A Projected Conjugate Gradient Method for Compressive Sensing

Yuanying Qiu[1], Wei Xue[2,⋆], and Gaohang Yu[2]

[1] School of Foreign Languages, Gannan Normal University, Ganzhou 341000, People's Republic of China
[2] School of Mathematics and Computer Science, Gannan Normal University, Ganzhou 341000, People's Republic of China
wxmaths@163.com

Abstract. Frequently, the most important information in a signal is much sparser than the signal itself. In this paper, we study a projected conjugate gradient method for finding sparse solutions to an undetermined linear system arising from compressive sensing. The construction of this method consists of two main phases: (1) reformulate a l_1 regularized least squares problem into an equivalent nonlinear system of monotone equations; (2) apply a conjugate gradient method with projection strategy to the resulting system. The derived method only needs matrix-vector products at each step and could be easily implemented. Global convergence result is established under some suitable conditions. Numerical results demonstrate that the proposed method can improve the computation time while obtaining similar reconstructed quality.

1 Introduction

Compressive sensing (CS) is an emerging field and is attracting considerable research interest in signal processing community. The fundamental principle of CS is that a sparse signal $\bar{x} \in R^n$ can be recovered from the undetermined linear system $y = \Phi\bar{x}$, where $\Phi \in R^{m\times n}$ (often $m \ll n$). By defining l_0 norm ($\|x\|_0$) of a vector as the number of nonzero elements in x, one natural way to recover $\bar{x}$ from the system is to solve the following problem

$$\min_{x\in R^n} \|x\|_0 \text{ s.t. } y = \Phi x. \tag{1}$$

However, the l_0 norm problem is computationally intractable. An alternative model 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) is the Basis Pursuit (BP) problem [1]

$$\min_{x\in R^n} \|x\|_1 \text{ s.t. } y = \Phi x. \tag{2}$$

Optimization methods often find a solution of (1) by solving the following closely related l_1 regularized least squares problem

$$\min_{x\in R^n} \frac{1}{2}\|y - \Phi x\|_2^2 + \mu\|x\|_1. \tag{3}$$

⋆ Corresponding author.

J. Yang, F. Fang, and C. Sun (Eds.): IScIDE 2012, LNCS 7751, pp. 398–406, 2013.

Here, $\mu > 0$ is related to the Lagrange multiplier of the constraint in (2).

It follows from some existing results that if a signal is sparse or approximately sparse in some orthogonal basis, then an accurate recovery can be obtained when Φ is a random matrix projections [3]. Various types of methods have been proposed to solve the l_1 regularized minimization problem. Recently, some first-order methods are popular for solving (3), such as the projection steepest descent method [2], the gradient projection algorithm (GPSR) proposed by Figueiredo et al. [5], and so on. In this paper, we mainly focus on developing an iterative method for solving l_1 regularized problem arising in CS. Among all the methods mentioned above, GPSR method firstly splits vector x into two vectors, reformulates (3) into a bound-constrained quadratic programming problem, and solves it by using the well-known BB stepsize. In [8], the authors notice that the quadratic programming problem is equivalent to a system of nonlinear equations. We use a projected conjugate gradient method to solve the resulting monotone equations in this paper. Our method has two main phases. In the first phase, a l_1 regularized least squares problem (3) is transformed into an equivalent nonlinear system of monotone equations. And then a projected conjugate gradient method is introduced to solve the equivalent system in the second phase.

The rest of this paper is organized as follows. We present the full description of the proposed algorithm in the next section. In Section 3, we establish its global convergence under some suitable conditions. We report some numerical experiments to illustrate the efficiency of the proposed method in Section 4. Some conclusions are drawn in Section 5.

2 Proposed Algorithm

We state our algorithm in this section. Firstly, we recall the approach of constructing a quadratic programming problem in [5]. Making a substitution, for any vector $x \in R^n$, it can be formulated as $x = u - v$, where $u \geq 0, u \in R^n$, $v \geq 0, v \in R^n$ and $u_i = \max\{0, x_i\}$, $v_i = \max\{0, -x_i\}$. Consequently, (3) can be formulated by the following bound-constrained quadratic programming

$$\min_{u,v \in R^n} \frac{1}{2}\|y - \Phi(u - v)\|_2^2 + \mu(\mathbf{I}_n^T u + \mathbf{I}_n^T v) \text{ s.t. } u \geq 0,\ v \geq 0, \tag{4}$$

where $\mathbf{I}_n^T$ represents the transpose of $\mathbf{I}_n$, and $\mathbf{I}_n = [1, 1, \ldots, 1]^T$ is a vector consisting of n ones. Particularly, it follows from [5] that (4) can be rewritten as the following form

$$\min_{p \in R^{2n}} \frac{1}{2} p^T \Gamma p + q^T p \text{ s.t. } p \geq 0, \tag{5}$$

where $p = [u\ v]^T$, $b = \Phi^T y$, $q = \mu \mathbf{I}_{2n} + [-b\ b]^T$ and $\Gamma = \begin{bmatrix} \Phi^T\Phi & -\Phi^T\Phi \\ -\Phi^T\Phi & \Phi^T\Phi \end{bmatrix}$.

Recently, Xiao et al. [8] pointed out that (5) can be transformed into the following form

$$F(p) = \min\{p, \Gamma p + q\} = 0, \tag{6}$$

where function F is vector value, 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 [6].

Lemma 1. *There exists a positive constant L such that*

$$\|F(x) - F(y)\| \leq L\|x - y\|, \ \forall \ x, y \in R^{2n}. \tag{7}$$

The following lemma shows that $F(\cdot)$ is monotone [8].

Lemma 2. *The mapping $F(\cdot)$ is monotone, i.e.,*

$$(F(x) - F(y))^T(x - y) \geq 0, \ \forall \ x, y \in R^{2n}. \tag{8}$$

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 [4,9,10].

In this paper, we propose a projected conjugate gradient method for the minimization of l_1 regularized minimization problem with application to CS. Particularly, the search direction is generated by the following way

$$d_k = \begin{cases} -F(p_1) & \text{if} \quad k = 1, \\ -F(p_k) + \alpha_k d_{k-1} - \beta_k y_{k-1} & \text{if} \quad k \geq 2, \end{cases} \tag{9}$$

where $\alpha_k = \frac{F(x_k)^T y_{k-1}}{\|F_{k-1}\|^2}$, $\beta_k = \frac{F(x_k)^T d_{k-1}}{\|F_{k-1}\|^2}$ and $y_{k-1} = F(x_k) - F(x_{k-1})$.

The full description of our method, PCG Algorithm (short for " projected conjugate gradient algorithm"), can be formally presented as follows now.

Algorithm 1. (*PCG Algorithm*)

Date: Give initial point $p_1 \in R^{2n}$, set parameters $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho \in (0, 1)$.
Convergence test: If $\|F(p_1)\| = 0$, then stop. Else set $d_1 = -F(p_1)$. Let $k := 1$.
Line search update: Determine the steplength λ_k and set $z_k = p_k + \lambda_k d_k$, where $\lambda_k = \sigma_1 \rho^{m_k}$ with m_k being the smallest nonnegative integer m satisfying

$$-F(z_k)^T d_k \geq \sigma_2 \sigma_1 \rho^m \|F(z_k)\| \|d_k\|^2. \tag{10}$$

Projection update: Compute

$$p_{k+1} = p_k - \frac{F(z_k)^T(p_k - z_k)}{\|F(z_k)\|^2} F(z_k). \tag{11}$$

If $\|F(p_{k+1})\| = 0$, then stop. Else let $k := k + 1$ and compute d_k defined by (9). Then go to the Convergence Test.

The following lemma states that PCG Algorithm is well-defined, which can be proved in a way similar to the proof of Lemma 1 in [10].

Lemma 3. *Suppose that $F(p_k) \neq 0$ for all k, then there exists a nonnegative integer m_k satisfying (10) for all k.*

3 Global Convergence of PCG Algorithm

We prepare to show our main global convergence result of PCG Algorithm. Throughout this section, we assume that the solution set of (6) is nonempty.

3.1 Some Properties

In this subsection, we derive some useful properties of PCG Algorithm.

Lemma 4. *Suppose that the sequence $\{p_k\}$ is generated by PCG Algorithm, then for any $\hat{p}$ such that $F(\hat{p}) = 0$, it holds that*

$$\lim_{k\to\infty} \lambda_k \|d_k\| = 0. \tag{12}$$

Proof. By the line search process (10), we have

$$\begin{aligned} F(z_k)^T(p_k - z_k) &= -\lambda_k F(z_k)^T d_k \geq \sigma_2 \lambda_k^2 \|F(z_k)\| \|d_k\|^2 \\ &= \sigma_2 \|F(z_k)\| \|p_k - z_k\|^2 > 0. \end{aligned} \tag{13}$$

By (11) and the monotonicity of F, it is easy to deduce that

$$\begin{aligned} \|p_{k+1} - \hat{p}\|^2 &= \|p_k - \tfrac{F(z_k)^T(p_k - z_k)}{\|F(z_k)\|^2} F(z_k) - \hat{p}\|^2 \\ &= \|p_k - \hat{p}\|^2 - 2F(z_k)^T(p_k - \hat{p}) \tfrac{F(z_k)^T(p_k - z_k)}{\|F(z_k)\|^2} + \tfrac{[F(z_k)^T(p_k - z_k)]^2}{\|F(z_k)\|^2} \\ &\leq \|p_k - \hat{p}\|^2 - 2F(z_k)^T(p_k - z_k) \tfrac{F(z_k)^T(p_k - z_k)}{\|F(z_k)\|^2} + \tfrac{[F(z_k)^T(p_k - z_k)]^2}{\|F(z_k)\|^2} \\ &= \|p_k - \hat{p}\|^2 - \tfrac{[F(z_k)^T(p_k - z_k)]^2}{\|F(z_k)\|^2} \\ &\leq \|p_k - \hat{p}\|^2 - \sigma_2^2 \|p_k - z_k\|^4. \end{aligned} \tag{14}$$

Hence the sequence $\{\|p_k - \hat{p}\|\}$ is decreasing and convergent. Furthermore, the sequence $\{\|p_k\|\}$ is bounded. By the Cauchy-Schwarz inequality and the monotonicity of F, we have

$$\|F(p_k)\| \geq \frac{F(p_k)^T(p_k - z_k)}{\|p_k - z_k\|} \geq \frac{F(z_k)^T(p_k - z_k)}{\|p_k - z_k\|} \geq \sigma_2 \|F(z_k)\| \|p_k - z_k\|. \tag{15}$$

Moreover, we obtain that the sequence $\{z_k\}$ is bounded too. It follows that

$$\sum_{k=1}^{\infty} \|p_k - z_k\|^4 \leq \frac{1}{\sigma_2^2} \sum_{k=1}^{\infty} (\|p_k - \hat{p}\|^2 - \|p_{k+1} - \hat{p}\|^2) < \infty, \tag{16}$$

which implies

$$\lim_{k\to\infty} \|p_k - z_k\| = 0, \text{ namely, } \lim_{k\to\infty} \lambda_k \|d_k\| = 0. \qquad \blacksquare \tag{17}$$

The following lemma can be proved in a way similar to the proof of Lemma 2.1 in [9].

Lemma 5. *Suppose that the sequences* $\{p_k\}$ *and* $\{z_k\}$ *are generated by PCG Algorithm, then it holds that*

$$\lambda_k \geq \min\{\sigma_1, \frac{\rho\|F(p_k)\|^2}{(L+\sigma_2\|F(z'_k)\|)\|d_k\|^2}\}, \tag{18}$$

where $z'_k = p_k + \lambda'_k d_k$ *and* $\lambda'_k = \lambda_k \rho^{-1}$.

The following lemmas come from Lemma 2.4 in [9] and Lemma 3.1 in [11], respectively.

Lemma 6. *Suppose that sequence* $\{p_k\}$ *is generated by PCG Algorithm,* $\hat{p}$ *satisfies* $F(\hat{p}) = 0$, $z'_k = p_k + \lambda'_k d_k$ *and* $\lambda'_k = \lambda_k \rho^{-1}$, *then there exists a constant* $M_1 > 0$ *such that* $\|F(p_k)\| \leq M_1$ *and* $\|F(z'_k)\| \leq M_1$.

Lemma 7. *If there exists a constant* $\varepsilon > 0$ *such that* $\|F(p_k)\| \geq \varepsilon$ *for all* k, *then there exists a constant* $M_2 > 0$ *such that* $\|d_k\| \leq M_2$ *for all* k.

3.2 Convergence Result

In this subsection, we establish the global convergence of the PCG Algorithm proposed in the previous section.

Theorem 1. *Suppose that the sequence* $\{p_k\}$ *is generated by PCG Algorithm, then it holds that*

$$\lim_{k\to\infty} \inf \|F(p_k)\| = 0. \tag{19}$$

Proof. Suppose that $\lim_{k\to\infty} \inf \|F(p_k)\| \neq 0$, then there exists a constant $\varepsilon > 0$ such that $\|F(p_k)\| > \varepsilon$, for $k \geq 1$. Notice that d_k defined by (9) satisfies $F(p_k)^T d_k = -\|F(p_k)\|^2$ and $\|F(p_k)\| \leq \|d_k\|$, which implies

$$\|d_k\| \geq \varepsilon, \text{ for } k \geq 2. \tag{20}$$

For all k sufficiently large, by Lemma 5, Lemma 6, Lemma 7, $\|F(p_k)\| \geq \varepsilon$ and (20), we deduce that

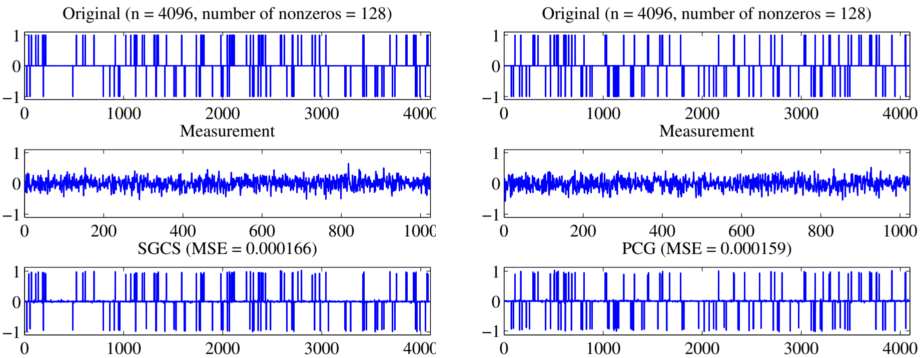
$$\begin{aligned} \lambda_k\|d_k\| &> \min\{\sigma_1, \tfrac{\rho\|F(p_k)\|^2}{(L+\sigma_2\|F(z'_k)\|)\|d_k\|^2}\}\|d_k\| \\ &= \min\{\sigma_1\|d_k\|, \tfrac{\rho\|F(p_k)\|^2}{(L+\sigma_2\|F(z'_k)\|)\|d_k\|}\} \\ &\geq \min\{\sigma_1\varepsilon, \tfrac{\rho\varepsilon^2}{(L+\sigma_2 M_1)M_2}\} \\ &> 0. \end{aligned} \tag{21}$$

Obviously, (21) contradicts with (12). Similarly, we can derive a contradiction when $k = 1$. Hence the proof is complete. ■

4 Experimental Results

In this section, numerical experiments are presented to show the performance of the PCG Algorithm for reconstructing sparse signals. These experiments are all tested in Matlab R2012a. 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.

In our experiments, we consider a typical compressive sensing scenario, the goal is to reconstruct a n length sparse signal from m observations. Random Φ is the Gaussian matrix whose elements are generated from shape *i.i.d.* normal distributions $\mathcal{N}(0, 1)$ (randn(m, n) in Matlab). For y, we add some noises such as $y = \Phi x + \eta$, where η is the Gaussian noise distributed as $\mathcal{N}(0, \sigma^2 I)$.



(a) SGCS (b) PCG

Fig. 1. (a) Top: original signal. Middle: noisy measurement with length 1024. Bottom: recovered signal by SGCS when $\sigma^2 = 10^{-2}$. (b) Top: original signal. Middle: noisy measurement with length 1024. Bottom: recovered signal by PCG when $\sigma^2 = 10^{-2}$.

It should be emphasized that we are mainly concerned with the speed of reconstructing the true signal $\bar{x}$ from the noisy measurement y in this paper. We restrict our attention to the penalized least squares model (3), and use $f(x) = \frac{1}{2}\|y - \Phi x\|_2^2 + \mu\|x\|_1$ as the merit function. Additionally, μ is forced to decrease as in [5] in order to avoid the solution of the quadratic penalty function (3) going to the BP problem while $\mu \to 0$. We compare the performance of PCG method with SGCS method [8]. According to [8], we let $\beta = 1.0$, $\rho = 0.1$, $\gamma = 1.2$ and $\xi = 10^{-4}$ in SGCS Algorithm. However, in PCG Algorithm, we let $\rho = 0.1$, $\sigma_1 = 0.95$ and $\sigma_2 = 0.93$. The common stopping criterion of both methods is

$$\frac{\|f(x_k) - f(x_{k-1})\|}{\|f(x_{k-1})\|} < 10^{-4}. \qquad (22)$$

We choose three different signals and four different values of σ^2 in our experiments. In order to test the speed of the algorithms more fairly, we list the average of the five results in the following tables, respectively. Numerical results are listed in Tables 1 2 3, 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 error to every original signal $\bar{x}$ (MSE) and the final objective function value (Obj). From Tables 1 2 3, we can see that the PCG method is faster than SGCS method, and the number of iteration of PCG method is less than that of the SGCS method. Moreover, we note that the MSE and Obj values attained by the PCG and SGCS method are very similar.

Table 1. SGCS v.s. PCG: performance of signal reconstruction. Original signal with length 1024 and 32 non-zero elements, noisy measurement with length 256.

σ^2	SGCS				PCG			
	Iter	Time	MSE	Obj	Iter	Time	MSE	Obj
10^{-4}	196	0.59	7.756e-06	7.365e-02	162	0.47	1.202e-05	6.725e-02
10^{-3}	245	0.62	1.671e-05	6.240e-02	166	0.49	1.304e-05	6.753e-02
10^{-2}	198	0.58	1.293e-04	7.738e-02	161	0.46	1.731e-04	6.829e-02
10^{-1}	255	0.66	1.357e-02	1.290e-01	193	0.61	1.666e-02	1.158e-01

Table 2. SGCS v.s. PCG: performance of signal reconstruction. Original signal with length 2048 and 64 non-zero elements, noisy measurement with length 512.

σ^2	SGCS				PCG			
	Iter	Time	MSE	Obj	Iter	Time	MSE	Obj
10^{-4}	213	2.53	8.252e-06	1.347e-01	160	1.97	1.181e-05	1.348e-01
10^{-3}	188	2.22	7.414e-05	1.333e-01	161	2.01	1.622e-05	1.391e-01
10^{-2}	201	2.43	1.327e-04	1.542e-01	169	2.15	1.647e-04	1.514e-01
10^{-1}	279	3.28	1.314e-02	2.524e-01	218	2.71	1.434e-02	2.415e-01

Table 3. SGCS v.s. PCG: performance of signal reconstruction. Original signal with length 4096 and 128 non-zero elements, noisy measurement with length 1024.

σ^2	SGCS				PCG			
	Iter	Time	MSE	Obj	Iter	Time	MSE	Obj
10^{-4}	191	7.95	7.061e-06	2.834e-01	161	6.82	1.169e-05	2.891e-01
10^{-3}	203	8.31	9.753e-05	2.685e-01	182	7.89	1.689e-05	2.961e-01
10^{-2}	191	8.03	1.459e-04	3.038e-01	175	7.63	1.838e-04	3.058e-01
10^{-1}	283	11.8	1.396e-02	6.235e-01	165	6.88	1.367e-02	4.952e-01

Fig. 1 shows simulation results of SGCS and PCG for a signal sparse reconstruction when $\sigma^2 = 10^{-2}$, respectively. As we can see from Figure 1 (b), all the original sparse signals are restored exactly by PCG method. These experiment results show that the PCG method can work well in an efficient manner.

5 Concluding Remarks

We have proposed a projected conjugate gradient method for solving a convex quadratic programming problem arising from compressed sensing. Our motivation for developing the method mainly comes from [8], where the authors point out that (5) can be transformed into an equivalent nonsmooth nonlinear system of monotone equations, namely, $F(p) = 0$. This system is monotone and Lipschitz continuous, and it can be solved efficiently with some derivative-free methods. In this paper, we adopt the recent conjugate gradient method of Zhang, Zhou and Li [11] with projection strategy of Solodov and Svaiter [7]. We name our method PCG (the abbreviation of " projected conjugate gradient") and establish its global convergence under some suitable conditions. Numerical results show that the PCG method can significantly improve the CPU time for solving the nonlinear system of monotone equations in sparse signals reconstruction while obtaining similar reconstructive quality.

Acknowledgments. This work was supported by the NSFC grants (No. 11001060 and No. 61262026), the NSF of Jiangxi Province (No. 2009GQS0007), the programm of JGZX (No. 20112BCB23027) and the Postgraduate Innovation Fund of Gannan Normal University (No. YCX10B006).

References

1. Chen, S., Donoho, D.L., Saunders, M.: Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing 20, 33–61 (1998)
2. Daubechies, I., Fornasier, M., Loris, I.: Accelerated projection gradient method for linear inverse problems with sparsity constraints. Journal of Fourier Analysis and Applications 14, 764–792 (2008)
3. Duarte, M.F., Eldar, Y.C.: Structured compressed sensing: from theory to applications. IEEE Transactions on Signal Processing 59, 4053–4085 (2011)
4. Grippo, L., Sciandrone, M.: Nonmonotone derivative-free methods for nonlinear equations. Computational Optimization and Applications 37, 297–328 (2007)
5. Figueiredo, M.A.T., Nowak, R.D., Wright, S.J.: Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE Journal of Selected Topics in Signal Processing 1, 586–598 (2007)
6. Pang, J.S.: Inexact Newton methods for the nonlinear complementary problem. Mathematical Programming 36, 54–71 (1986)
7. Solodov, M.V., Svaiter, B.F.: A globally convergent inexact Newton method for systems of monotone equations. In: Fukushima, M., Qi, L. (eds.) Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smooth Methods, pp. 355–369. Kluwer Academic Publishers (1998)
8. Xiao, Y.H., Wang, Q.Y., Hu, Q.J.: Non-smooth equations based method for l_1-norm problems with applications to compressed sensing. Nonlinear Analysis: Theory, Methods & Applications 74, 3570–3577 (2011)
9. Yan, Q.R., Peng, X.Z., Li, D.H.: A globally convergent derivative-free method for solving large-scale nonlinear monotone equations. Journal of Computational and Applied Mathematics 234, 649–657 (2010)

10. Yu, G.H., Niu, S.Z.: Multivariate spectral gradient projection method for large-scale nonlinear systems of monotone equations, http://www.paper.edu.cn/index.php/default/en_releasepaper/downPaper/201201-778
11. Zhang, L., Zhou, W.J., Li, D.H.: A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence. IMA Journal of Numerical Analysis 26, 629–640 (2006)