AN EFFICIENT ALGORITHM FOR NON-CONVEX SPARSE OPTIMIZATION

YONG WANG
School of Mathematics, Tianjin University
Tianjin 300072, China

WANQUAN LIU*
Department of Computing
Curtin University, WA, 6102, Australia

GUANGLU ZHOU
Department of Mathematics and Statistics
Curtin University, WA, 6102, Australia

(Communicated by Changzhi Wu)

Abstract. It is a popular research topic in computer vision community to find a solution for the zero norm minimization problem via solving its non-convex relaxation problem. In fact, there are already many existing algorithms to solve the non-convex relaxation problem. However, most of them are computationally expensive due to the non-Lipschitz property of this problem and thus these existing algorithms are not suitable for many engineering problems with large dimensions.

In this paper, we first develop an efficient algorithm to solve the non-convex relaxation problem via solving a sequence of non-convex sub-problems based on our recent work. To this end, we reformulate the minimization problem into another non-convex one but with non-negative constraint. Then we can transform the non-Lipschitz continuous non-convex problem with the non-negative constraint into a Lipschitz continuous problem, which allows us to use some efficient existing algorithms for its solution. Based on the proposed algorithm, an important relation between the solutions of relaxation problem and the original zero norm minimization problem is established from a different point of view. The results in this paper reveal two important issues: i) The solution of non-convex relaxation minimization problem converges to the solution of the original problem; ii) The general non-convex relaxation problem can be solved efficiently with another reformulated high dimension problem with non-negative constraint. Finally, some numerical results are used to demonstrate effectiveness of the proposed algorithm.

1. Introduction. Sparse signal representation with a prior basis has attracted much attention recently in both optimization and computer vision communities. Mathematically, it is to find a solution of the zero norm minimization problem with linear equality constraint. Such a model has been applied to many fields, such as signal processing [16, 17], image deconvolution and denoising [39, 6], face recognition.

2010 Mathematics Subject Classification. Primary: 65L09, 90C30, 65H10, 68W25, 68W01.
Key words and phrases. Non-Lipschitz optimization, SMM method, non-convex problem with non-negative constraint, sparse optimization, compressing sensing.

* Corresponding author: Wanquan Liu.
and compressive sensing [9]. However, this combinatorial optimization problem is proved to be NP-hard [29, 22] and efficient algorithms are required to solve it.

In order to find an approximate solution of the zero norm minimization problem, there are two classes of popular methods: convex relaxation method and non-convex relaxation method named by virtue of the convexity of the replacement of the zero norm, respectively. Many algorithms have been proposed to solve the convex relaxation problem [2, 8, 20, 23, 25, 38, 41] and the non-convex relaxation problem [34, 35, 26, 44, 19, 21, 30, 14, 10, 13, 28, 31, 7, 24, 43], respectively. It has been shown that the condition to find a sparse solution of the original problem by solving its non-convex relaxation $l_p$ problem is weaker than that by solving its convex relaxation one ($l_1$ problem) [36, 5]. However, due to the non-Lipschitz continuity of the non-convex relaxation $l_p$ problem, it is more difficult to solve than solving the convex relaxation one.

In our prior work [40], we have considered a class of non-convex $l_p$ minimization problems with non-negative constraint, which can be regarded as the non-convex relaxation of the $l_p$ norm minimization problem with non-negative constraint. In order to solve such a problem, an equivalent Lipschitz problem is presented in [40]. Based on this Lipschitz reformulation, an efficient algorithm is developed by using the classical limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method. The numerical results show that the proposed algorithm based on the reformulated Lipschitz problem is much faster than other non-convex relaxation algorithms and can produce more sparse solution than some typical convex relaxation algorithms.

In this paper, we aim to remove the non-negative constraint and will develop a new method to find a sparse solution of the $l_p$ norm minimization problem without non-negative constraint based on our previous work [40] and also we can establish a relationship between the zero norm minimization problem and its non-convex relaxation one in a different perspective from [18]. In detail, we can summarize the contribution of this paper as follows.

(i) We reformulate the non-convex relaxation model of the $l_p$ norm minimization problem into an augmented non-convex one with non-negative constraint, and prove the equivalence between them. Then, we propose an effective algorithm to solve the reformulated non-negative $l_p$ minimization problem and find its approximate solution.

(ii) For the internal relation between the non-convex relaxation minimization problem and the zero norm minimization problem, we prove that every solution to the non-convex relaxation problem with small parameter sufficiently close to zero also solves the zero norm minimization problem. This assertion was proved similarly in [18, 32] in a different framework.

(iii) We validate the proposed algorithm for signal reconstruction via experimental results. The numerical results show that the proposed method for the non-convex relaxation $l_p$ problem is more effective in terms of recovering the sparse signal and solving some large size problems in comparison with some typical existing non-convex relaxation $l_p$ algorithms.

The rest of the paper is organized as follows. In Sect. 2, we present some related optimization models and review some basic results in [40]. In Sect. 3, we reformulate the non-convex $l_p$ problem into a non-convex one with non-negative constraint. Then we prove the equivalence of these two problems and develop an efficient algorithm to solve the reformulated optimization problem. Finally the relationship
of the non-convex minimization problem and the zero norm minimization problem is established. In Sect. 4, some numerical experiments for signal reconstruction are conducted in order to show the efficiency of the proposed method and Sect. 5 concludes the paper.

2. Preliminary results. In this section, we present some related optimization models and review some main results on non-negative $l_p$ minimization problem in [40] and this paves the way for our main results in next section.

Throughout the paper, all vectors are column vectors; the superscript $T$ denotes the transpose. Let $n$ be a positive integer, we denote $\mathcal{I} = \{1, 2, \ldots, n\}$, $\mathbb{R}^n$ the space of $n$-dimensional real column vectors and $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n | x \geq 0\}$. For any $u \in \mathbb{R}^n$, $u_i$ denotes the $i$-th element of $u$. The symbol “$\circ$” denotes the Hadamard product; that is, for any vector $u, v \in \mathbb{R}^n$, $(u \circ v)_i = u_i v_i$, $i \in \mathcal{I}$. For simplicity, we use $(u, v)^T$ to denote the vector $(u^T, v^T)^T$. The bold $1$ is used to denote one vector with all elements of 1 with proper dimension.

A sparse signal representation problem can be modeled as the following optimization problem

$$\min_x \{ \|x\|_0 \ | \ Ax = b \}$$

where $A$ is the basis matrix with $m$ row and $n$ column, $b \in \mathbb{R}^m$ is the observed signal and $x \in \mathbb{R}^n$ is the representation of $b$. $\|x\|_0$ is called zero norm of $x$, which denotes the number of nonzero elements in the vector $x$.

The problem (1) is NP-hard [29, 22]. In order to solve it, there are usually two alternative schemes. One is to replace the $l_0$ norm of (1) with the $l_1$ norm and consider the following convex optimization problem [9, 15]:

$$\min_x \{ \|x\|_1 \ | \ Ax = b \}$$

where $\|x\|_1 = \sum_{i=1}^n |x_i|$. The problem (2) is convex and many methods have been proposed to solve it, such as $l_1$-magic [8], GPSR [20], FPC [23], LDM [38], ADMM [2, 41].

Another alternative scheme is to use the $l_p$ norm to approximate the $l_0$ norm and consider the following optimization problem: 

$$\min_x \{ \|x\|_p \ | \ Ax = b \}$$

where $0 < p < 1$ and $\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$. It is shown that recovering a sparse signal by the model (3) is more economical than that by the model (2) [36, 5]. However, since $0 < p < 1$, (3) is still a non-convex and non-Lipschitz optimization problem, which implies that it is more difficult to solve than (2). Currently there are some algorithms for solving (3) and they can be classified in three classes [26]: the iteratively thresholding method (ITM) [26, 34, 35, 44], the iterative reweighted $l_1$ (IRL1) minimization methods [14, 19, 21, 30] and the iterative reweighted least square (IRLS) methods [10, 13, 28, 31]. In [7], a smoothing algorithm based on a smoothing function for the absolute value function is proposed to solve (3). In addition, for some special values of $p$, analytic solutions of (3) are available [24, 43].

In particular, for the non-negative case, we have considered the following $l_p$ minimization problem in [40]

$$\min_x \left\{ \sum_{i=1}^n x_i^p \ | \ Ax = b, \ x \geq 0 \right\}$$

(4)
where $p \in (0, 1)$ and $x \geq 0$ denotes that each element of $x$ is non-negative. More generally, we can consider the following problem

$$\min_{x} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \tau \sum_{i=1}^{n} x_i^p \right\} \tag{6}$$

where $0 < p < 1$, $\tau > 0$ is the penalty factor. As for problems (4)-(6), the following lemma establishes a connection among them and the proof can be found in [40].

Lemma 2.1. If $x^*$ is a solution of (6), there exists an $\eta \geq 0$ such that $x^*$ is also a solution of (5); In particular, when $\eta = 0$, $x^*$ is a solution of (4).

Obviously, (6) is a non-convex and non-Lipschitz continuous problem. Therefore, it is difficult to solve (6) by some existing efficient algorithms, such as the Newton method, pseudo Newton method, etc. In order to solve problem (6), we propose the following mathematical transformation. For a vector $x \in \mathbb{R}^n_+$ and $0 < p < 1$, let $u = x^p = (x_1^p, \ldots, x_n^p)^T$, and $u^1 = (u_1^1, \ldots, u_n^1)^T$. With above transformation, one can consider the following optimization problem

$$\min_{u \geq 0} \left\{ \frac{1}{2} \|Au^\frac{1}{p} - b\|_2^2 + \tau \sum_{i=1}^{n} u_i \right\}. \tag{7}$$

Now, we have the following important result and its proof is given in [40].

Lemma 2.2. For any $0 < p < 1$, if $u^*$ is a solution of (7), then $x^* = (u^*)^p$ is a solution of (6); Conversely, if $x^*$ is a solution of (6), then $u^* = (x^*)^\frac{1}{p}$ is a solution of (7).

Notice that the objective function in problem (7) as given below

$$f(u) := \frac{1}{2} \|Au^\frac{1}{p} - b\|_2^2 + \tau \sum_{i=1}^{n} u_i,$$

is differentiable and its derivative is given by

$$\nabla f(u) = \frac{1}{p} u^{\frac{1}{p} - 1} \circ [A^T(Au^\frac{1}{p} - b)] + \tau 1,$$

With these knowledge, one can design an efficient algorithm to solve problem (4) as described in [40]. Keep in mind that though the problem (4) can be solved efficiently by the proposed LLpM algorithm in [40], it has a non-negative constraint which will restrict its broad applications. We will remove this constraint in next section.

3. Main results. Now we consider to solve problem (3). Notice that when $0 < p < 1$, (3) is a non-smooth and non-Lipschitz optimization problem. It is extremely difficult to find a solution of (3) with some existing optimization algorithms. In this section, we would reformulate (3) into a Lipschitz optimization problem, so that it can be solved by some efficient optimization algorithms.
Let $B = [A - A] \in \mathbb{R}^{n \times 2n}$ and consider the following augmented optimization problem

$$\min_{z} \{1^T z^p | Bz = b, \ z \in \mathbb{R}^{2n}_+\}. \quad (8)$$

Now we can obtain the following important result.

**Lemma 3.1.** If $z^* = (z^*_1, z^*_2)^T \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is a global optimal solution of (8), then it follows that $(z^*_i)_i (z^*_j)_j = 0, i = 1, 2, ..., n.$

**Proof.** Suppose that there exists $i \in \mathcal{I}$ such that $(z^*_i)_i (z^*_j)_j > 0.$ We define $\tilde{z} = (\tilde{z}_1, \tilde{z}_2)^T \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ as follows:

$$(\tilde{z}_1)_j = (z^*_i)_j \quad \text{and} \quad (\tilde{z}_2)_j = (z^*_j)_j \quad \text{if} \quad j \neq i,$$

$$(\tilde{z}_1)_i = (z^*_i)_i - (z^*_j)_i \quad \text{and} \quad (\tilde{z}_2)_i = 0 \quad \text{if} \quad (z^*_i)_i \geq (z^*_j)_i,$$

$$(\tilde{z}_1)_i = 0 \quad \text{and} \quad (\tilde{z}_2)_i = (z^*_j)_i - (z^*_i)_i \quad \text{if} \quad (z^*_i)_i < (z^*_j)_i.$$

Then we have $\tilde{z}_1 - \tilde{z}_2 = z^*_1 - z^*_2$ and

$$B \tilde{z} = A(\tilde{z}_1 - \tilde{z}_2) = A(z^*_1 - z^*_2) = Bz^* = b.$$

Recall that

$$|(z^*_i)_i - (z^*_j)_j|^p < (z^*_1)_i^p + (z^*_2)_i^p,$$

one can obtain

$$1^T \tilde{z}^p = \sum_{j \neq i} [(\tilde{z}_1)_j^p + (\tilde{z}_2)_j^p] + (\tilde{z}_1)_i^p + (\tilde{z}_2)_i^p = \sum_{j \neq i} [(z^*_1)_j^p + (z^*_2)_j^p] + |(z^*_i)_i - (z^*_j)_j|^p < \sum_{j \neq i} [(z^*_1)_j^p + (z^*_2)_j^p] + (z^*_1)_i^p + (z^*_2)_i^p = \sum_{j} [(z^*_1)_j^p + (z^*_2)_j^p] = 1^T (z^*)^p.$$

This contradicts the fact that $z^*$ is an optimal solution of (8). So, it is true that $(z^*_1)^T \tilde{z}_2 = 0.$ This completes the proof. $\square$

From the above lemma, we can obtain the following corollary.

**Corollary 1.** Let $\mathcal{D} := \{Bz = b, z = (z_1, z_2) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+, z_1^T z_2 = 0\}.$ If $z^* = (z^*_1, z^*_2) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is an optimal solution of (8), $z^*$ must be an optimal solution of the following problem

$$\min_{z} \{1^T z^p | Bz = b, \ z \in \mathcal{D}\}. \quad (9)$$

Conversely, if $z^*$ is a solution of (9), it must be a solution of (8).

**Proof.** From Lemma 3.1, the result is obvious, and thus we omit the proof here. $\square$

Let $x_+ = \max\{x, 0\} \in \mathbb{R}^n_+$ and $x_- = \max\{-x, 0\} \in \mathbb{R}^n_+.$ Then we have the following important result on the equivalence of problems (3) and (8) in terms of solutions.

**Theorem 3.2.** Let $x^* = x^*_+ - x^*_-$ be a solution of (3), then $z^* = (x^*_+)^T$ is a solution of (8); Conversely, if $z^* = (z^*_1, z^*_2)^T \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is a solution of (8), then $x^* = z^*_1 - z^*_2$ is a solution of (3).

**Proof.** (i) $(3) \Rightarrow (8).$ From Corollary 1, it is sufficient to prove that $z^* = (x^*_+)^T$ is a solution of (9).

For any $i \in \mathcal{I}$, it follows that $(x^*_i)_i + (x^*_i)_i = (x^*_i)_i^p + (x^*_i)_i^p$ for any $0 < p < 1$. Therefore, we have

$$1^T (z^*)^p = \|(x^*_+)^T\|^p = \sum_{i=1}^n [(x^*_i)_i^p + (x^*_i)_i^p] = \|x^*\|^p.$$

$\Box$
and
\[ Bz^* = [A, -A] \begin{pmatrix} x^*_+ \\ x^*_- \end{pmatrix} = A(x^*_+ - x^*_-) = Ax^* = b. \]

On the other hand, for any \( z \in \mathcal{D} \), one has \( x = z_1 - z_2 \in \mathbb{R}^n \) and \( Ax = A(z_1 - z_2) = Bz = b \). Moreover, \( \|x\|^p = 1^T z^p \). Therefore, we have
\[ 1^T (z^*)^p = \|x^*\|^p \leq \|x\|^p = 1^T z^p. \]

That is, \( z^* \) is a solution of (9), and then a solution of (8).

(ii) \( (8) \Rightarrow (3) \). From Lemma 3.1, we have \( (z^*_1)Tz^*_2 = 0 \). That is, for a fixed index \( i \in \mathcal{I} \), at most only one does not equal to zero between \((z^*_1)_i\) and \((z^*_2)_i\).

Therefore, it follows that
\[ |(z^*_1)_i - (z^*_2)_i|^p = (z^*_1)_i^p + (z^*_2)_i^p, \quad \text{for any } i \in \mathcal{I}, \]

and then
\[ \|x^*\|^p = \|z^*_1 - z^*_2\|^p = \sum_{i=1}^n |z^*_1_i - z^*_2_i|^p = \sum_{i=1}^n |(z^*_1)_i - (z^*_2)_i|^p = \sum_{i=1}^n (z^*_1)_i^p + (z^*_2)_i^p = 1^T (z^*)^p. \]

Moreover, for any \( x \in \mathbb{R}^n \), \( z = (x_+, x_-) \in \mathcal{D} \) and \( \|x\|^p = 1^T z^p \), one has
\[ \|x^*\|^p = 1^T (z^*)^p \leq 1^T z^p = \|x\|^p. \]

That is, \( x^* = z^*_1 - z^*_2 \) is a solution of (3).

In summary, we complete the proof. \( \square \)

By virtue of Theorem 3.2, we can solve the general \( l_p \) optimization problem (3) by solving the equivalent \( l_p \) problem (8) with the non-negative constraint. Although the size of the augmented optimization problem is doubled than that of the original one, this is acceptable in many applications. On the other hand, one can see that (8) is a special case of (4). Taking fully into account the structure information of the model (4) and noting that an effective LLpM algorithm to solve (4) has been proposed, we can easily find an efficient algorithm for (8). For such purpose, we consider the following penalty form of (8):
\[ \min_{x \geq 0} \left\{ \frac{1}{2} \|Az - b\|^2 + \lambda \sum_{i=1}^n z_i^p \right\}, \tag{10} \]

where \( \lambda > 0 \) is the penalty factor. In order to avoid the non-Lipschitz continuity of the objective function of (10), we can equivalently reformulate (10) into the following optimization problem
\[ \min_{u \geq 0} \left\{ \frac{1}{2} \|Au^\frac{1}{p} - b\|^2 + \lambda \sum_{i=1}^n u_i \right\}. \tag{11} \]

where \( u = z^p = (z_1^p, \ldots, z_n^p)^T \) and \( u^\frac{1}{p} = (u_1^\frac{1}{p}, \ldots, u_n^\frac{1}{p})^T \). With the above transformation, we have the following important theorem as in [40].

**Theorem 3.3.** For any \( 0 < p < 1 \), if \( u^* \) is a solution of (11), then \( z^* = (u^*)^p \) is a solution of (10); Conversely, if \( z^* \) is a solution of (10), then \( u^* = (z^*)^\frac{1}{p} \) is a solution of (11).

**Proof.** The proof can be found in [40]. We omit it here. \( \square \)
In comparison with problem (10), the objective function of (11) is Lipschitz continuous with $0 < p < 1$. It is well-known that the Lipschitz continuity is a necessary condition for many existing efficient algorithms. Therefore, the reformulated model (11) is pivotal and this allows us to use some existing efficient algorithms to solve it directly.

Next we will investigate the relationship between problems of (3) and (1). This problem was once investigated in [18] with $x$ having a constraint and it is recently further investigated with similar results in [32]. We will investigate this problem from a different perspective without any constraint.

Let $F := \{ z \in \mathbb{R}^{2n} | Bz = b, z \geq 0 \}$, i.e. the feasible field of problem (8), and recall that $z$ is an extreme point of $F$ if there do not exist two different points $x, y \in F$ and a scale $\alpha \in (0, 1)$ such that $z = \alpha x + (1 - \alpha)y$ [33]. Then we have the following result.

**Lemma 3.4.** All solutions of (8) are extreme points of the set $F$.

**Proof.** Let $z$ be a solution of (8). If $z$ is not an extreme point of $F$, then there exist two distinct points $x, y \in F$ and a scalar $\alpha \in (0, 1)$ such that $z = \alpha x + (1 - \alpha)y$. Since $t^p$ is strictly concave in $t$ for any $p \in (0, 1)$, we have

$$
\|z\|_p^p = \sum_{i=1}^{2n} (\alpha x_i + (1 - \alpha)y_i)^p > \alpha \sum_{i=1}^{2n} x_i^p + (1 - \alpha) \sum_{i=1}^{2n} y_i^p \geq \|z\|_p^p,
$$

where the last inequality holds since $z$ is a solution of (8). This yields a contradiction. Therefore, $z$ is an extreme point of $F$ and the proof is complete. \qed

By Lemma 3.4, there exists at least one $p$ such that a solution of (8) is a solution of (1). In fact, we have the following more important result.

**Theorem 3.5.** Let $u^* = (u_1^*, u_2^*)^T \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ be a solution of (8) and $v^*$ be a solution of (1), then there exists $\gamma \in (0, 1)$ such that $\|u_1^* - u_2^*\|_0 = \|v^*\|_0$ for any $p \in (0, \gamma)$.

**Proof.** Because the number of extreme points of $F$ is finite, we assume that $\{x^k, k = 1, 2, \cdots, \} \}$ is the set of extreme points of $F$. By Theorem 3.2 and assumption that $u^* = (u_1^*, u_2^*)^T \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is a solution of (8), we can conclude that $u_1^* - u_2^*$ is a solution of (3). Then for $v^*$, it holds that

$$
\|v^*\|_p^p \geq \min_k \|x_1^k - x_2^k\|_p^p = \|u_1^* - u_2^*\|_p^p. \tag{12}
$$

Suppose that there does not exist a number $\gamma \in (0, 1)$ such that $\|u_1^* - u_2^*\|_0 = \|v^*\|_0$ holds for any $p \in (0, \gamma)$. Then there exist sequences $\{p_i\}$ and $\{x^k_i\}$ such that $0 < p_i < \gamma, p_i \to 0$ as $i \to +\infty$ and each $x^k_i = (x_1^k_i, x_2^k_i) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is a solution of the following problem

$$
\min_z \{1^T z| | Bz = b, z \in \mathbb{R}^{2n}_+ \}, \tag{13}
$$

and then $x^k_i$ is an extreme point of $F$ by Lemma 3.4.

On one hand, we can get $A(x_1^k - x_2^k) = b$, i.e., $x_1^k - x_2^k$ is a feasible point of (1). Since $v^*$ is a solution of (1), together with the assumption, it holds

$$
\|x_1^k - x_2^k\|_0 > \|v^*\|_0.
$$

On the other hand, since $x^k_i$ is a solution of (13), (12) implies that

$$
\|v^*\|_0 = \lim_{p_i \to 0^+} \|v^*\|_{p_i} \geq \lim_{p_i \to 0^+} \|x_1^k - x_2^k\|_{p_i}^{p_i} = \|x_1^k - x_2^k\|_0.
$$
This is a contradiction and the proof is complete.

Let \( \{p_k\} \) be a sequence satisfying \( \lim_{k \to \infty} p_k = 0 \) and \( u^k = (u^k_1, u^k_2)^T \) be a solution of (8) for each \( p_k \). From Theorem 3.5, the limit of \( u^k_1 - u^k_2 \) is a solution of (1). Based on this idea, we propose a new algorithm to solve the problem (1) approximately. In detail, we describe the algorithm as follows.

**Algorithm 3.1. Sequential Minimization Method (SMM).**

Given \( A \in \mathbb{R}^{m \times n} \), \( b \in \mathbb{R}^m \), \( p_{-1} = 1, p_0 \in (0, 1) \), \( \delta \in (0, p_0) \) and \( \lambda > 0 \). Choose an initial point \( x^0 \in \mathbb{R}^n \), \( \epsilon > 0 \) and set \( k = 0 \).

**Step 0.** Let \( B = [A - A] \), \( u^k = (x^k_+, x^k_-)^T \).

**Step 1.** Let \( p = p_k \) and

\[
\begin{align*}
 f(u^k) &= \frac{1}{2} \|B(u^k)^\frac{1}{2} - b\|^2_2 + \lambda \sum_{i=1}^{n} (u^k)_i, \\
 \nabla f(u^k) &= \frac{1}{p}(u^k)^{\frac{1}{2}-1} \circ [B^T(B(u^k)^\frac{1}{2} - b)] + \lambda 1.
\end{align*}
\]

**Step 2.** Use L-BFGS algorithm to solve (11) with \( u^k \), \( f(u^k) \) and \( \nabla f(u^k) \), and obtain a solution \( u^{k+1} := (u^{k+1}_1, u^{k+1}_2)^T \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \).

**Step 3.** If \( \|u^{k+1}_1\|_p^p - \|u^{k}_1\|_p^{p_k} \leq \epsilon \) or \( p_k \leq \delta \), go to Step 4; otherwise, let \( p_{k+1} = p_k - \delta \), \( k := k + 1 \) and go to Step 1.

**Step 4.** Output a solution \( x^* = (u^{k+1}_1)^{\frac{1}{2}} - (u^{k+1}_2)^{\frac{1}{2}} \in \mathbb{R}^n \) of (1).

In the \( k \)-th iteration of this proposed algorithm, we have \( p_{k+1} = p_k - \delta > 0 \). So, Algorithm 3.1 is well defined. In addition, we have the following remark.

**Remark 1.**

i) As the objective function of (11) is non-convex when \( 0 < p < 1 \), the solution \( u^{k+1} \) obtained by the L-BFGS algorithm in Step 2 of Algorithm 3.1 may not be a global solution of (11). Instead, \( u^{k+1} \) may be a Karush-Kuhn-Tucker point of (11). Therefore, we could not prove \( x^* \) obtained by Algorithm 3.1 is a global solution of (1). However, our numerical results in Section 4 show Algorithm 3.1 is effective for sparse signal reconstruction problems.

ii) The critical value \( \gamma \) in Theorem 3.5 is important. If the value of \( \gamma \) is known, we only need to implement Algorithm 3.1 for some fixed value of \( p \) with \( p < \gamma \).

A similar problem is discussed in [32] for classical compressive sensing model.

4. Numerical illustrations. In this section, we use some typical examples to show the effectiveness of the proposed algorithm for sparse signal reconstruction problems. We implement the proposed algorithm in MATLAB Version R2014A by using a laptop with 2.10-GHz processor and 3.0-GB of memory.

**Example 4.1.** Sparse signal reconstruction. In this example, we randomly generate a coefficient matrix \( A \in \mathbb{R}^{m \times n} \), where \( m = 512, n = 2048 \). Then we randomly generate a sparse signal \( x_0 \in \{-1, 0, 1\}^n \), where the non-zero number of \( x_0 \) is 80. By adding random Gaussian noise with mean 0 and standard deviation \( \sigma = 0.1 \) (generated by Matlab code: \( noise = \sigma \ast \text{randn}(n, 1) \)), the observed signal \( b \) is generated. That is \( b = Ax_0 + noise \). Given the observed signal \( b \), we will reconstruct
the original signal \( x_0 \) by Algorithm 3.1. During this procedure, we set \( p = 0.9 \) and \( \delta = 0.1 \). For each value of \( p \), we set \( \lambda = 1 \). Some recovery signals with different values of \( p \) are shown in Fig. 1.

\[
\begin{align*}
\text{original signal: } m=512, n=2048, \text{nnz}=80 \\
\text{reconstructed signal: } p=0.8, \text{nnz}=384, \text{res}=3.2563 \\
\text{reconstructed signal: } p=0.6, \text{nnz}=327, \text{res}=2.7593 \\
\text{reconstructed signal: } p=0.4, \text{nnz}=186, \text{res}=1.6649 \\
\text{reconstructed signal: } p=0.2, \text{nnz}=80, \text{res}=1.0814
\end{align*}
\]

\textbf{Figure 1.} The original signal and the reconstructed signals with \( p = 0.8, 0.6, 0.4, 0.2 \), respectively.

In Fig. 1, suppose \( x^*_p \) is the obtained solution by Algorithm 3.1 with \( p \), we use \( \text{nnz} \) to denote the number of non-zero element of \( x^*_p \), \( \text{res} \) to denote \( \| x^*_p - x_0 \|_2 \). From Fig. 1, we can see that with the decreasing of the value of \( p \), the recovery signal becomes more and more clear. In fact, the \( \text{nnz} \) and \( \text{res} \) all become more and more small with the decreasing of the value of \( p \), and when \( p = 0.2 \), the \( \text{nnz} \) is 80, which is exactly the number of the non-zero element of the original signal \( x_0 \).

\textbf{Example 4.2.} Restoring an image. In this example, we choose a testing image. First, the testing image is transformed to the grey image. Next, by using a random matrix, the grey image is transformed to an observed vector \( b \). After adding Gaussian noise with mean 0 and standard deviation \( \sigma = 0.1 \) to \( b \), we aim to recover the original grey image by Algorithm 3.1. The original and obtained images with different values of \( p \) are shown in Fig. 2.

From Fig. 2, we observe that the image can be recovered successfully with different values of \( p \). With the decreasing of the value of \( p \), the recovered image looks more and more clear except for \( p = 0.2 \). From the signal-to-noise ratio (SNR), we also see that the SNR with \( p = 0.3 \) is the highest among them, and it increases with the decreasing of value of \( p \) until \( p = 0.3 \). This example shows that our proposed method is effective. However, it is worth mentioning that if we continue to reduce the value of \( p \), the recovered image becomes bad, which can be seen from the restored image with \( p = 0.2 \). Empirically, we think that some parameter
selections in Algorithm 3.1 may lead to this phenomena, which deserves further investigation in future.

**Example 4.3.** Comparison with different algorithms on large size problems. In order to obtain a solution of (3) by using our method, the size of the original problem must be doubled. How will it affect the performance of Algorithm 3.1? We now show the difference between Algorithm 3.1 and other existing algorithms to solve the $l_p$ problem (3).

As mentioned in [40], IRLS is failure because it needs the inverse of the coefficient matrix and $l_1$ will result in a very large cpu time because it needs to call an $l_1$ procedure repeatedly. So, we only compare the performance of Algorithm 3.1 and the algorithm ITM [34] in this section.

We use $m$ and $n$ to denote the row number and the column number of the coefficient matrix $A$, and let $n = 10^3, 2 \times 10^3, 5 \times 10^3, 10^4, 2 \times 10^4, 5 \times 10^4, 10^5$, in our experiments respectively. Then $A$ is randomly generated with about $3n$ nonzero elements, the original signal $x_0 \in \{-1, 0, 1\}^n$ is randomly generated with $n/8$ nonzero components, and the random noise is chosen as white Gaussian noise with standard deviation 001. We use Algorithm 3.1 and ITM to find a solution of the corresponding problems, respectively. In order to compare the speed of Algorithm 3.1 with ITM in simplicity, we only run Algorithm 3.1 with one given values of $p$ for each problem. The results are shown in Fig. 3.

From Fig. 3, we can see that Algorithm 3.1 can recover the large size signals successfully with advantages in cpu time and accuracy. In comparison with ITM, we
can see that the solutions obtained by Algorithm 3.1 have usually smaller sparsity and lower relative error than those by ITM with the similar costs of the cpu time. We can see from the third column in Fig. 3 that Algorithm 3.1 saves more cpu time than ITM with the increasing of the problem’s size. Taking the comprehensive performance into consideration, we think our method is more effective to solve the problem (3) and is promising for large size problems.

5. **Conclusions.** In this paper, we have developed an effective algorithm for the non-Lipschitz optimization problem (3) in general. As many existing methods to solve the non-convex problem (3) require expensive computational costs and this fact restricts their applications for large size engineering problems. In order to find a sparse solution of general $l_p$ optimization problem (3), we first transformed (3) into (8) with a nonnegative constraint, and established the equivalence between their solutions. Then, by taking into account the structure of the problem (8), we employed the L-BFGS method to design an efficient algorithm to solve the general non-convex minimization problem (8). Also the connection of the problems (3) and (1) is established based on the proposed algorithm. In the end, we implemented some numerical experiments to demonstrate the effectiveness of the proposed method. The numerical results show that by using the proposed method, an accurate solution of (3) can be obtained efficiently. In the future, we will try to consider the following things. One is to take into consideration both gradient
norm and the value of objective function as the new termination criterion for Algorithm 3.1, which is expected to improve its performance. Another is to prove the convergence of Algorithm 3.1 under some proper conditions.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China (NSFC 61602321, 11431002, 71572125).

REFERENCES

[1] R. H. Byrd, P. Lu and J. Nocedal, A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Stat. Comput.*, 16 (1995), 1190–1208.

[2] 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. Learning*, 3 (2010), 1–122.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[4] A. Cohen, W. Dahmen and R. DeVore, Compressed sensing and best -term approximation, *J. Amer. Math. Soc.*, 22 (2009), 211–231.

[5] R. Chartrand, Nonconvex compressed sensing and error correction, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2007), 889–892.

[6] A. Charkrabarti and F. Hirakawa, Effective separation of sparse and non-sparse image features for denoising, in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, (2008), 857–860.

[7] X. Chen, K. Ng, Michael and C. Zhang, Non-Lipschitz-Regularization and box constrained model for image restoration, *IEEE Trans. Image Processing*, 21 (2012), 4709–4721.

[8] E. J. Candès and J. Romberg, The code package l1-magic. Available from: http://statweb.stanford.edu/~candes/l1magic/.

[9] E. J. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pure Appl. Math.*, 59 (2006), 1207–1223.

[10] R. Chartrand and V. Staneva, Restricted isometry properties and nonconvex compressed sensing, *Inverse Problems*, 24 (2008), 035020, 14 pp.

[11] E. J. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory*, 51 (2005), 4203–4215.

[12] X. Chen and S. Xiang, Sparse solutions of linear complementarity problems, *Math. Program.*, 159 (2016), 539–556.

[13] R. Chartrand and W. Yin, Iteratively reweighted algorithms for compressive sensing, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2008), 3869–3872.

[14] X. Chen and W. Zhou, Convergence of Reweighted lq Minimization Algorithms and Unique Solution of Truncated lp Minimization, Tech. rep., Hong Kong Polytechnic University, 2010.

[15] D. L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory*, 52 (2006), 1289–1306.

[16] D. L. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inf. Theory*, 47 (2001), 2845–2862.

[17] M. Elad and A. M. Bruckstein, A generalized uncertainly principle and sparse representation in pairs of bases, *IEEE Trans. Inf. Theory*, 48 (2002), 2558–2567.

[18] G. Fung and O. Mangasarian, Equivalence of minimal l0— and lp—norm solutions of linear equalities, inequalities and linear programs for sufficiently small p, *J. Optim. Theory Appl.*, 151 (2011), 1–10.

[19] S. Foucart and M. Lai, Sparkest solutions of underdetermined linear systems via lq minimization for 0 < q < 1, *Applied and Computational Harmonic Analysis*, 26 (2009), 395–407.

[20] M. A. T. Figueiredo, R. D. Nowak and S. J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE J. Select. Top. Signal Process.*, 1 (2007), 585–597.

[21] G. Gasso, A. Rakotomamonjy and S. Canu, Recovering sparse signals with a certain family of nonconvex penalties and DC programming, *IEEE Trans. Signal Process.*, 57 (2009), 4686–4698.

[22] M. Hyder and K. Mahata, An approximate l0 norm minimization algorithm for compressed sensing, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2009), 3365–3368.
[23] E. T. Hale, W. Yin and Y. Zhang, A fixed-point continuation method for $\ell_1$-regularized minimization with applications to compressed sensing, CAAM Technical Report TR07-07, Rice University, Houston, TX, 2007.

[24] D. Krishnan and R. Fergus, Fast Image Deconvolution Using Hyper-Laplacian Priors, Neural Information Processing Systems., Cambridge, MA: MIT Press, 2009.

[25] K. Koh, S.-J. Kim and S. Boyd, The code package $\ell_1s$. Available from: http://www.standord.edu/ boyd/\ell_1s.

[26] Q. Lyu, Z. Lin, Y. She and C. Zhang, A comparison of typical $l_p$ minimization algorithms, Neurocomputing, 119 (2013), 413–424.

[27] D. C. Liu and J. Nocedal, On the limited memory method for large scale optimization, Mathematical Programming B, 45 (1989), 503–528.

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

[29] B. K. Natraajan, Sparse approximation to linear systems, SIAM J. Comput., 24 (1995), 227–234.

[30] P. Ochs, A. Dosovitskiy, T. Brox and T. Pock, An iterated $\ell_1$ algorithm for non-smooth non-convex optimization in computer vision, in Computer Vision and Pattern Recognition (CVPR), IEEE Conference, (2013), 1759–1766.

[31] J. K. Pant, W. S. Lu and A. Antoniou, New improved algorithms for compressive sensing based on $l_p$ norm, IEEE Trans. on Circuits and Systems-II: Express Briefs, 61 (2014), 198–202.

[32] J. Peng, S. Yue and H. Li, NP/CMP equivalence: A phenomenon hidden among sparsity models $l_0$ minimization and $l_p$ minimization for information processing, IEEE Trans. Inf. Theory, 61 (2015), 4028–4033.

[33] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.

[34] Y. She, Thresholding-based iterative selection procedures for model selection and shrinkage, Electron. J. Stat., 3 (2009), 384–415.

[35] Y. She, An iterative algorithm for fitting nonconvex penalized generalized linear models with grouped predictors, Comput. Statist. Data Anal., 9 (2012), 2976–2990.

[36] R. Saab, R. Chartrand and O. Yilmaz, Stable sparse approximations via nonconvex optimization, in IEEE International Conference on Acoustics, Speech and Signal Processing, (2008), 3885–3888.

[37] J. Wright, A. Yang, A. Ganesh, S. Sastry and Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Recogn. Anal. Mach. Intell., 31 (2009), 210–227.

[38] Y. J. Wang, G. L. Zhou, L. Caccetta and W. Q. Liu, An alternating direction algorithm for $\ell_1$ problems in compressive sensing, IEEE Trans. Signal Process., 59 (2011), 1895–1901.

[39] Y. Wang and Q. Ma, A fast subspace method for image deblurring, Appl. Math. Comput., 215 (2009), 2359–2377.

[40] Y. Wang, G. Zhou, X. Zhang, W. Liu and L. Caccetta, The non-convex sparse problem with nonnegative constraint for signal reconstruction, J. Optim. Theory Appl., 170 (2016), 1009–1025.

[41] A. Y. Yang, Z. Zhou, A. G. Balasubramanian, S. S. Sastry and Y. Ma, Fast-minimization algorithms for robust face recognition, IEEE Trans. Image Processing, 22 (2013), 3234–3246.

[42] F. Zou, H. Feng, H. Ling, C. Liu, L. Yan, P. Li and D. Li, Nonnegative sparse coding induced hashing for image copy detection, Neurocomputing 105 (2013), 81–89.

[43] J. Zeng, S. Lin, Y. Wang and Z. Xu, $L_{1/2}$ regularization: Convergence of iterative half thresholding algorithm, IEEE Trans. Signal Process., 62 (2014), 2317–2329.

[44] W. Zuo, D. Meng, L. Zhang, X. Feng and D. Zhang, A generalized iterated shrinkage algorithm for non-convex sparse coding, in IEEE International Conference on Computer Vision (ICCV), 2013.