Journal of Mathematical Modeling
Vol. 9, No. 3, 2021, pp. 375-390

JMM

An efficient conjugate gradient method with strong convergence properties for non-smooth optimization

Fahimeh Abdollahi, Masoud Fatemi*

Department of Mathematics, K. N. Toosi University of Technology, Tehran, Iran

Email(s): fabdollahi@email.kntu.ac.ir, smfatemi@kntu.ac.ir

Abstract. In this paper, we introduce an efficient conjugate gradient method for solving nonsmooth optimization problems by using the Moreau-Yosida regularization approach. The search directions generated by our proposed procedure satisfy the sufficient descent property, and more importantly, belong to a suitable trust region. Our proposed method is globally convergent under mild assumptions. Our numerical comparative results on a collection of test problems show the efficiency and superiority of our proposed method. We have also examined the ability and the effectiveness of our approach for solving some real-world engineering problems from image processing field. The results confirm better performance of our method.

Keywords: Conjugate gradient method, nonsmooth optimization, global convergence, image processing.
AMS Subject Classification 2010: 90C06, 90C26, 65Y20.

1 Introduction

Consider the following optimization problem

$$\min_{x\in\Phi} f(x). \tag{1}$$

This model depending on the types of objective function f (smooth or nonsmooth), and the set Φ (discrete or continuous) involves a variety of optimization problems.

Optimization problems are appeared in many research fields such as engineering, management, economics, medicine, pharmacy, astronomy and so on. They are highly regarded and also there exist many effective ways to solve them. A challenging issue regarding to (1) is solving large-scale nonsmooth problems, such as those in optimal control, image processing and machine learning problems.

*Corresponding author.
Received: 6 June 2020 / Revised: 22 December 2020 / Accepted: 23 December 2020
DOI: 10.22124/jmm.2020.16747.1452

Image processing is an active field of research and has many applications in photography, medicine, astronomy, industry, etc. One of its main proposes is reconstruction of a real image from a blurred or corrupted version of it.

The most commonly used model for damaged images due to blurring and noise is the following linear model:

$$y = Ax + \eta, \qquad x \in \mathbb{X},$$

which is used to retrieve the original image x from y destroyed by a blurring matrix A and a noise vector η. Matrix A can be obtained by a certain Point Spread Function (PSF). There is typically a PSF corresponding to each kind of blurring [16].

This problem can be cast as the following optimization problem

$$\min_{x \in \mathbb{X}} \frac{1}{2} \| Ax - y \|_2^2 + \nu \phi(x),$$

where ν is a regularization parameter and ϕ is a regularization term. We need to shed more light on the critical role of regularization term, which is necessary to obtain a satisfactory solution. On the other hand, it determines the type of the resulting optimization problem. For example choosing $\phi(x) = \|x\|_1$ or $\phi(x) = \|x\|_2$ results in a nonsmooth or smooth problem, respectively.

As we know, image deblurring is a large-scale inverse problem which is very ill-conditioned and hard to solve. Therefore, designing new methods for solving this problem is still an active field of research.

In this paper, we investigate optimization problems like (1), where f is a nonsmooth convex function and Φ is $\mathbb{R}^n$ (unconstrained case).

We know that there are a lot of methods for solving (1) with continuously differentiable objective function such as Newton, Quasi-Newton and conjugate gradient methods. All the introduced methods find the optimal solution by generating descent directions using exact or inexact line search procedures.

Conjugate gradient methods have been extensively employed by researchers in recent decades due to their strong local and global convergence properties and also low memory requirements for solving large-scale problems. Conjugate gradient iterations are generally defined as follows:

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

where $\alpha_k > 0$ is a step length and the search direction d_k is computed by

$$d_{k+1} = -\, g_{k+1} + \beta_k d_k \quad (d_0 = -g_0), \tag{3}$$

recursively, where $g_k =: \nabla f(x_k)$. The scalar β_k known as the conjugate gradient parameter is indeed one of the most important parameters of these methods. In fact, various choices of β_k lead to different conjugate gradient algorithms. Some previous works in this field show that a proper choice of this parameter led us to a better numerical performance. Some famous classical choices of β_k can be found in Hestenes and Stiefel (HS) [17], Fletcher and Reeves (FR) [11], Polak and Ribiere and Polyak (PRP) [31], Fletcher (CD) [10], Liu and Storey (LS) [26] and Dai and Yuan (DY) [8] as shown below:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \qquad \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \qquad \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2},$$
$$\beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-d_k^T g_k}, \qquad \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}, \qquad \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}$$

where $y_k := g_{k+1} - g_k$ and $\| \cdot \|$ denotes the Euclidean norm.

Andrei [2], introduced 40 different types of these parameters, and thus, 40 different types of conjugate gradient algorithms. There are many articles that addressed different types of these methods and compared them numerically, see for example [1,3,7,15,21]. Although these methods originally have been developed for solving smooth optimization problems, some researchers have used them recently to solve nonsmooth convex optimization problems.

Some well known methods available for solving nonsmooth convex optimization problems are subgradient and bundle methods [22]. But, our approach here to cast these problems is transforming them to equivalent smooth problems via Moreau-Yosida regularization technique.

Consider

$$\min_{x \in \mathbb{R}^n} F(x), \tag{4}$$

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

$$F(x) = \min_{z \in \mathbb{R}^n} \{f(z) + \frac{1}{2\lambda}\|z - x\|^2\}, \tag{5}$$

where λ is a positive parameter. The set of minimizers of (1) coincide with the set of minimizers of (4). Fortunately, F is a differentiable convex function even when the function f is nondifferentiable [30]. There are various iterative methods for solving (4), for example see [4, 12, 13, 20, 24, 25, 27, 33–41, 43].

The good features of conjugate gradient methods for smooth problems encouraged us to modify these methods for nonsmooth problems.

The main features of the method presented in this paper are:

- All search directions satisfy sufficient descent condition, indicating that the function is decreasing.
- All search directions lie in a trust region, yielding some good convergence results.
- Our proposed method has global convergence under specific assumptions and the numerical results show a better performance of the method than some existing standard methods.

The organization of the paper is as follow: after a brief review in Section 2, we introduce the new method in Section 3. Then, we discuss its convergence in Section 4. Finally in Section 5, we present the two types of comparisons. Firstly, we report the numerical efficiency of the new algorithm for solving some small and large scale nonsmooth optimization problems using different choices of the main parameters. The details of iterations along with the Dolan-Moré performance profiles are presented. Secondly, we consider an image debluring problem and present its numerical comparisons.

2 A brief review

In the sequel, we describe the main features of the Moreau-Yosida regularization (5).

We note that the right-hand side of (5) is well defined when f is convex and while it is strongly convex, it has a unique minimizer which is denoted by

$$p(x) = \arg \min_{z \in \mathbb{R}^n} \{f(z) + \frac{1}{2\lambda}\|z - x\|^2\}. \tag{6}$$

Therefore, F can be expressed by

$$F(x) = f(p(x)) + \frac{1}{2\lambda}\|p(x) - x\|^2, \tag{7}$$

where its gradient is

$$g(x) = \frac{x - p(x)}{\lambda}. \tag{8}$$

$F(x)$ and $g(x)$ have remarkable properties. We refer interested readers to [13, 19, 23, 29, 32]. In continuation, we present some of their properties without proof.

Theorem 1. *The function F in (5) is finite-valued, convex, everywhere differentiable and its gradient is $g(x) = \frac{x-p(x)}{\lambda}$.Furthermore, for all $x_1, x_2 \in \mathbb{R}^n$*

$$\|g(x_1) - g(x_2)\|^2 \leq \frac{1}{\lambda}(g(x_1) - g(x_2))^T(x_1 - x_2).$$

Consequently, g is globally Lipschitz continuous with the constant term $\frac{1}{\lambda}$, namely

$$\|g(x_1) - g(x_2)\| \leq \frac{1}{\lambda}\|x_1 - x_2\|. \tag{9}$$

Finally, the following statements are equivalent:
(i) x is the minimizer of f.
(ii) x is the minimizer of F.
(iii) $g(x) = 0$.
(iv) $x = p(x)$.

Proof. see [18]. □

By Rademacher's theorem, $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is almost everywhere differentiable and the set

$$\partial_B g(x) = \{V \in \mathbb{R}^{n \times n} | V = \lim_{x_i \longrightarrow x} \nabla g(x_i), x_i \in D_g\}$$

is nonempty and compact, where $D_g = \{x \in \mathbb{R}^n | g \text{ is differentiable in } x\}$. We note that every $V \in \partial_B g(x)$ is a symmetric positive semidefinite matrix, because g is the gradient of a convex function F [40].

To compute $F(x)$ and $g(x)$, we need to estimate $p(x)$ in (6). There are some efficient procedures for computing $p(x)$. We refer interested readers to [5, 6, 12] for a clear explanation of these procedures along with a brief analysis of its computational cost.

In order to provide an efficient conjugate gradient algorithm, Fatemi [9] introduced an optimization problem by combining the three conditions

$$\begin{aligned} d_{k+1}^T y_k &= 0, \\ d_i^T g_k &= 0, \qquad i = 0, 1, \ldots, k-1 \qquad \text{and} \\ d_k^T g_k &\leqslant -\eta \|g_k\|^2, \quad \eta > 0 \text{ is a constant}, \end{aligned}$$

that are familiar in the linear conjugate gradient theory. The problem was

$$\min_{\beta_k} \big[g_{k+1}^T d_{k+1} + M \big(\big(g_{k+2}^T s_k \big)^2 + \big(d_{k+1}^T y_k \big)^2 \big) \big], \tag{10}$$

where $s_k := x_{k+1} - x_k$ and M is a penalty parameter. By solving this problem and using the secant condition $B_{k+1} s_k = y_k$, a new β_k was proposed as follows

$$\beta_k = \frac{-1}{2M(1+t^2)} \frac{g_{k+1}^T d_k}{(y_k^T d_k)^2} + \frac{y_k^T g_{k+1}}{y_k^T d_k} - \frac{t}{(1+t^2)} \frac{s_k^T g_{k+1}}{y_k^T d_k}, \tag{11}$$

where $t > 0$ is a suitable approximation of the step length α_k. The author showed that the resulting method is globally converged for general functions. Moreover, numerical comparisons reported in the paper showed that the method is efficient in the sense of Dolan-Moré performance profile.

3 Preliminary results

The good features of the method presented in [9] inspired us to modify it for solving nonsmooth problems. Considering $g_k = \nabla F(x_k)$, we define

$$d_{k+1} = -\, g_{k+1} + \beta_k^N d_k \quad (d_0 = -g_0), \tag{12}$$

where

$$\beta_k^N = (y_k - \frac{1}{2M(1+t^2)} \frac{d_k}{T_k} - \frac{t}{1+t^2} s_k)^T \frac{g_{k+1}}{T_k}, \tag{13}$$

and

$$T_k = \max\{\gamma \|d_k\|.\|y_k\|, \|d_k\|.\|s_k\|, |d_k^T y_k|, \frac{d_k^T g_{k+1}}{2M\big((1+t^2) y_k - t s_k\big)^T g_{k+1}}\} \geq 0, \tag{14}$$

for some constant $\gamma > 0$.

Equality (13) is our modified version of (11) suitable for nonsmooth problems as we will show as follows.

Lemma 1. *Consider conjugate gradient iterations based on (2) and (12) with any step length $\alpha_k > 0$ and β_k in (13). Then, for a fixed positive scalar $0 < c < 1$ and the choice of*

$$M = \frac{2c}{(1+t^2)\|y_k\|^2}, \tag{15}$$

in (13), we have

$$d_{k+1}^T g_{k+1} \leqslant -(1-c)\|g_{k+1}\|^2. \tag{16}$$

Proof. Using (12) and (13), we have

$$\begin{aligned} d_{k+1}^T g_{k+1} = &- \|g_{k+1}\|^2 + \beta_k^N d_k^T g_{k+1} \\ &= -\|g_{k+1}\|^2 + \frac{(y_k^T g_{k+1})(d_k^T g_{k+1})}{T_k} - \frac{1}{2M(1+t^2)} \times \frac{(d_k^T g_{k+1})^2}{T_k^2} \\ &- \frac{t}{1+t^2} \times \frac{(s_k^T g_{k+1})(d_k^T g_{k+1})}{T_k}. \end{aligned}$$

Since α_k and t are positive, we deduce that

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + (y_k^T g_{k+1}) \times \frac{d_k^T g_{k+1}}{T_k} - \frac{1}{2M(1+t^2)} \times \frac{(d_k^T g_{k+1})^2}{T_k^2}$$

Now, using the fact that $ab \leq \frac{l}{4}a^2 + \frac{1}{l}b^2$, where a, b and l are positive scalars, we have

$$d_{k+1}^T g_{k+1} \leq - \|g_{k+1}\|^2 + \frac{l}{4}(y_k^T g_{k+1})^2 + \frac{1}{l}\frac{(d_k^T g_{k+1})^2}{T_k^2} - \frac{1}{2M(1+t^2)} \times \frac{(d_k^T g_{k+1})^2}{T_k^2}.$$

Now, let $l = 2M(1+t^2)$, so

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \frac{M(1+t^2)}{2} \times (y_k^T g_{k+1})^2$$

Finally, the Cauchy-schwartz inequality implies that

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \frac{M(1+t^2)}{2}\|y_k\|^2\|g_{k+1}\|^2.$$

The proof is now completed using (15) and the above inequality. □

It is easy to see that replacing (15) in (13) and (14) yields

$$\beta_k^N = (y_k - \frac{\|y_k\|^2}{4c}\frac{d_k}{T_k} - \frac{t}{1+t^2}s_k)^T \frac{g_{k+1}}{T_k}, \tag{17}$$

where

$$T_k = \max\{\gamma\|d_k\|.\|y_k\|, \|d_k\|.\|s_k\|, |d_k^T y_k|, \frac{(1+t^2)\|y_k\|^2 d_k^T g_{k+1}}{4c\big((1+t^2)y_k - ts_k\big)^T g_{k+1}}\}. \tag{18}$$

Lemma 2. *For the search direction d_k introduced by (12) and (17), we have*

$$\|d_k\| \leq (2 + \frac{1}{\gamma} + \frac{1}{4c\gamma^2})\|g_k\|. \tag{19}$$

Proof. Using (12), (17) and the Cauchy-schwartz inequality, we have

$$\begin{aligned}
\|d_{k+1}\| &= \| - g_{k+1} + \beta_k^N d_k\| \leq \|g_{k+1}\| + |\beta_k^N|\|d_k\| \\
&= \|g_{k+1}\| + \|y_k - \frac{\|y_k\|^2}{4c}\frac{d_k}{T_k} - \frac{t}{1+t^2}s_k\|\frac{\|g_{k+1}\|}{T_k}\|d_k\| \\
&\leq \|g_{k+1}\| + \frac{\|y_k\|\|d_k\|\|g_{k+1}\|}{T_k} + \frac{\|y_k\|^2\|d_k\|^2\|g_{k+1}\|}{4cT_k^2} + \frac{t}{1+t^2}\frac{\|s_k\|\|d_k\|\|g_{k+1}\|}{T_k}.
\end{aligned} \tag{20}$$

Moreover, using (18) we have

$$T_k \geq \gamma\|y_k\|\|d_k\| \Longrightarrow \frac{1}{T_k} \leq \frac{1}{\gamma\|y_k\|\|d_k\|}, \tag{21}$$

and

$$T_k \geq \|s_k\|\|d_k\| \Longrightarrow \frac{1}{T_k} \leq \frac{1}{\|s_k\|\|d_k\|}. \tag{22}$$

Finally, by replacing (21) and (22) in (20), we have

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \frac{1}{\gamma}\|g_{k+1}\| + \frac{1}{4c\gamma^2}\|g_{k+1}\| + \|g_{k+1}\| \Longrightarrow \|d_{k+1}\| \leq (2 + \frac{1}{\gamma} + \frac{1}{4c\gamma^2})\|g_{k+1}\|.$$

□

4 Global convergence

We are now in a position where we can sum up the contents of the previous sections to introduce our new conjugate gradient algorithm.

Algorithm 1. *New Conjugate Gradient Algorithm* (*NCG*)

- *Step1.* *Choose a starting point $x_0 \in \mathbb{R}^n$ and a suitable value for positive parameters λ, γ, δ, $0 < \sigma < 1$, $0 < c < 1$ and $0 < \epsilon < 1$. Compute $g_0 = \nabla F(x_0)$, set $d_0 = -g_0$ and $k = 0$.*

- *Step2.* *Check the stopping condition. if $\|g_k\| < \epsilon$ then stop; else go to step 3.*

- *Step3.* *Compute the step length α_k using the following Armijo-type line search:*

$$F(x_k + \alpha_k d_k) - F(x_k) \leq \sigma\alpha_k g_k^T d_k \tag{23}$$

  where $\alpha_k = \delta \times 2^{-i_k}$, $i_k \in \{0, 1, 2, \ldots\}$

- *Step4.* *Compute $x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = \nabla F(x_{k+1})$, $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$.*

- *Step5.* *Compute the conjugate gradient parameter* β_k^N *using (17) and (18).*
- *Step6.* *Compute the search direction* $d_{k+1} = -g_{k+1} + \beta_k^N d_k$.
- *Step7.* *Set* $k = k+1$ *and go to step 2.*

Our final task is investigation of the global convergence properties of algorithm (NCG). To do this, we consider the following necessary assumptions to prove convergence of the algorithm.

1. The level set $\mathit{\Gamma} = \left\{x \in \mathbb{R}^n \mid F(x) \leq F(x_0)\right\}$ is bounded.
2. The function F is bounded from below.
3. The sequence $\{V_i\}$ is bounded, that is, there is a constant L such that for each i

$$\|V_i\| \leq L, \tag{24}$$

where $V_i \in \partial_B g(x_i)$.

Lemma 3. *For the sequence* $\{x_k, \alpha_k\}$ *generated by algorithm NCG and in the presence of assumptions 1, 2 and 3, we have for sufficiently large* k*, that there exists constant* $\alpha_0 > 0$ *such that*

$$\alpha_k \geq \alpha_0 \tag{25}$$

Proof. Assume that α_k satisfies (23). If $\alpha_k = 1$ then the proof is completed. Otherwise, suppose that $\alpha_k^{'} = \alpha_k/2$. So we have

$$F(x_k + \alpha_k^{'} d_k) - F(x_k) > \sigma \alpha_k^{'} g_k^T d_k.$$

Using the Taylor expansion of F and Eq. (24) we conclude that

$$\begin{aligned}
\sigma \alpha_k^{'} g_k^T d_k &< F(x_k + \alpha_k^{'} d_k) - F(x_k) \\
&= \alpha_k^{'} d_k^T g_k + \frac{1}{2}(\alpha_k^{'})^2 d_k^T V(u_k) d_k \\
&\leq \alpha_k^{'} d_k^T g_k + \frac{L}{2}(\alpha_k^{'})^2 \|d_k\|^2
\end{aligned}$$

where $u_k = x_k + \theta \alpha_k^{'} d_k$ and $\theta \in (0,1)$. The upper bound on $V(u_k)$ comes from assumption 2 and the fact that the quadratic form $d^t V(x) d$ is generally convex for every d, because $V(x)$ is positive semidefinite for every x.

Now by replacing (16) and (19) in the above inequality we have

$$\begin{aligned}
\alpha_k^{'} &> \left[\frac{-(1-\sigma) d_k^T g_k}{\|d_k\|^2}\right] \frac{2}{L} \\
&> \left[\frac{(1-\sigma)(1-c)}{\left(2 + \frac{1}{\gamma} + \frac{1}{4c\gamma^2}\right)^2}\right] \frac{2}{L}.
\end{aligned}$$

Now, by considering $W = \left[\frac{(1-\sigma)(1-c)}{\left(2+\frac{1}{\gamma}+\frac{1}{4c\gamma^2}\right)^2}\right] \frac{2}{L}$, we have

$$\alpha_k > 2W.$$

The proof is completed by taking $\alpha_0 \in (0, 2W)$. $\square$

Now, it is the time to proof our main result.

Theorem 2. *Assume that the conditions in Lemma 3 hold. Then, we have*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{26}$$

Moreover, any accumulation point of x_k is an optimal solution of (1).

Proof. Assume that (26) does not hold. Then there exist $k_0 > 0$ and $\epsilon_0 > 0$ such that

$$\forall k > k_o : \quad \|g_k\| \geq \epsilon_0. \tag{27}$$

Using inequalities (16), (23), (25) and (27), for all $k > k_0$, we have

$$\begin{aligned} F(x_{k+1}) - F(x_k) &\leq \sigma\alpha_k d_k^T g_k \\ &\leq -(1-c)\sigma\alpha_k\|g_k\|^2 \\ &\leq -(1-c)\sigma\alpha_0\epsilon_0^2. \end{aligned} \tag{28}$$

Now inequality (28) and assumption 1 yield

$$\sum(1-c)\sigma\alpha_0\epsilon_0^2 < \infty.$$

It is clearly a contradiction due to the fact that $\sum(1-c)\sigma\alpha_0\epsilon_0^2 = \infty$. Then,

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

Now assume that x^* is an accumulation point of x_k. So there exists a subsequence x_{k_j} such that

$$\lim_{k_j\longrightarrow\infty} x_k = x^*.$$

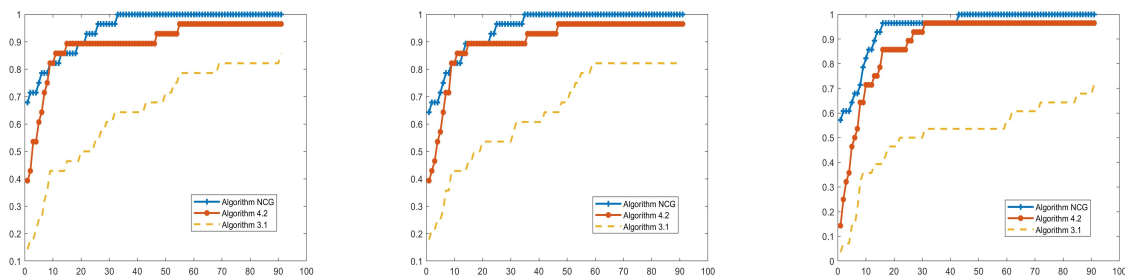
Therefore,

$$\|\nabla F(x^*)\| = \|g(x^*)\| = 0,$$

and since $g(x) = \nabla F(x) = \frac{x-p(x)}{\lambda}$, we have $x^* = p(x^*)$ and so by Theorem 1, x^* is an optimal solution for (1). □

5 Numerical experiments and comparisons

In this section, we investigate the numerical performance of Algorithm 1 (NCG) presented in section 4 by solving some unconstrained nonsmooth test problems. We have compared NCG by Algorithm 4.2 and Algorithm 3.1 in [38] and [37], respectively. Numerical results reported in [38] show that Algorithm 4.2 is more efficient than the MPRP method presented in [40]. Authors in [40] claim that MPRP method is more efficient than both the proximal bundle and bundle trust region methods. In addition, Algorithm 3.1 introduce a new three-term conjugate gradient method which is an interesting idea in this field. According to these facts, we have considered those algorithms in our comparisons.



(a) Total number of iterations (k). (b) Total number of function evaluations (i). (c) CPU time (t).

Figure 1: Performance profiles of these methods.

We initialized algorithms by setting $\lambda = 1$, $\gamma = 0.02$, $c = 0.02$, $\delta = 0.5$, $\sigma = 0.8$ and $t = 1$. The algorithms are terminated if either $\|g_k\| \leq 10^{-7}$ or the number of iterations are exceeded 3000. All tests were run on a 2.4 Intel Core(TM) i7-5500U processor computer with 8 GB of RAM and a 64-bit windows 10 operating system using MATLAB R2017b programming environment.

We considered small and large scale problems reported by [28] and [14] in our numerical tests. Tables 1 and 2 show the names, dimensions (Dim) and the optimal objective function value (f_{ops}) of test problems.

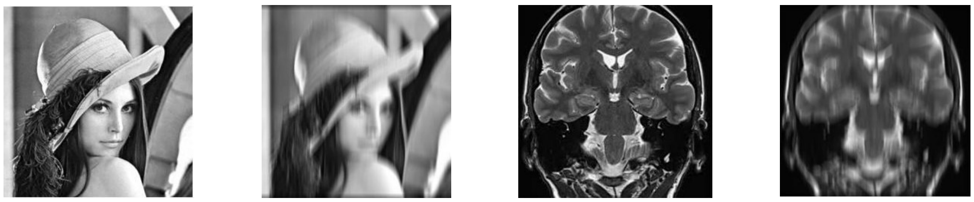
In Tables 3 and 4, we have reported the details of the results for NCG algorithm, Algorithm 4.2 and Algorithm 3.1. In both tables, k is the total number of iterations, i is the total number of function evaluations and t is the CPU time in second. Furthermore, f indicates the final function value and Dim demonstrates the problem dimension considered for comparisons.

The reported results express that the proposed algorithm can successfully solve all test problems. We can also see from Figure 1 that the NCG algorithm acts better than both Algorithm 4.2 and Algorithm 3.1 in the sense of Dolan-Moré performance profile. Therefore, NCG can be introduced as an acceptable and efficient way to solve nonsmooth problems.

We now investigate the efficiency of our algorithm for solving a highly regarded real word application from image processing field, namely, image debluring.

In figure 2, we have blurred 256×256, Lena, Brain, Cameraman and Megan images using Toeplitz matrices. In this figure, images (a) and (c) have been blurred with the linear motion of a camera by 15 pixels and images (e) and (g) have been blurred by the linear motion of a camera by 20 pixels. Then, using nonsmooth regularization term $\phi(x) = \| x\|_1$, the images have been recovered. According to [42] the best value of the regularization parameter ν is around 0.001 and the algorithm stops after 50 iterations.

It can be seen from Table 5 and Figures 3, 4 and 5 that all algorithms can successfully solve image restoration problem, and we can see less CPU time for NCG in comparison with other two algorithms. As a result, our method with the proper choices of parameters is also effective for solving real world problems.



(a) (b) (c) (d)

(e) (f) (g) (h)

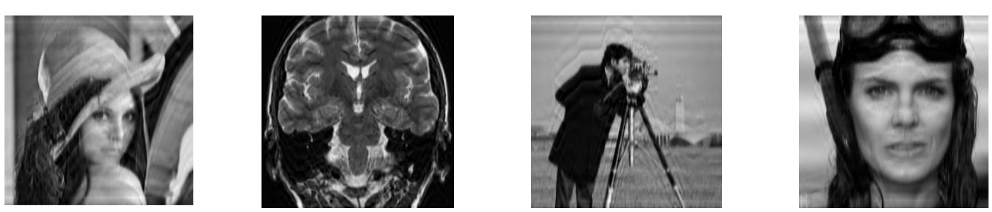
Figure 2: Original and blurred/noisy images.

(a) $t = 658.985$ (b) $t = 766.233$ (c) $t = 651.095$ (d) $t = 641.430$

Figure 3: Deblurring image with Algorithm NCG.

(a) $t = 835.226$ (b) $t = 688.648$ (c) $t = 723.232$ (d) $t = 747.763$

Figure 4: Deblurring image with Algorithm 4.2.



(a) $t = 3792.801$ (b) $t = 1526.456$ (c) $t = 1814.476$ (d) $t = 2854.751$

Figure 5: Deblurring image with Algorithm 3.1.

Table 1: Small-scale test problems [28].

Nr.	Problem's Name	Dim	f_{ops}
1	Crescent	2	0
2	CB2	2	1.952245
3	CB3	2	2
4	DEM	2	-3
5	QL	2	7.20
6	LQ	2	-1.4142136
7	Mifflin1	2	-1
8	Mifflin2	2	-1
9	Wolf	2	-8
10	Rosen-Suzuki	4	-44

Table 2: Large-scale test problems [14].

Nr.	Problem's Name
1	Generalization of MAXQ (convex)
2	Generalization of MAXHILB (convex)
3	Number of active faces (nonconvex)
4	Nonsmooth generalization of Brown function (nonconvex)
5	Chained cresent1 (nonconvex)
6	Chained cresent2 (nonconvex)

Table 3: Test results for small-scale problems.

Nr.	Algorithm NCG				Algorithm 4.2				Algorithm 3.1			
	k	i	t	f	k	i	t	f	k	i	t	f
1	86	463	21.525	6.5472310^(-4)	156	917	52.677	0.74999	106	530	22.184	0.74999
2	6	66	3.622	1.95350	6	66	7.765	1.95350	381	4478	207.447	1.95378
3	13	188	10.389	2.00145	13	188	10.919	2.00145	171	2121	134.351	2.00149
4	16	124	9.224	-2.99864	18	120	8.213	-2.99905	16	95	4.883	-2.99945
5	12	93	16.022	7.20362	12	93	16.685	7.20362	12	93	15.585	7.20362
6	14	98	4.74	-1.41388	14	98	5.39	-1.41388	14	98	4.852	-1.41388
7	21	170	289.819	-0.99923	409	4958	288.417	-0.99996	133	967	81.697	-0.99967
8	14	111	6.753	-0.99975	12	81	10.894	-0.99954	12	81	10.963	-0.99954
9	18	311	30.075	-7.99290	114	1739	106.909	-7.99201	139	2101	310.549	-7.99551
10	29	872	52.301	-43.97174	7	200	36.121	-43.94660	87	2831	110.121	-43.9049

6 Conclusion

In many top-notch domains like image and medical image processing, new mathematical methods that facilitate and speed up the processing step are highly appreciated. We introduced an efficient conjugate gradient method to solve nonsmooth optimization problems by using the Moreau-Yosida regularization technique. Our approach guarantees sufficient descent property and with mild assumptions has global convergence. In final section, we showed how our proposed method has higher performance and outperforms the older methods in this field. An implemen-

Table 4: Test results for large-scale problems.

Nr.	Dim	Algorithm NCG			Algorithm 4.2			Algorithm 3.1		
		k	i	t	k	i	t	k	i	t
1	1000	15	60	3.169	26	104	5.009	26	104	2.331
	5000	20	80	22.532	32	128	30.621	32	128	31.231
	10000	43	172	137.387	67	268	213.279	67	268	207.711
2	1000	9	70	22.696	7	55	12.637	31	270	118.749
	5000	10	87	438.84	12	103	540.842	41	441	4931.256
	10000	24	199	4021.326	7	60	788.047	25	236	10320
3	1000	11	55	6.94	11	55	7.481	11	55	6.559
	5000	13	65	52.223	13	65	53.426	13	65	53.158
	10000	13	64	212.012	13	64	212.731	13	64	209.936
4	1000	22	110	7.215	36	190	9.619	74	451	49.339
	5000	38	194	54.29	51	264	78.601	379	1979	1668.019
	10000	39	197	230.627	57	292	364.195	1280	6589	5301.382
5	1000	18	92	10.129	18	92	11.263	29	231	26.921
	5000	17	93	49.856	17	93	56.576	99	586	400.038
	10000	16	95	163.977	16	95	171.754	98	583	1626.941
6	1000	9	36	2.359	9	36	2.396	11	44	6.069
	5000	13	52	29.658	13	52	15.064	37	148	105.763
	10000	11	44	37.822	10	40	92.828	5	20	64.555

Table 5: CPU time comparisons.

	Lena	Brain	Cameraman	Megan
Algorithm NCG	658.985	766.233	651.095	641.430
Algorithm 4.2	835.226	688.648	723.232	747.763
Algorithm 3.1	3792.801	1526.456	1814.476	2854.751

tation of our proposed approach also employed to solve image deblurring problem. Our reported results showed that the proposed method is efficient in the sense of CPU time comparisons.

Acknowledgements

The authors thank the Research Council of K.N. Toosi University of Technology for supporting this work.

References

[1] N. Andrei, *Numerical comparison of conjugate gradient algorithms for unconstrained optimization*, Stud. Inform. Control. **16** (2007) 333–352.

[2] N. Andrei,*Conjugate gradient algorithms for unconstrained optimization. A survey on their definition,* Tech. rep., ICI Technical Report, 2008.

[3] N. Andrei, *A numerical study on efficiency and robustness of some conjugate gradient algorithms for large-scale unconstrained optimization*, Stud. Inform. Control. **22** (2013)259–284.

[4] X. Chen, M. Fukushima, *Proximal quasi-newton methods for nondifferentiable convex optimization*, Math. Program. **85** (1999) 313–334.

[5] A.R. Conn, N.I. Gould, P.L. Toint, *Trust Region Methods*, SIAM, Philadelphia, 2000.

[6] R. Correa, C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Math. Program. **62** (1993) 261–275.

[7] Y.H. Dai, L.Z. Liao, *New conjugacy conditions and related nonlinear conjugate gradient methods*, Appl. Math. Optim. **43** (2001) 87–101.

[8] Y.H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim. **10** (1999) 177–182.

[9] M. Fatemi, *A new efficient conjugate gradient method for unconstrained optimization*, J. Comput. Appl. Math. **300** (2016) 207–216.

[10] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, 2013.

[11] R. Fletcher, C.M. Reeves, *Function minimization by conjugate gradients*, Comput. J. **7** (1964) 149–154.

[12] M. Fukushima, *A descent algorithm for nonsmooth convex optimization*, Math. Program. **30** (1984) 163–175.

[13] M. Fukushima, L. Qi, *A globally and superlinearly convergent algorithm for nonsmooth convex minimization*, SIAM J. Optim. **6** (1996) 1106–1120.

[14] M. Haarala, K. Miettinen, M.M Mäkelä, *New limited memory bundle method for large-scale nonsmooth optimization*, Optim. Methods Softw. **19** (2004) 673–692.

[15] W.W. Hager, H. Zhang,*A survey of nonlinear conjugate gradient methods*, Pac. J. Optim. **2** (2006) 35–58.

[16] P.C. Hansen, J.G. Nagy, D.P. O'leary, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006.

[17] M.R. Hestenes, E. Stiefel, *Methods of Conjugate Gradients for Solving Linear Systems*, NBS Washington, DC, 1952.

[18] J.B. Hiriart-Urruty, C. Lemaréchal, *Acceleration of the Cutting-Plane Algorithm: Primal Forms of Bundle Methods*, Springer, 1993.

[19] J.B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms I: Fundamentals*, Springer science & business media, 2013.

[20] Y. Hu, *Multivariate spectral gradient algorithm for nonsmooth convex optimization problems*, E Math. Probl. Eng. **2015** (2015) 145323.

[21] X. Jiang, J. Jian, *Improved Fletcher–Reeves and Dai–Yuan conjugate gradient methods with the strong Wolfe line search*, J. Comput. Appl. Math. **348** (2019) 525–534.

[22] N. Karmitsa, A. Bagirov, M.M. Mäkelä, *Comparing different nonsmooth minimization methods and software*, Optim. Methods Softw. **27** (2012) 131–153.

[23] C. Lemaréchal, C. Sagastizábal, *Practical aspects of the Moreau–Yosida regularization: Theoretical preliminaries*, SIAM J. Optim. **7** (1997) 367–385.

[24] Q. Li, *Conjugate gradient type methods for the nondifferentiable convex minimization*, Optim. Lett **7** (2013) 533–545.

[25] Li, *A modified Fletcher-Reeves-type method for nonsmooth Convex Minimization*, Stat. Optim. Inform. Comput. **2** (2014) 200–210.

[26] Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms, part 1: Theory*, J. Optim. Theory Appl. **69** (1991) 129–137.

[27] S. Lu, Z. Wei, L. Li, *A trust region algorithm with adaptive cubic regularization methods for nonsmooth convex minimization*, Comput. Optim. Appl. **51** (2012) 551–573.

[28] L. Lukšan, J. Vlcek, *Test problems for nonsmooth unconstrained and linearly constrained optimization*, Technická zpráva, 2000.

[29] F. Meng, G. Zhao, *On second-order properties of the Moreau–Yosida regularization for constrained nonsmooth convex programs*, Numer. Funct. Anal. Optim. **25** (2004) 515–529.

[30] N. Parikh, S. Boyd, *Proximal algorithms*, Foundations Trends. Optim. **1** (2013) 123–231.

[31] B.T. Polyak, *The conjugate gradient method in extremal problems*, Comput. Math. Math. Phys. **9** (1969) 94–112.

[32] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res. **18** (1993) 227–244.

[33] A. Rauf, M. Fukushima, *Globally convergent BFGS method for nonsmooth convex optimization1*, J. Optim. Theory Appl. **104** (2000) 539–558.

[34] N. Sagara, M. Fukushima, *A trust region method for nonsmooth convex optimization*, Management **1** (2005) 171–180.

[35] Z. Wei, L. Ql, *Convergence analysis of a proximal newton method: Analysis of a proximal Newton method*, Numer. Funct. Anal. Optim. **17** (1996) 463–472.

[36] T.G. Woldu, H. Zhang, Y.H. Fissuh, *A modified nonlinear conjugate gradient algorithm for large-scale nonsmooth convex optimization*, J. Optim. Theory Appl. **185** (2020) 1–16.

[37] G. Yuan, T. Li, W. Hu, *A conjugate gradient algorithm and its application in large-scale optimization problems and image restoration*, J. Inequal. Appl. **1** (2019) 247–258.

[38] G. Yuan, Z. Meng, Y. Li, *A modified Hestenes and Stiefel conjugate gradient algorithm for large-scale nonsmooth minimizations and nonlinear equations*, J. Optim. Theory Appl. **168** (2016) 129–152.

[39] G. Yuan, Z. Wei, *The Barzilai and Borwein gradient method with nonmonotone line search for nonsmooth convex optimization problems*, Math. Model. Anal. **17** (2012) 203–216.

[40] G. Yuan, Z. Wei, G. Li, *A modified Polak–Ribière–Polyak conjugate gradient algorithm for nonsmooth convex programs*, J. Comput. Appl. Math. **255** (2014) 86–96.

[41] G. Yuan, Z. Wei, Z. Wang, *Gradient trust region algorithm with limited memory BFGS update for nonsmooth convex minimization*, Comput. Optim. Appl. **54** (2013) 45–64.

[42] L.U. Yuzhen, *Out-of-focus blur: Image deblurring*, arXiv:1710.00620, 2017.

[43] L. Zhang, *A new trust region algorithm for nonsmooth convex minimization*, Appl. Math. Comput. **193** (2007) 135–142.