Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 10, No. 1, pp. 215–228, January 2014

A PROJECTED CONJUGATE GRADIENT METHOD FOR SPARSE RECONSTRUCTION WITH APPLICATIONS TO COMPRESSED SENSING*

Gaohang Yu, Jianlei Yan and Yi Zhou

Abstract: Sparse reconstruction has been widely involved in signal processing, compressed sensing and statistical learning. In order to realise the task, one need to solve a l_1-norm regularized least squares problem. Due to the nonsmooth objective function, the resulting problem is challenging. This paper presents a modified Polak-Ribière-Polyak (PRP) conjugate gradient method for sparse reconstruction with applications to compressed sensing. The construction of the method consists of two main steps: (1) reformulate the l_1 regularized least squares problem into an equivalent nonlinear system of monotone equations; (2) apply a projected PRP conjugate gradient method for solving the resulting system of monotone equations. Under suitable conditions, its global convergence result could be established. Moreover, the algorithm is easily implemented, in which only matrix-vector inner product is required at each step. Due to the low memory requirement of conjugate gradient approaches, the proposed method is practical and effective for sparse reconstruction. Numerical experiments indicate that PCGM is promising and competitive to the well-known iterative soft-thresholding (IST) method.

Key words: *sparse reconstruction, l_1-norm regularized least squares, compressed sensing, monotone equations, projected conjugate gradient method*

Mathematics Subject Classification: *65F22, 65K05, 90C30, 94A12*

1 Introduction

In the last few years, algorithms for finding sparse solutions to under-determined linear systems of equations have been intensively investigated in signal processing and compressed sensing. The fundamental principle of compressed sensing (CS) is that a sparse signal $\bar{x} \in R^n$ can be recovered from the under-determined linear system $b = A\bar{x}$, where $A \in R^{m\times n}$ (often $m \ll n$).

By defining l_0-norm ($\|x\|_0$) of a vector as the number of nonzero components in x, one natural way to reconstruct $\bar{x}$ from the system $b = A\bar{x}$ is to solve the following problem

$$\min_{x\in R^n} \|x\|_0 \text{ s.t. } b = Ax \tag{1.1}$$

*This work was partly supported by the National Natural Science Foundation of China (No.11001060, 61262026, 61263011), NCET Programm of the Ministry of Education (NCET-13-0738), Natural Science Foundation of Jiangxi Province (20132BAB201026), JGZX programm of Jiangxi Province (20112BCB23027), the science and technology programm of Jiangxi Education Committee (LDJH12088), Program in Sun Yat-sen University for the Fundamental Research Funds for the Central Universities (11ykpy07), Natural Science Foundation of Guangdong Province (S2011010005309).

via certain reconstruction technique. However, the l_0-norm problem is combinatorial and generally computationally intractable. A fundamental decoding model in CS is to replace l_0-norm by l_1-norm, which is defined as $\|x\|_1 = \sum_{i=1}^{n} |x(i)|$. The resulting adaptation of (1.1) is the so-called basis pursuit (BP) problem [5]

$$\min_{x \in R^n} \|x\|_1 \text{ s.t. } b = Ax. \tag{1.2}$$

It is shown that, problem (1.2) shares common solutions with (1.1) under some reasonable conditions [6]. When b contains some noise in most practical applications, certain relaxation to the equality constraint in (1.2) is desirable. One of the relaxations is the unconstrained basis pursuit denoising problem:

$$\min_{x \in R^n} \frac{1}{2}\|b - Ax\|_2^2 + \mu\|x\|_1. \tag{1.3}$$

The regularization parameter $\mu > 0$ provides a tradeoff between fidelity to the measurements and noise sensitivity.

The recent results indicate that if a signal is sparse or approximately sparse in some orthogonal basis, then we can obtain an accurate recovery when A is a random matrix projections [8]. In CS, encoding matrices A is often generated by randomly taking a subset of rows from orthonormal transform matrices, such as discrete Fourier transform (DFT), discrete cosine transform (DCT), and discrete Walsh-Hadamard transform (DWT) in magnetic resonance imaging (MRI) [18]. Such encoding matrices do not require storage and enable fast matrix-vector multiplications.

In recent years, quite a few algorithms have been proposed and studied for solving the aforementioned l_1-norm regularization problems arising in CS. One most widely studied first-order method is the iterative shrinkage/thresholding (IST) method [9, 11, 20, 21], which is designed for wavelet-based image deconvolution. In [16], Hale, Yin and Zhang derive the IST fixed-point continuation algorithm (FPC) via an operator splitting technique. TwIST [3] and FISTA [2] speed up the performance of IST and have virtually the same complexity but with better convergence properties. Another closely related method is the sparse reconstruction algorithm called SpaRSA was also studied by Wright, Nowak and Figueiredo in [25].

Gradient based algorithm is also prevalent for solving problem (1.3). One of the earliest gradient projection method for sparse reconstruction (GPSR) was developed by Figueiredo et al. [13]. SPGL1 [10] solves a least-squares problem with l_1-norm constraint by the spectral gradient projection method with an efficient Euclidean projection on l_1-norm ball. In [29], Yun and Toh studied a block coordinate gradient descent (CGD) method for solving (1.3). The projection steepest descent (PSD) method [7] also have good performance. Recently, the alternating direction method (ADM) have received much attention for solving total variation regularization problems for image restoration, and is also capable of solving the l_1-norm regularization problems in CS [27, 28]. Among all the methods mentioned above, GPSR splits x into two vectors, formulates (1.3) for a bound-constrained QP problem, and solves subsequently by using Barzilai-Borwein gradient method [4] with an efficient nonmonotone line search [15]. The resulting convex QP problem can be formulated to an equivalent non-smooth monotone equation [26]. Thanks to the low memory requirement, conjugate gradient methods are popular in dealing with large-scale optimization problem. During the past few years, descent conjugate gradient methods have been widely developed for solving unconstrained optimization and systems of nonlinear equations [1, 19, 32, 33, 34, 31].

One main motivation of this paper is to extend the well-known Polak-Ribière-Polyak (PRP) conjugate gradient method to solve the l_1 regularized least squares problem arising from compressed sensing. The construction of the method consists of two main steps: (1) reformulate the l_1 regularized least squares problem into an equivalent nonlinear system of monotone equations; (2) apply a descent Polak-Ribière-Polyak (DPRP) conjugate gradient method [31] with the projection technique [24] for solving the resulting system of monotone equations. Under reasonable assumptions, we can establish its convergence result. Numerical experiments illustrate that the proposed method is efficient to recover a sparse signal arising in compressive sensing and performs better than its competitors.

This paper is organized as follows. A full description of the projected conjugate gradient method for CS is presented in the next section. Meanwhile, we establish its global convergence under some suitable conditions. In Section 3, we present some numerical experiments to show the practical performance of the proposed algorithm. Finally, we have a conclusion section.

2 Proposed Algorithm and the Global Convergence

Firstly, we briefly review the process for constructing a quadratic programming problem in [13]. Making a substitution, for any vector $x \in R^n$, it can be formulated for $x = u - v$, where $u \geq 0, u \in R^n$, $v \geq 0, v \in R^n$ and $u_i = \max\{0, x_i\}$, $v_i = \max\{0, -x_i\}$. Hence (1.3) can be rewritten as the following bound-constrained quadratic programming

$$\min_{u,v\in R^n} \frac{1}{2}\|b - A(u-v)\|_2^2 + \mu(\mathbf{I}_n^T u + \mathbf{I}_n^T v) \quad \text{s.t. } u \geq 0,\ v \geq 0, \tag{2.1}$$

where $\mathbf{I}_n^T$ represents the transpose of $\mathbf{I}_n$, and $\mathbf{I}_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ is a vector consisting of n ones. Particularly, it follows from [13] that (2.1) can be rewritten as the following form

$$\min_{z\in R^{2n}} \frac{1}{2} z^T H z + c^T z \quad \text{s.t. } z \geq 0, \tag{2.2}$$

where $z = \begin{bmatrix} u \\ v \end{bmatrix}$, $y = A^T b$, $c = \mu \mathbf{I}_{2n} + \begin{bmatrix} -y \\ y \end{bmatrix}$ and $H = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix}$.

Recently, Xiao et al. [26] indicated that (2.2) can be transformed into the following form

$$F(z) = \min\{z, Hz + c\} = 0, \tag{2.3}$$

where function F is vector-valued, and the "min" is interpreted as componentwise minimum. Without specific statements, $\|\cdot\|$ denotes the Euclidean norm in the following paper.

The following lemma shows that $F(\cdot)$ is Lipschitz continuous.

Lemma 2.1 (Lemma 3 in [22]). *There exists a positive constant L such that*

$$\|F(z_1) - F(z_2)\| \leq L\|z_1 - z_2\|,\ \forall\, z_1, z_2 \in R^{2n}. \tag{2.4}$$

The following lemma shows that $F(\cdot)$ is monotone.

Lemma 2.2 (Lemma 2.2 in [26]). *The mapping $F(\cdot)$ is monotone, i.e.,*

$$(F(z_1) - F(z_2))^T (z_1 - z_2) \geq 0,\ \forall\, z_1, z_2 \in R^{2n}. \tag{2.5}$$

The above two lemmas illustrate that the system of nonlinear equations has nice properties, and it can be solved efficiently by some derivative-free methods [14, 30, 33].

In this paper, based on a descent PRP conjugate gradient method [31], we propose a projected conjugate gradient method for the minimization of l_1-regularized minimization problem with applications to CS. In particular, the search direction is generated by the following way

$$d_k = \begin{cases} -F_0 & \text{if } \ k = 0, \\ -F_k + \beta_k d_{k-1} - \theta_k y_{k-1} & \text{if } \ k \geq 1, \end{cases} \tag{2.6}$$

where $F_k = F(z_k)$, $\beta_k = \lambda_{k-1}\beta_k^{DPRP}$, $\beta_k^{DPRP} = \beta_k^{PRP} - \frac{C\|y_{k-1}\|^2}{\|F_{k-1}\|^4} F_k^T d_{k-1}$ with parameter $C = 1$, and $\beta_k^{PRP} = \frac{F_k^T y_{k-1}}{\|F_{k-1}\|^2}$ is the well-known PRP conjugate gradient formula, $\theta_k = \frac{\lambda_{k-1}^2 (F_k^T y_{k-1})\|d_{k-1}\|^2}{\|F_{k-1}\|^4}$, $y_{k-1} = F(z_k) - F(z_{k-1})$ and λ_{k-1} is the steplength determined by a line search technique.

A full description of the projected conjugate gradient method (PCGM) can be presented as follows.

Projected Conjugate Gradient Method (PCGM) for solving (2.3)

Date: Give initial point $z_0 \in R^{2n}$, set parameters $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho \in (0, 1)$.

Convergence test: If $\|F(z_0)\| = 0$, then stop. Else set $d_0{=}{-F(z_0)}$. Let $k := 0$.

Line search update: Determine the steplength λ_k and set $t_k = z_k + \lambda_k d_k$, where $\lambda_k = \sigma_1 \rho^{m_k}$ with m_k being the smallest nonnegative integer m satisfying

$$-F(t_k)^T d_k \geq \sigma_2 \sigma_1 \rho^m \|F(t_k)\| \|d_k\|^2. \tag{2.7}$$

Projection update: Compute

$$z_{k+1} = z_k - \frac{F(t_k)^T (z_k - t_k)}{\|F(t_k)\|^2} F(t_k). \tag{2.8}$$

If $\|F(z_{k+1})\| = 0$, then stop. Else let $k := k + 1$ and compute d_k defined by (2.6). Then go to the Convergence Test.

The following lemma states that PCGM method satisfies the sufficient descent condition.

Lemma 2.3. *The introduced direction in the Algorithm PCGM satisfies the sufficient descent condition, i.e. there exists a constant $\tau > 0$ such that $F_k^T d_k \leq -\tau \|F_k\|^2$ holds.*

Proof. From (2.6), it follows that

$$\begin{aligned} F_k^T d_k = & \frac{-\|F_k\|^2 \|F_{k-1}\|^4 + \lambda_{k-1}(F_k^T d_{k-1})(F_k^T y_{k-1})\|F_{k-1}\|^2}{\|F_{k-1}\|^4} \\ & - \frac{\lambda_{k-1}\|y_{k-1}\|^2 (F_k^T d_{k-1})^2 + \lambda_{k-1}^2 (F_k^T y_{k-1})^2 \|d_{k-1}\|^2}{\|F_{k-1}\|^4}. \end{aligned} \tag{2.9}$$

Setting $u = \frac{1}{\sqrt{2}}\|F_{k-1}\|^2 F_k$ and $v = \sqrt{2}\lambda_{k-1}(F_k^T y_{k-1}) d_{k-1}$ in the second term of the numerator and using the inequality $u^T v \leq \frac{1}{2}(u^2 + v^2) = \frac{1}{2}(\frac{1}{2}\|F_{k-1}\|^4 \|F_k\|^2 + 2\lambda_{k-1}^2 (F_k^T y_{k-1})^2 \|d_{k-1}\|^2)$, it straightforwardly implys

$$\begin{aligned} F_k^T d_k \leq & \frac{-\|F_k\|^2 \|F_{k-1}\|^4 + \frac{1}{2}(\frac{1}{2}\|F_{k-1}\|^4 \|F_k\|^2 + 2\lambda_{k-1}^2 (F_k^T y_{k-1})^2 \|d_{k-1}\|^2)}{\|F_{k-1}\|^4} \\ & - \frac{\lambda_{k-1}^2 (F_k^T y_{k-1})^2 \|d_{k-1}\|^2}{\|F_{k-1}\|^4} = -\frac{3}{4}\|F_k\|^2. \end{aligned} \tag{2.10}$$

Therefore, the introduced direction satisfies the sufficient descent condition. □

The following lemma indicates that if the sequence $\{z_k\}$ be generated by Algorithm PCGM and $\overline{z}$ such that $F(\overline{z}) = 0$, then the sequence $\{z_k - \overline{z}\}$ is decreasing and convergent, thus the sequence $\{z_k\}$ is bounded.

Lemma 2.4. *Suppose the sequence $\{z_k\}$ be generated by Algorithm PCGM. For any $\overline{z}$ such that $F(\overline{z}) = 0$, we have that*

$$\|z_{k+1} - \overline{z}\|^2 \leq \|z_k - \overline{z}\|^2 - \|z_{k+1} - z_k\|^2. \tag{2.11}$$

In particular, the sequence $\{z_k\}$ is bounded and

$$\sum_{k=0}^{\infty} \|z_{k+1} - z_k\|^2 < \infty. \tag{2.12}$$

Proof. Using the monotonicity property of $F(z)$, for any $\overline{z}$ such that $F(\overline{z}) = 0$, it can be concluded that

$$F(t_k)^T(\overline{z} - t_k) \leq 0 \tag{2.13}$$

while $t_k = z_k + \lambda_k d_k$. On the other hand, due to the line search technique (2.7), we have

$$F(t_k)^T(z_k - t_k) > 0. \tag{2.14}$$

It means that the hyperplane

$$H_k = \left\{ z \in R^n | F(t_k)^T(z - t_k) = 0 \right\} \tag{2.15}$$

strictly separates the current iterate z_k from $\overline{z}$. It is also easy to verify that z_{k+1} is the projection of z_k onto the halfspace $\left\{ z \in R^n | F(t_k)^T(z - t_k) \leq 0 \right\}$. Since z_{k+1} belongs to this halfspace, it follows from the basic properties of the projection operator (see p.121 in [23]) that $\langle z_k - z_{k+1}, z_{k+1} - \overline{z} \rangle \geq 0$. Therefore

$$\begin{aligned} \|z_k - \overline{z}\|^2 &= \|z_k - z_{k+1}\|^2 + \|z_{k+1} - \overline{z}\|^2 + 2 \langle z_k - z_{k+1}, z_{k+1} - \overline{z} \rangle \\ &\geq \|z_k - z_{k+1}\|^2 + \|z_{k+1} - \overline{z}\|^2. \end{aligned}$$

Hence the sequence $\{\|z_k - \overline{z}\|\}$ is nonincreasing and convergent, therefore the sequence $\{z_k\}$ is bounded, and also

$$\lim_{x \to \infty} \|z_{k+1} - z_k\| = 0. \tag{2.16}$$

This completes the proof. □

Lemma 2.5. *Suppose the sequence of directions $\{d_k\}$ be generated by the Algorithm PCGM. Also there exists a constant $\varepsilon_0 > 0$ such that*

$$\|F_k\| \geq \varepsilon_0 \tag{2.17}$$

for all $k \in N \cup \{0\}$, then the directions $\{d_k\}$ are bounded, i.e. there exists a constant $M > 0$ such that

$$\|d_k\| \leq M \tag{2.18}$$

for all $k \in N \cup \{0\}$.

Proof. From (2.8), it can be concluded that

$$\|z_{k+1}-z_k\| = \frac{\left|F(t_k)^T(z_k-t_k)\right|}{\|F(t_k)\|} = \frac{\left|\lambda_k F(t_k)^T d_k\right|}{\|F(t_k)\|}. \tag{2.19}$$

Relation (2.7) together with (2.19) directly imply

$$\|z_{k+1}-z_k\| = \frac{\left|F(t_k)^T(z_k-t_k)\right|}{\|F(t_k)\|} = \frac{\left|\lambda_k F(t_k)^T d_k\right|}{\|F(t_k)\|} \geq \sigma_2\lambda_k^2\|d_k\|^2 \geq 0. \tag{2.20}$$

This fact and Lemma 4 suggest that

$$\lim_{k\to\infty} \lambda_k \|d_k\| = 0. \tag{2.21}$$

Under other circumstances, from (2.19) and the Cauchy–Schwartz inequality, we also have

$$\|z_{k+1}-z_k\| \leq \frac{\lambda_k \|F(t_k)\| \|d_k\|}{\|F(t_k)\|} = \lambda_k \|d_k\|. \tag{2.22}$$

Using (2.6), (2.17) and the Cauchy–Schwartz inequality, it follows that

$$\begin{aligned}\|d_k\| \leq \|F_k\| &+ \tfrac{\|F_k\|\|y_{k-1}\|}{\varepsilon_0^2} + \tfrac{\lambda_{k-1}\|y_{k-1}\|^2\|F_k\|\|d_{k-1}\|^2}{\varepsilon_0^4} \\ &+ \tfrac{\lambda_{k-1}^2\|y_{k-1}\|^2\|F_k\|\|d_{k-1}\|^2}{\varepsilon_0^4}\end{aligned} \tag{2.23}$$

for all $k \in N$. Now, by (2.4), there exist two constants $L > 0$ and $\kappa > 0$ such that

$$\|y_{k-1}\| \leq L\|z_{k+1}-z_k\| \;,\; \|F_k\| \leq \kappa. \tag{2.24}$$

As a consequence of (2.21), there exist a constant $\varepsilon_1 \in (0,1)$ and a positive integer k_0 such that

$$\lambda_{k-1}\|d_{k-1}\| < \sqrt[3]{\varepsilon_0^4\varepsilon_1/L^2\kappa},\; for\ all\ k > k_0. \tag{2.25}$$

These inequalities together with (2.23) lead to

$$\|d_k\| \leq \kappa + \frac{L\kappa\sqrt[3]{\varepsilon_0^4\varepsilon_1/L^2\kappa}}{\varepsilon_0^2} + \varepsilon_1\|d_{k-1}\| + \frac{L^2\kappa(\sqrt[3]{\varepsilon_0^4\varepsilon_1/L^2\kappa})^4}{\varepsilon_0^4}. \tag{2.26}$$

Let $r = \kappa + \frac{L\kappa\sqrt[3]{\varepsilon_0^4\varepsilon_1/L^2\kappa}}{\varepsilon_0^2} + \frac{L^2\kappa(\sqrt[3]{\varepsilon_0^4\varepsilon_1/L^2\kappa})^4}{\varepsilon_0^4}$, and we have, for any $k > k_0$,

$$\|d_k\| \leq r + \varepsilon_1\|d_{k-1}\| \leq r(1+\varepsilon_1+\varepsilon_1^2+\cdots+\varepsilon_1^{k-k_0-1}) + \varepsilon_1^{k-k_0}\|d_{k_0}\| \leq \frac{r}{1-\varepsilon_1} + \|d_{k_0}\|.$$

Then let $M = \max\left\{\|d_0\|, \|d_1\|, \cdots, \|d_{k_0}\|, \frac{r}{1-\varepsilon_1} + \|d_{k_0}\|\right\}$ to get (2.18). □

The following Lemma guarantees that the recommended line search in Algorithm PCGM is well-defined.

Lemma 2.6. *Suppose that all conditions of Lemma 5 hold. Then the line search procedure (2.7) in Algorithm PCGM is well-defined.*

Proof. Our aim is to show that the line search procedure (2.7) terminates finitely with a positive steplength λ_k. By contradiction, suppose that for some iterate indexes such as $\hat{k}$, the condition (2.7) is not held. As a result, by setting $\lambda_{\hat{k}}^{(m)} = \sigma_1 \rho^m$, it can be conclude

$$-F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}})^T d_{\hat{k}} < \sigma_2 \lambda_{\hat{k}}^{(m)} \left\| F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}}) \right\| \|d_{\hat{k}}\|^2$$

for all $m \in N \cup \{0\}$. This fact along with the Lipschitz continuous Lemma 1 and the sufficient descent condition Lemma 3 express

$$\begin{aligned} \tau \|F_{\hat{k}}\|^2 \le -F_{\hat{k}}^T d_{\hat{k}} &= \left[F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}}) - F_{\hat{k}}\right]^T d_{\hat{k}} - F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}})^T d_{\hat{k}} \\ &< \left[L + \sigma_2 \left\| F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}}) \right\| \right] \lambda_{\hat{k}}^{(m)} \|d_{\hat{k}}\|^2 \end{aligned} \tag{2.27}$$

for all $m \in N \cup \{0\}$. On the other hand, using Lemma 1, (2.24) and Lemma 5, it directly follows that

$$\begin{aligned} \left\| F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}}) \right\| &\le \left\| F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}}) - F_{\hat{k}} \right\| + \|F_{\hat{k}}\| \\ &\le L \lambda_{\hat{k}}^{(m)} \|d_{\hat{k}}\| + \kappa \ \le \sigma_2 L M + \kappa = \kappa_1 \end{aligned}$$

where $\kappa_1 = \sigma_2 L M + \kappa$. This inequality together with (2.17), (2.18) and (2.27) yield

$$\lambda_{\hat{k}}^{(m)} > \frac{\tau \|F_{\hat{k}}\|^2}{\left[L + \sigma_2 \left\| F(z_{\hat{k}} + \lambda_{\hat{k}}^{(m)} d_{\hat{k}}) \right\| \right] \|d_{\hat{k}}\|^2} \ge \frac{\tau \varepsilon_0^2}{(L + \sigma_2 \kappa_1) M^2} > 0$$

for all $m \in N \cup \{0\}$. This contradicts the definition of $\lambda_{\hat{k}}^{(m)}$. Consequently, the line search procedure (2.7) can attain a positive steplength λ_k in a finite number of backtracking repetitions in step **line search update**. Therefore, the line search is well-defined. □

Assume that the solution set of (2.3) is nonempty, we can establish the following global convergence theorem for PCGM method.

Theorem 2.7. *Suppose that the sequence $\{z_k\}$ is generated by Algorithm PCGM, then it holds that*

$$\lim_{k \to \infty} \inf \|F_k\| = 0. \tag{2.28}$$

Proof. By contradiction, we assume that (2.28) does not hold. Then there exists a constant $\varepsilon_0 > 0$ such that $\|F_k\| \ge \varepsilon_0$ holds, for all $k \in N \cup \{0\}$. From Lemma 3 we know

$$F_k^T d_k \le -\tau \|F_k\|^2 \tag{2.29}$$

and we also have

$$\tau \|F_k\|^2 \le -F_k^T d_k \le \|F_k\| \, \|d_k\|$$

for all $k \in N \cup \{0\}$. Hence

$$\|d_k\| \ge \tau \varepsilon_0 > 0$$

for all $k \in N \cup \{0\}$. According to this condition and (2.21), it follows that

$$\lim_{k \to \infty} \lambda_k = 0. \tag{2.30}$$

From Line search update of Algorithm PCGM, it is easy to see that for $\lambda_k^* = \rho^{-1}\lambda_k$, the condition (2.7) does not hold, i.e.

$$-F(z_k + \lambda_k^* d_k)^T d_k < \sigma_2 \lambda_k^* \|F(z_k + \lambda_k^* d_k)\| \, \|d_k\|^2. \tag{2.31}$$

As a result of the boundedness of $\{z_k\}$ in Lemma 4, there exist an accumulation point $\bar{z}$ and an infinite index set K_1 such that $\lim_{k\to\infty} z_k = \bar{z}$, for $k \in K_1$. Meanwhile, it follows from Lemma 5 that there exist an infinite index set $K_2 \subset K_1$ and an accumulation point $\bar{d}$ such that $\lim_{k\to\infty} d_k = \bar{d}$, for $k \in K_2$. Therefore, taking the limit as $k \to \infty$ in both sides of (2.31) for all $k \in K_2$ gives

$$F(\bar{z})^T \bar{d} > 0. \tag{2.32}$$

In the other point, by taking the limit as $k \to \infty$ in both sides of (2.29) for $k \in K_2$, obviously

$$F(\bar{z})^T \bar{d} \le 0.$$

This yields a contradiction with (30). Therefore, the proof is completed. □

3 Experimental Results

In this section, we present some numerical experiments to illustrate the practical performance of PCGM method, and compare PCGM with SGCS [26], DFPB1 [1], IST [11], MPRP [19] for sparse reconstruction. All experiments are tested in Matlab R2010a. We consider a typical compressive sensing scenario, where the goal is to reconstruct a n length sparse signal with k non-zero elements from m observations. We restrict our attention to the l_1 regularized least squares model (1.3), and use $f(x) = \frac{1}{2}\|b - Ax\|_2^2 + \mu\|x\|_1$ as the merit function. The random A is the Gaussian matrix whose elements are generated from shape $i.i.d.$ normal distributions $\mathcal{N}(0,1)$ (randn(m,n) in Matlab). In real applications, the measurement b is usually contaminated by noise, i.e., $b = Ax + \eta$, where η is the Gaussian noise distributed as $\mathcal{N}(0,\sigma^2 I)$.

In PCGM, we let $\rho = 0.1$, $\sigma_1 = 0.95$ and $\sigma_2 = 0.93$. The common stopping criterion is $\frac{\|f(x_k)-f(x_{k-1})\|}{\|f(x_{k-1})\|} < 10^{-4}$, where $f(x_k)$ denotes the function value at x_k. Mean squared error (MSE) is used to measure the quality of the reconstructive signals, which is defined as MSE $= \|\hat{x} - \bar{x}\|^2/n$, where $\hat{x}$ denotes the reconstructive signal, $\bar{x}$ denotes the original signal and n is the length of the signal.

Firstly, we test the practical performance of PCGM for reconstructing sparse signal. Fig.1 and Fig.2 report that the results of PCGM for a signal sparse reconstruction in CS. Comparing the first and the last plots in Fig.1, we can see that the original sparse signal is restored almost exactly. As shown in the right plot in Fig.2, all the blue dots are circled by the red circles, which illustrates that the original signal has been found almost exactly.

Secondly, we report some numerical comparison to some exists gradient methods for solving monotone equations arising from CS. In Fig.3, we visually illustrate the performance of PCGM with comparison to the recent solver SGCS [26] and two related conjugate gradient methods, MPRP [19], DFPB1 [1]. As we can see, PCGM requires less computing time than SGCS, MPRP and DFPB1 while obtaining similar reconstructive quality.

In the next experiment, we choose three different signals and four different values of noise level σ^2 to compare PCGM with the well known first-order method IST [11]. Numerical results are listed in Table 1, in which we report the number of iterations (Iter), the CPU time in seconds (Time) required for the whole reconstructing process, the means of squared

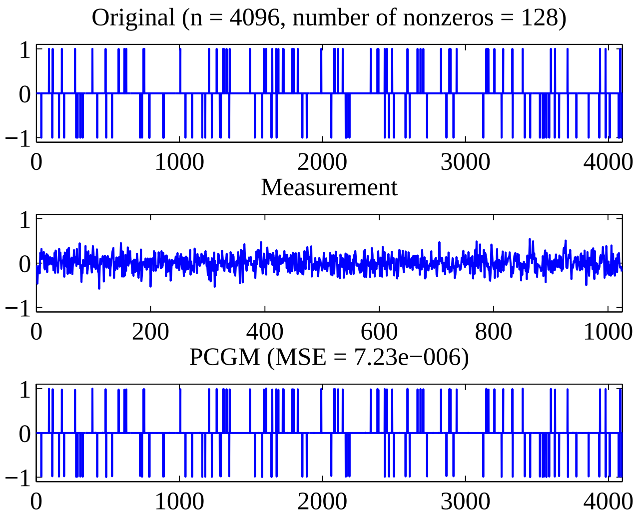

Figure 1: Top: original signal with length 4096 and 128 non-zero elements. Middle: noisy measurement with length 1024. Bottom: recovered signal by PCGM when $\sigma^2 = 10^{-3}$.

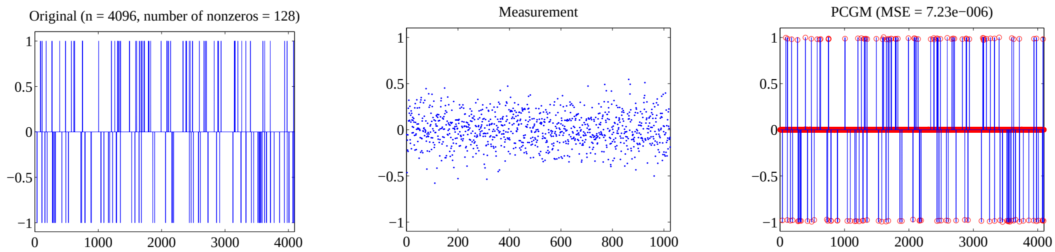

Figure 2: Left: original signal with length 4096 and 128 non-zero elements. Middle: noisy measurement with length 1024. Right: recovered signal by PCGM (red circle) versus original signal (blue peaks) when $\sigma^2 = 10^{-3}$, the error MSE= 7.231e-6, the CPU time is 3.67 seconds, and requires 143 iterations.

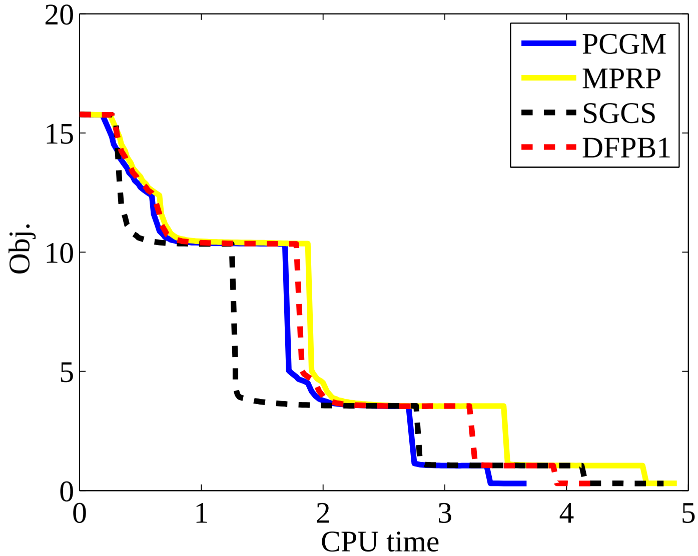

Figure 3: Comparison result of PCGM and other methods with n=4096, m=1024, k=128 and $\sigma^2 = 10^{-3}$. The x-axes represent the CPU time in seconds. The y-axes represent the function values.

error to every original signal $\bar{x}$ (MSE), where n is the length of the signal, m is the length of noisy measurement and k is the number of non-zero elements. As we can see from Table 1, PCGM requires less iterations and less CPU time than that of IST and SGCS while deriving the similar quality of restoration.

4 Concluding Remark

Sparse reconstruction has been widely involved in signal processing, compressed sensing and statistical learning. In order to realise the task, one need to solve a l1-norm regularized least squares problem. This paper presents a projected conjugate gradient method (PCGM) for sparse reconstruction with applications to compressed sensing. The construction of the method consists of two main steps: (1) reformulate the l1 regularized least squares problem into an equivalent nonlinear system of monotone equations; (2) apply a projected PRP conjugate gradient method for solving the resulting system of monotone equations. The algorithm is easily implemented, in which only matrix-vector inner product is required at each step. Due to the low memory requirement of conjugate gradient approaches, the proposed method is practical and effective for sparse reconstruction. Numerical experiments indicate that PCGM is promising and competitive to the well-known iterative soft-thresholding (IST) method. One future topic is to investigate the applications of PCGM in image restoration or image reconstruction.

Acknowledgements

The authors would like to thank three anonymous referees for their comments and suggestions on the first version of the article, which lead to significant improvements of the

Table 1: performance of signal reconstruction with different length of signals and different noise level σ^2 in the number of iterations, CPU time and the means of squared error.

σ^2	$n/m/k$	Iter			Time			MSE		
		PCGM	IST	SGCS	PCGM	IST	SGCS	PCGM	IST	SGCS
10^{-1}	$2^{10}/2^8/2^5$	117	417	288	0.17	0.33	0.42	1.14e-2	1.53e-2	1.13e-2
10^{-1}	$2^{11}/2^9/2^6$	125	403	305	0.84	1.84	1.98	1.32e-2	1.60e-2	1.34e-2
10^{-1}	$2^{12}/2^{10}/2^7$	207	419	303	5.02	7.30	7.41	1.06e-2	1.43e-2	1.07e-2
10^{-2}	$2^{10}/2^8/2^5$	102	788	202	0.20	0.77	0.31	1.28e-4	3.24e-4	1.44e-4
10^{-2}	$2^{11}/2^9/2^6$	121	686	178	0.92	3.22	1.48	1.08e-4	2.38e-4	1.18e-4
10^{-2}	$2^{12}/2^{10}/2^7$	163	769	207	3.91	13.61	5.30	1.22e-4	3.52e-4	1.36e-4
10^{-3}	$2^{10}/2^8/2^5$	95	750	201	0.19	0.77	0.33	6.73e-6	8.63e-6	9.57e-6
10^{-3}	$2^{11}/2^9/2^6$	108	698	182	0.81	3.39	1.39	5.34e-6	6.77e-6	7.44e-6
10^{-3}	$2^{12}/2^{10}/2^7$	143	655	198	3.67	11.53	4.80	7.23e-6	8.58e-6	9.50e-6
10^{-4}	$2^{10}/2^8/2^5$	125	885	215	0.23	0.83	0.31	4.35e-6	5.58e-6	6.68e-6
10^{-4}	$2^{11}/2^9/2^6$	129	658	207	0.91	3.11	1.38	7.93e-6	8.81e-6	1.01e-5
10^{-4}	$2^{12}/2^{10}/2^7$	163	810	218	4.09	13.67	5.22	5.88e-6	7.18e-6	8.05e-6

presentation.

References

[1] M. Ahookhosh, K. Amini and S. Bahrami, Two derivative-free projection approaches for systems of large-scale nonlinear monotone equations, *Numerical Algorithms* 64 (2013) 21–42.

[2] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences* 2 (2009) 83–202.

[3] J. Bioucas-Dias and M. Figueiredo, A new TwIST: Two-step iterative thresholding algorithm for image restoration, *IEEE Transactions on Image Processing* 16 (2007) 2992–3004.

[4] J. Barzilai and J. Borwein, Two point step size gradient methods, *IMA Journal of Numerical* 8 (1988) 141–148.

[5] S. Chen, D.L. Donoho and M. Saunders, Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing* 20 (1998) 33–61.

[6] D. Donoho, For most large underdetemind systems of linear equations the minimal l_1-norm solution is also the sparse stsolution, *Communications on pure and applied mathematics* 59 (2006) 797–829.

[7] I. Daubechies, M. Fornasier and I. Loris, Accelerated projection gradient method for linear inverse problems with sparsity constraints, *Journal of Fourier Analysis and Applications* 14 (2008) 764–792.

[8] M.F. Duarte and Y.C. Eldar, Structured compressed sensing: from theory to applications, *IEEE Transactions on Signal Processing* 59 (2011) 4053–4085.

[9] M. Figueiredo and R. Nowak, An EM algorithm for wavelet-based image restoration, *IEEE Transactions on Image Processing* 12 (2003) 906–916.

[10] M. Friedlander and E. VandenBerg, Probing the pareto frontier for basis pursuit solutions, *SIAM Journal on Scientific Computing* 31 (2008) 890–912.

[11] M.A. Figueiredo and R.D. Nowak, A bound optimization approach to wavelet-based image deconvolution in *IEEE International Conference on Image Processing. ICIP 2005*, vol. 2, 2005, pp. 782–785.

[12] G. Fasano, F. Lampariello and M. Sciandrone, A truncated nonmonotone Gauss-Newton method for large-scale nonlinear least-squares problems, *Computational Optimization and Applications* 34 (2006) 343–358.

[13] M.A.T. Figueiredo, R.D. Nowak and S.J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE Journal of Selected Topics in Signal Processing* 1 (2007) 586–598.

[14] L. Grippo and M. Sciandrone, Nonmonotone derivative-free methods for nonlinear equations, *Computational Optimization and Applications* 37 (2007) 297–328.

[15] L. Grippo, F. Lampariello, and S. Lucidi, A nonmonotone line search technique for Newtons method, *SIAM Journal on Numerical Analysis* 23 (1986) 707–716.

[16] E. Hale, W. Yin, and Y. Zhang, Fixed-point continuation for l_1-minimization: methodology and convergence, *SIAM Journal on Optimization* 19 (2008) 1107–1130.

[17] N.A. Iusem and V.M. Solodov, Newton-type methods with generalized distances for constrained optimization, *Optimization* 41 (1997) 257–278.

[18] M. Lustig, D. Donoho and J.M.Pauly, Sparse MRI: the application of compressed sensing for rapid MR Imageing, *Magnetic resonance in medicine* 58 (2007) 1182–1195.

[19] Q. Li and D.H. Li, A class of derivative-free methods for large-scale nonlinear monotone equations, *IMA journal of numerical analysis* 31 (2011) 1625–1635.

[20] C.De. Mol and M. Defrise, A note on wavelet-based inversion algorithms, *Contemporary Mathematics* 313 (2002) 85–96.

[21] R. Nowak and M. Figueiredo, Fast wavelet-based image deconvolution using the EM algorithm, in *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, vol. 1, 2001, pp. 371–375.

[22] J.S. Pang, Inexact Newton methods for the nonlinear complementary problem, *Mathematical Programming* 36 (1986) 54–71.

[23] B.T. Polyak, *Introduction to Optimization*, Optimization Software, Publications Division, New York, 1987.

[24] M.V. Solodov and B.F. Svaiter, A globally convergent inexact Newton method for systems of monotone equations, in *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, M. Fukushima, L. Qi (eds), Kluwer Academic Publishers, Boston, 1998, pp. 355–369.

[25] S. Wright, R. Nowak and M. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Transactions on Signal Processing* 57 (2009) 2479–2493.

[26] Y.H. Xiao, Q.Y. Wang and Q.J. Hu, Non-smooth equations based method for l_1-norm problems with applications to compressed sensing, *Nonlinear Analysis: Theory, Methods & Applications* 74 (2011) 3570–3577.

[27] J. Yang, Y. Zhang and W. Yin, A fast TV L1-L2 minimization algorithm for signal reconstruction from partial Fourier data, *IEEE Journal of Selected Topics in Signal Processing* 4 (2010) 288–297.

[28] J. Yang and Y. Zhang, Alternating direction algorithms for l_1-problems in compressive sensing, *SIAM journal on scientific computing* 33 (2011) 250–278.

[29] S. Yun and K.C. Toh, Acoordinate gradient descent method for l_1-regularized convex minimization, *Computational Optimization and Applications* 48 (2011) 273–307.

[30] G.H. Yu, S.Z. Niu and J.H. Ma, Multivariate spectral gradient projection method for nonlinear montone equations with convex constraints, *Journal of Industrial and Management Optimization* 9 (2013) 117–129.

[31] G.H. Yu, L.T. Guan and W.F. Chen, Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization, *Optimization Methods and Software* 23 (2008) 275–293.

[32] G.H. Yu, A derivative-free method for solving large-scale nonlinear systems of equations, *Journal of Industrial and Management Optimization* 6 (2010) 149–160.

[33] Q.R. Yan, X.Z. Peng and D.H. Li, A globally convergent derivative-free method for solving large-scale nonlinear monotone equations, *Journal of Computational and Applied Mathematics* 234 (2010) 649–657.

[34] L. Zhang, W. Zhou and D.H. Li, Some descent three-term conjugate gradient methods and their global convergence, *Optimization Methods and Software* 22 (2009) 697–711.

Manuscript received 31 March 2013
revised 5 November 2013
accepted for publication 2 December 2013

Gaohang Yu
School of Mathematics and Computer Sciences
GanNan Normal University Ganzhou, 341000, China
E-mail address: maghyu@163.com

Jianlei Yan
School of Mathematics and Computer Sciences
GanNan Normal University Ganzhou, 341000, China
E-mail address: mathyjl2012@163.com

Yi Zhou
Department of Biomedical Engineering, Zhongshan School of Medicine
Sun Yat-Sen University, Guangzhou, 510080, China
E-mail address: zhouyi@mail.sysu.edu.cn