Inverse Problems

PAPER

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

To cite this article: Liang Ding and Weimin Han 2020 *Inverse Problems* **36** 125012

IOP Publishing Inverse Problems

Inverse Problems **36** (2020) 125012 (30pp) https://doi.org/10.1088/1361-6420/abc857

A projected gradient method for $\alpha\ell_1-\beta\ell_2$ sparsity regularization[**]

Liang Ding[1,*] and **Weimin Han**[2]

[1] Department of Mathematics, Northeast Forestry University, Harbin 150040, People's Republic of China
[2] Department of Mathematics, University of Iowa, Iowa City, IA 52242, United States of America

E-mail: dl@nefu.edu.cn and weimin-han@uiowa.edu

Received 31 July 2020, revised 16 October 2020
Accepted for publication 6 November 2020
Published 3 December 2020

Abstract

The non-convex $\alpha\|\cdot\|_{\ell_1}-\beta\|\cdot\|_{\ell_2}$ $(\alpha\geqslant\beta\geqslant 0)$ regularization is a new approach for sparse recovery. A minimizer of the $\alpha\|\cdot\|_{\ell_1}-\beta\|\cdot\|_{\ell_2}$ regularized function can be computed by applying 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 solve problems with the classical ℓ_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 $\alpha\|\cdot\|_{\ell_1}-\beta\|\cdot\|_{\ell_2}$ sparsity regularization. Moreover, we discuss a strategy to determine the radius R of the ℓ_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

(Some figures may appear in colour only in the online journal)

1. Introduction

In this paper, we consider to solve an ill-posed operator equation of the form

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

[*]Author to whom any correspondence should be addressed.

[**]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).

where x is sparse, $A : \ell_2 \to Y$ is a linear and bounded operator from the ℓ_2 space to a Banach space Y. Norms in ℓ_2 and Y are denoted by $\|\cdot\|_{\ell_2}$ and $\|\cdot\|_Y$, respectively. Note that in the theory of classical Tikhonov regularization, Y is assumed to be a Hilbert space. Nevertheless, in the theory of modern nonlinear regularizations, e.g. sparsity regularization, total variation regularization, etc, one can discuss regularization properties in the more general setting of a Banach space ([4]). So for generality, we let Y be a Banach space here. In practice, the right-hand side y is known only approximately with an error up to a level $\delta \geqslant 0$. Therefore, we assume that we know $y^\delta \in Y$ with $\|y^\delta - y\|_Y \leqslant \delta$ for a given $\delta \geqslant 0$. The most commonly adopted technique to solve problem (1.1) is the ℓ_p-norm sparsity regularization with $1 \leqslant p < 2$, cf the monographs [17, 37] and the special issues [4, 13, 23, 24] for many developments on regularization properties and minimization schemes. Since the ℓ_p-norm regularization with $1 \leqslant p < 2$ does not always provide the 'sparsest' solution, the non-convex ℓ_p-norm sparsity regularization with $0 \leqslant p < 1$ has been proposed as alternatives. Iterative hard thresholding algorithms have been developed in [6, 7, 9, 19] for the ℓ_0 sparsity regularization. We refer the reader to [22, 25, 30] for some other types of alternatives to the ℓ_0-norm.

A non-convex regularization term of the form $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ ($\alpha \geqslant \beta \geqslant 0$) has attracted attention in the area of sparse recovery over the last five years, see [15, 26, 29, 44, 45] and references therein. In [15], we investigated the well-posedness and convergence rate of the non-convex sparsity regularization problem

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

where $x \in \ell_2$ space and

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

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

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

where

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

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

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

for problem (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 $s^k = 1$. An ISTA of the form

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

was first proposed in [12] to solve the classical ℓ_1 sparsity regularization problem

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

As an alternative to the ℓ_p-norm with $0 \leqslant p < 1$, the function $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ ($\alpha \geqslant \beta \geqslant 0$) has the desired property that it is a good approximation to a constant multiple of the ℓ_0-norm. The function has a simpler structure than the ℓ_0-norm from the perspective of computation. The ST-($\alpha\ell_1 - \beta\ell_2$) algorithm can easily be implemented, see [15, 20, 45] for several other algorithms for $\|\cdot\|_{\ell_1} - \|\cdot\|_{\ell_2}$ sparsity regularization. However, in general, the ST-($\alpha\ell_1 - \beta\ell_2$) algorithm, 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 is a popular research topic and some faster algorithms have been proposed. Several accelerated projected gradient methods can be found in [5, 14, 16, 43]. A comparison among several accelerated algorithms, including the 'fast ISTA' ([2]), is provided in [27]. Applying a smoothing technique from Nesterov ([31]), a fast and accurate first-order method is proposed in [3] to solve large-scale compressed sensing problems. 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 [32], a new iterative regularization procedure for inverse problems based on the use of Bregman distances is studied. Numerical results show that the proposed method offers significant improvement over the standard method. An explicit algorithm based on a primal-dual approach for the minimization of an ℓ_1-penalized least-squares function with a non-separable ℓ_1 term is proposed in [28]. An iteratively reweighted least squares algorithm and the corresponding convergence analysis for the regularization of linear inverse problems with sparsity constraints are investigated in [18]. For a projected gradient method of nonlinear ill-posed problems, see [38].

Unfortunately, the algorithms cited in the previous paragraph are limited to the classical ℓ_1-norm sparsity regularization. Though there is a 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 [34], the problem of minimizing a general continuously differentiable function subject to $\|x\|_0 \leqslant s$ is treated, where $s > 0$ is an integer, and $\|x\|_0$ is the ℓ_0-norm of x, which stands for the number of nonzero components in x. In this paper, we extend the projected gradient (PG) 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 ℓ_1-norm penalty, and it takes many steps to re-correct back. In other words, the algorithm generates a path $\{x_n | n \in \mathbb{N}\}$ that is initially fully contained in an ℓ_1-ball $B_R := \{x \in \ell_2 | \|x\|_{\ell_1} \leqslant R\}$. 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

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

1.2. Contribution and organization

In [15], we introduced the non-convex $\alpha\ell_1 - \beta\ell_2$ sparsity regularization method, with the primary interest in the regularization properties. In addition, we proposed an ST-($\alpha\ell_1 - \beta\ell_2$)

algorithm for the $\alpha\ell_1 - \beta\ell_2$ sparsity regularization. Though its formulation is simple and it can easily be implemented, ST-($\alpha\ell_1 - \beta\ell_2$) algorithm can be arbitrarily slow and it is computationally intensive. So, the aim of this paper is utilizing the PG method to solve $\alpha\ell_1 - \beta\ell_2$ sparsity regularization problems. Though there is a great potential for the PG method in solving the non-convex $\alpha\ell_1 - \beta\ell_2$ sparsity regularization problems, little result is available in the literature except the reference [34] where the PG method is utilized to solve ℓ_0 regularization problems. 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 problem (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 solving the optimization problem

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

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

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

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 the ℓ_1-ball should be chosen as $R = \|x^\delta_{\alpha,\beta}\|_{\ell_1}$ ([14]), where $x^\delta_{\alpha,\beta}$ is a minimizer of problem (1.2). However, in general, one cannot obtain the value of $\|x^\delta_{\alpha,\beta}\|_{\ell_1}$ 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 δ and the observed data y^δ. Moreover, we investigate the well-posedness of problem (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 $q = 2$ and $Y = \mathbb{R}^m$, problem (1.2) takes the form

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

where $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^m$. In the following, we remove the ℓ_1 constraint in (1.9) by considering a constrained optimization problem with an ℓ_1-ball constraint for a certain radius R. In analogy with the techniques on the projection in [14, 39], a natural strategy is to consider the constrained optimization problem

$$\begin{aligned} &\min \frac{1}{2}\|Ax - y^\delta\|_{\ell_2}^2 \quad \text{subject to } x \in B'_R \\ &\quad := \{x \in \mathbb{R}^n \,|\, \|x\|_{\ell_1} - \eta\|x\|_{\ell_2} \leqslant R\}, \quad 1 \geqslant \eta \geqslant 0. \end{aligned} \tag{1.10}$$

However, since B'_R is non-convex, it is challenging to analyze and solve this constrained optimization problem. We can remove the ℓ_1 term in (1.9) by considering instead the following optimization problem

$$\begin{aligned} &\min \mathcal{D}^\delta_\beta(x) = \frac{1}{2}\|Ax - y^\delta\|_{\ell_2}^2 - \beta\|x\|_{\ell_2} \quad \text{subject to} \\ &x \in B_R := \{x \in \mathbb{R}^n \,|\, \|x\|_{\ell_1} \leqslant R\} \end{aligned} \tag{1.11}$$

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

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

for problem (1.11), where $\lambda > 0$ is subject to some conditions, cf assumption 4.6 below.

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 projected gradient method. In section 3, we investigate an accelerated algorithm via the generalized conditional gradient method. Furthermore, we give a strategy to determine the radius R of the ℓ_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 discussing the accelerated algorithms, we briefly introduce some notation and results of the Tikhonov regularization and the PG method. Let

$$x^\delta_{\alpha,\beta} \in \arg\min_x\{\frac{1}{2}\|Ax - y^\delta\|^2_Y + \mathcal{R}_{\alpha,\beta}(x)\} \tag{2.1}$$

be a minimizer of the regularization function $\mathcal{J}^\delta_{\alpha,\beta}(x)$ in (1.2) with $q = 2$ and $\alpha \geqslant \beta \geqslant 0$. We denote by $\mathcal{L}^\delta_{\alpha,\beta}$ the set of all such minimizers $x^\delta_{\alpha,\beta}$, and by $x^\delta_{R,\beta}$ a solution of problem (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

$$Ax^\dagger = y \quad \text{and} \quad \mathcal{R}_\eta(x^\dagger) = \min\{\mathcal{R}_\eta(x)|x \in \ell_2, Ax = y\}.$$

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

Definition 2.2. An element $x \in \ell_2$ is called sparse if $\text{supp}(x) := \{i \in \mathbb{N}|x_i \neq 0\}$ is finite, where x_i is the ith component of x. $\|x\|_0 :=$ supp(x) is the cardinality of 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). Given $1 < \tau_1 \leqslant \tau_2$, choose $\alpha = \alpha(\delta, y^\delta) > 0$ such that

$$\tau_1\delta \leqslant \|Ax^\delta_{\alpha,\beta} - y^\delta\|_Y \leqslant \tau_2\delta \tag{2.2}$$

holds for some $x^\delta_{\alpha,\beta}$.

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 $e_i = (\underbrace{0, \ldots, 0, 1}_{i}, 0, \ldots)$, x_i is the ith component of x and

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

Definition 2.5. The projection onto the ℓ_1-ball is defined by

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

which gives the projection of an element $\hat{x}$ onto the ℓ_1-norm ball with radius $R > 0$.

We review two results from [14] on relations between the soft thresholding operator and the projection operator. For relations between the parameters α and R, see [14, figure 2]. For some countable index set Λ, denote $\ell_p = \ell_p(\Lambda)$, $1 \leqslant p < \infty$.

Lemma 2.6. *For any fixed $a \in \ell_2(\Lambda)$ and for $\tau > 0$, $\|\mathbb{S}_\alpha(a)\|_1$ is a piecewise linear, continuous, decreasing function of τ. Moreover, if $a \in \ell_1(\Lambda)$ then $\|\mathbb{S}_0(a)\|_1 = \|a\|_1$ and $\|\mathbb{S}_\alpha(a)\|_1 = 0$ for $\alpha \geqslant \max_i |a_i|$.*

Lemma 2.7. *If $\|a\|_{\ell_1} > R$, then the ℓ_2 projection of a on the ℓ_1 ball with radius R is given by $\mathbb{P}_R(a) = \mathbb{S}_\alpha(a)$ where α (depending on a and R) is chosen such that $\|\mathbb{S}_\alpha(a)\|_1 = R$. If $\|a\|_1 \leqslant R$ then $\mathbb{P}_R(a) = \mathbb{S}_0(a) = a$.*

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

Lemma 2.8. *For any $x \in \ell_2(\Lambda)$, $\mathbb{P}_R(x)$ is characterized as the unique vector in B_R such that*

$$\langle w - \mathbb{P}_R(x), x - \mathbb{P}_R(x)\rangle \leqslant 0, \quad \text{for all } w \in B_R.$$

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

$$\|\mathbb{P}_R(x) - \mathbb{P}_R(x')\| \leqslant \|x - x'\|$$

for all $x, y \in \ell_2(\Lambda)$.

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 (1.2) with $q = 2$ as

$$\min \mathcal{J}^\delta_{\alpha,\beta}(x) = F(x) + \Phi(x), \tag{3.1}$$

where

$$F(x) = \frac{1}{2}\|Ax - y^\delta\|^2_Y - \Theta(x),$$

$$\Phi(x) = \Theta(x) + \alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2},$$

$$\Theta(x) = \frac{\lambda}{2}\|x\|^2_{\ell_2} + \beta\|x\|_{\ell_2}, \quad \lambda > 0.$$

The ST-($\alpha\ell_1 - \beta\ell_2$) algorithm is stated in the form of algorithm 1. The first order optimization

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 $k = 0, 1, 2, \ldots$, do
 if $x^k = 0$ then
 $x^{k+1} = \operatorname{argmin} \frac{1}{2}\|Ax - y^\delta\|_Y^2 + \alpha\|x\|_{\ell_1}$
 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

condition of problem (3.1) is stated in lemma 3.1. Convergence of algorithm 1 is given in theorem 3.2; see [15, theorem 3.5] for its proof.

Lemma 3.1. *Let $0 \neq \hat{x} \in \ell_2$ be the minimizer of $\mathcal{J}^\delta_{\alpha,\beta}(x)$ in (3.1). Then*

$$\langle F'(\hat{x}), z - \hat{x}\rangle \geqslant \Phi(\hat{x}) - \Phi(z) \quad \textit{for all } z \in \ell_2 \tag{3.2}$$

This condition is equivalent to

$$\langle F'(\hat{x}), \hat{x}\rangle + \Phi(\hat{x}) = \min_{z\in\ell_2}\langle F'(\hat{x}), z\rangle + \Phi(z). \tag{3.3}$$

Proof. Since $0 \neq \hat{x} \in \ell_2$ is the minimizer of $\mathcal{J}^\delta_{\alpha,\beta}(x)$ in (3.1),

$$\begin{aligned} F(\hat{x}) + \Phi(\hat{x}) &\leqslant F(\hat{x} + t(z - \hat{x})) + \Phi(\hat{x} + t(z - \hat{x})) \\ &\leqslant F(\hat{x} + t(z - \hat{x})) + (1 - t)\Phi(\hat{x}) + t\Phi(z) \end{aligned}$$

for all $z \in \ell_2$ and $t \in [0, 1]$. Hence, we have

$$\Phi(\hat{x}) - \Phi(z) \leqslant \frac{F(\hat{x} + t(z - \hat{x})) - F(\hat{x})}{t}. \tag{3.4}$$

Taking the $\lim_{t\to 0^+}$ on (3.4), this implies that (3.2) holds. □

Theorem 3.2. *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 functional $\mathcal{J}^\delta_{\alpha,\beta}(x)$.*

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

$$\min \mathcal{C}^\delta_{\alpha,\beta,\lambda}(z, x^k) = \langle A^*(Ax^k - y^\delta) - \lambda x^k - \frac{\beta x^k}{\|x^k\|_{\ell_2}}, z\rangle + \frac{\lambda}{2}\|z\|_{\ell_2}^2 + \alpha\|z\|_{\ell_1}. \tag{3.5}$$

We may solve problem (3.5) by ([15])

$$z^k = \mathbb{S}_{\alpha/\lambda}\left(\left(\frac{\beta}{\lambda\|x^k\|_{\ell_2}} + 1\right)x^k - \frac{1}{\lambda}A^*(Ax^k - y^\delta)\right). \tag{3.6}$$

However, the iteration (3.6) is known to converge quite slowly. To accelerate the ST-($\alpha\ell_1 - \beta\ell_2$) algorithm, we transform problem (3.5) to an ℓ_1-ball constraint optimization problem of the form

$$\begin{cases} \min \mathcal{D}^{\delta}_{\beta,\lambda}(z, x^k) = \langle A^*(Ax^k - y^\delta) - \lambda x^k - \dfrac{\beta x^k}{\|x^k\|_{\ell_2}}, z\rangle + \dfrac{\lambda}{2}\|z\|^2_{\ell_2}, \quad \beta \geqslant 0, \\ \text{subject to } \ell_1 \text{ ball} \quad B_R := \{z \in \ell_2 | \|z\|_{\ell_1} \leqslant R\}. \end{cases} \tag{3.7}$$

Since $\mathcal{C}^{\delta}_{\alpha,\beta,\lambda}(z, x^k)$, $\mathcal{D}^{\delta}_{\beta,\lambda}(z, x^k)$ and B_R are convex with respect to the variable z, problem (3.7) is equivalent to problem (3.5) for a certain R ([35, theorem 27.4], [46, theorem 47.E]). In lemma 3.3, we show that equation (3.7) can be solved by a PG method of the form

$$z^k = \mathbb{P}_R\left(x^k + \frac{\beta x^k}{\lambda\|x^k\|_{\ell_2}} - \frac{1}{\lambda}A^*(Ax^k - y^\delta)\right). \tag{3.8}$$

Lemma 3.3. *An element $\hat{z} \in B_R$ is a minimizer of problem (3.7) if and only if*

$$\hat{z} = \mathbb{P}_R\left(x^k + \frac{\beta x^k}{\lambda\|x^k\|_{\ell_2}} - \frac{1}{\lambda}A^*(Ax^k - y^\delta)\right) \tag{3.9}$$

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

$$\left\langle x^k + \frac{\beta x^k}{\lambda\|x^k\|_{\ell_2}} - \frac{1}{\lambda}A^*(Ax^k - y^\delta) - \hat{z}, z - \hat{z}\right\rangle \leqslant 0 \quad \forall z \in B_R. \tag{3.10}$$

Proof. Note that $\hat{z} \in B_R$ is a solution of problem (3.7) if and only if for any $z \in B_R$, the function $f(t) = \mathcal{D}^{\delta}_{\beta,\lambda}((1-t)\hat{z} + tz, x^k)$ of $t \in [0, 1]$ attains its minimum at $t = 0$. Since $f(t)$ is quadratic and convex, a necessary and sufficient condition for $f(0) = \min_{0\leqslant t\leqslant 1} f(t)$ is $f'(0+) \geqslant 0$. 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 $f'(0+) \geqslant 0$ is equivalent to the inequality (3.10). □

The PG algorithm for problem (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 problem (3.5) is equivalent to problem (3.7) for a certain R. Before starting iteration (3.8), 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 ℓ_1-ball constraint by Morozov's discrepancy principle.

By lemma 2.7, for a given α in (3.5), R in (3.7) should be chosen such that $R = \|x^{\delta}_{\alpha,\beta}\|_{\ell_1}$. However, one does not know the value of $\|x^{\delta}_{\alpha,\beta}\|_{\ell_1}$ before starting the PG method (3.8). Of course, we can find an approximation of $x^{\delta}_{\alpha,\beta}$ 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 determine if a value of R is appropriate for problem (3.7). Recall that there exists a regularization parameter α depending on R such that problem (3.5) is equivalent to problem (3.7). So to determine an appropriate value of R, we may turn to check if the value of α is appropriate. One criterion is to check whether $\delta = O(\alpha)$. If $\delta = O(\alpha)$, then

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 $k = 0, 1, 2, \ldots,$ do
if $x^k = 0$ then
$x^{k+1} = \text{argmin}\frac{1}{2}\|Ax - y^\delta\|_Y^2 + \alpha\|x\|_{\ell_1}$
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

$x^\delta_{\alpha,\beta}$ is a regularized solution ([15, theorem 2.13]). However, by lemmas 2.6 and 2.7, we only know that α is a piecewise linear, continuous, decreasing function of R (see [14, figure 2]), and there is no explicit formula relating α and R. We cannot determine the value of α from the value of R directly. So we cannot ensure whether R is appropriate.

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

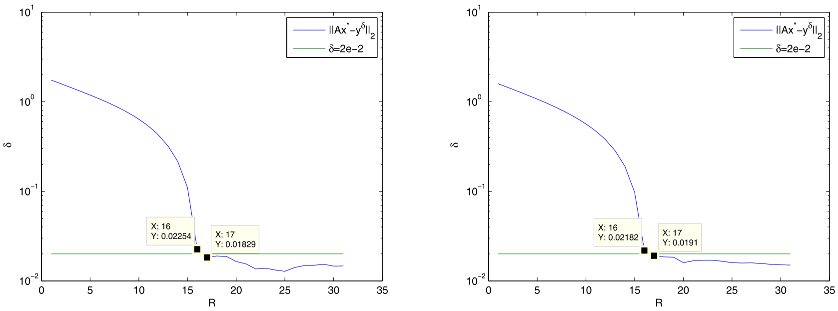
$$\tau_1\delta \leqslant \|Ax^\delta_{\alpha,\beta} - y^\delta\|_Y \leqslant \tau_2\delta, \quad 1 < \tau_1 \leqslant \tau_2.$$

For any fixed R, we need to compute $x^\delta_{\alpha,\beta}$ by algorithm 2 where z^k is determined by the PG method (3.8). Subsequently, we check whether $x^\delta_{\alpha,\beta}$ satisfies Morozov's discrepancy principle (2.2). For this strategy, we only need to know the observed data y^δ and the noise level δ. By lemma 3.7, the discrepancy $\|Ax^\delta_{\alpha,\beta} - y^\delta\|_Y$ is an increasing function of α. A commonly adopted technique is to try $\alpha_j = \alpha/2^j$, $j = 1, 2, \ldots$. We start with $j = 1$ and the value of j is increased by 1 each time until the calculated solution $x^\delta_{\alpha,\beta}$ satisfies ([40])

$$\tau_1\delta \leqslant \|Ax^\delta_{\alpha,\beta} - y^\delta\|_Y \leqslant \tau_2\delta.$$

Since α is a decreasing function of R, the discrepancy $\|Ax^\delta_{\alpha,\beta} - y^\delta\|_Y$ is a decreasing function of R, see lemma 2.6 and figure 1. So an alternative is to begin with a small value of R such that $x^\delta_{\alpha,\beta}$ satisfies Morozov's discrepancy principle (2.2). Subsequently, we increase the value of R by a fixed amount $c \in \mathbb{Z}^+$ until $x^\delta_{\alpha,\beta}$ fails to satisfy Morozov's discrepancy principle. Then we choose the maximal value of such 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 Morozov's discrepancy principle (2.2) is satisfied. Under Morozov's discrepancy principle, the PG algorithm for problem (1.2) based on GCGM is stated in the form of algorithm 3. A natural stopping criterion for the inner iteration in algorithm 3 is the change of the iterative solution. If it does not change for three subsequent iterations or the maximum number of 2000 iterations is reached, then a minimizer is considered to have been obtained.

Morozov's discrepancy principle is a method for determining the regularization parameter α. If it is chosen appropriately, then problem (1.2) is a regularization method. So, a natural question is whether problem (1.2) is a regularization method where the regularization parameter α is



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

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

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 $j = 0, 1, 2, \ldots,$ do
  for $k = 0, 1, 2, \ldots,$ do
    if $x^k = 0$ then
      $x^{k+1} = \operatorname{argmin} \frac{1}{2}\|Ax - y^\delta\|_Y^2 + \alpha\|x\|_{\ell_1}$
    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
check stopping criteria and return x^{k+1} as a solution or set $k = k + 1$
end for
    if Morozov's discrepancy principle (2.2) is not satisfied, set $R_{j+1} = R_j - c$, $c > 1$
otherwise stop iteration
end if
 $j = j + 1$
end for

determined by Morozov's discrepancy principle. It is known that 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, 33, 36, 40, 41]). If the regularized term is non-convex, some results can be found in [15, 42] 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 problem (1.2) is a regularization method. In this paper, we prove that if the non-convex regularization term satisfies some properties, e.g. coercivity, weakly lower semi-continuity and Radon–Riesz property, well-posedness of the regularized formulation remains valid.

We now consider the stability of algorithm 3. The stability analysis is restricted to the case of the finite dimensional space $\mathbb{R}^n$. If $\{x^k\}$ is bounded in ℓ_2, then $\{x^k\}$ has a weakly convergence subsequent $x^{k_j} \rightharpoonup x^*$. However, the challenge is that $x^{k_j} \rightharpoonup x^*$ does not ensure $\beta x^{k_j}/\|x^{k_j}\|_{\ell_2} \to \beta x^*/\|x^*\|_{\ell_2}$.

Theorem 3.4. *Let $x_{\delta_n}^{k_i} \to x_{\delta_n}^*$, $x_{\delta_n}^* \neq 0$ being a stationary point of problem (3.1) where y^δ is replaced by y^{δ_n} and $\{x_{\delta_n}^{k_i}\}$ being a subsequence of $\{x_{\delta_n}^k\}$ generated by algorithm 3. Assume $\delta_n \to \delta$, as $n \to \infty$. Then there exists a subsequence of $\{x_{\delta_n}^*\}$, still denoted by $\{x_{\delta_n}^*\}$, such that $x_{\delta_n}^* \to x^*$. If $x^* \neq 0$, then x^* is a stationary point of problem (3.1). Furthermore, if x^* is unique, then $\lim_{n\to\infty}\|x_{\delta_n}^* - x^*\|_2 = 0$.*

Proof. By the definition of $x_{\delta_n}^*$ and lemma 3.1, we have

$$\langle F'(x_{\delta_n}^*), z - x_{\delta_n}^* \rangle \geqslant \Phi(x_{\delta_n}^*) - \Phi(z) \quad \text{for all } z \in \ell_2, \tag{3.11}$$

i.e.

$$\left\langle A^*(Ax_{\delta_n}^* - y^{\delta_n}) - \lambda x_{\delta_n}^* - \frac{\beta x_{\delta_n}^*}{\|x_{\delta_n}^*\|_{\ell_2}}, z - x_{\delta_n}^* \right\rangle$$
$$\geqslant \Phi(x_{\delta_n}^*) - \Phi(z) \quad \text{for all } z \in \ell_2. \tag{3.12}$$

Since $\{x_{\delta_n}^*\}$ is bounded, there exist a convergent subsequence of $\{x_{\delta_n}^*\}$, still denoted by $\{x_{\delta_n}^*\}$, and an element x^* such that $x_{\delta_n}^* \to x^*$. Taking the limit $n \to \infty$ in (3.12), we have

$$\left\langle A^*(Ax^* - y^\delta) - \lambda x^* - \frac{\beta x^*}{\|x^*\|_{\ell_2}}, z - x^* \right\rangle \geqslant \Phi(x^*) - \Phi(z) \quad \text{for all } z \in \ell_2. \tag{3.13}$$

By lemma 3.1, this implies that x^* is a stationary point of problem (3.1). If x^* is unique, then every subsequence $\{x_{\delta_n}^*\}$ converges to x^*. So we have $\lim_{n\to\infty}\|x_{\delta_n}^* - x^*\|_2 = 0$. □

3.2. Well-posedness of regularization

In this section, we discuss the well-posedness of problem (1.2) under Morozov's discrepancy principle. First, we show that there exists at least one regularization parameter α in (1.2) such that Morozov's discrepancy principle (2.2) holds. We recall some properties of $\mathcal{R}_{\alpha,\beta}(x)$, needed in analyzing the well-posedness of problem (1.2), cf [15, lemmas 2.5, 2.8 and 2.10] for proofs.

Lemma 3.5. *If $\alpha > \beta \geqslant 0$, the function $\mathcal{R}_{\alpha,\beta}(x)$ in (1.2) has the following properties:*

(*i*) (Coercivity)$\|x\|_{\ell_2} \to \infty$ *implies* $\mathcal{R}_{\alpha,\beta}(x) \to \infty$.

(*ii*) (Weaklowersemi − continuity). *Let $M > 0$ be given. Then, for any $x_n \in \ell_2$ with $\mathcal{R}_{\alpha,\beta}(x_n) \leqslant M$, $\{x_n\}$ weakly converging to x in ℓ_2 implies* $\liminf_n \mathcal{R}_{\alpha,\beta}(x_n) \geqslant \mathcal{R}_{\alpha,\beta}(x)$.

(*iii*) (Radon–Rieszproperty). *Let $M > 0$ be given. Then, for any $x_n \in \ell_2$ with $\mathcal{R}_{\alpha,\beta}(x_n) \leqslant M$, if x_n converges weakly to x in ℓ_2 and $\mathcal{R}_{\alpha,\beta}(x_n) \to \mathcal{R}_{\alpha,\beta}(x)$, then x_n converges strongly to x in ℓ_2.*

Definition 3.6. For fixed δ and $\eta \in [0, 1]$, define

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

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

Let us list some properties of $m(\alpha)$, $F(x_{\alpha,\beta}^{\delta})$ and $\mathcal{R}_\eta(x_{\alpha,\beta}^{\delta})$ in lemmas 3.7 and 3.8. Since $\mathcal{R}_\eta(x_{\alpha,\beta}^{\delta})$ is weakly lower semi-continuous, the proofs are similar to that in [45, 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 3.7. *The function $m(\alpha)$ is continuous and non-increasing, i.e., $\alpha_1 > \alpha_2$ implies $m(\alpha_1) \leqslant m(\alpha_2)$. Moreover, for $\alpha_1 > \alpha_2$,*

$$\sup_{x_{\alpha_1,\beta_1}^{\delta} \in \mathcal{L}_{\alpha_1,\beta_1}^{\delta}} F(x_{\alpha_1,\beta_1}^{\delta}) \leqslant \inf_{x_{\alpha_2,\beta_2}^{\delta} \in \mathcal{L}_{\alpha_2,\beta_2}^{\delta}} F(x_{\alpha_2,\beta_2}^{\delta}),$$

$$\sup_{x_{\alpha_1,\beta_1}^{\delta} \in \mathcal{L}_{\alpha_1,\beta_1}^{\delta}} \mathcal{R}_\eta(x_{\alpha_1,\beta_1}^{\delta}) \geqslant \inf_{x_{\alpha_2,\beta_2}^{\delta} \in \mathcal{L}_{\alpha_2,\beta_2}^{\delta}} \mathcal{R}_\eta(x_{\alpha_2,\beta_2}^{\delta}).$$

Lemma 3.8. *For each $\bar{\alpha} > 0$ there exist $x', x'' \in \mathcal{L}_{\bar{\alpha},\bar{\beta}}^{\delta}$ such that*

$$\lim_{\alpha\to\bar{\alpha}^-} \left(\sup_{x_{\alpha,\beta}^{\delta} \in \mathcal{L}_{\alpha,\beta}^{\delta}} F(x_{\alpha,\beta}^{\delta}) \right) = F(x') = \inf_{x \in \mathcal{L}_{\bar{\alpha},\bar{\beta}}^{\delta}} F(x),$$

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

We now provide an existence result on the regularization parameter α, which can be proved similar to that in [1, 33] on Morozov's discrepancy principle for nonlinear ill-posed problems.

Lemma 3.9. *Assume $0 <_{\tau 2} \delta < \|y^\delta\|_Y$. Then there exist $\alpha_1, \alpha_2 \in \mathbb{R}^+$ such that*

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

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

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

This implies that there exists a small enough α_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 $\alpha_n \to \infty$. Then

$$\mathcal{R}_\eta(x_n) \leqslant \frac{1}{\alpha_n} m(\alpha_n) \leqslant \frac{1}{\alpha_n} \|A0 - y^\delta\|_Y \to 0. \tag{3.14}$$

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

$$\mathcal{R}_\eta(x) = (1-\eta)\,\|x\|_{\ell_1} + \eta\,(\|x\|_{\ell_1} - \|x\|_{\ell_2}). \tag{3.15}$$

Then a combination of (3.14) and (3.15) 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 3.5(ii), it follows from (3.14) that

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

By (3.15), this implies $x^* = 0$. Since $x_n \rightharpoonup 0$ and $\mathcal{R}_\eta(x_n) \to \mathcal{R}_\eta(0)$, lemma 3.5(iii) implies that $x_n \to 0$. Then

$$\|Ax_n - y^\delta\|_Y \to \|A0 - y^\delta\|_Y = \|y^\delta\|_Y > \tau_2\delta.$$

So there exists a large enough α_2 such that $\inf_{x^\delta_{\alpha_2,\beta_2}\in\mathcal{L}^\delta_{\alpha_2,\beta_2}} F(x^\delta_{\alpha_2,\beta_2}) > \tau_2\delta$. □

Note that we require $\|y^\delta\|_Y >_{\tau 2} \delta$ in lemma 3.9, 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 theorem 3.10 in [1]. The proof makes use of the properties stated in lemmas 3.8 and 3.9.

Theorem 3.10. *Assume $\|y^\delta\|_Y >_{\tau 2} \delta > 0$ and there is no $\alpha > 0$ with minimizers $x', x'' \in \mathcal{L}^\delta_{\alpha,\beta}$ such that*

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

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

Next, we give a convergence result for the $\alpha\ell_1 - \beta\ell_2$ regularization problem (1.2) under Morozov's discrepancy principle.

Theorem 3.11 (Convergence). *Let $x^{\delta_n}_{\alpha_n,\beta_n}$ be a minimizer of $\mathcal{J}^{\delta_n}_{\alpha_n,\beta_n}(x)$ defined by (2.1) with the data y^{δ_n} satisfying $\|y - y^{\delta_n}\| \leqslant \delta_n$, where $\delta_n \to 0$ if $n \to +\infty$ and y^{δ_n} belongs to the range of A. Let α_n be determined by Morozov's discrepancy principle,*

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

Moreover, assume that $\eta = \lim_{n\to\infty}\eta_n \in [0, 1)$ exists, where $\eta_n = \beta_n/\alpha_n$. Then there exists a subsequence of $\{x^{\delta_n}_{\alpha_n,\beta_n}\}$, denoted by $\{x^{\delta_{n_k}}_{\alpha_{n_k},\beta_{n_k}}\}$, that converges to an $\mathcal{R}_\eta$-minimizing solution $x^\dagger$ in ℓ_2. If, in addition, the $\mathcal{R}_\eta$-minimizing solution $x^\dagger$ is unique, then

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

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

$$\begin{aligned}
&\frac{1}{q}\|Ax_n - y_n\|^q_Y + \alpha_n\|x_n\|_{\ell_1} - \beta_n\|x_n\|_{\ell_2} \\
&\quad\leqslant \frac{1}{q}\|Ax^\dagger - y_n\|^q_Y + \alpha_n\|x^\dagger\|_{\ell_1} - \beta_n\|x^\dagger\|_{\ell_2} \\
&\quad\leqslant \frac{1}{q}\delta^q_n + \alpha_n\|x^\dagger\|_{\ell_1} - \beta_n\|x^\dagger\|_{\ell_2}.
\end{aligned} \tag{3.16}$$

Since $\tau_1\delta_n \leqslant \|Ax_n - y_n\|_Y$, it follows from (3.16) that

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

Then we have

$$\limsup_{n\to+\infty} \left(\|x_n\|_{\ell_1} - \eta_n\|x_n\|_{\ell_2}\right) \leqslant \|x^\dagger\|_{\ell_1} - \eta\|x^\dagger\|_{\ell_2}. \tag{3.17}$$

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 ℓ_2. By Morozov's discrepancy principle, we obtain

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

Therefore, weak lower semicontinuity of the norm gives

$$\|Ax^* - y\| \leqslant \liminf_{k\to\infty} \|Ax_{n_k} - y\|_Y = 0. \tag{3.18}$$

Meanwhile, by (3.17) and lemma 3.5(ii), we have

$$\begin{aligned} \|x^*\|_{\ell_1} - \eta\|x^*\|_{\ell_2} &\leqslant \liminf_{k} (\|x_{n_k}\|_{\ell_1} - \eta_{n_k}\|x_{n_k}\|_{\ell_2}) \\ &\leqslant \limsup_{k} (\|x_{n_k}\|_{\ell_1} - \eta_{n_k}\|x_{n_k}\|_{\ell_2}) \\ &\leqslant \limsup_{n} (\|x_n\|_{\ell_1} - \eta_n\|x_n\|_{\ell_2}) \leqslant \|x^\dagger\|_{\ell_1} - \eta\|x^\dagger\|_{\ell_2}. \end{aligned} \tag{3.19}$$

By the definition of $x^\dagger$, a combination of (3.18) and (3.19) implies that x^* is an $\mathcal{R}_\eta$-minimizing solution. Hence, $\lim_{k\to\infty} \mathcal{R}_\eta(x_{n_k}) \to \mathcal{R}_\eta(x^*)$. By lemma 3.5(iii), we have $x_{n_k} \to x^*$. 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 $\lim_{n\to+\infty} \|x_n - x^\dagger\|_{\ell_2} = 0$. □

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 a constant multiple of the ℓ_0-norm as $\beta/\alpha \to 1$. However, if $\alpha = \beta$, $\mathcal{R}_{\alpha,\alpha}(x)$ fails to satisfy the coercivity and the Radon–Riesz property, and we cannot ensure the convergence in ℓ_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 ℓ_2, i.e. $\|x\|_{\ell_2} \to \infty$ implies $\|Ax\|_Y \to \infty$, then the proof of the weak convergence is similar to that of theorem 3.11.

4. The projected gradient method via the surrogate function approach

In this section, we propose another projected gradient algorithm for problem (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 the optimization problem (1.11).

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

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

for any $\lambda > 0$, equivalently,

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

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,

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

i.e., (4.2) holds. □

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

Lemma 4.2. *For a given $\beta \geqslant 0$, define*

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

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

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

Then,

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

Since $w \mapsto \|w\|_{\ell_2}$ is convex, the matrix $(a_{i,j}(w))_{n\times n}$ is positive semi-definite. Thus, $\mathrm{eig}(w) \geqslant 0$, where $\mathrm{eig}(w)$ denotes the eigenvalues of $(a_{i,j}(w))_{n\times n}$. Moreover, $\max\{\mathrm{eig}(w)\}$ 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 $\beta \geqslant 0$ and a fixed nonzero $x \in B_R$, there exists $\lambda > 0$ such that $\lambda > \max\{\mathrm{eig}(\hat{w})\}$.*

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

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

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

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

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

In lemma 4.4, we assume $\lambda \geqslant \beta \max\{\mathrm{eig}(\hat{w})\}$. This condition is weaker than $\lambda \geqslant \max\{\mathrm{eig}(\hat{w})\}$. In general, the regularization parameter $\alpha \ll 1$ in the Tihkonov regularization. Since $\beta = \alpha\eta$ and $0 \leqslant \eta \leqslant 1$, we also have $\beta \ll 1$.

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

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

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

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

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

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

i.e.,

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

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

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

Define

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

If $w \neq 0$, we have

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

By (4.7), this implies that

$$0 \leqslant \langle J'(\hat{w}), w - \hat{w}\rangle = \lim_{t\to 0+}\frac{J(\hat{w} + t(w-\hat{w})) - J(\hat{w})}{t}. \tag{4.8}$$

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

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

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

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

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

The projected gradient algorithm based on the surrogate function is stated in the form of algorithm 4. The stopping criterion for the inner iteration in algorithm 4 is the same as that

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

Choose $x^0 \in \mathbb{R}^n$, $R_0 \in \mathbb{R}^+$, $\beta = O(\delta)$ and λ such that $\lambda > \beta \max\{\mathrm{eig}(x^0), \mathrm{eig}(x^\dagger)\}$
for $j = 0, 1, 2, \ldots,$ do
  for $k = 0, 1, 2, \ldots,$ 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)
    check stopping criteria and return x^{k+1} as a solution or set $k = k+1$
  end for
    if Morozov's discrepancy principle (2.2) is not satisfied, set $R_{j+1} = R_j - c$, $c > 1$
     otherwise stop iteration
    end if
$j = j+1$
end for

in algorithm 3. If the iterative solution does not change for three subsequent iterations or the maximum number of 2000 iterations is reached, then a minimizer is considered to have been obtained.

To prove the convergence of algorithm 4, we impose some restrictions on the operator A and λ.

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

(A1) $\|Ax\|_{\ell_2}^2 \leqslant \frac{\lambda r}{2}\|x\|_{\ell_2}^2$ for all $x \in \ell_2$
(A2) $\lambda \geqslant \beta \max\{\mathrm{eig}(x^k)\}$ 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 $A_{m\times n}x_n = y_m \to \left(\frac{1}{c}A_{m\times n}\right)x_n = \frac{1}{c}y_m$ 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 $\mathrm{eig}(x^\dagger)$ and $\mathrm{eig}(x^0)$, then we can give an approximation for the order of the maximal value of $\mathrm{eig}(x^\dagger)$ and $\mathrm{eig}(x^0)$. Subsequently, we choose λ such that λ 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 $A_{m\times n}x_n = y_m \to \left(\frac{1}{c}A_{m\times n}\right)(cx_n) = y_m$ 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 iteration (4.9). Then,*

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

and

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

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

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

Furthermore,

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

This implies

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

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 $k \geqslant 1$. So in the following we let $x^k \neq 0$ whenever $k \geqslant 1$.

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,

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

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 $\hat{w}^{l_k+1} = \Psi(\hat{w}^{l_k})$,

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

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 cannot ensure uniqueness of the fixed point. Indeed, due to the non-convexity of $\Phi_\lambda(w,x)$ in (4.3), the minimizer of $\Phi_\lambda(w,x)$ 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. *Let $\{x^k\}$ be the sequence generated by algorithm 4. Then $\{x^k\}$ has a subsequence which converges to a nonzero stationary point x^* of (1.11), i.e. x^* satisfies*

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

where R satisfies Morozov's discrepancy principle (2.2).

Proof. First, we prove the convergence of inner iteration

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

For any fixed $R_j > 0$, since $\{x^k\} \subset B_{R_j}$ is bounded, $\{x^k\}$ has a subsequence $\{x^{k_i}\}$ converging to an element x^* in B_{R_j}, i.e. $x^{k_i} \to x^*$ in B_{R_j}. Since A is linear and bounded, $A(x^{k_i}) \to A(x^*)$. By lemma 2.8 and the definition of x^{k+1}, we see that, for all $w \in B_{R_j}$,

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

This implies that

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

Taking the limit $i \to \infty$ in (4.11), we have

$$\lim_{i\to\infty} \left\langle x^{k_i} + \frac{\beta x^{k_i+1}}{\lambda \|x^{k_i+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^{k_i} - y^\delta) - x^{k_i+1}, w - x^{k_i+1} \right\rangle \leqslant 0. \tag{4.12}$$

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

$$\lim_{i\to\infty} |\langle x^{k_i} - x^{k_i+1}, w - x^{k_i+1} \rangle| = 0. \tag{4.13}$$

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

$$\lim_{i\to\infty} \left\langle \frac{\beta x^{k_i+1}}{\lambda \|x^{k_i+1}\|_{\ell_2}} - \frac{1}{\lambda} A^*(Ax^{k_i} - y^\delta), w - x^{k_i+1} \right\rangle \leqslant 0. \tag{4.14}$$

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

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

By lemma 4.1, x^* is a stationary point of $\mathcal{D}_\beta^\delta(x)$ on B_{R_j}. Moreover, for the outer iteration, if R_j satisfies Morozov's discrepancy principle (2.2), we rewrite R_j as R. So x^* is a stationary point of $\mathcal{D}_\beta^\delta(x)$ on B_R, where R satisfies Morozov's discrepancy principle. This proves the theorem. □

Finally, we show the stability of algorithm 4.

Theorem 4.11. *Let $x_{\delta_n}^{k_i} \to x_{\delta_n}^*$, $x_{\delta_n}^*$ being a stationary point of problem (1.11) where y^δ is replaced by y^{δ_n} and $\{x_{\delta_n}^{k_i}\}$ being a subsequence of $\{x_{\delta_n}^k\}$ generated by algorithm 4. Assume $\delta_n \to \delta$ as $n \to \infty$. Then there exists a subsequence of $\{x_{\delta_n}^*\}$, still denoted by $\{x_{\delta_n}^*\}$, such that $x_{\delta_n}^* \to x^*$. If $x^* \neq 0$, then it is a stationary point of problem (1.11). If x^* is unique, then* $\lim_{n\to\infty}\|x_{\delta_n}^* - x^*\|_2 = 0$.

Proof. By the assumption, $x_{\delta_n}^*$ is a stationary point of problem (1.11) where y^δ is replaced by y^{δ_n}. Thus,

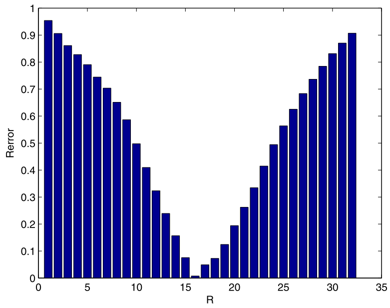
$$\left\langle \frac{\beta x_{\delta_n}^*}{\|x_{\delta_n}^*\|_{\ell_2}} - A^*(Ax_{\delta_n}^* - y^{\delta_n}), w - x_{\delta_n}^* \right\rangle \leqslant 0 \quad \forall\, w \in B_R. \tag{4.15}$$

Since $\{x_{\delta_n}^*\}$ is bounded, there exist a convergent subsequence of $\{x_{\delta_n}^*\}$, still denoted by $\{x_{\delta_n}^*\}$, and an element x^* such that $x_{\delta_n}^* \to x^*$. Taking the limit $n \to \infty$ in (4.15), we have

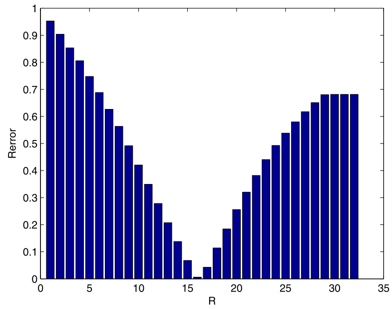
$$\left\langle \frac{\beta x^*}{\|x^*\|_{\ell_2}} - A^*(Ax^* - y^\delta), w - x^* \right\rangle \leqslant 0 \quad \forall\, w \in B_R. \tag{4.16}$$

By lemma 2.8, this implies that x^* is a stationary point of problem (1.11). If x^* is unique, then every subsequence $\{x_{\delta_n}^*\}$ converges to x^*. So we have $\lim_{n\to\infty}\|x_{\delta_n}^* - x^*\|_2 = 0$. □

Remark 4.12. 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 the case of ℓ_2 space. In theorem 4.10, if $\{x^k\}$ is defined in ℓ_2, then $\{x^k\}$ has a weak convergence subsequence $\{x^{k_j}\} \rightharpoonup x^*$. However, the challenge of the proof is that $x^{k_j} \rightharpoonup x^*$ cannot 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 $e_n = (\underbrace{0, \ldots, 0, 1}_{n}, 0, \ldots)$. Since $e_n \rightharpoonup 0$ in ℓ_2, $x_n \rightharpoonup x$ in ℓ_2. However, $\|x_n\|_{\ell_2} \nrightarrow \|x\|_{\ell_2}$. 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. $\|x_n\|_{\ell_2} \to \|x\|_{\ell_2}$, 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 $\|x_n\|_{\ell_2} \to \|x\|_{\ell_2}$ and $x_n \rightharpoonup x^*$ in ℓ_2 implies that $x_n \to x^*$. Moreover, the iterative algorithm in this paper has an implicit formulation, and we need to compute the iterative solution. However, in ℓ_2 space, we do not know whether the operator $\Phi(\hat{w})$ is weak-strong continuous. So we cannot ensure that the fixed point iteration is convergent.



(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. Rerror of reconstruction x^* with different values of η.

η	0.0	0.1	0.2	0.3	0.4	0.5	0.7	0.9	1.0
ST-$(\alpha\ell_1 - \beta\ell_2)$	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

5. Numerical examples

In this section, we present results from two numerical examples 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^*:

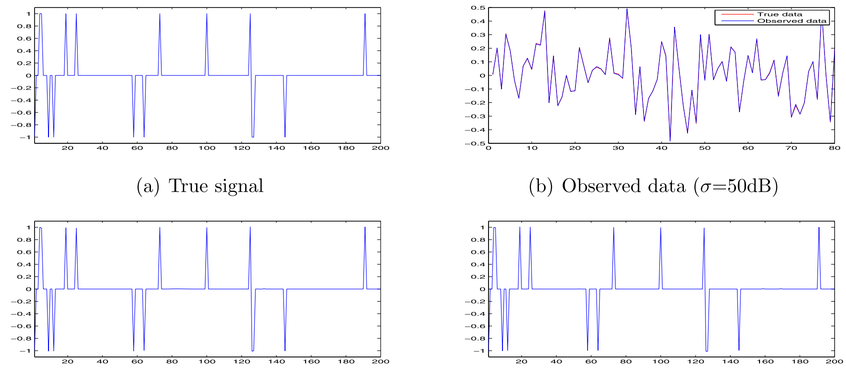
$$\text{Rerror} := \frac{\|x^* - x^\dagger\|_{\ell_2}}{\|x^\dagger\|_{\ell_2}},$$

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 ℓ_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 2.50 GHz workstation with 8 Gb RAM.

5.1. Example 1: compressive sensing

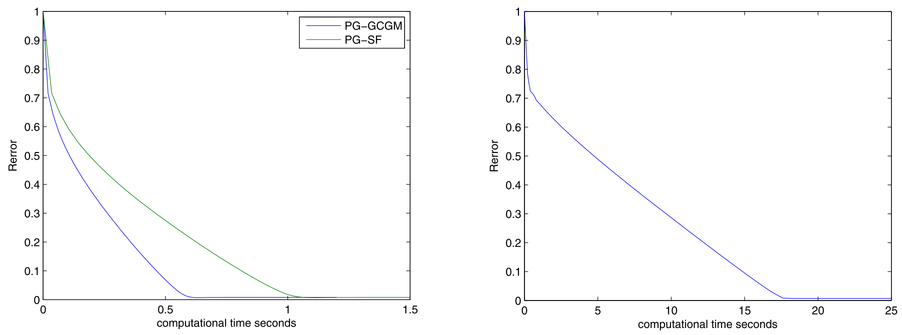
In the first example, we test compressive sensing with the commonly used random Gaussian matrix. The compressive sensing problem is defined as $A_{m\times n}x_n = y_m$, where $A_{m\times n}$ is a well conditioned random Gaussian matrix by calling $A = \text{randn(m, n)}$ in MATLAB. Exact data $y^\dagger$ 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 $\text{y}^\delta = \text{awgn}(Ax^\dagger, \sigma)$ in



(a) True signal (b) Observed data (σ=50dB)

(c) PG-GCGM algorithm, Rerror=0.0059 (d) PG-SF algorithm, Rerror=0.0089

Figure 3. True signal and its noisy observation together with reconstructions x^* for $\eta = 1$.

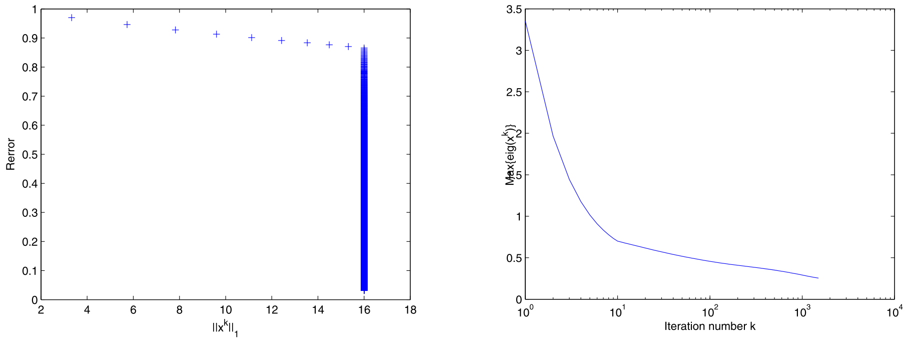


(a) PG-GCGM algorithm and PG-SF algorithm (b) ST-($\alpha\ell_1 - \beta\ell_2$) algorithm

Figure 4. (a) Convergence rate of PG-GCGM algorithm and PG-SF algorithm; (b) convergence rate of ST-($\alpha\ell_1 - \beta\ell_2$) algorithm.

MATLAB, where σ (measured in dB) measures the ratio between the true (noise free) data $y^\dagger$ or $Ax^\dagger$ and Gaussian noise. A larger value of σ corresponds to a smaller value of the noise level δ, where the noise level δ 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 $A_{m\times n} \to c * A_{m\times n}$, 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., λ, β 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 δ is added to exact data $y^\dagger$ by calling $\text{y}^\delta = \text{awgn}(Ax^\dagger, \sigma)$, where $\sigma = 50\text{dB}$, δ is around 0.02. We let $\lambda = 1$, $\eta = 1$ and the initial vector x^0 is generated by calling $x^0 = 0.01\text{ones}(n, 1)$. We choose the regularization parameter α by $\alpha = O(\delta) = 0.02$. Indeed, if $\alpha = O(\delta)$, we can prove $x^\delta_{\alpha,\beta} \to x^\dagger$. This implies that (1.9) is a regularization method, cf [15, theorem 2.13] for details. As for



(a) (b)

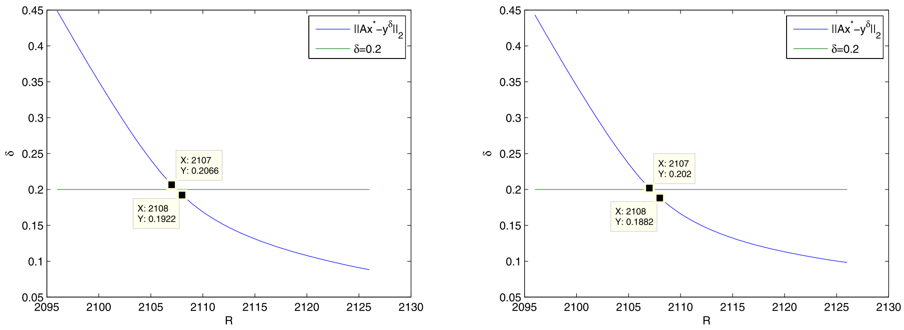
Figure 5. (a) Rerror for x^k, $1 \leqslant k \leqslant 1500$; (b) $\max\{\mathrm{eig}(x^k)\}$ for $1 \leqslant k \leqslant 1500$.

Table 2. Time of computation for the reconstruction x^* with different values of Rerror.

Rerror	ST-$(\alpha\ell_1 - \beta\ell_2)$ time (m)	PG-GCGM time (s)	PG-SF time (s)
0.8	9.7463	0.0214	0.0208
0.6	12.7113	0.1926	0.8573
0.4	14.9283	0.6995	3.2097
0.2	24.8903	1.6924	7.5099
0.1	39.2569	2.8578	11.1562
0.05	60.5784	4.9201	22.2830
0.02	102.8623	8.2870	41.2480

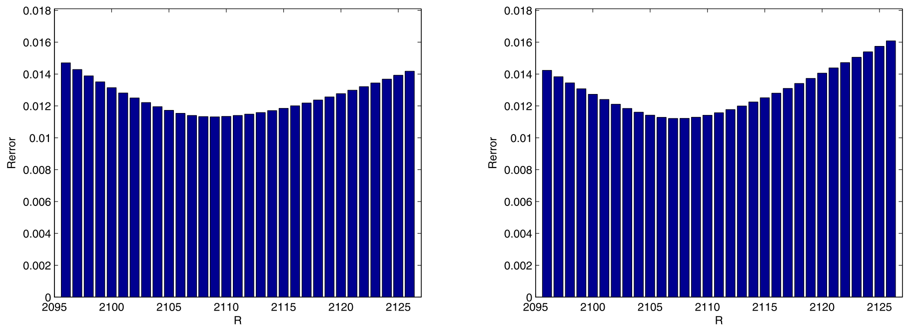
the parameter β, it is determined as $\beta = \alpha\eta = 0.02$. The choice of R_0 depends on the priori information of a true solution, i.e., the value of $\|x^\dagger\|_1$. If R_0 is close to the value of $\|x^\dagger\|_1$, one can obtain desired inversion results with less outer iteration numbers. On the contrary, if R_0 is far away from the value of $\|x^\dagger\|_1$, one needs more outer iteration numbers. In this part, we let $R_0 = 10$. We utilize discrepancy principle (2.2) to determine the radius R of the ℓ_1-ball constraint such that $R = \sup\{R > 0 | \delta \leqslant \|Ax^* - y^\delta\|_2\}$. It is shown that when a good estimate for the noise level δ is known, this method yields a good radius R. According to the priori information of $x^\dagger$, we choose an initial value of R_0 and compute x^*. If $\|Ax^* - y^\delta\|_2 > \delta$, we try $R_j = R_0 + j$, $j = 1, 2, \ldots$ until $\|Ax^* - y^\delta\|_2 \leqslant \delta$ is satisfied. With j increasing, we can find $R = \sup\{R > 0 | \delta \leqslant \|Ax^* - y^\delta\|_2\}$. On the contrary, for any initial R, if $\|Ax^* - y^\delta\|_2 < \delta$, we try $R_j = R_0 - j$, $j = 1, 2, \ldots$ until $\|Ax^* - y^\delta\|_2 \geqslant \delta$ is satisfied. Figure 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 $R = \sup\{R > 0 | \delta \leqslant \|Ax^* - y^\delta\|_2\}$. 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 $R = \|x^*\|_1$.

To test the stability of the PG Algorithms with respect to R, we choose several values of R in figure 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 is close to 16.



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

Figure 6. The value of the discrepancy $\|Ax^* - y^\delta\|_2$ with different R.



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

Figure 7. The relative error of reconstruction x^* by the two PG algorithms with different R.

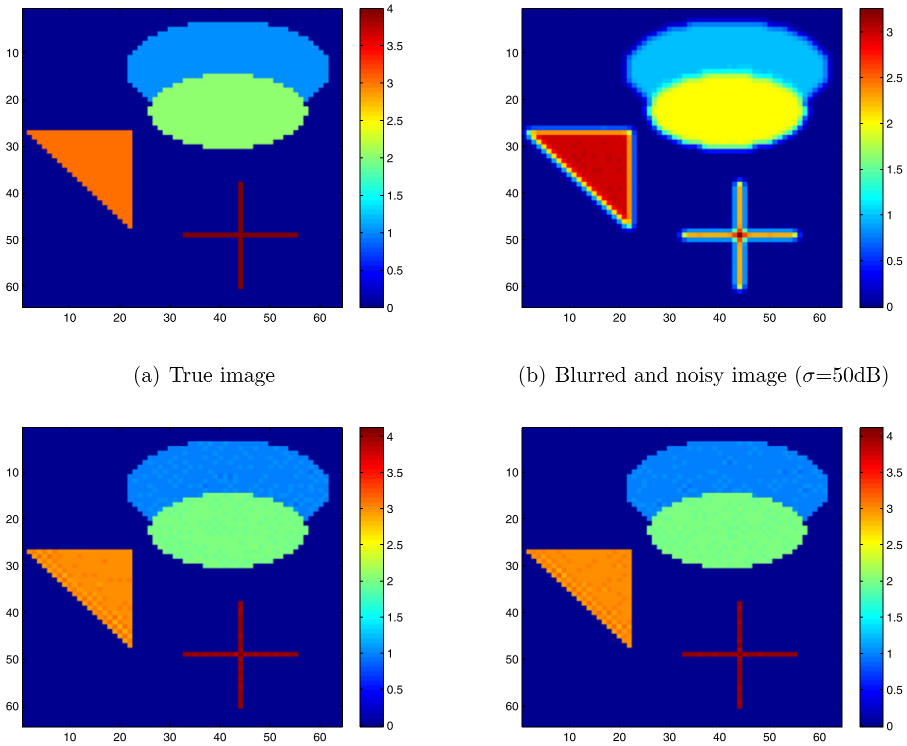
For $0 < \eta \leqslant 1$, $\mathcal{R}_\eta(x)$ is non-convex. To analyze the influence of η, we choose different values for the parameter η. From each row in table 1, we see that, Rerror of reconstruction gets better with η increasing which implies the non-convex regularization (case $\eta > 0$) has better performance compared to the classical ℓ_1 regularization (case $\eta = 0$). Figure 3 shows graphs of the reconstruction x^* by the PG-GCGM and PG-SF algorithm when $\eta = 1$.

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 Rerror. The results are shown in figure 4. To reach a level of 7×10^{-3} for the relative error, PG-GCGM algorithm takes 0.62 s, PG-SF algorithm 1.08 s, and ST-($\alpha\ell_1 - \beta\ell_2$) algorithm 18.40 s. The ST-($\alpha\ell_1 - \beta\ell_2$) algorithm procedure is significantly slower than the two PG algorithms.

Theoretically, we require assumption 4.6(A2), i.e. $\lambda \geqslant \beta \max\{\mathrm{eig}(x^k)\}$ for the convergence of the PG-SF algorithm. Next, we test whether λ satisfies this assumption. Figure 5(a) shows Rerror corresponding to the different reconstruction x^k, $1 \leqslant k \leqslant 1500$, whereas figure 5(b) shows the maximal eigenvalues $\max\{\mathrm{eig}(x^k)\}$. It is obvious that all $\max\{\mathrm{eig}(x^k)\}$ are less than 3.5. In this section, we let $\lambda = 1$ and $\beta = \alpha\eta$, where $\alpha = 0.02$ and $\eta = 1$. Thus, $\lambda \geqslant 3.5\beta$ and assumption 4.6(A2) is satisfied. Theoretically, we can let λ be any value greater than 3.5β.

Table 3. Rerror of reconstruction x^* with different values of η.

η	0.0	0.1	0.2	0.3	0.4	0.5	0.7	0.9	1.0
ST-($\alpha\ell_1 - \beta\ell_2$)	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



(a) True image

(b) Blurred and noisy image (σ=50dB)

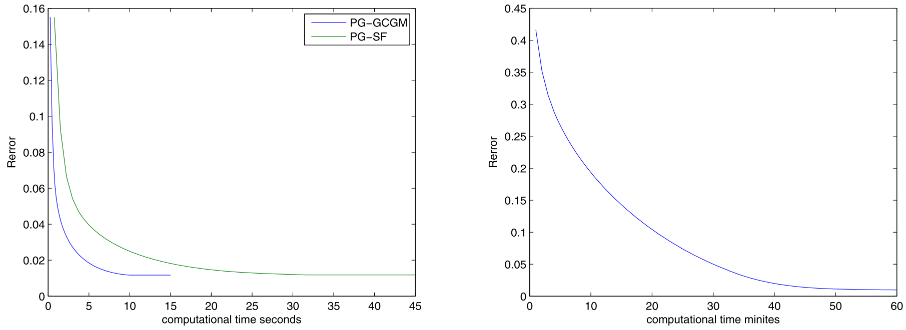
(c) PG-GCGM algorithm, Rerror=0.0130

(d) PG-SF algorithm, Rerror=0.0126

Figure 8. True image and its blurred and noisy observation together with reconstructions x^* for $\eta = 0.7$.

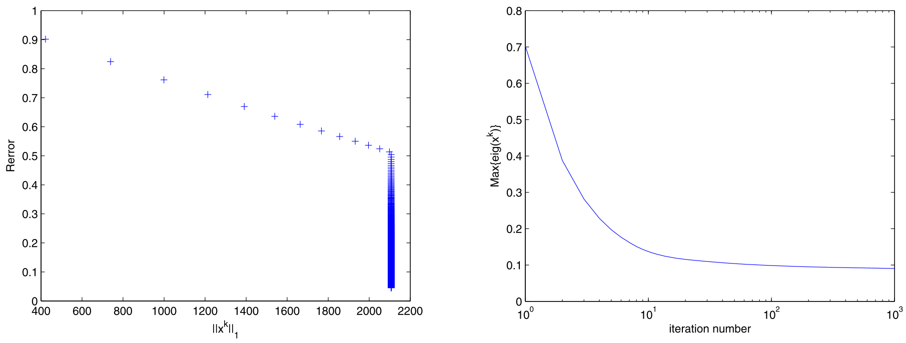
Nevertheless, a larger value of λ corresponds to a smaller iteration step, leading to a slower convergence rate.

Finally, we let $n = 1800$, $m = 0.4n$ and $s = 0.2m$, $\sigma = 50dB$. The coefficients λ and η remain the same as in the first test. The noise level δ is around 0.09; hence we let $\alpha = \beta = 0.1$. We test the convergence rate of the two PG algorithms and ST-($\alpha\ell_1 - \beta\ell_2$) algorithm in terms of the computational time with several different values of Rerror. We observe from table 2 that the ST-($\alpha\ell_1 - \beta\ell_2$) algorithm takes more than 100 min to reach a 2% relative error, whereas the two PG algorithms only need around 8 and 41 s to reach the same level of relative error. The PG algorithms converge much faster than the ST-($\alpha\ell_1 - \beta\ell_2$) algorithm.



(a) PG-GCGM algorithm and PG-SF algorithm (b) ST-($\alpha\ell_1 - \beta\ell_2$) algorithm

Figure 9. (a) Convergence rate of PG-GCGM algorithm and PG-SF algorithm; (b) convergence rate of ST-($\alpha\ell_1 - \beta\ell_2$) algorithm.



(a) (b)

Figure 10. (a) Rerror for x^k, $1 \leqslant k \leqslant 1000$; (b) $\max\{\mathrm{eig}(x^k)\}$ for $1 \leqslant k \leqslant 1000$.

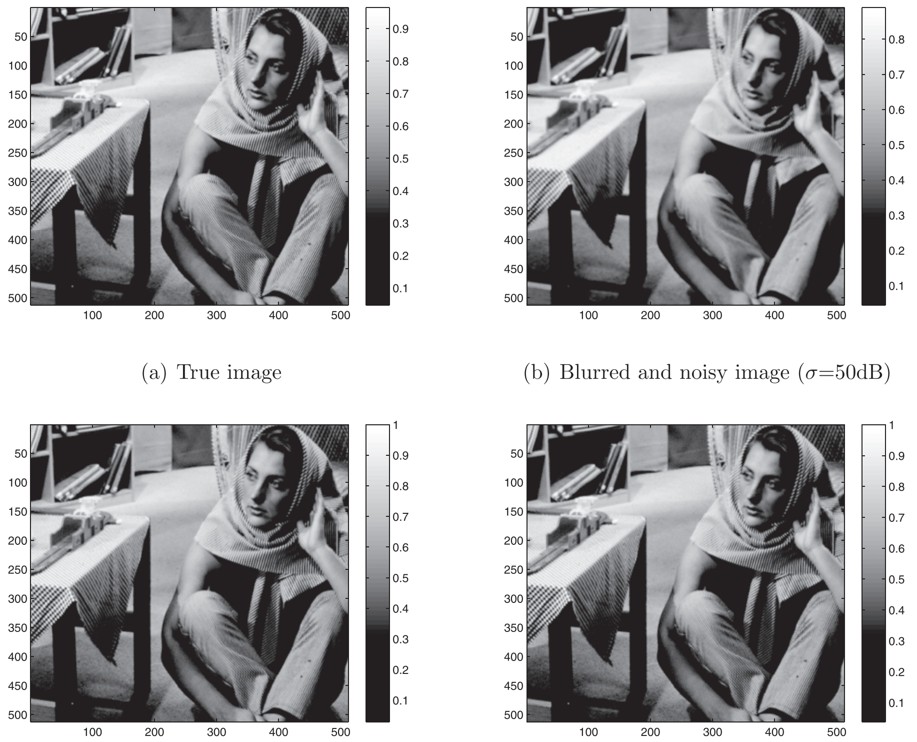
5.2. Example 2: image deblurring

In the second example, we consider an ill-conditioned image deblurring problem which is related to 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

$$\int_a^b K(s,t)\, f(t)\, \mathrm{d}t = g(s),$$

where $K(s,t)$ is the kernel function, $g(s)$ is the observed image and $f(t)$ is the true image. We utilize MATLAB regularization tools ([21]) by calling $[A, b, x^\dagger] = \mathrm{blur}(n, \mathrm{band}, \tau)$, where the Gaussian point-spread function is used as the kernel function

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



(a) True image

(b) Blurred and noisy image (σ=50dB)

(c) PG-GCGM algorithm, Rerror=0.0863

(d) PG-SF algorithm, Rerror=0.0872

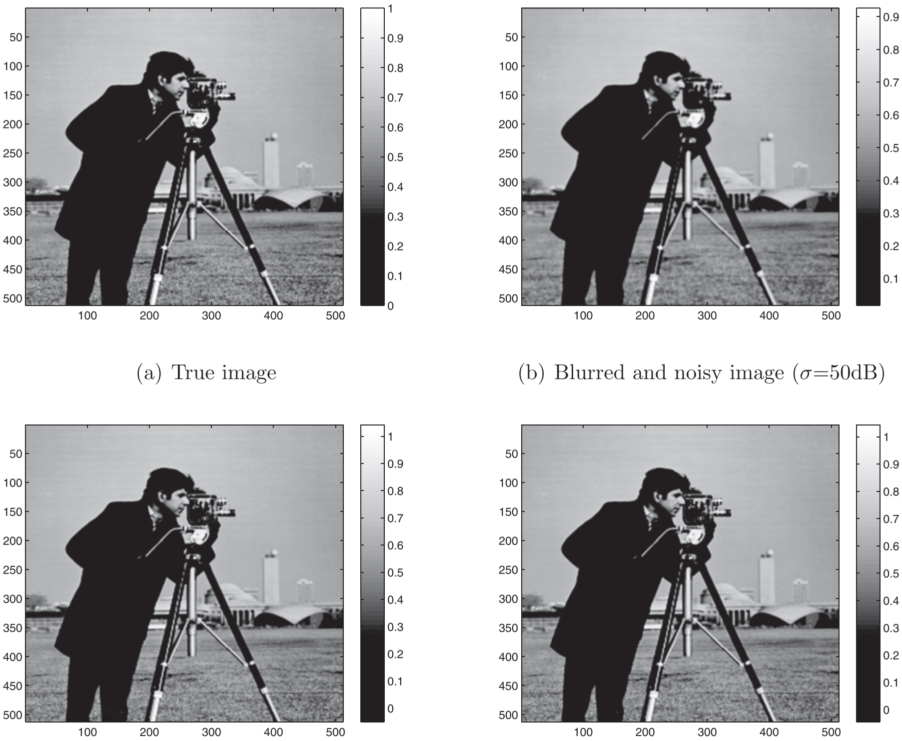
Figure 11. True image and its blurred and noisy observation together with reconstructions x^* for $\eta = 0.7$.

The symmetric $n^2 \times n^2$ Toeplitz matrix A 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

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

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

We choose $n = 64$, band $= 3$, $\tau = 0.7$. A noise δ is added to exact data $y^\dagger$ by calling $y^\delta = \mathrm{awgn}(Ax^\dagger, \sigma)$, where $\sigma = 50$dB and δ 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 $x^0 = 0.01\mathrm{ones}(n, 1)$. The value of $\|A\|_2$ is around 1 and the condition number is around 30. The initial value x^0 is generated by calling $x^0 = 0.01\mathrm{ones}(n \times n, 1)$. We let $R_0 = 2000$. Figure 6 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 $R = \sup\{R > 0 | \delta < \|Ax^* - y^\delta\|_2\}$. We choose an initial value of R_0 and compute x^*. If $\|Ax^* - y^\delta\|_2 > \delta$, we try $R_j = R_0 + j,\ j = 1, 2, \ldots$ until $\|Ax^* - y^\delta\|_2 \leqslant \delta$ is satisfied. From figure 6, we see that if $R \geqslant 2018$, then $\|Ax^* - y^\delta\|_2 < \delta$, and if $R \leqslant 2017$, then $\|Ax^* - y^\delta\|_2 > \delta$. So the numerical results suggest that R should be chosen as 2107. Figure 7 shows the performance of the PG algorithms with respect to R. It is



(a) True image
(b) Blurred and noisy image (σ=50dB)
(c) PG-GCGM algorithm, Rerror=0.0379
(d) PG-SF algorithm, Rerror=0.0355

Figure 12. True image and its blurred and noisy observation together with reconstructions x^* for $\eta = 0.7$.

shown that the two PG algorithms have good performance with appropriate radius *R*. Observe that for a fixed parameter η, Rerror of reconstruction x^* gets better if *R* is closer to 2107.

To analyze the influence of η, we choose different values for the parameter η. From each row in table 3, we see that the results of reconstruction get better with η increasing, implying that the non-convex regularization (for $\eta > 0$) has better performance than the classical ℓ_1 regularization (for $\eta = 0$). However, if η increases to near 1, the accuracy of recovery decreases and $\eta = 0.7$ is optimal. Figure 8 shows graphs of the reconstruction x^* by the PG-GCGM and PG-SF algorithm when $\eta = 0.7$.

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 figure 9. To get within a distance of the true minimizer corresponding to a 1.2×10^{-2} relative error, the PG-GCGM algorithm takes 10.12 s, PG-SF algorithm 36.26 s, and the ST-($\alpha\ell_1 - \beta\ell_2$) algorithm 58.54 min. The ST-($\alpha\ell_1 - \beta\ell_2$) algorithm procedure is significantly slower than the two PG algorithms.

Theoretically, we need assumption 4.6(A2), $\lambda \geqslant \beta \max\{\mathrm{eig}(x^k)\}$, for convergence of the PG-SF algorithm. In figure 10, we test whether λ satisfies this assumption. Figure 10 (a) shows Rerror corresponding to the different reconstruction x^k and figure 10 (b) shows the maximal eigenvalue $\max\{\mathrm{eig}(x^k)\}$. It is seen 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, $\lambda \geqslant 3.5\beta$, and assumption 4.6(A2) is satisfied.

Finally, we show the inversion results for the deblurred imaging problem with two synthetic images, 'Barbara' and 'cameraman'. The size of both images is 512×512. We let $\tau = 1.2$ and $\sigma = 50\text{d}B$. The coefficients λ and η remain the same as in the previous test. The noise level δ is around 0.8; hence we let $\alpha = 0.8$ and $\beta = \alpha\eta = 0.56$. Figures 11 and 12 show the true images and inversion results from the proposed algorithms. It is shown that we can obtain good results even for the synthetic images which have different features.

ORCID iDs

Liang Ding https://orcid.org/0000-0003-1543-8614

References

[1] Anzengruber S W and Ramlau R 2010 Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators *Inverse Problems* **26** 025001
[2] Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM J. Imaging Sci.* **2** 183–202
[3] Becker S, Bobin J and Candès E J 2011 NESTA: a fast and accurate first-order method for sparse recovery *SIAM J. Imaging Sci.* **4** 1–39
[4] Benning M and Burger M 2018 Modern regularization methods for inverse problems *Acta Numer.* **27** 1–111
[5] van den Berg E and Friedlander M P 2008 Probing the pareto frontier for basis pursuit solutions *SIAM J. Sci. Comput.* **31** 890–912
[6] Blumensath T and Davies M E 2008 Iterative thresholding for sparse approximations *J. Fourier Anal. Appl.* **14** 629–54
[7] Blumensath T and Davies M E 2009 Iterative hard thresholding for compressed sensing *Appl. Comput. Harmon. Anal.* **27** 265–74
[8] Bonesky T 2009 Morozov's discrepancy principle and Tikhonov-type functionals *Inverse Problems* **25** 015015
[9] Bredies K and Lorenz D A 2008 Iterated hard shrinkage for minimization problems with sparsity constraints *SIAM J. Sci. Comput.* **30** 657–83
[10] Chambolle A and Dossal C 2015 On the convergence of the iterates of the 'fast iterative shrinkage/thresholding algorithm' *J. Optim. Theory Appl.* **166** 968–82
[11] O'Donoghue B and Candès E 2015 Adaptive restart for accelerated gradient schemes *Found. Comput. Math.* **15** 715–32
[12] Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint *Comm. Pure Appl. Math.* **57** 1413–57
[13] Daubechies I, Defrise M and Mol C D 2016 Sparsity-enforcing regularisation and ISTA revisited *Inverse Problems* **32** 104001
[14] Daubechies I, Fornasier M and Loris I 2008 Accelerated projected gradient method for linear inverse problems with sparsity constraints *J. Fourier Anal. Appl.* **14** 764–92
[15] Ding L and Han W 2019 $\alpha\ell_1 - \beta\ell_2$ regularization for sparse recovery *Inverse Problems* **35** 125009
[16] Figueiredo M A T, Nowak R D and Wright S J 2007 Gradient projection for sparse reconstruction application to compressed sensing and other inverse problems *IEEE J. Sel. Top. Signal Process.* **1** 586–97
[17] Fornasier M (ed) 2010 *Theoretical Foundations and Numerical Methods for Sparse Recovery* (Berlin: De Gruyter & Co)
[18] Fornasier M, Peter S, Rauhut H and Worm S 2016 Conjugate gradient acceleration of iteratively re-weighted least squares methods *Comput. Optim. Appl.* **65** 205–59
[19] Fornasier M and Rauhut H 2008 Iterative thresholding algorithms *Appl. Comput. Harmon. Anal.* **25** 187–208
[20] Ge H, Wen J and Chen W 2018 The null space property of the truncated ℓ_{1-2}-minimization *IEEE Signal Process. Lett.* **25** 1261–5

[21] Hansen P C 2007 Regularization tools version 4.0 for matlab 7.3 *Numer. Algorithm.* **46** 189–94

[22] Huang X, Shi L and Yan M 2015 Nonconvex sorted ℓ_1 minimization for sparse approximation *J. Oper. Res. Soc. China* **3** 207–29

[23] Jin B and Maass P 2012 Sparsity regularization for parameter identification problems *Inverse Problems* **28** 123001

[24] Jin B, Maass P and Scherzer O 2017 Sparsity regularization in inverse problems *Inverse Problems* **33** 060301

[25] Lazzaro D, Piccolomini E L and Zama F 2019 A nonconvex penalization algorithm with automatic choice of the regularization parameter in sparse imaging *Inverse Problems* **35** 084002

[26] Li P, Chen W, Ge H and Ng M. K 2020 ℓ_1-$\alpha\ell_2$ minimization methods for signal and image reconstruction with impulsive noise removal *Inverse Problems* **36** 055009

[27] Loris I, Bertero M, De Mol C, Zanella R and Zanni L 2009 Accelerating gradient projection methods for ℓ_1-constrained signal recovery by steplength selection rules *Appl. Comput. Harmon. Anal.* **27** 247–54

[28] Loris I and Verhoeven C 2011 On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty *Inverse Problems* **27** 125007

[29] Lou Y and Yan M 2018 Fast L1-L2 minimization via a proximal operator *J. Sci. Comput.* **74** 767–85

[30] Montefusco L B, Lazzaro D and Papi S 2013 A fast algorithm for nonconvex approaches to sparse recovery problems *Signal Process.* **93** 2636–47

[31] Nesterov Y 2005 Smooth minimization of non-smooth functions *Math. Program.* **103** 127–52

[32] Osher S, Burger M, Goldfarb D, Xu J and Yin W 2005 An iterative regularization method for total variation-based image restoration *Multiscale Model. Simul.* **4** 460–89

[33] Ramlau R 2002 Morozov's discrepancy principle for Tikhonov regularization of nonlinear operators *Numer. Funct. Anal. Opt.* **23** 147–72

[34] Ramlau R and Zarzer C A 2012 On the minimization of a Tikhonov functional with a non-convex sparsity constraint *Electron. Trans. Numer. Anal.* **39** 476–507

[35] Rockafellar R T 1970 *Convex Analysis* (Princeton: Princeton University Press)

[36] Scherzer O 1993 The use of Morozov's discrepancy principle for Tikhonov regularization for solving non-linear ill-posed problems *SIAM J. Numer. Anal.* **30** 1796–838

[37] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (*Applied Mathematical Sciences* vol 167) (New York: Springer)

[38] Teschke G and Borries C 2010 Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints *Inverse Problems* **26** 025007

[39] Tibshirani R 1996 Regression shrinkage and selection via the Lasso *J. Roy. Stat. Soc.* B **58** 267–88

[40] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-Posed Problems* (Washington: V H Winston and Sons)

[41] Tikhonov A N, Leonov A S and Yagola A G 1998 *Nonlinear Ill-Posed Problems* (London: Chapman and Hall)

[42] Wang W, Lu S, Mao H and Cheng J 2013 Multi-parameter Tikhonov regularization with the ℓ_0 sparsity constrain *Inverse Problems* **29** 065018

[43] Wright S, Nowak R and Figueiredo M 2009 Sparse reconstruction by separable approximation *IEEE Trans. Signal Process.* **57** 2479–93

[44] Yan L, Shin Y and Xiu D 2017 Sparse approximation using ℓ_1-ℓ_2 minimization and its application to stochastic collocation *SIAM J. Sci. Comput.* **39** A229–254

[45] Yin P, Lou Y, He Q and Xin J 2015 Minimization of ℓ_{1-2} for compressed sensing *SIAM J. Sci. Comput.* **37** A536–63

[46] Zeidler E 1985 *Nonlinear Functional Analysis and its Application* vol 3 (New York: Springer)