A class of derivative-free CG projection methods for nonsmooth equations with an application to the LASSO problem

Min Sun[1,2], Maoying Tian[3]
1. School of Management, Qufu Normal University,
Shandong, 276826, P. R. China, E-mail:ziyouxiaodou@163.com
2. School of Mathematics and Statistics, Zaozhuang University,
Shandong, 277160, P. R. China
3. Department of Physiology, Shandong Coal Mining Health School,
Shandong, 277011, P. R. China, E-mail:hfmaoying@163.com

Abstract. In this paper, based on a modified Gram-Schmidt (MGS) process, we propose a class of derivative-free conjugate gradient (CG) projection methods for nonsmooth equations with convex constraints. Two attractive features of the new class of methods are: (i) its generated direction contains a free vector, which can be set as any vector such that the denominator of the direction does not equal to zero; (ii) it adopts a new line search which can reduce its computing cost. The new class of methods includes many efficient iterative methods for the studied problem as its special cases. When the underlying mapping is monotone, we establish its global convergence and convergence rate. Finally, preliminary numerical results about the LASSO problem show that the new class of methods is promising compared to some existing ones.

Keywords: Monotone constrained equations, derivative-free method, global convergence, the LASSO problem.

1 Introduction

Derivative-free projection methods are a class of very popular and widely used methods for solving the nonsmooth equations with convex constraints as follows

$$F(x^*) = 0, \quad x^* \in \mathcal{X}, \tag{1}$$

where $F : \mathcal{X} \to \mathcal{R}^m$ is a continuous monotone mapping (not necessarily smooth), i.e.,

$$(x-y)^\top (F(x) - F(y)) \geq 0, \quad \forall x, y \in \mathcal{X}, \tag{2}$$

and the set $\mathcal{X} \subseteq \mathcal{R}^n$ is a nonempty closed convex set. Obviously, problem (1) contains m independent nonlinear equations and n unknown variables. Then it is called an over-determined system when $m > n$ and an under-determined system when $m < n$ in the literature. In this paper, we consider the under-determined problem (1) with $m \ll n$, which often permits infinite solutions, and we assume that the solution set of problem (1), denoted by $\mathcal{X}^*$, is nonempty throughout this paper.

Problem (1) serves as a unified model of many mathematical problems encountered in different disciplines. For example, the famous variational inequality problems (VIPs), which aims to find a vector $x^* \in \mathcal{X}$ such that

$$(x - x^*)^\top F(x^*) \geq 0, \quad \forall x \in \mathcal{X},$$

can be transformed into the following fixed point equations

$$x^* = P_{\mathcal{X}}[x^* - \beta F(x^*)],$$

where $\beta > 0$ is a constant and $P_{\mathcal{X}}(x)$ denotes the orthogonal projection of a vector $x \in \mathcal{R}^n$ onto the convex set $\mathcal{X}$. Then the solution set of the VIPs coincides with that of problem (1) with

$$F(x) := x - P_{\mathcal{X}}[x - \beta F(x)], \quad \mathcal{X} := \mathcal{R}^n.$$

More applications of problem (1) in economic equilibrium analysis, chemical equilibrium systems, compressive sensing and control theory can be found in [1, 2, 3, 4, 5, 6] and reference therein.

In the era of big data, the scale of problem (1) is bigger and bigger. Then, the derivative-free iterative methods for problem (1) which only use the values of the mapping $F(x)$ become more appealing to practitioners, and during the past few decades, it has been a fascinating area of research and has drawn continuing interest from both researchers and practitioners. For example, in the seminal papers [7, 8], Cruz et al. proposed some spectral gradient projection methods and spectral residual methods for solving problem (1) with $\mathcal{X} = \mathcal{R}^n$. The most characteristics of these methods is that they not only do not need the first order derivative of the mapping $F(x)$ but also do not need to solve any linear equations, and thus they are suitable to solve large-scale nonlinear constrained equations (1). Later, Zhang and Zhou [9] developed a spectral gradient projection method for nonlinear equations, which combines the spectral gradient method [10] with the projection technique [11]. Subsequently, this method was extended by Yu et al. [12] to solve problem (1). Other spectral gradient-type methods for problem (1) can be found in [13, 14].

Similar to the spectral gradient method, the conjugate gradient (CG) method is also an effective first-order iterative method for solving unconstrained optimization problems. Then motivated by the numerical performance of the spectral gradient-type methods, researchers have tried to extend the conjugate gradient method to solve problem (1) and proposed some efficient derivative-free CG-type projection methods. For example, Cheng [15] firstly extended the Polak-Ribière-Polyak (PRP) method, one of the most efficient conjugate gradient methods for unconstrained optimization problem, to solve problem (1) with $\mathcal{X} = \mathcal{R}^n$. Later, many other CG methods for unconstrained optimization problem are extended successfully to solve problem (1), such as the modified PRP method [16], the modified Fletcher-Reeves method [17, 18], the CG_DESCENT method [4], the Hestenes-Stiefel method [19] and the hybrid conjugate gradient method [20] etc. For more details on the derivative-free CG projection methods for problem (1), the interested reader is referred to [21, 22, 23, 24] and the references therein.

Our concern now is the following: *Can we construct a class of derivative-free CG projection methods which includes some of the above methods as its special cases?*

In this paper, we give a positive answer to this question. Motivated and inspired by the works of Cheng et al. [25], Sun et al. [26, 27] and Feng et al. [28], we will introduce a new class of derivative-free CG projection methods for problem (1), whose main contribution is as follows: at the k-th ($k \geq 1$) iteration, the direction d_k is recursively defined by

$$d_k = -F(x_k) + \beta_k d_{k-1}. \tag{3}$$

Then, in order to ensure the direction d_k satisfy the property

$$F(x_k)^\top d_k = -\|F(x_k)\|^2, \tag{4}$$

we only need to choose a vector from the subspace $\Omega_k = \{v | F(x_k)^\top v = 0\}$ to replace the second term $\beta_k d_{k-1}$ of d_k, and get

$$d_k = -F(x_k) + v, \quad v \in \Omega_k.$$

For example, if we choose $v = 0 \in \Omega_k$, we get the steepest descent direction; if we choose

$$v = \beta_k \Big(d_{k-1} - \frac{F(x_k)^\top d_{k-1}}{\|F(x_k)\|^2} F(x_k) \Big) \in \Omega_k,$$

which is obviously motivated by the Gram-Schmidt (MGS) process, we get the direction used in [27, 28].

The remainder of the paper is organized as follows. In Section 2, we collect some definitions and results for further investigation, and propose a class of derivative-free CG projection methods for problem (1). In Section 3, we prove the global convergence and convergence rate of the proposed method. In Section 4, we provide some numerical experiments to illustrate the effectiveness of the proposed method in solving the LASSO problem. Some conclusions are summarized in the final section.

2 New method

This section is devoted to devising a new class of derivative-free CG projection methods for problem (1), and some properties of the proposed method are also presented.

The projection operator $P_{\mathcal{X}}[x]$, which is defined as a mapping from $\mathcal{R}^n$ to the nonempty closed convex subset $\mathcal{X}$:

$$P_{\mathcal{X}}[x] := \text{argmin}\{\|y-x\| | y \in \mathcal{X}\}, \quad \forall x \in \mathcal{R}^n.$$

The projection operator $P_\Omega[x]$ satisfies the following nice properties [29].

Lemma 2.1 Let $\mathcal{X}$ be a closed convex subset of $\mathcal{R}^n$. For any $x, y \in \mathcal{R}^n$, the following inequality holds

$$\|P_{\mathcal{X}}[x] - P_{\mathcal{X}}[y]\| \leq \|x - y\|. \tag{5}$$

Assumption 2.1 The mapping $F(x)$ satisfies the Lipschitz condition, i.e., there exists a constant $L > 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{X}. \tag{6}$$

Based on the Gram-Schmidt (GS) orthogonalization method, we first present a modified GS method as follows: Given two vectors $\alpha_1, \alpha_2 \in \mathcal{R}^n$, then the two vectors β_1, β_2 defined as

$$\begin{cases} \beta_1 = \alpha_1, \\ \beta_2 = \alpha_2 - \frac{\alpha_2^\top \beta_1}{\beta_1^\top v} v, \end{cases} \tag{7}$$

are orthogonal, where $v \in \mathcal{R}^n$ such that $\beta_1^\top v \neq 0$.

Example 2.1 Given $\alpha_1 = [1, 1, 2]^\top, \alpha_2 = [2, 4, 4]^\top$, by the MGS method, we have

(i) if we set $v = \alpha_1$, then $\beta_1 = [1, 1, 2]^\top, \beta_2 = [-1/3, 5/3, -2/3]^\top$;

(ii) if we set $v = [1, 1, 1]^\top$, then $\beta_1 = [1, 1, 2]^\top, \beta_2 = [-3/2, 1/2, 1/2]^\top$;

(iii) if we set $v = [1, 2, 1]^\top$, then $\beta_1 = [1, 1, 2]^\top, \beta_2 = [-4/5, -8/5, 6/5]^\top$.

Now we recall the iterative scheme of CG method for solving unconstrained optimization problem

$$\min f(x), \qquad x \in \mathcal{R}^n, \tag{8}$$

where $f : \mathcal{R}^n \to \mathcal{R}$ is a smooth function, whose gradient $\nabla f(x)$ denoted by $g(x)$. The conjugate gradient method is one of the most efficient methods for solving (8), which generates an iterative sequence $\{x_k\}$ by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \ldots, \tag{9}$$

where $\alpha_k > 0$ is the step size determined by some line search, and d_k is the search direction defined by

$$d_k = \begin{cases} -g(x_k), & \text{if} \quad k = 0, \\ -g(x_k) + \beta_k d_{k-1}, & \text{if} \quad k \geq 1. \end{cases}$$

Here β_k is the CG parameter, which is the main difference of different CG methods [30]. Recently, in the seminar papers [31, 32], Zhang et al. firstly presented two modified CG methods, i.e., the modified PRP method and the modified FR method, which satisfy the sufficient descent property naturally

$$g(x_k)^\top d_k = -\|g(x_k)\|^2, \tag{10}$$

i.e.,

$$g(x_k)^\top (g(x_k) + d_k) = 0,$$

and their directions are defined as follows

$$d_k^1 = \begin{cases} -g(x_k), & \text{if} \quad k = 0, \\ -g(x_k) + \beta_k^{\mathrm{PRP}} d_{k-1}^1 - \theta_k^1 y_{k-1}, & \text{if} \quad k \geq 1, \end{cases}$$

where

$$\theta_k^1 = \frac{g(x_k)^\top d_{k-1}}{\|g(x_{k-1})\|^2}, \quad \beta_k^{\mathrm{PRP}} = \frac{y_{k-1}^\top g(x_k)}{\|g(x_{k-1})\|^2}, \quad y_{k-1} = g(x_k) - g(x_{k-1})$$

and

$$d_k^2 = \begin{cases} -g(x_k), & \text{if} \quad k = 0, \\ -\theta_k^2 g(x_k) + \beta_k^{\mathrm{FR}} d_{k-1}^2, & \text{if} \quad k \geq 1, \end{cases}$$

where

$$\theta_k^2 = \frac{y_{k-1}^\top d_{k-1}}{\|g(x_{k-1})\|^2}, \quad \beta_k^{\mathrm{FR}} = \frac{\|g(x_k)\|^2}{\|g(x_{k-1})\|^2}.$$

Since d_k^1 and d_k^2 both satisfy the property (10), we have $g(x_k) + d_k^i \in \bar{\Omega}_k (i = 1, 2)$, where $\bar{\Omega}_k = \{v | g(x_k)^\top v = 0\}$. Furthermore, through aborative observation, we find that $g(x_k) + d_k^i (i = 1, 2)$ can be rewritten as the following unified form

$$g(x_k) + d_k = \beta_k \Big(d_{k-1} - \frac{g(x_k)^\top d_{k-1}}{g(x_k)^\top v} v \Big), \tag{11}$$

where $v \in \mathcal{R}^n$ such that $g(x_k)^\top v \neq 0$. In fact, setting $v = y_{k-1}$ and $g(x_k)$, respectively, we can get $g(x_k) + d_k^1$ and $g(x_k) + d_k^2$. Comparing (7) and (11), we can find that (11) is motivated by (7).

Based on the foregoing analysis, we now present a new class of derivative-free CG projection method for problem (1).

Algorithm 2.1: Derivative-free CG projection method, denoted by the DFCGPM

Step 0. Let parameters $C > 0, r \geq 0, \sigma > 0, \beta > 0, 0 < \rho < 1, 0 < \gamma < 2$, tolerance error $\varepsilon > 0$. Choose an initial point $x_0 \in \mathcal{X}$, and set $k = 0$.

Step 1. If $\|F(x_k)\| < \varepsilon$, then stop; otherwise go to step 2.

Step 2. Compute d_k by

$$d_k = \begin{cases} -F(x_k), & \text{if} \quad k = 0, \\ -F(x_k) + \beta_k \Big(d_{k-1} - \frac{F(x_k)^\top d_{k-1}}{F(x_k)^\top v_k} v_k \Big), & \text{if} \quad k \geq 1, \end{cases} \tag{12}$$

where β_k is a parameter satisfying

$$\frac{|\beta_k| \|d_{k-1}\| \|v_k\|}{|F(x_k)^\top v_k|} \leq \frac{C}{\|F(x_{k-1})\|^r}, \tag{13}$$

and the vector $v_k \in \mathcal{R}^n$ satisfies: (i) if the numerator of β_k contains the term $F(x_k)^\top v_k$ and the denominator of β_k contains the term v_k, then v_k can take any vector; (ii) if the numerator of β_k contains the term $F(x_k)^\top v_k$ and the denominator of β_k does not contain the term v_k, then v_k can take any nonzero vector; (iii) if the numerator of β_k does not contain the term $F(x_k)^\top v_k$, then v_k can take any vector such that $F(x_k)^\top v_k \neq 0$.

Step 3. Compute a temporal iterate $z_k = x_k + \alpha_k d_k$, where $\alpha_k = \beta\rho^{m_k}$ with m_k being the smallest nonnegative integer m such that

$$-\langle F(x_k + \beta\rho^m d_k), d_k\rangle \geq \sigma\beta\rho^m \min\{1, \|F(x_k + \beta\rho^m d_k)\|\}\|d_k\|^2. \tag{14}$$

Step 4. If $\|F(z_k)\| < \varepsilon$, then stop; otherwise compute the new iterate x_{k+1} by

$$x_{k+1} = P_{\mathcal{X}}[x_k - \gamma\xi_k F(z_k)], \tag{15}$$

where

$$\xi_k = \frac{\langle F(z_k), x_k - z_k\rangle}{\|F(z_k)\|^2}.$$

Set $k = k + 1$ and go to Step 1.

Remark 2.1 The direction defined by (14) satisfies the equality (4), by which and the Cauchy-Schwartz inequality, we have

$$\|d_k\| \geq \|F(x_k)\|. \tag{16}$$

Remark 2.2 If $\|F(x_k)\| \leq \varepsilon$, from $x_k \in \mathcal{X}$, the iterate x_k is an approximate solution of problem (1). If $\|F(z_k)\| < \varepsilon$, from the line search (14), we have $\|d_k\| \leq \varepsilon/\alpha_k$, which together with (16) implies $\|F(x_k)\| \leq \varepsilon/\alpha_k$, thus the iterate x_k also can be viewed as an approximate solution of problem (1). So the stopping criteria of the DFCGPM is reasonable.

Remark 2.3 From the definitions of d_k and β_k, for $k \geq 1$, it follows that

$$\begin{aligned}
&\|d_k\| \\
&\leq \|F(x_k)\| + |\beta_k|\|d_{k-1}\| + |\beta_k|\frac{|F(x_k)^\top d_{k-1}|}{|F(x_k)^\top v_k|}\|v_k\| \\
&\leq \|F(x_k)\| + |\beta_k|\|d_{k-1}\| + \frac{|\beta_k|\|d_{k-1}\|}{|F(x_k)^\top v_k|}\|v_k\|\|F(x_k)\| \\
&\leq \|F(x_k)\| + \frac{C}{\|F(x_{k-1})\|^r}\frac{|F(x_k)^\top v_k|}{\|v_k\|} + \frac{C}{\|F(x_{k-1})\|^r}\|F(x_k)\| \\
&\leq \left(1 + \frac{2C}{\|F(x_{k-1})\|^r}\right)\|F(x_k)\|.
\end{aligned} \tag{17}$$

Remark 2.4 Generally speaking, to find the smallest nonnegative positive m such that the inequality (14) holds, we often the trial method. That is, we first set $m = 0$ in (14), and if it doesn't hold, we set $m = 1, 2, \ldots$, until it holds. This indicates that the smaller the right hand-side of (14) is, the smaller the computing cost of the line search is. Therefore, comparing with the line search in [9], the right hand-side of (14) incorporates a term $\min\{1, \|F(x_k + \beta\rho^m d_k)\|\}$, which can decrease the computing cost of the line search when the iterate x_k is far from the solution of problem (1).

Remark 2.5 The DFCGPM includes many efficient derivative-free CG projection methods as its special cases.

(i) If we set $v_k = F(x_k)$ and

$$\beta_k = \beta_k^{\text{MFR}} := \frac{\|F(x_k)\|^2}{\|d_{k-1}\|^2},$$

the direction defined by (12) reduces to the direction used in [18]. Furthermore, we have

$$\frac{|\beta_k^{\text{MFR}}|\|d_{k-1}\|\|F(x_k)\|}{\|F(x_k)\|^2} = \frac{\|F(x_k)\|}{\|d_{k-1}\|} \leq 1,$$

where the second inequality comes from (16), then the parameter β_k^{MFR} satisfies (13) with $C = 1$ and $r = 0$.

(ii) If we set

$$v_k = z_{k-1} := y_{k-1} + \left(\max\left\{ 0, -\frac{d_{k-1}^\top y_{k-1}}{d_{k-1}^\top s_{k-1}} \right\} + t\|F(x_{k-1})\|^v \right) s_{k-1},$$

and

$$\beta_k = \beta_k^{\text{MHS}} := \frac{F(x_k)^\top z_{k-1}}{d_{k-1}^\top z_{k-1}},$$

where

$$s_{k-1} = \alpha_{k-1} d_{k-1} = z_{k-1} - x_{k-1}, y_{k-1} = F(z_{k-1}) - F(x_{k-1}), t > 0, v > 0,$$

the direction defined by (12) reduces to the direction used in [19]. Furthermore, we have

$$\frac{|\beta_k^{\text{MHS}}|\|d_{k-1}\|\|z_{k-1}\|}{|F(x_k)^\top z_{k-1}|} \leq \frac{\|z_{k-1}\|}{t\alpha_{k-1}\|F(x_{k-1})\|^v \|d_{k-1}\|} \leq \frac{M}{t\|F(x_{k-1})\|^v},$$

where the two inequalities follow from the definition of z_{k-1} and $M > 0$ is a constant (see (24) in [19]), thus the parameter β_k^{MHS} satisfies (13) with $C = M/t$ and $r = v > 0$.

(iii) If we set $v_k = y_{k-1} := F(x_k) - F(x_{k-1})$ and

$$\beta_k = \beta_k^{\text{MPRP}} := \frac{y_{k-1}^\top F(x_k)}{\min\{\mu\|d_{k-1}\|\|y_{k-1}\|, \|F(x_{k-1})\|^2\}},$$

where $\mu > 0$, the direction defined by (12) reduces to the direction used in [21]. Furthermore, we have

$$\frac{|\beta_k^{\text{MPRP}}|\|d_{k-1}\|\|y_{k-1}\|}{|F(x_k)^\top y_{k-1}|} \leq \frac{|y_{k-1}^\top F(x_k)|}{\mu\|d_{k-1}\|\|y_{k-1}\|} \frac{\|d_{k-1}\|\|y_{k-1}\|}{|F(x_k)^\top y_{k-1}|} \leq \frac{1}{\mu},$$

i.e., the parameter β_k^{PRP} satisfies (13) with $C = 1/\mu$ and $r = 0$.

(iv) If we set $v_k = F(x_k)$ and

$$\beta_k = \beta_k^{\text{MP}} := \frac{(y_{k-1} - \bar{s}_{k-1})^\top F(x_k)}{d_{k-1}^\top w_{k-1}},$$

where

$$w_{k-1} = y_{k-1} + \bar{\gamma}\bar{s}_{k-1}, \bar{\gamma} > 0, y_{k-1} = F(z_{k-1}) - F(x_{k-1}), \bar{s}_{k-1} = z_{k-1} - x_{k-1} = \alpha_{k-1} d_{k-1},$$

the direction defined by (12) reduces to the direction used in [22]. Furthermore, by the proof of Theorem 3.1 in [22], we have

$$|\beta_k^{\rm MP}| < \frac{(L+1)\|F(x_k)\|}{\bar{\gamma}\|d_{k-1}\|},$$

thus

$$\frac{|\beta_k^{\rm MP}|\|d_{k-1}\|\|F(x_k)\|}{\|F(x_k)\|^2} \le \frac{L+1}{\bar{\gamma}},$$

where the second inequality comes from (16), then the parameter β_k satisfies (13) with $C = (L+1)/\bar{\gamma}$ and $r = 0$.

(v) If we set $v_k = F(x_k)$ and β_k satisfies

$$|\beta_k| < v\frac{\|F(x_k)\|}{\|d_{k-1}\|},$$

where $v > 0$, the direction defined by (12) reduces to the directions used in [27, 28]. Furthermore, we have

$$\frac{|\beta_k|\|d_{k-1}\|\|F(x_k)\|}{\|F(x_k)\|^2} \le v,$$

where the second inequality comes from (16), then the parameter β_k satisfies (13) with $C = v$ and $r = 0$.

3 Convergence analysis

In this section, we analyze the convergence properties of the DFCGPM under the Assumption 2.1, including the global convergence and the convergence rate.

In what follows we assume that $\|F(x_k)\| \neq 0$ for all k, namely, the DFCGPM generates two infinite sequence $\{x_k\}$ and $\{z_k\}$. The following lemma indicates that the line search (14) is well defined.

Lemma 3.1 For each integer $k \ge 0$, there exists a nonnegative integer m_k satisfying the inequality (14).

Proof Obviously, if the Armijo line search (14) is executed, we have $\|F(x_k)\| \ge \varepsilon > 0$. Now we prove the lemma by contradiction. Assume that there exists an integer $k_0 \ge 0$ such that the inequality (14) does not hold for any nonnegative integer m, so we have the following inequality

$$-\langle F(x_{k_0} + \beta_{k_0}\rho^m d_{k_0}), d_{k_0}\rangle < \sigma\beta_{k_0}\rho^m \min\{1, \|F(x_{k_0} + \beta_{k_0}\rho^m d_{k_0})\|\}\|d_{k_0}\|^2, \quad \forall m \ge 0.$$

Letting $m \to +\infty$ on both sides of the above inequality and using the continuity of the mapping $F(x)$, we get

$$-\langle F(x_{k_0}), d_{k_0}\rangle \le 0.$$

This together with the inequality (4) gives $\|F(x_{k_0})\| = 0$, which contradicts the fact $\|F(x_{k_0})\| \ge \varepsilon$. This completes the proof.

The following lemma provides a positive lower bound of the step size α_k for all $k \geq 0$.

Lemma 3.2 Suppose that Assumption 2.1 holds. Let the sequence $\{x_k\}$ be generated by the DFCGPM. Then, for all $k \geq 0$, we have

$$\alpha_k \geq \min\left\{\beta, \frac{\rho\|F(x_k)\|^2}{(L+\sigma\min\{1, \|F(x_k+\alpha_k\rho^{-1}d_k)\|\})\|d_k\|^2}\right\}. \tag{18}$$

Proof According to principle of the Armijo line search, if $\alpha_k \neq \beta$, the positive number $\alpha'_k = \alpha_k/\rho$ does not satisfy the inequality (14), so we have the following inequality

$$-\langle F(x_k+\alpha'_k d_k), d_k\rangle < \sigma\alpha'_k \min\{1, \|F(x_k+\alpha'_k d_k)\|\}\|d_k\|^2.$$

By Assumption 2.1, (4) and (13), we get

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

which implies the inequality (18). This completes the proof.

The following lemma is motivated by [33]. For completeness, we give its detail proof.

Lemma 3.3 Let the sequences $\{x_k\}$ and $\{z_k\}$ be generated by the DFCGPM. Then, for any $x^* \in \mathcal{X}^*$, there exists a constant $c_1 > 0$ such that

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - c_1\|x_k - z_k\|^4, \tag{19}$$

and

$$\lim_{k\to\infty} \alpha_k\|d_k\| = 0. \tag{20}$$

Furthermore, both $\{x_k\}$ and $\{z_k\}$ are bounded.

Proof From the line search (14), it follows that

$$\langle F(z_k), x_k - z_k\rangle = -\alpha_k\langle F(z_k), d_k\rangle \geq \sigma\min\{1, \|F(z_k)\|\}\alpha_k^2\|d_k\|^2 = \sigma\min\{1, \|F(z_k)\|\}\|x_k - z_k\|^2. \tag{21}$$

Since $x^* \in \mathcal{X}^*$, $x_k \in \mathcal{X}$ and the mapping $F(x)$ is monotone, we get

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

So

$$\langle F(z_k), z_k - x_k\rangle \geq \langle F(z_k), x^* - x_k\rangle. \tag{22}$$

Then, by (5), (15), (21) and (22), it holds that

$$\begin{aligned}
&\|x_{k+1} - x^*\|^2 \\
&\le \|x_k - \gamma\xi_k F(z_k) - x^*\|^2 \\
&= \|x_k - x^*\|^2 - 2\gamma\xi_k\langle F(z_k), x_k - x^*\rangle + \gamma^2\xi_k^2\|F(z_k)\|^2 \\
&\le \|x_k - x^*\|^2 - 2\gamma\xi_k\langle F(z_k), x_k - z_k\rangle + \gamma^2\xi_k^2\|F(z_k)\|^2 \\
&= \|x_k - x^*\|^2 - \gamma(2-\gamma)\ \frac{\langle F(z_k), x_k - z_k\rangle^2}{\|F(z_k)\|^2} \\
&\le \|x_k - x^*\|^2 - \gamma(2-\gamma)\sigma^2\min\{1, \|F(z_k)\|\}^2\ \frac{\|x_k - z_k\|^4}{\|F(z_k)\|^2}\ ,
\end{aligned} \tag{23}$$

which implies that the sequence $\{\|x_k - x^*\|\}$ is decreasing and convergent, and thus the sequence $\{x_k\}$ is bounded, i.e., there exists $M_1 > 0$, such that $\|F(x_k)\| \le M_1$ for all $k \ge 0$. If $\|F(z_k)\| \le 1$, the inequality (19) holds with $c_1 = \gamma(2-\gamma)\sigma^2$. If $\|F(z_k)\| > 1$, from (21), it follows that

$$\langle F(z_k), x_k - z_k\rangle \ge \sigma\|x_k - z_k\|^2.$$

Then, from the Cauchy-Schwartz inequality, the monotonicity of $F(x)$ and the above inequality, we have

$$\|F(x_k)\| \ge \frac{\langle F(x_k), x_k - z_k\rangle}{\|x_k - z_k\|} \ge \frac{\langle F(z_k), x_k - z_k\rangle}{\|x_k - z_k\|} \ge \sigma\|x_k - z_k\|.$$

This and the boundness of $\{F(x_k)\}$ implies that the sequence $\{z_k\}$ is boundness, and the sequence $\{F(z_k)\}$ is also boundness, i.e., there exists a constant $M_2 > 0$, such that $\|F(z_k)\| \le M_2$ for all $k \ge 0$. Then, by (23), the inequality (19) holds with $c_1 = \gamma(2-\gamma)\sigma^2/M_2^2$.

Furthermore, from (19), it follows that

$$c_1\sum_{k=0}^{\infty}\|x_k - z_k\|^4 \le \sum_{k=0}^{\infty}(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2) = \|x_0 - x^*\|^2 < +\infty.$$

Thus,

$$\lim_{k\to\infty}\alpha_k\|d_k\| = \lim_{k\to\infty}\|x_k - z_k\| = 0,$$

which together with the boundness of the sequence $\{x_k\}$ implies that the sequence $\{z_k\}$ is also bounded. This completes the proof.

Based on the above two lemmas, we now are ready to establish the global convergence of the DFCGPM.

Theorem 3.1 Let $\{x_k\}$ be the sequence generated by the DFCGPM. Then,

$$\lim_{k\to\infty}\inf\|F(x_k)\| = 0. \tag{24}$$

Furthermore, the sequence $\{x_k\}$ converges to a solution of problem (1).

Proof We prove the conclusion (24) by using reduction to absurdity. Suppose that (24) doesn't hold. Then, there exists a constant $\varepsilon_0 > 0$ such that

$$\|F(x_k)\| \ge \varepsilon_0, \quad \forall k \ge 0.$$

By (16), it follows that

$$\|d_k\| \geq \|F(x_k)\| \geq \varepsilon_0, \quad \forall k \geq 0.$$

This together with (20) implies that

$$\lim_{k\to\infty} \alpha_k = 0. \tag{25}$$

On the other hand, by (20), the boundness of the sequence $\{x_k\}$ and the continuity of $F(x)$, there exists $M_3 > 0$, such that

$$\|F(x_k + \alpha_k \rho^{-1} d_k)\| \leq M_3, \quad \forall k \geq 0. \tag{26}$$

Then by the inequality (17), it holds that

$$\|d_k\| \leq \left(1 + \frac{2C}{\varepsilon_0^r}\right) \|F(x_k)\|.$$

The above two inequalities together with (18) implies that

$$\alpha_k \geq \min\left\{\beta, \frac{\rho\varepsilon_0^{2r}}{(L + \sigma \min\{1, M_3\})(\varepsilon_0^r + 2C)^2}\right\} > 0, \quad \forall k \geq 0,$$

which contradicts (25). Therefore the conclusion (24) holds. Therefore, there exists an infinite index set $K \subseteq \{0, 1, 2, \ldots\}$ such that

$$\lim_{k\to\infty, k\in K} \|F(x_k)\| = 0.$$

Since the sequence $\{x_k\}$ is bounded, it has at least one cluster, saying $\bar{x}$, and without lose of generality, we assume that the subsequences $\{x_k : k \in K\}$ converges to $\bar{x}$. By the continuity of the mapping $F(x)$, it follows that

$$\|F(\bar{x})\| = 0,$$

which indicates that $\bar{x}$ is a solution of problem (1). So, setting $x^* = \bar{x}$ in (19), we immediately have

$$\|x_{k+1} - \bar{x}\| \leq \|x_k - \bar{x}\|, \quad \forall k \geq 0.$$

Consequently, we obtain the whose sequence $\{x_k\}$ converging to $\bar{x}$. This completes the proof.

In the remainder of this section, we are going to analyze the convergence rate of the DFCGPM with $r = 0$, which needs the following assumption except Assumption 2.1.

Assumption 3.1. For any $x^* \in \mathcal{X}^*$, there exist two positive constants c_2 and c_3 such that

$$c_2 \text{dist}(x, \mathcal{X}^*) \leq \|F(x)\|, \quad \forall x \in N(x^*, c_3), \tag{27}$$

where $\text{dist}(x, \mathcal{X}^*)$ denotes the distance from x to the solution set $\mathcal{X}^*$, and

$$N(x^*, c_3) = \{x \in R^n | \|x - x^*\| \leq c_3\}.$$

The following lemma is adapted from Lemma 6 in Chapter 2 of [34], which can help us establish the convergence rate of the DFCGPM with $r = 0$.

Lemma 3.4 Let $u_k > 0$ and let

$$u_{k+1} \leq u_k - a u_k^2, \quad a \geq 0.$$

Then

$$u_k \leq \frac{u_0}{1 + aku_0}.$$

Lemma 3.5 Suppose that Assumption 2.1 and Assumption 3.1 hold. Then there exists a constant $c_4 > 0$ such that

$$\alpha_k \geq c_4, \quad \forall k \geq 0. \tag{28}$$

Proof Since $r = 0$, by setting

$$c_4 = \min\left\{\beta, \frac{\rho}{(L + \sigma \min\{1, M_3\})(1 + 2C)^2}\right\},$$

the proof follows immediately from the results (17), (18) and (26). This completes the proof.

The following theorem indicates that the sequence $\{\text{dist}(x_k, \mathcal{X}^*)\}$ has the $\mathcal{O}(1/\sqrt{k})$ convergence rate.

Theorem 3.2 Suppose that Assumption 2.1 and Assumption 3.1 hold. Then there is a constant $\omega_1 > 0$ such that for sufficiently large k, we have

$$\text{dist}(x_k, \mathcal{X}^*) \leq \frac{1}{\sqrt{\omega_1 k + \text{dist}^{-2}(x_0, \mathcal{X}^*)}}.$$

Proof From Theorem 3.1, we assume that $x_k \to \bar{x} \in \mathcal{X}^*$ as $k \to \infty$. Thus, from the continuity of the mapping $F(x)$, it holds that

$$\lim_{k\to\infty} F(x_k) = F(\bar{x}) = 0.$$

Then, from (16), (17) and $r = 0$, we get

$$\lim_{k\to\infty} d_k = 0.$$

So,

$$\lim_{k\to\infty} F(z_k) = \lim_{k\to\infty} F(x_k + \alpha_k d_k) = 0.$$

Therefore, for sufficiently large k, $\min\{1, F(z_k)\} = F(z_k)$. It follows from (19) that

$$\|x_{k+1} - \bar{x}_k\|^2 \leq \|x_k - \bar{x}_k\|^2 - c_1\|x_k - z_k\|^4. \tag{29}$$

Now, let us deal with the last term of (29). Let $\bar{x}_k \in \mathcal{X}^*$ be the closest solution to x_k. Namely,

$$\|x_k - \bar{x}_k\| = \text{dist}(x_k, \mathcal{X}^*). \tag{30}$$

It follows from (16), (27) and (28) that

$$\|x_k - z_k\| = \alpha_k\|d_k\| \geq \alpha_k\|F(x_k)\| \geq c_2 c_4 \text{dist}(x_k, \mathcal{X}^*).$$

Substituting the above inequality into (29), we have

$$\begin{aligned}
&\text{dist}^2(x_{k+1}, \mathcal{X}^*) \\
&\le \|x_{k+1} - \bar{x}_k\|^2 \\
&\le \text{dist}^2(x_k, \mathcal{X}^*) - c_1 c_2^4 c_4^4 \text{dist}^4(x_k, \mathcal{X}^*).
\end{aligned}$$

Then, let $\omega_1 = c_1 c_2^4 c_4^4$, and by Lemma 3.4, we get

$$\text{dist}(x_k, \mathcal{X}^*) \le \frac{\text{dist}(x_0, \mathcal{X}^*)}{\sqrt{1 + \omega_1 k \text{dist}^2(x_0, \mathcal{X}^*)}} = \frac{1}{\sqrt{\omega_1 k + \text{dist}^{-2}(x_0, \mathcal{X}^*)}}.$$

The proof is completed.

Now, let us analyze the $\mathcal{O}(1/\sqrt{k})$ convergence rate of the sequence $\{x_k\}$, which needs the following assumption except Assumption 2.1 and Assumption 3.1.

Assumption 3.2 For any $x^* \in \mathcal{X}^*$, there exists a constant $c_3 > 0$ such that the mapping $F(x)$ is strongly monotone with modulus $\mu > 0$ on $N(x^*, c_3)$, i.e.,

$$\langle F(x) - F(y), x - y \rangle \ge \mu \|x - y\|^2, \quad \forall x, y \in N(x^*, c_3). \tag{31}$$

Theorem 3.3 Suppose that Assumption 2.1, Assumption 3.1 and Assumption 3.2 hold. Then there is a constant $\omega_2 > 0$ such that for sufficiently large k, we have

$$\|x_k - \bar{x}\| \le \frac{1}{\sqrt{\omega_2 k + \|x_0 - \bar{x}\|^{-2}}},$$

where $\bar{x} \in \mathcal{X}^*$ is the limit of the sequence $\{x_k\}$.

Proof Since $\bar{x} \in \mathcal{X}^*$, it follows from (19) that

$$\|x_{k+1} - \bar{x}\|^2 \le \|x_k - \bar{x}\|^2 - c_1 \|x_k - z_k\|^4. \tag{32}$$

For sufficiently large k, it follows from (31) that

$$\|F(x_k)\| = \|F(x_k) - F(\bar{x})\| \ge \mu \|x_k - \bar{x}\|.$$

This together with (16), (28) implies that

$$\|x_k - z_k\| = \alpha_k \|d_k\| \ge c_4 \|F(x_k)\| \ge \mu c_4 \|x_k - \bar{x}\|.$$

Combining the above inequality with (32) gives

$$\|x_{k+1} - \bar{x}\|^2 \le \|x_k - \bar{x}\|^2 - c_1 \mu^4 c_4^4 \|x_k - \bar{x}\|^4,$$

Then, let $\omega_2 = c_1 \mu^4 c_4^4$, and by Lemma 3.4 again, we get

$$\|x_k - \bar{x}\| \le \frac{\|x_0 - \bar{x}\|}{\sqrt{1 + \omega_2 k \|x_0 - \bar{x}\|^2}} = \frac{1}{\sqrt{\omega_2 k + \|x_0 - \bar{x}\|^{-2}}}.$$

The proof is completed.

4 Numerical results

This section reports some numerical results to evaluate the performance of the proposed method on the least absolute shrinkage and selection operator (denoted by LASSO) problem [35], which can be formulated as an ℓ_1-norm minimization problem:

$$\min_x \frac{1}{2}\|Ax-b\|^2 + \rho\|x\|_1, \tag{33}$$

where $x \in \mathcal{R}^n, b \in \mathcal{R}^m, A \in \mathcal{R}^{m\times n}$, and ρ is a nonnegative parameter. In the following, the term $\frac{1}{2}\|Ax-b\|^2$ is denoted by $f(x)$. For $i = 1, 2, \dots, n$, the optimality conditions of problem (33) is

$$\begin{aligned} &\nabla_i f(x) + \rho \text{sign}(x_i) = 0, \quad |x_i| > 0, \\ &|\nabla_i f(x)| \le \rho, \quad |x_i| = 0, \end{aligned}$$

where sign(t) is the subdifferential of the absolute value function $|t|$ given by the signum function, that is

$$\partial|t| = \text{sign}(t) := \begin{cases} \{-1\}, & \text{if } \ t < 0, \\ [-1, 1], & \text{if } \ t = 0, \\ \{1\}, & \text{if } \ t > 0. \end{cases}$$

Then, the optimality conditions of problem (33) can be further written as a system of nonsmooth equations as follows [36]:

$$H^\tau(x) = 0, \tag{34}$$

where $H^\tau = (H_1^\tau, H_2^\tau, \dots, H_n^\tau)^\top : \mathcal{R}^n \to \mathcal{R}^n$, and

$$H_i^\tau(x) = \max\{\tau(\nabla_i f(x) - \rho), \min\{x_i, \tau(\nabla_i f(x) + \rho)\}\},$$

and $\tau \in (0, \tau^*]$ is a constant, $\tau^* = \min_i\{1/D_{ii}\}$ and D_{ii} is the ith diagonal element of $A^\top A$. Note that, Xiao et al. [4] also proposed a nonsmooth equation-based reformulation of problem (33), whose dimension is twice of the dimension of the nonsmooth equations (34).

In what follows, we will perform a sparse signal recovery experiment to demonstrate the efficiency of the DFCGPM, and give some comparisons with the related methods, including the spectral gradient projection method in [12] (denoted by SGPM) and the conjugate gradient method in [4] (denoted by CGM). The parameters in the three tested methods are listed as follows:

SGPM: $r = 10, \rho = 0.35, \sigma = 0.01, \beta = 1$.

CGM: $\rho = 0.35, \sigma = 0.01, \beta = 1$.

DFCGPM: $\rho = 0.35, \sigma = 0.01, \beta = 1, \gamma = 1.9$, $v_k = y_{k-1} - \bar{s}_{k-1}$, $\tau = \tau^*$ and

$$\beta_k = \beta_k^{\text{MP}} := \frac{(y_{k-1} - \bar{s}_{k-1})^\top F(x_k)}{d_{k-1}^\top w_{k-1}},$$

with

$$w_{k-1} = y_{k-1} + \bar{\gamma}\bar{s}_{k-1}, \bar{\gamma} > 0, y_{k-1} = F(z_{k-1}) - F(x_{k-1}), \bar{s}_{k-1} = z_{k-1} - x_{k-1} = \alpha_{k-1}d_{k-1}.$$

Therefore, the direction d_k is defined by

$$d_k = \begin{cases} -F(x_k), & \text{if } \ k = 0, \\ -F(x_k) + \beta_k^{\mathrm{MP}} d_{k-1} - \frac{F(x_k)^\top d_{k-1}}{d_{k-1}^\top w_{k-1}}(y_{k-1} - \bar{s}_{k-1}), & \text{if } \ k \geq 1. \end{cases}$$

Furthermore, the vector v_k and the parameter β_k satisfy the condition (13). In fact, it follows from Assumption 2.1 and the inequality (3.6) in [22] that

$$\begin{aligned} &\frac{|\beta_k^{\mathrm{MP}}|\|d_{k-1}\|\|y_{k-1} - \bar{s}_{k-1}\|}{|F(x_k)^\top (y_{k-1} - \bar{s}_{k-1})|} \\ &= \frac{\|d_{k-1}\|\|y_{k-1} - \bar{s}_{k-1}\|}{|d_{k-1}^\top w_{k-1}|} \leq \frac{\|d_{k-1}\|\|y_{k-1} - \bar{s}_{k-1}\|}{\bar{\gamma}\alpha_{k-1}\|d_{k-1}\|^2} \leq \frac{\|y_{k-1}\| + \|\bar{s}_{k-1}\|}{\bar{\gamma}\alpha_{k-1}\|d_{k-1}\|} \leq \frac{L\|\bar{s}_{k-1}\| + \|\bar{s}_{k-1}\|}{\bar{\gamma}\|\bar{s}_{k-1}\|} = \frac{L+1}{\bar{\gamma}}. \end{aligned}$$

In the experiment, we set $\bar{\gamma} = 10$. We stop the iteration if the following condition

$$\|H^\tau(x_k)\| \leq 10^{-6} \ \text{ or } \ \|H^\tau(z_k)\| \leq 10^{-6}$$

is satisfied or the number of iterations exceeds 1000. We consider a typical LASSO scenario and generate the synthetic data of problem (33) in the same way as [37]. More specifically, the original signal $\bar{x} \in \mathcal{R}^n$ contains k randomly placed ± 1 spikes. The matrix $A \in \mathcal{R}^{m\times n}$ is the Gaussian matrix whose elements are generated from *shape i.i.d.* normal distributions $\mathcal{N}(0,1)$ (family `randn(m,n)` in Matlab) then orthonormalizing the rows. Due to the storage limitations of PC, we test a small size signal with $n = 2^{12}, m = 2^{11}$, and the original contains $k = 120, 140, 160$ randomly non-zero elements. The restoration accuracy is measured by means of the mean-squared-error (MSE)$=\frac{1}{n}\|\bar{x} - x^*\|^2$, where x^* is the restored signal. The process is started with the initial signal $x_0 = A^\top b$, and the parameter ρ in problem (33) is set as $0.01\|b\|_\infty$.

All the numerical experiments are performed on an Thinkpad laptop with Intel Core 2 CPU 2.10 GHZ and RAM 4.00 GM. All the programs are written in Matlab R2014a. In the experiment, we consider the following two cases.

Case 1 The vector $b = A\bar{x}$, so no noise is assumed.

Case 2 The vector $b = A\bar{x} + w$ with white Gaussian noise of variance 10^{-4}.

Table 1 and Table 2 give the mean squared error of the SGPM of Yu et al. [12], the CGM of Xiao [4], and the DFCGPM as well as the number of iterations and the execution times.

From the numerical results in Table 1 and Table 2, we see that the DFCGPM is less time consuming that those of the SGPM [12] and the CGM [4], and thus it performs better than the latter two methods, which implies that it is computationally efficient for solving the LASSO problem.

To better understanding the performance of the three tested methods, the original signal, observed signal and recovered signals by the three tested methods are plotted in Figure 1 and Figure 2 for problem (33) with free noise or white noise.

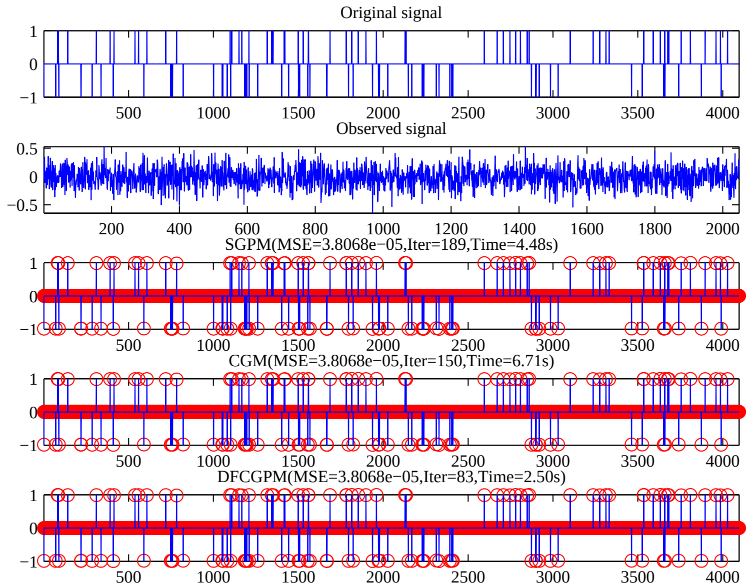

Figure 1: Problem (33) with $k = 120, m = 2048, n = 4096$ and free noise.

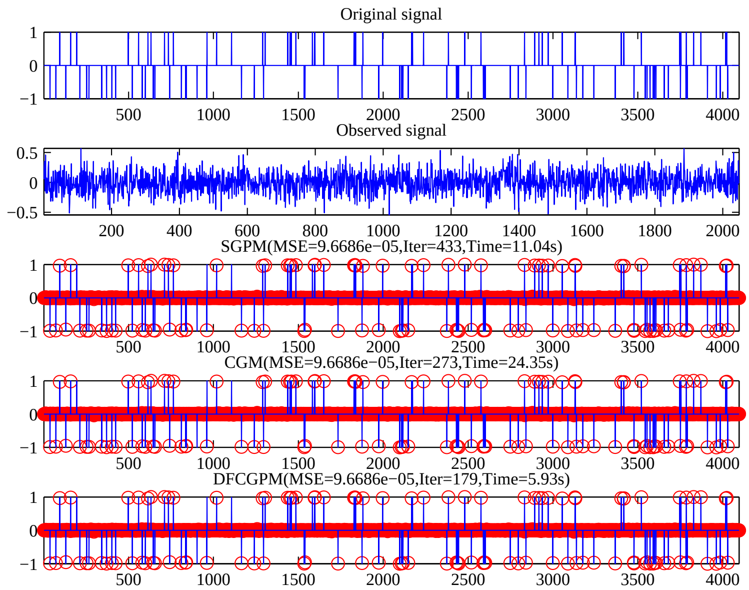

Figure 2: Problem (33) with $k = 120, m = 2048, n = 4096$ and white noise.

Table 1: Numerical results of problem (33) with $m = 2048, n = 4096$ and free noise

k-sparse signal	Methods	MSE	Iter	Times (s)
k=120	SGPM	3.7268e-05	191	4.18
	CGM	3.7268e-05	148	7.39
	DFCGPM	3.7268e-05	83	2.57
k=140	SGPM	4.4714e-05	213	4.80
	CGM	4.4714e-05	157	7.01
	DFCGPM	4.4714e-05	89	2.65
k=160	SGPM	4.9841e-05	208	5.51
	CGM	4.9841e-05	158	8.68
	DFCGPM	4.9841e-05	89	3.15

From the last three subplots in Figure 1 and Figure 2, the three tested methods all successfully recover the original signal. To further observe the convergence of the three tested methods, in Figure 3 and Figure 4 we visualize the evolution of convergence when the three methods are applied to solve problem (33).

From Figure 3 and Figure 4, we observe that the DFCGPM converges faster than those of the SGPM [12] and the CGM [4] before they reach the stable stage.

5 Conclusion

In this paper, we have proposed a class of derivative-free CG projection method for nonsmooth equations with convex constraints. Under the condition that the underlying mapping is monotone and Lipschitz continuous, we have established its global convergence, and under additional local error bound assumption, we have proved its convergence rate. Preliminary numerical results about the LASSO problem indicate that the proposed method is more efficient than some state-of-the-art solvers.

Funding

This research was partially supported by the National Natural Science Foundation of China and Shandong Province (No. 11671228, 11601475, ZR2016AL05).

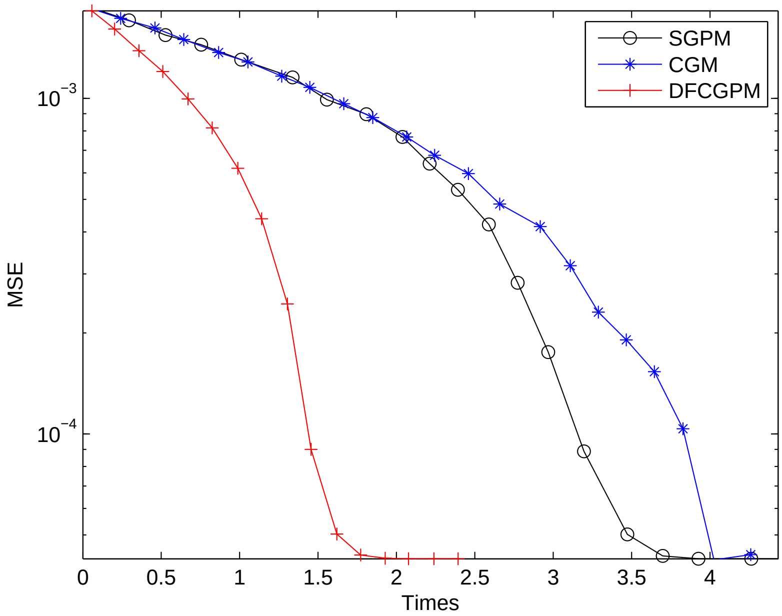

Figure 3: MSE of times for problem (33) with $k = 120, m = 2048, n = 4096$ and free noise.

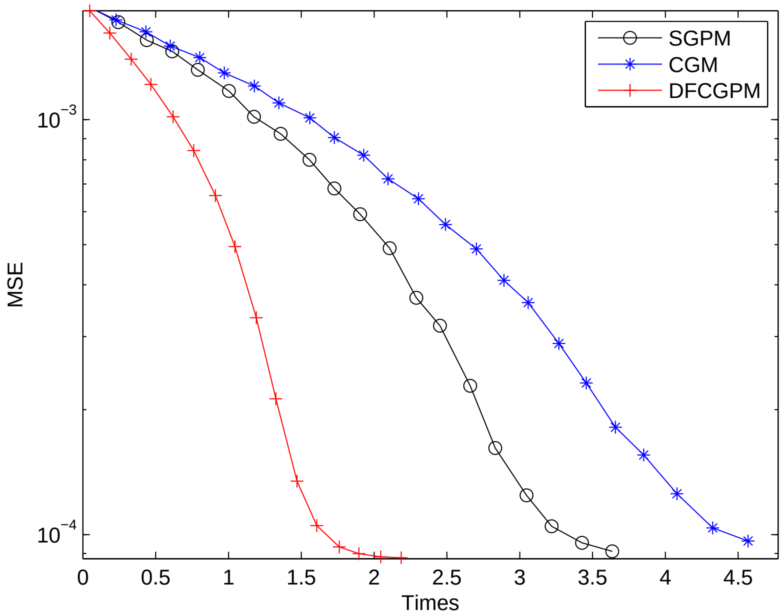

Figure 4: MSE of times for problem (33) with $k = 120, m = 2048, n = 4096$ and white noise.

Table 2: Numerical results of problem (33) with $m = 2048, n = 4096$ and white noise

k-sparse signal	Methods	MSE	Iter	Times (s)
k=120	SGPM	9.3172e-05	357	8.05
	CGM	9.3172e-05	208	14.18
	DFCGPM	9.3172e-05	149	4.26
k=140	SGPM	1.0448e-04	418	9.95
	CGM	1.0448e-04	309	34.36
	DFCGPM	1.0448e-04	171	5.47
k=160	SGPM	1.0229e-04	415	11.09
	CGM	1.0229e-04	283	30.29
	DFCGPM	1.0229e-04	169	6.32

References

[1] Zeidler E., Nonlinear functional analysis and its applications. Springer-Verlag, 1990.

[2] Wood A.J., Wollenberg B.F., Power generation, operation, and control. Wiley, New York, 1996.

[3] Meintjes K., Morgan A.P, A methodology for solving chemical equilibrium systems, Applied Mathematics and Computation, 1987, 22: 333-361.

[4] Xiao Y.H., Zhu H., A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing, Journal of Mathematical Analysis and Applications, 2013, 405: 310-319.

[5] Sun M., Liu J., The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization, Journal of Inequalities and Applications, December 2017, 2017: 19.

[6] Zhou B., Duam G.R., Lin Z., A parametric periodic Lyapunov equation with application in semi-global stabilization of discrete-time periodic systems subject to actuator saturation, Automatica, 2011, 47: 316-325.

[7] La Cruz W., Raydan M., Nonmonotone spectral methods for large-scale nonlinear systems, Optimization Methods and Software, 2003, 18: 583-599.

[8] La Cruz W., Martìnez J.M., Raydan M., Spectral residual method without gradient information for solving large- scale nonlinear systems of equations, Mathematics of Computation, 2006, 75: 1429-1448.

[9] Zhang L., Zhou W.J., Spectral gradient projection method for solving nonlinear monotone equations, Journal of Computational and Applied Mathematics, 2006, 196: 478-484.

[10] Barzilai J., Borwein J.M., Two-point step size gradient methods, IMA Journal of Numerical Analysis, 1988, 8: 141148. (1988)

[11] Solodov M.V., Svaiter B.F., A globally convergent inexact Newton method for system of monotone equations. In: Fukushima, M., Qi, L. (eds.) Reformulation: Nonsmooth, Piecewise Smooth, Semi- smooth and Smoothing Methods, pp. 355369. Kluwer Academic Publishers, Dordrecht (1999).

[12] Yu Z.S., Lin J., Sun J., Xiao Y.H., Liu L.Y., Li Z.H., Spectral gradient projectionmethod formonotone nonlinear equations with convex constraints, Applied Numerical Mathematics, 2009, 59: 2416-2423.

[13] Yu G.H., Niu S.Z., Ma J.H., Multivariate spectral gradient projectionmethod for nonlinear monotone equations with convex constraints, Journal of Industrial and Management Optimization, 2013, 9: 117-129.

[14] Liu J., Duan Y.R., Two spectral gradient projection methods for constrained equations and their linear convergence rate, Journal of Inequalities and Applications, 2015, 2015: 8.

[15] Cheng W.Y.. A PRP type method for systems of monotone equations, Mathematical and Computer Modelling, 2009, 50: 15-20.

[16] Li Q.N., Li D.H., A class of derivative-free methods for large-scale nonlinear monotone equations, IMA Journal of Numerical Analysis, 2011, 31: 1625-1635.

[17] Li D.H., Wang X.L., A modified Fletcher-Reeves-type derivative-free method for symmetric nonlinear equations, Numerical Algebra Control and Optimization, 2012, 1(1): 71-82.

[18] Sun X.L., A two-term Fletcher-Reeves conjugate gradient method for monotone constrained equations and its application in compressive sensing, ICIC Express Letters, 2015, 9(11): 2987-2992.

[19] Sun M., Liu J., A modified Hestenes-Stiefel projection method for constrained nonlinear equations and its linear convergence rate, Journal of Applied Mathematics and Computating, 2015, 49(1-2): 145-156.

[20] Sun M., Liu J., New hybrid conjugate gradient projection method for the convex constrained equations, Calcolo, 2016, 53: 399-411.

[21] Yuan G.L., Zhang M.J., A three-terms Polak-Ribière-Polyak conjugate gradient algorithm for large-scale nonlinear equations, Journal of Computational and Applied Mathematics, 2015, 286: 186-195.

[22] Dai Z.F., Chen X.H., Wen F.H., A modified Perrys conjugate gradient method-based derivative-free method for solving large-scale nonlinear monotone equations, Applied Mathematics and Computation, 2015, 270(1): 378-386.

[23] Liu J.K., Li S.J., Multivariate spectral DY-type projection method for convex constrained nonlinear monotone equations, Journal of Industrial and Management Optimization, 2017, 13(1): 283-295.

[24] Ou Y.G., Li J.Y., A new derivative-free SCG-type projection method for nonlinear monotone equations with convex constraints, Journal of Applied Mathematics and Computing, 2018, 56(12): 195-216.

[25] Cheng W.Y., A two-term PRP-based descent method, Numerical Functional Analysis and Optimization, 2007, 28: 1217-1230.

[26] Sun M., Bai Q.G., A new descent memory gradient method and its global convergence, Journal of Systems Science and Complexity, 2011 , 24(4): 784-794.

[27] Sun M., Liu J., Three derivative-free projection methods for nonlinear equations with convex constraints, Journal of Applied Mathematics and Computating, 2015, 47: 265-276.

[28] Feng D.X., Sun M., Wang X.Y., A family of conjugate gradient methods for large-scale nonlinear equations, Journal of Inequalities and Applications, 2017, 2017: 236.

[29] Zarantonello E.H. Projections on Convex Sets in Hilbert Space and Spectral Theory. Academic Press, New York (1971).

[30] Sun W.Y., Yuan Y.X., Optimization Theory and Methods : Nonlinear Programming. Springer Opti- mization and Its Applications, vol. 1. Springer, New York (2006).

[31] Zhang L., Zhou W.J., Li D.H., A descent modified Polak-Ribi‘ ere-Polyak conjugate gradient method and its global convergence, IMA Journal of Numerical Analysis, 2006, 26: 629640.

[32] Zhang L., Zhou W.J., Li D.H., Global convergence of a modified Fletcher-Reeves conjugate method with Armijo-type line search, Numerische Mathematik, 2006, 104: 561-572.

[33] Solodov M.V., Svaiter B.F., A globally convergent inexact Newton method for systems of monotone equations, in: M. Fukushima, L. Qi (Eds.), Reformulation: Nonsmooth, Piecewise smooth, Semismooth and Smoothing Methods, Kluwer Academic Publishers, 1998, pp. 335-369.

[34] Polyak B.T., Introduction to Optimization, Optimization Software Inc., Publications Division, New York, 1987 (Translated from Russian, with a foreword by Dimitri P. Bertsekas).

[35] Tibshirani R., Regression shrinkage and selection via the LASSO, Journal of the Royal Statistical Society Series B-Statistical Methodology, 1996, 58: 267-288.

[36] Wu L., Sun Z., New nonsmooth equations-based algorithms for ℓ_1-norm minimization and applications, Journal of Applied Mathematics, 2012, Article ID 139609, 14 pages.

[37] 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, 2007, 1(4): 586-597.