

The Barzilai and Borwein Gradient Method with Nonmonotone Line Search for Nonsmooth Convex Optimization Problems

Gonglin Yuan and Zengxin Wei

Guangxi University

530004 Nanning, Guangxi, China

E-mail(*corresp.*): glyuan@gxu.edu.cn

E-mail: zxwei@gxu.edu.cn

Received April 8, 2011; revised January 15, 2012; published online April 1, 2012

Abstract. The Barzilai and Borwein gradient algorithm has received a great deal of attention in recent decades since it is simple and effective for smooth optimization problems. Whether can it be extended to solve nonsmooth problems? In this paper, we answer this question positively. The Barzilai and Borwein gradient algorithm combined with a nonmonotone line search technique is proposed for nonsmooth convex minimization. The global convergence of the given algorithm is established under suitable conditions. Numerical results show that this method is efficient.

Keywords: nonsmooth convex minimization, spectral gradient, nonmonotone technique, global convergence.

AMS Subject Classification: 90C26.

1 Introduction

Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x), \quad (1.1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a possibly nonsmooth convex function. The following problem

$$\min_{x \in \mathbb{R}^n} F(x) \quad (1.2)$$

is the so-called Moreau–Yosida regularization of f , where $F(x) = \min_{z \in \mathbb{R}^n} \{f(z) + \frac{1}{2\lambda} \|z - x\|^2\}$, λ is a positive parameter, and $\|\cdot\|$ denotes the Euclidean norm. It is well known that problems (1.1) and (1.2) are equivalent in the sense that the solution sets of the two problems coincide are the same. Now we review some methods for nonsmooth optimization problems. The classical proximal point algorithm (see [37]) is regarded as a gradient method for solving problem (1.2). The gradient function of F can be proved to be semi-smooth under some reasonable conditions [22, 41]. Based on these features,

many algorithms were proposed for (1.2) (see [3, 22, 41]). The proximal methods have been proved to be effective in dealing with the difficulty of evaluating the function value of $F(x)$ and its gradient $\nabla F(x)$ at a given point x (see [2, 6, 9, 31, 48, 49]). Lemaréchal [32] and Wolfe [50] initiated a giant stride forward in nonsmooth optimization by the bundle concept, which can handle convex and nonconvex f . All bundle methods carry two distinctive features (see Lemaréchal [33] and Zowe [60] in detail). Kiwiel [30] proposed a bundle variant, which is close to bundle trust iteration method (see [45]). Some similar results can be found in [29, 30, 36]. In the past decades, many trust region algorithms to minimize nonsmooth objective function have been presented (see [7, 8, 18, 23, 28, 19, 39, 40, 57]).

The spectral gradient method (also named the two-point stepsize method) was originated in [1] for unconstrained optimization problems. This method consists essentially of a steepest descent method, where the choice of the stepsize along the antigradient direction is potentially derived from a two-point approximation to the secant equation underlying the quasi-Newton method [25]. If the objective function is a strictly convex quadratic, Raydan [42] proved that the two-point stepsize gradient method is globally convergent. For the nonquadratic case, Raydan [43] incorporated a globalization scheme of the two-point stepsize gradient method by using the technique of nonmonotone line search. Dai et al. extended this method to box-constrained quadratic programming [11] and unsymmetric linear equations [16], respectively. Some authors use the method for solving constrained optimization problems (see [38, 55]) and nonlinear equations (see [54, 58]). The effectiveness of the classical spectral gradient method has been significantly improved by incorporating it with new and fast nonmonotone line search techniques (e.g. [15]). The spectral gradient method does not guarantee a descent in the objective function at each iteration, but performs better than the classical steep descent (SD) method in practice. An interesting fact is that an alternating strategy that uses the SD step and spectral gradient step alternately can accelerate the convergent rate of the spectral gradient method. An important work on this scheme is due to the cycle Barzilai–Borwein (CBB) method [10], see also [12]. An implementation of the CBB method, combined with a nonmonotone line search, shows that this method performs better than the existing spectral gradient method. It is even competitive to some other well-known standard codes (see [12]). Due to its simplicity and numerical efficiency, the spectral gradient method has received a great deal of attention in recent decades (see [4, 13, 14, 21, 27, 51, 52]). However, the spectral gradient algorithms are only used to solve smooth optimization problems.

The first nonmonotone line search framework was developed by Grippo, Lampariello, and Lucidi in [24] for Newton's methods. Many subsequent papers have exploited nonmonotone line search techniques of this nature (see [5, 26, 34, 59]). There are some spectral gradient methods with nonmonotone line search technique for optimization problems (see [47, 53]). Although these nonmonotone technique work well in many cases, there are some drawbacks. First, a good function value generated in any iteration is essentially discarded due to the max in the nonmonotone line search technique. Second, in some

cases, the numerical performance is very dependent on the choice of M , where $M > 0$ is integer (see [24, 46]). In order to overcome these two drawbacks, Dai and Zhang [17] proposed an adaptive nonmonotone line search which is combined with the two-point gradient method for optimization problems. Moreover, Zhang and Hager [56] presented a new nonmonotone line search technique. Numerical results show that the new nonmonotone technique is better than the normal nonmonotone technique and the monotone technique.

It is well known that the trust region methods, the Newton and quasi-Newton methods, and the proximal gradient methods which were firstly used to solve smooth optimization problems are widely used in nonsmooth fields. The question is whether the spectral gradient method can be extended to nonsmooth problems. In this paper, we answer this question positively. Motivated by the above observations, we present a spectral gradient method which combines with a nonmonotone line search technique for nonsmooth optimization problems. The main attributes of this presented method are stated as follows.

- A spectral gradient method is introduced for nonsmooth problems (1.1) and (1.2).
- All search directions are sufficiently descent, which shows that the functions are decreasing. All search directions belong to a trust region, which hints that this method has a good convergent result.
- This method possesses the global convergence.
- Numerical results show that this method is more effective than the standard method.

This paper is organized as follows. In the next section, we briefly review some basic results about the objective function of (1.2). In Section 3, the new algorithm is stated. In Section 4, we prove the global convergence of the proposed method. Numerical results are reported in Section 5. Throughout this paper, without specification, $\|\cdot\|$ denotes the Euclidean norm of vectors or matrices.

2 Results of Convex Analysis and Nonsmooth Analysis

Some basic results in convex analysis and nonsmooth analysis, which will be used later, are reviewed in this section. Let

$$\theta(z) = f(z) + \frac{1}{2\lambda} \|z - x\|^2$$

and denote $p(x) = \operatorname{argmin} \theta(z)$. Then $p(x)$ is well-defined and unique since $\theta(z)$ is strongly convex. By the definition of $F(x)$, we have

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

In what follows, we denote the gradient of F by g . Some features about $F(x)$ can be found in [6, 9].

(i) The function F is finite-valued, convex, and everywhere differentiable with

$$g(x) = \nabla F(x) = (x - p(x))/\lambda.$$

Moreover, the gradient mapping $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is globally Lipschitz continuous with modulus λ , i.e.,

$$\|g(x) - g(y)\| \leq \|x - y\|/\lambda, \quad \forall x, y \in \mathbb{R}^n.$$

(ii) x is an optimal solution to (1.1) if and only if $\nabla F(x) = 0$, namely, $p(x) = x$.

(iii) By the Rademacher theorem and the Lipschitzian property of ∇F , for each $x \in \mathbb{R}^n$, we conclude that the set of generalized Jacobian matrices

$$\partial_B g(x) = \left\{ V \in \mathbb{R}^{n \times n} : V = \lim_{x_k \rightarrow x} \nabla g(x_k), x_k \in D_G \right\}$$

is nonempty and compact, where $D_g = \{x \in \mathbb{R}^n : g \text{ is differentiable at } x\}$. Since g is a gradient mapping of the convex function F , for each $x \in \mathbb{R}^n$, we deduce that every $V \in \partial_B g(x)$ is a symmetric positive semidefinite matrix.

(iv) If g is BD-regular at x , namely, all matrices $V \in \partial_B g(x)$ are nonsingular, then there exist constants $\mu_1 > 0$, $\mu_2 > 0$ and a neighborhood Ω of x such that

$$d^T V d \geq \mu_1 \|d\|^2, \quad \|V^{-1}\| \leq \mu_2, \quad \forall d \in \mathbb{R}^n, \quad V \in \partial_B g(x).$$

It is obviously that $F(x)$ and $g(x)$ can be obtained through the optimal solution of $\operatorname{argmin}_{z \in \mathbb{R}^n} \theta(z)$. However, $p(x)$ is difficult or even impossible to solve exactly. Thus we can not apply the exact value of $p(x)$ to define $F(x)$ and $g(x)$. Fortunately, for each $x \in \mathbb{R}^n$ and any $\varepsilon > 0$, there exists a vector $p^\alpha(x, \varepsilon) \in \mathbb{R}^n$ satisfying

$$f(p^\alpha(x, \varepsilon)) + \frac{1}{2\lambda} \|p^\alpha(x, \varepsilon) - x\|^2 \leq F(x) + \varepsilon. \tag{2.1}$$

Thus, we can use $p^\alpha(x, \varepsilon)$ to define approximations of $F(x)$ and $g(x)$ by

$$F^\alpha(x, \varepsilon) = f(p^\alpha(x, \varepsilon)) + \frac{1}{2\lambda} \|p^\alpha(x, \varepsilon) - x\|^2, \tag{2.2}$$

$$g^\alpha(x, \varepsilon) = (x - p^\alpha(x, \varepsilon))/\lambda, \tag{2.3}$$

respectively. Some implementable algorithms to find $p^\alpha(x, \varepsilon)$ for a nondifferentiable convex function are introduced in [8]. A remarkable feature of $F^\alpha(x, \varepsilon)$ and $g^\alpha(x, \varepsilon)$ is given as follows [22].

Proposition 1. *Let $p^\alpha(x, \varepsilon)$ be a vector satisfying (2.1). Suppose that $F^\alpha(x, \varepsilon)$ and $g^\alpha(x, \varepsilon)$ are defined by (2.2) and (2.3), respectively. Then we get*

$$F(x) \leq F^\alpha(x, \varepsilon) \leq F(x) + \varepsilon, \tag{2.4}$$

$$\|p^\alpha(x, \varepsilon) - p(x)\| \leq \sqrt{2\lambda\varepsilon}, \tag{2.5}$$

$$\|g^\alpha(x, \varepsilon) - g(x)\| \leq \sqrt{2\varepsilon/\lambda}. \tag{2.6}$$

The above proposition says that we can compute approximately $F^\alpha(x, \varepsilon)$ and $g^\alpha(x, \varepsilon)$, by choosing parameter ε small enough, which may be arbitrarily close to $F(x)$ and $g(x)$, respectively.

3 Algorithm

The following iterative formula is used by spectral gradient method

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 1, 2, \dots$$

where x_k is the current iterate point, $d_k = -g^\alpha(x_k, \varepsilon_k)$ is the search direction, and two choices of the scalar α_k are

$$\alpha_k^1 = s_k^T s_k / s_k^T y_k \quad \text{and} \quad \alpha_k^2 = s_k^T y_k / y_k^T y_k,$$

where $s_k = x_k - x_{k-1}$ and $y_k = g^\alpha(x_k, \varepsilon_k) - g^\alpha(x_{k-1}, \varepsilon_{k-1})$. These two formulas are motivated by [1].

Algorithm 1. Nonmonotone Spectral Gradient Algorithm.

Step 0. Initialization. Given $x_0 \in \mathbb{R}^n$, $\sigma \in (0, 1)$, $s > 0$, $\lambda > 0$, $\rho \in [0, 1]$, $E_0 = 1$, $\varepsilon_0 = 1$, $J_0 = F^\alpha(x_0, \varepsilon_0)$, $d_0 = -g^\alpha(x_0, \varepsilon_0)$, and $\epsilon \in (0, 1)$. Let $k = 0$.

Step 1. Termination Criterion. Stop if x_k satisfies termination condition $\|g^\alpha(x_k, \varepsilon_k)\| < \epsilon$. Otherwise go to the next step.

Step 2: Choose a scalar ε_{k+1} such that $0 < \varepsilon_{k+1} < \varepsilon_k$, and compute step size α_k by the following nonmonotone line search rule

$$F^\alpha(x_k + \alpha_k d_k, \varepsilon_{k+1}) - J_k \leq \sigma \alpha_k g^\alpha(x_k, \varepsilon_k)^T d_k, \tag{3.1}$$

where $\alpha_k = \max\{s, \alpha_k^1\} \times 2^{-i_k}$ (or $\alpha_k = \max\{s, \alpha_k^2\} \times 2^{-i_k}$), $i_k \in \{0, 1, 2, \dots\}$.

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k$. If $\|g^\alpha(x_{k+1}, \varepsilon_{k+1})\| < \epsilon$, then stop.

Step 4: Update J_k by

$$E_{k+1} = \rho E_k + 1, \quad J_{k+1} = (\rho E_k J_k + F^\alpha(x_k + \alpha_k d_k, \varepsilon_{k+1})) / E_{k+1}.$$

Step 5: Calculate the search direction by $d_{k+1} = -g^\alpha(x_{k+1}, \varepsilon_{k+1})$.

Step 6: Set $k := k + 1$, and go to Step 2.

Remark. It is not difficult to see that J_{k+1} is a convex combination of J_k and $F^\alpha(x_{k+1}, \varepsilon_{k+1})$. Since $J_0 = F^\alpha(x_0, \varepsilon_0)$, it follows that J_k is a convex combination of the function values $F^\alpha(x_0, \varepsilon_0), F^\alpha(x_1, \varepsilon_1), \dots, F^\alpha(x_k, \varepsilon_k)$. The choice of ρ controls the degree of nonmonotonicity. If $\rho = 0$, the line search is the usual monotone Armijo line search. If $\rho = 1$, $J_k = A_k$, where $A_k = \frac{1}{k+1} \sum_{i=0}^k F^\alpha(x_i, \varepsilon_i)$ is the average function value (these cases have been analyzed by Yu-Hong Dai or see [56]).

4 Properties and Global Convergence

In this section, we turn to the behavior of Algorithm 1 when it is applied to problem (1.1). In order to establish the global convergence result, the following assumptions are needed.

ASSUMPTION A. (i) The sequence $\{V_k\}$ is bounded, i.e., there exists a positive constant M such that

$$\|V_k\| \leq M, \quad \forall k, \tag{4.1}$$

where the matrix $V_k \in \partial_B g(x_k)$.

(ii) F is bounded from below.

(iii) For sufficiently large k , ε_k converges to zero.

From the definition of $d_k = -g^\alpha(x_k, \varepsilon_k)$, we can get

$$g^\alpha(x_k, \varepsilon_k)^T d_k = -\|g^\alpha(x_k, \varepsilon_k)\|^2, \tag{4.2}$$

$$\|d_k\| = \|g^\alpha(x_k, \varepsilon_k)\|. \tag{4.3}$$

The above two relations (4.2) and (4.3) show that the search direction possesses the sufficiently descent property and belongs to a trust region. Based on (4.2) and (4.3), similar to Lemma 1.1 in [56], it is not difficult to get the following lemma. So we only state it as follows but omit the proof.

Lemma 1. *Let Assumption A hold and the sequence $\{x_k\}$ be generated by Algorithm 1. Then, for each k , we have $F^\alpha(x_k, \varepsilon_k) \leq J_k \leq A_k$. Moreover, there exists α_k satisfying Armijo conditions of the line search update.*

The above lemma shows that Algorithm 1 is well-defined.

Lemma 2. *Let Assumption A hold and the sequence $\{x_k\}$ be generated by Algorithm 1. Suppose that $\varepsilon_k = o(\alpha_k^2 \|d_k\|^2)$ holds. Then, for sufficiently large k , there exists a constant $m > 0$ such that*

$$\alpha_k \geq m. \tag{4.4}$$

Proof. From Lemma 1, we conclude that there exists a α_k satisfying (3.1). If $\alpha_k \geq 1$, the proof is complete. Otherwise let $\alpha'_k = \alpha_k/2$, we deduce that the following relation

$$F^\alpha(x_k + \alpha'_k d_k, \varepsilon_{k+1}) - J_k > \sigma \alpha'_k g^\alpha(x_k, \varepsilon_k)^T d_k$$

holds. By $F^\alpha(x_k, \varepsilon_k) \leq J_k \leq A_k$ of Lemma 1, we get

$$\begin{aligned} F^\alpha(x_k + \alpha'_k d_k, \varepsilon_{k+1}) - F^\alpha(x_k, \varepsilon_k) &\geq F^\alpha(x_k + \alpha'_k d_k, \varepsilon_{k+1}) - J_k \\ &> \sigma \alpha'_k g^\alpha(x_k, \varepsilon_k)^T d_k. \end{aligned} \tag{4.5}$$

By (2.4), (4.5), and Taylor's formula, we have

$$\begin{aligned} \sigma \alpha'_k g^\alpha(x_k, \varepsilon_k)^T d_k &< F^\alpha(x_k + \alpha'_k d_k, \varepsilon_{k+1}) - F^\alpha(x_k, \varepsilon_k) \\ &\leq F(x_k + \alpha'_k d_k) - F(x_k) + \varepsilon_{k+1} \\ &= \alpha'_k d_k^T g(x_k) + \frac{1}{2} (\alpha'_k)^2 d_k^T V(\xi_k) d_k + \varepsilon_{k+1} \\ &\leq \alpha'_k d_k^T g(x_k) + \frac{M}{2} (\alpha'_k)^2 \|d_k\|^2 + \varepsilon_{k+1}, \end{aligned} \tag{4.6}$$

where $\xi_k = x_k + \theta \alpha'_k d_k$, $\theta \in (0, 1)$, and the last inequality follows (4.1). It follows that from (4.6)

$$\begin{aligned} \alpha'_k &> \left[\frac{(g^\alpha(x_k, \varepsilon_k) - g(x_k))^T d_k - (1 - \sigma) g^\alpha(x_k, \varepsilon_k)^T d_k - \varepsilon_{k+1} / (\alpha'_k)^2}{\|d_k\|^2} \right] \frac{2}{M} \\ &\geq \left[\frac{(1 - \sigma) \|g^\alpha(x_k, \varepsilon_k)\|^2 - \sqrt{2\varepsilon_k/\lambda} \|d_k\| - \varepsilon_k}{\|d_k\|^2} \right] \frac{2}{M} \\ &= [(1 - \sigma) - o(\alpha_k) / \sqrt{\lambda} - o(1)] \frac{2}{M} \geq \frac{1 - \sigma}{2M}, \end{aligned} \tag{4.7}$$

where the second inequality follows (2.6), (4.2) and $\varepsilon_{k+1} \leq \varepsilon_k$, and the equality follows $\varepsilon_k = o(\alpha_k^2 \|d_k\|^2)$ and (4.3). Therefore, we obtain $\alpha_k \geq (1 - \sigma)/M$. Let $m \in (0, (1 - \sigma)/M]$, we complete the proof. \square

Theorem 1. *Let the conditions in Lemma 2 hold. Then $\lim_{k \rightarrow \infty} \|g(x_k)\| = 0$ holds and any accumulation point of x_k is an optimal solution of (1.1).*

Proof. In order to get the results of this theorem, we first show that

$$\lim_{k \rightarrow \infty} \|g^\alpha(x_k, \varepsilon_k)\| = 0 \tag{4.8}$$

holds. Suppose that (4.8) is not true. Then there exist constants $\epsilon_1 > 0$ and $k_1 > 0$ satisfying

$$\|g^\alpha(x_k, \varepsilon_k)\| \geq \epsilon_1, \quad \forall k > k_1. \tag{4.9}$$

By (3.1), (4.2), (4.4), and (4.9), we have

$$\begin{aligned} F^\alpha(x_{k+1}, \varepsilon_{k+1}) - J_k &\leq \sigma \alpha_k g^\alpha(x_k, \varepsilon_k)^T d_k = -\sigma \alpha_k \|g^\alpha(x_k, \varepsilon_k)\|^2 \\ &\leq -\sigma m \epsilon_1, \quad \forall k > k_1. \end{aligned}$$

By the definition of J_{k+1} , we obtain

$$\begin{aligned} J_{k+1} &= \frac{\rho E_k J_k + F^\alpha(x_k + \alpha_k d_k, \varepsilon_{k+1})}{E_{k+1}} \\ &\leq \frac{\rho E_k J_k + J_k - \sigma m \epsilon_1}{E_{k+1}} = J_k - \frac{\sigma m \epsilon_1}{E_{k+1}}. \end{aligned} \tag{4.10}$$

Since $F^\alpha(x, \varepsilon)$ is bounded from below and $F^\alpha(x_k, \varepsilon_k) \leq J_k$ for all k , we conclude that J_k is bounded from below. From (4.10), we have

$$\sum_{k=k_0}^{\infty} \frac{\sigma m \epsilon_1}{E_{k+1}} < \infty. \tag{4.11}$$

By the definition of E_{k+1} , we have $E_{k+1} \leq k + 2$. Then the relation (4.11) contradicts to this case. So (4.8) holds. Using (2.6), we get

$$\|g^\alpha(x_k, \varepsilon_k) - g(x_k)\| \leq \sqrt{\frac{2\varepsilon_k}{\lambda}}.$$

Which in view of Assumption A(iii), yields

$$\lim_{k \rightarrow \infty} \|g(x_k)\| = 0. \tag{4.12}$$

Let x^* be an accumulation point of $\{x_k\}$, without loss of generality, there exists a subsequence $\{x_k\}_K$ such that

$$\lim_{k \in K, k \rightarrow \infty} x_k = x^*. \tag{4.13}$$

From properties of $F(x)$, we get $g(x_k) = (x_k - p(x_k))/\lambda$. By (4.12) and (4.13), we have $x^* = p(x^*)$. Therefore x^* is an optimal solution of (1.1). \square

Table 1. The test problems and global minimum values of the functions.

| Nr. | Problems | $f_{ops}(x)$ | Nr. | Problems | $f_{ops}(x)$ |
|-----|------------|--------------|-----|--------------|--------------|
| 1 | Rosenbrock | 0 | 8 | Mifflin 1 | -1.0 |
| 2 | Crescent | 0 | 9 | Mifflin 2 | -1.0 |
| 3 | CB2 | 1.9522245 | 10 | Wolfe | -8.0 |
| 4 | CB3 | 2.0 | 11 | Rosen-Suzuki | -44 |
| 5 | DEM | -3 | 12 | Shor | 22.600162 |
| 6 | QL | 7.20 | 13 | Colville | -32.348679 |
| 7 | LQ | -1.4142136 | 14 | HS78 | -2.9197004 |

5 Numerical Results

In this section, we test the numerical behavior of Algorithm 1. All the non-smooth problems with initial point of Table 1 can be found in [35]. Table 1 contains the names of the test problems and global minimum values of the functions, where $f_{ops}(x)$ is a global minimum value of a function. The algorithm is implemented by Matlab 7.6, all experiments are run on a PC with CPU Intel Pentium Dual E7500 2.93GHz, 2G bytes of SDRAM memory, and Windows XP operating system. The parameters were chosen as $s = 0.5$, $\lambda = 1$, $\rho = 0.75$, $\sigma = 0.9$, and $\varepsilon_k = 1/(NI + 2)^2$ (NI is the iteration number). The program is stopped when the condition $\|g^\alpha(x, \varepsilon)\| \leq 10^{-10}$ was satisfied. In order to show the performance of the given algorithm, we also list the recent results of paper [44] (New trust region method, BT(S-F)). For BT(S-F) method, the parameters were chosen as $\rho = 0.45$ and $\Delta = 0.5$.

The columns of Table 2 have the following meanings: No.: the name of the test problem, $f(x)$: the function evaluations at the final iteration, NI: the total number of iterations, NF: the iteration number of the function evaluations, time: CPU time in seconds, $f_{ops}(x)$: a global minimum value of a function.

Dolan and Moré [20] gave a new tool to analyze the efficiency of Algorithms. In order to show their performance, this technique will be used in this paper. From Table 2, it is easy to see that Algorithm 1 performs better than the BT(S-F) method for most of the test problems. Compared with the optimization value, the final function value is acceptable for both of these methods. Figures 1, 2, and 3 show that the performance of the iteration number, the function number, and the CPU time respectively. It is not difficult to see that the given algorithm is more competitive than the new trust region method. Overall, the preliminary numerical results indicate that the proposed method is competitive to the other method.

6 Conclusions

In this paper, we propose a spectral gradient method for nonsmooth convex minimization. The global convergence is established under suitable conditions. Numerical results show that this method is interesting. Considering the simplicity and numerical efficiency of the spectral gradient method, we propose Algorithm 1 for nonsmooth problems. The main work of this paper is to ex-

Table 2. Numerical results of the tested problems.

| Algorithm 1 (α_k^1) | | Algorithm 1 (α_k^2) |
|------------------------------|----------------------------|------------------------------|
| No. | NI/NF/ $f(x)$ /time | NI/NF/ $f(x)$ /time |
| 1 | 54/56/3.448409e-007/2.09 | 54/56/3.311278e-007/1.92 |
| 2 | 14/16/2.744977e-005/0.656 | 14/16/2.903955e-005/0.703 |
| 3 | 13/15/1.952225e+000/0.625 | 13/15/1.952225e+000/0.625 |
| 4 | 4/8/2.000008e+000/0.203 | 4/8/2.000008e+000/0.172 |
| 5 | 4/7/-2.999969e+000/0.2031 | 4/6/-2.999956e+000/0.188 |
| 6 | 22/25/7.200000e+000/1.28 | 19/22/7.200001e+000/1 |
| 7 | 6/7/-1.414214e+000/0.25 | 6/7/-1.414214e+000/0.25 |
| 8 | 3/6/-9.937901e-001/0.281 | 3/6/-9.954955e-001/0.203 |
| 9 | 12/13/-9.999222e-001/0.656 | 12/13/-9.999222e-001/0.641 |
| 10 | 9/12/-7.999999e+000/0.344 | 9/12/-7.999999e+000/0.313 |
| 11 | 8/9/-4.394932e+001/1.14 | 8/9/-4.394932e+001/1.11 |
| 12 | 9/10/2.260038e+001/1.750 | 9/10/2.260038e+001/1.52 |
| 13 | 5/6/-3.234794e+001/1 | 5/6/-3.234794e+001/0.968 |
| 14 | 7/9/-2.911583e+000/1.64 | 17/19/-2.911311e+000/3.79 |

| BT(S-F) | | |
|---------|----------------------------|--------------|
| No. | NI/NF/ $f(x)$ /time | $f_{ops}(x)$ |
| 1 | 20/39/6.729353e-003/0.812 | 0 |
| 2 | 11/21/4.412617e-002/0.719 | 0 |
| 3 | 31/61/1.952384e+000/1.48 | 1.9522245 |
| 4 | 3/5/2.000252e+000/0.109 | 2.0 |
| 5 | 4/7/-2.998072e+000/0.156 | -3 |
| 6 | 38/75/7.200323e+000/2.11 | 7.20 |
| 7 | 2/3/-1.207068e+000/0.078 | -1.4142136 |
| 8 | 2/3/5.848352e+000/0.094 | -1.0 |
| 9 | 9/17/-9.796397e-001/0.562 | -1.0 |
| 10 | 24/47/-7.956443e+000/1.05 | -8.0 |
| 11 | 8/15/-4.365725e+001/1.565 | -44 |
| 12 | 36/71/2.260088e+001/5.345 | 22.600162 |
| 13 | 17/33/-3.191816e+001/4.735 | -32.348679 |
| 14 | 39/77/-2.919162e+000/6.34 | -2.9197004 |

tend the spectral gradient method to solve nonsmooth problems. From Algorithm 1, it is easy to see that this method is not difficult to be performed. We can conclude that it may become one of the most simple and efficient methods for nonsmooth problems.

The parameters $\lambda > 0$ and $s > 0$ may influence the performance of the method, so the choice of the positive constants λ and s are our further work.

Acknowledgements

The authors would like to thank the referees for their helpful suggestions and comments, which improved this paper greatly. We also thank Guoyin Li of the University of New South Wales for his valuable suggestions on nonsmooth optimization problems. This work is supported by Program for Excellent Talents in Guangxi Higher Education Institutions, China NSF (Grant No. 11161003 and 71001015), Guangxi Education research project (Grant No. 201012MS013 and

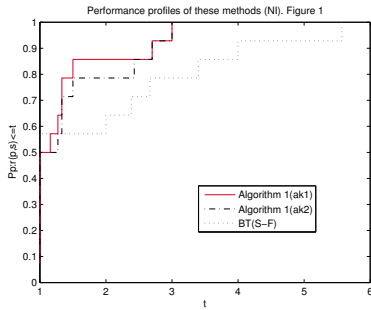


Figure 1. Performance of NI.

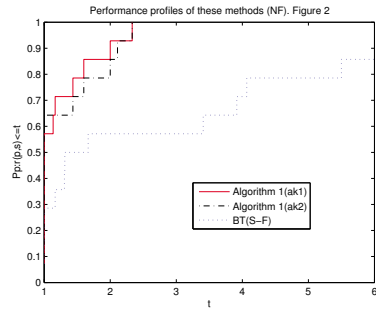


Figure 2. Performance of NF.

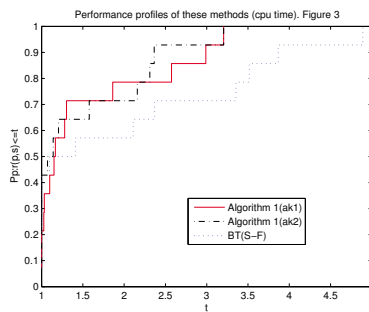


Figure 3. Performance of cpu time.

201010LX288), Guangxi SF (Grant No. 0991028), and the Scientific Research Foundation of GuangXi University (Grant No. XJZ110632).

References

- [1] J. Barzilai and J.M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, **8**:141–148, 1988. <http://dx.doi.org/10.1093/imanum/8.1.141>.
- [2] J.R. Birge, L. Qi and Z. Wei. Convergence analysis of some methods for minimizing a nonsmooth convex function. *J. Optim. Theory Appl.*, **97**:357–383, 1998. <http://dx.doi.org/10.1023/A:1022630801549>.
- [3] J.R. Birge, L. Qi and Z. Wei. A general approach to convergence properties of some methods for nonsmooth convex optimization. *Appl. Math. Optim.*, **38**:141–158, 1998. <http://dx.doi.org/10.1007/s002459900086>.
- [4] E.G. Birgin and J.M. Martinez. A spectral conjugate gradient method for unconstrained optimization. *Appl. Math. Optim.*, **43**:117–128, 2001. <http://dx.doi.org/10.1007/s00245-001-0003-0>.
- [5] E.G. Birgin, J.M. Martinez and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, **10**:1196–1221, 2000. <http://dx.doi.org/10.1137/S1052623497330963>.

- [6] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal and C.A. Sagastizábal. A family of variable metric proximal methods. *Math. Program.*, **68**:15–47, 1995. <http://dx.doi.org/10.1007/BF01585756>.
- [7] F.H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley, New York, 1983.
- [8] A.R. Conn, N.I.M. Gould and P.L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, USA, 2000.
- [9] R. Correa and C. Lemaréchal. Convergence of some algorithms for convex minimization. *Math. Program.*, **62**:261–273, 1993. <http://dx.doi.org/10.1007/BF01585170>.
- [10] Y.H. Dai. Alternate step gradient method. *Optimization*, **52**:395–415, 2003. <http://dx.doi.org/10.1080/02331930310001611547>.
- [11] Y.H. Dai and R. Fletcher. Projected Barzilai–Borwein methods for large-scale box-constrained quadratic programming. *Numer. Math.*, **100**:21–47, 2005. <http://dx.doi.org/10.1007/s00211-004-0569-y>.
- [12] Y.H. Dai, W.W. Hager, K. Schittkowski and H. Zhang. The cyclic Barzilai–Borwein method for unconstrained optimization. *IMA J. Numer. Anal.*, **26**:604–627, 2006. <http://dx.doi.org/10.1093/imanum/drl006>.
- [13] Y.H. Dai and L.Z. Liao. R-linear convergence of the Barzilai and Borwein gradient method. *IMA J. Numer. Anal.*, **22**:1–10, 2002. <http://dx.doi.org/10.1093/imanum/22.1.1>.
- [14] Y.H. Dai, L.Z. Liao and D. Li. An analysis of barzilai–borwein gradient method for unsymmetric linear equations. In K. Teo L. Qi and X. Yang(Eds.), *Optimization and Control with Applications*, pp. 183–211. Springer, 2005.
- [15] Y.H. Dai, J. Yuan and Y.X. Yuan. Modified two-point stepsize gradient method for unconstrained optimization. *Comput. Optim. Appl.*, **22**:103–109, 2002. <http://dx.doi.org/10.1023/A:1014838419611>.
- [16] Y.H. Dai and Y. Yuan. Analysis of monotone gradient methods. *J. Ind. Manag. Optim.*, **1**:181–192, 2005. <http://dx.doi.org/10.3934/jimo.2005.1.181>.
- [17] Y.H. Dai and H. Zhang. Adaptive two-point stepsize gradient algorithm. *Numer. Algorithms*, **27**:377–385, 2001. <http://dx.doi.org/10.1023/A:1013844413130>.
- [18] R.J.B. de Sampaio, J.Y. Yuan and W.Y. Sun. Trust region algorithm for nonsmooth optimization. *Appl. Math. Comput.*, **85**:109–116, 1997. [http://dx.doi.org/10.1016/S0096-3003\(96\)00112-9](http://dx.doi.org/10.1016/S0096-3003(96)00112-9).
- [19] J.E. Dennis Jr., S.B. Li and R.A. Tapia. A unified approach to global convergence of trust region methods for nonsmooth optimization. *Math. Program.*, **68**:319–346, 1995. <http://dx.doi.org/10.1007/BF01585770>.
- [20] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, **91**:201–213, 2002. <http://dx.doi.org/10.1007/s101070100263>.
- [21] A. Friedlander, J.M. Martínez and B. Molina. Gradient method with restarts and generalizations. *SIAM J. Numer. Anal.*, **36**:275–289, 1999. <http://dx.doi.org/10.1137/S003614299427315X>.
- [22] M. Fukushima and L. Qi. A global and superlinearly convergent algorithm for nonsmooth convex minimization. *SIAM J. Optim.*, **6**:1106–1120, 1996. <http://dx.doi.org/10.1137/S1052623494278839>.

- [23] S.A. Gabriel and J.S. Pang. A trust-region method for constrained nonsmooth equations. In W.W. Hager, D.W. Hearn and P.M. Pardalos(Eds.), *Large Scale Optimization-State of the Art*, Dordrecht, Holland, 1994. Kluwer Academic Publishers.
- [24] L. Grippo, F. Lampariello and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, **23**:707–716, 1986. <http://dx.doi.org/10.1137/0723046>.
- [25] L. Grippo and M. Sciandrone. Nonmonotone globalization techniques for the Barzilai–Borwein gradient method. *Comput. Optim. Appl.*, **23**:134–169, 2002. <http://dx.doi.org/10.1023/A:1020587701058>.
- [26] J.Y. Han and G.H. Liu. Global convergence analysis of a new nonmonotone BFGS algorithm on convex objective functions. *Comput. Optim. Appl.*, **7**:277–289, 1997. <http://dx.doi.org/10.1023/A:1008656711925>.
- [27] L. Han, G. Yu and L. Guan. Multivariate spectral gradient method for unconstrained optimization. *Appl. Math. Comput.*, **210**:621–630, 2008. <http://dx.doi.org/10.1016/j.amc.2007.12.054>.
- [28] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II*. Springer-Verlag, Berlin, Heidelberg, 1983.
- [29] K.C. Kiwiel. *Methods of Descent for Nondifferentiable Optimization*. Springer-Verlag, Berlin, New York, 1985.
- [30] K.C. Kiwiel. Proximity control in bundle methods for convex nondifferentiable optimization. *Math. Program.*, **46**:105–122, 1990. <http://dx.doi.org/10.1007/BF01585731>.
- [31] K.C. Kiwiel. Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities. *Math. Program.*, **69**:89–109, 1995. <http://dx.doi.org/10.1007/BF01585554>.
- [32] C. Lemaréchal. *Extensions diverses des méthodes de gradient et applications*. Thèse d'Etat, Paris, 1980.
- [33] C. Lemaréchal. Nondifferentiable optimization. In G.L. Nemhauser, A.H.G. Rinnooy Kan and M.J. Todd(Eds.), *Handbooks in Operations Research and Management Science, vol. 1, Optimization*, Amsterdam, 1989. North-Holland.
- [34] G.H. Liu and J.M. Peng. The convergence properties of a nonmonotonic algorithm. *J. Comput. Math.*, **1**:65–71, 1992.
- [35] L. Lukšan and J. Vlček. *Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization*. Technical Report No. 798, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2000.
- [36] L. Lukšan and J. Vlček. A bundle-Newton method for nonsmooth unconstrained minimization. *Math. Program.*, **83**:373–391, 1998. <http://dx.doi.org/10.1007/BF02680566>.
- [37] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française d'Aut. Inf. Rech. Opé.*, **4**:154–159, 1970.
- [38] J.M. Martínez, E.A. Pilotta and M. Raydan. Spectral gradient methods for linearly constrained optimization. *J. Optim. Theory Appl.*, **125**:629–651, 2005. <http://dx.doi.org/10.1007/s10957-005-2093-3>.
- [39] H. Qi, L. Qi and D. Sun. Solving KKT system via the trust region and the conjugate gradient method. *SIAM J. Optim.*, **14**:439–463, 2004. <http://dx.doi.org/10.1137/S105262340038256X>.

- [40] L. Qi. Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems. *Math. Oper. Res.*, **24**:440–471, 1999. <http://dx.doi.org/10.1287/moor.24.2.440>.
- [41] L. Qi and J. Sun. A nonsmooth version of Newton’s method. *Math. Program.*, **58**:353–367, 1993. <http://dx.doi.org/10.1007/BF01581275>.
- [42] M. Raydan. On the Barzilai and Borwein choce of steplength for the gradient method. *IMA J. Numer. Anal.*, **13**:321–326, 1993. <http://dx.doi.org/10.1093/imanum/13.3.321>.
- [43] M. Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.*, **7**:26–33, 1997. <http://dx.doi.org/10.1137/S1052623494266365>.
- [44] N. Sagara and M. Fukushima. A trust region method for nonsmooth convex optimization. *J. Ind. Manag. Optim.*, **1**:171–180, 2005. <http://dx.doi.org/10.3934/jimo.2005.1.171>.
- [45] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM J. Optim.*, **2**:121–152, 1992. <http://dx.doi.org/10.1137/0802008>.
- [46] P.L. Toint. An assessment of non-monotone line search techniques for unconstrained minimization problem. *SIAM J. Optim.*, **17**:725–739, 1996.
- [47] C. Wang, Q. Liu and X. Yang. Convergence properties of nonmonotone spectral projected gradient methods. *J. Comput. Appl. Math.*, **182**:51–66, 2005. <http://dx.doi.org/10.1016/j.cam.2004.10.018>.
- [48] Z. Wei and L. Qi. Convergence analysis of a proximal Newton method. *Numer. Funct. Anal. Optim.*, **17**:463–472, 1996. <http://dx.doi.org/10.1080/01630569608816705>.
- [49] Z. Wei, L. Qi and J.R. Birge. A new methods for nonsmooth convex optimization. *J. Inequal. Appl.*, **2**:157–179, 1998.
- [50] P. Wolfe. A method of conjugate subgradients for minimizing nondifferentiable convex functions. *Mathematical Programming Study*, **3**:145–173, 1975.
- [51] Y. Xiao, Q. Wang and D. Wang. Notes on the Dai–Yuan–Yuan modified spectral gradient method. *J. Comput. Appl. Math.*, **234**:2986–2992, 2010. <http://dx.doi.org/10.1016/j.cam.2010.04.012>.
- [52] G. Yu, J. Huang and Y. Zhou. A descent spectral conjugate gradient method for impulse noise removal. *Appl. Math. Lett.*, **23**:555–560, 2010. <http://dx.doi.org/10.1016/j.aml.2010.01.010>.
- [53] Z. Yu. Solving bound constrained optimization via a new nonmonotone spectral projected gradient method. *Appl. Numer. Math.*, **58**:1340–1348, 2008. <http://dx.doi.org/10.1016/j.apnum.2007.07.007>.
- [54] Z. Yu, J. Lin, J. Sun, Y. Xiao, L. Liu and Z. Li. Spectral gradient projection method for monotone nonlinear equations with convex constraintsl. *Appl. Numer. Math.*, **59**:2416–2423, 2009. <http://dx.doi.org/10.1016/j.apnum.2009.04.004>.
- [55] Z. Yu, J. Sun and Y. Qin. A multivariate spectral projected gradient method for bound constrained optimization. *J. Comput. Appl. Math.*, **235**:2263–2269, 2011. <http://dx.doi.org/10.1016/j.cam.2010.10.023>.

- [56] H.C. Zhang and W.W. Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.*, **14**:1043–1056, 2004. <http://dx.doi.org/10.1137/S1052623403428208>.
- [57] L. Zhang. A new trust region algorithm for nonsmooth convex minimization. *Appl. Math. Comput.*, **193**:135–142, 2007. <http://dx.doi.org/10.1016/j.amc.2007.03.059>.
- [58] L. Zhang and W. Zhou. Spectral gradient projection method for solving nonlinear monotone equations. *J. Comput. Appl. Math.*, **196**:478–484, 2006. <http://dx.doi.org/10.1016/j.cam.2005.10.002>.
- [59] J.L. Zhou and A.L. Tits. Nonmonotone line search for minimax problem. *J. Optim. Theory Appl.*, **76**:455–476, 1993. <http://dx.doi.org/10.1007/BF00939377>.
- [60] J. Zowe. Nondifferentiable optimization. In K. Schittkowski(Ed.), *Computational Mathematical Programming*, pp. 323–356, Berlin, Germany, 1985. Springer-Verlag.