

# A projected gradient method for $\alpha\ell_1 - \beta\ell_2$ sparsity regularization

Liang Ding<sup>1</sup> and Weimin Han<sup>2</sup>

**Abstract.** The non-convex  $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$  ( $\alpha \geq \beta \geq 0$ ) regularization has attracted attention in the field of sparse recovery. One way to obtain a minimizer of this regularization is the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm which is similar to the classical iterative soft thresholding algorithm (ISTA). It is known that ISTA converges quite slowly, and a faster alternative to ISTA is the projected gradient (PG) method. However, the conventional PG method is limited to the classical  $\ell_1$  sparsity regularization. In this paper, we present two accelerated alternatives to the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm by extending the PG method to the non-convex  $\alpha\ell_1 - \beta\ell_2$  sparsity regularization. Moreover, we discuss a strategy to determine the radius  $R$  of the  $\ell_1$ -ball constraint by Morozov's discrepancy principle. Numerical results are reported to illustrate the efficiency of the proposed approach.

**Keywords.** projected gradient method,  $\alpha\ell_1 - \beta\ell_2$  sparsity regularization, non-convex sparsity regularization, Morozov's discrepancy principle

## 1 Introduction

In this paper, we are interested in solving an ill-posed operator equation of the form

$$Ax = y, \tag{1.1}$$

where  $x$  is sparse,  $A : \ell_2 \rightarrow Y$  is a linear and bounded operator mapping between the  $\ell_2$  space and a Banach space  $Y$  with norms  $\|\cdot\|_{\ell_2}$  and  $\|\cdot\|_Y$ , respectively. In practice, the right-hand side  $y$  is known only approximately with an error up to a level  $\delta \geq 0$ . Therefore, we assume that we know  $\delta \geq 0$  and  $y^\delta \in Y$  with  $\|y^\delta - y\|_Y \leq \delta$ . The most commonly adopted technique to solve problem (1.1) is the  $\ell_p$ -norm sparsity regularization with  $1 \leq p < 2$ , see the monographs [18, 39] and the special issues [4, 13, 24, 25] for many developments on regularizing properties and minimization schemes. Since the  $\ell_p$ -norm regularization with  $1 \leq p < 2$  does not always provide the sparsest solution, the non-convex  $\ell_p$ -norm sparsity regularization with  $0 \leq p < 1$  was

---

<sup>1</sup>Department of Mathematics, Northeast Forestry University, Harbin 150040, China; e-mail: [d1@nefu.edu.cn](mailto:d1@nefu.edu.cn). The work of this author was supported by the Fundamental Research Funds for the Central Universities (no. 2572018BC02), Heilongjiang Postdoctoral Research Developmental Fund (no. LBH-Q16008), the National Nature Science Foundation of China (no. 41304093).

<sup>2</sup>Department of Mathematics, University of Iowa, Iowa City, IA 52242, USA; e-mail: [weimin-han@uiowa.edu](mailto:weimin-han@uiowa.edu).

proposed as alternatives. For the  $\ell_0$  sparsity regularization, see [6, 7, 9, 20] for the iterative hard thresholding algorithm. We refer the reader to [23, 26, 31] for some other types of alternatives to the  $\ell_0$ -norm.

The investigation of the non-convex  $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$  ( $\alpha \geq \beta \geq 0$ ) regularization has attracted attention in the field of sparse recovery over the last five years, see [15, 27, 30, 46, 47] and references therein. In [15], we investigated the well-posedness and convergence rate of the non-convex  $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$  ( $\alpha \geq \beta \geq 0$ ) sparsity regularization of the form

$$\min \mathcal{J}_{\alpha,\beta}^\delta(x) = \frac{1}{q} \|Ax - y^\delta\|_Y^q + \mathcal{R}_{\alpha,\beta}(x) \quad (1.2)$$

in the  $\ell_2$  space, where

$$\mathcal{R}_{\alpha,\beta}(x) := \alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2}, \quad \alpha \geq \beta \geq 0, \quad q \geq 1.$$

Denoting  $\eta = \beta/\alpha$ , we can equivalently express the function  $\mathcal{J}_{\alpha,\beta}^\delta(x)$  in (1.2) as

$$\frac{1}{q} \|Ax - y^\delta\|_Y^q + \alpha \mathcal{R}_\eta(x),$$

where

$$\mathcal{R}_\eta(x) := \|x\|_{\ell_1} - \eta\|x\|_{\ell_2}, \quad \alpha > 0, \quad 1 \geq \eta \geq 0.$$

For the particular case  $q = 2$ , we provided an ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm of the form

$$z^k = \mathbb{S}_{\frac{\alpha}{\lambda}} \left( \left( \frac{\beta}{\lambda\|x^k\|_{\ell_2}} + 1 \right) x^k - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right), \quad x^{k+1} = x^k + s^k(z^k - x^k) \quad (1.3)$$

for (1.2), where  $s^k$  is the step size and  $\lambda > 0$ . Obviously, the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm is similar to the classical ISTA when the step size  $s^k = 1$ . In [12], an ISTA of the form

$$x^{k+1} = \mathbb{S}_\alpha (x^k - A^*(Ax^k - y^\delta)) \quad (1.4)$$

was first proposed to solve the classical  $\ell_1$  sparsity regularization of the form

$$\min \mathcal{J}_\alpha^\delta(x) = \frac{1}{2} \|Ax - y^\delta\|_Y^2 + \alpha\|x\|_{\ell_1}. \quad (1.5)$$

As an alternative of the  $\ell_p$ -norm with  $0 \leq p < 1$ , the function  $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$  ( $\alpha \geq \beta \geq 0$ ) has the desired property that it is a good approximation of a multiple of the  $\ell_0$ -norm. The function has a simpler structure than the  $\ell_0$ -norm from the perspective of computation. The ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm can easily be implemented, see [15, 21, 47] for several other algorithms for  $\|\cdot\|_{\ell_1} - \|\cdot\|_{\ell_2}$  sparsity regularization. However, the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm, in general, can be arbitrarily slow and it is computationally intensive. So it is desirable to develop accelerated versions of the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm, especially for large-scale ill-posed inverse problems.

## 1.1 Some accelerated algorithms for ISTA

Searching for accelerated algorithms of the ISTA has become popular and some faster algorithms have been proposed. In [5, 14, 17, 45], several accelerated projected gradient methods have been provided. A comparison among several accelerated algorithms is provided in [28], including “fast ISTA” ([2]). Applying a smoothing technique from Nesterov ([32]), a fast and accurate first-order method is proposed for solving large-scale compressed sensing problems ([3]). In [11], a simple heuristic adaptive restart technique is introduced, which can dramatically improve the convergence rate of accelerated gradient schemes. In [10], convergence of the iterates of the Fast Iterative Shrinkage/Thresholding Algorithm is established. In [33], a new iterative regularization procedure for inverse problems based on the use of Bregman distances is studied. Numerical results show that the proposed method gives significant improvement over the standard method. An explicit algorithm based on a primal-dual approach for the minimization of an  $\ell_1$ -penalized least-squares function, with a non-separable  $\ell_1$  term, is proposed in [29]. An iteratively reweighted least squares algorithm and the corresponding convergence analysis for the regularization of linear inverse problems with sparsity constraints are investigated in [19]. For a projected gradient method of nonlinear ill-posed problems, see [40].

Unfortunately, the algorithms stated above are only limited to the classical  $\ell_1$ -norm sparsity regularization. Though there is great potential for accelerated algorithms in sparsity regularization with a non-convex penalty term, to the best of our knowledge, little work can be found in the literature. In [35], the authors treat the problem of minimizing a general continuously differentiable function subject to  $\|x\|_0 \leq s$ , where  $s > 0$  is an integer, and  $\|x\|_0$  is the  $\ell_0$ -norm of  $x$ , which counts the number of nonzero components in  $x$ . In this paper, we extend the projected gradient method to the non-convex  $\alpha\ell_1 - \beta\ell_2$  sparsity regularization. There are two reasons why we choose PG method. First, its formulation is simple and it can easily be implemented. Another reason is that it converges quite fast. So it is adequate for solving large-scale ill-posed problems.

The PG method was introduced in [14] to accelerate the ISTA. It is shown that the ISTA converges initially relatively fast, then it overshoots the  $\ell_1$ -norm penalty, and it takes many steps to re-correct back. It means that the algorithm generates a path  $\{x_n \mid n \in \mathbb{N}\}$  that is initially fully contained in the  $\ell_1$ -ball  $B_R := \{x \in \ell_2 \mid \|x\|_{\ell_1} \leq R\}$ . Then it gets out of the ball to slowly inch back to it in the limit. To avoid this long external detour, the authors of [14] proposed an accelerated algorithm by substituting the soft thresholding operation  $\mathbb{S}_\alpha$  by the projection  $\mathbb{P}_R$  which is defined in Definition 2.5. This leads to a projected gradient method of the form

$$x^{k+1} = \mathbb{P}_R(x^k - \gamma^k A^*(Ax^k - y^\delta)). \quad (1.6)$$

## 1.2 Contribution and organization

Since the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm (1.3) is similar to ISTA (1.4), inspired by [14], we propose two accelerated alternatives to (1.3) by extending the PG method to solve (1.2).

The first accelerated algorithm is based on the generalized conditional gradient method (GCGM).

In [15], based on GCGM, we proposed the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm where the crucial issue is to determine  $z^k$  by the optimization problem of the form

$$\min_z \langle A^*(Ax^k - y^\delta) - \lambda x^k - \frac{\beta x^k}{\|x^k\|_{\ell_2}}, z \rangle + \frac{\lambda}{2} \|z\|_{\ell_2}^2 + \alpha \|z\|_{\ell_1}. \quad (1.7)$$

In this paper, we show that the problem (1.7) can be solved by a PG method of the form

$$z^k = \mathbb{P}_R \left( x^k + \frac{\beta x^k}{\lambda \|x^k\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right). \quad (1.8)$$

With  $z^k$  at our disposal, we compute  $x^{k+1}$  by  $x^{k+1} = x^k + s^k(z^k - x^k)$ , where  $s^k$  is the step size.

Theoretically, the radius  $R$  of  $\ell_1$ -ball should be chosen by  $R = \|x_{\alpha,\beta}^\delta\|_{\ell_1}$  ([14]), where  $x_{\alpha,\beta}^\delta$  is a minimizer of (1.2). However, in general, one can not obtain the value of  $\|x_{\alpha,\beta}^\delta\|_{\ell_1}$  before starting the iteration (1.8). In this paper, we utilize Morozov's discrepancy principle to determine  $R$ . This method only requires knowledge of the noise level  $\delta$  and the observed data  $y^\delta$ . Moreover, we investigate the well-posedness of (1.2) under Morozov's discrepancy principle.

The second accelerated algorithm is based on the surrogate function approach. We investigate this algorithm in the finite dimensional space  $\mathbb{R}^n$ . For the case  $q = 2$ , (1.2) takes the form

$$\min \mathcal{J}_{\alpha,\beta}^\delta(x) = \frac{1}{2} \|Ax - y^\delta\|_{\ell_2}^2 + \alpha \|x\|_{\ell_1} - \beta \|x\|_{\ell_2}, \quad (1.9)$$

where  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear and bounded operator mapping between the  $\mathbb{R}^n$  and  $\mathbb{R}^m$  space with  $\|\cdot\|_{\ell_2}$  norms. In the following, we remove the  $\ell_1$  constraint in (1.9) and to consider a constrained optimization problem for a certain radius  $R$  of  $\ell_1$ -ball constraint. So, in analogy to the techniques about projection in [14, 41], a natural strategy is to consider the constrained optimization problem of the form

$$\min \frac{1}{2} \|Ax - y^\delta\|_{\ell_2}^2 \quad \text{subject to } x \in B'_R := \{x \in \mathbb{R}^n \mid \|x\|_{\ell_1} - \eta \|x\|_{\ell_2} \leq R\}, \quad 1 \geq \eta \geq 0. \quad (1.10)$$

However, since  $B'_R$  is non-convex, it is challenge to analyze and solve this constrained optimization problem. To utilize the theory of convex constraints, we remove the  $\ell_1$  constraint in (1.9) and to consider instead the following optimization problem of the form

$$\min \mathcal{D}_\beta^\delta(x) = \frac{1}{2} \|Ax - y^\delta\|_{\ell_2}^2 - \beta \|x\|_{\ell_2} \quad \text{subject to } x \in B_R := \{x \in \mathbb{R}^n \mid \|x\|_{\ell_1} \leq R\} \quad (1.11)$$

for a suitable  $R$ . We propose a projected gradient method of the form

$$x^{k+1} = \mathbb{P}_R \left( x^k + \frac{\beta x^{k+1}}{\lambda \|x^{k+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right) \quad (1.12)$$

for (1.11), where  $\lambda > 0$  satisfies some conditions, see Assumption 4.6.

An outline of the rest of this paper is as follows. In the next section we introduce the notation and review results of the Tikhonov regularization and the PG method. In Section 3, we investigate an accelerated algorithm via GCGM. Furthermore, we give a strategy to determine the radius  $R$  of  $\ell_1$ -ball constraint. In Section 4, we propose another accelerated algorithm via the surrogate function approach. Finally, we present results from numerical experiments on compressive sensing and image deblurring problems in Section 5.

## 2 Preliminaries

Before starting the discussion on the accelerated algorithms, we briefly introduce some notation and results of the Tikhonov regularization and the PG method. Let

$$x_{\alpha,\beta}^\delta = \arg \min_x \left\{ \frac{1}{2} \|Ax - y^\delta\|_Y^2 + \mathcal{R}_{\alpha,\beta}(x) \right\} \quad (2.1)$$

be a minimizer of the regularization function  $\mathcal{J}_{\alpha,\beta}^\delta(x)$  in (1.2) with  $q = 2$  for every  $\alpha \geq \beta \geq 0$ . We denote by  $\mathcal{L}_{\alpha,\beta}^\delta$  the set of all minimizers  $x_{\alpha,\beta}^\delta$ , and by  $x_{R,\beta}^\delta$  a solution of (1.11). We use the following definition of  $\mathcal{R}_\eta$ -minimum solution ([15]).

**Definition 2.1** *An element  $x^\dagger \in \ell_2$  is called an  $\mathcal{R}_\eta$ -minimum solution of the linear problem  $Ax = y$  if*

$$Ax^\dagger = y \quad \text{and} \quad \mathcal{R}_\eta(x^\dagger) = \min_x \{ \mathcal{R}_\eta(x) \mid Ax = y \}.$$

We recall the definition of sparsity ([12]).

**Definition 2.2** *An element  $x \in \ell_2$  is called sparse if  $\text{supp}(x) := \{i \in \mathbb{N} \mid x_i \neq 0\}$  is finite, where  $x_i$  is the  $i^{\text{th}}$  component of  $x$ .  $\|x\|_0 := \text{supp}(x)$  is the cardinality of  $\text{supp}(x)$ . If  $\|x\|_0 = s$  for some  $s \in \mathbb{N}$ , then  $x \in \ell_2$  is called  $s$ -sparse.*

**Definition 2.3** (Morozov's discrepancy principle) *For  $1 < \tau_1 \leq \tau_2$ , we choose  $\alpha = \alpha(\delta, y^\delta) > 0$  such that*

$$\tau_1 \delta \leq \|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y \leq \tau_2 \delta \quad (2.2)$$

*holds for some  $x_{\alpha,\beta}^\delta$ .*

Next we recall definitions of the soft thresholding and the projection operators ([5, 12]).

**Definition 2.4** *For a given  $\alpha > 0$ , the soft thresholding operator is defined as*

$$\mathbb{S}_\alpha(x) = \sum_i S_\alpha(x_i) e_i,$$

*where  $e_i = \underbrace{(0, \dots, 0, 1, 0, \dots)}_i$ ,  $x_i$  is the  $i^{\text{th}}$  component of  $x$  and*

$$S_\alpha(t) = \begin{cases} t - \alpha & \text{if } t \geq \alpha, \\ 0 & \text{if } |t| < \alpha, \\ t + \alpha & \text{if } t \leq -\alpha. \end{cases}$$

**Definition 2.5** *The projection onto the  $\ell_1$ -ball is defined by*

$$\mathbb{P}_R(\hat{x}) := \{\arg \min_x \|x - \hat{x}\|_{\ell_2} \text{ subject to } \|x\|_{\ell_1} \leq R\},$$

*which gives the projection of an element  $\hat{x}$  onto the  $\ell_1$ -norm ball with radius  $R > 0$ .*

Then we review two results from [14] on relations between the soft thresholding operator and the projection operator. For relations between the parameters  $\alpha$  and  $R$ , see [14, Fig. 2].

**Lemma 2.6** *For some countable index set  $\Lambda$ , denote  $\ell_p = \ell_p(\Lambda)$ ,  $1 \leq p < \infty$ . For any fixed  $a \in \ell_2(\Lambda)$  and for  $\alpha > 0$ ,  $\|\mathbb{S}_\alpha(a)\|_{\ell_1}$  is a piecewise linear, continuous, decreasing function of  $\alpha$ . Moreover, if  $a \in \ell_1(\Lambda)$  then  $\|\mathbb{S}_0(a)\|_{\ell_1} = \|a\|_{\ell_1}$  and  $\|\mathbb{S}_\alpha(a)\|_{\ell_1} = 0$  for  $\alpha \geq \max_i |a_i|$ .*

**Lemma 2.7** *If  $\|a\|_{\ell_1} > R$ , then the  $\ell_2$  projection of  $a$  on the  $\ell_1$ -ball with radius  $R$  is given by  $\mathbb{P}_R(a) = \mathbb{S}_\alpha(a)$ , where  $\alpha$  (depending on  $a$  and  $R$ ) is chosen such that  $\|\mathbb{S}_\alpha(a)\|_{\ell_1} = R$ . If  $\|a\|_{\ell_1} \leq R$  then  $\mathbb{P}_R(a) = \mathbb{S}_0(a) = a$ .*

Finally, recall the following properties of  $\mathbb{P}_R(x)$  ([14]).

**Lemma 2.8** *Let  $H$  be a Hilbert space with the inner product  $\langle \cdot, \cdot \rangle$  and norm  $\|\cdot\|_H$ . For any  $x \in H$ ,  $\mathbb{P}_R(x)$  is characterized as the unique vector in  $H$  such that*

$$\langle w - \mathbb{P}_R(x), x - \mathbb{P}_R(x) \rangle \leq 0 \quad \forall w \in B_R.$$

*Moreover, the projection  $\mathbb{P}_R$  is non-expansive:*

$$\|\mathbb{P}_R(x') - \mathbb{P}_R(x'')\|_H \leq \|x' - x''\|_H \quad \forall x', x'' \in H.$$

### 3 The projected gradient method via GCGM

In [15], we proposed an ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm for (1.2) based on GCGM. We rewrite  $\mathcal{J}_{\alpha,\beta}^\delta(x)$  in (1.2) as

$$\mathcal{J}_{\alpha,\beta}^\delta(x) = F(x) + \Phi(x),$$

where

$$\begin{aligned} F(x) &= \frac{1}{2} \|Ax - y^\delta\|_Y^2 - \Theta(x), \\ \Phi(x) &= \Theta(x) + \alpha \|x\|_{\ell_1} - \beta \|x\|_{\ell_2}, \\ \Theta(x) &= \frac{\lambda}{2} \|x\|_{\ell_2}^2 + \beta \|x\|_{\ell_2}, \quad \lambda > 0. \end{aligned}$$

The ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm is stated in the form of Algorithm 1. Convergence of Algorithm 1 is given in Theorem 3.1; see [15, Theorem 3.5] for its proof.

---

**Algorithm 1** ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm for problem (1.2) with  $q = 2$

---

Set  $k = 0$ ,  $x^0 \in \ell_2$  such that  $\Phi(x^0) < +\infty$ ,

for  $k = 0, 1, 2, \dots$ , do

if  $x^k = 0$  then

$$x^{k+1} = \arg \min \frac{1}{2} \|Ax - y^\delta\|_Y^2 + \alpha \|x\|_{\ell_1}$$

else

determine a descent direction  $z^k$  as a solution of

$$\min_z \langle A^*(Ax^k - y^\delta) - \lambda x^k - \frac{\beta x^k}{\|x^k\|_{\ell_2}}, z \rangle + \frac{\lambda}{2} \|z\|_{\ell_2}^2 + \alpha \|z\|_{\ell_1}$$

determine a step size  $s^k$  as a solution of

$$\min_{s \in [0,1]} F(x^k + s(z^k - x^k)) + \Phi(x^k + s(z^k - x^k))$$

$$x^{k+1} = x^k + s^k(z^k - x^k)$$

end if

$$k = k + 1$$

end for

---

**Theorem 3.1** Let  $\{x^k\}$  denote the sequence generated by Algorithm 1. Then  $\{x^k\}$  contains a convergent subsequence and every convergent subsequence of  $\{x^k\}$  converges to a stationary point of the function  $\mathcal{J}_{\alpha,\beta}^\delta(x)$  in (1.2) with  $q = 2$ .

A crucial step in Algorithm 1 is the determination of  $z^k$  as a solution of

$$\min \mathcal{C}_{\alpha,\beta,\lambda}^\delta(z, x^k) = \langle A^*(Ax^k - y^\delta) - \lambda x^k - \frac{\beta x^k}{\|x^k\|_{\ell_2}}, z \rangle + \frac{\lambda}{2} \|z\|_{\ell_2}^2 + \alpha \|z\|_{\ell_1}. \quad (3.1)$$

In [15], we solve (3.1) by

$$z^k = \mathbb{S}_{\alpha/\lambda} \left( \left( \frac{\beta}{\lambda \|x^k\|_{\ell_2}} + 1 \right) x^k - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right). \quad (3.2)$$

However, (3.2) is known to converge quite slowly. To accelerate the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm, we transform (3.1) to an  $\ell_1$ -ball constraint optimization problem of the form

$$\begin{cases} \min \mathcal{D}_{\beta,\lambda}^\delta(z, x^k) = \langle A^*(Ax^k - y^\delta) - \lambda x^k - \frac{\beta x^k}{\|x^k\|_{\ell_2}}, z \rangle + \frac{\lambda}{2} \|z\|_{\ell_2}^2, & \beta \geq 0, \\ \text{subject to } \ell_1 \text{ ball } B_R := \{z \in \ell_2 \mid \|z\|_{\ell_1} \leq R\}. \end{cases} \quad (3.3)$$

Since  $\mathcal{C}_{\alpha,\beta,\lambda}^\delta(z, x^k)$ ,  $\mathcal{D}_{\beta,\lambda}^\delta(z, x^k)$  and  $B_R$  are convex with respect to the variable  $z$ , problem (3.3) is equivalent to (3.1) for a certain  $R$  ([36, Theorem 27.4], [48, Theorem 47.E]). In Lemma 3.2, we

show that the problem (3.3) can be solved by a PG method of the form

$$z^k = \mathbb{P}_R \left( x^k + \frac{\beta x^k}{\lambda \|x^k\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right). \quad (3.4)$$

**Lemma 3.2** *An element  $\hat{z} \in B_R$  is a minimizer of (3.3) if and only if*

$$\hat{z} = \mathbb{P}_R \left( x^k + \frac{\beta x^k}{\lambda \|x^k\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right) \quad (3.5)$$

for any  $\lambda > 0$ , which is equivalent to

$$\left\langle x^k + \frac{\beta x^k}{\lambda \|x^k\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) - \hat{z}, z - \hat{z} \right\rangle \leq 0 \quad \forall z \in B_R. \quad (3.6)$$

**Proof.** Note that  $\hat{z} \in B_R$  is a solution of (3.3) if and only if for any  $z \in B_R$ , the function  $f(t) = \mathcal{D}_{\beta, \lambda}^\delta((1-t)\hat{z} + tz, x^k)$  of  $t \in [0, 1]$  attains its minimum at  $t = 0$ . Since  $f(t)$  is quadratic and convex, a necessary and sufficient condition for  $f(0) = \min_{0 \leq t \leq 1} f(t)$  is  $f'(0+) \geq 0$ . Easily,

$$f'(0+) = \langle A^*(Ax^k - y^\delta) - \lambda x^k - \frac{\beta x^k}{\|x^k\|_{\ell_2}} + \lambda \hat{z}, z - \hat{z} \rangle,$$

and  $f'(0+) \geq 0$  is equivalent to (3.6). ■

The PG algorithm for (1.2) based on GCGM is stated in the form of Algorithm 2.

### 3.1 Determination of the radius $R$

From the previous discussion, we know that (3.1) is equivalent to (3.3) for a certain  $R$ . Before starting iteration (3.4), we need to choose an appropriate value of  $R$  which is crucial for the computation, especially in practical application. In this section, we give a strategy to determine the radius  $R$  of the  $\ell_1$ -ball constraint by Morozov's discrepancy principle.

By Lemma 2.7, for a given  $\alpha$  in (3.1),  $R$  in (3.3) should be chosen such that  $R = \|x_{\alpha, \beta}^\delta\|_{\ell_1}$ . However, one does not know the value of  $\|x_{\alpha, \beta}^\delta\|_{\ell_1}$  before starting (3.4). Of course, we can find an approximation of  $x_{\alpha, \beta}^\delta$  by the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm (1.3). Nevertheless, this implies that an additional soft thresholding iteration (1.3) is needed in Algorithm 2. Then the resulting algorithm is no longer an accelerated one.

So a crucial issue is how to check whether a value of  $R$  is appropriate for (3.3). Recall that there exists a regularization parameter  $\alpha$  depending on  $R$  such that (3.1) is equivalent to (3.3). So to determine an appropriate  $R$ , we need to check whether the corresponding regularization parameter  $\alpha$  is appropriate. One criterion is to check whether  $\delta^2 = O(\alpha)$ . If  $\delta^2 = O(\alpha)$ , then  $x_{\alpha, \beta}^\delta$  is a regularized solution ([15, Theorem 2.13]). However, by Lemmas 2.6 and 2.7, we only know that  $\alpha$  is a piecewise linear, continuous, decreasing function of  $R$  (see [14, Fig. 2]), and there is



---

**Algorithm 2** PG algorithm for problem (1.2) based on GCGM

---

Choose  $x^0 \in \ell_2$ ,  $\beta = O(\delta)$ ,  $\Phi(x^0) < +\infty$ ,

for  $k = 0, 1, 2, \dots$ , do

if  $x^k = 0$  then

$$x^{k+1} = \arg \min \frac{1}{2} \|Ax - y^\delta\|_Y^2 + \alpha \|x\|_{\ell_1}$$

else

determine  $z^k$  by

$$z^k = \mathbb{P}_R \left( x^k + \frac{\beta x^k}{\lambda \|x^k\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right)$$

determine a step size  $s^k$  as a solution of

$$\min_{s \in [0,1]} F(x^k + s(z^k - x^k)) + \Phi(x^k + s(z^k - x^k))$$

$$x^{k+1} = x^k + s^k(z^k - x^k)$$

end if

$k = k + 1$

end for

---

no explicit formula relating  $\alpha$  and  $R$ . We can not determine the value of  $\alpha$  from the value of  $R$  directly. So we can not ensure whether the  $R$  is appropriate.

Another criterion is Morozov's discrepancy principle. For any given  $R$ , we should check whether the regularization parameter  $\alpha$  satisfies Morozov's discrepancy principle (2.2), i.e.

$$\tau_1 \delta \leq \|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y \leq \tau_2 \delta, \quad 1 < \tau_1 \leq \tau_2.$$

For any fixed  $R$ , we need to compute  $x_{\alpha,\beta}^\delta$  by Algorithm 2 where  $z^k$  is determined by the PG method (3.4). Subsequently, we check whether  $x_{\alpha,\beta}^\delta$  satisfies (2.2). For this strategy, we only need to know the observed data  $y^\delta$  and the noise level  $\delta$ . By Lemma 3.5, the discrepancy  $\|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y$  is an increasing function of  $\alpha$ . A commonly adopted technique is to try  $\alpha_j = \alpha 2^{-j}$ ,  $j = 1, 2, \dots$ . With  $j$  increasing, one calculates  $x_{\alpha,\beta}^\delta$  until one finds  $\alpha = \inf\{\alpha > 0 \mid \tau_1 \delta \leq \|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y \leq \tau_2 \delta\}$  ([42]). Since  $\alpha$  is a decreasing function of  $R$ , the discrepancy  $\|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y$  is a decreasing function of  $R$ , see Lemma 2.6 and Fig. 1. Hence  $R := \sup\{R > 0 \mid \tau_1 \delta \leq \|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y \leq \tau_2 \delta\}$  is a reasonable choice. We begin with a small  $R$  such that  $x_{\alpha,\beta}^\delta$  satisfies Morozov's discrepancy principle (2.2). Subsequently, we increase the value of  $R$  to  $R + c$ ,  $c \in \mathbb{Z}^+$ , until  $x_{\alpha,\beta}^\delta$  fails to satisfy Morozov's discrepancy principle. Then we can find a maximal  $R$  which satisfies Morozov's discrepancy principle (2.2). Of course, we can also begin with a large  $R$  and gradually reduce the value of  $R$  until  $R$  satisfies Morozov's discrepancy principle (2.2). Under Morozov's discrepancy principle, the PG algorithm for (1.2) based on GCGM is stated in the form of Algorithm 3.

A natural question is whether (1.2) combined with Morozov's discrepancy principle is a reg-

---

**Algorithm 3** The PG algorithm for problem (1.2) based on GCGM

---

Choose  $x^0 \in \ell_2$ ,  $R_0 \in \mathbb{R}^+$ ,  $\beta = O(\delta)$ ,  $\Phi(x^0) < +\infty$ ,

for  $j = 0, 1, 2, \dots$  do

  for  $k = 0, 1, 2, \dots$ , do

    if  $x^k = 0$  then

$$x^{k+1} = \arg \min \frac{1}{2} \|Ax - y^\delta\|_Y^2 + \alpha \|x\|_{\ell_1}$$

    else

      determine  $z^k$  by

$$z^k = \mathbb{P}_{R_j} \left( x^k + \frac{\beta x^k}{\lambda \|x^k\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right)$$

      determine a step size  $s^k$  as a solution of

$$\min_{s \in [0,1]} F(x^k + s(z^k - x^k)) + \Phi(x^k + s(z^k - x^k))$$

$$x^{k+1} = x^k + s^k(z^k - x^k)$$

    end if

$k = k + 1$

end for

  if (2.2) is satisfied, set  $R_{j+1} = R_j + c$ ,  $c > 1$

    otherwise stop iteration

  end if

$j = j + 1$

end for

---

ularization method. As we know, Tikhonv type functions combined with Morozov's discrepancy principle is a regularization method. However, this result is usually shown only when the regularized term is convex ([1, 8, 34, 38, 42, 43]). If the regularized term is non-convex, some results can be found in [15, 44] where Morozov's discrepancy principle is applied to derive the convergence rate. However, these results are obtained under additional source conditions on the true solution  $x^\dagger$ . To the best of our knowledge, no results are available on whether Morozov's discrepancy principle combined with (1.2) is a regularization method. In this paper, we prove that if the non-convex regularized term satisfies some properties, e.g. coercivity, weakly lower semi-continuity and Radon-Riesz property, the well-posedness of the regularization still holds.

### 3.2 Well-posedness of regularization

In this section, we discuss the well-posedness of (1.2) under Morozov's discrepancy principle. First, we show that there exists at least one regularization parameter  $\alpha$  in (1.2) such that Morozov's discrepancy principle (2.2) holds. We recall some properties of  $\mathcal{R}_{\alpha,\beta}(x)$  ([15]), needed in analyzing the well-posedness of (1.2).

**Lemma 3.3** *If  $\alpha > \beta$ , the function  $\mathcal{R}_{\alpha,\beta}(x)$  in (1.2) has the following properties:*

- (i) (Coercivity) *For  $x \in \ell_2$ ,  $\|x\|_{\ell_2} \rightarrow \infty$  implies  $\mathcal{R}_{\alpha,\beta}(x) \rightarrow \infty$ .*
- (ii) (Weak lower semi-continuity) *If  $x_n \rightharpoonup x$  in  $\ell_2$  and  $\{\mathcal{R}_{\alpha,\beta}(x_n)\}$  is bounded, then*

$$\liminf_n \mathcal{R}_{\alpha,\beta}(x_n) \geq \mathcal{R}_{\alpha,\beta}(x).$$

- (iii) (Radon-Riesz property) *If  $x_n \rightharpoonup x$  in  $\ell_2$  and  $\mathcal{R}_{\alpha,\beta}(x_n) \rightarrow \mathcal{R}_{\alpha,\beta}(x)$ , then  $\|x_n - x\|_{\ell_2} \rightarrow 0$ .*

**Definition 3.4** *For fixed  $\delta$  and  $\eta \in [0, 1]$ , define*

$$\begin{aligned} F(x_{\alpha,\beta}^\delta) &= \frac{1}{2} \|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y^2, \\ \mathcal{R}_\eta(x_{\alpha,\beta}^\delta) &= \|x_{\alpha,\beta}^\delta\|_{\ell_1} - \eta \|x_{\alpha,\beta}^\delta\|_{\ell_2}, \\ m(\alpha) &= \mathcal{J}_{\alpha,\beta}^\delta(x_{\alpha,\beta}^\delta) = \min \mathcal{J}_{\alpha,\beta}^\delta(x), \end{aligned}$$

where  $\alpha \in (0, \infty)$  and  $\beta = \alpha\eta$ .

In the following we give some properties of  $m(\alpha)$ ,  $F(x_{\alpha,\beta}^\delta)$  and  $\mathcal{R}_\eta(x_{\alpha,\beta}^\delta)$  in Lemmas 3.5 and 3.6. Since  $\mathcal{R}_\eta(x_{\alpha,\beta}^\delta)$  is weakly lower semi-continuous, the proofs are similar to that in [42, Section 2.6]. Note that  $\eta \in [0, 1]$  is fixed, and for given  $\alpha_1, \alpha_2 \in (0, \infty)$ , we write  $\beta_1 = \alpha_1\eta$  and  $\beta_2 = \alpha_2\eta$ .

**Lemma 3.5** *The function  $m(\alpha)$  is continuous and non-increasing, i.e.,  $\alpha_1 > \alpha_2$  implies  $m(\alpha_1) \leq m(\alpha_2)$ . Moreover, for  $\alpha_1 > \alpha_2$ ,*

$$\begin{aligned} \sup_{x_{\alpha_1, \beta_1}^\delta \in \mathcal{L}_{\alpha_1, \beta_1}^\delta} F(x_{\alpha_1, \beta_1}^\delta) &\leq \inf_{x_{\alpha_2, \beta_2}^\delta \in \mathcal{L}_{\alpha_2, \beta_2}^\delta} F(x_{\alpha_2, \beta_2}^\delta), \\ \sup_{x_{\alpha_1, \beta_1}^\delta \in \mathcal{L}_{\alpha_1, \beta_1}^\delta} \mathcal{R}_\eta(x_{\alpha_1, \beta_1}^\delta) &\geq \inf_{x_{\alpha_2, \beta_2}^\delta \in \mathcal{L}_{\alpha_2, \beta_2}^\delta} \mathcal{R}_\eta(x_{\alpha_2, \beta_2}^\delta). \end{aligned}$$

**Lemma 3.6** *For each  $\bar{\alpha} > 0$  there exist  $x', x'' \in \mathcal{L}_{\bar{\alpha}, \bar{\beta}}^\delta$  such that*

$$\lim_{\alpha \rightarrow \bar{\alpha}^-} \left( \sup_{x_{\alpha, \beta}^\delta \in \mathcal{L}_{\alpha, \beta}^\delta} F(x_{\alpha, \beta}^\delta) \right) = F(x') = \inf_{x \in \mathcal{L}_{\bar{\alpha}, \bar{\beta}}^\delta} F(x) \quad \text{and} \quad \lim_{\alpha \rightarrow \bar{\alpha}^+} \left( \inf_{x_{\alpha, \beta}^\delta \in \mathcal{L}_{\alpha, \beta}^\delta} F(x_{\alpha, \beta}^\delta) \right) = F(x'') = \sup_{x \in \mathcal{L}_{\bar{\alpha}, \bar{\beta}}^\delta} F(x).$$

In the following we provide an existence result on the regularization parameter  $\alpha$ . The proof is along the line of Morozov's discrepancy principle for nonlinear ill-posed problems ([1, 34]).

**Lemma 3.7** *Assume  $0 < c_2\delta < \|y^\delta\|_Y$ . Then there exist  $\alpha_1, \alpha_2 \in \mathbb{R}^+$  such that*

$$\sup_{x_{\alpha_1, \beta_1}^\delta \in \mathcal{L}_{\alpha_1, \beta_1}^\delta} F(x_{\alpha_1, \beta_1}^\delta) < \tau_1\delta \leq \tau_2\delta < \inf_{x_{\alpha_2, \beta_2}^\delta \in \mathcal{L}_{\alpha_2, \beta_2}^\delta} F(x_{\alpha_2, \beta_2}^\delta).$$

**Proof.** First, let  $\alpha_n \rightarrow 0$  and consider a sequence of corresponding minimizers  $x_n := x_{\alpha_n, \beta_n}^\delta \in \mathcal{L}_{\alpha_n, \beta_n}^\delta$ . By the definition of  $x_{\alpha, \beta}^\delta$  and  $x^\dagger$ , we have

$$F(x_n)^q \leq m(\alpha_n) \leq \mathcal{J}_{\alpha_n}(x^\dagger) \leq \delta^q + \alpha_n \mathcal{R}_\eta(x^\dagger) \rightarrow \delta^q < \tau_1^q \delta^q.$$

This implies that there exists a small enough  $\alpha_1$  such that  $\sup_{x_{\alpha_1, \beta_1}^\delta \in \mathcal{L}_{\alpha_1, \beta_1}^\delta} F(x_{\alpha_1, \beta_1}^\delta) < \tau_1\delta$ .

Next, let  $\alpha_n \rightarrow \infty$ . Then

$$\mathcal{R}_\eta(x_n) \leq \frac{1}{\alpha_n} m(\alpha_n) \leq \frac{1}{\alpha_n} \|A0 - y^\delta\|_Y \rightarrow 0. \quad (3.7)$$

From the definition of  $\mathcal{R}_\eta(x)$ ,

$$\mathcal{R}_\eta(x) = (1 - \eta) \|x\|_{\ell_1} + \eta (\|x\|_{\ell_1} - \|x\|_{\ell_2}). \quad (3.8)$$

Then a combination of (3.7) and (3.8) implies that  $\{\|x_n\|_{\ell_2}\}$  is bounded. Consequently,  $\{x_n\}$  has a convergent subsequence, again denoted by  $\{x_n\}$ , such that  $x_n \rightharpoonup x^*$  for some  $x^* \in \ell_2$ . By Lemma 3.3 (ii), it follows from (3.7) that

$$0 \leq \mathcal{R}_\eta(x^*) \leq \liminf \mathcal{R}_\eta(x_n) = \lim \mathcal{R}_\eta(x_n) = 0.$$

By (3.8), this implies  $x^* = 0$ . Since  $x_n \rightharpoonup 0$  and  $\mathcal{R}_\eta(x_n) \rightarrow \mathcal{R}_\eta(0)$ , Lemma 3.3 (iii) implies that  $x_n \rightarrow 0$ . Then

$$\|Ax_n - y^\delta\|_Y \rightarrow \|A0 - y^\delta\|_Y = \|y^\delta\|_Y > c_2\delta.$$

This implies that there exists a large enough  $\alpha_2$  such that  $\inf_{x_{\alpha_2, \beta_2}^\delta \in \mathcal{L}_{\alpha_2, \beta_2}^\delta} F(x_{\alpha_2, \beta_2}^\delta) > \tau_2 \delta$ . ■

Note that we require  $\|y^\delta\|_Y > c_2 \delta$  in Lemma 3.7, which is a reasonable assumption. Indeed, in applications, it is almost impossible to recover a solution from observed data of a size in the same order as the noise.

We state an existence result on the regularized parameter, similar to Theorems 3.10 in [1]. The proof makes use of the properties stated in Lemmas 3.6 and 3.7.

**Theorem 3.8** *Assume  $\|y^\delta\|_Y > c_2 \delta > 0$  and there is no  $\alpha > 0$  with minimizers  $x', x'' \in \mathcal{L}_{\alpha, \beta}^\delta$  such that*

$$\|Ax' - y^\delta\|_Y < \tau_1 \delta \leq \tau_2 \delta < \|Ax'' - y^\delta\|_Y.$$

*Then there exist  $\alpha = \alpha(\delta, y^\delta) > 0$  and  $x_{\alpha, \beta}^\delta \in \mathcal{L}_{\alpha, \beta}^\delta$  such that (2.2) holds.*

Next, we give the convergence of (1.2) under Morozov's discrepancy principle.

**Theorem 3.9** (Convergence) *Let  $x_{\alpha_n, \beta_n}^{\delta_n}$  be a minimizer of  $\mathcal{J}_{\alpha_n, \beta_n}^{\delta_n}(x)$  defined by (2.1) with the data  $y^{\delta_n}$  satisfying  $\|y - y^{\delta_n}\| \leq \delta_n$ , where  $\delta_n \rightarrow 0$  if  $n \rightarrow +\infty$  and  $y^{\delta_n}$  belongs to the range of  $A$ . Let  $\alpha_n$  be determined by Morozov's discrepancy principle,*

$$\tau_1 \delta_n \leq \|A(x_{\alpha_n, \beta_n}^{\delta_n}) - y^{\delta_n}\|_Y \leq \tau_2 \delta_n, \quad 1 < \tau_1 \leq \tau_2.$$

*Moreover, assume that  $\eta = \lim_{n \rightarrow \infty} \eta_n \in [0, 1)$  exists, where  $\eta_n = \beta_n / \alpha_n$ . Then there exists a subsequence of  $\{x_{\alpha_n, \beta_n}^{\delta_n}\}$ , denoted by  $\{x_{\alpha_{n_k}, \beta_{n_k}}^{\delta_{n_k}}\}$ , that converges to an  $\mathcal{R}_\eta$ -minimizing solution  $x^\dagger$  in  $\ell_2$ . If, in addition, the  $\mathcal{R}_\eta$ -minimizing solution  $x^\dagger$  is unique, then*

$$\lim_{n \rightarrow +\infty} \|x_{\alpha_n, \beta_n}^{\delta_n} - x^\dagger\|_{\ell_2} = 0.$$

**Proof.** Denote  $y_n := y^{\delta_n}$ ,  $x_n := x_{\alpha_n, \beta_n}^{\delta_n}$ ,  $\eta_n := \eta^{\delta_n}$ . By the definition of  $x_n$ , we obtain

$$\begin{aligned} \frac{1}{q} \|Ax_n - y_n\|_Y^q + \alpha_n \|x_n\|_{\ell_1} - \beta_n \|x_n\|_{\ell_2} &\leq \frac{1}{q} \|Ax^\dagger - y_n\|_Y^q + \alpha_n \|x^\dagger\|_{\ell_1} - \beta_n \|x^\dagger\|_{\ell_2} \\ &\leq \frac{1}{q} \delta_n^q + \alpha_n \|x^\dagger\|_{\ell_1} - \beta_n \|x^\dagger\|_{\ell_2}. \end{aligned} \quad (3.9)$$

Since  $\tau_1 \delta_n \leq \|Ax_n - y_n\|_Y$ , it follows from (3.9) that

$$\alpha_n \|x_n\|_{\ell_1} - \beta_n \|x_n\|_{\ell_2} \leq \alpha_n \|x^\dagger\|_{\ell_1} - \beta_n \|x^\dagger\|_{\ell_2}.$$

Then we have

$$\limsup_{n \rightarrow +\infty} (\|x_n\|_{\ell_1} - \eta_n \|x_n\|_{\ell_2}) \leq \|x^\dagger\|_{\ell_1} - \eta \|x^\dagger\|_{\ell_2}. \quad (3.10)$$

Since  $\|x_n\|_{\ell_2}$  is bounded, there exist an  $x^* \in \ell_2$  and a subsequence of  $\{x_{n_k}\}$  such that  $x_{n_k} \rightharpoonup x^*$  in  $\ell_2$ . By Morozov's discrepancy principle, we obtain

$$\|Ax_{n_k} - y\|_Y \leq \|Ax_{n_k} - y_{n_k}\|_Y + \|y - y_{n_k}\|_Y \leq (\tau_2 + 1)\delta_{n_k}.$$

Therefore, weak lower semicontinuity of the norm gives

$$\|Ax^* - y\| \leq \liminf_{k \rightarrow \infty} \|Ax_{n_k} - y\|_Y = 0. \quad (3.11)$$

Meanwhile, by (3.10) and Lemma 3.3 (ii), we have

$$\begin{aligned} \|x^*\|_{\ell_1} - \eta\|x^*\|_{\ell_2} &\leq \liminf_k (\|x_{n_k}\|_{\ell_1} - \eta_{n_k}\|x_{n_k}\|_{\ell_2}) \leq \limsup_k (\|x_{n_k}\|_{\ell_1} - \eta_{n_k}\|x_{n_k}\|_{\ell_2}) \\ &\leq \limsup_n (\|x_n\|_{\ell_1} - \eta_n\|x_n\|_{\ell_2}) \leq \|x^\dagger\|_{\ell_1} - \eta\|x^\dagger\|_{\ell_2}. \end{aligned} \quad (3.12)$$

By the definition of  $x^\dagger$ , a combination of (3.11) and (3.12) implies that  $x^*$  is an  $\mathcal{R}_\eta$ -minimizing solution. Hence,  $\lim_{k \rightarrow \infty} \mathcal{R}_\eta(x_{n_k}) \rightarrow \mathcal{R}_\eta(x^*)$ . By Lemma 3.3 (iii), we have  $x_{n_k} \rightarrow x^*$ . If the  $\mathcal{R}_\eta$ -minimizing solution is unique, then  $x^* = x^\dagger$ . This implies that, for every subsequence  $\{x_{n_k}\}$ ,  $x_{n_k}$  converges to  $x^\dagger$ , then we have  $\lim_{n \rightarrow +\infty} \|x_n - x^\dagger\|_{\ell_2} = 0$ . ■

The numerical experiments in [15] show that we can obtain satisfactory results even when  $\alpha = \beta$ . Indeed,  $\mathcal{R}_{\alpha,\beta}(x)$  behaves more and more like the  $\ell_0$ -norm as  $\beta/\alpha \rightarrow 1$ . Nevertheless, note that if  $\alpha = \beta$ ,  $\mathcal{R}_{\alpha,\alpha}(x)$  fails to satisfy the coercivity and the Radon-Riesz property, and we can not ensure the convergence in  $\ell_2$ -norm. Without the Radon-Riesz property, we may expect to have only weak convergence for the regularized solution. If we assume the operator  $A$  is coercive in  $\ell_2$ , i.e.  $\|x\|_{\ell_2} \rightarrow \infty$  implies  $\|Ax\|_Y \rightarrow \infty$ , then the proof of the weak convergence is similar to that of Theorem 3.9.

## 4 The projected gradient method via the surrogate function approach

In this section, we propose another projected gradient algorithm for (1.2) in the finite dimensional space  $\mathbb{R}^n$  based on the surrogate function approach. By the discussion in Subsection 1.2, we consider the optimization problem (1.11). The following result provides a first order optimality condition for (1.11).

**Lemma 4.1** *Let  $0 \neq \hat{w} \in \mathbb{R}^n$  be a minimizer of (1.11). Then*

$$\mathbb{P}_R \left( \hat{w} + \frac{\beta \hat{w}}{\lambda \|\hat{w}\|_{\ell_2}} - \frac{1}{\lambda} A^*(A\hat{w} - y^\delta) \right) = \hat{w} \quad (4.1)$$

for any  $\lambda > 0$ , equivalently,

$$\left\langle \frac{\beta \hat{w}}{\|\hat{w}\|_{\ell_2}} - A^*(A\hat{w} - y^\delta), w - \hat{w} \right\rangle \leq 0, \quad (4.2)$$

for all  $w \in B_R$ .

**Proof.** By the definition of  $\hat{w}$ , for any  $w \in B_R$ , the function

$$f(t) = \frac{1}{2} \|A((1-t)\hat{w} + tw) - y^\delta\|_{\ell_2}^2 - \beta \|(1-t)\hat{w} + tw\|_{\ell_2}, \quad t \in [0, 1]$$

has its minimum at  $t = 0$ . Thus,

$$f'(0+) = \langle A\hat{w} - y^\delta, A(w - \hat{w}) \rangle - \beta \|\hat{w}\|_{\ell_2}^{-1} \langle \hat{w}, w - \hat{w} \rangle \geq 0,$$

i.e., (4.2) holds. ■

Due to the non-convexity of  $\mathcal{D}_\beta^\delta(x)$ , (4.2) is only a necessary condition of (1.11).

**Lemma 4.2** For any fixed  $\beta \geq 0$ , define

$$\Phi_\lambda(w, x) := \frac{1}{2} \|Aw - y^\delta\|_{\ell_2}^2 - \beta \|w\|_{\ell_2} - \frac{1}{2} \|A(w - x)\|_{\ell_2}^2 + \frac{\lambda}{2} \|w - x\|_{\ell_2}^2, \quad w, x \in B_R. \quad (4.3)$$

Then for any fixed  $x \in B_R$ , there exists a minimizer  $\hat{w}$  of  $\Phi_\lambda(w, x)$  on  $B_R$ .

**Proof.** Being continuous, the function  $\Phi_\lambda(\cdot, x)$  has a minimum on the compact set  $B_R$ . ■

Note that a minimizer  $\hat{w}$  of  $\Phi_\lambda(w, x)$  depends on  $x$  in  $\Phi_\lambda(w, x)$ . For  $w \neq 0$ , we denote

$$a_{i,j}(w) = \frac{\partial^2 \|w\|_{\ell_2}}{\partial w_i \partial w_j}, \quad 1 \leq i, j \leq n.$$

Then,

$$a_{i,j}(w) = \frac{\delta_{ij}}{\|w\|_{\ell_2}} - \frac{w_i w_j}{\|w\|_{\ell_2}^3}, \quad 1 \leq i, j \leq n. \quad (4.4)$$

Since  $\|w\|_{\ell_2}$  is convex, the matrix  $(a_{i,j}(w))_{n \times n}$  is positive semi-definite. Thus,  $\text{eig}(w) \geq 0$ , where  $\text{eig}(w)$  denotes the eigenvalues of  $(a_{i,j}(w))_{n \times n}$ . Moreover,  $\max\{\text{eig}(w)\}$  is an increasing function of  $\|w\|_{\ell_2}$ .

**Lemma 4.3** Let  $\hat{w}$  be a minimizer of  $\Phi_\lambda(w, x)$ . For a fixed  $\beta \geq 0$  and a fixed nonzero  $x \in B_R$ , there exists  $\lambda > 0$  such that  $\lambda > \max\{\text{eig}(\hat{w})\}$ .

**Proof.** As  $\lambda \rightarrow +\infty$  in (4.3), all minimizers  $\hat{w}$  of  $\Phi_\lambda(w, x)$  converge to  $x$ . Then  $\text{eig}(\hat{w}) \rightarrow \text{eig}(x)$ . Since  $0 \neq x \in B_R$  is fixed, there exists a large enough  $\lambda$  such that  $\lambda \geq \max_n\{\text{eig}(\hat{w})\}$ . ■

**Lemma 4.4** For a nonzero minimizer  $\hat{w}$  of  $\Phi_\lambda(w, x)$  and a fixed  $\beta \geq 0$ , if  $\lambda \geq \beta \max\{\text{eig}(\hat{w})\}$ , then  $\Phi_\lambda(w, x)$  is locally convex.

**Proof.** By the definition of  $\Phi_\lambda(w, x)$ ,

$$\frac{\partial^2 \Phi_\lambda(w, x)}{\partial w_i \partial w_j} = \lambda \delta_{ij} - \beta a_{i,j}(w), \quad 1 \leq i, j \leq n.$$

By the assumption  $\lambda \geq \beta \max\{\text{eig}(\hat{w})\}$ , the Hessian matrix  $(\frac{\partial^2 \Phi_\lambda(w, x)}{\partial w_i \partial w_j} |_{w=\hat{w}})$  is positive semi-definite. This proves the lemma. ■

In Lemma 4.4, we assume  $\lambda \geq \beta \max\{\text{eig}(\hat{w})\}$ . This condition is weaker than  $\lambda \geq \max\{\text{eig}(\hat{w})\}$ . In general, the regularization parameter  $\alpha \ll 1$  in the Tikhonov regularization. Since  $\beta = \alpha\eta$  and  $0 \leq \eta \leq 1$ , we also have  $\beta \ll 1$ .

**Lemma 4.5** *Let  $0 \neq \hat{w} \in B_R$  and  $\lambda \geq \beta \max\{\text{eig}(\hat{w})\}$ . Then  $\hat{w}$  is a minimizer of  $\Phi_\lambda(w, x)$  on  $B_R$  if and only if*

$$\hat{w} = \mathbb{P}_R \left( x + \frac{\beta \hat{w}}{\lambda \|\hat{w}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax - y^\delta) \right). \quad (4.5)$$

**Proof.** By the definition of  $\hat{w}$ , for any  $w \in B_R$ , the function

$$\begin{aligned} f(t) &= \frac{1}{2} \|A((1-t)\hat{w} + tw) - y^\delta\|_{\ell_2}^2 - \beta \|(1-t)\hat{w} + tw\|_{\ell_2} \\ &\quad - \frac{1}{2} \|A((1-t)\hat{w} + tw - x)\|_{\ell_2}^2 + \frac{\lambda}{2} \|(1-t)\hat{w} + tw - x\|_{\ell_2}^2, \quad t \in [0, 1] \end{aligned}$$

has its minimum at  $t = 0$ . Thus,

$$\begin{aligned} f'(0+) &= \langle A\hat{w} - y^\delta, A(w - \hat{w}) \rangle - \beta \|\hat{w}\|_{\ell_2}^{-1} \langle \hat{w}, w - \hat{w} \rangle - \langle A\hat{w} - Ax, A(w - \hat{w}) \rangle + \lambda \langle \hat{w} - x, w - \hat{w} \rangle \\ &\geq 0, \end{aligned}$$

i.e.,

$$\left\langle \frac{1}{\lambda} A^*(Ax - y) + \hat{w} - x - \frac{\beta}{\lambda} \frac{\hat{w}}{\|\hat{w}\|_{\ell_2}}, w - \hat{w} \right\rangle \geq 0.$$

By Lemma 2.8, this implies (4.5).

On the other hand, let now  $\hat{w} \in B_R$  be such that (4.5) holds. By Lemma 2.8, we have

$$\left\langle x + \frac{\beta \hat{w}}{\lambda \|\hat{w}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax - y) - \hat{w}, w - \hat{w} \right\rangle \leq 0.$$

Define

$$J(w) := \Phi_\lambda(w, x) = \frac{1}{2} \|Aw - y\|_{\ell_2}^2 - \beta \|w\|_{\ell_2} - \frac{1}{2} \|A(w - x)\|_{\ell_2}^2 + \frac{\lambda}{2} \|w - x\|_{\ell_2}^2. \quad (4.6)$$

If  $w \neq 0$ , we have

$$J'(w) = A^*(Ax - y) + \lambda(w - x) - \beta \frac{w}{\|w\|_{\ell_2}}. \quad (4.7)$$



By (4.7), this implies that

$$0 \leq \langle J'(\hat{w}), w - \hat{w} \rangle = \lim_{t \rightarrow 0^+} \frac{J(\hat{w} + t(w - \hat{w})) - J(\hat{w})}{t}. \quad (4.8)$$

By assumption and Lemma 4.4,  $\Phi_\lambda(w, x)$  is locally convex at  $\hat{w}$ . This implies that

$$\begin{aligned} 0 \leq \langle J'(\hat{w}), w - \hat{w} \rangle &= \lim_{t \rightarrow 0^+} \frac{J(\hat{w} + t(w - \hat{w})) - J(\hat{w})}{t} \\ &\leq \lim_{t \rightarrow 0^+} \frac{tJ(w) + (1-t)J(\hat{w}) - J(\hat{w})}{t} = J(w) - J(\hat{w}) \end{aligned}$$

for all  $w \in B_R$ . This proves the lemma.  $\blacksquare$

Denoting by  $x^{k+1}$  the sequence generated by the formula

$$x^{k+1} = \mathbb{P}_R \left( x^k + \frac{\beta x^{k+1}}{\lambda \|x^{k+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right). \quad (4.9)$$

The projected gradient algorithm based on the surrogate function is stated in the form of Algorithm 4.

---

**Algorithm 4** PG algorithm for (1.11) based on the surrogate function approach

---

Choose  $x^0 \in \mathbb{R}^n$ ,  $R_0 \in \mathbb{R}^+$ ,  $\beta = O(\delta)$  and  $\lambda$  such that  $\lambda > \beta \max\{\text{eig}(x^0), \text{eig}(x^\dagger)\}$   
for  $j = 0, 1, 2, \dots$  do  
  for  $k = 0, 1, 2, \dots$  do  
     $x^{k+1} = \mathbb{P}_{R_j} \left( x^k + \frac{\beta x^{k+1}}{\lambda \|x^{k+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right)$  (by fixed point iteration)  
     $k = k + 1$   
  end for  
  if (2.2) is satisfied, set  $R_{j+1} = R_j + c$ ,  $c > 1$   
  otherwise stop iteration  
  end if  
   $j = j + 1$   
end for

---

To prove the convergence of Algorithm 4, we impose some restrictions on the operator  $A$  and  $\lambda$ .

**Assumption 4.6** Let  $r := \|A^*A\|_{L(\mathbb{R}^n, \mathbb{R}^n)} < 1$ . Assume that

(A1)  $\|Ax\|_{\ell_2}^2 \leq \frac{\lambda r}{2} \|x\|_{\ell_2}^2$  for all  $x \in \ell_2$

(A2)  $\lambda \geq \beta \max\{\text{eig}(x^k)\}$  for all  $k$ .

In Assumption 4.6, we let  $r := \|A^*A\|_{L(\mathbb{R}^n, \mathbb{R}^n)} < 1$ . In the classical theory of sparsity regularization, the value of  $\|A_{m \times n}\|_2$  is assumed to be less than 1 ([12]), where  $m$  denotes the number of rows

in the operator  $A$ . This requirement is still needed in this paper. If  $r > 1$ , we need to re-scale the original ill-posed problem by  $A_{m \times n} x_n = y_m \rightarrow (\frac{1}{c} A_{m \times n}) x_n = \frac{1}{c} y_m$  so that  $\frac{1}{c^2} \|A^* A\|_{L(\mathbb{R}^n, \mathbb{R}^n)} < 1$ , where  $c > 1$ . If  $r < 1$ , we let  $\lambda > 2$ ; then (A1) holds. As for (A2), it seems that we need to compute eigenvalues for every  $(a_{ij}(x^k))_{n \times n}$ . However, we can give an approximation for the eigenvalues of  $(a_{ij}(x^k))_{n \times n}$ . In this paper, we first estimate the value of  $\|x^\dagger\|_{\ell_2}$  and  $\|x^0\|_{\ell_2}$ , then we can give an approximation for the order of the maximal eigenvalues of  $\|x^\dagger\|_{\ell_2}$  and  $\|x^0\|_{\ell_2}$ . Subsequently, we choose  $\lambda$  such that  $\lambda$  is greater than the order of the maximal eigenvalues of  $\|x^\dagger\|_{\ell_2}$  and  $\|x^0\|_{\ell_2}$ . If the value of  $\|x^\dagger\|_{\ell_2}$  is too small, we can re-scale the original ill-posed problem by  $A_{m \times n} x_n = y_m \rightarrow (\frac{1}{c} A_{m \times n})(c x_n) = y_m$  to increase the value of  $\|x^\dagger\|_{\ell_2}$ , where  $c > 1$ . Meanwhile, this strategy can reduce the value of  $\|A_{m \times n}\|_2$ , see Section 5 for details.

**Lemma 4.7** *Let Assumption 4.6 hold with  $\{x^{k+1}\}$  generated by (4.9). Then,*

$$\mathcal{D}_\beta^\delta(x^{k+1}) \leq \mathcal{D}_\beta^\delta(x^k)$$

and

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|_{\ell_2} = 0.$$

**Proof.** By Lemma 4.5 and the definition of  $x^{k+1}$ , we see that  $x^{k+1}$  is a minimizer of  $\Phi_\lambda(w, x^k)$ . Then we have

$$\begin{aligned} \mathcal{D}_\beta^\delta(x^{k+1}) &\leq \mathcal{D}_\beta^\delta(x^{k+1}) + \frac{2-r}{2r} \|A(x^{k+1} - x^k)\|_Y^2 \\ &\leq \frac{1}{2} \|Ax^{k+1} - y\|_Y^2 - \beta \|x^{k+1}\|_{\ell_2} + \frac{1}{r} \|A(x^{k+1} - x^k)\|_Y^2 - \frac{1}{2} \|A(x^{k+1} - x^k)\|_Y^2 \\ &\leq \frac{1}{2} \|Ax^{k+1} - y\|_Y^2 - \beta \|x^{k+1}\|_{\ell_2} - \frac{1}{2} \|A(x^{k+1} - x^k)\|_Y^2 + \frac{\lambda}{2} \|x^{k+1} - x^k\|_{\ell_2}^2 \\ &= \Phi_\lambda(x^{k+1}, x^k) \leq \Phi_\lambda(x^k, x^k) = \mathcal{D}_\beta^\delta(x^k). \end{aligned}$$

Furthermore,

$$\begin{aligned} \Phi_\lambda(x^{k+1}, x^k) - \Phi_\lambda(x^{k+1}, x^{k+1}) &= \frac{\lambda}{2} \|x^{k+1} - x^k\|_{\ell_2}^2 - \frac{1}{2} \|A(x^{k+1} - x^k)\|_Y^2 \\ &\geq \frac{\lambda(2-r)}{4} \|x^{k+1} - x^k\|_{\ell_2}^2. \end{aligned}$$

This implies

$$\begin{aligned} \sum_{k=0}^N \|x^{k+1} - x^k\|_{\ell_2}^2 &\leq \frac{4}{\lambda(2-r)} \sum_{k=0}^N (\Phi_\lambda(x^{k+1}, x^k) - \Phi_\lambda(x^{k+1}, x^{k+1})) \\ &\leq \frac{4}{\lambda(2-r)} \sum_{k=0}^N (\Phi_\lambda(x^k, x^k) - \Phi_\lambda(x^{k+1}, x^{k+1})) \\ &= \frac{4}{\lambda(2-r)} (\Phi_\lambda(x^0, x^0) - \Phi_\lambda(x^{N+1}, x^{N+1})) \\ &\leq \frac{4}{\lambda(2-r)} (\Phi_\lambda(x^0, x^0) + \beta R). \end{aligned}$$

Since  $\sum_{k=0}^N \|x^{k+1} - x^k\|_{\ell_2}^2$  is uniformly bounded with respect to  $N$ , the series  $\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|_{\ell_2}^2$  converges. This proves the lemma. ■

**Remark 4.8** *To prove the convergence, we need to analyze the relation between  $x^k$  and 0. If  $0 = x^0 = x^1$ , then we stop the iteration and 0 is the iterative solution. Otherwise, by Lemma 4.7,  $\mathcal{D}_\beta^\delta(x^k)$  decreases, which implies that  $x^k \neq 0$  for  $k \geq 1$ . So in the following we let  $x^k \neq 0$  whenever  $k \geq 1$ .*

**Lemma 4.9** *Denote  $\Psi(\hat{w}) := \mathbb{P}_R \left( x + \frac{\beta \hat{w}}{\lambda \|\hat{w}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax - y^\delta) \right)$ . Then the fixed point iteration  $\hat{w}^{l+1} = \Psi(\hat{w}^l)$  has a subsequence which converges to an element  $\hat{w}$ . If  $\hat{w} \neq 0$ , then  $\hat{w}$  is a fixed point of  $\Psi(\hat{w})$ .*

**Proof.** By Lemma 2.8,  $\mathbb{P}_R(x)$  is non-expansive,

$$\|\Psi(\hat{w}_1) - \Psi(\hat{w}_2)\|_{\ell_2} \leq \left\| \frac{\beta \hat{w}_1}{\lambda \|\hat{w}_1\|_{\ell_2}} - \frac{\beta \hat{w}_2}{\lambda \|\hat{w}_2\|_{\ell_2}} \right\|_{\ell_2},$$

which implies  $\Psi(\hat{w})$  is continuous at any nonzero element  $w$ . Since  $\{\hat{w}^l\}$  is bounded, it has a subsequence  $\{\hat{w}^{l_k}\}$  which converges to an element  $\hat{w} \in B_R$ . Since  $\hat{w}^{l_k+1} = \Psi(\hat{w}^{l_k})$ ,

$$\lim_k \hat{w}^{l_k+1} = \lim_k \Psi(\hat{w}^{l_k}). \quad (4.10)$$

If  $\hat{w} \neq 0$ , it follows from (4.10) that  $\hat{w} = \Psi(\hat{w})$ . ■

Even though  $\mathbb{P}_R(x)$  is non-expansive, the map  $\Psi(\hat{w})$  is not necessarily non-expansive. So we only have the existence of a fixed point. We can not ensure uniqueness of the fixed point. Indeed, due to the non-convexity of  $\Phi_\lambda(w, x)$  in (4.3), the minimizer of (4.3) may be non-unique. Nevertheless, the convergence still holds and the limit depends on the choice of the initial vector  $x^0$ .

**Theorem 4.10** (Convergence) *Let  $\{x^k\}$  be the sequence generated by*

$$x^{k+1} = \mathbb{P}_R \left( x^k + \frac{\beta x^{k+1}}{\lambda \|x^{k+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) \right).$$

*Then  $\{x^k\}$  has a subsequence which converges to a nonzero stationary point  $x^*$  of (1.11), i.e.  $x^*$  satisfies*

$$\left\langle \frac{\beta x^*}{\|x^*\|_{\ell_2}} - A^*(Ax^* - y^\delta), w - x^* \right\rangle \leq 0 \quad \forall w \in B_R.$$

**Proof.** Since  $\{x^k\} \subset B_R$  is bounded,  $\{x^k\}$  has a subsequence  $\{x^{k_j}\}$  converging to an element  $x^*$  in  $B_R$ , i.e.  $x^{k_j} \rightarrow x^*$  in  $B_R$ . Since  $A$  is linear and bounded,  $A(x^{k_j}) \rightarrow A(x^*)$ . By Lemma 2.8 and the definition of  $x^{k+1}$ , we see that, for all  $w \in B_R$ ,

$$\left\langle x^k + \frac{\beta x^{k+1}}{\lambda \|x^{k+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^k - y^\delta) - x^{k+1}, w - x^{k+1} \right\rangle \leq 0.$$

This implies that

$$\left\langle x^{k_j} + \frac{\beta x^{k_j+1}}{\lambda \|x^{k_j+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^{k_j} - y^\delta) - x^{k_j+1}, w - x^{k_j+1} \right\rangle \leq 0. \quad (4.11)$$

Taking the limit in (4.11) as  $j \rightarrow \infty$ , we have

$$\lim_{j \rightarrow \infty} \left\langle x^{k_j} + \frac{\beta x^{k_j+1}}{\lambda \|x^{k_j+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^{k_j} - y^\delta) - x^{k_j+1}, w - x^{k_j+1} \right\rangle \leq 0. \quad (4.12)$$

Since  $\|x^{k_j} - x^{k_j+1}\|_{\ell_2} \rightarrow 0$  as  $j \rightarrow \infty$  and  $\{w - x^{k_j+1}\}$  is uniformly bounded, we have

$$\lim_{j \rightarrow \infty} |\langle x^{k_j} - x^{k_j+1}, w - x^{k_j+1} \rangle| = 0. \quad (4.13)$$

A combination of (4.12) and (4.13) shows that

$$\lim_{j \rightarrow \infty} \left\langle \frac{\beta x^{k_j+1}}{\lambda \|x^{k_j+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^{k_j} - y^\delta), w - x^{k_j+1} \right\rangle \leq 0. \quad (4.14)$$

Since  $x^{k_j} \rightarrow x^*$ , it follows from (4.14) that

$$\left\langle \frac{\beta x^*}{\|x^*\|_{\ell_2}} - A^*(Ax^* - y^\delta), w - x^* \right\rangle \leq 0.$$

by Lemma 4.1,  $x^*$  is a stationary point of  $\mathcal{D}_\beta^\delta(x)$  on  $B_R$ . ■

**Remark 4.11** *In this section, we restrict the analysis of the projected algorithm based on the surrogate function approach in the finite dimensional space  $\mathbb{R}^n$ . Actually, all results except Lemma 4.9 and Theorem 4.10 can be extended to  $\ell_2$  space. In Theorem 4.10, if  $\{x^k\}$  is defined in  $\ell_2$  space, then  $\{x^k\}$  has a weak convergence subsequence  $\{x^{k_j}\} \rightharpoonup x^*$ . However, the challenge of the proof is that  $x^{k_j} \rightharpoonup x^*$  can not ensure  $x^{k_j+1}/\|x^{k_j+1}\|_{\ell_2} \rightharpoonup x^*/\|x^*\|_{\ell_2}$ . For example, let  $x_n = \bar{x} + e_n$ , where  $e_n = (\underbrace{0, \dots, 0}_n, 1, 0, \dots)$ . Since  $e_n \rightharpoonup 0$  in  $\ell_2$ ,  $x_n \rightharpoonup x$  in  $\ell_2$ . However,  $\|x_n\|_{\ell_2} \not\rightarrow \|x\|_{\ell_2}$ . Hence,*

*$x_n/\|x_n\|_{\ell_2}$  does not converge to  $x^*/\|x^*\|_{\ell_2}$ . If we impose an additional condition on  $\{x_n\}$ , e.g.  $\|x_n\|_{\ell_2} \rightarrow \|x\|_{\ell_2}$ , then we have  $x_n/\|x_n\|_{\ell_2} \rightharpoonup \eta x^*/\|x^*\|_{\ell_2}$ . However, this condition is too restrictive, since a combination of  $\|x_n\|_{\ell_2} \rightarrow \|x\|_{\ell_2}$  and  $x_n \rightharpoonup x^*$  in  $\ell_2$  implies that  $x_n \rightarrow x^*$ . Moreover, the iterative algorithm in this paper has an implicit formulation, and we need to compute the iterative solution. However, in  $\ell_2$  space, we do not know whether the operator  $\Phi(\hat{w})$  is weak-strong continuity. So we can not ensure that the fixed point iteration is convergent.*

## 5 Numerical experiments

In this section, we present results from two numerical experiments to demonstrate the efficiency of the proposed algorithms. Comparisons between ST- $(\alpha\ell_1 - \beta\ell_2)$  and the two projected gradient algorithms are provided. For convenience, we write PG-GCGM algorithm to refer to the first projected gradient algorithm which is based on GCGM, and PG-SF algorithm for the second projected gradient algorithm which is based on the surrogate function approach. The relative error (Rerror) is utilized to measure the performance of the reconstruction  $x^*$ :

$$\text{Rerror} := \frac{\|x^* - x^\dagger\|_{\ell_2}}{\|x^\dagger\|_{\ell_2}},$$

where  $x^\dagger$  is a true solution.

We utilize the algorithm in [5, Section 4.2] to compute the projection defined in Definition 2.5. The MATLAB code oneProjector.m regarding the  $\ell_1$ -ball projection can be obtained at <http://www.cs.ubc.ca/labs/scl/spgl1>. The first example deals with a well-conditioned compressive sensing problem. The second example deals with an ill-conditioned image deblurring problem. All numerical experiments were tested in MATLAB R2010 on an i7-6500U 2.50GHz workstation with 8Gb RAM.

### 5.1 Example 1: Compressive sensing

In the first example, we test compressive sensing with the commonly used random Gaussian matrix. The compressive sensing problem is defined as  $A_{m \times n}x_n = y_m$ , where  $A_{m \times n}$  is a well conditioned random Gaussian matrix by calling  $A = \text{randn}(m, n)$  in MATLAB. Exact data  $y^\dagger$  is generated by  $y^\dagger = Ax^\dagger$ . The exact solution  $x^\dagger$  is an  $s$ -sparse signal supported on a random index set. White Gaussian noise is added to the exact data  $y^\dagger$  by calling  $y^\delta = \text{awgn}(Ax^\dagger, \sigma)$  in MATLAB, where  $\sigma$  (measured in dB) measures the ratio between the true (noise free) data  $y^\dagger$  or  $Ax^\dagger$  and Gaussian noise. A larger value of  $\sigma$  corresponds to a smaller value of the noise level  $\delta$ , where the noise level  $\delta$  is defined by  $\delta = \|y^\delta - y^\dagger\|_2$ .  $x^*$  denotes the reconstruction computed by the proposed algorithms. For compressive sensing, if the value of  $\|(A^*A)_{n \times n}\|_2$  is greater than 1, we rescale the matrix  $A_{m \times n}$  by  $A_{m \times n} \rightarrow c * A_{m \times n}$ , where  $c < 1$ . Then the original compressive sensing problem  $A_{m \times n}x_n = y_m$  can be rewritten as  $(c * A_{m \times n})x_n = c * y_m$ . Note that the condition number does not change under the matrix rescaling. To compare the performance of ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm, PG-GCGM algorithm and PG-SF algorithm, we choose the same initial setting, i.e.  $\lambda$ ,  $\beta$  and the initial vector  $x^0$ . Moreover, for each fixed point iteration in PG-SF algorithm, we choose  $x^0 = \text{ones}(n, 1)$  as the initial vector.

We choose  $n = 200$ ,  $m = 0.4n$ ,  $s = 0.2m$ , then  $\|x^\dagger\|_0 = 16$ . A noise  $\delta$  is added to exact data  $y^\dagger$  by calling  $y^\delta = \text{awgn}(Ax^\dagger, \sigma)$ , where  $\sigma = 50\text{dB}$ ,  $\delta$  is around 0.02. We let  $\lambda = 1$ ,  $\eta = 1$ ,  $\alpha = O(\delta) = 0.2$ ,  $\beta = \alpha\eta = 0.2$  and the initial vector  $x^0$  is generated by calling  $x^0 = 0.01\text{ones}(n, 1)$ . We utilize discrepancy principle (2.2) to determine the radius  $R$  of the  $\ell_1$ -ball constraint such that

$R = \sup\{R > 0 \mid \delta \leq \|Ax^* - y^\delta\|_2\}$ . It is shown that when a good estimate for the noise level  $\delta$  is known, this method yields a good radius  $R$ . According to the priori information of  $x^\dagger$ , we choose an initial value of  $R$  and compute  $x^*$ . If  $\delta < \|Ax^* - y^\delta\|_2$ , we try  $R_j = R + j$ ,  $j = 1, 2, \dots$  until  $\|Ax^* - y^\delta\|_2 \leq \delta$ . With  $j$  increasing, we can find  $R = \sup\{R > 0 \mid \delta \leq \|Ax^* - y^\delta\|_2\}$ . On the contrary, for any initial  $R$ , if  $\|Ax^* - y^\delta\|_2 \leq \delta$ , we try  $R_j = R - j$ ,  $j = 1, 2, \dots$  until  $\delta < \|Ax^* - y^\delta\|_2$ . Fig. 1 shows Morozov's discrepancy principle for determining the radius  $R$ . We see that the discrepancy  $\|Ax^* - y^\delta\|_2$  is a decreasing function of the radius  $R$ . According the strategy stated above,  $R$  should be chosen such that  $R = \sup\{R > 0 \mid \delta < \|Ax^* - y^\delta\|_2\}$ . It is obvious that  $R$  should be chosen as 16. Indeed, by ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm, we can obtain  $\|x^*\|_1 = 16.0153$ . Thus the experimental results confirm that the strategy proposed in this paper is feasible and they match the theoretical results stated in Subsection 3.1, i.e.  $R$  should be chosen by  $R = \|x^*\|_1$ .

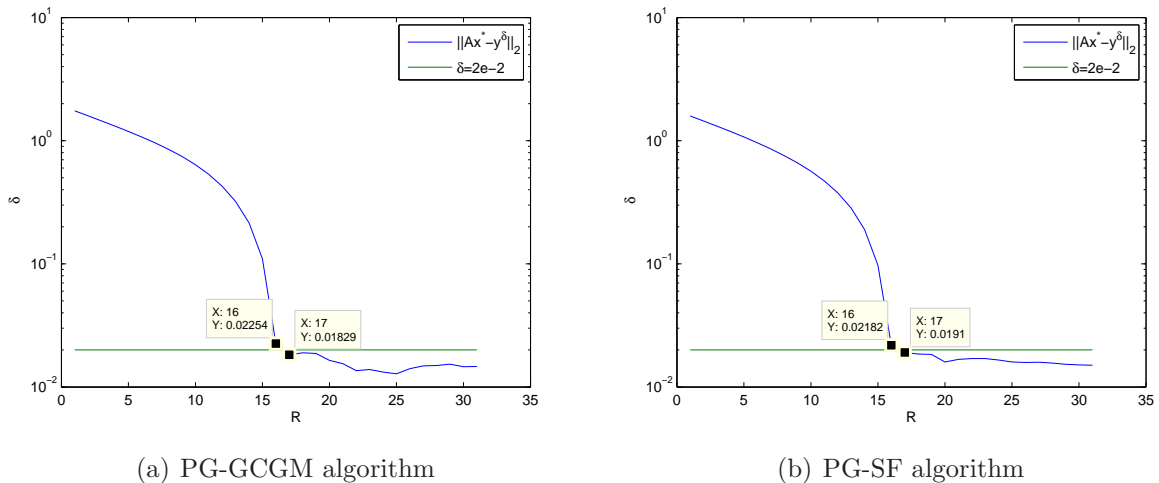
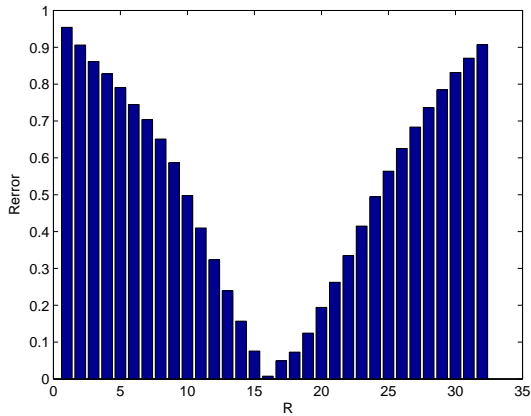


Figure 1: The discrepancy  $\|Ax^* - y^\delta\|_2$  vs.  $R$ .

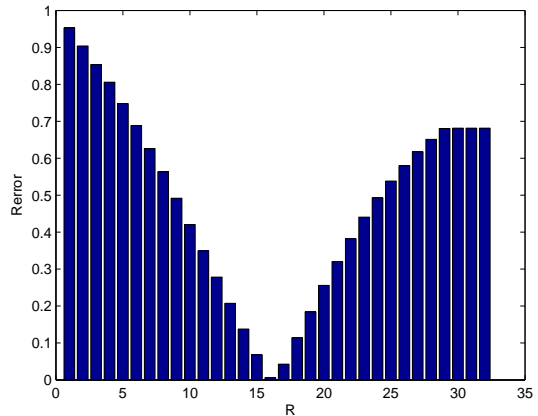
To test the stability of the PG Algorithms with respect to  $R$ , we choose several values of  $R$  in Fig. 2. It is shown that the two PG algorithms have good performance with the appropriate radius  $R$ . We see that the two PG algorithms are stable with respect to  $R$ . Furthermore, the results of reconstruction get better if  $R$  close to 16.

When  $0 < \eta \leq 1$ ,  $\mathcal{R}_\eta(x)$  is non-convex. To analyze the influence of  $\eta$ , we choose different values for the parameter  $\eta$ . From each row in Table 1, we see that, Error of reconstruction gets better with  $\eta$  increasing which implies the non-convex regularization (case  $\eta > 0$ ) has better performance compared to the classical  $\ell_1$  regularization (case  $\eta = 0$ ).

We test the convergence rate of the two PG algorithms and the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm. We are primarily interested in the time of computation corresponding to Error. The results are shown in Fig. 3. To get within a distance of the true minimizer corresponding to a  $7e-3$  relative error, PG-GCGM algorithm takes 0.62 second, PG-SF algorithm 1.08 seconds, and ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm 18.40 seconds. The ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm procedure is significantly slower than the



(a) PG-GCGM algorithm



(b) PG-SF algorithm

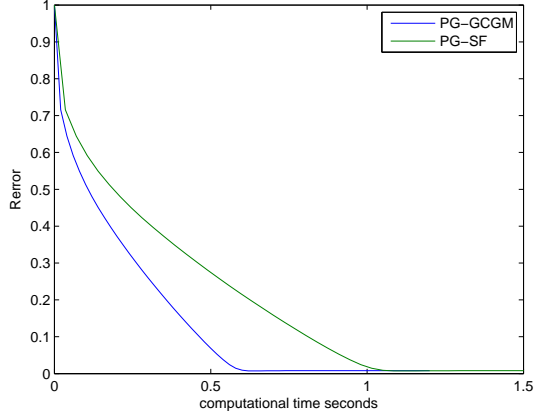
Figure 2: The relative error of reconstruction  $x^*$  by the two PG algorithms with different  $R$ .Table 1: Error of reconstruction  $x^*$  with different values of  $\eta$ .

$\eta$	0.0	0.1	0.2	0.3	0.4	0.5	0.7	0.9	1.0
ST- $(\alpha l_1 - \beta l_2)$	0.0250	0.0246	0.0147	0.0098	0.0086	0.0081	0.0073	0.0067	0.0064
PG-GCGM	0.0180	0.0126	0.0102	0.0089	0.0081	0.0074	0.0067	0.0061	0.0059
PG-SF	0.0356	0.0285	0.0197	0.0145	0.0121	0.0111	0.0096	0.0091	0.0089

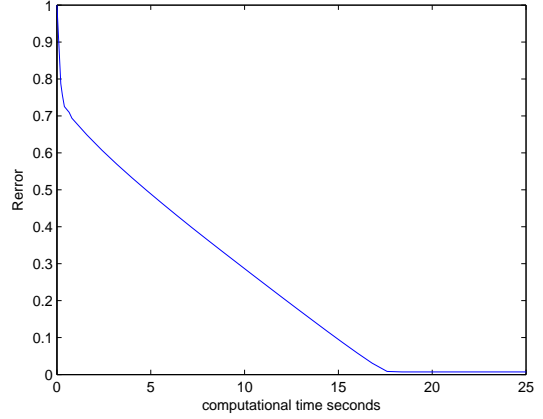
two PG algorithms.

Theoretically, for the PG-SF algorithm, we require that Assumption 4.6 (A2) holds, i.e.  $\lambda \geq \beta \max\{\text{eig}(x^k)\}$ . Next, we test whether  $\lambda$  satisfies this assumption. Fig. 4(a) shows Error corresponding to the different reconstruction  $x^k$ ,  $1 \leq k \leq 1500$  and Fig. 4(b) shows the maximal eigenvalues  $\max\{\text{eig}(x^k)\}$ . It is obvious that all  $\max\{\text{eig}(x^k)\}$  are less than 3.5. In this section, we let  $\lambda = 1$  and  $\beta = \alpha\eta$ , where  $\alpha = 0.02$  and  $\eta = 1$ . Thus,  $\lambda \geq 3.5\beta$ , which satisfies Assumption 4.6 (A2). Theoretically, we can let  $\lambda$  be any value greater than  $3.5\beta$ . Nevertheless, a larger value of  $\lambda$  corresponds to a smaller iteration step, and then we can not obtain a good convergence rate.

Finally, we let  $n = 1800$ ,  $m = 0.4n$  and  $s = 0.2m$ .  $\sigma = 50\text{dB}$ . The coefficients  $\lambda$  and  $\eta$  remain the same as in the first test. The noise level  $\delta$  is around 0.09, hence we let  $\beta = 0.1$ . We test the convergence rate of the two PG algorithms and ST- $(\alpha l_1 - \beta l_2)$  algorithm regarding computational time with several different values of Error. With the value of Error decreasing, when Error gets within each value, we check the computational time of the three algorithms. In Table 2, we see that the ST- $(\alpha l_1 - \beta l_2)$  algorithm takes more than 100 minutes to get within a distance of the true minimizer corresponding to a 2% relative error. The two PG algorithms only take around 8 and 41 seconds to reach the same level of relative error. The PG algorithms converge much faster than the ST- $(\alpha l_1 - \beta l_2)$  algorithm.

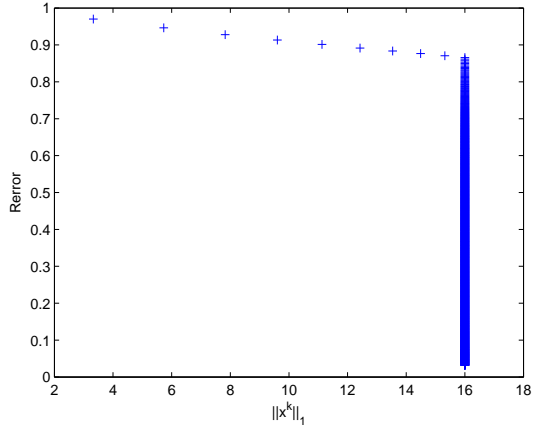


(a) PG-GCGM algorithm and PG-SF algorithm

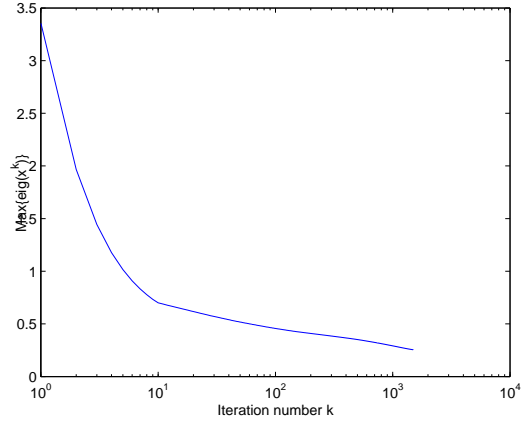


(b)  $ST-(\alpha l_1 - \beta l_2)$  algorithm

Figure 3: (a) Convergence rate of PG-GCGM algorithm and PG-SF algorithm; (b) Convergence rate of  $ST-(\alpha l_1 - \beta l_2)$  algorithm.



(a)



(b)

Figure 4: (a) Error for  $x^k$ ,  $1 \leq k \leq 1500$ ; (b)  $\max\{\text{eig}(x^k)\}$  for  $1 \leq k \leq 1500$ .

Table 2: Time of computation for the reconstruction  $x^*$  with different values of Error.

Error	$ST-(\alpha l_1 - \beta l_2)$ time	PG-GCGM time	PG-SF time
0.8	9.7463 m	0.0214 s	0.0208 s
0.6	12.7113 m	0.1926 s	0.8573 s
0.4	14.9283 m	0.6995 s	3.2097 s
0.2	24.8903 m	1.6924 s	7.5099 s
0.1	39.2569 m	2.8578 s	11.1562 s
0.05	60.5784 m	4.9201 s	22.2830 s
0.02	102.8623 m	8.2870 s	41.2480 s



## 5.2 Example 2: Image deblurring

In the second example, we test an ill-conditioned image deblurring problem which is the process of removing blurring artifacts from images, such as blur caused by defocus aberration or motion blur. The blur is typically modeled by a Fredholm integral equation of the first kind

$$\int_a^b K(s, t) f(t) dt = g(s),$$

where  $K(s, t)$  is the kernel function,  $g(s)$  is the observed image and  $f(t)$  is the true image. We utilize the blur problem from MATLAB Regularization Tools ([22]) by calling  $[A, b, x^\dagger] = \text{blur}(n, \text{band}, \tau)$ , where the Gaussian point-spread function is used as the kernel function

$$K(s, t) = \frac{1}{\pi\tau^2} \exp\left(-\frac{s^2 + t^2}{2\tau^2}\right).$$

The matrix  $A$  is a symmetric  $n^2 \times n^2$  Toeplitz matrix and is given by  $A = (2\pi\tau^2)^{-1}T \otimes T$ , where  $T$  is an  $n \times n$  symmetric banded Toeplitz matrix whose first row is obtained by calling

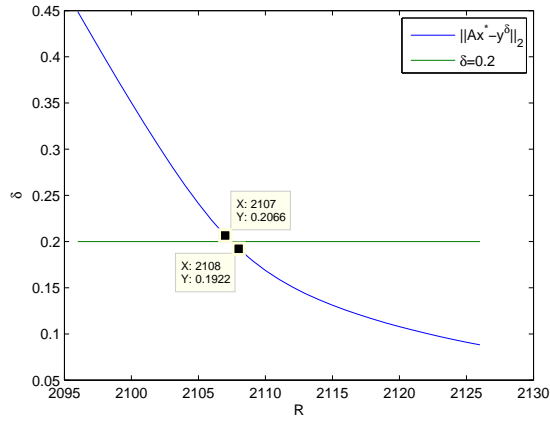
$$z = [\exp(-([0 : \text{band} - 1].^2)/(2\tau^2)); \text{zeros}(1, N - \text{band})].$$

The parameter  $\tau$  controls the shape of the Gaussian point spread function and thus the amount of smoothing (the larger the value of  $\tau$ , the wider the function, and the less ill-posed the problem).

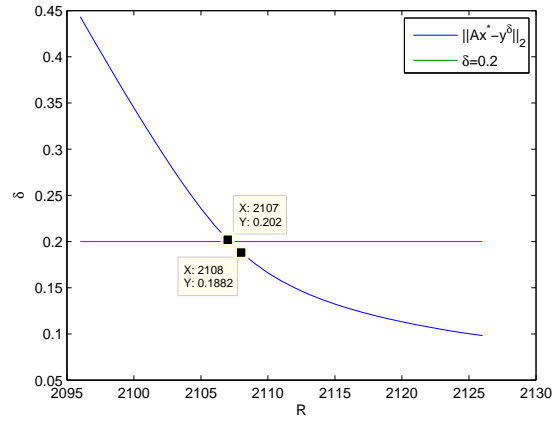
We choose  $n = 64$ ,  $\text{band} = 3$ ,  $\tau = 0.7$ . A noise  $\delta$  is added to exact data  $y^\dagger$  by calling  $y^\delta = \text{awgn}(Ax^\dagger, \sigma)$ , where  $\sigma = 50\text{dB}$ ,  $\delta$  is around 0.2. We let  $\lambda = 5$ ,  $\eta = 0.7$ ,  $\alpha = O(\delta) = 0.2$ ,  $\beta = \alpha\eta = 0.14$  and generate the initial vector  $x^0$  by calling  $x^0 = 0.01\text{ones}(n, 1)$ . The value of  $\|A\|_2$  is around 1 and the condition number is around 30. The initial value  $x^0$  is generated by calling  $x^0 = 0.01\text{ones}(n \times n, 1)$ . Fig. 5 shows Morozov's discrepancy principle for determining the radius  $R$ . We see that the value of the discrepancy  $\|Ax^* - y^\delta\|_2$  decreases with increasing radius  $R$ . According to the strategy stated previously,  $R$  should be chosen such that  $R = \sup\{R > 0 \mid \delta < \|Ax^* - y^\delta\|_2\}$ . It is obvious that  $R$  should be chosen as 2107. Actually, the optimal  $R$  is 2108 (see Fig. 6), thus the results of the experiment testify the theory, i.e.  $R$  should be chosen by  $R = \|x^*\|_1$ . Note that  $\|x^\dagger\|_1 = 2111$ . Fig. 6 shows the performance of the PG algorithms with respect to  $R$ . It is shown that the two PG algorithms have good performance with appropriate radius  $R$ . Observe that for a fixed parameter  $\eta$ , Error of reconstruction  $x^*$  gets better if  $R$  close to 2107.

To analyze the influence of  $\eta$ , we choose different values for the parameter  $\eta$ . From each row in Table 3, we see that the results of reconstruction get better with  $\eta$  increasing, implying that the non-convex regularization (for  $\eta > 0$ ) has better performance than the classical  $\ell_1$  regularization (for  $\eta = 0$ ). However, if  $\eta$  increases to near 1, the accuracy of recovery decreases and  $\eta = 0.7$  is optimal.

We test the convergence rate of the two PG algorithms and the ST- $(\alpha\ell_1 - \beta\ell_2)$  algorithm, focusing on the computation time corresponding to Rerror. The results are shown in Fig. 7. To get within a distance of the true minimizer corresponding to a 1.2e-2 relative error, the PG-GCGM

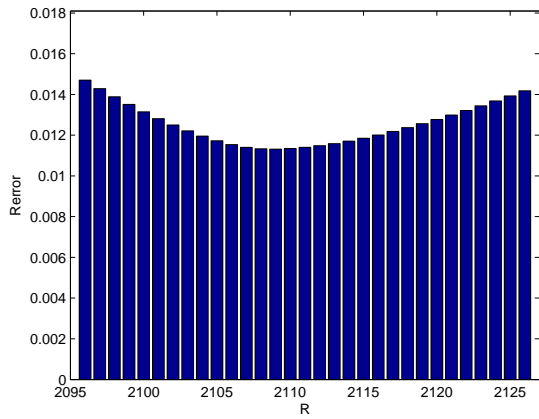


(a) PG-GCGM algorithm

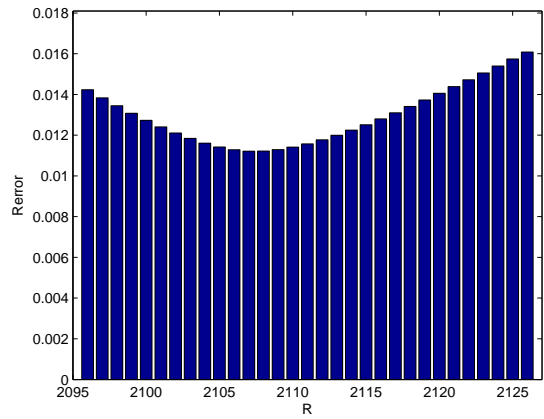


(b) PG-SF algorithm

Figure 5: The value of the discrepancy  $\|Ax^* - y^\delta\|_2$  with different  $R$ .



(a) PG-GCGM algorithm



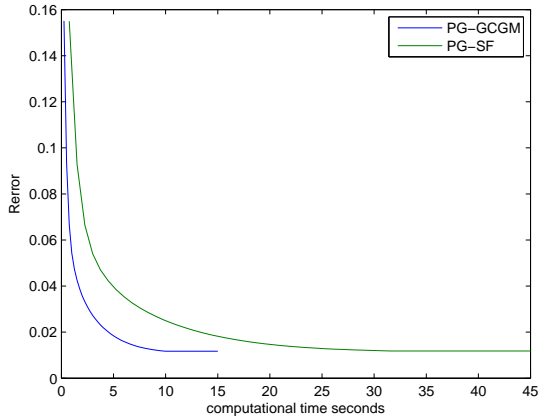
(b) PG-SF algorithm

Figure 6: The relative error of reconstruction  $x^*$  by the two PG algorithms with different  $R$ .

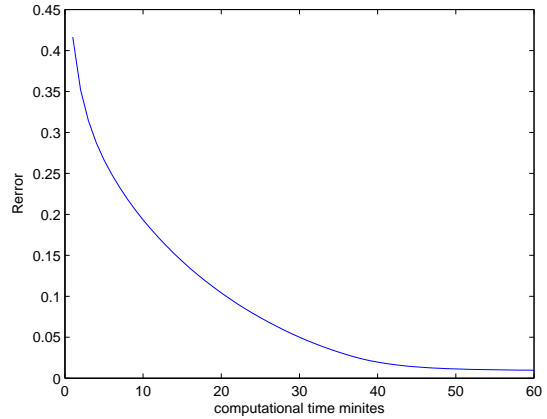
Table 3: Error of reconstruction  $x^*$  with different values of  $\eta$ .

$\eta$	0.0	0.1	0.2	0.3	0.4	0.5	0.7	0.9	1.0
ST- $(\alpha l_1 - \beta l_2)$	0.0265	0.0253	0.0231	0.0205	0.0163	0.0144	0.0125	0.0138	0.0198
PG-GCGM	0.0278	0.0263	0.0242	0.0225	0.0198	0.0162	0.0130	0.0152	0.0205
PG-SF	0.0296	0.0271	0.0237	0.0231	0.0204	0.0156	0.0126	0.0147	0.0203

algorithm takes 10.12 seconds, PG-SF algorithm 36.26 seconds, and the ST- $(\alpha l_1 - \beta l_2)$  algorithm 58.54 minutes. The ST- $(\alpha l_1 - \beta l_2)$  algorithm procedure is significantly slower than the two PG algorithms.



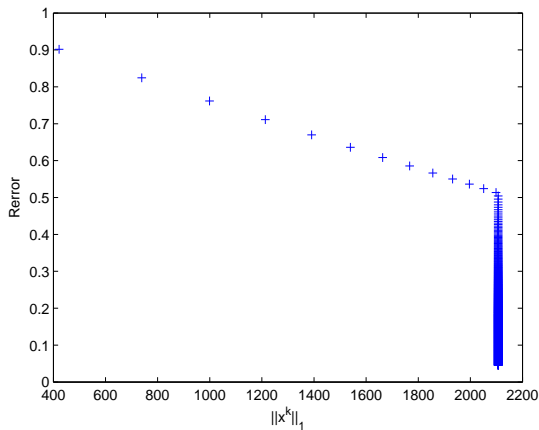
(a) PG-GCGM algorithm and PG-SF algorithm



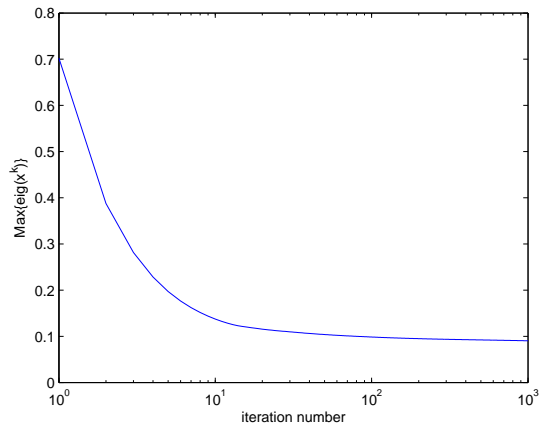
(b) ST- $(\alpha l_1 - \beta l_2)$  algorithm

Figure 7: (a) Convergence rate of PG-GCGM algorithm and PG-SF algorithm; (b) Convergence rate of ST- $(\alpha l_1 - \beta l_2)$  algorithm.

Theoretically, for the PG-SF algorithm, we require that Assumption 4.6 (A2) holds, i.e.  $\lambda \geq \beta \max\{\text{eig}(x^k)\}$ . In Fig. 8, we test whether  $\lambda$  satisfies this assumption. Fig. 8(a) shows Error corresponding to the different reconstruction  $x^k$  and Fig. 8(b) shows the maximal eigenvalue  $\max\{\text{eig}(x^k)\}$ . It is obvious that the maximal eigenvalue of all  $x^k$  is less than 0.45. We let  $\lambda = 1$  and  $\beta = \alpha\eta = 0.14$ , where  $\alpha = 0.2$  and  $\eta = 0.7$ . Thus,  $\lambda \geq 3.5\beta$ , and Assumption 4.6 (A2) is satisfied.



(a)



(b)

Figure 8: (a) Error for  $x^k$ ,  $1 \leq k \leq 1000$ ; (b)  $\max\{\text{eig}(x^k)\}$  for  $1 \leq k \leq 1000$ .

## References

- [1] Anzengruber S W and Ramlau R. *Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators*. Inverse Problems, 2010, 26: 025001.
- [2] Beck A and Teboulle M. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*. SIAM Journal on Imaging Sciences, 2009, 2: 183–202.
- [3] Becker S, Bobin J and Cands E J. *NESTA: A fast and accurate first-order method for sparse recovery*. SIAM Journal on Imaging Sciences, 2011, 4: 1–39.
- [4] Benning M and Burger M. *Modern regularization methods for inverse problems*. Acta Numerica, 2018, 1–111.
- [5] van den Berg E and Friedlander M P. *Probing the pareto frontier for basis pursuit solutions*. SIAM Journal on Scientific Computing, 2008, 31: 890–912.
- [6] Blumensath T and Davies M E. *Iterative thresholding for sparse approximations*. Journal of Fourier Analysis and Applications, 2008, 14: 629–654.
- [7] Blumensath T and Davies M E. *Iterative hard thresholding for compressed sensing*. Applied and Computational Harmonic Analysis, 2009, 27: 265–274.
- [8] Bonesky T. *Morozov's discrepancy principle and Tikhonov-type functionals*. Inverse Problems, 2009, 25: 015015.
- [9] Bredies K and Lorenz D A. *Iterated hard shrinkage for minimization problems with sparsity constraints*. SIAM Journal on Scientific Computing, 2008, 30: 657–683.
- [10] Chambolle A and Dossal C. *On the convergence of the iterates of the “fast iterative shrinkage/thresholding algorithm”*. Journal of Optimization Theory and Applications, 2015, 166: 968–982.
- [11] O'Donoghue B and Candès E. *Adaptive restart for accelerated gradient schemes*. Foundations of Computational Mathematics, 2015, 15: 715–732.
- [12] Daubechies I, Defrise M and De Mol C. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*. Communications on Pure and Applied Mathematics, 2004, 57(11): 1413–1457.
- [13] Daubechies I, Defrise M, and De Mol C. *Sparsity-enforcing regularisation and ISTA revisited*. Inverse Problems, 2016, 32: 104001.
- [14] Daubechies I, Fornasier M and Loris I. *Accelerated projected gradient method for linear inverse problems with sparsity constraints*. Journal of Fourier Analysis and Applications, 2008, 14: 764–792.

- [15] Ding L and Han W.  $\alpha\ell_1 - \beta\ell_2$  regularization for sparse recovery, *Inverse Problems*, 2019, 35: 125009.
- [16] Engl H W, Hanke M, and Neubauer A. *Regularization of Inverse Problems*. Mathematics and its Applications vol 375: Dordrecht: Kluwer, 1996.
- [17] Figueiredo M, Nowak R and Wright S. *Gradient projection for sparse reconstruction Application to compressed sensing and other inverse problems*. *IEEE Journal of Selected Topics in Signal Processing*, 2007, 1: 586–597.
- [18] Fornasier M, eds. *Theoretical Foundations and Numerical Methods for Sparse Recovery*. De Gruyter, 2010.
- [19] Fornasier M, Peter S, Rauhut H and Worm S. *Conjugate gradient acceleration of iteratively re-weighted least squares methods*. *Computational Optimization and Applications*, 2016, 35: 205–259.
- [20] Fornasier M and Rauhut H. *Iterative thresholding algorithms*. *Applied and Computational Harmonic Analysis*, 2008, 25: 187–208.
- [21] Ge H, Wen J and Chen W. 2018 *The null space property of the truncated  $\ell_{1-2}$ -minimization*. *IEEE Signal Process. Lett.*, 2018, 8: 1261–1265.
- [22] Hansen P C. *Regularization Tools Version 4.0 for Matlab 7.3*. *Numerical Algorithms*, 2007, 46: 189–194.
- [23] Huang X, Shi L, and Yan M. *Nonconvex sorted  $\ell_1$  minimization for sparse approximation*. *J. Oper. Res. Soc. China*, 2015, 3: 207–229.
- [24] Jin B and Maass P. *Sparsity regularization for parameter identification problems*. *Inverse Problems*, 2012, 28(12): 123001.
- [25] Jin B, Maass P, and Scherzer O. *Sparsity regularization in inverse problems*. *Inverse Problems*, 2017, 33: 060301.
- [26] Lazzaro D, Piccolomini E L, and Zama F. *A nonconvex penalization algorithm with automatic choice of the regularization parameter in sparse imaging*. *Inverse Problems*, 2019, 35: 084002.
- [27] Li P, Chen W, Ge H, and K Ng M.  *$\ell_1$ - $\alpha\ell_2$  minimization methods for signal and image reconstruction with impulsive noise removal*. *Inverse Problems*, 2020, 36: 055009.
- [28] Loris I, Bertero M, De Mol C, Zanella R and Zanni L. *Accelerating gradient projection methods for  $\ell_1$ -constrained signal recovery by steplength selection rules*. *Applied and Computational Harmonic Analysis*, 2009, 27: 247–254.
- [29] Loris I and Verhoeven C. *On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty*. *Inverse Problems*, 2011, 27: 125007

- [30] Lou Y and Yan M. *Fast  $L_1$ - $L_2$  minimization via a proximal operator*. Journal of Scientific Computing, 2018, 74: 767–785.
- [31] Montefusco L B, Lazzaro D, and Papi S. *A fast algorithm for nonconvex approaches to sparse recovery problems*. Signal Proc., 2013, 93: 2636–2647.
- [32] Nesterov Y. *Smooth minimization of non-smooth functions*. Mathematical Programming, 2005, 103: 127–152.
- [33] Osher S, Burger M, Goldfarb D, Xu J and Yin W. *An iterative regularization method for total variation-based image restoration*. Multiscale Modeling & Simulation, 2005, 4: 460–489.
- [34] Ramlau R. *Morozovs discrepancy principle for Tikhonov regularization of nonlinear operators*. Numer. Funct. Anal. and Opt., 2002, :23: 147–172.
- [35] Ramlau R and Zarzer C A. *On the minimization of a Tikhonov functional with a non-convex sparsity constraint*. Electronic Transactions on Numerical Analysis, 2012, 39: 476–507.
- [36] Rockafellar R T. *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [37] Rockafellar R T and Wets R J-B. *Variational Analysis*, Berlin: Springer, 1998.
- [38] Scherzer O. *The use of Morozovs discrepancy principle for Tikhonov regularization for solving non-linear ill-posed problems*. SIAM J. Numer. Anal., 1993, 30: 1796–1838.
- [39] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F. *Variational Methods in Imaging*. Applied Mathematical Sciences, Vol.167, Newyork: Springer, 2009.
- [40] Teschke G and Borries C. *Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints*. Inverse Problems, 2010, 26: 025007.
- [41] Tibshirani R. *Regression shrinkage and selection via the Lasso*. Journal of the Royal Statistical Society: Series B, 1996, 58: 267–288.
- [42] Tikhonov A N and Arsenin V Y. *Solutions of Ill-posed Problems*. Washington, DC: V. H. Winston & Sons, 1977.
- [43] Tikhonov A N, Leonov A S and Yagola A G. *Nonlinear Ill-posed Problems*. London: Chapman & Hall, 1998.
- [44] Wang W, Lu S, Mao H, and Cheng J. *Multi-parameter Tikhonov regularization with the  $\ell_0$  sparsity constrain*. Inverse Problems, 2013, 29: 065018.
- [45] Wright S, Nowak R and Figueiredo M. *Sparse reconstruction by separable approximation*. IEEE Transactions on Signal Processing, 2009, 57: 2479–2493.
- [46] Yan L, Shin Y and Xiu D. *Sparse approximation using  $\ell_1$ - $\ell_2$  minimization and its application to stochastic collocation*. SIAM Journal of Scientific Computing, 2017, 39: A229-254.

- [47] Yin P, Lou Y, He Q, and Xin J. *Minimization of  $\ell_{1-2}$  for compressed sensing*. SIAM Journal on Scientific Computing, 2015, 37(1): A536–A563.
- [48] Zeidler E. *Nonlinear Functional Analysis and its Application*. Volume 3, New York: Springer, 1985.