Iterative Reweighted Methods for $\ell_1 - \ell_p$ Minimization

Xianchao Xiu*, Lingchen Kong†, Yan Li‡, and Houduo Qi§

(First version: Dec. 16, 2014; Final version: November 12, 2017)

Abstract

In this paper, we focus on the $\ell_1 - \ell_p$ minimization problem with $0 < p < 1$, which is challenging due to the ℓ_p norm being non-Lipschizian. In theory, we derive computable lower bounds for nonzero entries of the generalized first-order stationary points of $\ell_1 - \ell_p$ minimization, and hence of its local minimizers. In algorithms, based on three locally Lipschitz continuous ϵ-approximation to ℓ_p norm, we design several iterative reweighted ℓ_1 and ℓ_2 methods to solve those approximation problems. Furthermore, we show that any accumulation point of the sequence generated by these methods is a generalized first-order stationary point of $\ell_1 - \ell_p$ minimization. This result, in particular, applies to the iterative reweighted ℓ_1 methods based on the new Lipschitz continuous ϵ-approximation introduced by Lu [20], provided that the approximation parameter ϵ is below a threshold value. Numerical results are also reported to demonstrate the efficiency of the proposed methods.

Keywords: $\ell_1 - \ell_p$ minimization, generalized first-order stationary point, lower bound, iterative reweighted ℓ_1 method, iterative reweighted ℓ_2 method

1 Introduction

Compressed Sensing (CS) has been used in the fields of finance, econometrics, signal processing, machine learning, and so on. The mathematical model of CS can be described as the following ℓ_0 minimization

$$\min \|x\|_0 \quad \text{s.t. } Ax = b, \tag{1}$$

where $\|x\|_0$ is ℓ_0 function which counts the number of nonzero entries of decision variable x, and $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are given measurement matrix and vector, respectively. However, the problem is provably NP-hard due to the combinatorial nature of ℓ_0 function. Fortunately, Candès and Tao [7], and Donoho [13] creatively recommended to surrogate ℓ_0 function with ℓ_1 norm, which leads to ℓ_1 minimization

$$\min \|x\|_1 \quad \text{s.t. } Ax = b, \tag{2}$$

where $\|x\|_1$ is the ℓ_1 norm defined to be the sum of absolute values of all entries. Many researchers have made lots of contributions related to the theory, algorithms as well as applications to problem (1) or (2). See, e.g., [5, 6, 7, 8, 13, 14, 24, 25].

* Xianchao Xiu, Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, P. R. China (e-mail: xianchaoxiu@163.com)

† Lingchen Kong, Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, P. R. China (e-mail: konglchen@126.com)

‡ Yan Li, School of Insurance and Economics, University of International Business and Economics, Beijing 100029, P. R. China (e-mail:liyan2010@uibe.edu.cn)

§School of Mathematics, University of Southampton, UK (e-mail: hdqi@soton.ac.uk).

In fact, ℓ_1 norm is a loose approximation of ℓ_0 function and often leads to an over-regularized problem. Consequently, some further improvements are required. Among such efforts, a very natural improvement is the suggestion of the use of ℓ_p $(0 < p < 1)$ norm, which yields ℓ_p minimization

$$\min \|x\|_p^p \quad \text{s.t. } Ax = b,$$

where $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. As we know, ℓ_p is not a true norm, and it is generally nonconvex, nonsmooth, non-Lipschitz, and NP-hard (see [9, 17]). It is ℓ_0 function when $p \to 0$, and ℓ_1 norm when $p \to 1$.

Based on the idea of regularization, the above optimization problem can be modified as the unconstrained $\ell_2 - \ell_p$ minimization

$$\min \ \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_p^p \tag{3}$$

for some $\lambda > 0$. Over the past few years various researchers have made systematic attempts for $\ell_2 - \ell_p$ minimization (3), including the sparse recovery conditions and the algorithms for approximate sparse solution (see, e.g., [9, 10, 11, 12, 16, 17, 18, 19, 20]). In particular, Chen et al. [10] derived lower bounds for nonzero entries of local minimizers and also proposed a hybrid orthogonal matching pursuit-smoothing gradient method. Since $\|x\|_p^p$ is non-Lipschitz continuous, Chen and Zhou [11], Lai and Wang [18], Lai et al. [19] considered different approximations of $\|x\|_p^p$, and proposed iterative reweighted ℓ_1 and ℓ_2 methods to solve this approximation problem. Recently, Lu [20] provided a unified convergence analysis for the generalized $\ell_2 - \ell_p$ minimization

$$\min \ f(x) + \lambda\|x\|_p^p, \tag{4}$$

where $f(x)$ is convex, smooth, and has L_f-Lipschitz-continuous gradient. It is obvious that (3) is a special case of (4) when $f(x) = \frac{1}{2}\|Ax - b\|_2^2$. He proposed new variants, where each subproblem has a closed form solution. Moreover, Lu [20] developed new iterative reweighted methods for (4) and showed that any accumulation point of the sequence generated by these methods is a first-order stationary point, provided that the approximation parameter is below a threshold value. It is worth mentioning that ϵ is not necessarily approaching zero. The natural question is whether we can generalize the results in Lu [20] when $f(x)$ is a nonsmooth function, especially $f(x) = \|Ax - b\|_1$?

This question is meaningful because $f(x) = \|Ax - b\|_1$ is the least absolute deviation in statistics and it has been widely used in dealing with the situation where noises are heavy-tailed or heterogeneous. That is to say, compared with ℓ_2 minimization, ℓ_1 minimization is more robust where the noises are not normal distribution. For details, see, e.g. [2, 3, 15, 21, 22] and references therein. Note that $\|Ax-b\|_1$ is convex but not smooth, and it is a special case of the quantile function. Zhang et al. [26] studied the quantile function by introducing its smoothing functions, and established a smoothing iterative method for quantile regression with nonconvex ℓ_p penalty.

In this paper, we mainly concentrate on the following $\ell_1 - \ell_p$ minimization

$$\min \ \|Ax - b\|_1 + \lambda\|x\|_p^p. \tag{5}$$

In order to generalize the results in Lu [20], we first characterize the definition of the generalized first-order stationary point of (5). We then derive its lower bounds and local minimizers. We consider three locally Lipschitz continuous ϵ-approximation to ℓ_p norm. According to these approximation functions, we design their corresponding iterative reweighted ℓ_1 and ℓ_2 methods to solve these problems, and variants of them with closed form subproblem solutions. We establish that any accumulation point of the sequence generated by these methods is a generalized first-order stationary point. In particular, for the iterative reweighted ℓ_1 methods based on the new Lipschitz continuous ϵ-approximation introduced by Lu [20], we prove that any accumulation point of the sequence generated by these methods is a generalized first-order stationary point provided that the approximation parameter ϵ is below a threshold value.

The rest of this paper is organized as follows. We discuss the ϵ-approximation problems in Section 2. Several iterative reweighted ℓ_1 and ℓ_2 minimization methods are proposed in Section 3 and Section 4 including their convergence results. Section 5 is devoted to the numerical results to demonstrate the efficiency of our methods. Section 6 concludes the paper.

2 Preliminary

This section studies the generalized first-order stationary point of $\ell_1 - \ell_p$ minimization (5) and establishes its lower bound for nonzero entries. We then introduce its Locally Lipschtiz continuous ϵ-approximation by a nonconvex $\ell_{(p,\epsilon)}$-regularized function of $\|x\|_p^p$.

2.1 Notation

Given $x^* \in \mathbb{R}^n$, $T = \{i : x_i^* \neq 0\}$ denotes the support set of x^*, and $\overline{T}$ denotes its complement in $\{1, \ldots, n\}$. Let ϵ and ϵ_i be positive scalars in $\mathbb{R}$, while ϵ^k is a sequence of positive vectors in $\mathbb{R}^n$. Suppose $h(x)$ is convex, then the subdifferential of $h(x)$ at x is defined as

$$\partial h(x) = \{\xi \in \mathbb{R}^n \mid h(y) \geq h(x) + \xi^T(y - x)\}.$$

In addition, we denote $A = [A_1, A_2, \cdots, A_n] \in \mathbb{R}^{m\times n}$, where $A_i \in \mathbb{R}^m$ for $i = 1, 2, \cdots, n$. Let $\text{sgn}(x) = (\text{sgn}(x_1), \text{sgn}(x_2), \cdots, \text{sgn}(x_n))$, where the sgn operator is given by

$$\text{sgn}(x_i) = \begin{cases} 1 & \text{if } x_i > 0, \\ [-1,\, 1] & \text{if } x_i = 0, \\ -1 & \text{if } x_i < 0. \end{cases}$$

From the chain rule of subdifferential (e.g., [4, Theorem 3.3.5]), we can derive that

$$\partial(\|Ax - b\|_1) = A^T \text{sgn}(Ax - b).$$

In particular, we will often use the notation $(A^T\text{sgn}(Ax^* - b))_i$, which is the set of consisting of all the ith elements in the set $A^T\text{sgn}(Ax^* - b)$. It is worth to point out that this set in general can not be characterized by $A_i^T\text{sgn}(Ax^* - b)$. That is,

$$(A^T\text{sgn}(Ax^* - b))_i \neq A_i^T\text{sgn}(Ax^* - b).$$

2.2 Lower bound for nonzero entries of stationary point of (5)

We below consider the lower bound for nonzero entries of the generalized first-order stationary point of $\ell_1 - \ell_p$ minimization (5), which is motivated by Chen et al. [10] and Lu [20] for $\ell_2 - \ell_p$ minimization. Different from the smooth function $\|Ax - b\|_2$, $\|Ax - b\|_1$ is nonsmooth. Hence, we begin with defining the generalized first-order stationary point of problem (5).

Definition 2.1 *Suppose that x^* is a vector in $\mathbb{R}^n$. We say that x^* is a generalized first-order stationary point of* (5) *if*

$$0 \in (A^T \mathrm{sgn}(Ax^* - b))_i x_i^* + \lambda p |x_i^*|^p, \quad i = 1, 2, \cdots, n. \tag{6}$$

We are able to show that any local minimizer of (5) is also a generalized first-order stationary point as defined above.

Proposition 2.2 *Suppose that x^* is a local minimizer of (5). Then x^* is a generalized first-order stationary point, that is, (6) holds at x^*.*

Proof Suppose that x^* is a local minimizer of (5). It is easy to see that x^* is also a local minimizer of

$$\min_{x \in \mathbb{R}^n} \{\|Ax - b\|_1 + \lambda \|x\|_p^p \mid x_i = 0, \ i \notin T\}. \tag{7}$$

Note that the objective function of (7) is locally Lipschitz continuous at x^* in the subspace T. Applying the first-order optimality condition of (7), we obtain that

$$0 \in (A^T \mathrm{sgn}(Ax^* - b))_i + \lambda p |x_i^*|^{p-1} \mathrm{sgn}(x_i^*), \ \forall i \in T.$$

Then, multiplying x_i^* on both sides of the above relation gives

$$0 \in (A^T \mathrm{sgn}(Ax^* - b))_i x_i^* + \lambda p |x_i^*|^p, \ \forall i \in T.$$

From the fact $x_i^* = 0$ for $i \notin T$, we clarify that the above relation also holds for $i \notin T$. Hence, (6) holds and the proof is completed. □

We next establish a lower bound for the nonzero entries of the generalized first-order stationary points, and hence of local minimizers of problem (5). For problems (3) and (4), Chen et al. [10] and Lu [20] derived some interesting lower bounds for the nonzero entries of local minimizers, respectively.

Theorem 2.3 *Suppose that x^* is a generalized first-order stationary point of problem (5). Then the following statement holds:*

$$|x_i^*| \ \geq \ (\frac{\lambda p}{\|A_i\|_1})^{\frac{1}{1-p}}, \quad \forall i \in T. \tag{8}$$

Proof From the definition of the generalized first-order stationary point, we obtain that

$$0 \in (A^T \mathrm{sgn}(Ax^* - b))_i + \lambda p |x_i^*|^{p-1} \mathrm{sgn}(x_i^*), \ \forall i \in T.$$

For $i \in T$, we have $\mathrm{sgn}(x_i^*) \neq 0$ and $|\mathrm{sgn}(x_i^*)| = 1$. This together with the above relation derives

$$-\lambda p |x_i^*|^{p-1} \in \mathrm{sgn}(x_i^*)(A^T \mathrm{sgn}(Ax^* - b))_i, \quad \forall i \in T,$$

and

$$|\mathrm{sgn}(x_i^*)(A^T \mathrm{sgn}(Ax^* - b))_i| = |(A^T \mathrm{sgn}(Ax^* - b))_i| \leq \|A_i\|_1, \forall i \in T.$$

Hence it holds that

$$\lambda p |x_i^*|^{p-1} \leq \|A_i\|_1, \forall i \in T,$$

which implies

$$|x_i^*| \geq \left(\frac{\lambda p}{\|A_i\|_1}\right)^{\frac{1}{1-p}}, \forall i \in T.$$

Thus we have completed the proof. □

2.3 Locally Lipschitz continuous ϵ-approximation to (5)

Since $\|x\|_p^p$ is non-Lipschitz continuous at some points which contain zero entries, we know that its Clarke subdifferential does not exist at theses points and this poses difficulties to design algorithms for solving (5). In this subsection we introduce the smoothing $\ell_{(p,\epsilon)}$-regularized version based on an ϵ-approximation function of $\|x\|_p^p$. We consider the following ϵ-approximations to $\|x\|_p^p$: $\sum_{i=1}^n(|x_i|+\epsilon_i)^p$, $\lambda\sum_{i=1}^n(|x_i|^2+\epsilon_i)^{p/2}$ and $\lambda\sum_{i=1}^n h_{u_\epsilon}(x_i)$ where

$$h_{u_\epsilon}(x_i) := \min_{0\le s\le u_\epsilon} p\left(|x_i|s - \frac{p-1}{p}s^{\frac{p}{p-1}}\right), \qquad u_\epsilon := \left(\frac{\epsilon}{\lambda n}\right)^{\frac{p-1}{p}}. \tag{9}$$

As a consequence, by replacing $\|x\|_p^p$ with its ϵ-approximations, we obtain the $\ell_1 - \ell_{(p,\epsilon)}$ problems

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_1 + \lambda\sum_{i=1}^n(|x_i|+\epsilon_i)^p,$$

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_1 + \lambda\sum_{i=1}^n(|x_i|^2+\epsilon_i)^{p/2},$$

and

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_1 + \lambda\sum_{i=1}^n h_{u_\epsilon}(x_i).$$

For (3), Chen and Zhou [11], Lai and Wang [18], Lai et al. [19] considered the corresponding ϵ-approximation problems. They also proposed iterative reweighted ℓ_1 minimization methods to solve them. It is worth noting that Lu [20] extended those methods to (4) by introducing the new ϵ-approximation function $h_{u_\epsilon}(x_i)$ of $|x_i|^p$, which has the following interesting proposition.

Proposition 2.4 *(Lemma 2.4 (3) in [20]) The Clarke subdifferential of h_{u_ϵ}, denoted by ∂h_{u_ϵ}, exists everywhere, and it is given by*

$$\partial h_{u_\epsilon}(x_i) = \begin{cases} p|x|^{p-1}\mathrm{sgn}(x_i), & if\ |x_i| > u_\epsilon^{\frac{1}{p-1}}, \\ pu_\epsilon\mathrm{sgn}(x_i), & if\ |x_i| \le u_\epsilon^{\frac{1}{p-1}}. \end{cases}$$

The next theorem shows that when ϵ is below a threshold value, the generalized stationary point of the corresponding ϵ-approximation problems is also that of the original problem (5).

Theorem 2.5 *Let ϵ be a constant such that*

$$0 < \epsilon < \lambda n\left(\frac{\|A_i\|_1}{\lambda p}\right)^{\frac{p}{p-1}}. \tag{10}$$

Suppose that x^* is a generalized first-order stationary point of

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_1 + \lambda\sum_{i=1}^n h_{u_\epsilon}(x_i).$$

Then, x^* is also a generalized first-order stationary point of (5), i.e., (6) holds at x^*. Moreover, the nonzero entries of x^* satisfy the lower bound property (8).

Proof From the assumption, we know

$$0 \in \partial[\|Ax^* - b\|_1 + \lambda \sum_{i=1}^{n} h_{u_\epsilon}(x_i^*)],$$

that is

$$0 \in (A^T \text{sgn}(Ax^* - b))_i + \lambda \partial h_{u_\epsilon}(x_i^*), \quad \forall i \in T \tag{11}$$

which implies

$$|\partial h_{u_\epsilon}(x_i^*)| \le \frac{\|A_i\|_1}{\lambda}, \quad \forall i \in T. \tag{12}$$

Next, we will prove that $|x_i^*| > u_\epsilon^{\frac{1}{p-1}}$ for all $i \in T$. Suppose $0 < |x_i^*| \le u_\epsilon^{\frac{1}{p-1}}$ for some $i \in T$, then following from Proposition 2.4, we have $|\partial h_{u_\epsilon}(x_i^*)| = pu_\epsilon$. This together with (10) and the definition of u_ϵ derives

$$|\partial h_{u_\epsilon}(x_i^*)| \; = pu_\epsilon = p\left(\frac{\epsilon}{\lambda n}\right)^{\frac{p-1}{p}} \; > \; \frac{\|A_i\|_1}{\lambda}.$$

This is in contraction to (12), hence $|x_i^*| > u_\epsilon^{\frac{1}{p-1}}$ for all $i \in T$. So we conclude that $\partial h_{u_\epsilon}(x_i^*) = p|x_i^*|^{p-1}\text{sgn}(x_i^*)$ for every $i \in T$. It follows from (11) that

$$0 \in (A^T \text{sgn}(Ax^* - b))_i + \lambda p |x_i^*|^{p-1} \text{sgn}(x_i^*), \quad \forall i \in T.$$

Multiplying x_i^* on both sides of the above relation, we get

$$0 \in (A^T \text{sgn}(Ax^* - b))_i x_i^* + \lambda p |x_i^*|^p, \quad \forall i \in T.$$

From the fact $x_i^* = 0$ for $i \notin T$, we clarify that the above relation also holds for $i \notin T$. Hence, (6) holds. Moreover, applying this relation and Theorem 2.3, we immediately obtain the second part of this theorem. □

We end this section with the following corollary.

Corollary 2.6 *Suppose that x^* is a local minimizer of*

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1 + \lambda \sum_{i=1}^{p} h_{u_\epsilon}(x_i).$$

Then x^* is a generalized first-order stationary point of (5), i.e., (6) holds at x^*. Moreover, the nonzero entries of x^* satisfy the lower bound (8).

Proof Since x^* is a local minimizer of

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1 + \lambda \sum_{i=1}^{p} h_{u_\epsilon}(x_i),$$

we know that x^* is one of its stationary points. Then the desired conclusion immediately follows from Theorem 2.5. □

3 Iterative reweighted methods based on $F_{\alpha,\epsilon}$

In order to solve the nonsmooth, nonconvex, and non-Lipschitz $\ell_1 - \ell_p$ minimization (5), we employ its ϵ-approximation minimization problems as mentioned in the previous section. We will give three iterative reweighted methods. We now begin with introducing the related nonconvex $\ell_1 - \ell_{(p,\epsilon)}$ problems as follows

$$\min_{x\in\mathbb{R}^n} F_{\alpha,\epsilon}(x) \quad := \quad \|Ax-b\|_1 + \lambda\sum_{i=1}^{n}(|x_i|^\alpha + \epsilon_i)^{\frac{p}{\alpha}}, \tag{13}$$

$$\min_{x\in\mathbb{R}^n} F_{\epsilon}(x) \quad := \quad \|Ax-b\|_1 + \lambda\sum_{i=1}^{n} h_{u_\epsilon}(x_i), \tag{14}$$

where $\alpha = 1$ or 2. For solving the $\ell_2 - \ell_p$ minimization (3), two types of iterative reweighted methods have been proposed in the literature [11, 12, 16, 18]. Recently, Lu [20] provided a unified convergence analysis for the generalized version (4).

This section will focus on the iterative reweighted methods based on the problem (13) and we will consider the problem (14) in the next section. Here, we give two types of iterative reweighted ℓ_α methods, as well as the corresponding convergence analysis. In what follows, we denote $f(x) := \|Ax-b\|_1$ for easy of description, and its subdifferential $\partial f(x)$ at $x = x^*$ with $(\partial f(x^*))_i = (A^T\mathrm{sgn}(Ax^* - b))_i$.

3.1 The first type IRL$_\alpha$ method

The main idea is that problem (5) can be solved by applying the iterative reweighted ℓ_α methods to a sequence of problems (13) with $\epsilon = \epsilon^k \to 0$ as $k \to \infty$, where $\{\epsilon^k\}$ is a sequence of positive vectors. These extend the iterative reweighted methods for $\ell_2 - \ell_p$ minimization (3) proposed in [11, 18], which apply an iterative reweighted ℓ_1 or ℓ_2 method to solve a sequence of ϵ-approximation minimization problems, to $\ell_1 - \ell_p$ minimization (5).

Before giving the first type iterative reweighted ℓ_α method for (5), we need to introduce an iterative reweighted ℓ_α method for solving problem (13).

Algorithm 1: An iterative reweighted ℓ_α method for (13)

Choose an arbitrary x^0. Set $k = 0$.

1) Solve the weighted ℓ_α minimization problem

$$x^{k+1} \in \operatorname{argmin}\left\{f(x) + \tfrac{\lambda p}{\alpha}\textstyle\sum_{i=1}^n s_i^k|x_i|^\alpha\right\},$$

where $s_i^k = (|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}$ for all i.

2) Set $k \leftarrow k+1$ and go to step 1).

End

Since $f(x) := \|Ax-b\|_1$ is nonsmooth, we define the generalized first-order stationary point for the problem (13) by making use if the subdifferential of f.

Definition 3.1 *We say that a vector $x^* \in \Re^n$ is a generalized first-order stationary point of (13) if*

$$0 \in x_i^*(\partial f(x^*))_i + \lambda p(|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}|x_i^*|^\alpha, \quad i = 1, 2, \cdots, n.$$

We will show that any accumulation point of $\{x^k\}$ generated by Algorithm 1 is a generalized first-order stationary point of (13).

Theorem 3.2 *Let the sequence $\{x^k\}$ be generated by Algorithm 1. Suppose that x^* is an accumulation point of $\{x^k\}$. Then x^* is a generalized first-order stationary point of (13).*

Proof The proof is inspired by that of Theorem 3.1 in Lu [20], but we adopt the technique of nonsmooth analysis. Let q be such that

$$\frac{\alpha}{p} + \frac{1}{q} = 1. \tag{15}$$

For any $\delta > 0$, it is easy to find that

$$(|x|^\alpha + \delta)^{\frac{p}{\alpha}} = \frac{p}{\alpha} \min_{s \geq 0} \left\{ (|x|^\alpha + \delta)s - \frac{s^q}{q} \right\}, \quad \forall x \in \mathbb{R}.$$

Indeed, the minimum is achieved at $s = (|x|^\alpha + \delta)^{\frac{1}{q-1}}$. Together with the definition of s^k and (15), we have

$$s^k = \text{argmin}_{s \geq 0}\, G_{\alpha,\epsilon}(x^k, s), \quad x^{k+1} \in \text{argmin}_x\, G_{\alpha,\epsilon}(x, s^k),$$

where

$$G_{\alpha,\epsilon}(x, s) = f(x) + \frac{\lambda p}{\alpha} \sum_{i=1}^n \left[(|x_i|^\alpha + \epsilon_i)s_i - \frac{s_i^q}{q} \right].$$

We can obtain that $F_{\alpha,\epsilon}(x^k) = G_{\alpha,\epsilon}(x^k, s^k)$. Thus,

$$F_{\alpha,\epsilon}(x^{k+1}) = G_{\alpha,\epsilon}(x^{k+1}, s^{k+1}) \leq G_{\alpha,\epsilon}(x^{k+1}, s^k) \leq G_{\alpha,\epsilon}(x^k, s^k) = F_{\alpha,\epsilon}(x^k). \tag{16}$$

Thus, $\{F_{\alpha,\epsilon}(x^k)\}$ is non-increasing. Since x^* is an accumulation point of $\{x^k\}$, there is a subsequence $\{x^k\}_K$ such that $\{x^k\}_K \to x^*$. Moreover, $F_{\alpha,\epsilon}$ is continuous and monotonic, hence $F_{\alpha,\epsilon}(x^k) \to F_{\alpha,\epsilon}(x^*)$. By the definition of s^k, we have $\{s^k\}_K \to s^*$, where $s_i^* = (|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}$. Also, it is not hard to see that $F_{\alpha,\epsilon}(x^*) = G_{\alpha,\epsilon}(x^*, s^*)$. From (16) and $F_{\alpha,\epsilon}(x^k) \to F_{\alpha,\epsilon}(x^*)$, we obtain that $G_{\alpha,\epsilon}(x^{k+1}, s^k) \to F_{\alpha,\epsilon}(x^*) = G_{\alpha,\epsilon}(x^*, s^*)$. Moreover, it holds by (16) that

$$G_{\alpha,\epsilon}(x, s^k) \geq G_{\alpha,\epsilon}(x^{k+1}, s^k) \quad \forall x \in \mathbb{R}^n.$$

Taking limits on both sides of this inequality as $k \in K \to \infty$, we have

$$G_{\alpha,\epsilon}(x, s^*) \geq G_{\alpha,\epsilon}(x^*, s^*) \quad \forall x \in \mathbb{R}^n.$$

That is to say, $x^* \in \text{argmin}_{x \in \mathbb{R}^n}\, G_{\alpha,\epsilon}(x, s^*)$. Using the first-order optimality condition and the definition of x^*, we can derive

$$0 \in (\partial f(x^*))_i + \lambda p(|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1} |x_i^*|^{\alpha-1} \text{sgn}(x_i^*), \quad \forall i. \tag{17}$$

Therefore, we complete the proof by multiplying x^*. □

Proposition 3.3 *Let $\delta > 0$ be arbitrarily given, and let the sequence $\{x^k\}$ be generated by Algorithm 1 for (13). Suppose that $\{x^k\}$ has at least one accumulation point. Then, there exist some x^k and $g_k \in \partial f(x^k)$ such that*

$$|x_i^k (g_k)_i + \lambda p |x_i^k|^\alpha (|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}| \ \le \ \delta, \quad \forall i.$$

Proof Without loss of generality, let $\lim_{k\to\infty}\{x^k\} = x^*$. From Theorem 3.2, we know that for the sequence $\{x^k\}$

$$0 \in (\partial f(x^k))_i + \lambda p(|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}|x_i^k|^{\alpha-1}\mathrm{sgn}(x_i^k), \quad \forall i.$$

Then there exists $g_k \in \partial f(x^k)$ such that

$$(g_k)_i + \lambda p(|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}|x_i^k|^{\alpha-1}\mathrm{sgn}(x_i^k) = 0, \quad \forall i.$$

For $i \in T$, since $\lim_{k\to\infty} x_i^k = x_i^*$ and $x_i^* \neq 0$, we know that $\lim_{k\to\infty}(g_k)_i$ exists and

$$\lim_{k\to\infty}(g_k)_i + \lambda p(|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}|x_i^*|^{\alpha-1}\mathrm{sgn}(x_i^*) = 0, \quad \forall i \in T.$$

Multiplying x_i^* on both sides of the above relation, we obtain

$$\lim_{k\to\infty}(g_k)_i x_i^* + \lambda p(|x_i^*|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}|x_i^*|^{\alpha} = 0, \ \forall i \in T.$$

That is, there exist some x^k and $g_k \in \partial f(x^k)$ such that

$$|x_i^k (g_k)_i + \lambda p |x_i^k|^\alpha (|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}| \ \le \ \delta, \quad \forall i \in T.$$

Clearly, the above relation also holds for $i \notin T$ since $\lim_{k\to\infty} x_i^k = x_i^*$ and $x_i^* = 0$ for $i \notin T$. □

Applying the above proposition, we are ready to establish the first type of iterative reweighted ℓ_α method for problem (5) with each subproblem being in the form of (13) and being solved by the iterative reweighted ℓ_α method. Combining Algorithm 1 and Proposition 3.3, we have the following Algorithm 2.

Algorithm 2: The first type iterative reweighted ℓ_α method for (5)

Let $\{\delta_k\}$ be a sequence of positive scalars and let $\{\epsilon^k\}$ be a sequence of positive vectors.

Set $k = 0$.

1) Apply the Algorithm 1 to problem (13) with $\epsilon = \epsilon^k$ for finding x^k satisfying

$$|x_i^k (g_k)_i + \lambda p |x_i^k|^\alpha (|x_i^k|^\alpha + \epsilon_i)^{\frac{p}{\alpha}-1}| \ \le \ \delta_k, \quad \forall i,$$

where $g_k \in \partial f(x^k)$.

2) Set $k \leftarrow k + 1$ and go to step 1).

End

For Algorithm 2, we establish its convergence result as follows.

Theorem 3.4 *Let $\{\delta_k\}$ and $\{\epsilon^k\}$ be sequences of positive scalars and vectors such that $\{\delta_k\} \to 0$ and $\{\epsilon^k\} \to 0$, respectively. Let $\{x^k\}$ be a sequence of points generated by Algorithm 2. Suppose that x^* is an accumulation point of $\{x^k\}$. Then x^* is a generalized first-order stationary point of (5), i.e., (6) holds at x^*.*

Proof From Algorithm 2, we have

$$\left| x_i^k (g_k)_i + \lambda p |x_i^k|^\alpha (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}-1} \right| \leq \delta_k, \quad \forall i \in T, \tag{18}$$

where $g_k \in \partial f(x^k)$. Because x^* is an accumulation point of $\{x^k\}$, there exists a subsequence $\{x^k\}_K$ such that $\{x^k\}_K \to x^*$. Taking limits on both sides of (18) as $k \in K \to \infty$, we obtain that

$$0 \in x_i^* (\partial f(x^*))_i + \lambda p |x_i^*|^p, \quad \forall i \in T.$$

From the fact $x_i^* = 0$ for $i \notin T$, we clarify that the above relation also holds for $i \notin T$. Hence, x^* satisfies (6) and it is also a generalized first-order stationary point of (5). □

3.2 The second type iterative reweighted ℓ_α method for (5)

We continue with stating the second type iterative reweighted ℓ_α methods for $\ell_1 - \ell_p$ minimization (5). Indeed, problem (5) can also be solved by a sequence ϵ^k-approximation problems, which is motivated by the idea employed in the iterative reweighted ℓ_1 and ℓ_2 methods for $\ell_2 - \ell_p$ minimization (3) in [12, 16], and the generalized version (4) in [20].

Algorithm 3: The second type iterative reweighted ℓ_α method for (5)

Let $\{\epsilon^k\}$ be a sequence of positive vectors in $\mathbb{R}^n$. Choose an arbitrary x^0. Set $k = 0$.

1 Solve the weighted ℓ_α minimization problem

$$x^{k+1} \in \operatorname{argmin} \left\{ f(x) + \tfrac{\lambda p}{\alpha} \textstyle\sum_{i=1}^n s_i^k |x_i|^\alpha \right\},$$

where $s_i^k = (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}-1}$ for all i.

2) Set $k \leftarrow k+1$ and go to step 1).

End

We below show that any accumulation point of $\{x^k\}$ generated by Algorithm 3 is a generalized stationary point of (5).

Theorem 3.5 *Suppose that $\{\epsilon^k\}$ is a sequence of non-increasing positive vectors in $\mathbb{R}^n$ and $\epsilon^k \to 0$ as $k \to \infty$. Let $\{x^k\}$ be the sequence generated by Algorithm 3. Suppose that x^* is an accumulation point of $\{x^k\}$. Then, x^* is a stationary point of (5).*

Proof From the definition of $G_{\alpha,\epsilon}(\cdot,\cdot)$, one can observe that $G_{\alpha,\epsilon^k}(x^{k+1}, s^k) \leq G_{\alpha,\epsilon^k}(x^k, s^k)$ and $G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^{k+1}) = \inf_{s \geq 0} G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s)$. Hence,

$$G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^{k+1}) \leq G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^k).$$

In addition, since $s^k > 0$ and $\{\epsilon^k\}$ is component-wise non-increasing, it is easy to see that $G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^k) \leq G_{\alpha,\epsilon^k}(x^{k+1}, s^k)$. Using these relations, we have

$$G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^{k+1}) \;\leq\; G_{\alpha,\epsilon^{k+1}}(x^{k+1}, s^k) \;\leq\; G_{\alpha,\epsilon^k}(x^{k+1}, s^k) \;\leq\; G_{\alpha,\epsilon^k}(x^k, s^k), \quad \forall k \geq 0. \tag{19}$$

Therefore, $\{G_{\alpha,\epsilon^k}(x^k, s^k)\}$ is non-increasing. Because x^* is an accumulation point of $\{x^k\}$, there exists a subsequence $\{x^k\}_K$ such that $\{x^k\}_K \to x^*$. From the definition of s^k, it holds that

$$G_{\alpha,\epsilon^k}(x^k, s^k) = f(x^k) + \lambda \sum_{i=1}^n (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}}.$$

It then follows from $\{x^k\}_K \to x^*$ and $\epsilon^k \to 0$ that $\{G_{\alpha,\epsilon^k}(x^k, s^k)\}_K \to f(x^*) + \lambda \|x^*\|_p^p$. Together with the monotonicity of $\{G_{\alpha,\epsilon^k}(x^k, s^k)\}$, we obtain $G_{\alpha,\epsilon^k}(x^k, s^k) \to f(x^*) + \lambda \|x^*\|_p^p$. Combining this relation and (19), we derive

$$G_{\alpha,\epsilon^k}(x^{k+1}, s^k) \to f(x^*) + \lambda \|x^*\|_p^p. \tag{20}$$

We claim that

$$x^* \in \operatorname{argmin}_{x_{\overline{T}}=0} \left\{ f(x) + \frac{\lambda p}{\alpha} \sum_{i \in T} |x_i|^\alpha |x_i^*|^{p-\alpha} \right\}. \tag{21}$$

Thus, $\{s_i^k\}_K \to |x_i^*|^{p-\alpha}, \ \forall i \in T$. Furthermore, we observe that

$$0 \;\leq\; \frac{p}{\alpha} \sum_{i \in \overline{T}} \left[\epsilon_i^k s_i^k - \frac{(s_i^k)^q}{q} \right] \;\leq\; \frac{p}{\alpha} \sum_{i \in \overline{T}} \left[(|x_i^k|^\alpha + \epsilon_i^k) s_i^k - \frac{(s_i^k)^q}{q} \right] \;=\; \sum_{i \in \overline{T}} (|x_i^k|^\alpha + \epsilon_i^k)^{\frac{p}{\alpha}},$$

which, together with $\epsilon^k \to 0$ and $\{x_i^k\}_K \to 0$ for $i \in \overline{T}$, means that

$$\lim_{k \in K, k \to \infty} \sum_{i \in \overline{T}} \left[\epsilon_i^k s_i^k - \frac{(s_i^k)^q}{q} \right] = 0. \tag{22}$$

In addition, we know that $G_{\alpha,\epsilon^k}(x, s^k) \;\geq\; G_{\alpha,\epsilon^k}(x^{k+1}, s^k)$. Then for every $x \in \mathbb{R}^n$ such that $x_{\overline{T}} = 0$, we have

$$f(x) + \frac{\lambda p}{\alpha} \sum_{i \in T} \left[(|x_i|^\alpha + \epsilon_i^k) s_i^k - \frac{(s_i^k)^q}{q} \right] + \frac{\lambda p}{\alpha} \sum_{i \in \overline{T}} \left[\epsilon_i^k s_i^k - \frac{(s_i^k)^q}{q} \right] \;=\; G_{\alpha,\epsilon^k}(x, s^k) \;\geq\; G_{\alpha,\epsilon^k}(x^{k+1}, s^k).$$

Taking limits on both sides of this inequality as $k \in K \to \infty$, and using (20), (22) and the fact that $\{s_i^k\}_K \to |x_i^*|^{p-\alpha}, \ \forall i \in T$, we obtain that

$$f(x) + \frac{\lambda p}{\alpha} \sum_{i \in T} \left[|x_i|^\alpha |x_i^*|^{p-\alpha} - \frac{|x_i^*|^{q(p-\alpha)}}{q} \right] \;\geq\; f(x^*) + \lambda \|x^*\|_p^p$$

for all $x \in \mathbb{R}^n$ such that $x_{\overline{T}} = 0$. From inequality and (15), we immediately have (21). The first-order optimality condition of (21) and (15) lead to

$$0 \in x_i^* (\partial f(x^*))_i + \lambda p |x_i^*|^p \quad \forall i \in T.$$

Since $x_i^* = 0$ for $i \in \overline{T}$, we observe that the above relation also holds for $i \in \overline{T}$. Hence, x^* satisfies (6) and the desired conclusion follows. □

4 Iterative reweighted methods based on F_ϵ

This section will establish our iterative reweighted methods based on problem (14). As shown in Section 3, problem (14) has a locally Lipschitz continuous objective function and it is an ϵ-approximation to (2). Moreover, when ϵ is below a computable threshold value, a certain stationary point of (14) is also that of (5). According to this property, we propose an iterative reweighted ℓ_1 methods for solving (5), which can be viewed as the iterative reweighted ℓ_1 methods directly applied to problem (14). The novelty of these methods is in that the parameter ϵ is chosen only once and then fixed throughout all iterations, while the iterative reweighted ℓ_α methods based on $F_{\alpha,\epsilon}$ studied in the preceding section require that the parameter ϵ be dynamically adjusted and approach zero. However, it still can be verified that any accumulation point of the sequence generated by these methods is a generalized first-order stationary point of (5). This shows that the interesting properties of iterative reweighted ℓ_1 methods by Lu [20] still hold for the corresponding methods of $\ell_1 - \ell_p$ minimization (5).

We now state the iterative reweighted ℓ_1 methods based on F_ϵ as follows. Here, we set q as

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{23}$$

Algorithm 4: The iterative reweighted ℓ_1 method for (5)

Let q be defined in (23). Choose an arbitrary $x^0 \in \mathbb{R}^n$ and ϵ such that (10) holds.

Set $k = 0$.

1) Solve the weighted ℓ_1 minimization problem

$$x^{k+1} \in \operatorname{argmin}_x \left\{ f(x) + \lambda p \textstyle\sum_{i=1}^n s_i^k |x_i| \right\},$$

where $s_i^k = \min\left\{ (\frac{\epsilon}{\lambda n})^{\frac{1}{q}}, |x_i^k|^{\frac{1}{q-1}} \right\}$ for all i.

2) Set $k \leftarrow k+1$ and go to step 1).

End

The following theorem claims that any accumulation point of $\{x^k\}$ generated by Algorithm 4 is a generalized first-order stationary point of (5).

Theorem 4.1 *Let $\{x^k\}$ be the sequence generated by Algorithm 4. Assume that ϵ satisfies (10). Suppose that x^* is an accumulation point of $\{x^k\}$. Then x^* is a generalized first-order stationary point of (5), i.e., (6) holds at x^*. Moreover, the nonzero entries of x^* satisfy the lower bound (8).*

Proof Let $u_\epsilon = (\frac{\epsilon}{\lambda n})^{1/q}$ and

$$G(x,s) = f(x) + \lambda p \sum_{i=1}^n \left[|x_i| s_i - \frac{s_i^q}{q} \right]. \tag{24}$$

It is not hard to see that

$$s^k = \operatorname{argmin}_{0 \le s \le u_\epsilon} G(x^k, s), \quad x^{k+1} \in \operatorname{argmin}_x G(x, s^k). \tag{25}$$

Furthermore, $F_\epsilon(x) = \min_{0\le s\le u_\epsilon} G(x,s)$ and $F_\epsilon(x^k) = G(x^k, s^k)$. It then follows that

$$F_\epsilon(x^{k+1}) \;=\; G(x^{k+1}, s^{k+1}) \;\le\; G(x^{k+1}, s^k) \;\le\; G(x^k, s^k) \;=\; F_\epsilon(x^k). \tag{26}$$

Hence, $\{F_\epsilon(x^k)\}$ is non-increasing. Because x^* is an accumulation point of $\{x^k\}$, there exists a subsequence $\{x^k\}_K$ such that $\{x^k\}_K \to x^*$. Moreover, F_ϵ is continuous and monotonic, hence $F_\epsilon(x^k) \to F_\epsilon(x^*)$. Let $s_i^* = \min\{u_\epsilon, |x_i^*|^{\frac{1}{q-1}}\}$ for all i. We then observe that $\{s^k\}_K \to s^*$ and $F_\epsilon(x^*) = G(x^*, s^*)$. Using (26) and $F_\epsilon(x^k) \to F_\epsilon(x^*)$, we see that $G(x^{k+1}, s^k) \to F_\epsilon(x^*) = G(x^*, s^*)$. In addition, it follows from (25) that

$$G(x, s^k) \;\ge\; G(x^{k+1}, s^k) \quad \forall x \in \mathbb{R}^n.$$

Taking limits on both sides of this inequality as $k \in K \to \infty$, we obtain

$$G(x, s^*) \;\ge\; G(x^*, s^*) \quad \forall x \in \mathbb{R}^n,$$

which yields

$$x^* \in \operatorname{argmin}\left\{ f(x) + \lambda p \sum_{i=1}^n s_i^* |x_i| \right\}. \tag{27}$$

The first-order optimality condition of (27) is

$$0 \in (\partial f(x^*))_i + \lambda p s_i^* \operatorname{sgn}(x_i^*), \quad \forall i. \tag{28}$$

Recall that $s_i^* = \min\{u_\epsilon, |x_i^*|^{\frac{1}{q-1}}\}$, which together with (23) implies that for all i,

$$s_i^* = \begin{cases} |x_i^*|^{p-1}, & \text{if } |x_i^*| > u_\epsilon^{q-1}, \\ u_\epsilon, & \text{if } |x_i^*| \le u_\epsilon^{q-1}. \end{cases}$$

Replacing it into (28), gives

$$0 \in (\partial f(x^*))_i + \lambda \partial h_{u_\epsilon}(x_i^*), \quad \forall i.$$

It then follows that x^* is a generalized first-order stationary point of F_ϵ. Using these results and Theorem 2.5, we conclude that x^* is a generalized first-order stationary point of (5). The rest of conclusion immediately follows from Theorem 2.3. □

5 Numerical Study

In this section, we conduct numerical experiments to compare the performance of the iterative reweighted ℓ_1 methods (Algorithm 2, Algorithm 3, Algorithm 4) studied in Subsections 3.1 and 3.2 and Section 4. For the sake of fairness, we all run 100 times to illustrate the efficiency of the above algorithms. Meanwhile, we use Alg. to denote Algorithm for convenience of presentation. All experiments are implemented in MATLAB on a desktop computer with Intel Core I5 2.60GHz CPU and 8GB of RAM.

We now address the initialization and the termination criteria for our proposed iterative reweighted ℓ_1 methods applied to the $\ell_1 - \ell_p$ minimization (5). In particular, we use the following MATLAB code to generate the decision variable x, the measurement matrix A, the log-normal noise w, and the measurement vector b:

$x = zeros(n,1);\ \ v = randperm(n);\ \ x(v(1:k)) = 2 * randn(k,1);$

$A = randn(m,n);\ \ A = orth(A);\ \ w = 10^{-3} * lognrnd(0,1,m,1);$

$b = A * x + w.$

The same initial point x^0 is used for all methods. Notice that, we choose x^0 as a solution of

$$\min \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1,$$

which can be computed by a variety of methods. In this paper, we adopt FISTA proposed in Beck and Teboulle [1]. To validate these methods, all methods terminate according to the following criterion

$$\|A^T \text{sgn}(Ax - b)x + \lambda p|x|^p\|_\infty \leq 10^{-4}.$$

We set $\epsilon^k = 0.1^k e$ and $\delta_k = 0.1^k$ for Alg. 2, and $\epsilon^k = 0.5^k e$ for Alg. 3, respectively, where e is the all-ones vector. As remarked in Section 2, for Alg. 4, ϵ is chosen as the one satisfying (10). The regularization parameter λ is chosen to be 10^{-3}. In addition, each subproblem of the algorithms is solved by the spectral projected gradient (SPG) method [23] with $\alpha = 1$.

We next conduct numerical experiments to test the performance of Alg. 2 – Alg. 4 for solving (5). We compare the objective value and CPU time obtained by these methods. The computational results of the three iterative reweighted ℓ_1 methods and their variants are presented in Table 1 for $p = 0.1$ and Table 2 for $p = 0.5$, respectively. In detail, the dimensions of each data set are given in the first two columns. The middle four columns list the objective value which is defined above. In addition, the CPU time is reported in the last four columns. From the above numerical experiments, we can observe that: the new iterative reweighted ℓ_1 method (i.e, Alg. 4) has smaller objective value than FISTA and the other two iterative reweighted ℓ_1 methods, namely, Alg. 2 and Alg. 3. Furthermore, the new iterative reweighted ℓ_1 method (i.e, Alg. 4) is usually faster than FISTA, Alg. 2 and Alg. 3.

Table 1: Comparison of FISTA and three iterative reweighted ℓ_1 methods with $p = 0.1$

Scale		Objective Value				CPU Time			
m	n	FISTA	Alg. 2	Alg. 3	Alg. 4	FISTA	Alg. 2	Alg. 3	Alg. 4
100	500	1.7657	1.5666	1.3445	0.8233	0.0149	0.0171	0.0140	0.0119
200	1000	2.8363	2.5920	1.7763	1.3731	0.0224	0.0238	0.0198	0.0188
300	1500	3.2738	3.2067	2.9865	2.0257	0.0398	0.0449	0.0395	0.0387
400	2000	4.4914	5.1432	3.6068	2.6324	0.0572	0.0650	0.0582	0.0530
500	2500	6.3543	6.5100	4.5689	3.4269	0.0845	0.0872	0.0841	0.0825
600	3000	7.6703	7.8139	6.5679	4.0010	0.1074	0.1102	0.0924	0.0898
700	3500	8.9000	9.0091	7.6402	4.7255	0.1220	0.1269	0.1207	0.1140
800	4000	10.3874	10.4071	8.7117	5.0410	0.1598	0.1706	0.1468	0.1413
900	4500	11.0344	11.6670	9.7155	5.9991	0.1843	0.1947	0.1763	0.1707
1000	5000	12.7837	13.2954	11.7446	6.7973	0.2192	0.2637	0.2020	0.1937

Table 2: Comparison of FISTA and three iterative reweighted ℓ_1 methods with $p = 0.5$

Scale		Objective Value				CPU Time			
m	n	FISTA	Alg. 2	Alg. 3	Alg. 4	FISTA	Alg. 2	Alg. 3	Alg. 4
100	500	0.5836	0.6080	0.5762	0.5221	0.0149	0.0152	0.0089	0.0030
200	1000	0.8514	1.1474	1.1770	0.7563	0.0224	0.0215	0.0171	0.0121
300	1500	1.7702	1.8072	1.6325	1.1288	0.0398	0.0364	0.0252	0.0215
400	2000	2.3643	2.5384	2.1090	1.5618	0.0572	0.0569	0.0388	0.0362
500	2500	3.0050	3.0225	2.8328	1.8968	0.0845	0.0705	0.0570	0.0521
600	3000	3.4831	3.5585	3.1589	2.2723	0.1074	0.0914	0.0754	0.0715
700	3500	4.0306	4.0773	3.6546	2.6107	0.1220	0.1224	0.1123	0.0825
800	4000	4.4564	4.7423	4.2087	3.0702	0.1598	0.1413	0.1140	0.1088
900	4500	5.2710	5.7123	5.5077	3.5192	0.1843	0.1864	0.1610	0.1459
1000	5000	6.0417	6.2578	6.0745	4.2749	0.2192	0.2365	0.1758	0.1651

6 Concluding remarks

In this paper we have dealt with $\ell_1 - \ell_p$ minimization (5). By introducing the definition of generalized first-order stationary points, we have derived lower bounds of their nonzero entries and local minimizers of (5). We then proposed the iterative reweighted ℓ_1 and ℓ_2 methods and their variants for $\ell_1 - \ell_p$ minimization (5), and established their convergent results. In particular, for the proposed iterative reweighted ℓ_1 methods based on ϵ-approximation function h_{u_ϵ} introduced by Lu [20], we have proved that any accumulation point of the sequence generated by these methods is a generalized first-order stationary point, provided that the approximation parameter ϵ is below a threshold value. We demonstrated the efficiency of our proposed methods through numerical experiments and comparison

Acknowledgments The authors would like to thank two anonymous referees and the associate editor for their numerous insightful comments and suggestions, which have greatly improved the paper. In particular, we would like to thank a referee for clarifying the chain rule for computing the subdifferential of the ℓ_1 norm composited with a linear mapping. The work was supported in part by the National Natural Science Foundation of China (11431002, 11671029), the 111 Project of China (B16002) and the Fundamental Research Funds for the Central Universities (KSJB16013536)

References

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

[2] A. Belloni and V. Chernozhukov. L1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.*, 39: 82-130, (2011).

[3] J. Bradic, J. Fan and W. Wang. Penalized composite quasi-likelihood for ultrahigh-dimensional variable selection. *J. Royal Statist. Soc., Ser. B*, 73: 325-349, (2011).

[4] J.M. Borwein and A.S Lewis. Convex Analysis and Nonlinear Optimization: Theory and Examples. Canadian Mathematical Society, Springer 2000.

[5] A. Bruckstein, D. Donoho and A. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51: 34-81, (2009).

[6] E. 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, (2006).

[7] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12): 4203-4215, (2005).

[8] E. Candès, M. Walkin and S. Boyd. Enhancing sparsity by reweighted l_2 minimization. *J. Fourier Anal. Appl.*, 14(5-6): 877-905, (2008).

[9] X. Chen, D. Ge, Z. Wang and Y. Ye. Complexity of unconstrained L_2-L_p minimization. *Math. Program.*, 143(1-2): 371-383, (2014).

[10] X. Chen, F. Xu and Y. Ye. Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. *SIAM J. Sci. Comput.*, 32: 2832-2852, (2010).

[11] X. Chen and W. Zhou. Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization. Department of Applied Mathematics, The Hong Kong Polytechnic University, (2010).

[12] I. Daubechies, R. DeVore, M. Fornasier and C. Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure. Appl. Math.*, 63: 1-38, (2010).

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

[14] Y. Eldar and G. Kutyniok. Compressed sensing: theory and applications. Cambridge University Press, (2012).

[15] Y. Fan, J. Fan and E. Barut. Adaptive robust variable selection. *Ann. Statist.*, 42(1): 324-351, (2014).

[16] S. Foucart and M. Lai. Sparsest solutions of underdetermined linear systems via ℓ_p-minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, 26: 395-407, (2009).

[17] D. Ge, X. Jiang and Y. Ye. A note on the complexity of ℓ_p minimization. *Math. Program.*, 21: 1721-1739, (2011).

[18] M. Lai and J. Wang. An unconstrained l_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems. *SIAM J. Optim.*, 21: 82-101, (2010).

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

[20] Z. Lu. Iterative reweighted minimization methods for ℓ_p regularized unconstrained nonlinear programming. *Math. Program.*, 147(1-2): 277-307,(2014).

[21] L. Wang. L1 penalized LAD estimator for high dimensional linear regression. *J. Multi. Anal.*, 120: 135-151, (2013).

[22] L. Wang, Y. Wu and R. Li. Quantile regression for analyzing heterogeneity in ultrahigh dimension. *J. Ameri. Statist. Assoc.*, 107: 214-222, (2012).

[23] S. Wright, R. Nowak and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Image Process.*, 57: 24792493, (2009).

[24] X. Xiu and L. Kong. Rank-min-one and sparse tensor decomposition for surveillance video. *Pac. J. Optim.*, 11(2): 403-418, (2015).

[25] W. Yin, S. Osher, D. Goldfarb and J. Darbonet. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM J. Imaging Sci.*, 1(1): 143-168, (2008).

[26] S. Zhou, Y. Li, L. Kong and L. Zhang. A smoothing iterative method for quantile regression with nonconvex ℓ_p penalty. *J. Ind. Manag. Optim.*, 13(1): 93-112, (2017).