

# An Efficient Projected Gradient Method for Convex Constrained Monotone Equations with Applications in Compressive Sensing

Yaping Hu, Yujie Wang

College of Science, Tianjin University of Science and Technology, Tianjin, China

Email: huyaping@tust.edu.cn

**How to cite this paper:** Hu, Y.P. and Wang, Y.J. (2020) An Efficient Projected Gradient Method for Convex Constrained Monotone Equations with Applications in Compressive Sensing. *Journal of Applied Mathematics and Physics*, 8, 983-998.  
<https://doi.org/10.4236/jamp.2020.86077>

**Received:** May 6, 2020

**Accepted:** May 28, 2020

**Published:** June 1, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In this paper, a modified Polak-Ribière-Polyak conjugate gradient projection method is proposed for solving large scale nonlinear convex constrained monotone equations based on the projection method of Solodov and Svaiter. The obtained method has low-complexity property and converges globally. Furthermore, this method has also been extended to solve the sparse signal reconstruction in compressive sensing. Numerical experiments illustrate the efficiency of the given method and show that such non-monotone method is suitable for some large scale problems.

## Keywords

Projection Method, Monotone Equations, Conjugate Gradient Method, Compressive Sensing

## 1. Introduction

This paper is dedicated to solving the following nonlinear convex constrained monotone equations:

$$F(x) = 0, x \in \Omega, \quad (1)$$

where  $F: R^n \rightarrow R^n$  is a continuous nonlinear mapping and the feasible region  $\Omega \subset R^n$  is a nonempty closed convex set, e.g. an  $n$ -dimensional box, namely,  $\Omega = x \in R^n : l \leq x \leq u$ . Monotone means that

$$\langle F(x) - F(y), x - y \rangle \geq 0, \forall x, y \in R^n, \quad (2)$$

where the  $\langle \cdot, \cdot \rangle$  denotes the inner product of vectors. The problems (1) emerges in many fields such as economic equilibrium problems [1], chemical equilibrium

systems [2] and the power flow equations [3]. Based on the work of Solodov and Svaiter [4], Wang *et al.* [5] proposed a projection type method to solve Equation (1). The obtained method in [5] possesses global convergence property without any regularity assumptions. Nevertheless the method needs to solve a linear equation at each iteration. To avoid solving the linear equation and improving the effectiveness, some projected conjugate gradient methods [6] [7] [8] [9] are studied based on the projection technique of Solodov and Svaiter [4]. The numerical results gained in [6] [7] [8] [9] indicate that the projected conjugate gradient type methods for solving problem (1) are indeed efficient and promising. In this paper, by combining the well-known Polak-Ribière-Polyak [10] [11] method with the projection technique of Solodov and Svaiter [4], a conjugate gradient projected method with fast convergent property is proposed for the nonlinear monotone equations with convex constraints. Under some mild conditions, the global convergent results are established for the given method. The obtained method possesses the following three beneficial properties: 1) The search direction satisfies the sufficient descent condition, 2) The global convergence is independent of any merit function, and 3) It is derivative-free method and is effective for large scale nonlinear convex constrained monotone equations (with a maximum dimension of 100,000). Furthermore, the obtained method is extended to solve the  $l_1$ -norm problem by reformulating it as non-smooth monotone equations.

In Section 2, the modified PRP-type conjugate gradient projected method is proposed, and some preliminary properties are studied. The global convergence results are established in Section 3. The numerical experiments, and the applications of the obtained method for  $l_1$ -norm regularized compressive sensing problems are discussed in Section 4. Finally, we have a conclusion section.

## 2. The Proposed Method and Corresponding Algorithm

We firstly introduce the definition of the projection operator  $P_\Omega[\cdot]$  which is defined as the mapping from  $R^n$  to  $\Omega$ ,

$$P_\Omega[x] = \arg \min \{\|y - x\| \mid y \in \Omega\}, \quad \forall x \in R^n,$$

where  $\|\cdot\|$  denotes the Euclidean norm of vectors,  $\Omega$  is a nonempty closed convex subset of  $R^n$ .

The projection operator is non-expansive, namely, for any  $x, y \in R^n$ , the following condition holds

$$\|P_\Omega[y] - P_\Omega[x]\| \leq \|x - y\|. \quad (3)$$

Let's review the Polak-Ribière-Polyak [10] [11] conjugate gradient method briefly. The PRP method is firstly designed for solving the unconstrained optimization problem:

$$\min \{f(x) \mid x \in R^n\}, \quad (4)$$

where  $f : R^n \rightarrow R$  is continuously differentiable. It generates the iteration sequence  $\{x_k\}$  in the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{5}$$

where  $x_k$  is the current iteration point,  $\alpha_k > 0$  is a step-length, and  $d_k$  is the search direction given by

$$d_k = \begin{cases} -g_k + \beta_{k-1}^{PRP} d_{k-1}, & \text{if } k > 0, \\ -g_k, & \text{if } k = 0, \end{cases} \tag{6}$$

where  $\beta_{k-1}^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}$ ,  $y_{k-1} = g_k - g_{k-1}$ .

Combining the projected technique of Solodov and Svaiter [4] with the PRP method formed by Equation (5) and Equation (6), the following modified PRP formula is defined given in this paper

$$d_k = \begin{cases} -g_k + \frac{g_k^T y_{k-1} d_{k-1} - d_{k-1}^T g_k y_{k-1}}{\max\{2\gamma \|d_{k-1}\| \|y_{k-1}\|, d_{k-1}^T y_{k-1}, \|g_{k-1}\|^2\}}, & \text{if } k > 0 \\ -g_k, & \text{if } k = 0, \end{cases} \tag{7}$$

where  $y_{k-1} = g_k - g_{k-1}$  and  $\gamma > 0$  is a constant.

It is show be noted that the proposed direction formula Equation (7) reduces to PRP formula if the exact line search is used. Furthermore, the sufficient descent condition automatically holds for all  $k$ , since  $d_k^T g(x_k) = -\|g(x_k)\|^2$ . There are some conjugate gradient methods with similar idea concerning Equation (7) have been studied in the papers [12]-[19].

The corresponding modified PRP conjugate gradient projection algorithm for solving problem (1) starts as follows.

**Algorithm 1:**

**Step 0** Choose any initial point  $x_0 \in \Omega$ , and select constants  $\rho \in (0,1)$ ,  $\gamma > 0$ ,  $\sigma > 0$ ,  $\xi > 0$ ,  $\epsilon \in (0,1)$  and  $d_0 = -F(x_0)$ . Let  $k := 0$ .

**Step 1** If  $\|F(x_k)\| \leq \epsilon$ , stop. Otherwise compute search direction  $d_k$  by Equation (7) with  $g_k$  and  $g_{k-1}$  replaced by  $F_k$  and  $F_{k-1}$ , respectively.

**Step 2** Let  $z_k = x_k + \alpha_k d_k$ , where  $\alpha_k = \max\{\xi \rho^i \mid i = 0, 1, \dots\}$  such that

$$-\langle F(x_k + \alpha_k d_k), d_k \rangle \geq \sigma \alpha_k \|d_k\|^2. \tag{8}$$

**Step 3** If  $\|F(z_k)\| \leq \epsilon$ , stop and let  $x_{k+1} = z_k$ . Otherwise compute the next iteration by

$$x_{k+1} = P_\Omega [x_k - \beta_k F(z_k)], \tag{9}$$

where

$$\beta_k = \frac{\langle F(z_k), x_k - z_k \rangle}{\|F(z_k)\|^2} \tag{10}$$

**Step 4** Let  $k := k + 1$ , and go to Step 1.

**Remark 1:** In the algorithm 1, the step size  $\alpha_k$  given by Equation (8) satisfies

$$\langle F(z_k), x_k - z_k \rangle > 0,$$

where  $z_k = x_k + \alpha_k d_k$ ,  $d_k$  is the search direction. Moreover, for any  $x^*$  such that  $F(x^*) = 0$ ,

$$\langle F(z_k), x^* - z_k \rangle \leq 0.$$

comes from the monotonicity property of  $F(x)$ . This means that the hyperplane

$$H_k = \{x \in R^n \mid \langle F(z_k), x - z_k \rangle = 0\}$$

strictly separates the current point  $x_k$  from the solution set of the problem. The above facts and Step 3 indicate that the next iteration  $x_{k+1}$  is computed by projecting  $x_k$  onto the intersection of the feasible set  $\Omega$  with the halfspace  $H_k$ .

### 3. Convergence Analysis

In this section, we are going to discuss the convergence property of the given method. Before that, there are some basic assumptions on problem (1) needs to be given.

**Assumption 1:** The mapping  $F$  is Lipschitz continuous with constant  $L > 0$  in a set  $\Omega$ , written  $F \in \text{Lip}(\Omega)$ , for every  $x, y \in \Omega$ ,

$$\|F(x) - F(y)\| \leq L\|x - y\|. \tag{11}$$

**Assumption 2:** The solution set of the problem (1), denoted by  $S$ , is nonempty convex.

For conjugate gradient method, the sufficient descent property is essential in the convergence analysis, the following lemma shows that the search direction  $\{d_k\}$  generated by Algorithm 1 satisfies the sufficient descent condition independent of line search.

**Lemma 1:** Let the sequence  $\{x_k\}$  and  $\{d_k\}$  be generated by Algorithm 1. Then, for all  $k \geq 0$ ,

$$F(x_k)^T d_k = -\|F(x_k)\|^2, \tag{12}$$

and

$$\|d_k\| \leq \left(1 + \frac{1}{\gamma}\right) \|F(x_k)\|. \tag{13}$$

**Proof:** For  $k = 0$ , Equation (12) and Equation (13) follows from the direct application of  $d_0 = -g(x_0)$ . For  $k \geq 1$ , using Equation (7), the definition of the search direction  $d_{k+1}$ , it follows that

$$d_{k+1}^T F_{k+1} = -\|F_{k+1}\|^2 + \left[ \frac{F_{k+1}^T y_k d_k - d_k^T F_{k+1} y_k}{\max\{2\gamma \|d_k\| \|y_k\|, d_k^T y_k, \|F_k\|^2\}} \right]^T F_{k+1} = -\|F_{k+1}\|^2,$$

similarly,

$$\begin{aligned} \|d_{k+1}\| &= \left\| -F_{k+1} + \frac{F_{k+1}^T y_k d_k - d_k^T F_{k+1} y_k}{\max\{2\gamma \|d_k\| \|y_k\|, d_k^T y_k, \|F_k\|^2\}} \right\| \\ &\leq \|F_{k+1}\| + \frac{\|F_{k+1}\| \|y_k\| \|d_k\| + \|d_k\| \|F_{k+1}\| \|y_k\|}{\max\{2\gamma \|d_k\| \|y_k\|, d_k^T y_k, \|F_k\|^2\}} \\ &\leq \left(1 + \frac{1}{\gamma}\right) \|F_{k+1}\|, \end{aligned}$$

where the last inequality follows from the fact

$$\max\{2\gamma \|d_k\| \|y_k\|, \|F_k\|^2\} \geq 2\gamma \|d_k\| \|y_k\|.$$

In the remaining part of this paper, we assume that  $F_k \neq 0$  for all  $\forall k \geq 0$ , otherwise, the solution of the problem (1) has been found.

**Lemma2:** Let the sequence  $\{x_k\}$  and  $\{z_k\}$  be generated by Algorithm 1. Suppose that the Assumption 1 holds. Then there exists a positive number  $\alpha_k$  satisfying Equation (8) for all  $k \geq 0$ .

**Proof:** The line search ensure that if  $\alpha_k \neq \xi$ , then  $\alpha'_k = \rho^{-1} \alpha_k$  does not satisfy Equation (8), namely,

$$-\langle F(z'_k), d_k \rangle < \sigma \alpha'_k \|d_k\|^2,$$

where  $z'_k = x_k + \alpha'_k d_k$ . From Equation (12) and Assumption 1 we have

$$\begin{aligned} \|F_k\|^2 &= -\langle F_k, d_k \rangle = \langle F(z'_k) - F(x_k), d_k \rangle - \langle F(z'_k), d_k \rangle \\ &\leq L \alpha'_k \|d_k\|^2 + \sigma \alpha'_k \|d_k\|^2 \leq \rho^{-1} \alpha_k (L + \sigma) \|d_k\|^2 \end{aligned}$$

which means that

$$\alpha_k \geq \min \left\{ \xi, \frac{\rho}{L + \sigma} \frac{\|F_k\|^2}{\|d_k\|^2} \right\}. \tag{14}$$

The above result Equation (14) shows that the line search procedure Equation (8) always terminates in a finite number of steps.

**Lemma3:** Let sequences  $\{x_k\}$  and  $\{z_k\}$  be generated by Algorithm 1. Suppose that Assumptions 1 and 2 hold. Then both  $\{x_k\}$  and  $\{z_k\}$  are bounded. Moreover, we have

$$\lim_{k \rightarrow \infty} \|x_k - z_k\| = 0, \tag{15}$$

and

$$\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0. \tag{16}$$

Particularly, Equation (15) implies that

$$\lim_{k \rightarrow \infty} \alpha_k \|d_k\| = 0. \tag{17}$$

**Proof:**  $x^* \in S$  denotes any arbitrary solution of the problem (1). The monotonicity of  $F$  and the line search Equation (8) deduce

$$\langle F(z_k), x_k - x^* \rangle \geq \langle F(z_k), x_k - z_k \rangle \geq \sigma \alpha_k^2 \|d_k\|^2 \geq 0. \tag{18}$$

Equation (3), Equation (9) and Equation (18) imply

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|P_\Omega [x_k - \beta_k F(z_k)] - x^*\|^2 \leq \|x_k - \beta_k F(z_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - 2\beta_k \langle F(z_k), x_k - x^* \rangle + \beta_k^2 \|F(z_k)\|^2 \\ &\leq \|x_k - x^*\|^2 - 2\beta_k \langle F(z_k), x_k - z_k \rangle + \beta_k^2 \|F(z_k)\|^2 \\ &\leq \|x_k - x^*\|^2 - \frac{\langle F(z_k), x_k - z_k \rangle^2}{\|F(z_k)\|^2} \\ &\leq \|x_k - x^*\|^2 - \frac{\sigma^2 \|x_k - z_k\|^4}{\|F(z_k)\|^2}. \end{aligned} \tag{19}$$

Since the sequence  $\{\|x_k - x^*\|\}$  is decreasing and convergent, the sequence  $\{x_k\}$  is bounded. Equation (19) shows that  $\|x_k - x^*\| \leq \|x_0 - x^*\|$  for all  $k$ . Then, by Assumption 1, we have

$$\|F(x_k)\| = \|F(x_k) - F(x^*)\| \leq L \|x_k - x^*\| \leq L \|x_0 - x^*\|. \tag{20}$$

Let  $M_1 = L \|x_0 - x^*\|$ ,

$$\|F(x_k)\| \leq M_1, \forall k \geq 0. \tag{21}$$

From the Cauchy-Schwarz inequality, the line search Equation (8), the monotonicity of  $F$  and Equation (18), it follows that

$$\begin{aligned} 0 < \sigma \|x_k - z_k\|^2 &\leq \langle F(z_k), x_k - z_k \rangle \leq \langle F(x_k), x_k - z_k \rangle \leq \|F(x_k)\| \|x_k - z_k\|. \\ \sigma \|x_k - z_k\| &\leq \|F(x_k)\| \leq M_1, \end{aligned} \tag{22}$$

which shows that the sequence  $\{z_k\}$  is bounded. Furthermore, the sequence  $\{\|z_k - x^*\|\}$  is also bounded, there exists  $M_2 > 0, k_0 \geq 0$ , such that

$$\|z_k - x^*\| \leq M_2, \forall k \geq k_0. \tag{23}$$

Based on Equation (23) and Assumption 1 it follows

$$\|F(z_k)\| = \|F(z_k) - F(x^*)\| \leq L \|z_k - x^*\| \leq LM_2. \tag{24}$$

Substituting the above relationship into Equation (19), it deduces

$$\frac{\sigma^2}{(LM_2)^2} \sum_{k=0}^\infty \|x_k - z_k\|^4 \leq \sum_{k=0}^\infty (\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2) < \infty, \tag{25}$$

which implies

$$\lim_{k \rightarrow \infty} \|x_k - z_k\| = 0.$$

From the definition of  $z_k$  and Equation (15), it holds that

$$\lim_{k \rightarrow \infty} \alpha_k \|d_k\| = 0.$$

Combining the definition of  $\beta_k$ , Equation (3), and the Cauchy-Schwarz in-

equality, we have

$$\begin{aligned}\|x_{k+1} - x_k\| &= \|P_\Omega [x_k - \beta_k F(z_k)] - x_k\| \\ &\leq \|x_k - \beta_k F(z_k) - x_k\| \\ &= \frac{\langle F(z_k), x_k - z_k \rangle}{\|F(z_k)\|} \\ &\leq \|x_k - z_k\|\end{aligned}$$

which together with Equation (15), proves Equation (16).

**Theorem1:** Let sequences  $\{x_k\}$  and  $\{z_k\}$  be generated by Algorithm 1. Suppose that Assumptions 1 and 2 hold. Then

$$\liminf_{k \rightarrow \infty} \|F_k\| = 0. \quad (26)$$

**Proof:** We prove this Theorem by contradiction. Assume that Equation (26) does not hold, namely, there exists  $\varepsilon > 0$  such that

$$\|F_k\| \geq \varepsilon, \quad \forall k \geq 0. \quad (27)$$

From Equation (12) and Equation (27),

$$\begin{aligned}\|d_k\|^2 &= \|d_k + F_k - F_k\|^2 \\ &= \|d_k + F_k\|^2 - 2\langle d_k + F_k, F_k \rangle + \|F_k\|^2 \\ &\geq -2\langle d_k, F_k \rangle - \|F_k\|^2 \\ &= \|F_k\|^2,\end{aligned}$$

which implies

$$\|d_k\| \geq \varepsilon, \quad \forall k \geq 0. \quad (28)$$

On the other hand, Equation (13), Equation (21) and the definition of  $d_k$  deduce

$$\|d_k\| \leq \left(1 + \frac{1}{\gamma}\right) \|F_k\| \leq \left(1 + \frac{1}{\gamma}\right) M_1, \quad \forall k \geq 0.$$

Finally, from Equation (14), Equation (27) and Equation (28),

$$\begin{aligned}\alpha_k \|d_k\| &\geq \min \left\{ \xi, \frac{\rho}{L + \sigma} \frac{\|F_k\|^2}{\|d_k\|^2} \right\} \|d_k\| \\ &\geq \min \left\{ \xi \varepsilon, \frac{\rho \varepsilon^2}{(L + \sigma)(1 + \gamma^{-1}) M_1} \right\}\end{aligned}$$

which contradicts with Equation (17). Thus, Equation (26) holds.

## 4. Numerical Experiments

The numerical performances of the proposed Algorithm 1 for large scale non-linear convex constrained monotone equations with various dimensions and different initial points are studied in this section. Furthermore, the given Algorithm 1 is extended to solve the  $l_1$ -norm regularized problems which de-

code a sparse signal in compressive sensing. The algorithm is coded in MATLAB R2015a and run on a PC with Core i5 CPU and 4 GB memory.

### 4.1. Experiments on Nonlinear Convex Constrained Monotone Equations

The testing problems are listed as follows.

**Problem 1.** (Wang *et al.* [5]) The elements of  $F(x)$  are given by

$$F_i(x) = e^{-x_i} - 1, \quad i = 1, 2, 3, \dots, n.$$

and  $\Omega = R_+^n$ .

**Problem 2.** The example is taken from [7]. The elements of  $F(x)$  are given by

$$F_i(x) = 2x_i - \sin(x_i), \quad i = 1, 2, 3, \dots, n.$$

and  $\Omega = R_+^n$ .

**Problem 3.** The example is taken from [9].

$$\begin{aligned} g_1(x) &= x_1 - e^{\cos\left(\frac{x_1+x_2}{n+1}\right)}, \\ g_i(x) &= x_i - e^{\cos\left(\frac{x_{i-1}+x_i+x_{i+1}}{n+1}\right)}, \quad i = 2, 3, \dots, n-1, \\ g_n(x) &= x_n - e^{\cos\left(\frac{x_{n-1}+x_n}{n+1}\right)}. \end{aligned}$$

and  $\Omega = R_+^n$ .

**Problem 4.** The example is taken from [20].

$$F_i(x) = x_i - \sin(|x_i - 1|), \quad i = 1, 2, 3, \dots, n.$$

and  $\Omega = \left\{x \in R^n \mid \sum_{i=1}^n x_i \leq n, x_i \geq -1, i = 1, 2, \dots, n\right\}$ .

For convenience, MPRP denotes the proposed Algorithm 1. We compare the MPRP method with CGD method [8] on problems 1-4. For both methods, set  $\xi = 1$ ,  $\rho = 0.4$ ,  $\sigma = 10^{-4}$ . In order to evaluate the efficiency and the robustness of both methods, we test the Problems 1-4 with various dimensions

$$n = 10000, 50000, 100000 \quad \text{and different initial points: } x_1 = \left(1, 0.5, \dots, \frac{1}{n}\right)^T,$$

$x_2 = \frac{1}{n} \text{ones}(n, 1)$ ,  $x_3 = \text{ones}(n, 1)$ ,  $x_4 = 2\text{ones}(n, 1)$ ,  $x_5 = \text{rand}(n, 1)$ , where  $\text{ones}(n, 1)$  returns a  $n$ -by-1 array of ones and  $\text{rand}(n, 1)$  returns a  $n$ -by-1 array of rand values in MATLAB.

Numerical results are shown in **Tables 1-4**, in which Init (Dim), NI and NF denote initial points (dimension), the number of iterations and the number of function evaluations respectively.  $\|F(x)\|$  is the final Euclidean norm of the function values, and CPU-time in seconds.

**Tables 1-4** indicate that the dimension of the problem has little effect on the number of iterations of the algorithm. However, the computing time is relatively large in high dimension cases. Moreover, we can see from the results of **Tables**

**1-4** that Algorithm 1 is more competitive than CGD algorithm as Algorithm 1 can get the solution of all the test data at a smaller number of iterations and smaller CPU time. So the results of **Tables 1-4** show that our method is very efficient.

The numerical performances of the both methods are also evaluated by using the performance profile tool of Dolan and Moré [21]. **Figure 1** shows the performance of two methods, it is obviously that the proposed MPRP method is more efficient and robust than CGD method.

**Table 1.** Numerical results for MPRP/CGD on problem 1.

Init (Dim)	MPRP		CGD	
	NI/ NF/ $\ f(x)\ $	Time	NI/ NF/ $\ f(x)\ $	Time
$x_1(10000)$	11/25/3.84830e-006	0.13	16/185/5.84301e-006	0.50
$x_2(10000)$	5/11/9.76367e-006	0.08	5/11/9.59824e-006	0.09
$x_3(10000)$	11/24/2.98167e-006	0.15	10/21/4.63815e-006	0.13
$x_4(10000)$	11/27/6.30895e-006	0.14	12/28/5.36760e-006	0.16
$x_5(10000)$	14/31/8.16510e-006	0.15	30/61/9.55151e-006	0.28
$x_1(50000)$	11/25/3.84897e-006	0.42	19/261/5.42831e-006	2.92
$x_2(50000)$	5/11/4.36715e-006	0.21	5/11/4.29309e-006	0.25
$x_3(50000)$	11/24/6.66722e-006	0.45	11/23/2.07424e-006	0.49
$x_4(50000)$	12/29/3.52681e-006	0.56	13/30/2.40047e-006	0.62
$x_5(50000)$	15/33/5.67816e-006	0.63	32/65/7.69276e-006	1.26
$x_1(100000)$	11/25/3.84906e-006	0.78	16/192/7.19276e-006	3.26
$x_2(100000)$	5/11/3.08810e-006	0.39	5/11/3.03573e-006	0.47
$x_3(100000)$	11/24/9.42888e-006	0.86	11/23/2.93342e-006	0.94
$x_4(100000)$	12/29/4.98767e-006	1.03	13/30/3.39477e-006	1.21
$x_5(100000)$	15/33/8.06362e-006	1.22	33/67/6.52318e-006	2.61

**Table 2.** Numerical results for MPRP/CGD on problem 2.

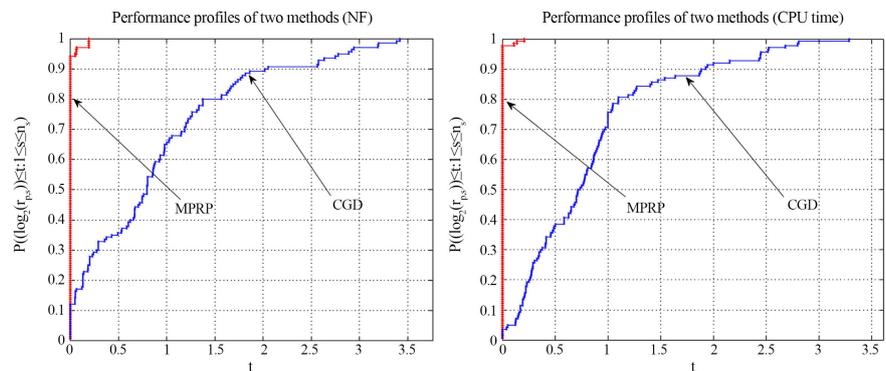
Init (Dim)	MPRP		CGD	
	NI/ NF/ $\ f(x)\ $	Time	NI/ NF/ $\ f(x)\ $	Time
$x_1(10000)$	10/21/5.34065e-006	0.09	17/141/7.20891e-006	0.25
$x_2(10000)$	5/11/3.20000e-006	0.07	5/11/9.60000e-006	0.07
$x_3(10000)$	10/21/3.73273e-006	0.09	11/23/4.43164e-006	0.11
$x_4(10000)$	9/20/5.61741e-006	0.11	11/23/4.83649e-006	0.10
$x_5(10000)$	13/27/3.34488e-006	0.11	31/63/9.40955e-006	0.21
$x_1(50000)$	10/21/5.34094e-006	0.21	16/125/9.89342e-006	0.61
$x_2(50000)$	4/9/7.15542e-006	0.12	5/11/4.29325e-006	0.14
$x_3(50000)$	10/21/8.34663e-006	0.21	11/23/9.90944e-006	0.24
$x_4(50000)$	10/22/2.51218e-006	0.22	12/25/2.16294e-006	0.28
$x_5(50000)$	13/27/7.50144e-006	0.27	19/67/3.81389e-006	0.47
$x_1(100000)$	10/21/5.34097e-006	0.12	16/125/8.77333e-006	1.17
$x_2(100000)$	4/9/5.05964e-006	0.06	5/11/3.03579e-006	0.22
$x_3(100000)$	11/23/2.36078e-006	0.14	12/25/2.80281e-006	0.51
$x_4(100000)$	10/22/3.55276e-006	0.13	12/25/3.05886e-006	0.48
$x_5(100000)$	14/29/2.96451e-006	0.18	19/95/7.73838e-006	1.03

**Table 3.** Numerical results for MPRP/CGD on problem 3.

Init (Dim)	MPRP		CGD	
	NI/ NF/ $\ f(x)\ $	Time	NI/ NF/ $\ f(x)\ $	Time
$x_1(10000)$	13/27/4.52180e-006	0.24	13/62/5.75833e-006	0.40
$x_2(10000)$	13/27/4.52370e-006	0.25	13/70/3.46424e-006	0.43
$x_3(10000)$	13/27/2.86185e-006	0.23	20/41/4.30230e-006	0.34
$x_4(10000)$	12/25/4.29188e-006	0.23	12/46/6.14493e-007	0.32
$x_5(10000)$	13/27/4.47547e-006	0.24	14/62/7.25107e-006	0.42
$x_1(50000)$	13/27/7.25565e-006	1.00	13/54/8.81290e-006	1.61
$x_2(50000)$	13/27/7.25860e-006	0.98	12/39/9.12282e-006	1.23
$x_3(50000)$	13/27/4.59974e-006	0.99	13/43/2.72069e-006	1.36
$x_4(50000)$	12/25/6.91496e-006	0.93	13/49/6.63736e-006	1.52
$x_5(50000)$	13/27/8.32203e-006	1.00	13/64/4.25399e-006	1.82
$x_1(100000)$	13/27/9.92424e-006	1.91	13/38/3.45225e-006	2.42
$x_2(100000)$	14/29/2.85975e-006	2.06	13/34/1.70785e-006	2.30
$x_3(100000)$	13/27/6.47855e-006	1.96	13/41/8.71814e-006	2.60
$x_4(100000)$	12/25/9.70603e-006	1.82	13/51/2.66088e-006	2.99
$x_5(100000)$	14/29/2.82988e-006	2.11	14/46/6.28570e-006	2.80

**Table 4.** Numerical results for MPRP/CGD on problem 4.

Init (Dim)	MPRP		CGD	
	NI/ NF/ $\ f(x)\ $	Time	NI/ NF/ $\ f(x)\ $	Time
$x_1(10000)$	16/49/3.81881e-006	0.17	17/83/4.26081e-006	0.21
$x_2(10000)$	11/34/4.33069e-006	0.11	18/54/5.36860e-006	0.17
$x_3(10000)$	11/34/2.49346e-006	0.11	19/56/7.93477e-006	0.17
$x_4(10000)$	11/32/8.70634e-006	0.10	20/58/9.36855e-006	0.17
$x_5(10000)$	13/40/3.26840e-006	0.12	31/91/6.47378e-006	0.23
$x_1(50000)$	16/49/9.28932e-006	0.37	18/66/4.06062e-006	0.52
$x_2(50000)$	11/34/9.68593e-006	0.26	19/57/4.81014e-006	0.49
$x_3(50000)$	11/34/5.57554e-006	0.25	20/59/7.11258e-006	0.50
$x_4(50000)$	12/35/4.06815e-006	0.28	21/61/8.39779e-006	0.51
$x_5(50000)$	13/40/7.33466e-006	0.36	43/127/9.90609e-006	0.95
$x_1(100000)$	17/52/4.28438e-006	0.25	19/59/1.32958e-006	0.95
$x_2(100000)$	12/37/2.86250e-006	0.18	19/57/6.80218e-006	0.98
$x_3(100000)$	11/34/7.88501e-006	0.18	21/62/4.03228e-006	0.98
$x_4(100000)$	12/35/5.75324e-006	0.17	22/64/4.76089e-006	1.09
$x_5(100000)$	14/43/2.89653e-006	0.20	18/52/7.84045e-006	0.89



**Figure 1.** Performance profiles for two methods MPRP and CGD, where the left and the right figures are represented as the number of function evaluations and the CPU time, respectively.

## 4.2. Experiments on the $l_1$ -Norm Regularization Problem

The problem of the combination of  $l_2$  and  $l_1$  norms in the cost function often emerges for the signal reconstruction, *i.e.*:

$$\min \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1, \quad (28)$$

where  $\|\cdot\|_2$  is the Euclidean norm, and

$$\|x\|_1 = \sum_{j=1}^m |x_j|$$

is the  $l_1$  norm,  $A$  is a system matrix,  $y \in R^m$  is the observed data,  $x \in R^n$  is the signal to be reconstructed, and  $\lambda$  is a positive regularization parameter.

The optimization problems of the form Equation (28) appear in several signal reconstruction problems, such as sparse signal de-blurring [22], medical image reconstructions [23], compressed sensing [24], and super-resolution [25]. Iterative line search method or fixed point iteration schemes are commonly used to solve problem (28). By using the technique proposed by Figueiredo *et al.* [26], we can reformulate problem (28) as a convex quadratic program problem. Let  $x = u - v$ ,  $u \geq 0$ ,  $v \geq 0$ , where  $u, v \in R^n$ ,  $u_i = \max(0, x_i)$  for all  $i = 1, \dots, n$  and  $v_i = -\min(0, x_i)$  for all  $i = 1, \dots, n$ . The  $l_1$  norm can be formulated as  $\|x\|_1 = e_n^T u + e_n^T v$ , where  $e_n = (1, 1, \dots, n)^T$ . The problem (28) is expressed as the bound-constrained quadratic program:

$$\min_{u, v} \frac{1}{2} \|y - A(u - v)\|_2^2 + \lambda e_n^T u + \lambda e_n^T v, \text{ s.t. } u \geq 0, v \geq 0. \quad (29)$$

Furthermore, the problem (29) can be rewritten as a standard convex quadratic program problem:

$$\min_z \frac{1}{2} z^T Bz + c^T z, \text{ s.t. } z \geq 0, \quad (30)$$

where

$$z = \begin{pmatrix} u \\ v \end{pmatrix}, \quad c = \lambda e_{2n} + \begin{pmatrix} u \\ v \end{pmatrix}, \quad b = A^T y, \quad B = \begin{pmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{pmatrix},$$

$B$  is a semi-definite positive matrix. Recently, the problem (30) was reformulated as a linear variable inequality (LVI) problem by Xiao *et al.* [8] [27]. They pointed out that this LVI problem is equivalent to a linear complementary problem, and  $z$  is a solution of the linear complementary problem if and only if it is a solution of the following nonlinear monotone equations:

$$F(z) = \min\{z, Bz + c\} = 0, \quad (31)$$

where  $F(z)$  is Lipschitz continuous. This result indicates that problem (28) can be solved by MPRP projection method.

In this part of numerical experiments, a compressive sensing scenario is considered, which aims to reconstruct a length- $n$  sparse signal from significantly fewer  $m$  observations, where  $m \ll n$ . The quality of restoration is measured by the mean of squared error (MSE) to the original signal  $\bar{x}$ , that is

$$\text{MSE} = \frac{1}{n} \|\bar{x} - x^*\|,$$

where  $x^*$  is the restored signal. In practice,  $n = 2^{12}$  and  $m = 2^{10}$ , and the original contains  $2^6$  randomly non-zero elements.  $A$  is the Gaussian matrix generated by Matlab's code `rand(m,n)`, the measurement  $y$  contains noise,

$$y = A\bar{x} + \omega,$$

where  $\omega$  is the Gaussian noise distributed as  $N(0, 10^{-4})$ . The merit function is

$$f(x) = \frac{1}{2} \|y - Ax\|_2^2 + \tau \|x\|_1,$$

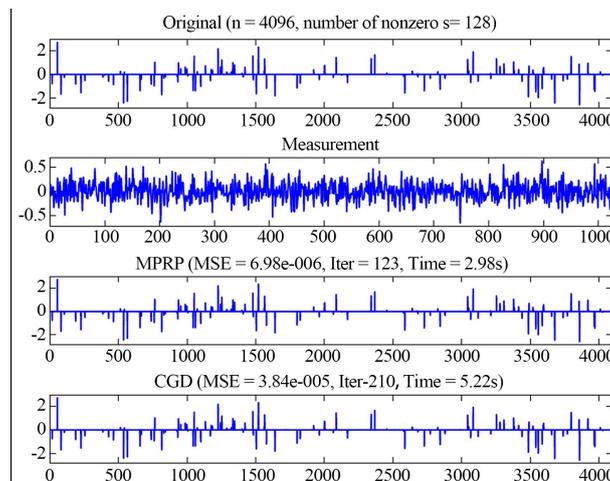
where  $\tau$  is forced to decrease as the measure in. The experiment starts at the measurement image, *i.e.*  $x_0 = A^T y$ , and terminates when the relative change of the iteration satisfies:

$$\text{Tol} = \frac{\|f_k - f_{k-1}\|}{\|f_{k-1}\|} < 10^{-5},$$

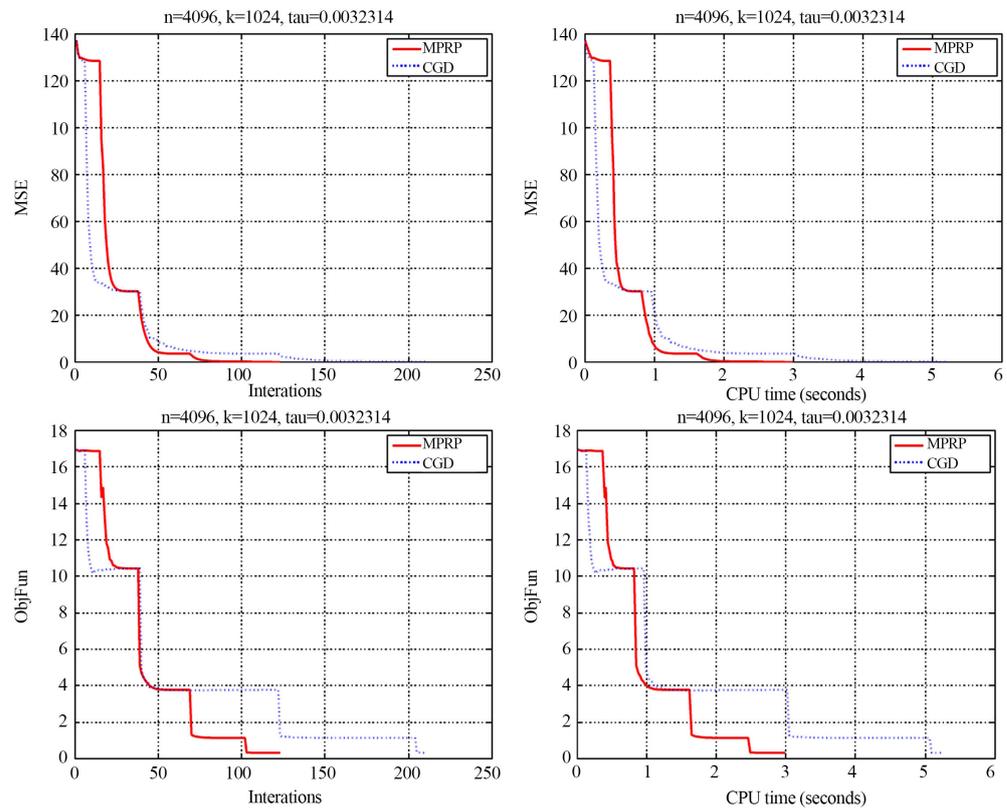
where  $f_k$  is the function value at  $x_k$ .

We compare the proposed MPRP method with CGD method for this problem. In both methods, the parameters are taken as  $\xi = 10$ ,  $\sigma = 10^{-4}$  and  $\rho = 0.5$ . The same initial point and continuation technique on parameter  $\tau$  are used in both methods.

**Figure 2** shows simulation results of MPRP and CGD for a signal sparse reconstruction. As we can see in **Figure 2**, the original sparse signal is restored highly exactly both by MPRP and CGD. **Figure 3** provides a series of comparisons among the objective function values and relative error as the iteration numbers and computing time increase. As we can see in **Figure 3**, the descent rates of MSE and objective function values of MPRP method are faster. The experiments are repeated for 15 random different noise samples in **Table 5**. We report the



**Figure 2.** From top to bottom: the original signal, the measurement, and the recovery signals by two methods MPRP and CGD, respectively.



**Figure 3.** Comparison results of MPRP and CGD methods. From left to right: the changed trends of MSE and the changed trends of the objective function values goes along with the number of iterations and CPU time in seconds, respectively.

**Table 5.** The experiment results for MPRP/CGD on  $l_1$ -norm regularization problem.

MSE	Niter	CPU(s)	MSE	Niter	CPU(s)
9.152e-006	119	2.69	2.278e-005	227	6.73
1.562e-005	120	3.23	6.210e-005	172	4.72
6.780e-006	127	3.47	2.520e-005	209	5.83
8.236e-006	124	3.05	3.367e-005	236	7.48
1.446e-005	120	3.09	8.207e-005	167	4.64
9.091e-006	110	2.25	4.870e-005	221	6.09
8.346e-006	122	3.31	5.382e-005	174	5.47
8.669e-006	117	3.23	4.233e-005	216	5.91
6.977e-006	123	3.33	3.839e-005	210	5.78
8.973e-006	122	3.70	3.789e-005	225	6.88
1.050e-005	119	3.30	5.531e-005	208	5.86
1.204e-005	128	2.63	5.370e-005	204	5.27
6.265e-006	111	3.52	1.873e-005	202	6.66
8.977e-006	129	3.70	3.035e-005	222	6.28
7.975e-006	126	3.47	6.946e-005	172	4.78

number of iterations (Niter) and the CPU time (in second) required for the whole testing process. From **Table 5**, we can see that MPRP method is better than CGD method. For example, the new method's iteration number and CPU time are much less than those of the CGD method. To summarize, these experiment results show that the proposed algorithm MPRP can work well in an efficient manner.

## 5. Conclusion

In this paper, we proposed a conjugate gradient projection algorithm for solving large-scale nonlinear convex constrained monotone equations based on the well-known Polak-Ribière-Polyak conjugate gradient method which is one of the most effective conjugate gradient methods to solve the unconstrained optimization problems. The algorithm combines CG technique with projection scheme and is a derivative-free method, so it can be applied to solve large-scale non-smooth equations for its low storage requirement. Under some technical conditions, we have established the global convergence. Another contribution of this paper is to use the given method to solve the  $l_1$ -norm regularized problems in compressive sensing.

## Acknowledgements

This work was supported by the Scientific Research Project of Tianjin Education Commission (No. 2019KJ232).

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Dirkse, S.P. and Ferris, M.C. (1995) MCPLIB: A Collection of Nonlinear Mixed Complementarity Problems. *Optimization Methods & Software*, **5**, 319-345. <https://doi.org/10.1080/10556789508805619>
- [2] Meintjes, K. and Morgan, A.P. (1990) Chemical Equilibrium Systems as Numerical Test Problems. *ACM Transactions on Mathematical Software*, **16**, 143-151. <https://doi.org/10.1145/78928.78930>
- [3] Wood, A.J. and Wollenberg, B.F. (1996) Power Generations, Operations and Control. Wiley, New York.
- [4] Solodov, M.V. and Svaiter, B.F. (1998) A Globally Convergent Inexact Newton Method for Systems of Monotone Equations. In: Fukushima, M. and Qi, L., Eds., *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic, 355-369. [https://doi.org/10.1007/978-1-4757-6388-1\\_18](https://doi.org/10.1007/978-1-4757-6388-1_18)
- [5] Wang, C.W. and Wang, Y.J. (2009) A Superlinearly Convergent Projection Method for Constrained Systems of Nonlinear Equations. *Journal of Global Optimization*, **44**, 283-296. <https://doi.org/10.1007/s10898-008-9324-8>
- [6] Hu, Y.P. and Wei, Z.X. (2015) Wei-Yao-Liu Conjugate Gradient Projection Algorithm for Nonlinear Monotone Equations with Convex Constraints. *International*

*Journal of Computer Mathematics*, **92**, 2261-2272.

<https://doi.org/10.1080/00207160.2014.977879>

- [7] Liu, J.K. and Li, S.J. (2015) A Projection Method for Convex Constrained Monotone Nonlinear Equations with Applications. *Computers and Mathematics with Applications*, **70**, 2442-2453. <https://doi.org/10.1016/j.camwa.2015.09.014>
- [8] Xiao, Y.H. and Zhu, H. (2013) A Conjugate Gradient Method to Solve Convex Constrained Monotone Equations with Applications in Compressive Sensing. *Journal of Mathematical Analysis and Applications*, **405**, 310-319. <https://doi.org/10.1016/j.jmaa.2013.04.017>
- [9] Yu, G.H., Niu, S.Z. and Ma, J.H. (2013) Multivariate Spectral Gradient Projection Method for Nonlinear Monotone Equations with Convex Constraints. *Journal of Industrial and Management Optimization*, **9**, 117-129. <https://doi.org/10.3934/jimo.2013.9.117>
- [10] Polak, E. (1969) The Conjugate Gradient Method in Extreme Problems. *USSR Computational Mathematics and Mathematical Physics*, **9**, 94-112. [https://doi.org/10.1016/0041-5553\(69\)90035-4](https://doi.org/10.1016/0041-5553(69)90035-4)
- [11] Polak, E. and Ribière, G. (1969) Note sur la convergence de méthodes de directions conjuguées. *Revue Francaise d'Informatique et de Recherche Opératinelle*, **3**, 35-43. <https://doi.org/10.1051/m2an/196903R100351>
- [12] Zhang, L. and Li, J.L. (2011) A New Globalization Technique for Nonlinear Conjugate Gradient Methods for Nonconvex Minimization. *Applied Mathematics and Computation*, **217**, 10295-10304. <https://doi.org/10.1016/j.amc.2011.05.032>
- [13] Hu, Y.P. and Wei, Z.X. (2014) A Modified Liu-Storey Conjugate Gradient Projection Algorithm for Nonlinear Monotone Equations. *International Mathematical Forum*, **9**, 1767-1777. <https://doi.org/10.12988/imf.2014.411197>
- [14] Yuan, G.L. and Hu, W.J. (2018) A Conjugate Gradient Algorithm for Large-Scale Unconstrained Optimization Problems and Nonlinear Equations. *Journal of Inequalities and Applications*, **1**, Article No.: 113. <https://doi.org/10.1186/s13660-018-1703-1>
- [15] Yuan, G.L., Meng, Z.H. and Li, Y. (2016) A Modified Hestenes and Stiefel Conjugate Gradient Algorithm for Large-Scale Nonsmooth Minimizations and Nonlinear Equations. *Journal of Optimization Theory and Applications*, **168**, 129-152. <https://doi.org/10.1007/s10957-015-0781-1>
- [16] Yuan, G.L., Wei, Z.X. and Li, G.Y. (2014) A Modified Polak-Ribière-Polyak Conjugate Gradient Algorithm for Nonsmooth Convex Programs. *Journal of Computational and Applied Mathematics*, **255**, 86-96. <https://doi.org/10.1016/j.cam.2013.04.032>
- [17] Yuan, G.L. and Zhang, M.J. (2015) A Three-Terms Polak-Ribière-Polyak Conjugate Gradient Algorithm for Large-Scale Nonlinear Equations. *Journal of Computational and Applied Mathematics*, **286**, 186-195. <https://doi.org/10.1016/j.cam.2015.03.014>
- [18] Yuan, G.L. and Zhang, M.J. (2013) A Modified Hestenes-Stiefel Conjugate Gradient Algorithm for Large-Scale Optimization. *Numerical Functional Analysis and Optimization*, **34**, 914-937. <https://doi.org/10.1080/01630563.2013.777350>
- [19] Yuan, G.L., Wei, Z.X. and Zhao, Q.M. (2014) A Modified Polak-Ribière-Polyak Conjugate Gradient Algorithm for Large-Scale Optimization Problems. *IIE Transactions*, **46**, 397-413. <https://doi.org/10.1080/0740817X.2012.726757>
- [20] Yu, Z.S., Lin, J., Sun, J., Xiao, Y.H., Liu, L.Y. and Li, Z.H. (2009) Spectral Gradient Projection Method for Monotone Nonlinear Equations with Convex Constraints. *Applied Numerical Mathematics*, **59**, 2416-2423.

- <https://doi.org/10.1016/j.apnum.2009.04.004>
- [21] Dolan, E.D. and Moré, J.J. (2002) Benchmarking Optimization Software with Performance Profiles. *Mathematical Programming*, **91**, 201-213.  
<https://doi.org/10.1007/s101070100263>
- [22] Elad, M. (2010) *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer Science & Business Media, LCC, Berlin.
- [23] Fessler, J.A. (2010) Model-Based Image Reconstruction for MRI. *IEEE Signal Processing Magazine*, **27**, 81-89. <https://doi.org/10.1109/MSP.2010.936726>
- [24] Romberg, J.K. (2008) Imaging via Compressive Sampling. *IEEE Signal Processing Magazine*, **25**, 14-20. <https://doi.org/10.1109/MSP.2007.914729>
- [25] Yang, J.C., Wright, J., Huang, T.S. and Ma, Y. (2010) Image Super-Resolution via Sparse Representation. *IEEE Transactions on Image Processing*, **19**, 2861-2873.  
<https://doi.org/10.1109/TIP.2010.2050625>
- [26] Figueiredo, M., Nowak, R. and Wright, S.J. (2007) Gradient Projection for Sparse Reconstruction, Application to Compressed Sensing and Other Inverse Problems. *IEEE Journal of Selected Topics in Signal Processing*, **1**, 586-597.  
<https://doi.org/10.1109/JSTSP.2007.910281>
- [27] Xiao, Y.H., Wang, Q.Y. and Hu, Q.J. (2011) Non-Smooth Equations Based Method for  $l_1$ -Norm Problems with Applications to Compressed Sensing. *Nonlinear Analysis: Theory, Methods & Applications*, **74**, 3570-3577.  
<https://doi.org/10.1016/j.na.2011.02.040>