An iterative support shrinking algorithm for
$\ell_p$-$\ell_q$ minimization

Zhifang Liu, Yanan Zhao and Chunlin Wu *
School of Mathematical Sciences, Nankai University, Tianjin 300071, China
January 31, 2018

Abstract. We present an iterative support shrinking algorithm for $\ell_p$-$\ell_q$ minimization ($0 < p < 1 \leq q < \infty$). This algorithm guarantees the nonexpensiveness of the signal support set and can be easily implemented after being proximally linearized. The subproblem can be very efficiently solved due to its convexity and reducing size along iteration. We prove that the iterates of the algorithm globally converge to a stationary point of the $\ell_p$-$\ell_q$ objective function. In addition, we show a lower bound theory for the iteration sequence, which is more practical than the lower bound results for local minimizers in the literature.

Keywords. nonconvex nonsmooth regularization, non-Lipschitz optimization, support shrinking, sparse signal reconstruction, lower bound theory

Mathematics subject classification (2010). 49M05,49K30,90C26,94A12,90C30

1 Introduction

Sparse reconstruction plays an important role in various applications such as signal and image processing, compressed sensing, model selection, variable selection, and many others [24, 9, 45, 36, 41]. This problem can described as follows. Given an $M \times N$ measurement matrix $A$ with $M < N$, we consider to recover the sparse signal $x \in \mathbb{R}^N$ from an observed signal $y = Ax + n \in \mathbb{R}^M$, where $n$ represents the measurement noise. There are many different types of noise. Two typical and important examples are Gaussian noise and heavy-tailed noise [38], which obey Gaussian distribution and heavier-than-Gaussian tails distribution [29], respectively. To obtain the sparest solution, one naturally proposes to solve the following $\ell_0$ minimization problem

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \text{ subject to } \|Ax - y\|_q \leq \varepsilon,$$

or its unconstrained counterpart

$$\min_{x \in \mathbb{R}^N} \|x\|_0 + \frac{1}{q\alpha} \|Ax - y\|_q^q,$$

(1.1)

where $\|\cdot\|_0$ denotes the $\ell_0$ “norm” that returns the number of nonzero entries of its argument, $\|\cdot\|_q$ for $q \in [1, \infty)$ is the $\ell_q$ norm, and $\alpha \in (0, \infty)$ is a parameter that balances the regularization and the fidelity. The second term in (1.1), named as the fidelity term, is constructed using the noise distribution and the Maximum Likelihood principle. As well known, for Gaussian noise, people use the $\ell_2$ fidelity term ($q = 2$). For heavy-tailed noise such as impulsive noise, the $\ell_1$ fidelity term ($q = 1$) is a good choice [23, 38]. Since the $\ell_0$ minimization is NP-hard [37], numerous methods have been proposed to approximate it. Two common ways are to replace the $\ell_0$ “norm” with the $\ell_1$ norm [12, 24] and the $\ell_p$ quasi-norm ($0 < p < 1$) [24, 16, 25, 22, 40], where the $\ell_p$...
quasi-norm is defined as \( \|x\|_p = (\sum_{j=1}^{N} |x_j|^p)^{1/p} \). In sparse reconstruction, the nonconvex \( \ell_p \) quasi-norm has some advantages \([15,25,40,32]\) over the convex \( \ell_1 \) norm.

In this paper, we focus on the following \( \ell_p-\ell_q \) minimization problem

\[
\min_{x \in \mathbb{R}^n} \mathcal{E}(x) := \|x\|_p^p + \frac{1}{q\alpha} \|Ax - y\|_q^q,
\]

where \( p \in (0,1), q \in [1, \infty) \) and \( \alpha \in (0, \infty) \). The objective function \( \mathcal{E} \) in (1.2) is nonsmooth, nonconvex and non-Lipschitz, which results in a great challenge for optimization. We now review some existing methods. As can been seen, most of them were designed for \( \ell_1-\ell_2 \) minimization.

One class of approaches is smoothing approximate methods \([19,17,18,4]\), which are based on the special structure of the nonsmooth function \( \mathcal{E} \). By a smoothing function \( \varphi(x, \theta) \) for the absolute value function \( |x| \), the \( \ell_p, p \in (0,1) \) regularization term can be smoothed. Two choices of \( \varphi(x, \theta) \) in \([17,18,4]\) are

\[
\varphi_1(x, \theta) = \left\{ \begin{array}{ll}
\frac{|x|}{\theta} & \text{if } |x| > \theta, \\
\frac{x^2}{\theta^2} + \frac{\theta}{2} & \text{if } |x| \leq \theta,
\end{array} \right. \text{ and } \varphi_2(x, \theta) = \sqrt{x^2 + 4\theta^2}.
\]

Based on this technique, hybrid orthogonal matching pursuit-smoothing gradient (OMP-SG) method \([19]\), smoothing quadratic regularization (SQR) algorithm \([4]\), and smoothing trust region Newton method \([18]\) have been proposed for \( \ell_p, p \in (0,1) \) regularized problems with smooth fidelity terms with convergence guarantee. They essentially reformulate the non-Lipschitz problem to be lipschitz ones by a smoothing parameter, which controls the approximate accuracy and need to be updated progressively to zero.

The second class of approaches is general iterative shrinkage-thresholding algorithms (GISA) for \( \ell_p-\ell_2 \) minimization \([43,46,8]\). GISA was inspired by the great success of soft thresholding and iterative shrinkage-thresholding algorithms (ISTA) \([21,3]\) for convex \( \ell_1-\ell_2 \) minimization problem. Specifically, the general step of GISA is

\[
x^{(k+1)} = T_{\alpha\beta}(x^{(k)} + \beta A^T(y - Ax^{(k)})),
\]

where \( \beta > 0 \) is an appropriate stepsize and \( T_{\alpha\beta} : \mathbb{R}^N \rightarrow \mathbb{R}^N \) is a shrinkage-thresholding operator. GISA is easy to implement, but it applies only to the case \( q = 2 \). Even for \( q = 2 \), the operator \( T_{\alpha\beta} \) have analytical expression only for \( p = 1/2 \) and \( p = 2/3 \) \([43,30]\). For a general \( 0 < p < 1 \), the operator \( T_{\alpha\beta} \) needs to be computed via numerical methods \([46,8]\).

The third class of approaches is iterative reweighted minimization methods; see, e.g. \([27,32,33,13,20]\). There are iterative reweighted least squares (IRLS) and iterative reweighted \( \ell_1 \) (IRL1) minimization methods. One can refer to \([35]\) for a systematic review. In \([32,33]\), the authors considered a smoothed \( \ell_p-\ell_2 \) minimization

\[
\sum_{j=1}^{N} (x_j^2 + \theta^2)^{p/2} + \frac{1}{2\alpha} \|Ax - y\|_2^2
\]

and proposed IRLS algorithms to solve this approximate problem. In \([20]\), Chen and Zhou considered the following approximation to \( \ell_p-\ell_2 \) minimization

\[
\sum_{j=1}^{N} (|x_j| + \theta)^p + \frac{1}{2\alpha} \|Ax - y\|_2^2
\]

for some small \( \theta > 0 \). An IRL1 algorithm was proposed to slove this approximate problem. Both IRLS and IRL1 are stable. Actually reweighted methods reformulate the original non-Lipschitz \( \ell_p-\ell_2 \) to lipschitz ones by a de-singularizing parameter.

In this paper, we consider (1.2) from a different perspective. We first obtain a proposition from the first order optimality condition. Motivated by this proposition, we propose an iterative algorithm with constraints on the
support set of the signals. The core idea is to guarantee that the signal support set will not expand in the iterative procedure. After constraints elimination and proximally linearized, this algorithm can be easily implemented. The subproblem therein is convex and with reducing size along iteration. It is solved inexactly by alternating direction method of multipliers (ADMM). Furthermore, we establish the global convergence of the iterates to a stationary point of (1.2). We also prove a new lower bound theory for the iteration sequence, which is more practical than those lower bounds for local minimizers in the literature. Numerical examples show the good performance of our proposed algorithm for both \(\ell_p-\ell_2\) and \(\ell_p-\ell_1\) restoration.

The rest of this paper is organized as follows. In section 2, we give some basic notation and preliminaries. In section 3, we describe the motivation, and propose our algorithms. In section 4, the convergence analysis is provided and the lower bound property of the iteration sequence is discussed. In section 5, we give implementation details. The numerical experiments are shown in section 6. Section 7 concludes the paper.

2 Some notations and preliminaries

Denote \(I = \{1, 2, \ldots, M\}\) and \(J = \{1, 2, \ldots, N\}\). For a vector \(x \in \mathbb{R}^N\), we refer to \(x_j\) as its \(j\)th entry and denote the support set of \(x\) by

\[\text{supp}(x) := \{j \in J : x_j \neq 0\} .\]

We assume that all vectors are column vectors. For a matrix \(A \in \mathbb{R}^{M \times N}\), we write its \(i\)th row as \(A^T_i\), which is the vector transpose of \(A_i \in \mathbb{R}^N\). Then we have

\[A = \begin{bmatrix} A^T_1 \\ \vdots \\ A^T_M \end{bmatrix} .\]

Let \(S\) be a subset of \(J\). We denote \(x_S\) be the subvector of \(x\) indexed by \(S\), which consists of the nonzero entries of \(x\) when \(S = \text{supp}(x)\). Similarly, we denote \(B = A_S\) to be the column submatrix of \(A\) consisting of the columns indexed by \(S\). Let \(B^T_i\) be the \(i\)th row of \(B\), we have \(B_i = (A_i)_S\).

Define \(\phi : [0, \infty) \rightarrow [0, \infty)\) by \(\phi(x) = x^p(0 < p < 1)\). We state some useful properties for \(\phi(\cdot)\).

**Proposition 2.1.** The function \(\phi(\cdot)\) has the following properties:

(i) \(\phi(0) = 0\) and \(\phi'(x) = px^{p-1} > 0\) on \((0, \infty)\).

(ii) \(\phi(x)\) is concave and the following inequality holds,

\[\phi(y) \leq \phi(x) + \phi'(x)(y - x), \quad \forall x \in (0, \infty), y \in [0, \infty) .\]  

(iii) For any \(c > 0\), \(\phi'(x)\) is \(L_c\)-Lipschitz continuous on \([c, \infty)\), i.e., there exists a constant \(L_c > 0\) determined by \(c\), such that \(\forall x, y \in [c, \infty)\),

\[|\phi'(x) - \phi'(y)| \leq L_c |x - y| .\]  

(iv) The subdifferential of \(\phi(|x|)\) at \(x\) is given by

\[\partial \phi(|x|) = \begin{cases} (-\infty, \infty), & x = 0, \\ \{\text{sgn}(x)\phi'(|x|)\}, & x \neq 0, \end{cases} \]

where \(\text{sgn}(x)\) is the signum function.
Using the definition of \( \phi(\cdot) \), we have \( \|x\|_p = \sum_{j \in J} \phi(|x_j|) \). Thus the objective function \( E \) in (1.2) reads
\[
E(x) = \sum_{j \in J} \phi(|x_j|) + \frac{1}{q\alpha} \sum_{i \in I} |A_i^T x - y_i|^q,
\]
which is bounded below and coercive.

Now, we drive the subdifferential of \( E \) at \( x \). For \( 1 < q < \infty \), by [39, Exercise 8.8 and Proposition 10.5], we get
\[
\partial E(x) = \partial \left( \sum_{j \in J} \phi(|x_j|) \right) + \frac{1}{q\alpha} \nabla \left( \sum_{i \in I} |A_i^T x - y_i|^q \right). \tag{2.4}
\]
where \( \partial (\sum_{j \in J} \phi(|x_j|)) = \partial \phi(|x_1|) \times \cdots \times \partial \phi(|x_N|) \). For \( q = 1 \), we have checked the regularity requirement by [39, Corollary 10.9], indicating
\[
\partial E(x) = \partial \left( \sum_{j \in J} \phi(|x_j|) \right) + \partial \left( \frac{1}{\alpha} \sum_{i \in I} |A_i^T x - y_i| \right). \tag{2.5}
\]

Throughout this paper, we say that \( x^* \) is a stationary point of (1.2) if \( x^* \) satisfies
\[
0 \in \partial E(x^*). \tag{2.6}
\]
If \( \bar{x} \) is a local minimizer of (1.2), then the first-order optimality condition (2.6) holds.

3 Motivation and the proposed algorithm

Proposition 3.1. Given \( x \in \mathbb{R}^N \). Suppose that \( x \) is sufficiently close to a local minimizer (or a stationary point) \( x^* \) of (1.2). Then it holds that
\[
x_j^* = 0, \quad \forall j \in \Omega_0 = J \setminus \text{supp}(x). \tag{3.1}
\]

Proof. We prove (3.1) by contradiction. For the case of \( 1 < q < \infty \). As \( x^* \) is a local minimizer (or a stationary point) of \( E \), the condition (2.6) implies that for any \( j \in J \), we have
\[
0 \in \partial \phi(|x_j^*|) + \frac{1}{\alpha} \left( \sum_{i \in I} \text{sgn}(A_i^T x^* - y_i) |A_i^T x^* - y_i|^{q-1} A_i \right)_j. \nonumber
\]
Assume that there is \( j' \in \Omega_0 \) such that \( x_{j'}^* \neq 0 \). Then we have
\[
0 = \text{sgn}(x_j^*) \phi'(|x_j^*|) + \frac{1}{\alpha} \left( \sum_{i \in I} \text{sgn}(A_i^T x^* - y_i) |A_i^T x^* - y_i|^{q-1} A_i \right)_{j'}. \tag{3.2}
\]
The second term on the right side of (3.2) is bounded. Since \( x \) is sufficiently close to \( x^* \), \( x_{j'}^* \) can be sufficiently close to \( x_{j'} = 0 \). Then the equation (3.2) is impossible to be true. This is a contradiction.

For the case of \( q = 1 \), the condition (2.6) can be rewritten as for any \( j \in J \),
\[
0 \in \partial \phi(|x_j^*|) + \frac{1}{\alpha} \left( \sum_{i \in I} \partial \cdot |(A_i^T x^* - y_i) A_i \right)_{j}. \tag{3.3}
\]
Using the boundedness of the second term on the right side of (3.3), we can prove the results similarly. \( \blacksquare \)
Motivated by Proposition 3.1, we propose to solve the problem (1.2) by an iteration process, which generates a sequence with nonincreasing support set. Suppose that \( x^{(k)} \) is an approximate solution in the \( k \)th iteration. In the next iteration, we minimize the objective function with the restriction of zero entries outside the support set of \( x^{(k)} \). This idea yields the following iterative support shrinking algorithm (ISSA) for solving (1.2).

### ISSA: Iterative Support Shrinking Algorithm

**Initialization:** Select \( x^{(0)} \in \mathbb{R}^N \).

**Iteration:** For \( k = 0, 1, \ldots \) until convergence:

1. Set \( S^{(k)} = \text{supp}(x^{(k)}) \).
2. Compute \( x^{(k+1)} \) by solving

\[
\begin{align*}
\min_{x \in \mathbb{R}^N} & \quad \sum_{j \in S^{(k)}} \phi(|x_j|) + \frac{1}{q\alpha} \|Ax - y\|_q^q, \\
\text{s. t.} & \quad x_j = 0, \; \forall j \in \Omega_0^{(k)} = J \setminus S^{(k)}. 
\end{align*}
\]

(\( P_x \))

In fact, the problem (\( P_x \)) amounts to minimize the objective function respect to only \#\( S^{(k)} \) entries of \( x \), with the remaining components being null. Note that \( S^{(k)} \) is the support of \( x^{(k)} \). Given a vector \( x \in \mathbb{R}^N \) with \( \text{supp}(x) \subseteq S^{(k)} \), we let \( z^{(k)} = x_{S^{(k)}}^{(k)}, B^{(k)} = A_{S^{(k)}} \) and \( z = x_{S^{(k)}} \). It follows that

\[
B^{(k)}z = Ax, \quad \|z - z^{(k)}\|_2 = \|x - x^{(k)}\|_2. \tag{3.4}
\]

These relationships help to reformulate the problem (\( P_x \)) to an unconstrained problem with \( z \) as the unknowns. At the same time, each term \( \phi(|x_j|), j \in S^{(k)} \) can be linearized at \( |x_j^{(k)}| \neq 0 \). Together with a proximal technique, we present an iterative support shrinking algorithm with proximal linearization (ISSAPL) to solve (1.2).

### ISSAPL: Iterative Support Shrinking Algorithm with Proximal Linearization

**Initialization:** Select \( x^{(0)} \in \mathbb{R}^N \) and \( \beta > 0 \).

**Iteration:** For \( k = 0, 1, \ldots \) until convergence:

1. Set \( S^{(k)} = \text{supp}(x^{(k)}) \).
2. Generate \( x^{(k+1)} \) as follows:
   1. Set \( z^{(k)} = x_{S^{(k)}}^{(k)} \) and \( B^{(k)} = A_{S^{(k)}} \).
   2. Compute \( \hat{z}^{(k+1)} \) by solving

\[
\min_{z} \hat{E}^{(k)}(z) = \sum_{j \in S^{(k)}} \phi'(|x_j^{(k)}|)z_j + \frac{1}{q\alpha} \|B^{(k)}z - y\|_q^q + \frac{\beta}{2} \|z - z^{(k)}\|_2^2. \tag{P_z}
\]

2.3. Set

\[
x_j^{(k+1)} = \begin{cases} 
0, & j \in \Omega_0^{(k)} = J \setminus S^{(k)}, \\
\hat{z}_j^{(k+1)}, & j \in S^{(k)}.
\end{cases} \tag{3.5}
\]

The problem (\( P_z \)) in ISSAPL has a unique optimal solution due to strong convexity of \( \hat{E}^{(k)} \). Although the problem (\( P_x \)) is a convex optimization problem, it needs to be solved by iteration. In practical, we solve the
Then for any $j \in \Omega^{(k)}$ inexactly. Now we present our inexact iterative support shrinking algorithm with proximal linearization (InISSAPL) to solve $\mathcal{P}_2$.

### InISSAPL: Inexact Iterative Support Shrinking Algorithm with Proximal Linearization

| Initialization: | Select $\mathbf{x}^{(0)} \in \mathbb{R}^N$, $\beta > 0$ and $0 \leq \varepsilon < 1$. |
|-----------------|----------------------------------------------------------------------------------------------------------------|
| Iteration:      | For $k = 0, 1, \ldots$ until convergence:                                                                  |
| 1. Set $S^{(k)} = \text{supp}(\mathbf{x}^{(k)})$.                                                          |
| 2. Generate $\mathbf{x}^{(k+1)}$ as follows:                                                                  |
| 2.1. Set $\mathbf{z}^{(k)} = \mathbf{x}^{(k)}_{S^{(k)}}$ and $\mathbf{B}^{(k)} = \mathbf{A}_{S^{(k)}}$.       |
| 2.2. Find $\hat{\mathbf{z}}^{(k+1)} \approx \arg\min_{\mathbf{z}} \hat{E}(\mathbf{z})$ and $\hat{\mathbf{u}}^{(k+1)} \in \partial \hat{E}(\hat{\mathbf{z}}^{(k+1)})$, such that |
| $\|\hat{\mathbf{u}}^{(k+1)}\|_2 \leq \frac{\beta}{2\varepsilon}\|\hat{\mathbf{z}}^{(k+1)} - \mathbf{z}^{(k)}\|_2$. (3.6) |
| 2.3. Set $x_j^{(k+1)} = 0$, $j \in \Omega_{\mathbf{u}}^{(k)} = J \setminus S^{(k)}$, $x_j^{(k+1)} = \hat{x}_j^{(k+1)}$, $j \in S^{(k)}$. (3.7) |

**Remark.** The condition (3.6) in InISSAPL is motivated by [2]. It corresponds to an inexact optimality condition and a guide to select the approximate solution for $\mathcal{P}_2$. Due to the strong convexity of the problem $\mathcal{P}_2$, it can be solved to any given accuracy. Therefore, the condition (3.6) in InISSAPL can hold, as long as the problem $\mathcal{P}_2$ is sufficiently solved.

We have some useful representations of $\hat{\mathbf{u}}^{(k+1)}$. Since $\hat{\mathbf{u}}^{(k+1)} \in \partial \hat{E}(\hat{\mathbf{z}}^{(k+1)})$, we have

$$
\hat{\mathbf{u}}^{(k+1)} \in \partial \left( \sum_{j \in S^{(k)}} \phi'(|x_j^{(k)}|) |\hat{x}_j^{(k+1)}| \right) + \beta(\hat{\mathbf{z}}^{(k+1)} - \mathbf{z}^{(k)}) + \frac{1}{q\alpha} \partial \left( \sum_{i \in I} \left| (B_i^{(k)})^T \hat{\mathbf{z}}^{(k+1)} - y_i \right|^q \right).
$$

Then for any $j \in S^{(k)}$ and $i \in I$, there are $\xi_j \in \partial |.|(\hat{x}_j^{(k+1)}) = \partial |.|(x_j^{(k+1)})$ and $\eta_i = \partial |.|((B_i^{(k)})^T \hat{\mathbf{z}}^{(k+1)} - y_i) = \partial |.|(A_i^T \mathbf{x}^{(k+1)} - y_i)$, such that when $1 < q < \infty$,

$$
\hat{u}_j^{(k+1)} = \xi_j \phi'(|x_j^{(k)}|) + \beta(\hat{x}_j^{(k+1)} - x_j^{(k)})
+ \frac{1}{\alpha} \left( \sum_{i \in I} \text{sgn}((B_i^{(k)})^T \hat{\mathbf{z}}^{(k+1)} - y_i) \left( (B_i^{(k)})^T \hat{\mathbf{z}}^{(k+1)} - y_i \right)^{q-1} B_i^{(k)} \right)_j
= \xi_j \phi'(|x_j^{(k)}|) + \beta(x_j^{(k+1)} - x_j^{(k)})
+ \frac{1}{\alpha} \left( \sum_{i \in I} \text{sgn}(A_i^T \mathbf{x}^{(k+1)} - y_i) \left( A_i^T \mathbf{x}^{(k+1)} - y_i \right)^{q-1} A_i \right)_j,
$$

(3.9)
and when $q = 1$,

$$
\hat{u}_j^{(k+1)} = \xi_j \phi'(|z_j^{(k)}|) + \beta(z_j^{(k+1)} - z_j^{(k)}) + \frac{1}{\alpha} \left( \sum_{i \in I} \eta_i B_i \right) j.
$$

$$
= \xi_j \phi'(|x_j^{(k)}|) + \beta(x_j^{(k+1)} - x_j^{(k)}) + \frac{1}{\alpha} \left( \sum_{i \in I} \eta_i A_i \right) j.
$$

(3.10)

### 4 Convergence analysis

In this section, we establish the global convergence result of the sequence by the proposed InISSAPL. These results also hold for ISSAPL.

From the iteration process of InISSAPL, we can see that it generates a nonincreasing sequence of support set. A basic lemma for $\{S^{(k)}\}$ is showed in the following.

**Lemma 4.1.** The sequence $\{S^{(k)}\}$ converges in a finite number of iterations, i.e., there exists an integer $K > 0$ such that if $k \geq K$, then $S^{(k)} = S^{(K)}$.

**Proof.** Since $J \supseteq S^{(0)} \supseteq \cdots \supseteq S^{(k)} \supseteq \cdots$, $\{S^{(k)}\}$ converges in a finite number of iterations. ■

Lemma 4.1 plays a key role in the convergence analysis for the three vector sequences, $\{x^{(k)}\}$, $\{z^{(k)}\}$ and $\{\hat{z}^{(k)}\}$, generated by InISSAPL. Note that $\hat{z}^{(k)} = x^{(k)}_{S^{(k)}}$. By (3.7), $\hat{z}^{(k)}$ and $x^{(k)}$ have exactly the same nonzero entries. From Lemma 4.1 we can claim that after a certain $K$ numbers of iteration, the support of $x^{(k)}$ is fixed, i.e., if $k \geq K$, then supp$(x^{(k)}) = S^{(K)}$, from which we can directly obtain that

$$
z^{(k)} = x^{(k)}_{S^{(K)}} = \hat{z}^{(k)}, \forall k > K.
$$

In the next, we establish the global convergence of the sequence $\{x^{(k)}\}$. For the convenience of description, we introduce an auxiliary function

$$
\mathcal{F}^{(k)}(x) = \sum_{j \in S^{(k)}} \phi(|x_j^{(k)}|) + \phi'(|x_j^{(k)}|) \left(|x_j^{(k)}| - |x_j^{(k)}|\right) + \frac{1}{q\alpha} \|Ax - y\|_q^q
$$

$$
+ \frac{\beta}{2} \|x - x^{(k)}\|_2^2.
$$

(4.1)

**Lemma 4.2.** For any $\beta > 0$ and $0 \leq \varepsilon < 1$, let $\{x^{(k)}\}$ be a sequence generated by InISSAPL. Then

(i) The sequence $\{\mathcal{E}(x^{(k)})\}$ is nonincreasing and satisfies

$$
\mathcal{E}(x^{(k+1)}) + \frac{\beta}{2} (1 - \varepsilon) \|x^{(k+1)} - x^{(k)}\|_2^2 \leq \mathcal{E}(x^{(k)}).
$$

(4.2)

(ii) The sequence $\{x^{(k)}\}$ is bounded and satisfies $\lim_{k \to \infty} \|x^{(k+1)} - x^{(k)}\|_2 = 0$. 

7
Proof. For the case of $1 < q < \infty$. Due to the fact that $\phi(0) = 0$, we have

$$\mathcal{F}^{(k)}(x^{(k)}) = \sum_{j \in S^{(k)}} \phi(|x_j^{(k)}|) + \frac{1}{q_\alpha} \|Ax^{(k)} - y\|_q^q$$

$$= \sum_{j \in J} \phi(|x_j^{(k)}|) + \frac{1}{q_\alpha} \|Ax^{(k)} - y\|_q^q = \mathcal{E}(x^{(k)}). \quad (4.3)$$

When $x \in \mathbb{R}^N$ and $\text{supp}(x) \subseteq S^{(k)}$, we obtain

$$\mathcal{F}^{(k)}(x) = \sum_{j \in \text{supp}(x)} \phi(|x_j|) + \frac{1}{q_\alpha} \|Ax - y\|_q^q + \frac{\beta}{2} \|x - x^{(k)}\|_2^2$$

\[ \text{[by (2.1)]} \geq \sum_{j \in S^{(k)}} \phi(|x_j|) + \frac{1}{q_\alpha} \|Ax - y\|_q^q + \frac{\beta}{2} \|x - x^{(k)}\|_2^2 \]

$$= \sum_{j \in J} \phi(|x_j|) + \frac{1}{q_\alpha} \|Ax - y\|_q^q + \frac{\beta}{2} \|x - x^{(k)}\|_2^2$$

$$= \mathcal{E}(x) + \frac{\beta}{2} \|x - x^{(k)}\|_2^2. \quad (4.4)$$

The subdifferential of $\mathcal{F}^{(k)}$ at $x$ is defined as

$$\partial \mathcal{F}^{(k)}(x) = \partial \left( \sum_{j \in S^{(k)}} \phi'(|x_j^{(k)}|) x_j \right) + \beta(x - x^{(k)})$$

$$+ \frac{1}{q_\alpha} \nabla \left( \sum_{i \in I} |A_i^T x - y_i|^q \right)$$

$$= \partial \left( \sum_{j \in S^{(k)}} \phi'(|x_j^{(k)}|) x_j \right) + \beta(x - x^{(k)})$$

$$+ \frac{1}{\alpha} \sum_{i \in I} \text{sgn}(A_i^T x - y_i) A_i^T x - y_i^{q-1} A_i.$$
with \( \hat{u}_j^{(k+1)} \) in (3.9). Then \( u^{(k+1)} \in \partial \mathcal{F}^{(k)}(x^{(k+1)}) \). Since for any \( j \in J \setminus S^{(k)} \), \( x_j^{(k+1)} = x_j^{(k)} = 0 \), we have

\[
\langle u^{(k+1)}, x^{(k)} - x^{(k+1)} \rangle = \sum_{j \in S^{(k)}} u_j^{(k+1)}(x_j^{(k)} - x_j^{(k+1)})
\]

\[
= \sum_{j \in S^{(k)}} u_j^{(k+1)}(z_j^{(k)} - z_j^{(k+1)})
\]

\[
\geq -\|u^{(k+1)}\|_2\|z^{(k)} - z^{(k+1)}\|_2
\]

(4.5)

Putting (4.3), (4.4) and (4.5) together, we obtain

\[
\mathcal{E}(x^{(k)}) = \mathcal{F}^{(k)}(x^{(k)}) \geq \mathcal{F}^{(k)}(x^{(k+1)}) + \langle u^{(k+1)}, x^{(k)} - x^{(k+1)} \rangle
\]

\[
\geq \mathcal{F}^{(k)}(x^{(k+1)}) - \frac{\beta}{2}\varepsilon\|x^{(k+1)} - x^{(k)}\|_2
\]

\[
\geq \mathcal{E}(x^{(k+1)}) + \frac{\beta}{2}(1 - \varepsilon)\|x^{(k+1)} - x^{(k)}\|_2^2.
\]

With the fact that \( \mathcal{E}(x) \) is bounded from below and \( \frac{\beta}{2}(1 - \varepsilon) > 0 \), it follows that \( \{\mathcal{E}(x^{(k)})\} \) is nonincreasing and converges to a finite value as \( k \to \infty \). Thus \( \lim_{k \to \infty} \|x^{(k+1)} - x^{(k)}\|_2 = 0 \).

Because \( \mathcal{E}(x) \) is coercive, we know that \( \{x^{(k)}\} \) is bounded.

For the case of \( q = 1 \), the subdifferential of \( \mathcal{F}^{(k)} \) at \( x \) is given by

\[
\partial \mathcal{F}^{(k)}(x) = \partial \left( \sum_{j \in S^{(k)}} \phi'(|x_j^{(k)}|)x_j \right) + \beta(x - x^{(k)}) + \frac{1}{\alpha} \sum_{i \in I} |A_i^T x - y_i|
\]

\[
= \partial \left( \sum_{j \in S^{(k)}} \phi'(|x_j^{(k)}|)x_j \right) + \beta(x - x^{(k)}) + \frac{1}{\alpha} \sum_{i \in I} \partial |(A_i^T x - y_i)|A_i.
\]

Let \( u^{(k+1)} = (u_1^{(k+1)}, \ldots, u_N^{(k+1)})^T \) with

\[
u_j^{(k+1)} = \begin{cases} \hat{u}_j^{(k+1)}, & j \in S^{(k)} \\ \frac{1}{\alpha} \left( \sum_{i \in I} \eta_i A_i \right), & \text{otherwise}, \end{cases}
\]

where \( \hat{u}_j^{(k+1)} \) is as in (3.10) and \( \eta_i \in \partial |(A_i^T x - y_i)|. \) Then \( u^{(k+1)} \in \partial \mathcal{F}^{(k)}(x^{(k+1)}) \). In a similar way, we can prove that (i)(ii) holds.

Recall the results of Lemma 4.1, we now focus on the iteration number \( k \geq K \) to get the convergence of the sequence \( \{x^{(k)}\} \). Then the entries of \( u^{(k+1)} \) in (3.8) can be written as for any \( j \in S^{(k)} \), when \( 1 < q < \infty \),

\[
\hat{u}_j^{(k+1)} = \text{sgn}(x_j^{(k)})\phi'(|x_j^{(k)}|) + \beta(x_j^{(k+1)} - x_j^{(k)}) + \frac{1}{\alpha} \left( \sum_{i \in I} \text{sgn}(A_i^T x^{(k+1)} - y_i)\left|A_i^T x^{(k+1)} - y_i\right|^{q-1} A_i \right).
\]
and when \( q = 1 \),
\[
\tilde{u}^{(k+1)}_j = \text{sgn}(x_j^{(k+1)}) \phi'(|x_j^{(k)}|) + \beta (x_j^{(k+1)} - x_j^{(k)}) + \frac{1}{\alpha} \left( \sum_{i \in I} \gamma_i A_i \right)_{j'}.
\] (4.7)

The condition (3.6) reads
\[
\| \tilde{u}^{(k+1)} \|_2 \leq \frac{\beta}{2} \varepsilon \| z^{(k+1)} - z^{(k)} \|_2 = \frac{\beta}{2} \varepsilon \| x^{(k+1)} - x^{(k)} \|_2,
\] (4.8)
for \( 1 \leq q < \infty \).

The following is a bound theory on the iteration sequence, which is important to establish the convergence of \( \{ x^{(k)} \} \).

**Theorem 4.3.** There are \( 0 < c < C < \infty \) such that
\[
either \quad x_j^{(k)} = 0 \quad or \quad c \leq |x_j^{(k)}| \leq C, \forall j \in J, \forall k \geq K.
\] (4.9)

**Proof.** From Lemma 4.1 for any \( j \in S^{(K)} \) and \( k \geq K \), \( x_j^{(k)} \neq 0 \). We now prove by contradiction that \( |x_j^{(k)}| \) has nonzero lower and upper bound for any \( j \in S^{(K)} \), \( \forall k \geq K \).

For the case of \( 1 < q < \infty \), assume there exists \( j' \in S^{(K)} \) such that \( x_j^{(k)} \neq 0 \) and \( \lim_{k \to \infty} x_j^{(k)} = 0 \). Note that, if necessary, we can pass to a subsequence of \( x_j^{(k)} \). By letting
\[
\zeta_{j'} = \beta (x_j^{(k+1)} - x_j^{(k)}) + \frac{1}{\alpha} \left( \sum_{i \in I} \text{sgn}(A_i x_j^{(k+1)} - y_i) \right) \left( A_i x_j^{(k+1)} - y_i \right)^{q-1} A_i_{j'}
\] according to (4.6). It follows from the boundness of \( \{ x^{(k)} \} \) (Lemma 4.2) that \( |\zeta_{j'}| \) is bounded. The condition (4.8) implies that \( |\tilde{u}^{(k+1)}_{j'}| \) is also bounded. Thus the equation (4.10) is impossible to hold when \( k \to \infty \).

For the case of \( q = 1 \), by letting
\[
\zeta_{j'} = \beta (x_j^{(k+1)} - x_j^{(k)}) + \frac{1}{\alpha} \left( \sum_{i \in I} \gamma_i A_i \right)_{j'}
\] we have
\[
|\phi'(|x_j^{(k)}|)| \leq |\tilde{u}^{(k+1)}_{j'}| + |\zeta_{j'}|
\] (4.10)
according to (4.7). Using the boundness of the right-hand side, we can prove the results similarly. \( \Box \)

Compared with the lower bound theory for local minimizers in the literature [19], the bound theory in Theorem 4.3 is for the iterative sequence and more practical. Theorem 4.3 indicates that when \( k \geq K \), there exists \( L_c > 0 \), such that for any \( j \in S^{(K)} \),
\[
|\phi'(|x_j^{(k+1)}|) - \phi'(|x_j^{(k)}|)| \leq L_c |x_j^{(k+1)}| - |x_j^{(k)}| \leq L_c |x_j^{(k+1)} - x_j^{(k)}|.
\] (4.11)

We now derive a subgradient lower bound for the iterates gap.
Lemma 4.4. For each $k \geq K$, there exists $v^{(k+1)} \in \partial E(x^{(k+1)})$ such that

$$\|v^{(k+1)}\|_2 \leq \left( L_c + \frac{\beta}{2} (\varepsilon + 2) \right) \|x^{(k+1)} - x^{(k)}\|_2.$$ \hspace{1cm} (4.12)

Proof. For the case of $1 < q < \infty$, denote

$$v_j^{(k+1)} = \operatorname{sgn}(x_j^{(k+1)}) \phi'(|x_j^{(k+1)}|)$$

$$+ \frac{1}{\alpha} \left( \sum_{i \in I} \operatorname{sgn}(A_i^T x^{(k+1)} - y_i) \left| A_i^T x^{(k+1)} - y_i \right|^{q-1} A_i \right)_j, \forall j \in S(K),$$

$$v_j^{(k+1)} = 0, \forall j \in J \setminus S(K);$$

$$v_j^{(k+1)} = \operatorname{sgn}(x_j^{(k+1)}) \phi'(|x_j^{(k)}|)$$

$$+ \frac{1}{\alpha} \left( \sum_{i \in I} \operatorname{sgn}(A_i^T x^{(k+1)} - y_i) \left| A_i^T x^{(k+1)} - y_i \right|^{q-1} A_i \right)_j, \forall j \in S(K),$$

$$v_j^{(k+1)} = 0, \forall j \in J \setminus S(K).$$

Since $\partial \phi(|0|) = (-\infty, \infty)$, we have $v^{(k+1)} = (v_1^{(k+1)}, \ldots, v_N^{(k+1)})^T \in \partial E(x^{(k+1)})$ and

$$\|\tilde{v}^{(k+1)}\|_2 = \sqrt{\sum_{j \in S(K)} |\tilde{v}_j^{(k+1)}|^2} = \sqrt{\sum_{j \in S(K)} |\tilde{u}_j^{(k+1)} - \beta (x_j^{(k+1)} - x_j^{(k)})|^2}$$

$$\leq \|\hat{u}^{(k+1)}\|_2 + \beta \|x^{(k+1)} - x^{(k)}\|_2 \hspace{1cm} (4.13)$$

[by (4.8)]

$$\leq \frac{\beta}{2} (\varepsilon + 2) \|x^{(k+1)} - x^{(k)}\|_2.$$

Form (4.11), it follows that

$$\|v^{(k+1)} - \tilde{v}^{(k+1)}\|_2$$

$$= \sqrt{\sum_{j \in S(K)} \left| \operatorname{sgn}(x_j^{(k+1)}) \phi'(|x_j^{(k+1)}|) - \operatorname{sgn}(x_j^{(k+1)}) \phi'(|x_j^{(k)}|) \right|^2}$$

[by (4.11)]

$$\leq L_c \sqrt{\sum_{j \in S(K)} \left| x_j^{(k+1)} - x_j^{(k)} \right|^2} \hspace{1cm} (4.14)$$

$$\leq L_c \|x^{(k+1)} - x^{(k)}\|_2.$$

Combining (4.13) and (4.14) yields:

$$\|v^{(k+1)}\|_2 \leq \|v^{(k+1)} - \tilde{v}^{(k+1)}\|_2 + \|\tilde{v}^{(k+1)}\|_2$$

$$\leq \left( L_c + \frac{\beta}{2} (\varepsilon + 2) \right) \|x^{(k+1)} - x^{(k)}\|_2.$$
For the case of \( q = 1 \), denote
\[
\nu_j^{(k+1)} = \text{sgn}(x_j^{(k+1)}) \phi'(|x_j^{(k+1)}|) + \frac{1}{\alpha} \left( \sum_{i \in I} \eta_i A_i \right)_j, \quad \forall j \in S^{(k)},
\]
\[
\nu_j^{(k+1)} = 0, \quad \forall j \notin J \setminus S^{(k)};
\]
\[
\psi_j^{(k+1)} = \text{sgn}(x_j^{(k+1)}) \phi'(|x_j^{(k+1)}|) + \frac{1}{\alpha} \left( \sum_{i \in I} \eta_i A_i \right)_j, \quad \forall j \in S^{(k)},
\]
\[
\psi_j^{(k+1)} = 0, \quad \forall j \notin J \setminus S^{(k)}.
\]

In a similar way, we can prove that (4.12) holds.

Finally, we establish our main convergence result. An important tool for establishing the convergence is based on the so-called Kurdyka-Łojasiewicz (KL) property, which has attracted a lot of attention in recent years. Related preliminaries have been provided in Appendix 8.

**Theorem 4.5.** The sequences \( \{ x^{(k)} \} \) generated by InISSAPL converges globally to the limit point \( x^* \), which is a stationary point of \( E \).

**Proof.** Since \( \{ x^{(k)} \} \) is bounded (Lemma 4.2), there exists a subsequence \( \{ x^{(k_i)} \} \) and \( x^* \) such that
\[
x^{(k_i)} \to x^* \text{ and } E(x^{(k_i)}) \to E(x^*), \quad \text{as } i \to \infty.
\]
(4.15)

The function \( E \) satisfies the KL property [2]. Combing (4.2), (4.4) and (4.15), and by Theorem 2.9 in [2], the sequence \( \{ x^{(k)} \} \) converges globally to the limit point \( x^* \), which is a stationary point of \( E \).

## 5 Algorithm implementation

The subproblem in InISSAPL is a weighted \( \ell_1 \) minimization. Some standard methods like ADMM [7, 42, 28, 44], split Bregman method [26, 10, 11] and primal-dual algorithm [14] can be used to efficiently solve it. We here adopt ADMM. For clarity of description in this section, we refer to \( S^{(k)}, \phi'(|z_j^{(k)}|), B^{(k)} \) and \( z^{(k)} \) by \( S, w_j, B \) and \( z \), respectively. Consequently, (5.2) becomes
\[
\min_{x} \sum_{j \in S} w_j |z_j| + \frac{1}{\alpha} \| Bz - y \|_q^q + \frac{\beta}{2} \| z - \bar{z} \|_2^2.
\]
(5.1)

We rewrite (5.1) to the following constrained optimization problem:
\[
\min_{s, t} \sum_{j \in S} w_j |s_j| + \frac{1}{\alpha} \| t \|_q^q + \frac{\beta}{2} \| z - \bar{z} \|_2^2,
\]
(5.2)
\[
s. t. \quad z = s, Bz - y = t,
\]
and define the augmented Lagrangian functional for the problem (5.2) as follows:
\[
\mathcal{L}(z, s, t; \lambda, \mu) = \sum_{j \in S} w_j |s_j| + \frac{1}{\alpha} \| t \|_q^q + \frac{\beta}{2} \| z - \bar{z} \|_2^2 + \langle \lambda, z - s \rangle
\]
\[
+ \langle \mu, (Bz - y) - t \rangle + \frac{\gamma}{2} \| z - s \|_2^2 + \frac{\delta}{2} \| Bz - y - t \|_2^2,
\]
where \( \gamma, \delta > 0 \) are the penalty parameters and \( \lambda, \mu \) are the Lagrangian multipliers. The ADMM for solving (5.1) is described as follows.
Initialization: Start with \( z^{(0)} = \bar{z}, \lambda^{(0)} = 0, \mu^{(0)} = 0 \).

Iteration: For \( l = 0, 1, \ldots, \text{MAXit} \),

1. Compute
\[
\begin{align*}
(s^{(l+1)}, t^{(l+1)}) &= \arg \min_{s, t} \mathcal{L}(z^{(l)}, s, t; \lambda^{(l)}, \mu^{(l)}). 
\end{align*}
\] (5.3)

2. Compute
\[
\begin{align*}
z^{(l+1)} &= \arg \min_{z} \mathcal{L}(z, s^{(l)}, t^{(l)}; \lambda^{(l)}, \mu^{(l)}).
\end{align*}
\] (5.4)

3. Update
\[
\begin{align*}
\lambda^{(l+1)} &= \lambda^{(l)} + \gamma (z^{(l+1)} - s^{(l+1)}), \\
\mu^{(l+1)} &= \mu^{(l)} + \delta (Bz^{(l+1)} - y) - t^{(l+1)}.
\end{align*}
\] (5.5) (5.6)

ADMM can solve (5.1) to any accuracy. Considering the computational efficiency, we utilize in practice, the following stopping criterion [7]:

\[
\begin{align*}
\|\tau^{(l+1)}\|_2 &\leq \sqrt{M} \epsilon_{\text{abs}} + \epsilon_{\text{rel}} \max \left\{ \left\| \begin{bmatrix} I \\ \bar{B} \end{bmatrix} \right\|_2, \left\| \begin{bmatrix} -s^{(l+1)} \\ -t^{(l+1)} \end{bmatrix} \right\|_2 \right\}, \\
\|v^{(l+1)}\|_2 &\leq \sqrt{N} \epsilon_{\text{abs}} + \epsilon_{\text{rel}} \left\| \begin{bmatrix} I \\ \bar{B} \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} \lambda^{(l+1)} \\ \mu^{(l+1)} \end{bmatrix} \right\|_2,
\end{align*}
\]

where \( \tau^{(l+1)} = \left[ \begin{array}{c} z^{(l+1)} - s^{(l+1)} \\ Bz^{(l+1)} - y - t^{(l+1)} \end{array} \right] \) and \( v^{(l+1)} = \left[ \begin{array}{c} I \\ \bar{B} \end{array} \right] \left[ \begin{array}{c} \gamma (s^{(l)} - s^{(l+1)}) \\ \delta (t^{(l)} - t^{(l+1)}) \end{array} \right] \) are primal and dual residuals, respectively, at the \( l \)th iteration. \( \epsilon_{\text{abs}} > 0 \) is an absolute tolerance and \( \epsilon_{\text{rel}} \) is a relative tolerance.

The subproblems (5.3) and (5.4) can be efficiently solved.

1. For (5.3), the minimization with respect to \( s \) and \( t \) is
\[
\begin{align*}
\min_{s, t} \sum_{j \in S} w_j |s_j| + \frac{1}{q \alpha} \|t\|_q^q - \langle \lambda^{(l)}, s \rangle - \langle \mu^{(l)}, t \rangle + \frac{\gamma}{2} \|z^{(l)} - s\|_2^2 + \frac{\delta}{2} \|Bz^{(l)} - y - t\|_2^2,
\end{align*}
\]

which can be separated into two independent subproblems.

1.1. The minimization (5.3) with respect to \( s \)
\[
\begin{align*}
\min_{s} \sum_{j \in S} w_j |s_j| + \frac{\gamma}{2} \|s - z^{(l)} - \frac{\lambda^{(l)}}{\gamma}\|_2^2,
\end{align*}
\]

has the following closed form solution:
\[
\begin{align*}
s_j^{(l+1)} &= \text{sgn} \left( z_j^{(l)} + \frac{\lambda_j^{(l)}}{\gamma} \right) \max \left\{ \left| z_j^{(l)} + \frac{\lambda_j^{(l)}}{\gamma} \right| - \frac{w_j}{\gamma}, 0 \right\}, \forall j \in S.
\end{align*}
\]

1.2. The minimization (5.3) with respect to \( t \) is
\[
\begin{align*}
\min_{t} \frac{1}{q \alpha} \|t\|_q^q + \frac{\delta}{2} \|t - (Bz^{(l)} - y) - \frac{\mu^{(l)}}{\delta}\|_2^2.
\end{align*}
\]
For \( q = 1 \), \( t_i^{(l+1)} = \operatorname{sgn}(\langle \mathbf{B} z^{(l)} - y \rangle_i + \frac{\nu_i^{(l)}}{\alpha}) \max \left\{ \left| (\mathbf{B} z^{(l)} - y)\right|_i + \frac{\nu_i^{(l)}}{\alpha} - \frac{1}{\alpha}, 0 \right\}, \forall i \in I. \) For \( q = 2 \), \( t_i^{(l+1)} = \frac{\delta \alpha (\mathbf{B} z^{(l)} - y)_i + \alpha u_i^{(l)}}{\alpha}, \forall i \in I. \) For other \( 1 < q < \infty \), we can find \( t_i^{(l+1)} \) via any numerical procedure such as Newton’s method.

2. For (5.4), the minimization with respect to \( z \) is a quadratic optimization problem,

\[
\min_z \frac{\beta}{2} \| z - \bar{z} \|_2^2 + \langle \lambda^{(l)}, z \rangle + \langle \mu^{(l)}, \mathbf{B} z \rangle + \frac{\gamma}{2} \| z - s^{(l+1)} \|_2^2 + \frac{\delta}{2} \| (\mathbf{B} z - y) - t^{(l+1)} \|_2^2.
\]

Its optimality condition gives a linear system

\[
(\beta + \gamma + \delta \mathbf{B}^T \mathbf{B}) z = \beta \bar{z} + \gamma s^{(l+1)} + \delta \mathbf{B}^T (y + t^{(l+1)}) - \lambda^{(l)} - \mathbf{B}^T \mu^{(l)},
\]

which can be solved efficiently [7].

**Remark.** For \( q = 2 \), we actually only need to introduce one new variable \( s \).

## 6 Numerical experiments

In this section, we present numerical experiments to demonstrate the efficiency of the InISSAPL algorithm. All the tests were performed using Windows 10 and MATLAB R2016a 64-bit on a HP Z228 microtower workstation with an Intel(R) Core(TM) i7-4790 CPU @3.60GHz and 8GB memory.

In our experiments, we generated the true signal \( \mathbf{x}^* \) of the sparsity \( \kappa \) supported on a random index set with independently and identically distributed Gaussian entries. For the InISSAPL algorithm, we chose \( \text{MAXit} = 1000, \epsilon_{\text{abs}} = 10^{-7} \) and \( \epsilon_{\text{abs}} = 10^{-5} \) in the inner ADMM and adopted the following stopping criteria for the outer iteration

\[
\frac{\| x^{(k+1)} - x^{(k)} \|_2}{\| x^{(k)} \|_2} \leq 10^{-3}.
\]

### 6.1 Choice of \( p \)

In our first example, we tested the InISSAPL algorithm for \( \ell_p-\ell_2 \) minimization to recover sparse vectors with \( p \) varying among \{0.1, 0.3, 0.5, 0.7, 0.9\}. We used a 3000 \times 5000 random Gaussian matrix \( \mathbf{A} \) and a true signal \( \mathbf{x}^o \) of the sparsity \( \kappa = 500 \). The Gaussian noises with \( \sigma = 0.01 \) and \( \sigma = 0.1 \) were added to the clean signal \( \mathbf{A} \mathbf{x}^o \) to simulate the measurements \( y \). The InISSAPL algorithm was applied to get recovered signals \( \mathbf{x}^* \). To show the performance of our algorithm, we chose the results of LASSO as the benchmarks, which is solved by ADMM-lasso [7]. For ADMM-lasso, we also set \( \epsilon_{\text{abs}} = 10^{-7} \) and \( \epsilon_{\text{abs}} = 10^{-5} \). We show the relative \( L_2 \) error \( \frac{\| \mathbf{x}^* - \mathbf{x}^o \|_2}{\| \mathbf{x}^o \|_2} \) in Table 1. As can be seen, for the low level noise, our InISSAPL algorithm with smaller \( p \) generates better results. However, with \( p = 0.5 \), it is more robust to different levels of noise.

**Table 1:** Relative \( L_2 \) errors of LASSO and InISSAPL for \( \ell_p-\ell_2 \) minimization with different \( p \).

| \( \kappa = 500 \) | LASSO | \( p = 0.1 \) | \( p = 0.3 \) | \( p = 0.5 \) | \( p = 0.7 \) | \( p = 0.9 \) |
|------------------|--------|-------------|-------------|-------------|-------------|-------------|
| \( \sigma = 0.01 \) | 0.05633 | 0.01379 | 0.01669 | 0.01719 | 0.01589 | 0.02118 |
| \( \sigma = 0.1 \) | 0.26325 | 0.20197 | 0.17989 | 0.17310 | 0.18830 | 0.21276 |
6.2 Robust recovery from measurements with heavy-tailed noise

In the third example, we tested our InISSAPL algorithm for \( \ell_p-\ell_1 \) minimization. We chose \( p = 0.5 \). We generated a \( 256 \times 512 \) random Gaussian \( A \) and a true signal \( x^o \) of the sparsity \( \kappa = 25 \). The \( Ax^o \) was corrupted by impulsive noise obeying the standard \((0, 1)\)-Cauchy distribution, which is scaled by a factor of \( 10^{-2} \). Figure 1(a) shows both the noiseless and noisy observations. The noisy observed signal approximates closely the noiseless observation almost everywhere except two outliers at the entries 124 and 249. Figure 1(b)-(d) show the recovered signals by LASSO, InISSAPL for \( \ell_p-\ell_2 \) minimization, and InISSAPL for \( \ell_p-\ell_1 \) minimization. Note that LASSO and \( \ell_p-\ell_2 \) minimization are very sensitive to the outliers and failed to reconstruct the signal. As can be seen in Figure 1(d), the \( \ell_p-\ell_1 \) minimization is able to recover the sparse signal with high accuracy.

![Graphs showing noiseless and noisy observed signals, recovered signals by LASSO, \( \ell_p-\ell_2 \) via InISSAPL, and \( \ell_p-\ell_1 \) via InISSAPL.](image)

Figure 1: Reconstructions of a sparse signal from measurements corrupted by impulsive noise using different methods. \( \ell_p-\ell_1 \) minimization via InISSAPL is more robust to this case.
7 Conclusions

We proposed an iterative support shrinking algorithm for non-Lipschitz $\ell_p$-$\ell_q$ minimization. The proposed algorithm overcomes the non-Lipschitzian by iteratively adding constraints on the support to the original problem. It is a new type of reweighted $\ell_1$ algorithm. The algorithm is easy to implement. The subproblem in each iteration is solved inexactly by ADMM. We proved the global convergence of the iterative sequence, whose limit is a stationary point of the $\ell_p$-$\ell_q$ objective function. We also showed a more practical lower bound theory of the iterates. Numerical experiments demonstrated the performance of the algorithm. Due to the successive size reduction of the subproblem, it has good potentials in applications for large scale sparse signal recovery problems.

8 Appendix

We recall some definitions and results here.

Definition 8.1 (Subdifferentials [39]). Let $h : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous function.

(i) The regular subdifferential of $h$ at $x \in \text{dom } h = \{x \in \mathbb{R}^N : h(x) < +\infty\}$ is defined as

$$\hat{\partial}h(x) := \left\{ v \in \mathbb{R}^N : \liminf_{x \rightarrow \bar{x}} \frac{h(x) - h(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\};$$

(ii) The (limiting) subdifferential of $h$ at $\bar{x} \in \text{dom } h$ is defined as

$$\partial h(\bar{x}) := \left\{ v \in \mathbb{R}^N : \exists x^{(k)} \rightarrow \bar{x}, h(x^{(k)}) \rightarrow h(\bar{x}), v^{(k)} \in \hat{\partial}h(x^{(k)}), v^{(k)} \rightarrow v \right\}.$$

Remark. Form Definition 8.1 the following properties hold:

(i) For any $\bar{x} \in \text{dom } h$, $\hat{\partial}h(\bar{x}) \subset \partial h(\bar{x})$. If $h$ is continuously differentiable at $\bar{x}$, then $\hat{\partial}h(\bar{x}) = \partial h(\bar{x}) = \{\nabla h(\bar{x})\}$;

(ii) For any $\bar{x} \in \text{dom } h$, the subdifferential set $\partial h(\bar{x})$ is closed, i.e.,

$$\left\{ v \in \mathbb{R}^N : \exists x^{(k)} \rightarrow \bar{x}, h(x^{(k)}) \rightarrow h(\bar{x}), v^{(k)} \in \partial h(x^{(k)}), v^{(k)} \rightarrow v \right\} \subset \partial h(\bar{x}).$$

The foundational works on the Kurdyka-Łojasiewicz (KL) property are due to Łojasiewicz [34] and Kurdyka [31]. For the development of the application of KL property in optimization theory, see [5,12,6] and reference therein.

Definition 8.2 (Kurdyka-Łojasiewicz Property [11]). A proper function $h$ is said to have the Kurdyka-Łojasiewicz property at $\bar{x} \in \text{dom } \partial h = \{x \in \mathbb{R}^N : \partial h(x) \neq \emptyset\}$ if there exist $\zeta \in (0, +\infty)$, a neighborhood $U$ of $\bar{x}$, and a continuous concave function $\varphi : [0, \zeta) \rightarrow \mathbb{R}_+$ such that

(i) $\varphi(0) = 0$;

(ii) $\varphi(0)$ is $C^1$ on $(0, \zeta)$;

(iii) for all $s \in (0, \zeta)$, $\varphi'(s) > 0$;

(iv) for all $x \in U$ satisfying $h(\bar{x}) < h(x) < h(\bar{x}) + \zeta$, the Kurdyka-Łojasiewicz inequality holds:

$$\varphi'(h(x) - h(\bar{x})) \text{dist}(0, \partial h(x)) \geq 1,$$

where $\text{dist}(0, \partial h(x)) = \min\{\|v\| : v \in \partial h(x)\}$. 

16
A proper, lower semicontinuous function $h$ satisfying the KL property at all points in $\text{dom } \partial h$ is called a KL function. One can refer to \cite{1,2,6} for examples of KL functions. For this paper, the function $E$ satisfies the KL property \cite{2}.

References

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res., 35(2):438–457, April 30 2010.

[2] H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program., 137(1-2):91–129, 2013.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci., 2(1):183–202, 2009.

[4] W. Bian and X. Chen. Worst-case complexity of smoothing quadratic regularization methods for non-Lipschitzian optimization. SIAM J. Optim., 23(3):1718–1741, 2013.

[5] J. Bolte, A. Daniilidis, and A. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim., 17(4):1205–1223, 2006.

[6] J. B. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program., 146(1-2):459–494, 2014.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn., 3(1):1–122, Jan. 2011.

[8] K. Bredies, D. A. Lorenz, and S. Reiterer. Minimization of non-smooth, non-convex functionals by iterative thresholding. J. Optim. Theory Appl., 165(1):78–112, Apr. 2015.

[9] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev., 51(1):34–81, 2009.

[10] J. Cai, S. Osher, and Z. Shen. Convergence of the linearized bregman iteration for $\ell_1$-norm minimization. Math. Comput., 78(268):2127–2136, 2009.

[11] J. Cai, S. Osher, and Z. Shen. Linearized bregman iterations for compressed sensing. Math. Comput., 78(267):1515–1536, 2009.

[12] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory, 52(2):489–509, Feb 2006.

[13] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted $\ell_1$ minimization. J. Fourier Anal. Appl., 14(5):877–905, Dec 2008.

[14] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vision, 40(1):120–145, 2011.

[15] R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. Inverse Problems, 24(3):035020, 14, 2008.

[16] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pages 3869–3872, March 2008.
[17] X. Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Math. Program.*, 134(1):71–99, 2012.

[18] X. Chen, L. Niu, and Y. Yuan. Optimality conditions and a smoothing trust region newton method for nonLipschitz optimization. *SIAM J. Optim.*, 23(3):1528–1552, July 2013.

[19] X. Chen, F. Xu, and Y. Ye. Lower bound theory of nonzero entries in solutions of $\ell_2$-$\ell_p$ minimization. *SIAM J. Sci. Comput.*, 32(5):2832–2852, 2010.

[20] X. Chen and W. Zhou. Convergence of the reweighted $\ell_1$ minimization algorithm for $\ell_2$-$\ell_p$ minimization. *Comput. Optim. Appl.*, 59(1):47–61, Oct 2014.

[21] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[22] I. Daubechies, R. A. Devore, M. Fornasier, and C. S. Gunter. Iteratively reweighted least squares minimization for sparse recovery. *Comm. Pure Appl. Math.*, 63(1):1–38, 2010.

[23] T. E. Dielman. Least absolute value regression: recent contributions. *J. Stat. Comput. Simul.*, 75(4):263–286, 2005.

[24] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[25] S. Foucart and M.-J. Lai. Sparest solutions of underdetermined linear systems via $\ell_q$-minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, 26(3):395 – 407, May 2009.

[26] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–343, 2009.

[27] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: a reweighted minimum norm algorithm. *IEEE Trans. Signal Process.*, 45(3):600–616, Mar 1997.

[28] B. He and X. Yuan. On the $O(1/n)$ convergence rate of the douglas-crachford alternating direction method. *SIAM J. Numer. Anal.*, 50(2):700–709, 2012.

[29] H. P. J. Robust statistics. Wiley, New York, 1981.

[30] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *Proc. 22nd Int. Conf. Neural Information Processing Systems*, pages 1033–1041, 2009.

[31] K. Kurdyka. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier (Grenoble)*, 48(3):769–783, 1998.

[32] M.-J. Lai and J. Wang. An unconstrained $\ell_q$ minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems. *SIAM J. Optim.*, 21(1):82–101, 2011.

[33] M.-J. Lai, Y. Xu, and W. Yin. Improved iteratively reweighted least squares for unconstrained smoothed $\ell_q$ minimization. *SIAM J. Numer. Anal.*, 51(2):927–957, 2013.

[34] S. Lojasiewicz. Une propriétée topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles (Paris, 1962)*, pages 87–89. Éditions du Centre National de la Recherche Scientifique, Paris, 1963.

[35] Z. Lu. Iterative reweighted minimization methods for $\ell_p$ regularized unconstrained nonlinear programming. *Math. Program.*, 147(1):277–307, Oct 2014.
[36] J. Lv and Y. Fan. A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.*, 37(6A):3498–3528, 2009.

[37] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.

[38] J. L. Paredes and G. R. Arce. Compressive sensing signal reconstruction by weighted median regression estimates. *IEEE Trans. Signal Process.*, 59(6):2585–2601, 2011.

[39] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1998.

[40] Q. Sun. Recovery of sparsest signals via $\ell_q$-minimization. *Appl. Comput. Harmon. Anal.*, 32(3):329 – 341, May 2012.

[41] J. A. Tropp and S. J. Wright. Computational methods for sparse solution of linear inverse problems. *Proc. IEEE*, 98(6):948–958, June 2010.

[42] C. Wu and X.-C. Tai. Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Imaging Sci.*, 3(3):300–339, 2010.

[43] Z. Xu, X. Chang, F. Xu, and H. Zhang. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Trans. Neural Netw. Learn. Syst.*, 23(7):1013–1027, Jul 2012.

[44] M. Yan and W. Yin. Self equivalence of the alternating direction method of multipliers. In R. Glowinski, S. J. Osher, and W. Yin, editors, *Splitting Methods in Communication, Imaging, Science, and Engineering*. Scientific Computation. Springer, Cham, 2016.

[45] H. Zou and R. Li. One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.*, 36(4):1509–1533, Aug. 2008.

[46] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang. A generalized iterated shrinkage algorithm for non-convex sparse coding. In *Proc. IEEE Int. Conf. Computer Vision*, pages 217–224, Dec 2013.