



A Modified Fletcher-Reeves-Type Method for Nonsmooth Convex Minimization

Qiong Li^{1, *}

¹*College of Science, China Three Gorges University, Yichang, Hubei, 443002, P. R. China.*

Received 15 January 2014; Accepted 25 May 2014

Editor: Junfeng Yang

Abstract Conjugate gradient methods are efficient for smooth optimization problems, while there are rare conjugate gradient based methods for solving a possibly nondifferentiable convex minimization problem. In this paper by making full use of inherent properties of Moreau-Yosida regularization and descent property of modified conjugate gradient method we propose a modified Fletcher-Reeves-type method for nonsmooth convex minimization. It can be applied to solve large-scale nonsmooth convex minimization problem due to lower storage requirement. The algorithm is globally convergent under mild conditions.

Keywords nonsmooth convex minimization, Moreau-Yosida regularization, a modified Fletcher-Reeves-type method, global convergence

DOI: 10.19139/soic.v2i3.64

1. Introduction

Let $f : R^n \rightarrow R$ be a possibly nondifferentiable convex function and consider unconstrained optimization problem of the form

$$\min_{x \in R^n} f(x). \quad (1)$$

Associated with problem (1) is the problem

$$\min_{x \in R^n} F(x), \quad (2)$$

*Correspondence to: College of Science, China Three Gorges University, Yichang, Hubei, 443002, P. R. China. E-mail: liqiongmanj@163.com.

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

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

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

The function F has some good properties: problems (1) and (2) are equivalent in the sense that the solution sets of the two problems coincide with each other [9]. F is a differentiable convex function, even though the function f is nondifferentiable. Moreover F has a Lipschitz continuous gradient [9]. These features motivate us to solve problem (1) through the Moreau-Yosida regularization, particularly when f is nondifferentiable.

Existing methods for minimizing Moreau-Yosida regularization F are mainly Newton-type methods (see, e.g., [2, 3, 4, 7, 10, 11, 13, 14, 15, 16] and the references therein) and trust region methods (see, e.g., [12, 17, 19] and the references therein). The algorithms studied in [4, 7, 14, 15, 16, 17] are implementable in the sense that they utilize inexact values of the Moreau-Yosida regularization and its gradient. Rauf and Fukushima in [16] made a direct application of the BFGS method to the Moreau-Yosida regularization, Sagara-Fukushima in [17] proposed an implementable trust-region method. [16] and [17] studied global convergence under the assumption of strong convexity of the objective function.

The conjugate gradient methods are welcome methods for smooth unconstrained optimization problems. They are particularly efficient for large-scale problems due to their simplicity and low storage [8]. To the best of our knowledge, there are rare conjugate gradient based methods for nonsmooth convex minimization, which motivates this paper. We propose a conjugate gradient based method for minimizing Moreau-Yosida regularization F , with a line search on approximate value of the function F instead of its exact value. The line search rule is different from that in [7, 16]. In this paper, we will focus on the MFR method which is a descent conjugate gradient method, proposed by Zhang, Zhou and Li [20] for solving unconstrained optimization. Under mild conditions, we prove the global convergence of the method. Note that we do not require the strong convexity assumption on the objective function f as [16] and [17] required. Recently, Yuan, Wei and Li [18] propose a modified Polak-Ribière-Polyak conjugate gradient algorithm for nonsmooth convex programs, [18] and this paper have common feature that they both propose algorithms for problem (1) by means of Moreau-Yosida regularization and the search directions satisfy the sufficient descent property, but they have different line search technique.

The paper is organized as follows. In section 2, we derive the algorithm. Section 3 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, and $g(x)$ denotes the gradient of $F(x)$.

2. Derivation of MFR Type Algorithm

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

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

since $\theta_1(z) = \frac{1}{2\lambda} \|z - x\|^2$ is strongly convex and $f(z)$ is convex, by the definition of strongly convex function we know that $\theta(z)$ is strongly convex. Hence $p(x) = \arg \min_{z \in R^n} \theta(z)$ is well defined and unique. Then $F(x)$ can be expressed by

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

Some features about $F(x)$ can be seen in [9].

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 $g : R^n \rightarrow R^n$ is globally Lipschitz continuous with modulus $\frac{1}{\lambda}$, i.e.,

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

2. x is an optimal solution to (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

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

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

$$F^a(x, \varepsilon) = f(p^a(x, \varepsilon)) + \frac{1}{2\lambda} \|p^a(x, \varepsilon) - x\|^2, \quad (3)$$

and

$$g^a(x, \varepsilon) = \frac{1}{\lambda}(x - p^a(x, \varepsilon)). \tag{4}$$

Implementable algorithms that are designed to find $p^a(x, \varepsilon)$ are introduced in [1, 5, 6]. The following Proposition deriving from Fukushima and Qi [7] shows that with $p^a(x, \varepsilon)$ we can compute approximations $F^a(x, \varepsilon)$ and $g^a(x, \varepsilon)$ to $F(x)$ and $g(x)$, respectively with any desired accuracy.

Proposition 1

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

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

and $F^a(x, \varepsilon)$ and $g^a(x, \varepsilon)$ be given by (3) and (4), respectively. Then we have

$$F(x) \leq F^a(x, \varepsilon) \leq F(x) + \varepsilon, \tag{5}$$

$$\| p^a(x, \varepsilon) - p(x) \| \leq \sqrt{2\lambda\varepsilon}, \tag{6}$$

$$\| g^a(x, \varepsilon) - g(x) \| \leq \sqrt{\frac{2\varepsilon}{\lambda}}. \tag{7}$$

Now we state our algorithm.

Algorithm 1: MFR 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

$$d^k = \begin{cases} -g^a(x^k, \varepsilon_k) & \text{if } k = 0, \\ -\theta_k g^a(x^k, \varepsilon_k) + \beta_k^{FR} d^{k-1} & \text{if } k \geq 1, \end{cases} \tag{8}$$

where

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

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

$$F^a(x^k + \rho^i d^k, \varepsilon_{k+1}) \leq F^a(x^k, \varepsilon_k) + \sigma_1 \rho^i \langle g^a(x^k, \varepsilon_k), d^k \rangle - \sigma_2 \rho^{2i} \| d^k \|^2 + \varepsilon_k. \tag{9}$$

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

Proposition 2

The search direction (8) in Algorithm 1 satisfies

$$\langle g^a(x^k, \varepsilon_k), d^k \rangle = -\|g^a(x^k, \varepsilon_k)\|^2.$$

Proof. We omit the detail.

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

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

$$F^a(x^k + \tau d^k, \varepsilon_{k+1}) \leq F^a(x^k, \varepsilon_k) + \sigma_1 \tau \langle g^a(x^k, \varepsilon_k), d^k \rangle - \sigma_2 \tau^2 \|d^k\|^2 + \varepsilon_k, \quad (10)$$

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

Proof. By proposition 1, we have

$$F^a(x_k + \tau d^k, \varepsilon_{k+1}) \leq F(x^k + \tau d^k) + \varepsilon_{k+1}, \quad (11)$$

$$F(x^k) \leq F^a(x^k, \varepsilon_k), \quad (12)$$

$$\begin{aligned} \langle g(x^k), d^k \rangle - \langle g^a(x^k, \varepsilon_k), d^k \rangle &= \langle g(x^k) - g^a(x^k, \varepsilon_k), d^k \rangle \\ &\leq \|g(x^k) - g^a(x^k, \varepsilon_k)\| \|d^k\| \\ &\leq \sqrt{\frac{2\varepsilon_k}{\lambda}} \|d^k\|. \end{aligned} \quad (13)$$

Adding (11), (12) and (13) multiplied by τ , we obtain

$$\begin{aligned} F^a(x_k + \tau d^k, \varepsilon_{k+1}) &\leq F^a(x^k, \varepsilon_k) + \tau \langle g^a(x^k, \varepsilon_k), d^k \rangle + \varepsilon_{k+1} + \tau \sqrt{\frac{2\varepsilon_k}{\lambda}} \|d^k\| \\ &\quad + F(x^k + \tau d^k) - F(x^k) - \tau \langle g(x^k), d^k \rangle. \end{aligned} \quad (14)$$

If $d^k = 0$, then (13) implies that (10) holds for any $\tau > 0$, because $\varepsilon_{k+1} < \varepsilon_k$. Consider the case $d^k \neq 0$. Then, we have

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

so that

$$\langle g^a(x^k, \varepsilon_k), d^k \rangle \leq \sigma \langle g^a(x^k, \varepsilon_k), d^k \rangle. \quad (15)$$

Denote

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

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

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

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

$$\varepsilon_{k+1} + \tau \sqrt{\frac{2\varepsilon_k}{\lambda}} \|d^k\| + F(x^k + \tau d^k) - F(x^k) - \tau \langle g(x^k), d^k \rangle < -\sigma_2 \tau^2 \|d^k\|^2 + \varepsilon_k \tag{16}$$

for all $\tau \in (0, \bar{\alpha}_k)$. That (10) holds for all $\tau \in (0, \bar{\alpha}_k)$ follows from (14)-(16).

3. Global Convergence of MFR Type Algorithm

In this section, we prove the global convergence under the following assumptions:

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 [16, 17], 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 [11] (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

$$\|g(x)\| \leq \gamma, \quad \forall x \in \Omega. \tag{1}$$

Combining (1) with Proposition 1, we obtain the conclusion that there exists a constant $\gamma_1 > 0$ such that

$$\|g^a(x, \varepsilon)\| \leq \gamma_1, \quad \forall x \in \Omega. \tag{2}$$

Lemma 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

$$\sum_{k=0}^{\infty} \sqrt{\varepsilon_k} < +\infty. \tag{3}$$

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

$$\sum_{k \geq 0} \alpha_k \|g^a(x^k, \varepsilon_k)\|^2 < +\infty. \tag{4}$$

Proof. By the line search rule, it holds that

$$F^a(x^{k+1}, \varepsilon_{k+1}) \leq F^a(x^k, \varepsilon_k) + \sigma_1 \alpha_k \langle g^a(x^k, \varepsilon_k), d^k \rangle - \sigma_2 \alpha_k^2 \|d^k\|^2 + \varepsilon_k \quad (5)$$

for all k . Since

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

it follows that

$$F^a(x^{k+1}, \varepsilon_{k+1}) \leq F^a(x^k, \varepsilon_k) + \varepsilon_k, \quad (6)$$

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 [7] that the whole sequence $\{F^a(x^k, \varepsilon_k)\}$ is convergent.

Applying the inequality (5) recursively, we have

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

That is,

$$\sigma_1 \sum_{i=0}^{i=k} -\alpha_i \langle g^a(x^i, \varepsilon_i), d^i \rangle + \sigma_2 \sum_{i=0}^{i=k} \alpha_i^2 \|d^i\|^2 \leq F^a(x^0, \varepsilon_0) - F^a(x^{k+1}, \varepsilon_{k+1}) + \sum_{i=0}^{i=k} \varepsilon_i. \quad (7)$$

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

$$\sum_{k=0}^{\infty} \varepsilon_k < +\infty \quad \text{and} \quad -\langle g^a(x^k, \varepsilon_k), d^k \rangle \geq 0,$$

by taking the limit in (7) we have

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

□

Lemma 2

Assume the conditions of Lemma 3.1 hold. Then

$$\sum_{k \geq 0} \frac{\|g^a(x^k, \varepsilon_k)\|^4}{\|d^k\|^2} < +\infty. \tag{8}$$

Proof. Now we prove (8) by considering the following two cases.

Case 1. $\alpha_k = 1$. We get from Proposition 2 that $\|g^a(x^k, \varepsilon_k)\| \leq \|d^k\|$. Hence

$$\sum_{k \geq 0} \frac{\|g^a(x^k, \varepsilon_k)\|^4}{\|d^k\|^2} \leq \sum_{k \geq 0} \|g^a(x^k, \varepsilon_k)\|^2 < +\infty.$$

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 (9). This means

$$F^a(x^k + \rho^{-1}\alpha_k d^k, \varepsilon_{k+1}) - F^a(x^k, \varepsilon_k) > \sigma_1 \rho^{-1}\alpha_k \langle g^a(x^k, \varepsilon_k), d^k \rangle - \sigma_2 \rho^{-2}\alpha_k^2 \|d^k\|^2 + \varepsilon_k. \tag{9}$$

By the mean-value theorem and inequality (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 (5), we get

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

Substituting this inequality into (9), we get

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

Due to (7), we get

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

which implies

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

where $c = \frac{\rho}{\frac{1}{\lambda} + \sigma_2}$. From (4), we obtain

$$\sum_{k \geq 0} \left((1 - \sigma_1) \frac{\|g^a(x^k, \varepsilon_k)\|^4}{\|d^k\|^2} - \sqrt{\frac{2\varepsilon_k}{\lambda}} \frac{\|g^a(x^k, \varepsilon_k)\|^2}{\|d^k\|} \right) < +\infty.$$

Since

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

the second inequality is due to (2). Hence (8) holds. □

Theorem 1

Assume the conditions of Lemma 3.1 hold. We have

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0.$$

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

$$\|g_k\| \geq \epsilon, \quad \forall k \geq 0. \tag{10}$$

From (7), we obtain there exists a constant $\epsilon_* > 0$ such that

$$\|g^a(x^k, \varepsilon_k)\| \geq \epsilon_*, \quad \forall k \geq 0. \tag{11}$$

We get from (8) that

$$\|d^k\|^2 = (\beta_k^{FR})^2 \|d^{k-1}\|^2 - 2\theta_k \langle d^k, g^a(x^k, \varepsilon_k) \rangle - \theta_k^2 \|g^a(x^k, \varepsilon_k)\|^2.$$

Dividing both sides of this equality by $\langle g^a(x^k, \varepsilon_k), d^k \rangle^2$, we get from Proposition 2 and (11) that

$$\begin{aligned} \frac{\|d^k\|^2}{\|g^a(x^k, \varepsilon_k)\|^4} &= \frac{\|d^k\|^2}{\langle g^a(x^k, \varepsilon_k), d^k \rangle^2} \\ &= (\beta_k^{FR})^2 \frac{\|d^{k-1}\|^2}{\langle g^a(x^k, \varepsilon_k), d^k \rangle^2} - \frac{2\theta_k}{\langle d^k, g^a(x^k, \varepsilon_k) \rangle} - \frac{\theta_k^2 \|g^a(x^k, \varepsilon_k)\|^2}{\langle g^a(x^k, \varepsilon_k), d^k \rangle^2} \\ &= \frac{\|d^{k-1}\|^2}{\|g^a(x^{k-1}, \varepsilon_{k-1})\|^4} - \frac{(\theta_k - 1)^2}{\|g^a(x^k, \varepsilon_k)\|^2} + \frac{1}{\|g^a(x^k, \varepsilon_k)\|^2} \\ &\leq \frac{\|d^{k-1}\|^2}{\|g^a(x^{k-1}, \varepsilon_{k-1})\|^4} + \frac{1}{\|g^a(x^k, \varepsilon_k)\|^2} \\ &\leq \sum_{i=0}^{k-1} \frac{1}{\|g^a(x^i, \varepsilon_i)\|^2} \leq \frac{k}{\epsilon_*^2}. \end{aligned}$$

The last inequality implies that

$$\sum_{k \geq 0} \frac{\|g^a(x^k, \varepsilon_k)\|^4}{\|d^k\|^2} \geq \epsilon_*^2 \sum_{k \geq 1} \frac{1}{k} = +\infty.$$

which contradicts (8). The proof is then complete. □

4. Concluding remarks

In this paper, via Moreau-Yosida regularization, by introducing a new line search on the approximation to the Moreau-Yosida regularization, we extend MFR method for smooth unconstrained optimization to nonsmooth convex minimization. The global convergence is established under mild conditions.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.11401343) and the Science Foundation of China Three Gorges University (No KJ2013B017).

REFERENCES

1. A. Auslender, "Numerical methods for nondifferentiable convex optimization," *Mathematical Programming Study*, vol. 30, pp. 102-126, 1987.
2. 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.
3. 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.
4. X. Chen, and M. Fukushima, "Proximal quasi-Newton methods for nondifferentiable convex optimization," *Mathematical Programming*, vol. 85, pp. 313-334, 1999.
5. R. Correa and C. Lemarechal, "Convergence of some algorithms for convex minimization," *Mathematical Programming*, vol. 62, pp. 261-275, 1993.
6. M. Fukushima, "A descent algorithm for nonsmooth convex optimization," *Mathematical Programming*, vol. 30, pp. 163-175, 1984.
7. 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.
8. W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific Journal of Optimization*, vol. 2, pp. 35-58, 2006.
9. J.-B. Hiriart-Urruty and C. Lemarechal, "Convex Analysis and Minimization Algorithms," Springer Verlag, Berlin, Germany, 1993.
10. 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. Yuvoor, Springer Verlag, New York, NY, pp. 144-162, 1994.
11. 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.
12. S. Lu, Z. Wei and L. Li, "A trust region algorithm with adaptive cubic regularization methods for nonsmooth convex minimization," *Computational Optimization and Applications*, vol. 51, no. 2, pp. 551-573, 2012.
13. R. Mifflin, "A quasi-second-order proximal bundle algorithm," *Mathematical Programming*, vol. 73, pp. 51-72, 1996.
14. 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.
15. L. Qi and X. Chen, "A preconditioned proximal Newton method for nondifferentiable convex optimization," *Mathematical Programming*, vol. 76, pp. 411-429, 1997.

16. 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.
17. 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.
18. G. Yuan, Z. Wei, G. Li, "A modified Polak-Ribière-Polyak conjugate gradient algorithm for nonsmooth convex programs," *Journal of Computational and Applied Mathematics*, vol. 255, pp. 86-96, 2014.
19. L. Zhang, "A new trust region algorithm for nonsmooth convex minimization," *Applied Mathematics and Computation*, vol. 193, pp. 135-142, 2007.
20. L. Zhang, W. Zhou and D. Li, "Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search", *Numerische Mathematik*, vol. 104, pp. 561-572, 2006.