Smoothing Nonlinear Conjugate Gradient Method for Image Restoration using Nonsmooth Nonconvex Minimization

Xiaojun Chen* and Weijun Zhou†

7 November 2008, Revised 7 May 2009, 15 April 2010

Abstract

Image restoration problems are often converted into large-scale, nonsmooth and nonconvex optimization problems. Most existing minimization methods are not efficient for solving such problems. It is well-known that nonlinear conjugate gradient methods are preferred to solve large-scale smooth optimization problems due to their simplicity, low storage, practical computation efficiency and nice convergence properties. In this paper, we propose a smoothing nonlinear conjugate gradient method where an intelligent scheme is used to update the smoothing parameter at each iteration and guarantees that any accumulation point of a sequence generated by this method is a Clarke stationary point of the nonsmooth and nonconvex optimization problem. Moreover, we present a class of smoothing functions and show their approximation properties. This method is easy to implement without adding any new variables. Three image restoration problems with different pixels and different regularization terms are used in numerical tests. Experimental results and comparison with the continuation method in [M.Nikolova et al, SIAM J. Imaging Sciences, 1(2008), pp.2-25] show the efficiency of the proposed method.

Keywords. Image restoration, regularization, nonsmooth and nonconvex optimization, nonlinear conjugate gradient method, smooth approximation, potential function

AMS subject classification. 65F22, 65F10, 65K05

1 Introduction

The image restoration problem is that of reconstructing an image of an unknown scene from an observed image. This problem plays an important role in medical sciences, biological engineering and other areas of science and engineering [1, 4, 31]. The most common image degradation model can be represented by the following system:

$$b = Ax + \eta, \tag{1.1}$$

where $\eta \in R^m$ represents the noise, A is an $m \times n$ blurring matrix, $x \in R^n$ and $b \in R^m$ are the underlying and observed images respectively. In many cases, A is a matrix of block

*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China (maxjchen@polyu.edu.hk). This author's work was supported in part by a Hong Kong Research Grant Council and a HK PolyU grant.

†College of Mathematics and Computational Science, Changsha University of Science and Technology, Changsha 410076, China (weijunzhou@126.com). Current address: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. This author's work was supported by the Hong Kong Polytechnic University Postdoctoral Fellowship Scheme and the NSF foundation (10701018 and 10771057) of China.

Toeplitz with Toeplitz blocks (BTTB) when zero boundary conditions are applied and block Toeplitz-plus-Hankel with Toeplitz-plus-Hankel blocks (BTHTHB) when Neumann boundary conditions are used [22].

Technically we are not solving (1.1) since η is unknown. We are instead solving

$$\min_{x\in R^n} \|b - Ax\|.$$

Solving this problem alone will not get a satisfactory solution since the system is very sensitive to the noise and lack of imformation. The following smooth least square problem:

$$\min_{x\in R^n} \|b - Ax\|_2^2 + \beta\|Dx\|_2^2,$$

is often used, where D is an operator and β is the regularization parameter that controls the trade-off between the data-fitting term and the regularization term. For the regularization term, there has been a growing interest in using l_1 norm [6, 15, 23]. The l_1 solution tends to have better statistical properties than the l_2 solution. In [13], Fu et al considered the mixed $l_2 - l_1$ norm form:

$$\min_{x\in R^n} \|b - Ax\|_2^2 + \beta\|Dx\|_1, \tag{1.2}$$

and the $l_1 - l_1$ norm form:

$$\min_{x\in R^n} \|b - Ax\|_1 + \beta\|Dx\|_1. \tag{1.3}$$

These two minimization problems are convex but nonsmooth. In [24], Nikolova et al considered the following more general form:

$$\min_{x\in R^n} \Theta(b - Ax) + \beta\Phi(x), \tag{1.4}$$

where Θ forces closeness to data and Φ embodies the priors. The mixed $l_2 - l_1$ norm form (1.2) and the $l_1 - l_1$ norm form (1.3) are special forms of (1.4). Minimization methods for these forms were proposed in [13]. However, (1.4) can be nonconvex and nonsmooth. A class of regularization functions is of the form

$$\Phi(x) = \sum_{i=1}^{r} \varphi(d_i^T x),$$

where φ is called a potential function and $\{d_1, \cdots, d_r\}$ is a set of vectors of R^n. The role of Φ is to push the solution to exhibit some priori expected features, such as the presence of edges, smooth regions, and textures. As proven in [23], although convex potential functions such as $\varphi(t) = |t|$ are often used for the regularization term, nonconvex regularization functions such as $\varphi(t) = \dfrac{\alpha|t|}{1+\alpha|t|}$ with $\alpha > 0$ provides better possibilities for restoring images with neat edges. For this reason, many image restoration problems are often converted into nonsmooth, nonconvex optimization problems. Moreover, the optimization problems are large-scale because the discretized scenes usually have a large number $n = l \times l$ of pixels. Several efficient algorithms for image restoration problems are proposed in [13, 24], which use linear or quadratic programming reformulation and interior point methods. Fu et al [13] considered nonsmooth and convex problems, and Nikolova et al [24] considered a continuation method for nonsmooth

and nonconvex problems for arbitrary A in (1.4). The continuation method proposed in [24] is to approximate the minimizer of the objective function. However, there is no guarantee for the convergence of the continuation method. A drawback of these methods is the use of $4n + 2m$ additional variables, which makes these methods impractical for solving large-scale problems. The aim of this paper is to present an efficient optimization method for large-scale nonsmooth, nonconvex image restoration problems. This method ensures that from any starting point in R^n, any accumulation point of the sequence generated by the method is a Clarke stationary point. Moreover, this method does not increase the dimension, which is important for large-scale problems.

For convenience, in this paper, we first consider the following nonsmooth and nonconvex optimization problem in an abstract form:

$$\min_{x \in R^n} f(x). \tag{1.5}$$

Although large-scale nonsmooth and nonconvex optimization problems occur frequently in practice [3, 20, 32], efficient existing methods are rare. Burke et al [3] introduced a robust gradient sampling algorithm for solving nonsmooth, nonconvex unconstrained minimization problem. Kiwiel [19] slightly revised this algorithm and showed that any accumulation point generated by the algorithm is a Clarke stationary point with probability one. Encouraging numerical results for some small problems are reported in [3, 19].

It is well-known that nonlinear conjugate gradient methods such as the Polak-Ribière-Polyak (PRP) method [26, 27] are very efficient for large-scale smooth optimization problems due to their simplicity and low storage. Moreover, nonlinear conjugate gradient methods such as the PRP+ method [17], and conjugate gradient methods with suitable line search [16, 35, 36] are proposed, for nonconvex minimization problems which ensure any accumulation point generated by the algorithm is a stationary point. However, nonlinear conjugate gradient methods for solving nonsmooth optimization have not been studied. Moreover, we notice that most models of image restoration have some symmetric character in the Hessian matrix at points where the objective function is differentiable. The methods in [16, 35, 36] do not have the symmetric feature. To develop an efficient optimization method for nonsmooth and nonconvex minimization problems arising from image restoration, we first present a globally convergent nonlinear conjugate gradient method for smooth nonconvex minimization problems where the search direction can be presented by the gradient with a symmetric and uniformly positive definite matrix. Next, we extend the method to solve nonsmooth and nonconvex optimization by adopting smoothing functions.

This paper is organized as follows. In the next section, we present a globally convergent smoothing method for nonsmooth and nonconvex minimization problems. To present our approach clearly, we first give a smooth version of this method and prove the convergence for the case where f is differentiable. In Section 3, we present a class of smoothing functions and show their nice approximation properties for image restoration. Moreover, we show that all assumptions for the convergence of the smoothing conjugate gradient method hold and any accumulation point generated by the method is a Clarke stationary point, when the method is applied to several common models of image restoration. In section 4, we present numerical results for three images with $n \times n$ ($n = 128 \times 128$ to 256×256) blurring matrices to show the effectiveness and the efficiency of the proposed method. Moreover, comparison with the

continuation method in [24] is reported. Numerical results show that the continuation method in [24] can find smaller function values, and our method can obtain better psnr values with less CPU time.

Throughout the paper, $\|\cdot\|$ denotes the l_2 norm and $\|\cdot\|_1$ denotes the l_1 norm. R_+ denotes the set of all nonnegative real numbers. R_{++} denotes the set of all positive real numbers.

2 Algorithms description

In this paper, we consider the general iterative scheme for solving (1.5):

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1 \cdots, \tag{2.1}$$

where stepsize α_k is a positive scalar and d_k is a search direction given by some formula. In order to describe our algorithms conveniently, we divide this section into two subsections for two cases: (a) f is smooth and nonconvex; (b) f is nonsmooth and nonconvex.

2.1 Smooth case

Based on the CG-Descent method in [16], the three-term descent method in [35], the symmetry of the limited memory BFGS method (L-BFGS) [25, 30] and the modified BFGS method [21], we propose a new symmetric descent nonlinear conjugate gradient method which converges to a stationary point from any starting point for nonconvex minimization problems. We consider the search direction

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k d_{k-1} + \theta_k z_{k-1}, & \text{if} \quad k > 0, \end{cases} \tag{2.2}$$

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k and

$$\beta_k = \frac{g_k^T z_{k-1}}{d_{k-1}^T z_{k-1}} - \frac{2\|z_{k-1}\|^2 g_k^T d_{k-1}}{(d_{k-1}^T z_{k-1})^2}, \qquad \theta_k = \frac{g_k^T d_{k-1}}{d_{k-1}^T z_{k-1}}, \tag{2.3}$$

$$z_{k-1} = y_{k-1} + t_k s_{k-1}, \qquad t_k = \varepsilon_0 \|g_k\|^r + \max\Big\{0, -\frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}\Big\}, \tag{2.4}$$

with

$$y_{k-1} = g_k - g_{k-1}, \qquad s_{k-1} = x_k - x_{k-1} = \alpha_{k-1} d_{k-1},$$

and for some constants $\varepsilon_0 > 0$ and $r \geq 0$.

We can claim that (2.2) is well-defined, that is, d_k is finite-valued for $g_{k-1} \neq 0$ and $\alpha_k > 0$. Indeed, since we have from (2.4) that

$$\frac{s_{k-1}^T z_{k-1}}{\|s_{k-1}\|^2} = \varepsilon_0 \|g_k\|^r + \frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2} + \max\Big\{0, -\frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}\Big\} \geq \varepsilon_0 \|g_k\|^r > 0, \tag{2.5}$$

which implies

$$d_{k-1}^T z_{k-1} = \frac{s_{k-1}^T z_{k-1}}{\alpha_{k-1}} > 0.$$

Hence β_k and θ_k are finite-valued and thus d_k is finite-valued.

It is worth noticing that (2.1)-(2.2) can be written as a Newton-like method

$$x_{k+1} = x_k - \alpha_k H_k g_k,$$

where $H_0 = I$ and for $k \geq 1$

$$H_k = I - \frac{d_{k-1}z_{k-1}^T + z_{k-1}d_{k-1}^T}{d_{k-1}^T z_{k-1}} + \frac{2\|z_{k-1}\|^2}{(d_{k-1}^T z_{k-1})^2} d_{k-1}d_{k-1}^T. \tag{2.6}$$

This means that the search direction d_k defined in (2.2) can be given as

$$d_k = -H_k g_k. \tag{2.7}$$

This can be verified as follows. It is obviously true for $k = 0$. For $k > 0$, we have

$$\begin{aligned} -H_k g_k &= -g_k + \frac{d_{k-1}z_{k-1}^T g_k + z_{k-1}d_{k-1}^T g_k}{d_{k-1}^T z_{k-1}} - \frac{2\|z_{k-1}\|^2}{(d_{k-1}^T z_{k-1})^2} d_{k-1}d_{k-1}^T g_k \\ &= -g_k + \Big(\frac{z_{k-1}^T g_k}{d_{k-1}^T z_{k-1}} - \frac{2\|z_{k-1}\|^2 d_{k-1}^T g_k}{(d_{k-1}^T z_{k-1})^2}\Big) d_{k-1} + \frac{d_{k-1}^T g_k}{d_{k-1}^T z_{k-1}} z_{k-1} \\ &= -g_k + \beta_k d_{k-1} + \theta_k z_{k-1} \\ &= d_k = \frac{1}{\alpha_k}(x_{k+1} - x_k). \end{aligned}$$

Remark 2.1 If f is strictly convex, we can choose $\varepsilon_0 = 0$. Then from $s_{k-1}^T y_{k-1} = (x_k - x_{k-1})^T(g_k - g_{k-1}) > 0$, we have that $t_k \equiv 0$, and H_k has the following well-defined version

$$H_k = I - \frac{d_{k-1}y_{k-1}^T + y_{k-1}d_{k-1}^T}{d_{k-1}^T y_{k-1}} + \frac{2\|y_{k-1}\|^2}{(d_{k-1}^T y_{k-1})^2} d_{k-1}d_{k-1}^T. \tag{2.8}$$

Moreover, the following equality holds

$$-H_k g_k = -g_k + \Big(\frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} - \frac{2\|y_{k-1}\|^2 d_{k-1}^T g_k}{(d_{k-1}^T y_{k-1})^2}\Big) d_{k-1} + \frac{d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} y_{k-1}.$$

In this case, if the exact line search ($\alpha_{k-1} = \text{argmin}_{\alpha>0} f(x_{k-1} + \alpha d_{k-1})$) is used, then we have $\nabla_\alpha f(x_{k-1} + \alpha d_{k-1}) = g_k^T d_{k-1} = 0$, and thus

$$-H_k g_k = -g_k + \frac{y_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} d_{k-1}.$$

This is the well-known Hestenes-Stiefel conjugate gradient method and satisfies the conjugacy condition $d_k^T y_{k-1} = (-H_k g_k)^T y_{k-1} = 0$, see [17]. Therefore, we call this method (2.1)-(2.2) a nonlinear conjugate gradient method as it reduces to the standard conjugate gradient method when f is strictly convex and the exact line search is used. However, for nonconvex functions, (2.8) is not well-defined, because $d_{k-1}^T y_{k-1} = 0$ can happen. It is significate to introduce the additional term $t_k s_{k-1}$ in z_{k-1} and use $d_{k-1}^T z_{k-1}$ instead of $d_{k-1}^T y_{k-1}$, which not only ensures that H_k is well-defined but also possesses the property (2.5).

The next lemma lists an important property of H_k defined by (2.6).

Lemma 2.1. *Matrices H_k are symmetric positive definite and satisfy*

$$\|H_k^{-1}\| \le 2, \qquad k \ge 1. \tag{2.9}$$

Proof Obviously H_k is symmetric. Moreover for any $k \ge 1$ and any $x \in R^n$, we have

$$\begin{aligned} x^T H_k x &= x^T x - \frac{2(x^T d_{k-1})(z_{k-1}^T x)}{d_{k-1}^T z_{k-1}} + \frac{2\|z_{k-1}\|^2 (x^T d_{k-1})^2}{(d_{k-1}^T z_{k-1})^2} \\ &= x^T x - 2\Big(\frac{\sqrt{2}x^T d_{k-1}}{d_{k-1}^T z_{k-1}} z_{k-1}\Big)^T \Big(\frac{1}{\sqrt{2}} x\Big) + \frac{2\|z_{k-1}\|^2 (x^T d_{k-1})^2}{(d_{k-1}^T z_{k-1})^2} \\ &\ge x^T x - \Big(\Big\|\frac{\sqrt{2}x^T d_{k-1}}{d_{k-1}^T z_{k-1}} z_{k-1}\Big\|^2 + \|\frac{1}{\sqrt{2}} x\|^2\Big) + \frac{2\|z_{k-1}\|^2 (x^T d_{k-1})^2}{(d_{k-1}^T z_{k-1})^2} \\ &= x^T x - \frac{1}{2} x^T x = \frac{1}{2}\|x\|^2. \end{aligned}$$

Hence H_k is positive definite and its smallest eigenvalue is greater than $\frac{1}{2}$, which gives (2.9). □

Lemma 2.1 ensures that H_k are symmetric positive definite and satisfy $\|H_k g_k\| \ge \frac{1}{2}\|g_k\|$. Existing nonlinear conjugate gradient methods [16, 21, 25, 30, 35] have no such nice property. For instance, in the conjugate gradient method proposed in [16], the search direction is given by

$$d_k = -H_{HZ_k} g_k, \qquad \text{where} \quad H_{HZ_k} = I - \frac{1}{d_{k-1}^T y_{k-1}} d_{k-1} y_{k-1}^T + \frac{2\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} d_{k-1} d_{k-1}^T.$$

Obviously, H_{HZ_k} is not symmetric.

Following directly from the above lemma and (2.7), we find that the search direction is a descent direction.

Lemma 2.2. *Let $\{x_k\}$ and $\{d_k\}$ be generated by the method (2.1) and (2.2). We have*

$$d_k^T g_k \le -\frac{1}{2}\|g_k\|^2, \qquad k \ge 0. \tag{2.10}$$

Now we present an algorithm for smooth and noncovex minimization problems.

Algorithm 2.1.

Step 0. Choose constants $\varepsilon_0 > 0$, $r \ge 0$. Choose $\delta \in (0, \sigma), \sigma \in (\delta, 1), \rho \in (0, 1)$ and initial point $x_0 \in R^n$. Let $k := 0$.

Step 1. Compute d_k by (2.2)–(2.4) with the chosen parameters $\varepsilon_0 > 0$, $r \ge 0$.

Step 2. Determine α_k by the Armijo line search, that is, $\alpha_k = \max\{\rho^0, \rho^1, \cdots\}$ satisfying

$$f(x_k + \rho^i d_k) \le f(x_k) + \delta \rho^i g_k^T d_k; \tag{2.11}$$

or the Wolfe line search, that is, α_k satisfying

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k, \\ d_k^T g(x_k + \alpha_k d_k) \ge \sigma d_k^T g_k. \end{cases} \tag{2.12}$$

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. Set $k := k + 1$. Go to Step 1.

Remark 2.2 The Armijo line search and the Wolfe line search in Algorithm 2.1 are well-defined, since the search directions are descent from Lemma 2.2.

To ensure the convergence of Algorithm 2.1, we need the following standard assumption.

Assumption A.

(i) For any $\hat{x} \in R^n$, the level set

$$S(\hat{x}) = \{x \in R^n | f(x) \leq f(\hat{x})\}$$

is bounded.

(ii) f is continuously differentiable and there exists a constant $L > 0$ such that for any $\hat{x} \in R^n$, the gradient of f satisfies

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad x, y \in S(\hat{x}). \tag{2.13}$$

Assumption A is often used in analysis of convergence of conjugate gradient methods. In the next section, we give some functions which satisfies Assumption A.

Throughout this subsection, we always suppose that Assumption A holds. We can get from Assumption A that there exists a constant $\gamma > 0$ such that all x in the level set $S(x^0)$ satisfy

$$\|g(x)\| \leq \gamma. \tag{2.14}$$

Hence, we have

$$\begin{aligned} \|z_{k-1}\| &\leq \|y_{k-1}\| + t_k\|s_{k-1}\| \\ &\leq 2\|y_{k-1}\| + \varepsilon_0\|g_k\|^r\|s_{k-1}\| \\ &\leq (2L + \varepsilon_0\|g_k\|^r)\|s_{k-1}\| \leq (2L + \varepsilon_0\gamma^r)\|s_{k-1}\|, \end{aligned} \tag{2.15}$$

where the first and second inequalities use (2.4), the third inequality uses (2.13) and the last inequality uses (2.14).

The next result shows that d_k is bounded.

Lemma 2.3. *Let $\{x_k\}$ be generated by Algorithm 2.1. If there exists a positive constant ε such that for all $k \geq 0$,*

$$\|g_k\| \geq \varepsilon, \tag{2.16}$$

then there exists a constant $c_0 > 0$ such that

$$\|d_k\| \leq c_0\|g_k\|, \quad k \geq 0. \tag{2.17}$$

Proof By the definition (2.6) for H_k, we have

$$\begin{aligned}
\|H_k\| &= \|I - \frac{d_{k-1}z_{k-1}^T + z_{k-1}d_{k-1}^T}{d_{k-1}^T z_{k-1}} + \frac{2\|z_{k-1}\|^2}{(d_{k-1}^T z_{k-1})^2} d_{k-1}d_{k-1}^T\| \\
&\leq \|I\| + \frac{2\|d_{k-1}\|\|z_{k-1}\|}{d_{k-1}^T z_{k-1}} + \frac{2\|z_{k-1}\|^2\|d_{k-1}\|^2}{(d_{k-1}^T z_{k-1})^2} \\
&\leq \|I\| + \frac{2(2L+\varepsilon_0\gamma^r)\|d_{k-1}\|\|s_{k-1}\|}{\varepsilon_0\|g_{k-1}\|^r d_{k-1}^T s_{k-1}} + \frac{2(2L+\varepsilon_0\gamma^r)^2\|s_{k-1}\|^2\|d_{k-1}\|^2}{(\varepsilon_0\|g_{k-1}\|^r d_{k-1}^T s_{k-1})^2} \\
&\leq \|I\| + \frac{2(2L+\varepsilon_0\gamma^r)}{\varepsilon_0\varepsilon^r} + \frac{2(2L+\varepsilon_0\gamma^r)^2}{(\varepsilon_0\varepsilon^r)^2} =: c_0,
\end{aligned}$$

where the second inequality uses (2.15), (2.5) and $s_{k-1} = \alpha_{k-1}d_{k-1}$, and the third inequality uses $d_{k-1}^T s_{k-1} = \alpha_{k-1}\|d_{k-1}\|^2$.

It follows from the above inequality and (2.7) that $\|d_k\| \leq c_0\|g_k\|$. □

The following theorem shows that the convergence of Algorithm 2.1 when the objective function f is smooth and nonconvex.

Theorem 2.4. *Let $\{x_k\}$ be generated by Algorithm 2.1. If all stepsize α_k are computed by a single type line search either the Armijo line search (2.11) or the Wolfe line search (2.12), then we have*

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{2.18}$$

Proof (i) If all stepsize α_k are computed by the Armijo line search (2.11), then the conclusion follows directly from Theorem 3.1 of [14] and Lemma 2.2.

(ii) If all stepsize α_k are computed by the Wolfe line search, then we have from Assumption A and the first inequality in (2.12) and (2.10) that

$$\lim_{k\to\infty} \alpha_k\|g_k\|^2 = 0. \tag{2.19}$$

Now if (2.18) does not hold, then there exists a positive constant ε such that (2.16) holds for all $k \geq 0$. It follows from the second inequality in (2.12) that

$$L\alpha_k\|d_k\|^2 \geq (g_{k+1} - g_k)^T d_k \geq (1-\sigma)(-g_k^T d_k).$$

This together with (2.10), (2.14) and (2.17) implies that

$$\alpha_k \geq \frac{(1-\sigma)(-g_k^T d_k)}{L\|d_k\|^2} \geq \frac{(1-\sigma)\|g_k\|^2}{2L\|d_k\|^2} \geq \frac{(1-\sigma)\varepsilon^2}{2Lc_0^2\gamma^2} > 0.$$

Hence from (2.19), we obtain

$$\lim_{k\to\infty} \|g_k\| = 0,$$

which leads to a contradiction. This shows (2.18) holds. □

Remark 2.3 Theorem 2.4 shows that from any initial point $x_0 \in R^n$, Algorithm 2.1 converges to a stationary point x^* of f. If f is strongly pseudo convex at x^*, then x^* is a local minimizer of f. A function f is said to be strongly pseudo convex at x^* if there is a neighborhood Ω of x^* such that for every $\xi \in \partial f(x^*)$ and every $y \in \Omega$, $\xi^T(y - x^*) \geq 0 \Rightarrow f(y) \geq f(x^*)$. Pseudo convexity is weaker than convexity. For example, $\varphi_1(t) = \frac{\alpha|t|}{1+\alpha|t|}$ is strongly pseudo convex at all $t \in R$, but it is not convex.

2.2 Nonsmooth case

From now on, throughout the paper, we assume that f is locally Lipschitz continuous but not necessarily differentiable. According to the Rademacher theorem [11], f is differentiable almost everywhere in R^n. The subdifferential $\partial f(x)$, called the generalized gradient of f at x is defined by

$$\partial f(x) = \mathrm{conv}\{\lim_{\substack{x_i \to x \\ x_i \in D_f}} \nabla f(x_i)\},$$

where "conv" denotes the convex hull of a set and D_f is the set of points at which f is differentiable [10].

Based on the idea of the smoothing Newton method for nonsmooth equations in [7, 8], Zhang and Chen [34] proposed a smoothing projected gradient(SPG) method for nonsmooth and nonconvex optimization problems on a closed convex set. The SPG method is very simple and suitable for large-scale problems. To accelerate the convergence rate, in this subsection, we extend Algorithm 2.1 to the nonsmooth case by using the conjugate gradient of the smoothing function as the search direction. Firstly we define a class of smoothing functions of f, which are more general than that used for nonsmooth equations in [5, 9, 7, 8]. A smoothing function can be considered as a special smooth approximation of f, which uses a scalar smoothing parameter to play a key role in convergence analysis of the smoothing method.

Definition 2.5. *Let $f : R^n \to R$ be a locally Lipschitz continuous function. We call $\tilde{f} : R^n \times R_+ \to R$ a smoothing function of f, if $\tilde{f}(\cdot, \mu)$ is continuously differentiable in R^n for any fixed $\mu \in R_{++}$, and*

$$\lim_{\mu \downarrow 0} \tilde{f}(x, \mu) = f(x)$$

for any fixed $x \in R^n$.

Now we denote $\nabla \tilde{f}(x, \mu) = \nabla_x \tilde{f}(x, \mu)$, $\tilde{g}_k = \nabla \tilde{f}(x_k, \mu_k)$, then we can present the following smoothing conjugate gradient method for nonsmooth and nonconvex optimization.

Algorithm 2.2: (Smoothing conjugate gradient method)

Step 0. Choose constants $\varepsilon_0 > 0$, $r \geq 0$. Choose $\delta \in (0, 1)$, $\rho, \gamma_1 \in (0, 1)$, $\mu_0 > 0$, $\gamma > 0$, and initial point $x_0 \in R^n$.

Let $d_0 = -\tilde{g}_0$. Set $k := 0$.

Step 1. Compute the stepsize α_k by the Armijo line search, $\alpha_k = \max\{\rho^0, \rho^1, \cdots\}$ satisfying

$$\tilde{f}(x_k + \rho^m d_k, \mu_k) \leq \tilde{f}(x_k, \mu_k) + \delta \rho^m \tilde{g}_k^T d_k,$$

set

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

Step 2. If $\|\nabla \tilde{f}(x_{k+1}, \mu_k)\| \geq \gamma \mu_k$, then set $\mu_{k+1} = \mu_k$; otherwise, choose $\mu_{k+1} = \gamma_1 \mu_k$.

Step 3. Compute d_{k+1} by the following formula

$$d_{k+1} = -\tilde{g}_{k+1} + \Big(\frac{\tilde{g}_{k+1}^T \tilde{z}_k}{d_k^T \tilde{z}_k} - \frac{2\|\tilde{z}_k\|^2 \tilde{g}_{k+1}^T d_k}{(d_k^T \tilde{z}_k)^2}\Big) d_k + \frac{\tilde{g}_{k+1}^T d_k}{d_k^T \tilde{z}_k} \tilde{z}_k,$$

where $\tilde{z}_k = \tilde{y}_k + \Big(\varepsilon_0 \|\tilde{g}_{k+1}\|^r + \max\{0, -\frac{s_k^T \tilde{y}_k}{s_k^T s_k}\}\Big) s_k$, $\tilde{y}_k = \tilde{g}_{k+1} - \tilde{g}_k$ and $s_k = x_{k+1} - x_k$.

Step 4. Set $k := k + 1$. Go to Step 1.

Theorem 2.6. *Suppose that $\tilde{f}(\cdot, \mu)$ is a smoothing function of f. If for every fixed $\mu > 0$, $\tilde{f}(\cdot, \mu)$ satisfies Assumption A, then a sequence $\{x^k\}$ generated by Algorithm 2.2 satisfies*

$$\lim_{k\to\infty} \mu_k = 0 \quad \text{and} \quad \liminf_{k\to\infty} \|\nabla \tilde{f}(x_k, \mu_{k-1})\| = 0.$$

Proof Denote $K = \{k \,|\, \mu_{k+1} = \gamma_1 \mu_k\}$. If K is finite, then there exists an integer $\bar{k}$ such that for all $k > \bar{k}$

$$\|\nabla \tilde{f}(x_k, \mu_{k-1})\| \geq \gamma \mu_{k-1} \tag{2.20}$$

and $\mu_k = \mu_{\bar{k}} =: \bar{\mu}$ in Step 2 of Algorithm 2.2. Since $\tilde{f}(\cdot, \bar{\mu})$ is a smooth function, Algorithm 2.2 reduces to Algorithm 2.1 for solving

$$\min_{x\in R^n} \tilde{f}(x, \bar{\mu}).$$

Hence, by Assumption A on $\tilde{f}(\cdot, \bar{\mu})$, we have from Theorem 2.4 that

$$\liminf_{k\to\infty} \|\nabla \tilde{f}(x_k, \bar{\mu})\| = 0$$

which contradicts with (2.20). This shows that K must be infinite and $\lim_{k\to\infty} \mu_k = 0$.

Since K is infinite, we can assume that $K = \{k_0, k_1, \cdots\}$ with $k_0 < k_1 < \cdots$. Then we have

$$\lim_{i\to\infty} \|\nabla \tilde{f}(x_{k_i+1}, \mu_{k_i})\| \leq \gamma \lim_{i\to\infty} \mu_{k_i} = 0.$$

□

Remark 2.4 We can replace the Armijo line search in Algorithm 2.2 by the Wolfe line search to have Theorem 2.6. In the next section, we give a class of smoothing functions which satisfy Assumption A for every fixed $\mu > 0$. Note that Theorem 2.6 does not need that the limit of the Lipschitz constant for $\nabla \tilde{f}(\cdot, \mu)$ exists. This is the novelty of the smoothing nonlinear conjugate gradient method.

Smoothing functions of f can be defined in many ways. We present a class of smoothing functions for image restoration in the next section. In general, we can use a kernel function $\rho : R^n \to R_+$ to define a sequence of mollifiers which are bounded and continuous, and satisfy

$$\rho_\mu(x) = \mu^n \rho(\mu, x), \quad \mu > 0.$$

Using it, we define a smoothing function of f

$$\tilde{f}(x, \mu) = \int_{R^n} f(x - y)\rho_\mu(y)dy = \int_{R^n} f(y)\rho_\mu(x - y)dy. \tag{2.21}$$

(See [28] and Example 7.19 [29]). By Theorem 9.67 in [29], we have

$$G(x^*) \subseteq \partial f(x^*)$$

with

$$G(x^*) = \{v \,|\, v = \lim_{\substack{i\to\infty \\ i\in K}} \nabla_x \tilde{f}(x_i, \mu_i), x_i \to x^*, \mu_i \downarrow 0\},$$

where K is a subset of the set of all natural numbers.

By Theorem 2.6, any accumulation point x^* of $\{x_k\}$ generated by Algorithm 2.2 with a smoothing function defined by mollifiers satisfies $0 \in \partial f(x^*)$, that is, x^* is a Clarke stationary point of f [10].

3 Smoothing functions

Many potential functions $\varphi(t)$ in image restoration are continuously differentiable on R except at the origin. For instance, the following potential functions [12, 24]

$$\varphi_1(t) = \frac{\alpha|t|}{1+\alpha|t|}, \quad \varphi_2(t) = \log(1+\alpha|t|), \quad \varphi_3(t) = (|t|+\alpha)^p, \quad \alpha > 0,\ p \in (0,1). \tag{3.1}$$

It is clear that these functions are nonsmooth and nonconvex. In this paper, we consider the following class of nonsmooth potential functions.

Assumptions on φ: We assume that $\varphi : R \to R_+$ is a continuous function and satisfies

(i) $\varphi(t) \geq 0$, $\varphi(t) = \varphi(-t)$, that is, φ is nonnegative and symmetric.

(ii) φ is continuously differentiable on R except at 0, and

$$\lim_{t\downarrow 0} \varphi'(t) = \varphi'(0^+) = -\lim_{t\uparrow 0} \varphi'(t) = -\varphi'(0^-) \in (0,\infty). \tag{3.2}$$

(iii) φ is twice differentiable on R except at 0 and there is a constant ν_0 such that $|\varphi''(t)| \leq \nu_0$, for $t \neq 0$.

It is easy to see that φ_i, $i = 1, 2, 3$ satisfy Assumptions on φ. Now we present a class of smoothing potential functions which combines the splitting of φ in [24] and the integration convolution smoothing technique [5, 28, 29].

Nikolova et al [24] split the function φ into a smooth term and a nonsmooth convex term as

$$\varphi(t) = \psi(t) + \varphi'(0^+)|t|. \tag{3.3}$$

Consider the function

$$\psi(t) = \varphi(t) - \varphi'(0^+)|t|. \tag{3.4}$$

Since the derivative of ψ satisfies

$$\psi'(0^+) = 0 = \psi'(0^-),$$

by (iii) of Assumption on φ, ψ is continuously differentiable at 0 and thus on R. We can built a smoothing function of $|t|$ by integration convolution. Let $\rho \in [0,\infty)$ be a piecewise continuous density function with a finite number of pieces satisfying

$$\rho(\tau) = \rho(-\tau) \quad \text{and} \quad \kappa := \int_{-\infty}^{\infty} |\tau|\rho(\tau)d\tau < \infty. \tag{3.5}$$

The integration convolution smoothing function of $|t|$ is defined as

$$s_\mu(t) = \int_{-\infty}^{\infty} |t - \mu\tau|\rho(\tau)d\tau. \tag{3.6}$$

For instance, if we choose the density function

$$\rho(\tau) = \begin{cases} 0 & \text{if } |\tau| > 0.5, \\ 1 & \text{othersise,} \end{cases}$$

then we obtain a smoothing function of $|t|$

$$s_\mu(t) = \begin{cases} |t| & \text{if} \quad |t| > \dfrac{\mu}{2}, \\ \dfrac{t^2}{\mu} + \dfrac{\mu}{4} & \text{if} \quad |t| \le \dfrac{\mu}{2}. \end{cases} \tag{3.7}$$

This function is also known as Huber potential function [18, 2] which is very often used in robust statistics. It is worth noticing that ρ is not necessarily continuous. From our numerical experiment, a piecewise continuous density function performs better than a continuous one. However, smoothing functions defined by (2.21) require the mollifier to be continuous in [28, 29].

Combining (3.3) and (3.6), we define a class of smoothing functions for a potential function φ as

$$\varphi_\mu(t) = \psi(t) + \varphi'(0^+)s_\mu(t). \tag{3.8}$$

The following proposition shows that φ_μ has nice smooth approximation properties.

Proposition 3.1. *A function φ_μ defined by (3.8) with s_μ defined by (3.5) and (3.6) has the following properties.*

(i) $\varphi_\mu(t) = \varphi_\mu(-t)$ for $t \in R$, that is, φ_μ is symmetric.

(ii) φ_μ is continuously differentiable on R, and its derivative can be given as

$$\varphi'_\mu(t) = \psi'(t) + 2\varphi'(0^+) \int_0^{\frac{t}{\mu}} \rho(\tau)d\tau. \tag{3.9}$$

(iii) φ_μ converges uniformly to φ on R with

$$|\varphi_\mu(t) - \varphi(t)| \le \varphi'(0^+)\kappa\mu.$$

(iv) The set of limits of gradient $\varphi'_\mu(t)$ coincides to the Clarke generalized gradient, that is,

$$\{\lim_{\mu\downarrow 0, t\to 0} \varphi'_\mu(t)\} = \partial\varphi(0), \quad \text{and} \quad \lim_{\mu\downarrow 0, t\to t^*} \varphi'_\mu(t) = \varphi'(t^*), \quad t^* \neq 0. \tag{3.10}$$

Moreover, we have

$$\lim_{\mu\downarrow 0} \varphi'_\mu(t) = \begin{cases} \varphi'(t) & \text{if} \quad t \neq 0, \\ 0 & \text{if} \quad t = 0. \end{cases} \tag{3.11}$$

(v) For any fixed $\mu > 0$, φ'_μ is Lipschitz continuous on R, that is, there is a constant $\nu_\mu > 0$ such that

$$|\varphi'_\mu(t_1) - \varphi'_\mu(t_2)| \le \nu_\mu |t_1 - t_2|. \tag{3.12}$$

Proof (i) We have from (3.8) that

$$\varphi_\mu(t) = \psi(t) + \varphi'(0^+)s_\mu(t) = \varphi(t) + \varphi'(0^+)(s_\mu(t) - |t|).$$

By variable transformation $u = -\tau$, we have from $\rho(\tau) = \rho(-\tau)$ that

$$s_\mu(-t) = \int_{-\infty}^{\infty} |-t - \mu\tau|\rho(\tau)d\tau = \int_{-\infty}^{\infty} |t - \mu u|\rho(u)du = s_\mu(t), \tag{3.13}$$

which shows that s_μ is symmetric in R. This, together with (i) of Definition 3.1, implies that φ_μ is symmetric.

(ii) From the definition of $s_\mu(t)$, we have

$$\begin{aligned} s_\mu(t) &= \int_{-\infty}^{+\infty} |t-\mu\tau|\rho(\tau)d\tau \\ &= \int_{-\infty}^{\frac{t}{\mu}} (t-\mu\tau)\rho(\tau)d\tau - \int_{\frac{t}{\mu}}^{+\infty} (t-\mu\tau)\rho(\tau)d\tau \\ &= t\Big(\int_{-\infty}^{\frac{t}{\mu}} \rho(\tau)d\tau - \int_{\frac{t}{\mu}}^{+\infty} \rho(\tau)d\tau\Big) + \mu\Big(\int_{\frac{t}{\mu}}^{+\infty} \tau\rho(\tau)d\tau - \int_{-\infty}^{\frac{t}{\mu}} \tau\rho(\tau)d\tau\Big) \\ &= 2t\int_0^{\frac{t}{\mu}} \rho(\tau)d\tau + \mu\Big(\int_{\frac{t}{\mu}}^{+\infty} \tau\rho(\tau)d\tau - \int_{-\infty}^{\frac{t}{\mu}} \tau\rho(\tau)d\tau\Big), \end{aligned}$$

where the last equality uses $\rho(\tau) = \rho(-\tau)$. Note that $\rho(\tau) \geq 0$ and $\int_{-\infty}^{\infty} |\tau|\rho(\tau)d\tau < \infty$. By the integral mean value theorem, we obtain

$$\begin{aligned} s'_\mu(t) &= \lim_{\Delta t\to 0} \frac{s_\mu(t+\Delta t) - s_\mu(t)}{\Delta t} \\ &= \lim_{\Delta t\to 0} 2\int_0^{\frac{t+\Delta t}{\mu}} \rho(\tau)d\tau + \lim_{\Delta t\to 0} \frac{2t\int_{\frac{t}{\mu}}^{\frac{t+\Delta t}{\mu}} \rho(\tau)d\tau - 2\mu\int_{\frac{t}{\mu}}^{\frac{t+\Delta t}{\mu}} \tau\rho(\tau)d\tau}{\Delta t} \\ &= 2\int_0^{\frac{t}{\mu}} \rho(\tau)d\tau + \lim_{\Delta t\to 0} 2\mu\frac{\int_{\frac{t}{\mu}}^{\frac{t+\Delta t}{\mu}} \Big(\frac{t}{\mu}-\tau\Big)\rho(\tau)d\tau}{\Delta t} \\ &= 2\int_0^{\frac{t}{\mu}} \rho(\tau)d\tau + \lim_{\Delta t\to 0} 2\mu\frac{\frac{t}{\mu}-\xi}{\Delta t}\int_{\frac{t}{\mu}}^{\frac{t+\Delta t}{\mu}} \rho(\tau)d\tau \\ &= 2\int_0^{\frac{t}{\mu}} \rho(\tau)d\tau, \end{aligned} \tag{3.14}$$

where $\xi \in [\frac{t}{\mu},\ \frac{t+\Delta t}{\mu}]$. This gives (3.9).

(iii) By (3.8), we have

$$\begin{aligned} |\varphi_\mu(t) - \varphi(t)| &= \varphi'(0^+)|s_\mu(t) - |t|| \\ &= \varphi'(0^+)|\int_{-\infty}^{+\infty} (|t-\mu\tau| - |t|)\rho(\tau)d\tau| \\ &\leq \varphi'(0^+)\int_{-\infty}^{+\infty} |t-\mu\tau-t|\rho(\tau)d\tau \\ &= \varphi'(0^+)\int_{-\infty}^{+\infty} \mu|\tau|\rho(\tau)d\tau \\ &= \varphi'(0^+)\kappa\mu, \end{aligned}$$

where κ is specified by (3.5).

(iv) It follows from (3.14) and $\rho(\tau) = \rho(-\tau)$ that

$$\lim_{\mu\downarrow 0, t\to t^*} s'_\mu(t) = \lim_{\mu\downarrow 0, t\to t^*} 2\int_0^{\frac{t}{\mu}} \rho(\tau)d\tau = \lim_{\mu\downarrow 0, t\to t^*} \int_{-\frac{t}{\mu}}^{\frac{t}{\mu}} \rho(\tau)d\tau.$$

Then for $|t^*| \neq 0$, we have

$$\lim_{\mu\downarrow 0, t\to t^*} s'_\mu(t) = \begin{cases} 1 & \text{if} \quad t^* > 0, \\ -1 & \text{if} \quad t^* < 0. \end{cases}$$

For $t^* = 0$, we have

$$\lim_{\mu\downarrow 0, t\to 0} s'_\mu(t) \quad \in \quad \begin{cases} \{1\} & \text{if} \quad \dfrac{t}{\mu} \to +\infty, \\ \{\alpha\} & \text{if} \quad \lim\limits_{\mu\downarrow 0, t\to 0} |\dfrac{t}{\mu}| < +\infty, \\ \{-1\} & \text{if} \quad \dfrac{t}{\mu} \to -\infty. \end{cases}$$

where $\alpha \in [-1, 1]$. This shows that $\lim_{\mu\downarrow 0, t\to 0} s'_\mu(t) \subseteq [-1, 1]$. Define $\alpha(\lambda) = 2\int_0^\lambda \rho(\tau)d\tau$, then $\alpha(\lambda) \in [-1, 1]$ is continuous in R since ρ is piecewise continuous. Therefore, for any $\alpha_0 \in (-1, 1)$, there exists λ_0 such that $\alpha_0 = \alpha(\lambda_0)$. If we choose $\frac{t_k}{\mu_k} = \lambda_0$ and $\mu_k \downarrow 0$, then we have

$$\alpha_0 = \lim_{\mu_k\downarrow 0, t_k\to 0} s'_{\mu_k}(t_k).$$

This shows that

$$[-1, 1] \subseteq \lim_{\mu\downarrow 0, t\to 0} s'_\mu(t).$$

Therefore, we have $\lim_{\mu\downarrow 0, t\to 0} s'_\mu(t) = [-1, 1]$. By the continuity of $\psi'(t)$, (3.14) and (3.9), we obtain (3.10). From $\varphi'_\mu(0) = 0$, we get (3.11).

(v) We first show s'_μ is Lipschitz continuous. Since ρ is piecewise continuous with a finite number of pieces, there exists a constant κ_0 such that $\rho(t) \leq \kappa_0$ for any $t \in R$. For any $t_1, t_2 \in R$, we have

$$|s'_\mu(t_1) - s'_\mu(t_2)| = 2\Big|\int_0^{\frac{t_1}{\mu}} \rho(\tau)d\tau - \int_0^{\frac{t_2}{\mu}} \rho(\tau)d\tau\Big| = 2\Big|\int_{\frac{t_2}{\mu}}^{\frac{t_1}{\mu}} \rho(\tau)d\tau\Big| \leq \frac{2\kappa_0}{\mu}|t_1 - t_2|.$$

Now we show ψ' is Lipschitz continuous. Since $\psi(t) = \varphi(t) - \varphi'(0^+)|t|$, then we have from (iii) in Assumptions on φ that for $t_1 t_2 > 0$,

$$|\psi'(t_1) - \psi'(t_2)| = |\varphi'(t_1) - \varphi'(t_2)| \leq \nu_0|t_1 - t_2|.$$

If $t_1 \neq 0, t_2 = 0$, then we have from $\psi'(0) = 0$ that

$$|\psi'(t_1) - \psi'(0)| = |\varphi'(t_1) - \varphi'(0^+)| \leq \nu_0|t_1 - 0|.$$

If $t_1 t_2 < 0$, we may assume $t_2 < 0 < t_1$, then we have $\varphi'(t_2) = -\varphi'(-t_2)$ by (i) in Assumptions on φ. Hence we have

$$\begin{aligned} |\psi'(t_1) - \psi'(t_2)| &= |(\varphi'(t_1) - \varphi'(0^+)) - (\varphi'(t_2) + \varphi'(0^+))| \\ &= |(\varphi'(t_1) - \varphi'(0^+)) + (\varphi'(-t_2) - \varphi'(0^+))| \\ &\leq |\varphi'(t_1) - \varphi'(0^+)| + |\varphi'(-t_2) - \varphi'(0^+)| \\ &\leq \nu_0 t_1 + \nu_0(-t_2) \\ &= \nu_0|t_1 - t_2|. \end{aligned}$$

Let $\nu_\mu = \nu_0 + \frac{2\varphi'(0^+)\kappa_0}{\mu}$, we obtain (3.12). □

Now we are ready to define a class of smoothing functions and use the smoothing conjugate gradient method (Algorithm 2.2) for nonsmooth and nonconvex image restoration problems.

In the remain part of this section as well as in the next section, we consider the minimization problem (1.5) with the objective function

$$f(x) = \|Ax - b\|^2 + \beta \sum_{i=1}^{r} \varphi(d_i^T x), \tag{3.15}$$

where $A \in R^{m\times n}$ is a blurring matrix, $b \in R^m$ is a vector containing observed data, φ is a potential function satisfying Assumptions on φ, $\beta > 0$ is a constant and $d_i \in R^n$, $i = 1, \ldots, r$ are the row vectors of a difference matrix.

Using the smoothing function φ_μ for φ, we define a class of smoothing functions for f as follows

$$\tilde{f}(x, \mu) = \|b - Ax\|^2 + \beta \sum_{i=1}^{r} \varphi_\mu(d_i^T x). \tag{3.16}$$

To show the smoothing function $\tilde{f}$ has nice approximation properties and any accumulation point of a sequence generated by Algorithm 2.2 with $\tilde{f}$ is a Clarke stationary point, we need the definition of a regular function [10, Definition 2.3.4] and some results from Proposition 2.3.3, Proposition 2.3.6, Theorem 2.3.9 and Corollary 3 in [10, Chapter 2]. We give the definition of a regular function and summarize these results as follows.

Definition 3.2. *[10] A function f is said to be regular at x provided*

(i) For all v, the usual one-sided directional derivative $f'(x; v)$ exists.

(ii) For all v, $f'(x; v) = \limsup_{\substack{y\to x \\ t\downarrow 0}} \dfrac{f(y + tv) - f(y)}{t}$.

Lemma 3.3. *[10]*

(i) Suppose that $g_i : R^n \to R$, $i = 1, \ldots, m$ are Lipschitz continuous near x. Then their sum $g = \sum_{i=1}^{m} g_i$ is also Lipschitz continuous near x and

$$\partial g(x) = \partial\Big(\sum_{i=1}^{m} g_i\Big)(x) \subseteq \sum_{i=1}^{m} \partial g_i(x).$$

If each g_i is regular at x, equality holds.

(ii) Let $g(x) = h(F(x))$, where $F : R^n \to R^m$ is Lipschitz continuous near x and where $h : R^m \to R$ is Lipschitz continuous near $F(x)$. Then g is Lipschitz continuous near x and

$$\partial g(x) \subset \overline{conv}\Big\{ \sum \alpha_i \zeta_i | \zeta_i \in \partial F_i(x), \alpha = (\alpha_1, \cdots, \alpha_m)^T \in \partial h(F(x)) \Big\}.$$

If h is regular at $F(x)$ and F is continuously differentiable at x (in this case the $\overline{conv}$ is superfluous), equality holds.

(iii) A Lipschitz continuous function $g : R^n \to R$ is regular at x if g is convex or g is continuously differentiable at x.

Theorem 3.4. *Let f and $\tilde{f}(\cdot, \mu)$ be defined by (3.15) and (3.16) respectively. Then*

(i) $\tilde{f}(\cdot, \mu)$ is continuously differentiable for any fixed $\mu > 0$, and there exists a constant $\kappa_1 > 0$ satisfying

$$|\tilde{f}(x, \mu) - f(x)| \leq \kappa_1 \mu.$$

(ii) $\tilde{f}(\cdot, \mu)$ satisfies the gradient consistent property, that is,

$$\{\lim_{\mu \downarrow 0, x \to x^*} \nabla \tilde{f}(x, \mu)\} = \partial f(x^*).$$

(iii) If A has full column rank, then for any $\hat{x} \in R^n$, the level set $S_\mu(\hat{x}) = \{x \in R^n | \tilde{f}(x, \mu) \leq \tilde{f}(\hat{x}, \mu)\}$ is bounded.

(iv) For any fixed $\mu > 0$, the gradient of $\tilde{f}(x, \mu)$ is Lipschitz continuous, that is, for any $x, y \in S_\mu(\hat{x})$, there exists a constant L_μ such that

$$\|\nabla \tilde{f}(x, \mu) - \nabla \tilde{f}(y, \mu)\| \leq L_\mu \|x - y\|.$$

Proof From the definitions of φ and φ_μ, we can write f and $\tilde{f}$ as the following

$$f(x) = \|b - Ax\|^2 + \beta \sum_{i=1}^{r} \psi(d_i^T x) + \beta \varphi'(0^+) \sum_{i=1}^{r} |d_i^T x|, \tag{3.17}$$

and

$$\tilde{f}(x, \mu) = \|b - Ax\|^2 + \beta \sum_{i=1}^{r} \psi(d_i^T x) + \beta \varphi'(0^+) \sum_{i=1}^{r} s_\mu(d_i^T x). \tag{3.18}$$

(i) It follows from (3.17) and (3.18) and Proposition 3.1 that $\tilde{f}(\cdot, \mu)$ is continuously differentiable for any fixed $\mu > 0$, and

$$|\tilde{f}(x, \mu) - f(x)| \leq \beta r \varphi'(0^+) \kappa \mu.$$

Set $\kappa_1 = \beta r \varphi'(0^+) \kappa \mu$, then (i) holds.

(ii) Set $h(t) = |t|$ and $F(x) = d^T x$. Then by (ii) in Lemma 3.3, $g(x) = |d^T x|$ is regular at x and $\partial g(x) = \partial |d^T x| d$. Using Lemma 3.3 again, we have

$$\partial f(x) = 2A^T(Ax - b) + \beta \sum_{i=1}^{r} (\psi'(d_i^T x)) d_i + \beta \varphi'(0^+) \sum_{i=1}^{r} \partial(|d_i^T x|) d_i.$$

The gradient of the smoothing function $\tilde{f}(\cdot, \mu)$ is given by

$$\nabla \tilde{f}(x, \mu) = 2A^T(Ax - b) + \beta \sum_{i=1}^{r} \psi'(d_i^T x) d_i + \beta \varphi'(0^+) \sum_{i=1}^{r} s'_\mu(d_i^T x) d_i.$$

By (iv) of Proposition 3.1 and Lemma 3.3, we have

$$\begin{aligned}
&\left\{\lim_{\mu\downarrow 0, x\to x^*} \nabla\tilde{f}(x,\mu)\right\} \\
= &\left\{\lim_{x\to x^*}\left(2A^T(Ax-b)+\beta\sum_{i=1}^r \psi'(d_i^Tx)d_i\right)+\beta\varphi'(0^+)\sum_{i=1}^r \lim_{\mu\downarrow 0, x\to x^*} s'_\mu(d_i^Tx)d_i\right\} \\
= &\ 2A^T(Ax^*-b)+\beta\sum_{i=1}^r \psi'(d_i^Tx^*)d_i+\beta\varphi'(0^+)\sum_{i=1}^r \partial|d_i^Tx^*|d_i = \partial f(x^*).
\end{aligned}$$

This shows that the gradient consistent property holds for the smoothing function $\tilde{f}$.

(iii) If $S_\mu(\hat{x})$ is unbounded, then there exists a sequence $\{x_k\}\subset S_\mu(\hat{x})$ such that $\|x_k\|\to\infty$. We have from (i) in Assumptions on φ and (iii) in Proposition 3.1 that

$$\begin{aligned}
\tilde{f}(x,\mu) &= x^TA^TAx-2(Ax)^Tb+\|b\|^2+\beta\sum_{i=1}^r \varphi_\mu(d_i^Tx) \\
&\geq x^TA^TAx-2(Ax)^Tb+\|b\|^2+\beta\sum_{i=1}^r\Big(\varphi(d_i^Tx)-|\varphi(d_i^Tx)-\varphi_\mu(d_i^Tx)|\Big), \\
&\geq x^TA^TAx-2(Ax)^Tb+\|b\|^2+\beta\sum_{i=1}^r(\varphi(d_i^Tx)-\varphi'(0^+)\kappa\mu). \qquad (3.19)
\end{aligned}$$

Since A^TA is symmetric positive definite, $\|x_k\|\to\infty$ implies $\tilde{f}(x_k,\mu)\to+\infty$. Hence (iii) is true.

(iv) Using the expression

$$\nabla\tilde{f}(x,\mu)=2A^T(Ax-b)+\beta\sum_{i=1}^r\varphi'_\mu(d_i^Tx)d_i,$$

we have from (v) of Proposition 3.1 that

$$\begin{aligned}
\|\nabla\tilde{f}(x,\mu)-\nabla\tilde{f}(y,\mu)\| &\leq 2\|A^TA\|\|x-y\|+\beta\sum_{i=1}^r\|d_i\|\|\varphi'_\mu(d_i^Tx)-\varphi'_\mu(d_i^Ty)\| \\
&\leq 2\|A^TA\|\|x-y\|+\beta\sum_{i=1}^r\|d_i\|\nu_\mu\|d_i^Tx-d_i^Ty\| \\
&\leq (2\|A^TA\|+\beta\sum_{i=1}^r\|d_i\|^2\nu_\mu)\|x-y\|.
\end{aligned}$$

Set $L_\mu=2\|A^TA\|+\beta\sum_{i=1}^r\|d_i\|^2\nu_\mu$, then (iv) holds. □

Remark 3.1 Theorem 3.4 shows that the smoothing function $\tilde{f}$ has very nice approximation properties and satisfies all assumptions of the convergence theorem (Theorem 2.6) for the smoothing conjugate gradient method (Algorithm 2.2). The most significant one is the gradient consistent property, which ensures that any accumulation point of a sequence generated by Algorithm 2.2 is a Clarke stationary point. Using smooth approximations to solve nonsmooth

Table 1: Test results with $\beta = 0.001, \alpha = 1$, and $f(x_{orig}) = 1.0849$. We stopped the iteration if $\|x_k - x_{orig}\|/\|x_k\| \leq 0.06$ or the number of iterations exceeds 120.

	CM			SCG		
Initial point	time	psnr	$f(x_k)$	time	psnr	$f(x_k)$
0.0001	120.2273	28.4635	0.6734	64.6418	29.8679	0.7551
0.001	119.2405	28.4635	0.6734	69.2970	29.8726	0.7529
0.01	120.0629	28.4635	0.6734	64.7127	29.8564	0.7517
0.1	118.6640	28.4635	0.6734	62.0299	29.8333	0.7513
0.2	119.1084	28.4635	0.6734	87.2522	29.8317	0.7545
0.4	119.6038	28.4635	0.6734	52.1802	29.8424	0.7544
0.6	123.0495	28.4635	0.6734	94.9425	29.8630	0.7500
0.8	119.5355	28.4635	0.6734	89.9128	29.8450	0.7547
observed	121.2054	28.4635	0.6734	49.4008	29.8413	0.7583
random	120.4385	28.4635	0.6734	109.0128	29.8350	0.7517
average	120.1135			74.3383		

optimization problems have been studied in many papers [23, 24]. However, there is no guarantee for convergence to a generalized stationary point of the nonsmooth optimization problems in [23, 24].

Remark 3.2 The smoothing functions studied in this section can be applied to other nonsmooth image restoration models. For example, approximating $|t|$ by $s_\mu(t)$ in the $l_1 - l_1$ model (1.3) and potential functions (3.1). From (3.19), it is not difficult to see that the assumption that A has full column rank in (iii) of Theorem 3.4 can be replaced by $\mathcal{N}(A) \cap \mathcal{N}(D) = \{0\}$ and φ is coercive, that is, $t \to \infty \Rightarrow \varphi(t) \to \infty$, where $\mathcal{N}(A)$ and $\mathcal{N}(D)$ are null spaces of A and D, respectively. See Assumption 1 in [33]

4 Numerical results

In this section, we present numerical results to show the efficiency of the smoothing conjugate gradient method(Algorithm 2.2, abbreviated by SCG). The numerical testing was carried out on a Lenovo PC (3.00GHz, 2.00GB of RAM) with the use of Matlab 7.4.

In our numerical experiment, the objective function has the form (3.15) and its smoothing function is defined by (3.16) and (3.7). We test three often used images: Lena image, Cameraman image and Phantom image, which are all gray level images with intensity values ranging from 0 to 1. The Lena and Cameraman images are of size $n = 128 \times 128$. The Phantom image is of size from $n = 128 \times 128$ to $n = 256 \times 256$. The pixels of the observed images are contaminated by Gaussian white noise with signal-to-noise ratios of 45 dB for Table 1 and 60 dB for the other tables with blurring. The blurring function is chosen to be a two-dimensional Gaussian,

$$a(i, j) = e^{-2(i/3)^2 - 2(j/3)^2},$$

truncated such that the function has a support of 7×7.

For the regularization term, we use different potential functions φ_i for $i = 1, 2, 3$. We choose

Table 2: Test results with $\beta = 0.02, \alpha = 0.5$, and $f(x_{orig}) = 11.22$.

	CM			SCG		
Initial point	time	psnr	$f(x_k)$	time	psnr	$f(x_k)$
0.0001	371.95	26.53	6.05	272.23	27.01	6.27
0.001	383.56	26.53	6.05	273.68	27.03	6.28
0.01	379.69	26.53	6.05	287.85	27.06	6.28
0.1	377.07	26.53	6.05	325.01	27.04	6.28
0.2	380.21	26.53	6.05	312.39	27.17	6.30
0.4	372.64	26.53	6.05	291.19	27.04	6.28
0.6	397.50	26.53	6.05	354.22	27.11	6.29
0.8	398.92	26.53	6.05	451.35	27.04	6.28
observe	426.33	26.53	6.05	271.22	27.11	6.29
random	400.08	26.53	6.05	462.94	27.07	6.28
average	388.80			330.21		

D to be the identity matrix $D_0 = I$, or a matrix of a first-order difference operator:

$$D_1 = \begin{pmatrix} L_1 \otimes I \\ I \otimes L_1 \end{pmatrix} \text{ with } L_1 = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix},$$

or a matrix of a second-order difference operator with the Neumann boundary condition

$$D_2 = \begin{pmatrix} L_2 \otimes I \\ I \otimes L_2 \end{pmatrix} \text{ with } L_2 = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}.$$

4.1 Comparison with the continuation method in [24]

In this subsection, we make a comparison between the smoothing conjugate gradient method (Algorithm 2.2) and the continuation method [24] for the Lena image. Tables 1-2 list the numerical results for the Lena image restoration problems with different regularization parameters, where we choose φ_1 and D_1 as the potential function and the difference operator, and set $\beta = 0.001$, $\alpha = 1$ in Table 1, and $\beta = 0.02$, $\alpha = 0.5$ in Table 2. In Table 1, we stop the iteration if $\|x_k - x_{orig}\|/\|x_k\| \leq 0.06$ or the number of iterations exceeds 120. In Table 2, for the CM method, we stop after computing 21 exterior iterations $J_{\varepsilon_0}, \cdots, J_{\varepsilon_{20}}$ and 15 interior iterations for every $J_{\varepsilon_k}, k = 0, \cdots, 20$; for the SCG method we stop the iteration if the total number of iterations exceeds 120 or the inequality $\|\nabla \tilde{f}(x_k, \mu_{k-1})\| < 0.05$ holds.

- CM: the continuation method proposed by Nikolova et al. in [24].
- SCG: Algorithm 2.2, in which we set parameters $\rho = 0.4$, $\gamma = 2$, $\gamma_1 = 0.5$, $\varepsilon_0 = 10^{-10}$, $\mu_0 = 1$, $\delta = 0.1$.

- Initial point: In both methods, we use 10 different initial points. The first 8 initial guesses are flat image (for example: 0.1 means all the pixel values are 0.1), and the ninth initial guess is the observed image and the last one is a random image.
- time: the CPU time in second.
- $f(x_{orig})$ and $f(x_k)$: the function values of f at the original image and the stopping point, respectively.

Figure 1: The left and the right are the original Lena image and the observed Lena image, respectively.

- psnr: peak signal-to-noise ratios of the restored images, which is defined by

$$\text{psnr} = -10 * \log_{10} \frac{\|x_k - x_{orig}\|^2}{mn},$$

  where x_{orig} is the original image, and $\|\cdot\|$ is the l_2 norm.

Figures 1-2 show the original, observed and restored Lena images by these two methods. We can see that the SCG method performs better than the CM method in the sense that the SCG method obtains higher psnr of the restored image and needs less CPU time. On the other hand, the CM method has less function value at the terminated point. The ultimate value of μ_k at the stopping point for the SCG method is about 0.0313.

4.2 Test results for Algorithm 2.2 with different potential functions and difference operators

In this subsection, we test the Cameraman image and the Phantom image by using the smoothing conjugate gradient method (SCG) with different potential functions and difference operators. We summarize numerical results in Tables 3-5, where

- SCG: We set parameters $\rho = 0.4$, $\gamma = 2$, $\gamma_1 = 0.5$, $\varepsilon_0 = 10^{-10}$, $\mu_0 = 1$, $\delta = 0.1$, $\beta = 0.001$, $\alpha = 1$. We stop the iteration if the inequality $\|\nabla \tilde{f}(x_k, \mu_{k-1})\| < 0.1$ holds. The observed image is used as the initial guess.
- iter: the number of iterations.

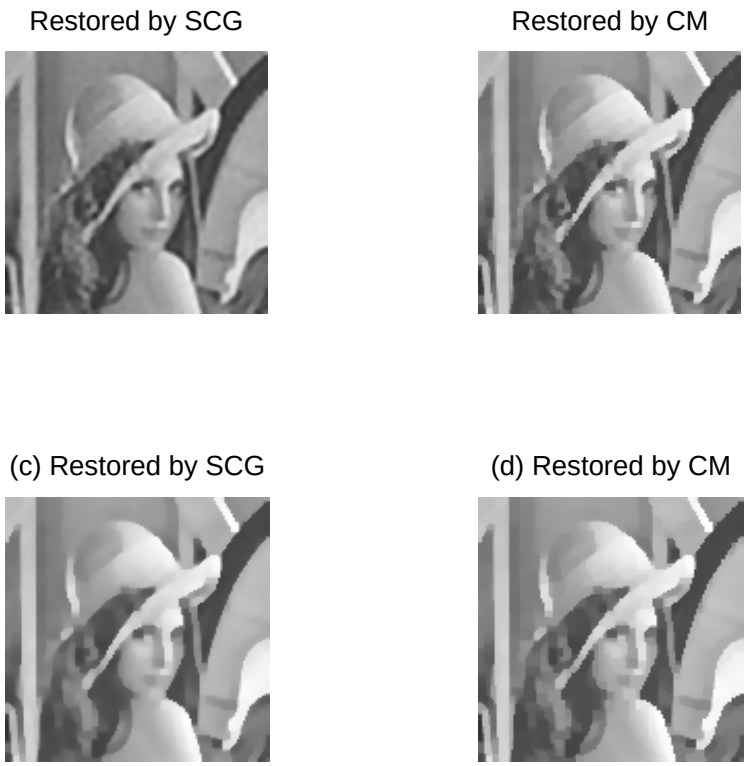

Figure 2: Lena image: (a) and (b) are the restoration images by SCG and CM methods with $\beta = 0.001$ and $\alpha = 1$ and the observed image as the initial point; (c) and (d) are the restoration images by SCG and CM methods with $\beta = 0.02$ and $\alpha = 0.5$ and the observed image as the initial point.

- $f(x_k)$ and $\|\nabla \tilde{f}_k\|$: the function value $f(x_k)$ and the l_2 norm of $\nabla \tilde{f}(x_k, \mu_{k-1})$ at the stopping point, respectively.

Figures 3-5 show the original, observed and restored images with different pixels, different potential functions and different linear operators D. We only list some restored images since the quality of the other restored images is very similar. The detailed numerical results are reported in Tables 3-5.

Numerical results show that the SCG method can efficiently reduce objective function values, improve peak signal-to-noise ratios of the image restorations and produce piecewise constant images. We can see from Figures 3-5 that the SCG method can preserve neat edge of the image. This effect is especially clear in Figures 4-5. Moreover, the SCG method can deal with large-scale image problems with $n = 256 \times 256 = 65536$ pixels.

Table 3: Test results for Cameraman image with $n = 128 \times 128$.

φ	D	iter	time	$f(x_{orig})$	$f(x_k)$	$\|\nabla \tilde{f}_k\|$	psnr
φ_1	D_0	214	53.04	2.97	3.08	0.0930	26.53
φ_2	D_0	217	53.24	3.36	3.43	0.0805	26.94
φ_3	D_0	224	59.30	15.97	16.05	0.0970	26.86
φ_1	D_1	292	539.39	1.94	1.97	0.0976	26.42
φ_2	D_1	255	446.97	2.36	2.32	0.0817	26.70
φ_3	D_1	245	459.78	8.81	8.77	0.0808	26.51
φ_1	D_2	217	398.30	1.94	1.96	0.0893	26.46
φ_2	D_2	232	411.85	2.36	2.33	0.0806	26.59
φ_3	D_2	236	440.84	8.81	8.78	0.0969	26.46

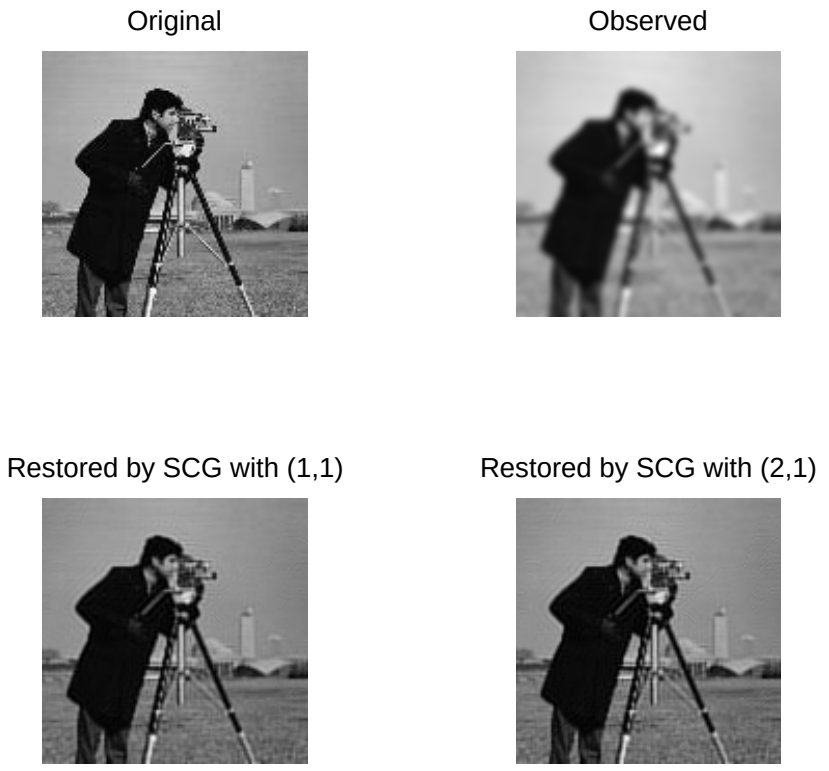

Figure 3: Cameraman image: the first row contains the original image(left) and the observed image(right); the second row contains the restoration images(left: psnr=26.42dB with φ_1 and D_1, and right: psnr=26.70dB with φ_2 and D_1).

Table 4: Test results for Phantom image with $n = 128 \times 128$.

φ	D	iter	time	$f(x_{orig})$	$f(x_k)$	$\|\nabla \tilde{f}_k\|$	psnr
φ_1	D_0	217	53.10	0.83	0.94	0.0845	26.00
φ_2	D_0	211	49.99	0.91	1.05	0.0924	25.77
φ_3	D_0	198	52.09	12.80	12.95	0.0980	25.82
φ_1	D_1	289	179.85	0.87	0.92	0.0853	25.79
φ_2	D_1	265	164.68	1.04	1.08	0.0904	25.74
φ_3	D_1	294	188.42	7.14	7.18	0.0849	25.65
φ_1	D_2	318	215.90	0.87	0.91	0.0896	25.85
φ_2	D_2	273	169.39	1.04	1.08	0.0931	25.76
φ_3	D_2	250	157.91	7.14	7.17	0.0911	25.72

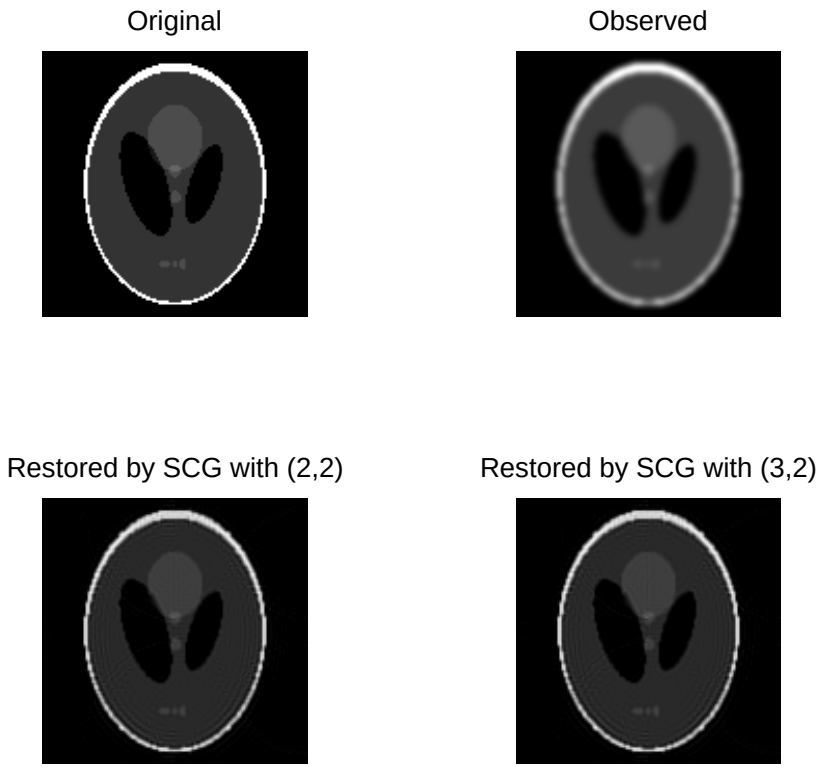

Figure 4: Phantom image with 128×128 pixels: the first row contains the original image(left) and the observed image(right); the second row contains the restoration images(left: psnr=25.76dB with φ_2 and D_2, and right: psnr=25.72dB with φ_3 and D_2).

Table 5: Test results for Phantom image with $n = 256 \times 256$.

φ	D	iter	time	$f(x_{orig})$	$f(x_k)$	$\|\nabla \tilde{f}_k\|$	psnr
φ_1	D_0	275	267.32	3.36	3.54	0.0857	28.80
φ_2	D_0	228	220.75	3.69	3.86	0.0960	28.83
φ_3	D_0	319	336.79	51.23	51.40	0.0696	29.03
φ_1	D_1	316	8627.62	3.32	3.37	0.0986	28.56
φ_2	D_1	283	7721.57	3.99	4.03	0.0975	28.45
φ_3	D_1	341	9420.75	28.32	28.31	0.0936	28.52
φ_1	D_2	376	10277.73	3.32	3.37	0.0939	28.56
φ_2	D_2	364	9831.85	3.99	3.98	0.0943	28.59
φ_3	D_2	302	8173.62	28.32	28.30	0.0967	28.54

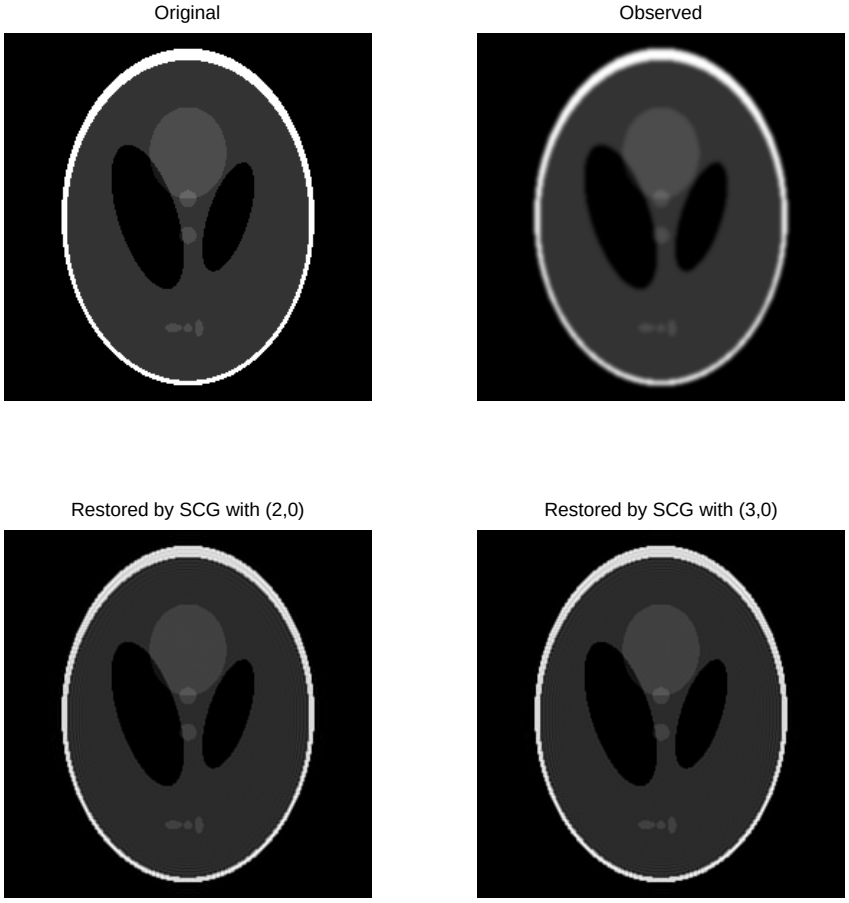

Figure 5: Phantom image with 256×256 pixels: the first row contains the original image(left) and the observed image(right); the second row contains the restoration images(left: psnr=28.83dB with φ_2 and D_0, and right: psnr=29.03dB with φ_3 and D_0).

5 Conclusions

Nonsmooth and nonconvex minimization problem has been widely used in image restoration. However, existing minimization algorithms are not efficient for solving such problem. In this paper, we present an efficient smoothing nonlinear conjugate gradient method for large-scale, nonsmooth and nonconvex image restoration problems. This method is very easy to implement without adding any new variables, and ensures that any accumulation point of a sequence generated by this method is a Clarke stationary point.

Acknowledgments
The authors would like to thank Professor Michael Elad and the three anonymous referees for their valuable suggestions and comments. We are grateful to Professor Michael K. Ng for his helpful comments and providing us the code of [24] for numerical test.

References

[1] M. R. Banham and A. K. Katsaggelos, Digital image restoration, IEEE Signal Processing Magazine, 14 (1997), pp. 24-41.

[2] M. J. Black and A. Rangarajan, On the unification of line processes, outlier rejection and robust statistics with applications in early vision, Int. J. Comp. Vision, 19 (1996), pp. 57-91.

[3] J. V. Burke, A. S. Lewis, and M. L. Overton, A robust gradient sampling algorithm for nonsmooth, nonconvex optimization, SIAM J. Optim., 15 (2005), pp. 751-779.

[4] C. L. Chan, A. K. Katsaggelos and A. V. Sahakian, Image sequence filtering in quantum-limited noise with applications to low-dose fluoroscopy, IEEE Trans. Med. Imag., 12 (1993), pp. 610-621.

[5] C. Chen and O. L. Mangasarian, A class of smoothing functions for nonlinear and mixed complementarity problems, Comp. Optim. Appl., 5 (1996), pp. 97-138.

[6] T. Chan and S. Esedoglu, Aspects of Total Variation Regularized l_1 Function Approximation, Technical report, University of California at Los Angeles, 2004.

[7] X. Chen, L. Qi and D. Sun, Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities, Math. Comp., 67 (1998), pp. 519-540.

[8] X. Chen and Y. Ye, On homotopy-smoothing methods for box-constrained variational inequalities, SIAM J. Control Optim., 37 (1999), pp. 589-616.

[9] X. Chen, Smoothing methods for complementarity problems and their applications: a survey, J. Oper. Res. Soc. Japan 43 (2000), pp. 32–47.

[10] F. H. Clarke, Optimization and Nonsmooth Analysis, John Wiley and Sons, Inc., New York, 1983.

[11] L. C. Evans and R. F. Gariepi, Measure Theory and Fine Properties of Functions, CRC Press, Boca Raton, 1992.

[12] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc., 96 (2001), pp. 1348-1360.

[13] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, Efficient minimization methods of mixed $l_2 - l_1$ and $l_1 - l_1$ norms for image restoration, SIAM J. Sci. Comput., 27 (2006), pp. 1881-1902.

[14] L. Grippo, F. Lampariello and S. Lucidi, A nonmonotone line search technique for Newton's method, SIAM J. Numer. Anal., 23 (1986), pp. 707-716.

[15] A. Guitton and D. J. Verschuur, Adaptive subtraction of multiples using the l_1 norm, Geophys. Prospecting, 52 (2004), pp. 1-27.

[16] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM J. Optim., 16 (2005), pp. 170-192.

[17] W. W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific J. Optim., 2 (2006), pp. 35-58.

[18] P. J. Huber, Robust Statistics, John Wiley and Sons, New York, 1981.

[19] K. C. Kiwiel, Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization, SIAM J. Optim., 18 (2007), pp. 379-388.

[20] A.S. Lewis and M.L. Overton, Nonsmooth optimization via BFGS, Revised Version Submitted to SIAM J. Optimization, 2010.

[21] D. H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, J. Comput. Appl. Math., 129 (2001), pp. 15-35.

[22] M. Ng, R. Chan and W. Tang, A fast algorithm for deblurring models with Neumann boundary conditions, SIAM J. Sci. Comput., 21 (1999), pp. 851-866.

[23] M. Nikolova, Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers, SIAM J. Numer. Anal., 40 (2002), pp. 965-994.

[24] M. Nikolova, M. K. Ng, S. Zhang and W. Ching, Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization, SIAM J. Imaging Sciences, 1 (2008), pp. 2-25.

[25] J. Nocedal, Updating quasi-Newton matrixes with limited storage, Math. Comp., 35 (1980), pp. 773-782.

[26] B. Polak and G. Ribiere, Note surla convergence des méthodes de directions conjuguées, Rev. Francaise Imformat Recherche Opertionelle, 16 (1969), pp. 35-43.

[27] B. T. Polyak, The conjugate gradient method in extreme problems, USSR Comp. Math. Math. Phys., 9 (1969), pp. 94-112.

[28] L. Qi and X. Chen, A globally convergent successive approximation method for severely nonsmooth equations, SIAM J. Control Optim., 33 (1995), pp. 402-418.

[29] R. T. Rockafellar and R. J-B. Wets, Variational Analysis, Springer, Berlin, 1998.

[30] D. F. Shanno, Conjugate gradient methods with inexact searches, Math. Oper. Res., 3 (1978), pp. 244-256.

[31] C. H. Slump, Real-time image restoration in siagnostic X-ray imaging, the effects on quantum noise, in Proceedings of the 11th IAPR International Conference on Pattern Recognition, Vol. II, Conference B: Pattern Recongnition Methodology and Systems, 1992, pp. 693-696.

[32] J. Vlcek and L. Luksan, Globally convergent variable metric method for nonconvex nondifferentiable unconstrained minimization, J. Optim. Theory Appl., 111(2001), pp. 407-430.

[33] W. Wang, J. Yang, W. Yin and Y. Zhang, A new alternationg minimization algortihm for total variation image reconstruction, SIAM J. Imaging Sciences, 1 (2008), pp. 248-272.

[34] C. Zhang and X. Chen, Smoothing projected gradient method and its application to stochastic linear complementarity problems, SIAM J. Optim. 20 (2009), pp. 627-649.

[35] L. Zhang, W. Zhou and D. Li, A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, IMA J. Numer. Anal., 26 (2006), pp. 629-640.

[36] L. Zhang, W. Zhou and D. Li, Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search, Numer. Math., 104 (2006), pp. 561-572.