cam.2013.04.032

[31] Yuan, G., Wei, Z. and Li, Y. (2015) A Modified Hestenes and Stiefel Conjugate Gradient Algorithm for Large-Scale Nonsmooth Minimizations and Nonlinear Equations. *Journal of Optimization Theory and Applications*, **168**, 129-152. https://doi.org/10.1007/s10957-015-0781-1

[32] Yuan, G. and Wei, Z. (2015) A Modified PRP Conjugate Gradient Algorithm with Nonmonotone Line Search for Nonsmooth Convex Optimization Problems. *Journal of Applied Mathematics and Computing*, **51**, 397-412. https://doi.org/10.1007/s12190-015-0912-8

[33] Yuan, G., Sheng, Z. and Liu, W. (2016) The Modified HZ Conjugate Gradient Algorithm for Large-Scale Nonsmooth Optimization. *PLoS ONE*, **11**, e0164289. https://doi.org/10.1371/journal.pone.0164289

[34] Yuan, G. and Wei, Z. (2012) The Barzilai and Browein Gradient Method with Nonmonotone Line Search for Nonsmooth Convex Optimization Problems. *Mathematical Modelling and Analysis*, **17**, 203-216. https://doi.org/10.3846/13926292.2012.661375

[35] Cui, Z., Yuan, G., Sheng, Z., Liu, W., Wang, X. and Duan, X. (2015) A Modified BFGS Formula Using a Trust Region Model for Nonsmooth Convex Minimizations. *PLoS ONE*, **10**, e0140606. https://doi.org/10.1371/journal.pone.0140606

[36] Burke, J.V. and Qian, M. (2000) On the Superlinear Convergence of the Variable Metric Proximal Point Algorithm Using Broyden and BFGS Matrix Secant Updating. *Mathematical Programming*, **88**, 157-181. https://doi.org/10.1007/PL00011373

[37] Chen, X. and Fukushima, M. (1999) Proximal Quasi-Newton Methods for Nondifferentiale Convex Optimization. *Mathematical Programming*, **85**, 313-334. https://doi.org/10.1007/s101070050059

[38] Sagara, N. and Fukushima, M. (2005) A Trust Region Method for Nonsmooth Convex Optimization. *Journal of Industrial & Management Optimization*, **1**, 171-180. https://doi.org/10.3934/jimo.2005.1.171

[39] Ou, Y. and Zhou, X. (2018) A Modified Scaled Memoryless BFGS Preconditioned Conjugate Gradient Algorithm for Nonsmooth Convex Optimization. *Journal of Industrial & Management Optimization*, **14**, 785-801. https://doi.org/10.3934/jimo.2017075

[40] Hiriart-Urruty, J.B. and Lemaréchal, C. (1993) Convex Analysis and Minimization Algorithms. Springer, Berlin. https://doi.org/10.1007/978-3-662-02796-7

[41] Fukushima, M. and Qi, L. (1996) A Global and Superlinearly Convergent Algorithm for Nonsmooth Convex Minimization. *SIAM Journal on Optimization*, **6**, 1106-1120.

https://doi.org/10.1137/S1052623494278839

[42] Qi, L. and Sun, J. (1993) A Nonsmooth Version of Newton's Method. *Mathematical Programming*, **58**, 353-367. https://doi.org/10.1007/BF01581275

[43] Miffin, R. (1996) A Quasi-Second-Order Proximal Bundle Algorithm. *Mathematical Programming*, **73**, 51-72. https://doi.org/10.1007/BF02592098

[44] Bonnans, J.F., Gilbert, J.C., Lemaréchal, C. and Sagastizábal, C. (1995) A Family of Variable-Metric Proximal Methods. *Mathematical Programming*, **68**, 15-47. https://doi.org/10.1007/BF01585756

[45] Lemaréchal, C. and Sagastizábal, C. (1997) Practical Aspects of the Moreu-Yosida Regularization: Theoretical Preliminaries. *SIAM Journal on Optimization*, **7**, 367-385. https://doi.org/10.1137/S1052623494267127

[46] Rauf, A.I. and Fukushima, M. (2000) A Globally Convergent BFGS Method for Nonsmooth Convex Optimization. *Journal of Optimization Theory and Applications*, **104**, 539-558. https://doi.org/10.1023/A:1004633524446

[47] Conn, A.R., Gould, N.I.M. and Toint, P.L. (2000) Trust Region Methods. SIAM, Philadelphia. https://doi.org/10.1137/1.9780898719857

[48] Li, D.H. and Fukushima, M. (1999) On the Global Convergence of BFGS Method for Nonconvex Unconstrained Optimization Problems. *SIAM Journal on Optimization*, **11**, 1054-1064. https://doi.org/10.1137/S1052623499354242

[49] Haarala, M. and Mäkelä, M.M. (2004) New Limited Memory Bundle Method for Large-Scale Nonsmooth Optimization. *Optimization Methods and Software*, **19**, 673-692. https://doi.org/10.1080/10556780410001689225

[50] Dolan, E.D. and Moré, J.J. (2002) Benchmarking Optimization Software with Performance Profiles. *Mathematical Programming*, **91**, 201-213. https://doi.org/10.1007/s101070100263