AMO - Advanced Modeling and Optimization, Volume 14, Number 1, 2012

A Spectral-Type Conjugate Gradient Method for Nonsmooth Convex Minimization *

Qiong Li†

Abstract. Conjugate gradient methods are efficient for smooth optimization problems, while few researchers study conjugate gradient based methods for nonsmooth convex minimization problems. In this paper by making full use of inherent properties of Moreau-Yosida regularization we propose a spectral-type conjugate gradient method for nonsmooth convex minimization, with a new line search on approximate value of the Moreau-Yosida regularization function instead of its exact value. This algorithm is globally convergent under mild conditions.

Key words. nonsmooth convex minimization, spectral-type conjugate gradient method, Moreau-Yosida regularization, global convergence

1 Introduction

Let $f : R^n \to R$ be a nonsmooth convex function and consider unconstrained optimization problems of the form

$$\min_{x \in R^n} f(x). \tag{1.1}$$

The nonsmooth convex minimization problems are encounted in many application areas: for instance, in economics [26], mechanics [23], and machine learning [31], etc. The most of methods for solving nonsmooth convex problems may be divided in four main groups: subgradient methods [7, 14, 30], bundle methods [19, 22], smoothing technology [8, 11, 21, 24, 25, 27, 28] and hybrid methods [2, 3, 4, 13]. In this paper we restrict our attention to a smoothing technique technique. We adopt Moreau-Yosida regularization to convert nonsmooth problem (1.1) into a smooth problem

$$\min_{x \in R^n} F(x), \tag{1.2}$$

*Supported by the NSF of China grant 11071087.

†College of Science, China Three Gorges University, Yichang, Hubei, 443002, P. R. China.
E-mail: liqiongmanj@yahoo.com.

AMO - Advanced Modeling and Optimization. ISSN: 1841-4311

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

$$F(x) = \min_{z \in R^n} \{ f(z) + \frac{1}{2\lambda} \parallel z - x \parallel^2 \},$$

where λ is a positive parameter and $\parallel \cdot \parallel$ denotes the Euclidean norm.

The function F has some good properties: problems (1.1) and (1.2) are equivalent in the sense that the solution sets of the two problems coincide with each other. F is a differentiable convex function and has a Lipschitz continuous gradient even when the function f is nondifferentiable. Thus, (1.1) can be solved by (1.2).

Existing methods for (1.2) are mainly Newton-type methods (see, e.g., [5, 6, 8, 11, 16, 17, 20, 21, 27, 28] and the references therein) and trust region method (see, e.g., [18, 29, 32] and the references therein). The algorithms studied in [8, 11, 21, 27, 28, 29] are implementable in the sense that they utilize inexact values of the Moreau-Yosida regularization and its gradient. Rauf and Fukushima in [28] make a direct application of the BFGS method to the Moreau-Yosida regularization, which is globally convergent under the assumption of strong convexity of the objective function. Sagara-Fukushima in [29] proposed an implementable trust-region method, and proved its global convergence under the strong convexity assumption on the function to be minimized.

It is well-known that nonlinear conjugate gradient methods are very efficient for large-scale smooth optimization problems due to their simplicity and low storage. As far as we know, few researchers study conjugate gradient based method for nonsmooth convex minimization, which motivate us to propose a conjugate gradient based method for minimizing Moreau-Yosida regularization F, with a new line search on approximate value of the function F instead of its exact value. The line search rule is similar to but different from that in [11, 28]. In this paper, we will focus on a spectral-type Fletcher-Reeves method (SVFR) which is a descent conjugate gradient method, recently proposed by Lu et al [15] for solving smooth unconstrained optimization. Under mild conditions, we prove the global convergence of the method. Note that the objective function f need not be strongly convex.

The paper is organized as follows. In the next section, we briefly review some known results about the objective function of (1.2) and some basic results in convex analysis and nonsmooth analysis. In section 3, we derive our algorithm. Section 4 is devoted to proving its global convergence. The last section contains some concluding remarks.

Throughout this paper, $\langle \cdot, \cdot \rangle$ denotes inner product of two vectors.

2 Preliminaries

In this section, we recall some basic results in convex analysis which are useful in the subsequent discussions. Let $\theta : R^n \to R$ be a function such that

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

Clearly $\theta(z)$ is strongly convex and hence $p(x) = \arg\min_{z \in R^n} \theta(z)$ is well defined and unique. Then $F(x)$ can be expressed by

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

Denote the gradient of $F(x)$ by $g(x)$. Some features about $F(x)$ can be seen in [12].

Properties

1. The function F is finite-valued, convex and everywhere differentiable with gradient

$$g(x) = \frac{1}{\lambda}(x - p(x)).$$

Moreover, the gradient mapping $g : R^n \to R^n$ is globally Lipschitz continuous with modulus $\frac{1}{\lambda}$, i.e.,

$$\parallel g(x) - g(y) \parallel \leq \frac{1}{\lambda} \parallel x - y \parallel, \quad \forall x, y \in R^n. \tag{2.1}$$

2. x is an optimal solution to (1.1) if and only if $g(x) = 0$, namely $p(x) = x$.

It is obvious that $F(x)$ and $g(x)$ can be obtained by $p(x)$. However $p(x)$ is difficult or even impossible to obtain. Fortunately, for each $x \in R^n$ and any $\varepsilon > 0$, there exists a vector $p^a(x, \varepsilon) \in R^n$, where the superscript character a means the approximation, such that

$$f(p^a(x, \varepsilon)) + \frac{1}{2\lambda} \parallel p^a(x, \varepsilon) - x \parallel^2 \leq F(x) + \varepsilon. \tag{2.2}$$

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

$$F^a(x, \varepsilon) = f(p^a(x, \varepsilon)) + \frac{1}{2\lambda} \parallel p^a(x, \varepsilon) - x \parallel^2, \tag{2.3}$$

and

$$g^a(x, \varepsilon) = \frac{1}{\lambda}(x - p^a(x, \varepsilon)). \tag{2.4}$$

Implementable algorithms that are designed to find $p^a(x, \varepsilon)$ are introduced in [1, 9, 10]. The following Proposition deriving from Fukushima and Qi [11] shows that with $p^a(x, \varepsilon)$ we can compute approximation $F^a(x, \varepsilon)$ and $g^a(x, \varepsilon)$ to $F(x)$ and $g(x)$, respectively with any desired accuracy.

Proposition 2.1 *Let $p^a(x,\varepsilon)$ be a vector satisfying*

$$F^a(x,\varepsilon) \le F(x) + \varepsilon,$$

and $F^a(x,\varepsilon)$ and $g^a(x,\varepsilon)$ be given by (2.3) *and* (2.4)*, respectively. Then we have*

$$F(x) \le F^a(x,\varepsilon) \le F(x) + \varepsilon, \tag{2.5}$$

$$\| p^a(x,\varepsilon) - p(x) \| \le \sqrt{2\lambda\varepsilon}, \tag{2.6}$$

$$\| g^a(x,\varepsilon) - g(x) \| \le \sqrt{\frac{2\varepsilon}{\lambda}}. \tag{2.7}$$

3 SVFR Type Algorithm Description

In this section, based on the idea of [15] we propose SVFR type method for solving (1.1), in which the direction is defined by

$$d^k = \begin{cases} -g^a(x^k,\varepsilon_k) & \text{if} \quad k = 0, \\ -\theta_k g^a(x^k,\varepsilon_k) + \beta_k^{SVFR} d_{k-1} & \text{if} \quad k \ge 1, \end{cases} \tag{3.1}$$

where

$$\begin{aligned} &y_{k-1} = g^a(x^k,\varepsilon_k) - g^a(x^{k-1},\varepsilon_{k-1}), \\ &\beta_k^{SVFR} = \tfrac{\|g^a(x^k,\varepsilon_k)\| |\langle g^a(x^{k-1},\varepsilon_{k-1}), g^a(x^k,\varepsilon_k)\rangle|}{\|g^a(x^{k-1},\varepsilon_{k-1})\|^3}, \\ &\theta_k = \tfrac{|\langle d^{k-1}, g^a(x^k,\varepsilon_k)\rangle| - \langle d^{k-1}, g^a(x^{k-1}\rangle,\varepsilon_{k-1})}{\|g^a(x^{k-1},\varepsilon_{k-1})\|^2}. \end{aligned}$$

The complete algorithm is described as follows.

Algorithm 1: SVFR type algorithm

Step 0 Given constants $\sigma_1 \in (0,1)$, $\rho \in (0,1)$, and $\sigma_2 > 0$ and an initial point $x_0 \in R^n$ and $\varepsilon_0 > 0$. Let $k := 0$.

Step 1 Compute $p^a(x^k,\varepsilon_k)$. Compute the search direction (3.1).

Step 2 Choose a scalar ε_{k+1} such that $0 < \varepsilon_{k+1} < \varepsilon_k$. Let i_k be the smallest nonnegative integer i such that

$$F^a(x^k + \rho^i d^k, \varepsilon_{k+1}) \le F^a(x^k,\varepsilon_k) + \sigma_1 \rho^i \langle g^a(x^k,\varepsilon_k), d^k\rangle - \sigma_2 \rho^{2i} \| d^k \|^2 + \varepsilon_k. \tag{3.2}$$

Set $\alpha_k = \rho^{i_k}$ and $x^{k+1} = x^k + \alpha_k d^k$. Let $k := k+1$. Go to **Step 1**.

Lemma 3.1 *Suppose the direction d^k is defined by* (3.1)*, then we have*

$$\langle g^a(x^k,\varepsilon_k), d^k\rangle \le -\|g^a(x^k,\varepsilon_k)\|^2. \tag{3.3}$$

Proof. It is obvious that (3.3) holds for $k=0$. Suppose that d^{k-1} satisfies (3.3) for $k \geq 1$. Due to

$$0 \leq \beta_k^{SVFR} \leq \frac{\|g^a(x^k,\varepsilon_k)\|^2}{\|g^a(x^{k-1},\varepsilon_{k-1})\|^2} = \beta_k^{FR}.$$

We have

$$\begin{aligned} \langle d^k, g^a(x^k,\varepsilon_k)\rangle &= -\theta_k\|g^a(x^k,\varepsilon_k)\|^2 + \beta_k^{SVFR}\langle d^{k-1}, g^a(x^k,\varepsilon_k)\rangle \\ &\leq -\theta_k\|g_k\|^2 + \beta_k^{FR} \mid \langle d^{k-1}, g^a(x^k,\varepsilon_k)\rangle \mid \\ &= \beta_k^{FR}\langle d^{k-1}, g^a(x^{k-1},\varepsilon_{k-1})\rangle \\ &\leq -\beta_k^{FR}\|g^a(x^{k-1},\varepsilon_{k-1}\|^2 \\ &= -\|g^a(x^k,\varepsilon_k)\|^2. \end{aligned}$$

□

Note $\beta_k^{FR} \leq \frac{\langle d^k, g^a(x^k,\varepsilon_k)\rangle}{\langle d^{k-1}, g^a(x^{k-1},\varepsilon_{k-1})\rangle}$. We obtain

$$\beta_k^{SVFR} \leq \frac{\langle d^k, g^a(x^k,\varepsilon_k)\rangle}{\langle d^{k-1}, g^a(x^{k-1},\varepsilon_{k-1})\rangle}. \tag{3.4}$$

The following proposition ensures that, at each iteration k of the algorithm, α_k is well defined and can be determined finitely in Step 2.

Proposition 3.2 *For every k, there exists $\bar{\alpha}_k > 0$ such that*

$$F^a(x^k+\tau d^k,\varepsilon_{k+1}) \leq F^a(x^k,\varepsilon_k) + \sigma_1\tau\langle g^a(x^k,\varepsilon_k), d^k\rangle - \sigma_2\tau^2 \parallel d^k \parallel^2 + \varepsilon_k, \tag{3.5}$$

for all $\tau \in (0,\bar{\alpha}_k)$.

Proof. By Proposition 2.1, we have

$$F^a(x_k+\tau d^k,\varepsilon_{k+1}) \leq F(x^k+\tau d^k) + \varepsilon_{k+1}, \tag{3.6}$$

$$F(x^k) \leq F^a(x^k,\varepsilon_k), \tag{3.7}$$

$$\begin{aligned} \langle g(x^k), d^k\rangle - \langle g^a(x^k,\varepsilon_k), d^k\rangle &= \langle g(x^k) - g^a(x^k,\varepsilon_k), d^k\rangle \\ &\leq \parallel g(x^k) - g^a(x^k,\varepsilon_k) \parallel\parallel d^k \parallel \\ &\leq \sqrt{\frac{2\varepsilon_k}{\lambda}} \parallel d^k \parallel . \end{aligned} \tag{3.8}$$

Adding (3.6), (3.7) and (3.8) multiplied by τ, we obtain

$$\begin{aligned} F^a(x_k+\tau d^k,\varepsilon_{k+1}) \leq\ & F^a(x^k,\varepsilon_k) + \tau\langle g^a(x^k,\varepsilon_k), d^k\rangle + \varepsilon_{k+1} + \tau\sqrt{\frac{2\varepsilon_k}{\lambda}} \parallel d^k \parallel \\ & + F(x^k+\tau d^k) - F(x^k) - \tau\langle g(x^k), d^k\rangle. \end{aligned} \tag{3.9}$$

If $d^k = 0$, then (3.8) implies that (3.5) holds for any $\tau > 0$, because $\varepsilon_{k+1} < \varepsilon_k$. Consider the case $d^k \neq 0$. By Lemma 3.1,

$$\langle g^a(x^k, \varepsilon_k), d^k \rangle \leq - \parallel g^a(x^k, \varepsilon_k) \parallel^2 \leq 0,$$

so that

$$\langle g^a(x^k, \varepsilon_k), d^k \rangle \leq \sigma \langle g^a(x^k, \varepsilon_k), d^k \rangle. \tag{3.10}$$

Given x^k and d^k, denote

$$\phi(\tau) = \tau \sqrt{\frac{2\varepsilon_k}{\lambda}} \parallel d^k \parallel + F(x^k + \tau d^k) - F(x^k) - \tau (g(x^k), d^k) + \sigma_2 \tau^2 \parallel d^k \parallel^2 .$$

Since F is continuous and $\varepsilon_{k+1} < \varepsilon_k$, we have

$$\lim_{\tau \to 0} \phi(\tau) = 0 < \varepsilon_k - \varepsilon_{k+1}.$$

Therefore there exists $\bar{\alpha}_k > 0$ such that $\phi(\tau) < \varepsilon_k - \varepsilon_{k+1}$ for all $\tau \in (0, \bar{\alpha}_k)$. That is,

$$\varepsilon_{k+1} + \tau \sqrt{\frac{2\varepsilon_k}{\lambda}} \parallel d^k \parallel + F(x^k + \tau d^k) - F(x^k) - \tau (g(x^k), d^k) < -\sigma_2 \tau^2 \parallel d^k \parallel^2 + \varepsilon_k \tag{3.11}$$

for all $\tau \in (0, \bar{\alpha}_k)$. That (3.5) holds for all $\tau \in (0, \bar{\alpha}_k)$ follows from (3.9)-(3.11).

4 Global Convergence

In this section, we establish the global convergence of Algorithm 1 under Assumption A.

A1 f is bounded from below.

A2 $\Omega = \{x \in R^n \mid F(x) \leq F(x^0) + \sum_{i=0}^{\infty} \varepsilon_i\}$ is bounded.

We note that this assumption is a weaker condition than the strong convexity of f as required in [28, 29], which can be verified by the fact that the property of strong convexity of f is transmitted to the Moreau-Yosida regularization F: If f is strongly convex, then F is strongly convex [17] (Theorem 2.2), so we can deduce that the strong convexity of f implies the boundedness of Ω. It is clear that the sequence $\{x^k\}$ generated by Algorithm 1 are contained in Ω. Combining this assumption with the Lipschitz continuous property of the gradient g, we have that there exists a constant $\gamma > 0$ such that

$$\parallel g(x) \parallel \leq \gamma, \quad \forall x \in \Omega. \tag{4.1}$$

Combining (4.1) with Proposition 2.1, we obtain the conclusion that there exists a constant $\gamma_1 > 0$ such that

$$\parallel g^a(x, \varepsilon) \parallel \leq \gamma_1, \quad \forall x \in \Omega. \tag{4.2}$$

Lemma 4.1 *Let $\{x^k\}$ and $\{d^k\}$ be generated by Algorithm 1. If the sequence $\{\varepsilon_k\}$ of strictly decreasing positive numbers satisfies the condition*

$$\sum_{k=0}^{\infty} \sqrt{\varepsilon_k} < +\infty. \tag{4.3}$$

Then the whole sequence $\{F^a(x^k, \varepsilon_k)\}$ is convergent, and

$$\sum_{k\geq 0} -\alpha_k \langle g^a(x^k, \varepsilon_k), d^k \rangle < +\infty. \tag{4.4}$$

Proof. By the line search rule, it holds that

$$F^a(x^{k+1}, \varepsilon_{k+1}) \leq F^a(x^k, \varepsilon_k) + \sigma_1 \alpha_k (g^a(x^k, \varepsilon_k), d^k) - \sigma_2 \alpha_k^2 \parallel d^k \parallel^2 + \varepsilon_k \tag{4.5}$$

for all k. By Lemma 3.1

$$\langle g^a(x^k, \varepsilon_k), d^k \rangle \leq - \parallel g^a(x^k, \varepsilon_k) \parallel^2 \leq 0,$$

it follows that

$$F^a(x^{k+1}, \varepsilon_{k+1}) \leq F^a(x^k, \varepsilon_k) + \varepsilon_k, \tag{4.6}$$

and hence

$$F^a(x^k, \varepsilon_k) \leq F^a(x^0, \varepsilon_0) + \sum_{i=0}^{k-1} \varepsilon_i,$$

which together with the assumption $\sum_{k=0}^{\infty} \sqrt{\varepsilon_k} < +\infty$ implies that the sequence $F^a(x^k, \varepsilon_k)$ is bounded from above. On the other hand, f is bounded from below by assumption, and hence F is also bounded from below. Since $F^a(x^k, \varepsilon_k) \geq F(x^k)$ for all k, the sequence $\{F^a(x^k, \varepsilon_k)\}$ is bounded from below. Therefore the sequence $\{F^a(x^k, \varepsilon_k)\}$ has at least one accumulation point. In fact, it can be shown in a way similar to the first part of the proof of Theorem 4.1 in [11] that the whole sequence $\{F^a(x^k, \varepsilon_k)\}$ is convergent. Applying the inequality (4.5) recursively, we have

$$F^a(x^{k+1}, \varepsilon_{k+1}) \leq F^a(x^0, \varepsilon_0) + \sigma_1 \sum_{i=0}^{i=k} \alpha_i (g^a(x^i, \varepsilon_i), d^i) - \sigma_2 \sum_{i=0}^{i=k} \alpha_i^2 \parallel d^i \parallel^2 + \sum_{i=0}^{i=k} \varepsilon_i.$$

That is,

$$\sigma_1 \sum_{i=0}^{i=k} -\alpha_i (g^a(x^i, \varepsilon_i), d^i) + \sigma_2 \sum_{i=0}^{i=k} \alpha_i^2 \parallel d^i \parallel^2 \leq F^a(x^0, \varepsilon_0) - F^a(x^{k+1}, \varepsilon_{k+1}) + \sum_{i=0}^{i=k} \varepsilon_i. \tag{4.7}$$

Since the whole sequence $\{F^a(x^k, \varepsilon_k)\}$ is convergent, by taking the limit in (4.7) we have

$$\sum_{k\geq 0} (-\sigma_1 \alpha_k \langle g^a(x^k, \varepsilon_k), d^k \rangle + \sigma_2 \alpha_k^2 \|d^k\|^2) < +\infty.$$

This implies

$$\sum_{k\geq 0} -\alpha_k \langle g^a(x^k, \varepsilon_k), d^k\rangle < +\infty.$$

$\square$

Lemma 4.2 *Let $\{x^k\}$ and $\{d^k\}$ be generated by Algorithm* 1. *Then*

$$\sum_{k\geq 0} \frac{\langle g^a(x^k, \varepsilon_k), d^k\rangle^2}{\|d^k\|^2} < +\infty. \tag{4.8}$$

Proof. Now we prove (4.8) by considering the following two cases.

Case 1. $\alpha_k = 1$. We get from (3.3) $\|g^a(x^k, \varepsilon_k)\| \leq \| d^k \|$. Hence

$$\sum_{k\geq 0} \frac{\langle g^a(x^k, \varepsilon_k), d^k\rangle^2}{\|d^k\|^2} \leq \sum_{k\geq 0} \|g^a(x^k, \varepsilon_k)\|^2 \leq \sum_{k\geq 0} -\langle g^a(x^k, \varepsilon_k), d^k\rangle < +\infty.$$

Case 2. $\alpha_k < 1$. By the line search step, i.e., Step 2 of Algorithm 1, $\rho^{-1}\alpha_k$ does not satisfy inequality (3.2). This means

$$F^a(x^k + \rho^{-1}\alpha_k d^k, \varepsilon_{k+1}) - F^a(x^k, \varepsilon_k) > \sigma_1 \rho^{-1}\alpha_k \langle g^a(x^k, \varepsilon_k), d^k\rangle - \sigma_2 \rho^{-2}\alpha_k^2 \| d^k \|^2 + \varepsilon_k. \tag{4.9}$$

By the mean-value theorem and inequality (2.7), there is a $t_k \in (0, 1)$ such that $x^k + t_k \rho^{-1}\alpha_k d^k \in \Omega$ and

$$\begin{aligned} F(x^k + \rho^{-1}\alpha_k d^k) - F(x^k) &= \rho^{-1}\alpha_k \langle g(x^k + t_k\rho^{-1}\alpha_k d^k), d^k\rangle \\ &\leq \rho^{-1}\alpha_k g_k^T d^k + \frac{1}{\lambda}\rho^{-2}\alpha_k^2 \|d^k\|^2. \end{aligned}$$

Combining this with (2.5), we get

$$F^a(x^k + \rho^i d^k, \varepsilon_{k+1}) - F^a(x^k, \varepsilon_k) \leq \rho^{-1}\alpha_k g_k^T d^k + \frac{1}{\lambda}\rho^{-2}\alpha_k^2\|d^k\|^2 + \varepsilon_{k+1}.$$

Substituting this inequality into (4.9), we get

$$\sigma_1 \langle g^a(x^k, \varepsilon_k), d^k\rangle - \sigma_2 \rho^{-1}\alpha_k \|d^k\|^2 \leq g_k^T d^k + \frac{1}{\lambda}\rho^{-1}\alpha_k\|d^k\|^2.$$

Due to (2.7), we get

$$-(1-\sigma_1)\langle g^a(x^k, \varepsilon_k), d^k\rangle - \sqrt{\frac{2\varepsilon_k}{\lambda}}\|d^k\| \leq (\frac{1}{\lambda} + \sigma_2)\rho^{-1}\alpha_k\|d^k\|^2,$$

which implies

$$\alpha_k > c(1-\sigma_1)\frac{-\langle g^a(x^k, \varepsilon_k), d^k\rangle}{\|d^k\|^2} - c\sqrt{\frac{2\varepsilon_k}{\lambda}}\frac{1}{\|d^k\|},$$

where $c = \frac{\rho}{\frac{1}{\lambda}+\sigma_2}$. From (4.4), we obtain

$$\sum_{k\geq 0}((1-\sigma_1)\frac{\langle g^a(x^k,\varepsilon_k),d^k\rangle^2}{\|d^k\|^2} + \sqrt{\frac{2\varepsilon_k}{\lambda}}\frac{\langle g^a(x^k,\varepsilon_k),d^k\rangle}{\|d^k\|}) < +\infty.$$

Since

$$\sum_{k\geq 0} -\sqrt{\frac{2\varepsilon_k}{\lambda}}\frac{\langle g^a(x^k,\varepsilon_k),d^k\rangle}{\|d^k\|} \leq \sum_{k\geq 0}\sqrt{\frac{2\varepsilon_k}{\lambda}}\|g^a(x^k,\varepsilon_k)\| \leq \sum_{k\geq 0}\gamma_1\sqrt{\frac{2\varepsilon_k}{\lambda}} < +\infty,$$

the second inequality is due to (4.2). So (4.8) holds. □

Theorem 4.3 *Let $\{x^k\}$ and $\{d^k\}$ be generated by Algorithm* 1. *We have*

$$\liminf_{k\to\infty}\|g_k\| = 0. \tag{4.10}$$

Proof. For the sake of contradiction, we suppose that the conclusion is not true. Then there exists a constant $\epsilon > 0$ such that

$$\|g_k\| \geq \epsilon, \quad \forall k \geq 0. \tag{4.11}$$

From (2.7), we obtain there exists a constant $\epsilon_* > 0$ such that

$$\|g^a(x^k,\varepsilon_k)\| \geq \epsilon_*, \quad \forall k \geq 0. \tag{4.12}$$

We get from (3.1) and (3.4) that

$$\begin{aligned}
\|d^k\|^2 &= (\beta_k^{SVFR})^2\|d^{k-1}\|^2 - 2\theta_k\langle d^k, g^a(x^k,\varepsilon_k)\rangle - \theta_k^2\|g^a(x^k,\varepsilon_k)\|^2 \\
&\leq \frac{\langle d^k, g^a(x^k,\varepsilon_k)\rangle^2}{\langle d^{k-1}, g^a(x^{k-1},\varepsilon_{k-1})\rangle}^2\|d^{k-1}\|^2 - 2\theta_k\langle d^k, g^a(x^k,\varepsilon_k)\rangle - \theta_k^2\|g^a(x^k,\varepsilon_k)\|^2
\end{aligned}$$

Dividing both sides of above inequality by $\langle g^a(x^k,\varepsilon_k),d^k\rangle^2$, we get from (4.12) that

$$\begin{aligned}
\frac{\|d^k\|^2}{\langle g^a(x^k,\varepsilon_k),d^k\rangle^2} &\leq \frac{\|d^{k-1}\|^2}{\langle g^a(x^{k-1},\varepsilon_{k-1}),d^k\rangle^2} - \frac{2\theta_k}{\langle d^k, g^a(x^k,\varepsilon_k)\rangle} - \frac{\theta_k^2\|g^a(x^k,\varepsilon_k)\|^2}{\langle g^a(x^k,\varepsilon_k),d^k\rangle^2} \\
&= \frac{\|d^{k-1}\|^2}{\langle g^a(x^{k-1},\varepsilon_{k-1}),d^k\rangle^2} - (\frac{\theta_k\|g^a(x^k,\varepsilon_k)\|}{\langle d^k, g^a(x^k,\varepsilon_k)\rangle} + \frac{1}{\|g^a(x^k,\varepsilon_k)\|})^2 + \frac{1}{\|g^a(x^k,\varepsilon_k)\|^2} \\
&\leq \frac{\|d^{k-1}\|^2}{\langle g^a(x^{k-1},\varepsilon_{k-1}),d^k\rangle^2} + \frac{1}{\|g^a(x^k,\varepsilon_k)\|^2} \\
&\leq \sum_{i=0}^{k}\frac{1}{\|g^a(x^i,\varepsilon_i)\|^2} \leq \frac{k+1}{\epsilon_*^2}.
\end{aligned}$$

The last inequality implies that

$$\sum_{k\geq 0}\frac{\langle g^a(x^k,\varepsilon_k),d^k\rangle^2}{\|d^k\|^2} \geq \epsilon_*^2\sum_{k\geq 0}\frac{1}{k+1} = +\infty.$$

which contradicts (4.8). The proof is then complete. □ .

5 Concluding remarks

In this paper, by means of Moreau-Yosida regularization, by introducing a new line search on the approximation to the Moreau-Yosida regularization, we propose a spectral-type conjugate gradient method for nonsmooth convex minimization. The global convergence is established under mild conditions.

References

[1] A. Auslender, "Numerical methods for nondifferentiable convex optimization," Mathematical Programming Study, vol. 30, pp. 102-126, 1987.

[2] A. M., Bagirov, and A. N. Ganjehlou, "A quasisecant method for minimizing nonsmooth functions," Optimization Methods and Software, vol. 25, no. 1, pp. 3-18, 2009.

[3] A. M., Bagirov, and A. N. Ganjehlou, A secant method for nonsmooth optimization. Submitted, 2009.

[4] A. M., Bagirov, B., Karasozen, and M. Sezer, "Discrete gradient method: A derivative free method for nonsmooth optimization," Journal of Optimization Theory and Applications, vol. 137, pp. 317-334, 2008.

[5] J. F., Bonnans, J. C., Gilbert, C., Lemarechal and C., Sagastizabal, "A family of variable-metric proximal methods," Mathematical Programming, vol. 68, pp. 15-47, 1995.

[6] J. V., Burke and M., Qian, "On the superlinear convergence of the variable- metric proximal point-algorithm using Broyden and BFGS matrix secant updating," Mathematical Programming, vol. 88, pp. 157-181, 2000.

[7] I. V. Burke, and A. S. Lewis, "The speed of Shor's r-algorithm," IMA Journal of Numerical Analysis, vol. 28, pp. 711-720, 2008.

[8] X., Chen, and M., Fukushima, "Proximal quasi-Newton methods for nondifferentiable convex optimization," Mathematical Programming, vol. 85, pp. 313-334, 1999.

[9] R. Correa and C. Lemarechal, "Convergence of some algorithms for convex minimization," Mathematical Programming, vol. 62, pp. 261-275, 1993.

[10] M. Fukushima, "A descent algorithm for nonsmooth convex optimization," Mathematical Programming, vol. 30, pp. 163-175, 1984.

[11] M., Fukushima and L., Qi, "A globally and superlinearly convergent algorithm for nonsmooth convex minimization," SIAM Journal on Optimization, vol. 6, pp. 1106-1120, 1996.

[12] J.-B., Hiriart-Urruty and C., Lemarechal, "Convex Analysis and Minimization Algorithms," Springer Verlag, Berlin, Germany, 1993.

[13] N., Haarala, K., Miettinen, and M. M. Mäkelä, "Globally convergent limited memory bundle method for large-scale nonsmooth optimization," Mathematical Programming, vol. 109, no. 1, pp. 181-205, 2007.

[14] F., Kappel, and A., Kuntsevich, "An implementation of Shor's r-algorithm," Computational Optimization and Applications, vol. 15, pp. 193-205, 2000.

[15] A., Lu, H., Liu, X., Zheng and W., Cong, "A variant spectral-type FR conjugate gradient method and its global convergence," Applied Mathematics and Computation, 2010.

[16] C., Lemarechal and C., Sagastizabal, "An approach to variable-metric bundle methods," Proceedings of the 16th IFIP-TC7 Conference on Systems Modelling and Optimization, edited by J. Henry and J. P. Yuvor, Springer Verlag, New York, NY, pp. 144-162, 1994.

[17] C., Lemarechal and C., Sagastizabal, "Practical aspects of the Moreau-Yosida regularization, I: theoretical preliminaries," SIAM Journal on Optimization, vol. 7, pp. 367-385, 1997.

[18] S., Lu, Z., Wei and L., Li, "A trust region algorithm with adaptive cubic regularization methods for nonsmooth convex minimization," Computational Optimization and Applications, published online, 19 October, 2010.

[19] L., Lukšan, and J. Vlček, "A bundle-Newton method for nonsmooth unconstrained minimization," Mathematical Programming, vol. 83, pp. 373-391, 1998.

[20] R., Mifflin, "A quasi-second-order proximal bundle algorithm," Mathematical Programming, vol. 73, pp. 51-72, 1996.

[21] R., Mifflin, D., Sun, and L., Qi, "Quasi-Newton bundle-type methods for nondifferentiable convex optimization," SIAM Journal on Optimization, vol. 8, pp. 583-603, 1998.

[22] M. M., Mäkelä, and P. Neittaanmäki, "Nonsmooth optimization: analysis and algorithms with applications to optimal control," World Scientific Publishing Co., Singapore, 1992.

[23] J. J., Moreau, P.D., Panagiotopoulos, and G., Strang, "Topics in nonsmooth mechanics," Birkhäuser Verlag, Basel, 1988.

[24] YU. Nesterov, "Excessive gap technique in nonsmooth convex minimization," SIAM Journal on Optimization, vol. 16, no. 1, pp. 235-249, 2005.

[25] YU. Nesterov, "Smooth minimization of nonsmooth functions," Mathematical Programming, Ser. A, vol. 103, pp. 127-152, 2005.

[26] J., Outrata, M., Kočvara, and J. Zowe, "Nonsmooth approach to optimization problems with equilibrium constraints," Theory, Applications and Numerical Results, Kluwert Academic Publisher, Dordrecht, 1998.

[27] L., Qi and X., Chen, "A preconditioned proximal Newton method for nondifferentiable convex optimization," Mathematical Programming, vol. 76, pp. 411-429, 1997.

[28] A. I., Rauf and M., Fukushima, "Global convergent BFGS method for nonsmooth convex optimization," Journal of Optimization Theory and Application, vol. 104, no. 3, pp. 539-558, 2000.

[29] N. Sagara and M. Fukushima, "A trust region method for nonsmooth convex optimization," Journal of Industrial and Management Optimization, vol. 1, no. 2, pp. 171-180, 2005.

[30] N. Z. Shor, "Minimization methods for non-differentiable functions," Springer-Verlag, Berlin, 1985.

[31] J. Yu, S. V. N. Vishwanathan, S. Günter, and N. N. Schraudolph, "A quasi-Newton approach to nonsmooth convex optimization in machine learning," Journal of Machine Learning Research, vol. 11, pp. 1145-1200, 2010.

[32] L. Zhang, "A new trust region algorithm for nonsmooth convex minimization," Applied Mathematics and Computation, vol. 193, pp. 135-142, 2007.

[33] L. Zhang, W. Zhou and D. Li, "Global convergence of a modified FletcherCReeves conjugate gradient method with Armijo-type line search", Numerische Mathematik, vol. 104, pp. 561-572, 2006.